Normal and Aberrant TCRγδ+ T Cells and T Cell Large Granular Lymphocyte Leukemia

Martine Joanna Kallemeijn

The research for this thesis was performed within the framework of the Erasmus MC Postgraduate School Molecular Medicine.

The studies described in this thesis were performed at the Laboratory for Medical Immunology, Department of Immunology, Erasmus MC, Rotterdam, the Netherlands.

The studies were financially supported by an unrestricted grant from Roche granted to dr. A.W. Langerak.

Publishing of this thesis was supported by Roche Diagnostics Deutschland GmbH.

ISBN:	978-94-91811-19-7
Cover and invitation photo:	Willem Kallemeijn
Cover, invitation and thesis lay-out design:	Martine Kallemeijn
Thesis lay-out:	Daniëlle Korpershoek
Printing:	Haveka BV, Hendrik-Ido-Ambacht

Normal and Aberrant TCRγδ+ T Cells and T Cell Large Granular Lymphocyte Leukemia

Normale en afwijkende TCRγδ+ T cellen en T cel large granular lymfocyt leukemie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 20 juni 2018 om 09.30 uur

Martine Joanna Kallemeijn

geboren te Vlissingen

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotoren
Prof.dr. J.J.M. van Dongen

Overige leden
Prof.dr. A.M.H. Boots
Prof.dr. E.F. Eldering
Dr. J.N. Samsom

Copromotor
Dr. A.W. Langerak

Luctor ...

CONTENTS

PROLOGUE

Approximately 3.8 billion years ago it is thought that life has started on Earth, of which the last 450 million years sharks dominated the oceans (reviewed by Sadava *et al.* 2007) [1]. Together with the evolution of other jawed vertebrates the adaptive immune system developed. All four T cell receptor (TCR) genes were identified in *Raja erinacea*, the small skate (Fig. 1a). These TCR loci were also traced back in early vertebrate phylogeny [2]. The complete origin of the adaptive immune system receptors lies in NAR-TCRs (new antigen receptor-TCRs) as identified from studies on nurse sharks *Ginglymostoma cirratum* (Fig. 1b). All currently known Immunoglobulin (Ig) and TCRs evolved from this single NAR-TCR-receptor type (Fig. 2) [3], which showed usage and thus support by the current T cell receptor delta variable region (TRDV) exons in all kinds of receptors (Fig. 3). The largest proportion of the shark immune system consists of cells expressing delta receptors (>50%, [3]), as also identified in birds (±25% of all lymphocytes, [4,5]), cattle (75%, [6]), sheep (30-60%, [7]), alpacas and llamas (16%, [8]), mice (5% of all T cells, [9]) and ultimately also in humans (1-5% of all circulating T cells, [10-13]).

This means that T cells bearing the γδ-TCR are highly evolutionary conserved, and that they are possibly the ancestors of all cells from our adaptive immune system. Not only the close relation with the NAR-TCRs suggests this, but also their combinatorial role in coupling innate with adaptive immunity, by sharing functions of both immune systems [12,13]. TCRγδ+ T cells are highly polyfunctional and exhibit important functions and features: from expression of a unique and highly evolutionary conserved TCR, to therefore unique antigen specificities (by both being dependent and independent from the unique TCR), to the broad and complex cellular interactions.

As also illustrated by Pierre Vantourout & Adrian Hayday in 2013 (Fig. 4), TCRγδ+ T cells have pleiotropic functions by producing many types of cytokines (which could be Th1/2/17 related), prime other T cells, regulate the epithelium, help B cells in producing the right antibodies, help dendritic cells to skew their maturation in the right way, and they can actively kill infected and stressed cells. Infected cells include infections with bacteria, especially *Mycobacterium tuberculosis*: specific TCRγδ+ T cells bearing the Vγ9/Vδ2 receptor recognize phosphoantigens which come from bacteria [14,15]. But also fungi, parasites, such as *Plasmodium spp.* [16], viruses; for instance, Herpesviruses such as Cytomegalovirus [17] and Epstein-Barr virus [18] can be recognized by TCRγδ+ T cells. During the last Ebola outbreak, investigators even found a protective role for effector Vδ2 cells in Ebola survivors [19]; therefore, the picture and arrow with "pathogens" should be added to the polyfunctionality of TCRγδ+ T cells (Fig. 4). Stressed cells could be for instance infected cells, dysregulated cells or transformed cells such as (ear-

ly) cancer cells, which could be readily killed by TCRγδ+ T cells (reviewed by Fleming *et al.* 2017) [20]: a novel therapeutic option in cancer treatment.

By using all these properties at the same time, makes TCRγδ+ T cells very – even though the importance of their existence is often denied – promising when it comes to disease pathology and therapy.

Hopefully, with this introduction and booklet, the Reader could be convinced about the importance of the humble, great and versatile TCRγδ+ T cells.

PROLOOG

Ongeveer 3.8 miljard jaar geleden startte het leven op aarde, waarvan 450 miljoen jaar de haaien de oceanen domineren (beschreven in Sadava *et al*. 2007) [1]. Samen met de evolutie van gewervelde dieren ontwikkelde ook het verworven immuunsysteem. Allevier de T cel receptor (TCR) genen werden geïdentificeerd in *Raja erinacea*, de kleine rog (Fig. 1a). Deze TCR loci werden tevens teruggevonden in de fylogenie van vroege gewervelde dieren [2]. De complete oorsprong van de receptoren in het verworven immuunsysteem ligt in die van de zogenaamde NAR-TCRs (nieuwe antigeen receptor-TCRs) zoals geïdentificeerd in studies met de verpleegsterhaai *Ginglymostoma cirratum* (Fig. 1b). Alle momenteel bekende Immuunglobuline (Ig) en TCRs evolueerden uit deze enkele NAR-TCR-receptor type (Fig. 2) [3]. Deze liet gebruik en steun vanuit de nu bekende T cel receptor delta variabele exonen zien in allerlei soorten receptoren (Fig.3). Het grootste deel van het haaien immuunsysteem bestaat uit cellen met een delta receptor (>50%, [3]), wat later ook is gevonden in vogels (±25% van alle lymfocyten, [4,5]), vee (75%, [6]), schapen (30-60%,[7]), alpaca's en lama's (16%, [8]), muizen (5% van alle T cellen, [9]), en tenslotte ook in mensen (1-5% van circulerende T cellen, [10-13]).

Dit betekent tevens dat T cellen met een γδ-TCR evolutionair zeer geconserveerd zijn, en dat zij mogelijk de voorouders van ons volledige verworven immuunsysteem zijn. Niet alleen de hoge correlatie met NAR-TCRs suggereert dit, maar ook hun rol in het koppelen van de aangeboren met verworven immuniteit middels het bezitten van functies van beiden [12,13]. TCRγδ+ T cellen zijn zeer polyfunctioneel en bezitten belangrijke eigenschappen: van de expressie van de unieke en geconserveerde receptor, tot unieke antigeen specificiteiten (door zowel afhankelijk en onafhankelijk te zijn van de unieke TCR), tot het brede scala aan complexe cellulaire interacties.

Zoals ook geïllustreerd door Pierre Vantourout & Adrian Hayday in 2013 (Fig. 4), hebben TCRγδ+ T cellen pleiotrope functies door de productie van verschillende cytokines (welke van Th1/2/17 origine kunnen zijn), het activeren van andere T cellen, epitheel regulatie, B cel hulp voor de juiste antilichaamproductie, dendritische cellen in de juiste richting sturen, en ze kunnen actief geïnfecteerde en gestresste cellen doodmaken. Cellen kunnen bijvoorbeeld geïnfecteerd zijn met bacteriën, vooral *Mycobacterium tuberculosis* geeft een specifieke reactie van Vγ9/Vδ2 cellen, omdat ze fosfo-antigenen afkomstig van bacteriën kunnen herkennen [14,15]. Maar ook schimmels, parasieten, zoals *Plasmodium spp.* [16], virussen; bijvoorbeeld Herpesviridae zoals Cytomegalovirus [17], en Epstein-Barr virus [18] worden door TCRγδ+ T cellen herkend. Tijdens de recente Ebola uitbraak vonden onderzoekers zelfs een beschermende rol van effector Vδ2 cellen in Ebola overlevenden [19]; daarom zou deze figuur kunnen worden aangepast met nog een extra pijl voor "pathogenen" (Fig. 4). Gestresste cellen zijn bijvoorbeeld

geïnfecteerde, ontregelde of getransformeerde (vroege) kankercellen [20], welke direct kunnen worden aangevallen door TCRγδ+ T cellen: een nieuwe optie voor kankertherapie.

Door al deze eigenschappen op hetzelfde moment te kunnen gebruiken, maakt TCRγδ+ T cellen zeer – alhoewel het belang van hun bestaan vaak wordt genegeerd – veelbelovend wanneer het om ziekte pathologie en therapie gaat.

Hopelijk, met deze introductie en dit boekje, kan de Lezer worden overtuigd van het belang van de bescheiden, grootse en veelzijdige TCRγδ+ T cellen.

REFERENCES

1. Sadava, D.E., Heller, H.C., Orians, G.H., Purves, W.K. & Hillis, D.M. Life: the science of biology. Macmillan (2007).
2. Rast, J.P., Anderson, M., Strong, S.J., Luer, C., Litman, R.T. & Litman, G.W. α, β, γ and δ T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* **6**, 1-11 (1997).
3. Criscitiello, M.F., Saltis, M. & Flajnik, M.F. An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc. Natl. Acad. Sci. U S A* **103(13)**, 5036-5041 (2006).
4. Chen, C-L. *et al.* Characterization of an avian (*Gallus gallus domesticus*) T cell receptor α/δ gene locus. *J. Immunol.* **163**, 3858-3866 (1999).
5. Pieper, J., Methner, U. & Berndt, A. Heterogeneity of avian γδ T cells. *Vet. Immunol. Immunopathol.* **124(3-4)**, 241-252 (2008).
6. Hein, W.R. & Mackay, C.R. Prominence of γδ T cells in the ruminant immune system. *Immunology Today* **12(1)**, 30-34 (1991).
7. Mackay, C.R., Beya, M.F. & Matzinger, B.P. Gamma/delta T cells express a unique surface molecule appearing late during thymic development. *Eur. J. Immunol.* **19**, 1477-1483 (1989).
8. Davis, W.C. *et al.* Flow cytometric analysis of an immunodeficiency disorder affecting juvenile llamas. *Vet. Immunol. Immunopathol.* **74(1-2)**, 103-120 (2000).
9. Maki, K. *et al.* Interleukin 7 receptor-deficient mice lack γδ T cells. *Proc. Natl. Acad. Sci. U S A* **93**, 7172-7177 (1996).
10. Hayday, A.C. *et al.* Structure, organization and somatic rearrangement of T cell gamma genes. *Cell* **40**, 259-269 (1985).
11. Lefranc, M-P. *et al.* Molecular mapping of the human T cell receptor gamma (TRG) genes and linkage of the variable and constant regions. *Eur. J. Immunol.* **19**, 989-994 (1989).
12. Hayday, A.C. γδ cells: a right time and a right place for a conserved third way of protection. Annu. Rev. Immunol. **18**, 975-1026 (2000).
13. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. Nat. Rev. Immunol. **13(2)**, 88-100 (2013).

14. Morita, C.T. et al. Direct presentation of nonpeptide prenyl pyrophosphate antigens to human γδ T cells. Immunity **3**, 495-507 (1995).
15. Morita, C.T., Jin, C., Sarikonda, G. & Wang, H. Nonpeptide antigens, presentation mechanisms, and immunological memory of human Vγ2/Vδ2 T cells: discriminating friend from foe through the recognition of prenyl pyrophosphate antigens. Immunol. Rev. **215**, 59-76 (2007).
16. Behr, C. et al. Plasmodium falciparum stimuli for human gammadelta T cells are related to phosphorylated antigens of mycobacteria. Infect. Immun. **64(8)**, 2892 (1996).
17. Déchanet, J. et al. Implication of gammadelta T cells in the human response to cytomegalovirus. J. Clin. Invest. **103(10)**, 1437-1449 (1999).
18. Djaoud, Z. et al. Two alternate strategies for innate immunity to Epstein-Barr virus: one using NK cells and the other NK cells and γδ T cells. J. Exp. Med. **214(6)**, 1827-1841 (2017).
19. Cimini, E. et al. Different features of Vδ2 T and NK cells in fatal and non-fatal human Ebola infections. PLoS Negl. Trop. Dis. **11(5)**, e0005645 (2017).
20. Fleming, C., Morrissey, S., Cai, Y. & Yan, J. γδ T cells: unexpected regulators of cancer development and progression. Trends Cancer **3(8)**, 561-570 (2017).
21. Criscitiello, M.F. Chapter 12: Shark T-Cell Receptors in Immunobiology of the shark. CRC Press. ISBN 9781466595743, pp. 237-246 (2014).

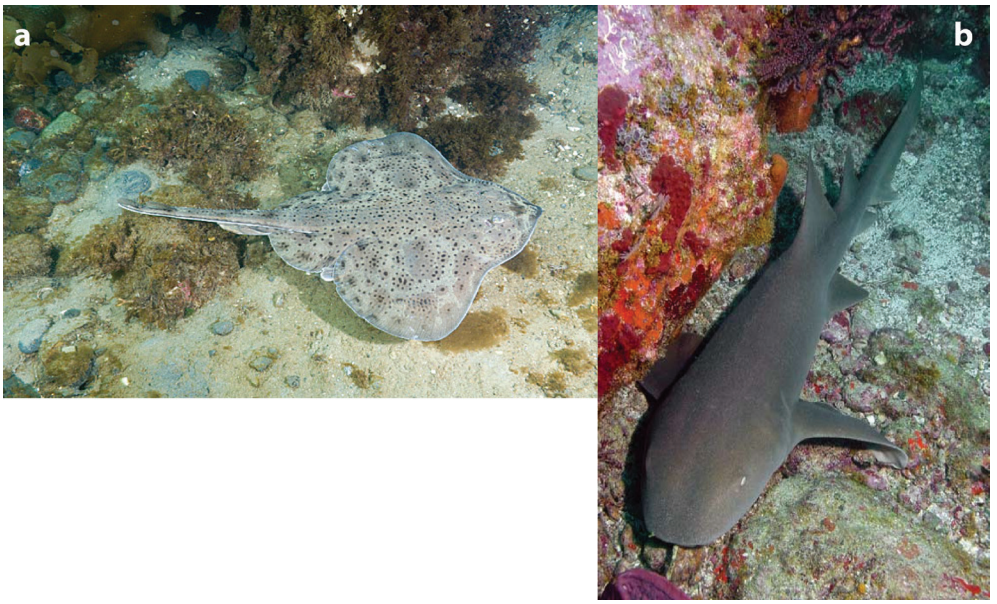


Figure 1. ***Raja erinacea*** **(a) and** ***Ginglymostoma cirratum*** **(b).**
The little skate (**a**) and the nurse shark (**b**). Images adapted from www.elasmodriver.com.

Figuur 1. ***Raja erinacea*** **(a) en** ***Ginglymostoma cirratum*** **(b).**
De kleine rog (**a**) en de verpleegsterhaai (**b**). Beelden afkomstig van www.elasmodriver.com.

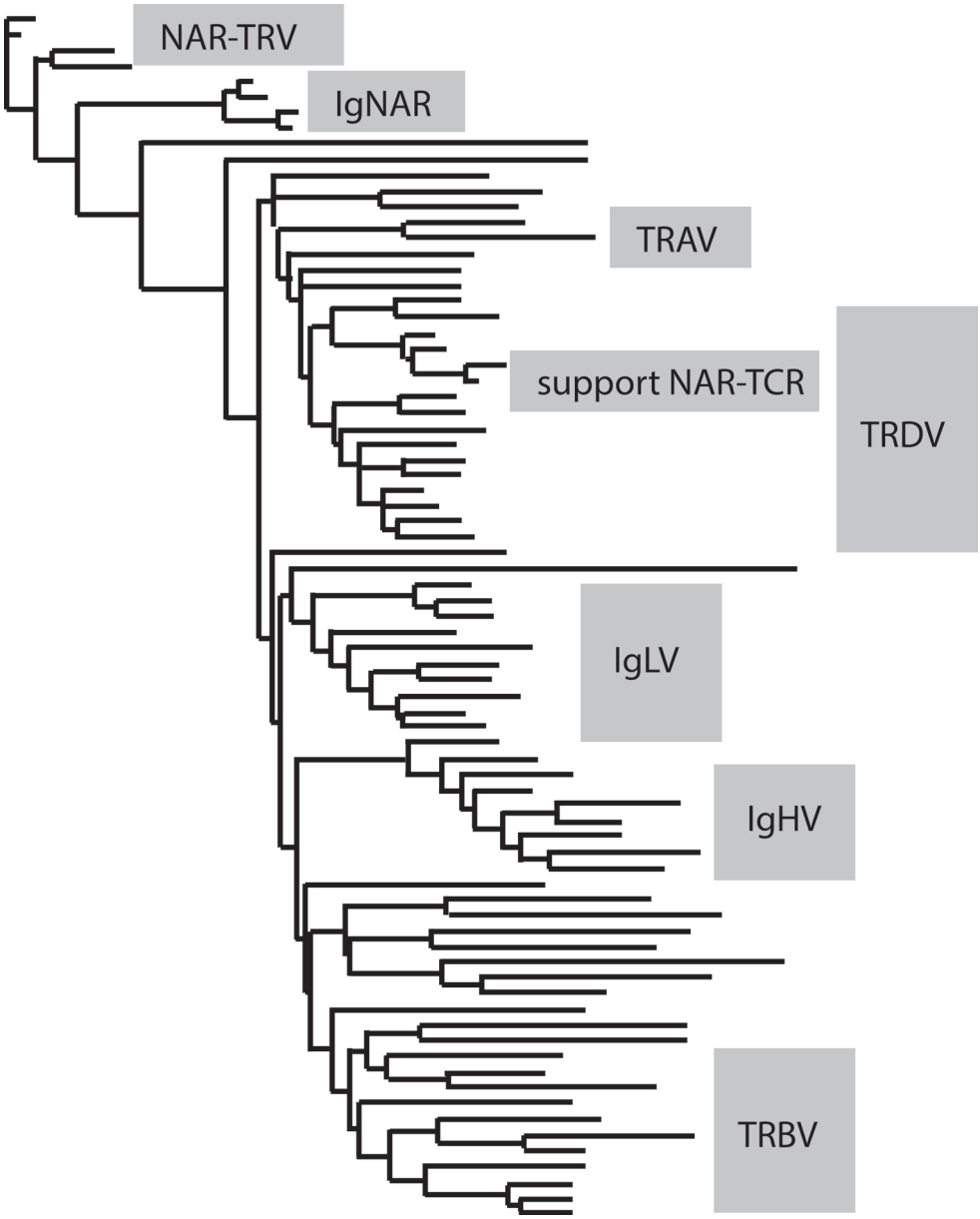


Figure 2. Adaptive immunity receptor phylogeny.
Phylogenetic tree of the Ig V domains showing early development of the NAR-TCRV which is supported by the TRDV locus. Figure adapted from Criscitiello *et al*. 2006.

Figuur 2. Verworven immuniteit receptor verwantschap.
Fylogenetische stamboom van de Ig V domeinen, welke de vroege ontwikkeling van de NAR-TCRV illustreert. Deze wordt gesteund door het TRDV locus. Figuur afkomstig van Criscitiello *et al*. 2006.

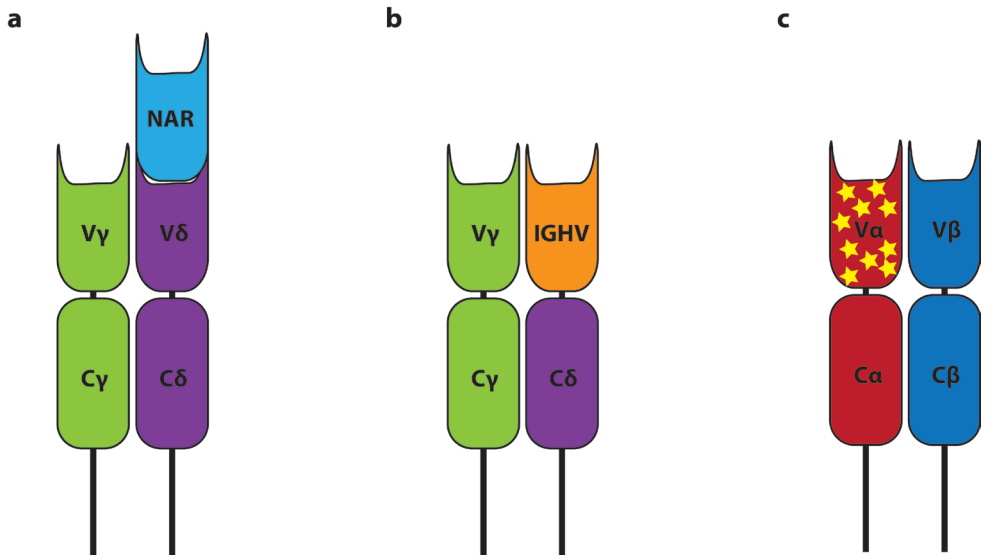


Figure 3. Receptor diversity and TRDV usage in the shark immune system.
Three different immune tricks in the shark immune system with NAR-TCR extensions of the TRD chain (**a**), trans-rearrangements combining TRG with IGHV (**b**) and somatic hypermutation in T cells (**c**). Figure adapted from Criscitiello 2014 [21].

Figuur 3. Receptor diversiteit en TRDV gebruik in het immuun systeem van haaien.
Drie verschillende manoeuvres van het immuun systeem in de haai met NAR-TCR verlengingen van de TRD keten (**a**), trans-herschikkingen met TRG en IGHV combinaties (**b**) en somatische hypermutatie in T cellen (**c**). Figuur afkomstig van Criscitiello 2014 [21].

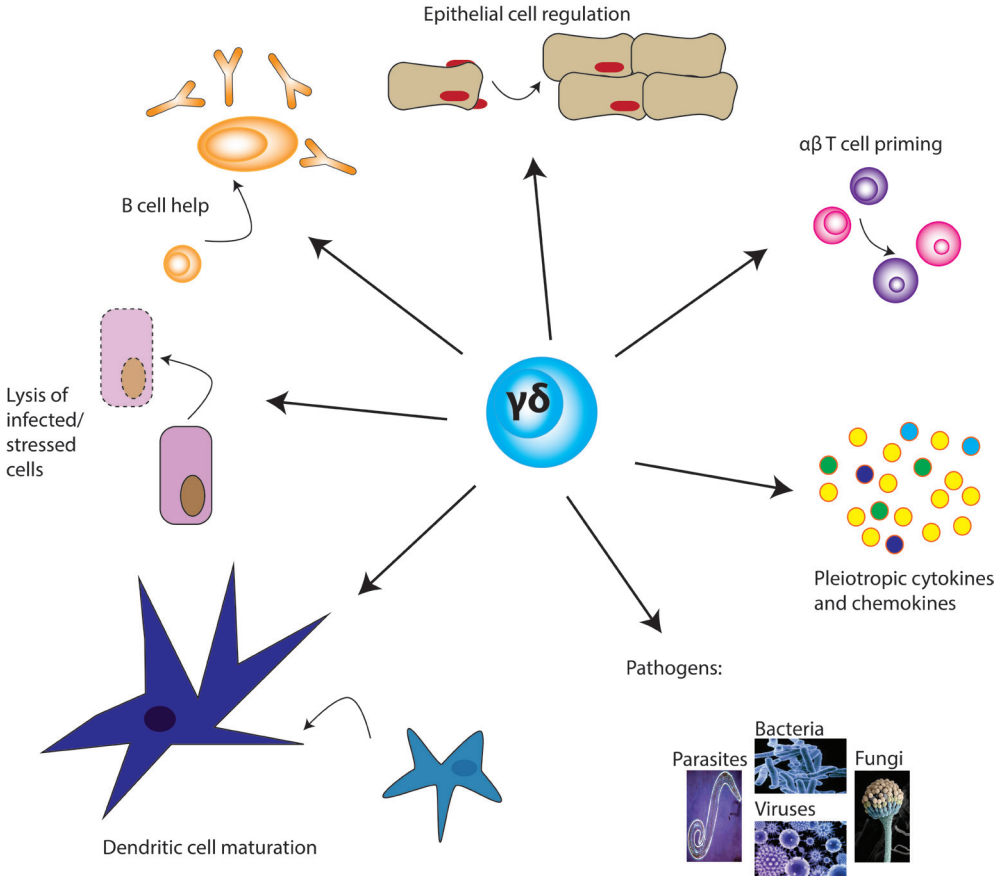


Figure 4. TCRγδ+ T cells' pleiotropic functions.
TCRγδ+ T cells are able to prime TCRαβ+ T cells, dendritic cells and B cells, regulate the epithelium, lyse infected and stressed cells, produce various cytokines and chemokines, and respond to pathogens (bacteria, fungi, parasites, viruses). Figure adapted from Vantourout & Hayday 2013.

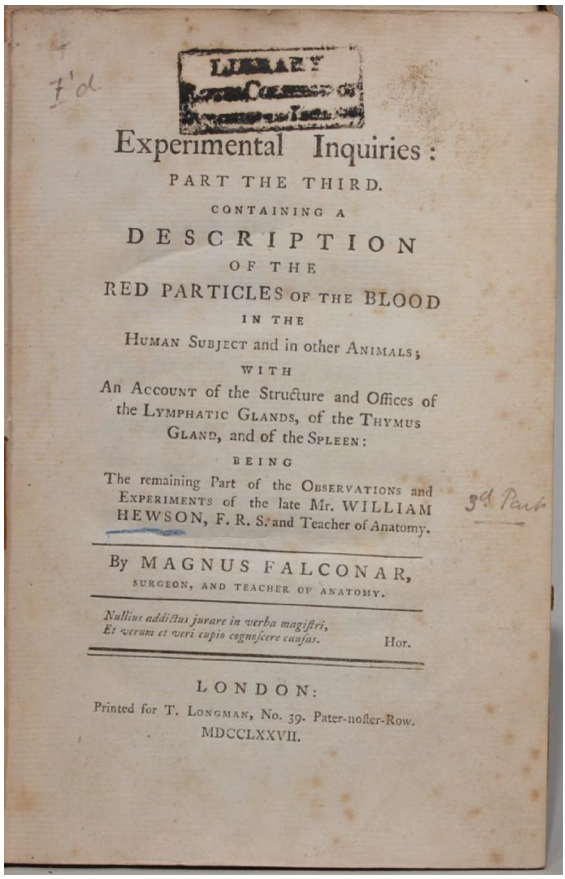
Figuur 4. De veelzijdige functies van TCRγδ+ T cellen.
TCRγδ+ T cellen kunnen TCRαβ+ T cellen, dendritische cellen en B cellen activeren, epitheelcellen reguleren, geïnfecteerde en gestresste cellen lyseren, vele verschillende cytokines en chemokines produceren en op pathogenen reageren (bacteriën, schimmels, parasieten, virussen). Figuur afkomstig van Vantourout & Hayday 2013.

Chapter 1

General Introduction

A HISTORICAL PERSPECTIVE OF T CELLS

The thymus has already been recognized and described by the ancient Greeks, who called it "θυμός", which means heart, or life, based on the anatomic position of the organ in mammals (summarized in [1,2]). Centuries later in 1769 William Hewson wrote about a unified lymphatic system, to which nodes, the spleen and thymus were attached. In 1777 his investigation on the thymus was published by his friend Marcus Falconer, in which it was described that the organ was filled with particles, in much the same way as blood was [3,4] (Fig. 1). It was stated that the identified particles were necessary only in early stages of life, and that therefore only in that period the thymus would



Experimental Inquiries:

PART THE THIRD.

CONTAINING A

DESCRIPTION

OF THE

RED PARTICLES OF THE BLOOD

IN THE

HUMAN SUBJECT and in other ANIMALS;

WITH

An ACCOUNT of the Structure and Offices of the LYMPHATIC GLANDS, of the THYMUS GLAND, and of the SPLEEN:

BEING

The remaining Part of the OBSERVATIONS and EXPERIMENTS of the late Mr. WILLIAM HEWSON, F. R. S. and Teacher of Anatomy.

By MAGNUS FALCONAR,

SURGEON, AND TEACHER OF ANATOMY.

Nullius addictus jurare in verba magistri,
Et verum et veri cupio cognoscere causas. Hor.

LONDON:

Printed for T. LONGMAN, No. 39. Pater-noster-Row.

MDCCLXXVII.

Figure 1. Original first page of the chapter containing the remaining part of the observations and experiments of the late Mr. William Hewson.
Original Chapter page from William Hewson's work on the lymphatic glands, thymus and spleen, continued, finished and published by Magnus Falconer in 1777 (Hewson, 1777).

exist [5]. In the years after, many speculations have risen on the origin, development, importance and function of the thymus in general. In 1954 J. Vadász described in his article "*die Entstehung der Thymozyten in Gewebekulturen*" the origin of thymocytes from the epithelium of the organ; under a microscope he noticed the development of small and rapidly dividing cells from the reticulum of the thymus [6]. A few years later, in 1961, Jacques Miller discovered the involvement of the thymus in cellular immunity through formation of lymphocytes, when he performed thymectomy in neonatal mice and observed poorly developed lymphoid organs, impaired immune responses and high susceptibility to infections [7]. Furthermore, he found that nude mice did not reject skin grafts, while immunocompetent mice did [2]. This implied that the thymus generates lymphocytes which have undergone a selection process, and upon maturation migrate to different parts of the body. The poorly developed lymphoid organs and impaired immune responses in thymectomized neonatal mice could be rescued by the transplantation of thymus grafts, again indicating that the particles or lymphocytes must originate from the thymus and/or the thymic epithelia [5,6]. In the years after, researchers could identify that thymocytes do not develop from the thymic epithelium itself, but the thymic epithelium nurtures and teaches stem cells originating from the bone marrow to become T cells [7]. In fact, the progenitors entering the human thymus still have the potential to differentiate into B lymphoid, myeloid and erythroid lineages [8]. Upon entering the thymus, the progenitors receive Notch signals leading to full commitment to the T cell lineage (reviewed in [9]).

A central event during T cell lineage commitment and further T cell development is the formation of the T cell receptor (TCR). The T cell receptor itself was discovered in 1983, along with the importance of antigen recognition via the TCR in the context of major histocompatibility complex (MHC) molecules [10]. A few years later also the genes encoding the T cell receptor were identified [11] (reviewed in [12-15]). First, the TCRαβ heterodimer (Fig. 2a) was identified when MHC-molecules were investigated in the context of T cells and vice versa. Later, also the T cell receptor γ (TRG) and T cell receptor δ (TRD) loci were mapped [16,17], which led to the discovery and recognition of a novel, alternative type of T cells, the TCRγδ+ T cells [18-20]. The TRD locus is located on chromosome 14 and is intertwined with the TRA locus. The locus is larger, although consisting of fewer genes than the TRG locus (Fig. 2b). The TRG locus is a small locus and located on chromosome 7. The genes group into several families (Fig. 2c).

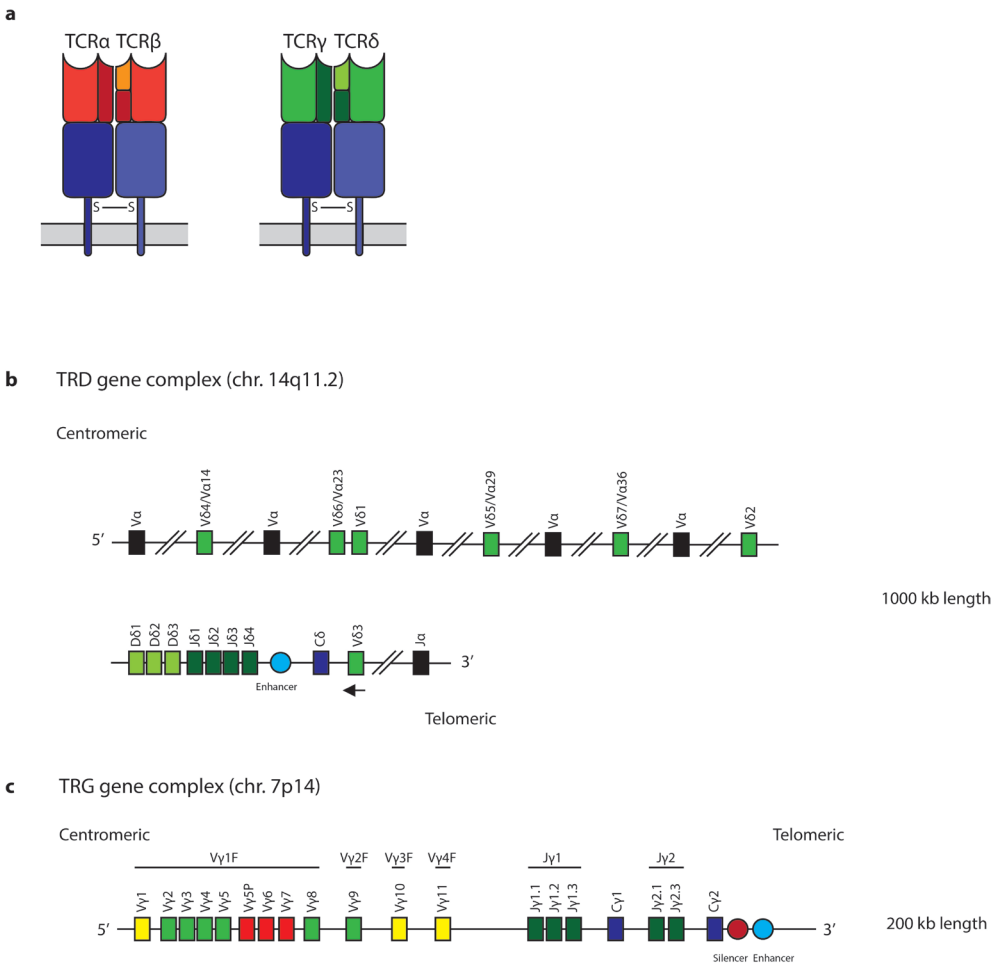


Figure 2. The T cell receptor heterodimer and TRD/TRG gene loci.
(**a**) T cell receptor as heterodimer, either consisting of a TCRα and TCRβ chain, or TCRγ and TCRδ chain. TCRβ and TCRδ chains are encoded after VDJ recombination, involving V, D and J genes, whilst TCRα and TCRγ chains only contain V and J gene encoded sequences. (**b**) TRD gene locus located on chromosome 14q11.2, intertwined with the TRA locus thus preventing nonfunctional αδ or γβ receptor combinations. All TRD genes are functional (green), although TRDV4-7 can also be used as TRAV genes (TRAV14, 29, 23 and 36 respectively). TRDV1-3 are most common, with TRDV3 reversely located in the locus. An enhancer is located in between TRDJ4 and TRDC. The locus is approx. 1000 kb in length. Black boxes indicate regions of TRAV and TRAJ genes. (**c**) TRG gene locus is located on chromosome 7p14. The locus is approx. 200 kb in length. The V-genes are divided into 4 families, with families 1 and 2 containing functional genes (green). TRGV1, 10 and 11 (yellow) are open reading frames (ORF), TRGV5P, 6 and 7(red) are pseudogenes. All J-genes are functional genes (dark green).

PRECURSOR T CELL DEVELOPMENT AND TCR RECOMBINATION

TCR recombination

Both TCRαβ+ and TCRγδ+ T cells develop in the thymus, undergoing consecutive stages of development characterized by expression of different combinations of markers (reports reviewed by [21]). During the developmental stages, from double-negative (DN) for CD4 and CD8 stages 1-4, to immature single-positive (ISP, CD4 SP), to double-positive (DP) and ultimately CD4 or CD8 single-positive (SP) they undergo several steps of maturation, involving TCR rearrangements and checkpoints, proliferation and selection. During stages DN3 – DN4 immature thymocytes undergo rearrangements of their Variable (V), Diversity (D) and Joining (J) genes (TRD, TRB), or only V and J genes (TRG, TRA), following the hierarchical order TRD > TRG > TRB > TRA [22,23] (Fig. 3). Since the TRD and TRA loci are intertwined, there will never be a (functional) TCRαδ receptor formed. V(D)J recombination is induced by the RAG1 and RAG2 proteins through the binding of these proteins to the recombination signal sequences (RSSs) [24]. The repertoire diversity is increased dramatically by the terminal deoxynucleotidyl transferase (TdT) enzyme [25]. It adds random N nucleotides at the V-D, D-J and V-J junctional sites [26,27]. This could potentially lead to an increase of the total receptor diversities from 10^6 [28] up to over 10^{13} for both TCRαβ+ and TCRγδ+ T cell receptors [29]. Only one functionally recombined TRB allele will pair with a TRA allele, and only one TRD allele with one TRG allele will be expressed, a phenomenon also referred to as allelic exclusion [30].

TCRαβ development

During human TCRαβ+ T cell development the rearranged β-chain is checked for functionality with the pre-Tα chain (β-selection, from DN2-stage). Upon formation of a TCR that is expressed on the cell surface, developing thymocytes are positively selected for the functionality of the receptor. For TCRαβ+ T cells up-regulation of first CD4 (ISP) and then also CD8 (DP) is also part of the development. During positive selection of the functional TCR, TCRαβ+ T cells are still CD4 and CD8 double-positive, but already have committed machineries to one or the other cell type [22,23,31] (Fig. 3). Positive selection is mediated through cortical thymic epithelial cells (cTECs), which provide thymocytes with functional TCRs signals for further proliferation, differentiation and maturation [32]. After positive selection, TCRαβ+ T cells become either CD4 or CD8 single-positive, and negative selection occurs with the help of medullary thymic epithelial cells (mTECs), which express the transcription factor autoimmune regulator (AIRE), which, in turn, induces expression of self-antigens [33]. Upon recognition and strong

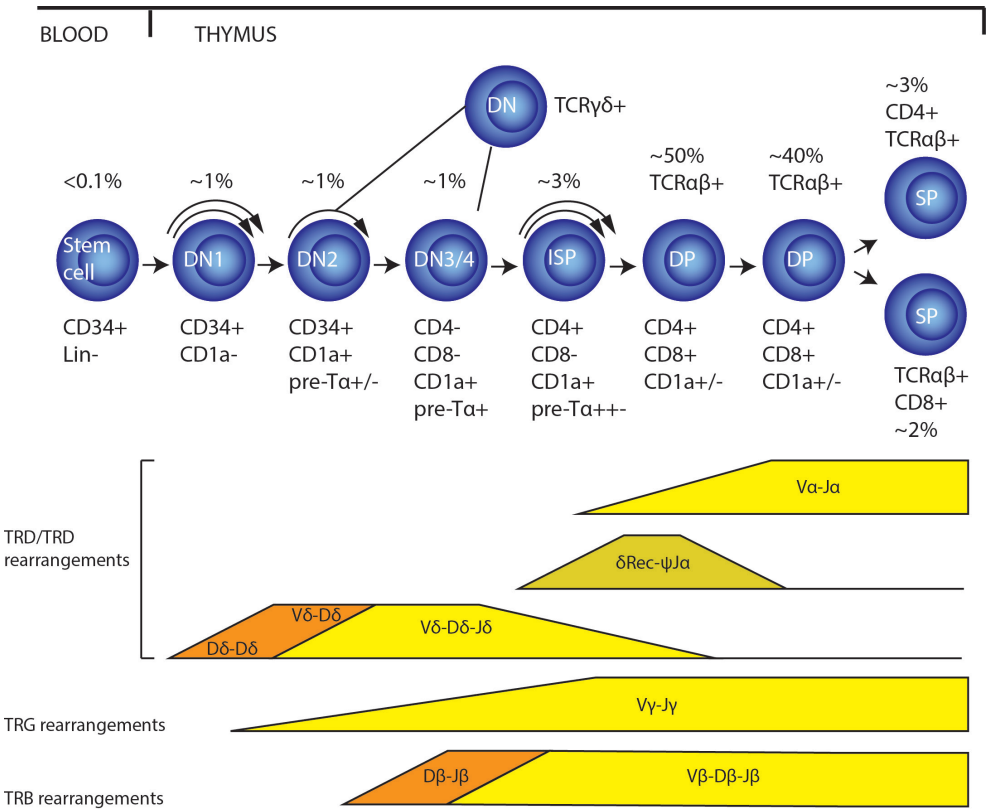


Figure 3. Thymic TCRαβ+ and TCRγδ+ development.
Summary of normal human precursor T cell development stages and T cell receptor (TCR) rearrangements. A CD34+ stem cell from the PB enters the thymus upon which the stem cell receives signals to commit to the T cell lineage. Each developmental stage is characterized by different marker expression and TCR rearrangements that occur in the order TRD > TRG > TRB > TRA. The figure also includes incomplete rearrangements. Figure adapted from Staal *et al.* 2007.

association with self-antigens, developing thymocytes will undergo apoptosis in order to maintain immune tolerance. After these processes, mature naive CD4 and CD8 single-positive TCRαβ+ T cells move to the periphery [31].

TCRγδ development

In case of TCRγδ+ T cell development processes such as pre-TCR (tentatively proposed to be called "pre-Tγ") signaling and positive and negative selection are not as evident as for TCRαβ+ T cell development. Commitment to either the TCRαβ+ or TCRγδ+ lineage appears to start early during development through Notch1 signaling: higher expression of Notch1 is in favor of TCRγδ+ T cell development (reviewed in [34]). Also,

TCR signaling strength determines αβ or γδ lineage commitment: high TCR signaling favors γδ, weak signaling favors αβ lineage commitment [35-37]. However, signaling is not only ligand-dependent, as ligand-independent signaling is also important for TCRγδ development, given the lack of a pre-Tγ chain. In mice, efficient pairing of TRG and TRD chains above a certain threshold also determines TCRγδ+ T cell commitment fate [38]. TCRαβ+ and TCRγδ+ T cells follow rather different paths of development and maturation with respect to positive and negative selection. The only proposed type of selection or checkpoint during TCRγδ+ T cell development is the TCR dimerization of the δ-chain [39]. Furthermore, it has been found that thymic selection contributes little to nothing in constraining TCRγδ+ T cell antigenic specificity. In fact, TCRγδ+ T cells do not need to encounter (self-)ligands in the thymus in order to mature. Therefore, most TCRγδ+ T cells that are found in the periphery are still naive [40]. Also in mice, thymic development affects more the effector function. Lack of positive selection and TCR ligand encounter leads to a large fraction of antigen-naive TCRγδ+ T cells, with robust interleukin-17 (IL-17) production. TCRγδ+ T cells that did have encountered TCR ligand in the thymic medulla rather produce interferon (IFN)-γ [41].

PERIPHERAL T CELL DEVELOPMENT

Peripheral TCRαβ+ T cells

Mature TCRαβ+ thymocytes leave the thymus to go to the periphery, into the circulation and/or secondary lymphoid organs. This process is regulated via the sphingosine-1-phosphate (S1P) signaling; mature thymocytes migrate to S1P due to their high S1P-receptor 1 (S1PR1) expression [42]. The S1P-S1PR1 axis in the periphery is also used for entering tissues [43]. In the periphery CD4+ and CD8+TCRαβ+ T cells respond to well-defined peptide antigens in the context of MHC-II and MHC-I molecules respectively, and typically need additional stimulation via co-stimulatory molecules and skewing via cytokines to become fully activated and functional [44]. After receiving these three activation signals the antigen-specific T cells start to proliferate, also called the *clonal expansion phase*. When the antigen is cleared, the T cells then undergo activation induced cell death (AICD) to restrict the inflammation, also called the *contraction phase*. A few antigen-specific T cells survive the contraction phase, and become long-lived memory T cells which either remain in the secondary lymphoid organs as central memory T cells, or they continue to circulate in the periphery as effector memory T cells. AICD occurs via the extrinsic apoptosis signaling pathway, which is Fas/FasL mediated [45,46]. The induction of the extrinsic apoptosis pathway could additionally further induce downstream the intrinsic pathway via up-regulation of Bim and down-regulation

of Bcl-2 [47,48], or via the activation of Bid through Caspase-8, which ultimately leads to mitochondrial outer membrane permeabilization, cytochrome-c release and formation of the apoptosome [49]. In case of CD8+ cytotoxic T lymphocytes (CTL) AICD can also be the result of auto- and paracrine killing through perforins and granzymes [50].

Peripheral TCRγδ+ T cells

As TCRγδ+ thymocytes do not undergo extensive positive and negative selection in the thymus, they egress faster from the thymus than their TCRαβ+ counterparts, and even without S1P/S1PR1 signaling [40].

Once in the periphery, TCRγδ+ T cells exert different unique functions and features when compared to TCRαβ+ T cells, which are more restricted when it comes to recognizing and responding to antigens [51]. As reviewed by Vantourout and Hayday in 2013 at least six unique features of TCRγδ+ T cells can be described, related to the type of antigens, antigenic recognition, antigenic responses, response kinetics, responses at anatomical sites and ontogeny [52]. TCRγδ+ T cells are non-MHC-restricted T cells, and therefore their (antigenic) specificity is rather unknown. Some molecules have been described as antigens, such as hydroxymethyl-but-2-enyl-pyrophosphate (HM-BPP), a non-mevalonate intermediate of the cholesterol synthesis pathway [53], isopentenyl pyrophosphate, a mevalonate intermediate of the canonical cholesterol synthesis pathway [54], other bisphosphonates such as aminobisphosphonates (nBPs) [55], or butyrophilins (BTNs) and other lipids that could be presented via CD1 molecules [56,57]. These antigens are often from specific pathogens, such as bacteria [58], but also eukaryotic organisms such as *Plasmodium spp.* [59] and mammals, or even cancer cells due to high cellular turnover [60-62].

The TCRγδ+ T cell response to the above mentioned antigens is in a CTL and NK cell like manner, with a fast cytokine release. Different cytokine-producing TCRγδ+ T cells have differential epigenetic and transcriptional signatures, although the cells show high plasticity when they encounter specific antigens [63]. The type of antigens encountered, but also the TCR signaling strength determines the cytokine production. High TCR signaling strength in TCRγδ+ T cells leads to IFN-γ producers, while weak or attenuated TCR signaling leads to IL-17 producing cells [64]. The ability to readily produce significant amounts of IL-17 during early infections in the acute response additionally forms a bridge to the adaptive immune response by recruiting CD4+TCRαβ+ T helper cells and B cells to the site of infection, and increases the crosstalk between these immune cells [40].

In order to respond, TCRγδ+ T cells do not necessarily require clonal expansion, and are therefore sometimes referred to as sensors of immune- or tissue dysregulation [52]. In case of such dysregulation, the so-called “epimmunome” can be observed: a

combination of cell surface molecules that are up-regulated in response to dysregulation [65]. Upon activation TCRγδ+ T cells have a broad spectrum of cytokines which they can produce, also in order to communicate with other immune or neighboring cells, to form a bridge between the innate, afferent phase, and the adaptive, efferent phase of the immune response. The type of response that TCRγδ+ T cells can exert also largely depends on the anatomical site and the type of tissue in which they are located. Both the recognition of antigens, the response to antigen and the anatomical location largely depends on the type of receptor the TCRγδ+ T cell is expressing. In mice, the occupation of different tissues is described to occur in consecutive waves [66]: Vγ5+ cells home to the skin as dendritic epidermal T cells (DETC), Vγ6+ colonize the uterus, tongue, lungs and peritoneum [67], Vγ4+ colonize the peritoneum, and Vγ7+ cells develop extrathymically, based on epithelial support from the gut [68]. These cells (except the DETCs) are IL-17 producers.

In human TCRγδ+ T cell development no Vγ-waves have been described; in contrast, in humans TCRγδ+ T cell definition is based on Vδ-usage [69]. TCRγδ+ T cells migrate mostly to epithelial tissues such as the gut, skin, urogenital tract and respiratory tract (Fig. 4). These TCRγδ+ T cells are largely Vδ1+ IL-17-producing cells in humans. A large fraction remains in the circulation, forming the common Vγ9/Vδ2 IFN-γ-producing population [37,70]. TCRγδ+ T cells expressing Vδ-genes 3-6 are extremely rare, especially in the circulation, although Vδ3+ cells can be found in the gut and synovial tissue [71-73], or clonally expanded in the peripheral blood (PB) in the context of a CD4+ T cell deficiency [74]. Vδ1+ cells can also be found in high titers in the human fetal and neonatal cord blood (CB), which readily produce IL-17 in order to provide protection for the neonate [75]. This protection by Vδ1+ cells in neonates is also reflected in the higher numbers of PB Vδ1+ cells compared to Vγ9/Vδ2 cells [76] (Fig. 4). TCRγδ+ T cells, especially Vγ9/Vδ2 cells, are highly plastic and heterogeneous, not only in case of infections or tumors, but also in homeostasis in healthy individuals. In humans, approximately six Vδ2 profiles could be defined based on expression of CD27, CD28 and CD16, alongside chemokine receptors and perforins and granzymes [77]. Among these profiles different relative distributions of four Vδ2+ subsets were identified: CD28+CD27+CD16- (γδ28+), CD28-CD27+CD16- (γδ28-) highly proliferative cells, and CD28-CD27-CD16- (γδ16-) and CD28-CD27-CD16+ (γδ16+) cells with co-expression of CD56, CD57, Perforins, Granzyme B, indicative of active large granular lymphocyte (LGL) functions of these cells.

Large granular lymphocytes

Upon (antigenic) stimulation CD8+TCRαβ+ CTL and TCRγδ+ T cells become activated, as well as a very small fraction of CD4+TCRαβ+ T-LGL cells [78]. This leads to the increased cellular activity of the cells and the increased production of granules to induce

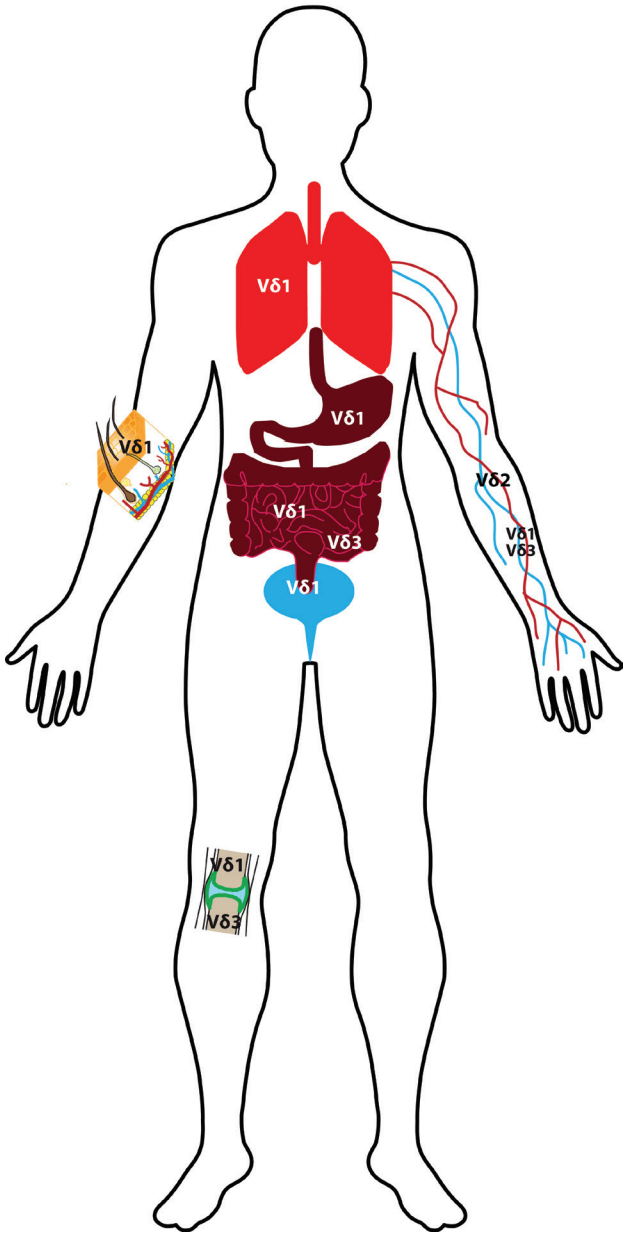


Figure 4. Anatomical distribution of TCRγδ+ T cells.
TCRγδ+ T cells are most abundant in epithelial tissues as intraepithelial lymphocytes. These epithelial tissues include the respiratory, gastrointestinal and urogenital tracts, but also the skin. TCRγδ+ T cells occupying the epithelial tissues mainly contain the Vδ1-chain. Vδ3 cells can also be found in the gut and synovial tissue. On the contrary, circulating peripheral blood TCRγδ+ T cells mostly possess Vγ9/Vδ2 receptors, based on a different antigenic exposure, albeit that Vδ1 and Vδ3 cells can be found in the circulation as well.

killing of target cells that are infected with specific pathogens, or cells that are malignantly transformed. The T cells then become LGLs, with a characteristic morphology of eccentric nuclei and an abundance of cytotoxic azurophilic granules in their cytoplasm [79,80] (Fig. 5). LGLs comprise 10-15% of total T cells in the healthy PB [81], and are mainly present during infection or inflammation in order to clear the pathogen or antigen. Thus they can form small clonal expansions, even in healthy individuals [82]. These clonal expansions have also been shown to be age-related [83].

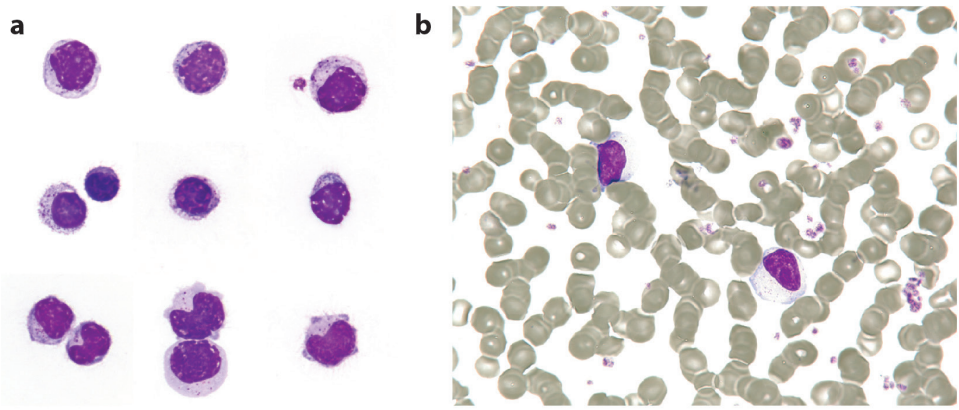


Figure 5. Cytological smears of TCRγδ+ T cell large granular lymphocytes.
(**a**) Normal CD3+ TCRγδ+ T-LGLs sorted from a healthy individual. Magnifications 63x. (**b**) TCRγδ+ T-LGL leukemia cells in the PB of a patient with TCRγδ+ T-LGL leukemia. Cellular cytoplasm is typically enlarged showing eccentric nucleus and azurophilic granules. Pictures adapted from Sandberg *et al.* 2006.

ANTIGENIC STIMULATION AND SHAPING OF THE TCR REPERTOIRE

Antigenic selection

Upon antigenic exposure the T cell repertoire is shaped, mirroring and reflecting processes such as infections, autoimmunity, cancer and ageing. In order to study the immune repertoire traditionally the spectratyping technique has been used [84,85]. Due to recent advances in the field of next-generation sequencing technologies deeper analysis of the repertoire of T cells can be performed. For example, Robins *et al.* applied in 2009 Genomic Analyzer technology on the TRB locus to study the diversity of the complementarity determining region-3 (CDR3) sequence – the antigen-binding part of

the immune receptor – between effector and memory CD4+ and CD8+ compartments [85]. Also in 2009, Freeman *et al.* were one of the first to apply deep sequencing on the investigation of the TRB repertoire diversity. In this study, the authors were able to identify 10-fold more distinct CDR3-β sequences than initially known, showing the promising capabilities of novel advances in DNA sequencing [86]. More recently, Yoshida *et al.* attempted to sequence the peripheral T cell receptors of the TRB locus in ageing individuals, and found that the TCR diversity of CD8+TCRαβ+ T cells decreased with 0.99% per year, and that many (harmless) clones persisted and were able to expand over 20 years, indicating the presence of antigenic selection [87].

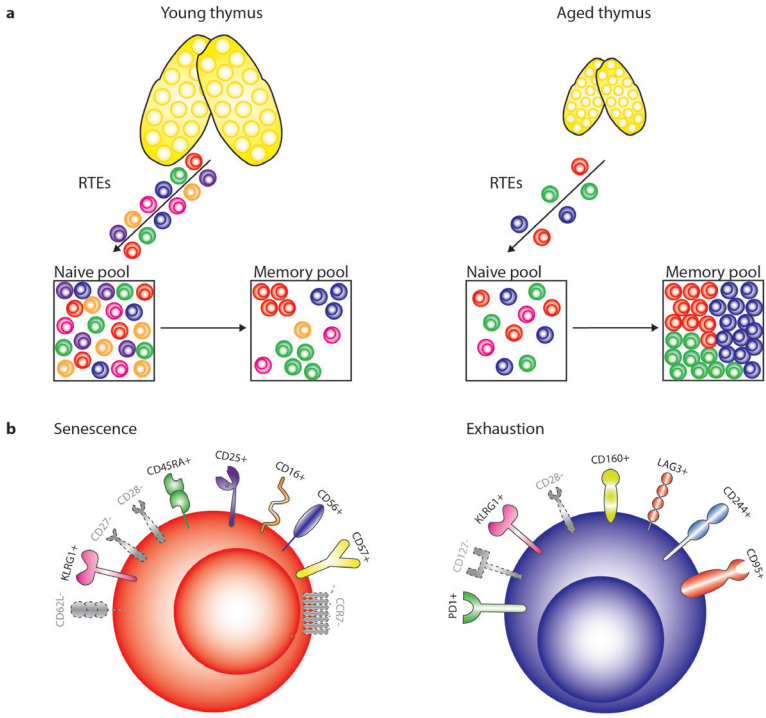
The effect of antigenic selection is even more evident in the TCRγδ+ T cell compartment. This effect is visible in already an early stage during development. In the fetus the epithelial tissues and blood are predominantly populated by the naive IL-17-producing Vδ1+ cells. Readily after birth a switch occurs in the predominant TCRγδ+ T cell population from Vδ1+ to Vγ9/Vδ2, indicating antigenic selection upon antigenic exposure during and after birth [72].

So far, the human TRG/TRD repertoire has hardly been studied. Many different methodologies are available for sequencing immune receptor repertoires, starting from the material, to the amplification, to the data analysis, meaning there is no golden standard available yet (reviewed in [88]). Recently, Chen *et al.* attempted to sequence the CDR3-region of the γδ-receptor by applying IR-SEQ technology on total peripheral blood mononuclear cells (PBMCs) [89]. The disadvantage of this study was the use of total PBMCs or whole blood, while it has been known that also TCRγδ+ T cells exist in different subsets, based on the Vδ-receptor usage and antigen-specificity, but also on maturation levels. Therefore many different receptor specificities and subsets differences could remain unidentified.

Ageing of T cells

The effect of antigenic selection is thought to be extended throughout a life time, since also the immune system is heavily subjected to ageing, often referred to as immunological ageing or immunosenescence. Cellular ageing as a process is defined as a result of damage inflicted by reactive oxygen species (ROS) during e.g. oxidative stress in mitochondria [90]. Damage induced by ROS could lead to genomic instability and without proper DNA proofreading machinery this could contribute to carcinogenesis [91-93]. Since the immune system has a high turnover of cells, which also includes the self-renewal and maintenance of long-lived memory cells, the effects of ageing are obvious at the level of both innate and adaptive immunity. Most striking events are a reduced dendritic cell (DC) activation, impaired humoral responses [94,95] and for T cells most importantly thymic involution, as already described during the early discoveries of the

thymus by Galen [96] (Fig. 6a). Possible explanations for thymic involution could be the (side) effects of circulating (sex) hormones and cytokines, the loss of thymic architecture [97], and the reduced quantity and quality of bone marrow T cell precursors [98]. Eventually, thymic involution is believed to lead to a steady decline in the production of



c Functional characteristics of senescent and exhausted cells

	Senescent cell	Exhausted cell
DNA damage	++	++
Telomere length	-	-
Telomerase activity	-	-
Proliferation	-	-
Apoptosis	-	+++
IFN-γ production	+++	-
TNF-α production	+++	-
Cytotoxicity	+++	-

Figure 6. Effects of ageing on the immune system.
(**a**) Thymic involution results in a decreased number of recent thymic emigrants (RTE) and thus a smaller naive T cell pool, compensated by an increased memory T cell pool. (**b**) T cell ageing is characterized by either senescence, characterized by a specific membrane marker profile, or exhaustion, with a different profile. Up-regulated markers are indicated in color, down-regulated markers are grey. (**c**) Functional changes and characteristics upon senescence or exhaustion. Senescent cells maintain their polyfunctional and cytotoxic potential, exhausted cells lose proliferative and cytotoxic abilities and undergo apoptosis. Figure adapted from Nikolich-Zugich 2008 and reviewed reports in Wherry 2011.

naive, undifferentiated T cells and thus to limited TCR repertoire [99]. Most T cells eventually become memory T cells, which are, due to their long-living character, subjected to ageing processes such as clonal expansion, but also senescence and/or exhaustion (Fig. 6a).

Senescence and exhaustion of T cells

Immunological ageing goes in parallel with the phenomenon of senescence and exhaustion. Senescence is marked by increased DNA damage, short telomeres, loss of proliferative capacity and low telomerase enzyme activity, but maintaining the ability to perform (polyfunctional) effector functions [100,101]. Exhaustion is mainly induced by persistent viral infections, primarily affects CD8+TCRαβ+ cytotoxic T lymphocytes (CTL), and is characterized by the gradual loss of effector functions and proliferative potential, and the gradual increase of the expression of inhibitory molecules (reviewed in [102]) (Fig. 6b, 6c). The final stage of exhaustion is the induction of apoptosis, either spontaneously or via Fas-signaling. Upon chronic stimulation, CD8+TCRαβ+ CTLs shift towards more late-stage differentiated effector phenotypes [103]; an effect that is also observed in TCRγδ+ T cells [104]. However, recently it has been shown that even though similar effects of senescence and exhaustion can be observed between functionally similar CD8+TCRαβ+ and TCRγδ+ T cells [105], Vδ2+ cells do follow different paths of ageing when compared to the TCRαβ+ T cell compartment in terms of differentiation status and cytokine production [106]. TCRγδ+ T cell exhaustion is observed and described more in the context of chronic infections, such as *M. tuberculosis* infection, where the exhaustion is regulated by cytokines, rather than continuous antigenic stimulation with for instance HM-BPP. More downstream mechanisms, such as reduced phosphorylated STAT3, are indicative of TCRγδ+ T cell exhaustion [107]. However, contradicting results were found in a study into the Vδ2-deficiency in granulomatosis polyangiitis (GPA, formerly known as Wegener's granulomatosis). A selective depletion of Vδ2 cells was observed, possibly due to prolonged exposure to one of the etiologic agents of GPA: *S. aureus.* The microbial products of *S. aureus* could be a chronic stimulus for Vγ9/Vδ2 cells, inducing cellular exhaustion [108]. Exhaustion in TCRγδ+ T cells has not been fully defined yet.

CHRONIC T CELL LARGE GRANULAR LYMPHOCYTE LEUKEMIA

Both cellular ageing and persistent, chronic or continuous (antigenic) stimulation do not only affect the exhaustion of T cells, but also impact on immunopathology of

especially lymphoproliferations and ultimately leukemias. Under normal circumstances, CD8+TCRαβ+ and TCRγδ+ T cells become activated upon a stimulus (antigen, cellular damage), undergo proliferation (the expansion phase), and produce granzymes and perforin for killing of target cells. After clearing the antigen, functional memory T cells are formed; the other activated T cells undergo AICD in order to keep the immune response under control. Normally, in case of a continuous, chronic (antigenic) stimulation T cells can undergo the process of exhaustion, the state of functional "anergy" [102], with apoptosis as end-stage. However, in rare situations T cells persist, do not become exhausted, and do not undergo apoptosis; instead they remain actively present, chronically activated and proliferating. This could lead to a proliferation, which eventually could develop into leukemia. In 1985 LGL leukemia was described for the first time, in combination with chromosomal abnormalities, mainly trisomy 8 and 14, and cytopenias such as neutropenia, thrombocytopenia and anemia [109]. The disease was divided into two groups based on either T or NK cell origins [81]. Recently, the World Health Organization (WHO) revised their classification on lymphoid neoplasms, defining T-LGL leukemia, aggressive NK cell leukemia and chronic lymphoproliferative disorder of NK cells [110,111]. Since it results from chronic stimulation, the disease course often follows an indolent and chronic course. These LGL proliferations can be seen as a spectrum, ranging from normal activated and expanded T-LGL cells, to a T cell clonopathy of undetermined significance (TCUS), ultimately ending in a large monoclonal component, also referred to as leukemia [112,113] (Fig. 7). The clinical spectrum is highly heterogeneous in terms of severity and aggression [112]. LGL patients often present with symptoms that are commonly observed among chronic leukemia patients, such as cytopenias, recurrent infections and B symptoms, although a considerable fraction of LGL patients remains asymptomatic. The aggressiveness of the disease was later also associated with the origin of the leukemia: the T-LGL leukemias are mainly indolent, but the NK-LGL origin reflects more malignantly aggressive disease [80,110,111]. T-LGL leukemia is typically presented by a persistent (>6 months) large monoclonal CD3+/CD57+ population of $>2x10^9$/L LGL cells in the PB (Fig. 7) [114]. T-LGL can be divided based on immunophenotype in three separate entities: CD4+TCRαβ+, CD8+TCRαβ+ and TCRγδ+ (Table 1). In case of CD8+TCRαβ+ and TCRγδ+ T-LGL leukemias the disease is often associated with underlying autoimmune diseases, especially rheumatoid arthritis (RA)[115] and malignancies [116]; CD4+TCRαβ+ T cells from a separate group, which also showed a distinctive antigen involved: CMV [78]. However, levels of cytopenias, clinical symptoms and therapies are highly heterogeneous [115]. The etiology of both CD8+TCRαβ+ and TCRγδ+ T-LGL leukemias remain largely unknown, although several underlying mechanisms have been described including chromosomal aberrancies [109],

STAT-mutations through whole-exome analysis [117,118], and neutral loss of heterozygosity at 17q11.2q25.3 with additional aberrations in oncogenes such as NF1 [119].

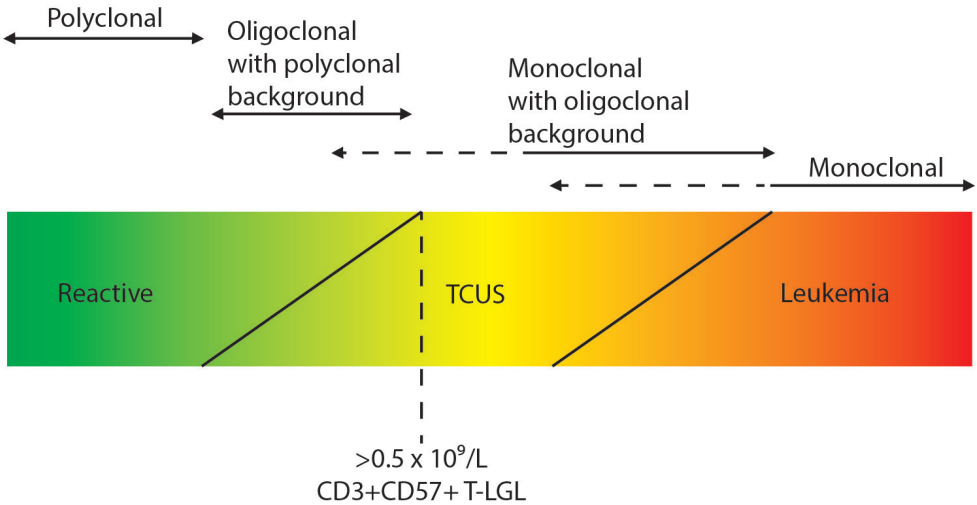


Figure 7. Spectrum of T cell large granular lymphocyte leukemias.
In the healthy situation T cells are polyclonally expanded and reactive in order to respond to and clear infections. During an infection cells with a specific receptor recognizing the specific epitope(s) are selected, activated and start to proliferate, leading to an oligoclonal population. The diverse polyclonal background is still present, and the oligoclonal population undergoes activation induced cell death after the antigen is cleared. Upon prolonged stimulation this oligoclonal population may persist, and continue to proliferate, leading to a monoclonal population called T cell clonopathy of undetermined significance (TCUS), still in the context of poly- or oligoclonal cells. Following chronic stimulation the monoclonal population persists, proliferates and represses the normal background, leading to a large monoclonal cell population. This ultimately leads to a clinically malignant situation: a chronic T cell large granular lymphocytic (T-LGL) leukemia. Diagnosis of T-LGL leukemia is based on a persistent (>6 months) monoclonal CD3+/CD57+ population with an absolute count of >0.5 x 10^9 cells per liter peripheral blood. Figure adapted from Langerak *et al.* 2003.

Table 1. T-LGL leukemia characteristics.

	CD4+TCRαβ+ T-LGL	**CD8+TCRαβ+ T-LGL**	**TCRγδ+ T-LGL**
Frequency (% of all T-LGL leukemia cases)	<5%	5-10%	80-90%
Associated autoimmune disease	<5% cytopenias 10% autoimmune diseases	80% cytopenias 20% autoimmune diseases	60% cytopenias 30% autoimmune diseases
Associated malignancies	20%	20%	20%

Characteristics based on Sandberg et al. 2014 (Chapter 5 in this book).

SCOPE OF THE THESIS

Based on this chronic character and their antigen-experienced effector profiles T-LGL leukemia is believed to arise from normal T-LGL cells in the adult or elderly blood, upon chronic (antigenic) stimulation followed by secondary events in either signaling, gene expression or genome. Throughout this thesis different aspects of T cells and T-LGL leukemias – with special emphasis on TCRγδ-variants – are studied (see for graphical abstract Fig. 8). Immunological events during a life time are described in a spectrum starting from normal developing, to healthy existing, to finally aberrancies causing a chronic mature leukemia.

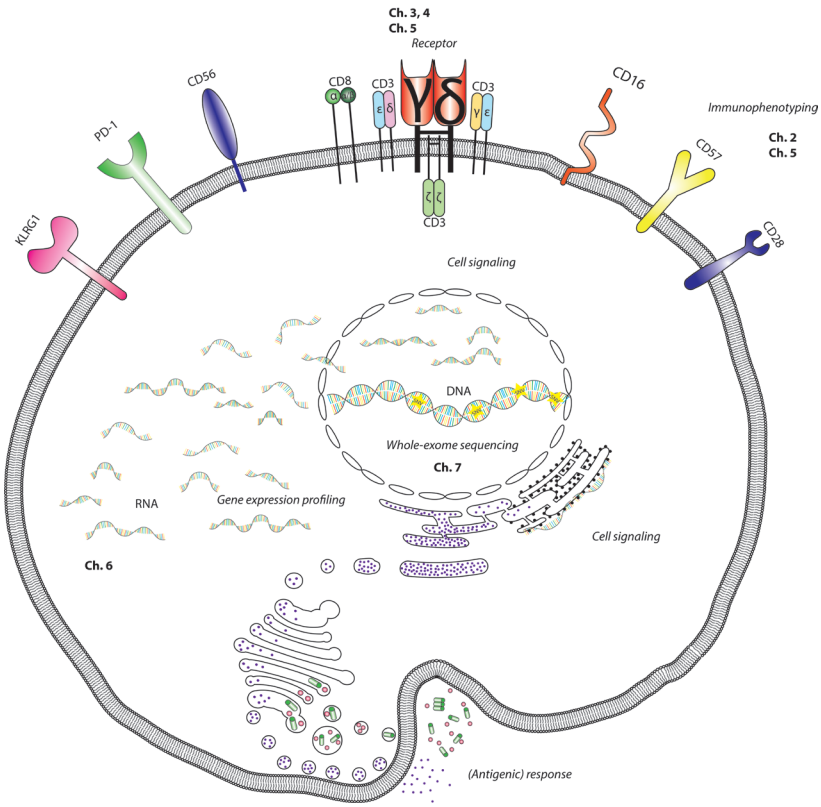


Figure 8. Graphical abstract of this thesis.
Chapter 2 addresses the immunophenotype of normal TCRγδ+ T cells in the context of ageing. In Chapters 3 and 4 TRG/TRD repertoire development in TCRγδ+ T cells is discussed. Chapter 5 is a review on the clinical features and immunophenotype, and clonotypic TCR repertoire of T-LGL leukemia, including TCRγδ+ T-LGL leukemia. Chapter 6 investigates gene expression profiles in T-LGL leukemia indicative of dysregulated processes, and Chapter 7 describes the deepest level in which aberrancies could occur, the genomic level, these last two Chapters have mainly focused on the TCRγδ+ variant.

In **Chapter 2** healthy developing and ageing TCRγδ+ T cells are studied with respect to surface marker expression. The effect of immunological ageing – also referred to as immunosenescence – with the additional effect of persistent viruses like CMV on the TCRγδ+ T cell population is discussed in this chapter. **Chapters 3** and **4** address the healthy ageing TCRγδ+ T cell receptor repertoire. In **Chapter 3** a detailed description is given of the novel next generation sequencing technique that was applied to study the TRG and TRD repertoires. Next, the development of and further shaping of the TCRγδ+ T cell receptor repertoire is reported in **Chapter 4**, including novel insights on development and ageing on both cell surface marker expression and antigen receptor diversity.

Chapter 5 provides an overview of all types of chronic mature T-LGL leukemias that can arise from the T cell type LGL: CD4+TCRαβ+, CD8+TCRαβ+ and TCRγδ+ T-LGL leukemias, and compares the different pathogenic events, but also the different clinical presentations and associated diseases. Alterations in TCRγδ+ T-LGL leukemia cells on the level of gene expression are described in **Chapter 6**, in comparison to different healthy TCRγδ+ T cell subsets. **Chapter 7** reports on the deepest level of cellular aberrancies, namely the genetic abnormalities that could contribute to the TCRγδ+ T-LGL leukemogenesis. Finally, in the **General Discussion** in **Chapter 8** an integrated view of all chapters is given, together with all (novel) findings and data reported in this thesis and previous literature on the corresponding topics. Future prospects on how to address newly identified issues are also included.

REFERENCES

1. Miller, J.F.A.P. Immunological function of the thymus. *Lancet,* **2(7205)**, 749-749 (1961).
2. Miller, J.F.A.P. The discovery of thymus function and of thymus-derived lymphocytes. *Immunol. Rev.* **185**, 7-14 (2002).
3. Hewson, W. Experimental inquiries, part the third, containing a description of the red particles of the blood in the human subject and in other animals, with an account of the structure and offices of the lymphatic glands, of the thymus gland, and of the spleen; being the remaining part of the observations and experiments of the late Mr. William Hewson, FRS and teacher of anatomy; by Magnus Falconer, surgeon and teacher of anatomy. Printed for T. Longman, No. **39**, Paternoster Row, London (1777).
4. Doyle, D. William Hewson (1739-74): the father of haematology. *Br. J. Haematol.* **133**, 375-381 (2006).
5. Gulliver, G. In: The Works of William Hewson. London: Sydenham Society (1846).
6. Vadász, J. The origin of thymocytes. *Acta. Morphol. Acad. Sci. Hung.* **4(3)**, 289-292 (1954).
7. Miller, J.F.A.P. & Mitchell, G.F. The thymus and the precursors of antigen-reactive cells. *Nature* **216**, 659-663 (1967).

8. Weerkamp, F. *et al.* Human thymus contains multipotent progenitors with T/B lymphoid, myeloid and erythroid lineage potential. *Blood* **107(8)**, 3131-3137 (2006).
9. Weerkamp, F., Pike-Overzet, K. & Staal, F.J.T. T-sing progenitors to commit. *Trends Immunol.* **27(3)**, 125-131 (2006).
10. Haskins, K. *et al.* The major histocompatibility complex-restricted antigen receptor on T cells. *J. Exp. Med.* **157**, 1149-1169 (1983).
11. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-581 (1983).
12. Kronenberg, M., Siu, G., Hood, L.E. & Shastri, N. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Ann. Rev. Immunol.* **4**, 529-591 (1986).
13. Davis, M.M. & Bjorkman, P.L. T-cell antigen receptor genes and T-cell recognition. *Nature* **334(4)**, 395-402 (1988).
14. Lefranc, M.P. & Rabbitts, T.H. The human T-cell receptor γ (TRG) genes. *TIBS* **14**, 214-218 (1989).
15. Lefranc, M.P. & Lefranc, G. The T cell receptor factsbook. Academic Press, ISBN 0-12-441352-8 (2001).
16. Saito, H., Kranz, D.M., Takagaki, Y., Hayday, A.C., Eisen, H.N. & Tonegawa, S. Complete primary structure of a heterodimeric T-cell receptor deduced from cDNA sequences. *Nature* **309**, 757-762 (1984).
17. Hayday, A.C. *et al.* Structure, organization, and somatic rearrangement of T cell gamma genes. *Cell* **40**, 259-269 (1985).
18. Bank, I., DePinho, R.A., Brenner, M.B., Cassimeris, J., Alt, F.W. & Chess, L. A functional T3 molecule associated with a novel heterodimer on the surface of immature human thymocytes. *Nature* **322**, 179-181 (1986).
19. Born, W. *et al.* Peptide sequences of T-cell receptor delta and gamma chains are identical to predicted X and gamma proteins. *Nature* **330**, 572-574 (1987).
20. Loh, E.J. *et al.* Identification and sequence of a fourth human T cell antigen receptor chain. *Nature* **330**, 569-592 (1987).
21. Engel, I. & Murre, C. The function of E- and id proteins in lymphocyte development. *Nat. Rev. Immunol.* **1**, 193-199 (2001).
22. Dik, W.A. *et al.* New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **2-1(11)**, 1715-1723 (2005).
23. Staal, F.J.T., van Dongen, J.J.M. & Langerak, A.W. Novel insights into the development of T-cell acute lymphoblastic leukemia. *Curr. Hematol. Malig. Rep.* **2**, 176-182 (2007).
24. Oettinger, M.A. Activation of V(D)J recombination by RAG1 and RAG2. *Trends Genet.* **8(12)**, 413-416 (1992).
25. Turner, S.J., Doherty, P.C., McCluskey, J. & Rossjohn, J. Generation of T-cell receptor (TCR) diversity by somatic recombination of TCR gene segments. *Nat. Rev. Immunol.* **6**, 883-894 (2006).
26. Bogue, M., Gilfillan, S., Benoist, C. & Mathis, D. Regulation of N-region diversity in antigen receptors through thymocyte differentiation and thymus ontogeny. *Proc. Natl. Acad. Sci. USA.* **89**, 11011-11015 (1992).

27. Benedict, C.L., Gilfillan, S., Thai, T-H. & Kearney, J.F. Terminal deoxynucleotidyl transferase and repertoire development. *Immunol. Rev.* **175**, 150-157 (2000).
28. Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J. & Kourilsky, P. A direct estimate of the human αβ T cell receptor diversity. *Science* **286**, 958-961 (1999).
29. Breit, T.M., Wolvers-Tettero, I.L.M. & van Dongen, J.J.M. Receptor diversity of human TCR-γδ expressing cells. *Prog. Histochem. Cytochem.* **26**, 182-193 (1992).
30. Schatz, D.G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **11**, 251-263 (2011).
31. Germain, R.N. T-cell development and the CD4-CD8 lineage decision. *Nat. Rev. Immunol.* **2**, 309-322 (2002).
32. Akiyama, T., Tateishi, R., Akiyama, N., Yoshinaga, R. & Kobayashi, T.J. Positive and negative regulatory mechanisms for fine-tuning cellularity and functions of medullary thymic epithelial cells. *Front. Immunol.* **6**, 461 (2015).
33. Heino, M. *et al.* Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla. *Biochem. Biophys. Res. Commun.* **257**, 821-825 (1999).
34. Pennington, D.J., Silva-Santos, B. & Hayday, A.C. γδ T cell development – having the strength to get there. *Curr. Opin. Immunol.* **17**, 108-115 (2005).
35. Haks, M.C. *et al.* Attenuation of γδ TCR signaling efficiently diverts thymocytes to the αβ lineage. *Immunity* **22**, 595-606 (2005).
36. Hayes, S.M., Li, L. & Love, P.E. TCR signal strength influences αβ/γδ lineage fate. *Immunity* **22**: 583-593 (2005).
37. Pang, D.J., Neves, J.F., Sumaria, N. & Pennington, D.J. Understanding the complexity of γδ T-cell subsets in mouse and human. *J. Immunol.* **136(3)**, 283-290 (2012).
38. Turchinovich, G. & Pennington, D.J. T cell receptor signaling in γδ cell development: strength isn't everything. *Trends Immunol.* **32(12)**, 567-573 (2011).
39. Prinz, I., Sansoni, A., Kissenpfennig, A., Ardouin, L., Malissen, M. & Malissen, B. Visualization of the earliest steps of gammadelta T cell development in the adult thymus. *Nat. Immunol.* **7**,: 995-1003 (2006).
40. Jensen, K.D.C. & Chien, Y. Thymic maturation determines γδ T cell function but not their antigen-specificities. *Curr. Opin. Immunol.* **21**, 140-145 (2009).
41. Jensen, K.D. *et al.* Thymic selection determines γδ T cell effector fate: antigen-naive cells make Interleukin-17 and antigen-experienced make Interferon γ. *Immunity* **29**, 90-100 (2008).
42. Resop, R.S., Douaisi, M., Craft, J., Jachimowski, L.C.M., Blom, B. & Uittenbogaart, C.H. Sphingosine-1-phosphate/sphingosine-1-phosphate receptor 1 signaling is required for migration of naive human T cells from the thymus to the periphery. *J. Allergy Clin. Immunol.* **138**, 551-557 (2016).
43. Baeyens, A., Fang, V., Chen, C. & Schwab, S.R. Exit strategies: S1P signaling and T cell migration. *Trends Immunol.* **36(12)**, 778-787 (2015).
44. Murphy, K.M. Janeway's Immunobiology. 8th ed. New York: Garland Science. 888p (2012).

45. Sikora, E. Activation-induced and damage-induced cell death in ageing human T cells. *Mech. Ageing Dev.* **151**, 85-92 (2015).
46. Alderson, M.R. *et al.* Fas ligand mediates activation-induced cell death in human T lymphocytes. *J. Exp. Med.* **181**, 71-77 (1995).
47. Green, D.R., Droin, N. & Pinkoski, M. Activation-induced cell death in T cells. *Immunol. Rev.* **193**, 70-81 (2003).
48. Michishita, Y. *et al.* Age-associated alteration of γδ T-cell repertoire and different profiles of activation-induced death of Vδ1 and Vδ2 T cells. *Int. J. Hematol.* **94**, 230-240 (2011).
49. Vucic, D., Dixit, V.M. & Wertz, I.E. Ubiquitinylation in apoptosis: a post-translational modification at the edge of life and death. *Nat. Rev. Mol. Cell. Biol.* **12**, 439-452 (2011).
50. Van den Brink, M.R.M. & Burakoff, S.J. Cytolytic pathways in haematopoietic stem-cell transplantation. *Nat. Rev. Immunol.* **2**, 273-281 (2002).
51. Chien, Y.H., Jores, R. & Crowley, M.P. Recognition by γ/δ T cells. *Annu. Rev. Immunol.* **14**, 511-532 (1996).
52. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13**, 88-100 (2013).
53. Morita, C.T., Jin, C., Sarikonda, G. & Wang, H. Nonpeptide antigens, presentation mechanisms, and immunological memory of human Vγ2Vδ2 T cells: discriminating friend from foe through the recognition of prenyl pyrophosphate antigens. *Immunol. Rev.* **215**, 59-76 (2007).
54. Morita, C.T. *et al.* Direct presentation of nonpeptide prenyl pyrophosphate antigens to human γδ T cells. *Immunity* **3**, 495-507 (1995).
55. Hewitt, R.E., Lissina, A., Green, A.E., Slay, E.S., Price, D.A. & Sewell, A.K. The bisphosphonate acute phase response: rapid and copious production of proinflammatory cytokines by peripheral blood γδ T cells in response to aminobisphosphonates is inhibited by statins. *Clin. Exp. Immunol.* **139**, 101-111 (2005).
56. Spada, F.M. *et al.* Self-recognition of CD1 by γ/δ T cells: implications for innate immunity. *J. Exp. Med.* **191**, 937-948 (2000).
57. Bai, L. *et al.* The majority of CD1d-sulfatide specific T cells in human blood use a semi-invariant Vδ1 TCR. *Eur. J. Immunol.* **42(9)**, 2505-2510 (2012).
58. Tanaka, Y., Morita, C.T., Tanaka, Y., Nieves, E., Brenner, M.B. & Bloom, B.R. Natural and synthetic non-peptide antigens recognized by human γδ T cells. *Nature* **375**, 155-158 (1995).
59. Behr, C. *et al.* Plasmodium falciparum stimuli for human gammadelta T cells are related to phosphorylated antigens of mycobacteria. *Infect. Immun.* **64(8)**, 2892-2896 (1996).
60. Watanabe, N. *et al.* Type I IFN-mediated enhancement of anti-leukemic cytotoxicity of γδ T cells expanded from peripheral blood cells by stimulation with zoledronate. *Cytotherapy* **8(2)**, 118-129 (2000).
61. Saitoh, A. *et al.* Anti-tumor cytotoxicity of γδ T cells expanded from peripheral blood cells of patients with myeloma and lymphoma. *Med. Oncol.* **25**, 137-147 (2008).

62. Maniar, A. *et al.* Human γδ T lymphocytes induce robust NK cell-mediated antitumor cytotoxicity through CD137 engagement. *Blood* **116**, 1726-1733 (2010).
63. Schmolka, N. *et al.* Epigenetic and transcriptional signatures of stable versus plastic differentiation of proinflammatory γδ T cell subsets. *Nat. Immunol.* **14(10)**, 1093-1100 (2013).
64. Muñoz-Ruiz, M. *et al.* TCR signal strength controls thymic differentiation of discrete proinflammatory γδ T cell subsets. *Nat. Immunol.* **17(6)**, 721-727 (2016).
65. Swamy, M., Jamora, C., Havran, W. & Hayday, A. Epithelial decision markers: in search of the 'epimmunome'. *Nat Immunol* **11**, 656-665 (2010).
66. Prinz, I., Silva-Santos, B. & Pennington, D.J. Functional development of γδ T cells. *Eur. J. Immunol.* **43**, 1988-1994 (2013).
67. Lu, Y., Cao, X., Zhang, X. & Kovalovsky, D. PLZF controls the development of fetal-derived IL-17+ Vγ6+ γδ T cells. *J Immunol* **195**, 4273-4281 (2015).
68. Hayday, A. & Gibbons, D. Brokering the peace: the origin of intestinal T cells. *Mucosal Immunol.* **1(3)**, 172-174 (2008).
69. Holtmeier, W., Pfänder, M., Hennemann, A., Zollner, T.M., Kaufmann, R. & Caspary, W. The TCRδ repertoire in normal human skin is restricted and distinct from the TCRδ repertoire in the peripheral blood. *J. Invest. Dermatol.* **116**, 275-280 (2001).
70. De Rosa, S.C. *et al.*Ontogeny of γδ T cells in humans. *J. Immunol*. **172**, 1637-1645 (2004).
71. Porcelli, S., Brenner, M.B. & Band, H. Biology of the human γδ T-cell receptor. *Immunol. Rev.* **120**, 137-183 (1991).
72. Kabelitz, D. Function and specificity of human γ/δ-positive T cells. *Crit. Rev. Immunol.* **11**, 281-303 (1992).
73. Söderstrom, K. *et al.* High expression of Vγ8 is a shared feature of human γδ T cells in the epithelium of the gut and in the inflamed synovial tissue. *J. Immunol.* **152**, 6017-6027 (1994).
74. Kabelitz, D. *et al.*Clonal expansion of Vγ3/Vδ3-expressing γδ T cells in an HIV-1/2-negative patient with CD4 T-cell deficiency. *Br. J. Haematol.* **96**, 266-271 (1997).
75. Gibbons, D.L., *et al.* Neonates harbor highly active γδ T cells with selective impairments in preterm infants. *Eur. J. Immunol.* **39**, 1794-1806 (2009).
76. Sandberg, Y. *et al.* TCRγδ+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRγδ+ T cells. *Leukemia* **20**, 505-513 (2006).
77. Ryan, P.L. *et al.* Heterogeneous yet stable Vδ2(+) T-cell profiles define distinct cytotoxic effector potentials in healthy human individuals. *Proc. Natl. Acad. Sci. USA.* **113(50)**, 14378-14383 (2016).
78. Garrido, P. *et al.* Monoclonal TCR-Vβ13.1+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis: evidence for an antigen-driven chronic T-cell stimulation origin. *Blood* **109**, 4890-4898 (2005).
79. Rose, M.G. & Berliner, N. T-cell large granular lymphocyte leukemia and related disorders. *Oncologist* **9(3)**, 247-258 (2004).
80. Sokol, L. & Loughran Jr, T.P. Large granular lymphocyte leukemia. *Oncologist* **11**, 263-273 (2006).
81. Loughran Jr, T.P. Clonal diseases of large granular lymphocytes. *Blood* **82**, 1-14 (1993).

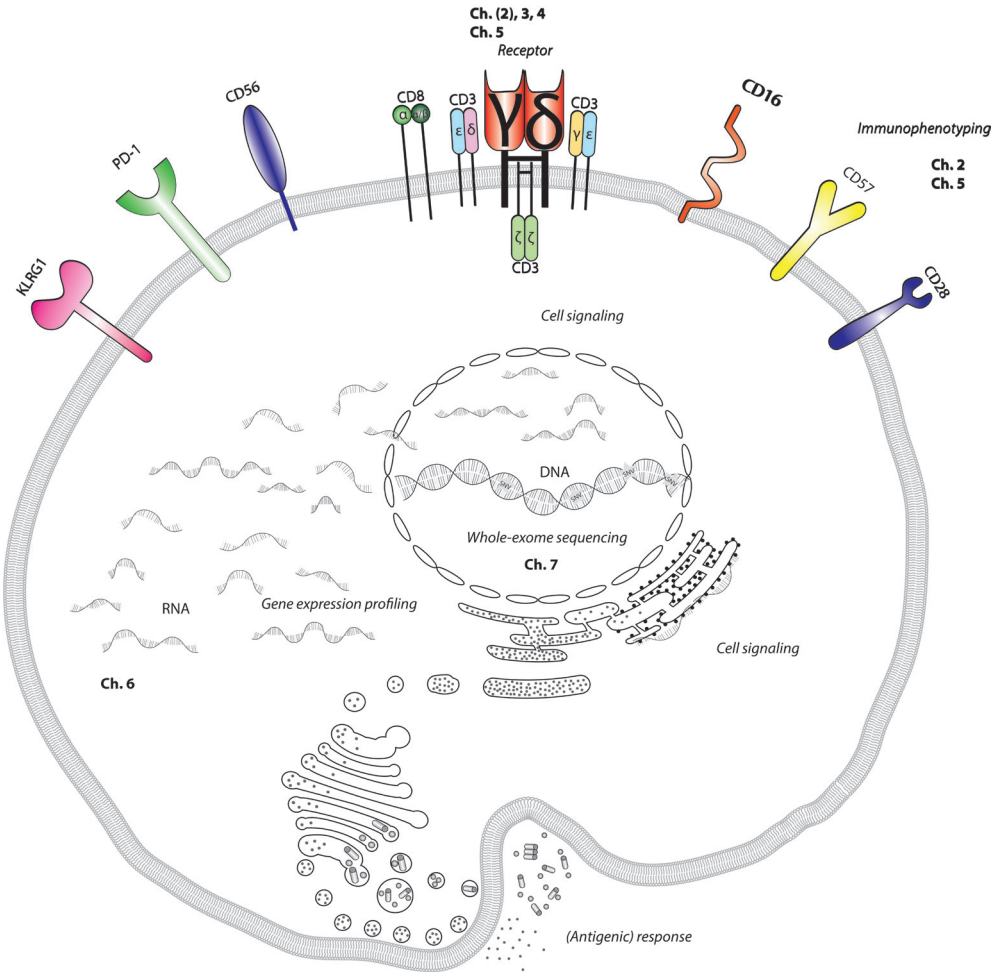
82. Fitzgerald, J.E. *et al.* Analysis of clonal CD8+ T cell expansions in normal individuals and patients with rheumatoid arthritis. *J. Immunol.* **154**, 3538-3547 (1995).
83. Wack, A. *et al.* Age-related modifications of the human alphabeta T cell receptor due to different clonal expansions in the CD4+ and CD8+ subsets. *Int. Immunol.* **10(9)**, 1281-1288 (1998).
84. Van Dongen, J.J.M. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* **17**, 2257-2317 (2003).
85. Robins, H.S. *et al.* Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114(19)**, 4099-4107 (2009).
86. Freeman, J.D., Warren, R.L., Webb, J.R., Nelson, B.H. & Holt, R.A. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* **19(10)**, 1817-1824 (2009).
87. Yoshida, K. *et al.* Aging-related changes in human T-cell repertoire over 20years delineated by deep sequencing of peripheral T-cell receptors. *Exp. Gerontol.* **96**, 29-37 (2017).
88. Rosati, E., Dowds, C.M., Liaskou, E., Henriksen, E.K.K., Karlsen, T.H. & Franke, A. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* **17(1)**, 61 (2017).
89. Chen, H., Zou, M., Teng, D., Zhang, J. & He, W. Characterization of the diversity of T cell receptor γδ complementary determinant region 3 in human peripheral blood by Immune Repertoire Sequencing. *J. Immunol. Methods* **443**, 9-17. (2017).
90. Lee, H.C. & Wei, Y.H. Mitochondria and ageing. *Adv. Exp. Med. Biol.* **942**, 311-327 (2012).
91. Gilbert, S.F. Ageing: the biology of senescence. Developmental biology. 6th edition. Sunderland (MA): Sinauer Associates (2000).
92. Sosa, V., Moliné, T., Somoza, R., Paciucci, R., Kondoh, H. & Leonart, M.E. Oxidative stress and cancer: an overview. *Ageing Res. Rev.* **12(1)**, 376-390 (2013).
93. Lauri, A., Pompilio, G. & Capogrossi, M.C. The mitochondrial genome in ageing and senescence. *Ageing Res. Rev.* **18**, 1-15 (2014).
94. Weiskopf, D., Weinberger, B. & Grubeck-Loebenstein, B. The ageing of the immune system. *Transp. Int.* **22**, 1041-1050 (2009).
95. Boraschi, D. *et al.* The gracefully ageing immune system. *Ageing* **5(185)**, 185ps8 (2013).
96. Aw, D., Silva, A.B. & Palmer, D.B. Immunosenescence: emerging challenges for an ageing population. *J. Immunol.* **120**, 435-446 (2007).
97. Gowans, J.L., Gesner, B.M. & McGregor, D.D. The immunological activities of lymphocytes. In: Wolstenholme GEW, O'Connor M, editors. Biological activities of leucocyte. London: Churchill, 32-34. Ciba Foundation Study Group (1961).
98. Ma, D., Wei, Y. & Feng, L. Regulatory mechanisms of thymus and T cell development. *Dev. Comp. Immunol.* **39**, 91-102 (2013).
99. Zdrojewicz, Z., Pachura, E. & Pachura, P. The thymus: a forgotton, but very important organ. *Adv. Clin. Exp. Med.* **25(2)**, 369-375 (2016).

100. Nikolich-Zugich, J. Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections. *Nat. Rev. Immunol.* **8**, 512-522 (2008).
101. Pereira, B.I. & Akbar, A.N. Convergence of innate and adaptive immunity during human ageing. *Front. Immunol.* **7**, 445 (2016).
102. Wherry, E.J. T cell exhaustion. *Nat. Immunol.* **12(6):** 492-499 (2011).
103. Purnama, C.P., Camous, X. & Larbi, A. An overview of T cell subsets and their potential use as markers of immunological ageing. *Int. Trends Immun.* **1(4)**, 21-32 (2013).
104. Wistuba-Hamprecht, K., Haehnel, K., Janssen, N., Demuth, I. & Pawelec, G. Peripheral blood T-cell signatures from high-resolution immune phenotyping of γδ and αβ T cells in younger and older subjects in the Berlin Ageing Study II. *Immun. Ageing* **12**:25 (2015).
105. Bonneville, M., O'Brien, R.L. & Born, W.K. γδ T cell effector functions: a blend of innate programming and acquired plasticity. *Nat. Rev. Immunol.* **10**, 467-478 (2010).
106. Tan, C.T.Y. *et al.* Vδ2+ and α/β T cells show divergent trajectories during human ageing. *Oncotarget* **7(29)**, 44906-44918 (2016).
107. Shen, H. *et al.* Selective destruciton of iL-23 induced expansion of major antigen-specific γδ T-cell subset in TB patients. *J. Infect. Dis.* Epub **511**, doi 10.1093 (2016).
108. Fazio, J. *et al.* Vδ2 deficiency in granulomatosis with polyangiitis (Wegener's granulomatosis). *Clin. Immunol.* **149(1)**, 65-72 (2013).
109. Loughran Jr, T.P. *et al.* Leukemia of large granular lymphocytes: association with clonal chromosomal abnormalities and autoimmune neutropenia, thrombocytopenia, and hemolytic anemia. *Ann. Intern. Med.* **102(2)**, 169-175 (1985).
110. Swerdlow, S.H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375-2390 (2016).
111. Lamy, T., Moignet, A. & Loughran Jr, T.P. LGL leukemia: from pathogenesis to treatment. *Blood* 2016-08-692590, (2017).
112. Semenzato, G. *et al.* The lymphoproliferative disease of granular lymphocytes: a heterogeneous disorder ranging from indolent to aggressive conditions. *Cancer* **60**, 2971-2978 (1987).
113. Singleton, T.P., Yin, B., Teferra, A. & Mao, J.Z. Spectrum of clonal large granular lymphocytes (LGLs) of αβ T cells. *Am. J. Clin. Pathol.* **144**, 137-144 (2015).
114. Langerak, A.W., Sandberg, Y. & van Dongen, J.J.M. Spectrum of T-large granular lymphocyte lymphoproliferations: ranging from expanded activated effector T cells to T-cell leukaemia. *Br. J. Haematol.* **123(3)**, 561-562 (2003).
115. Dhodapkar, M.V., Li, C-Y., Lust, J.A., Tefferi, A. & Phyliky, R.L. Clinical spectrum of clonal proliferations of T-large granular lymphocytes: a T-cell clonopathy of undetermined significance? *Blood* **84(5)**, 1620-1627 (1994).
116. Sandberg, Y. *et al.* Lack of common TCRA and TCRB clonotypes in CD8+/TCRαβ+ T-cell large granular lymphocyte leukemia: a review on the role of antigenic selection in the immunopathogenesis of CD8+ T-LGL. *Blood Cancer J.* **4**, e172 (2014).

117. Koskela, H.L.M. *et al.* Somatic STAT3 mutations in large granular lymphocytic leukemia. *N. Engl. J. Med.* **366**, 1905-1913 (2012).
118. Rajala, H.L.M. *et al.* Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. *Blood* **121**, 4541-4550 (2013).
119. Zhang, L., Ramchandrem, R., Papenhausen, P., Loughran Jr, T.P. & Sokol, L. Transformed aggressive γδ variant T-cell large granular lymphocytic leukemia with acquired copy neutral loss of heterozygosity at 17q11.2q25.3 and additional aberrations. *Eur. J. Haematol.* **93**, 260-264 (2014).

Part I

Normal TCRγδ+ T cells


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 7
RNA
Gene expression profiling
Cell signaling
Ch. 6
(Antigenic) response

Chapter 2

Ageing and latent CMV infection impact on maturation, differentiation and exhaustion profiles of T cell receptor gammadelta T cells

Martine J. Kallemeijn[1], Anne Mieke H. Boots[2], Michèle Y. van der Klift[1], Elisabeth Brouwer[2], Wayel H. Abdulahad[2], Jan A.N. Verhaar[3], Jacques J.M. van Dongen[1], Anton W. Langerak[1]

[1]: Department of Immunology, Laboratory for Medical Immunology, Erasmus MC, Rotterdam 3000 CA, the Netherlands.
[2]: Department of Rheumatology, University Medical Centre Groningen, Groningen 9700 RB, the Netherlands.
[3]: Department of Orthopedics, Erasmus MC, Rotterdam 3000 CA, the Netherlands.

Sci. Rep. **7(1)**, 5509 (2017)

ABSTRACT

Ageing is a broad cellular process, largely affecting the immune system, especially T lymphocytes. Additionally to immunosenescence alone, cytomegalovirus (CMV) infection is thought to have major impacts on T cell subset composition and exhaustion. These impacts have been studied extensively in TCRαβ+ T cells, with reduction in naive, increase in effector (memory) subsets and shifts in CD4/CD8-ratios, in conjunction with morbidity and mortality in elderly. Effects of both ageing and CMV on the TCRγδ+ T cell compartment remain largely elusive.

In the current study we investigated Vγ- and Vδ-usage, maturation, differentiation and exhaustion marker profiles of both CD4 and CD8 double-negative (DN) and CD8+TCRγδ+ T cells in 157 individuals, age range 20-95.

We observed a progressive decrease in absolute numbers of total TCRγδ+ T cells in blood, affecting the predominant Vγ9/Vδ2 population. Aged TCRγδ+ T cells appeared to shift from naive to more (late-stage) effector phenotypes, which appeared more prominent in case of persistent CMV infections. In addition, we found effects of both ageing and CMV on the absolute counts of exhausted TCRγδ+ T cells. Collectively, our data show a clear impact of ageing and CMV persistence on DN and CD8+TCRγδ+ T cells, similar to what has been reported in CD8+TCRαβ+ T cells, indicating that they undergo similar ageing processes.

Keyword(s): ageing, immunosenescence, TCRγδ+ T cells, CMV, differentiation, exhaustion, flow cytometry.

INTRODUCTION

Ageing is a general cellular process, defined as the result of damage created by reactive oxygen species (ROS) during oxidative stress in mitochondria [1]. ROS can cause cell membrane, protein, nucleic acid damage [2], and most importantly genome damage which leads to genomic instability, shortening of telomere length, and thus an increasing chance of cancer development [3,4]. The process of ageing particularly affects the immune system, due to its high metabolic rate and high cellular turnover for maintaining homeostasis, and for protecting the host against infections and cancer. Immunological ageing (also called immunosenescence) is defined at different levels: desensitization of dendritic cells (DCs) leading to reduced TLR responses, low bone marrow (BM) output of naive B cells, insufficient T cell help in the spleen and lymph nodes (LN), resulting in decreased memory B cell expansions and antibody secretion, and decreased thymopoiesis in the thymus [5,6]. Clinically this results in an inadequate response to infections in elderly, caused by reduced innate responses of macrophages, neutrophils and NK cells [6,7]. DCs are constitutively activated, which gives rise to an increased basal level of inflammation with increased tissue damage [8,9]. Defective antigen presentation and a reduced B cell repertoire lead to a reduced humoral response [10] and a reduced vaccine response [6,11].

A central feature in immunosenescence is involution of the thymus, which is characterized by thymic shrinkage and a significantly reduced naive T cell output [5,6,12]. This leads to a reduced T cell dependent antigen-specific response and thus fewer interactions with other immune cell types, such as reduced help to B cells in germinal centers [11]. During immune ageing, another major event has been described, which is referred to as T cell exhaustion. This exhaustion process is characterized by the progressive loss of robust effector functions and eventually the induction of apoptosis. T cell exhaustion is most clearly seen in chronic infections, e.g. in persistent viral infections, and in cancers. Due to continuous stimulation, T cells start to lose their effector functions in a hierarchical manner, starting with reduced IL-2 production, followed by reduced cytokine and chemokine productions, ending with the high expression of inhibitory molecules and eventually the induction of apoptosis [13,14]. Many different markers for exhausted CD8+ CTLs have been described, ranging from NK cell markers such as CD57 [15-17], killer cell lectin-like receptor G1 (KLRG1) [13,18,19], and 2B4, also known as CD244 [20-22], to cell death-associated markers such as Programmed cell death 1 (PD1) which is a marker of early exhaustion [23-25], and FAS (CD95) [13,26]. Loss of the self-renewal-associated marker IL-7 receptor α subunit (CD127) is associated with an early stage of exhaustion [27,28].

The process of immunological ageing, including expression of the above markers has been extensively studied in CD8+TCRαβ+ T cells, but less so in TCRγδ+ T cells, which show functional overlap with the CD8+TCRαβ+ CTLs with respect to high levels of cytotoxicity [29], cytokine release – mainly IFN-γ and IL-17 based on antigen experience – [30,31], induction of inflammation, immunoregulation and cytoprotection upon antigen recognition. However, TCRγδ+ T cells form a distinctive group of unconventional T cells with features of both innate and adaptive immune cells [32]. TCRγδ+ T cells recognize antigens directly without major histocompatibility molecules (MHC), or in the context of CD1-molecules [33-35]. TCRγδ+ T cells thus have the ability to directly respond to specific pathogens, and readily form a bridge between the innate and adaptive systems. Upon ageing, TCRγδ+ T cells also tend to decrease in total numbers [36,37], leading to a possibly reduced response to pathogens. This relates not only to the blood, but also to epithelial tissues where they reside as intra-epithelial and innate-like lymphocytes [33,38]. Furthermore, TCRγδ+ T cells can specifically bind to viruses, such as human lymphotrophic virus type I (HTLV-I) and Epstein-Barr virus (EBV) [39], through the Vγ9/Vδ2 receptor. Non-Vγ9/Vδ1 cells specifically respond to cytomegalovirus (CMV) [40], increase upon ageing and can be expanded and stimulated with CMV *ex vivo* [41]. As CMV is one of the persistent herpesviruses [42], CMV infection has a high impact on immunosenescence and exhaustion [43-45]. CMV is known for altering TCRαβ+CD4+ and CD8+ maturation subsets (reviewed in [46]), and recently it has been found that CMV seropositivity in elderly individuals is associated with a lower percentage of Vδ2+, and an increased percentage of Vδ1+TCRγδ+ T cells, of which the latter has a late-stage differentiated effector phenotype [37]. However, the full profile of phenotypic alterations of TCRγδ+ T cells upon ageing in the presence or absence of persistent CMV infections remains elusive.

Since TCRγδ+ T cells have innate features and show functional similarities to CD8+TCRαβ+ CTLs, we hypothesized that immunological ageing, especially in the presence of CMV, would similarly influence the TCRγδ+ T cell immune system with respect to subset compositions and exhaustion profiles. In the current study we included 157 healthy subjects from different age groups to investigate the effect of both ageing and CMV seropositivity on TCRγδ+ T cells. Our data illustrate the impact of immunological ageing on TCRγδ+ T cells, with a clear enhancing effect of CMV, as opposed to the more marginal contribution of CMV infection to increased TCRγδ+ T cell exhaustion in elderly.

MATERIALS & METHODS

Study subjects

The NWO ageing study cohort consisted of immunologically healthy patients from the Orthopedics outpatient clinic, Erasmus MC, complemented with (immunologically) healthy controls in the younger age groups. After applying exclusion criteria (auto-immune or –inflammatory diseases at present or in the past; malignancies; usage of anti-inflammatory or immunosuppressive drugs; surgery in the past 30 days; alcohol or drug abuse) a total of 121 subjects were included. To increase the number of subjects in especially the age groups of >60 years, an additional 36 subjects participating in the SENEX study of healthy elderly in the Dutch region of Groningen were included via the UMC Groningen. Participants in the NWO Ageing Study gave written informed consent and the study was approved by the Medical Ethics Committee of the Erasmus MC under number MEC-2011-409 and MEC-2016-202. From subjects participating in the SENEX Study of the UMC Groningen written informed consent was obtained and approval for the study was provided by the Medical Ethics Committee of the UMCG under protocol number 2012375. All experimental studies were conducted in accordance with relevant guidelines and principles of the Declaration of Helsinki. Samples were divided into five age groups, 40 to 50 (mean age 45), 50 to 60 (mean age 55), 60 to 70 (mean age 66), 70-plus (mean age 77) versus a control group of samples from healthy adults age 20 to 40 (mean age 25) (Table 1). Fresh peripheral blood mononuclear cells (PBMC) samples of a total of 157 participants of the NWO and SENEX studies were analyzed after lysis with ammonium chloride. CMV serostatus was determined on plasma with the use of the anti-CMV ELISA (IgG) according to the manufacturer's protocol (EuroImmun, Lübeck, Germany). Remaining peripheral blood after analysis and plasma storage was subjected to Ficoll-Paque (density 1.077 g/ml, Pharmacia, Uppsala, Sweden) density gradient separation and cryopreserved in Iscove's Modified Dulbecco's Medium (IMDM, Lonza, Basel, Switzerland) with dimethyl sulfoxide in vials at -180 ^{0}C until further use.

Flow cytometric immune phenotyping

Freshly obtained blood was lysed with ammonium chloride and washed with phosphate buffered saline (PBS) pH 7.8 containing fetal bovine serum (FBS, 30% w/v) and sodium azide. Samples were stained using three antibody panels according to Supplementary Table 1. Data analysis and gating strategies were based on the standardized protocols from the Generation R study [70]. Defining viable cells was based on FSC/SSC gating strategies and validated with negative expression of Annexin V of viable cells in a small series of additional samples (data not shown). With the use of tube 1 Vγ- and Vδ-usage could be determined. With the use of T cell maturation markers CD45RO,

Table 1. Age group characteristics of study subjects.

	Controls (20-40)	40-50	50-60	60-70	>70
Total number	N=30	N=24	N=29	N=40	N=34
Age (mean ± SD)	25.3 (4.2)	45.2 (2.6)	55.1 (2.6)	65.8 (2.5)	76.9 (5.5)
Range	20-40	40-49	51-59	61-69	70-95
Males (n; %)	11 (36.7%)	10 (41.7%)	11 (37.9%)	14 (35%)	12 (40%)
CMV- (n; %)	6 (54.5%)	8 (80%)	7 (63.6%)	7 (50.0%)	4 (33.3%)
CMV+ (n; %)	5 (45.5%)	2 (20%)	4 (36.3%)	7 (50.0%)	8 (66.7%)
Females (n; %)	19 (63.3%)	14 (58.3%)	18 (62.1%)	26 (65%)	22 (60%)
CMV- (n; %)	15 (78.9%)	8 (57.1%)	10 (55.6%)	9 (34.6%)	7 (31.8%)
CMV+ (n; %)	4 (21.1%)	6 (42.9%)	8 (44.2%)	17 (65.4%)	15 (68.2%)

Values are means (SD) and absolute numbers (percentages). Percentages of CMV negative and positive individuals are from total males or females.

CD197, CD27 and CD28 in the second tube maturation and differentiation statuses of TCRγδ+ T cell populations could be determined: naive (CD45RO-CD197+), central memory (CD45RO+CD197+), circulating effector memory (CD45RO+CD197-, also known

as TemRO cells), and effector (CD45RO-CD197-, also known as TemRA cells) TCRγδ+ T cells. Furthermore, CD27 and CD28 were used for subdivision into early-, intermediate- and late-stage differentiated effector memory and effector cells. Tube 3 included markers for the evaluation of exhaustion profiles. Cells were acquired using the Fortessa LSR flow cytometer (BD Biosciences, San Jose, CA, USA). Compensation was based on single color controls. Data were analyzed with FACSDiva software (BD Biosciences). The gating strategy applied for the various tubes is displayed in Supplementary Figure 1. The absolute cell count per microliter of a particular TCRγδ+ T cell population was calculated based on the total percentage and the total absolute lymphocyte count. The latter was calculated with the use of the number of events in the lymphocyte gate using the TruCount tube (BD Biosciences) and the Canto II flow cytometer (BD Biosciences) (Supplementary Fig. 1a), the initial white blood cell count and the number of leukocyte events.

Statistical analysis

The non-parametric one-way ANOVA Kruskal-Wallis test was performed to compare absolute numbers and frequencies between different age groups and between young and elderly CMV- and CMV+ groups on single immune subsets. The Dunn's test was applied for correction for multiple testing. P-values of <0.05 were considered statistically significant. The statistical analyses were performed in Prism 5 (GraphPad, La Jolla, CA, USA).

RESULTS

Clear decline in absolute numbers of the most common Vγ9/Vδ2 TCRγδ+ T cell subset in peripheral blood of elderly subjects

When studying absolute numbers of TCRγδ expressing T cells, a significant decrease was observed with ageing, which was already apparent at age group 40-50 (Fig. 1a). In contrast, absolute numbers of TCRαβ+ T cells were hardly or not affected, although variation was high in elderly (Supplementary Fig. 2a). As a consequence the overall distribution of TCRαβ versus TCRγδ expressing T cells showed a significant increase in TCRαβ+ T cell frequencies and a significant decrease in TCRγδ+ T cell frequencies, again already at age group 40-50 (Supplementary Fig. 2b). When focusing more on subsets with specific Vδ receptor usage no significant differences in the absolute numbers of Vδ1+ cells were found (Fig. 1b). However the significant decrease in total TCRγδ+ T cells was rather paralleled by a significant decrease in Vδ2+ cells (Fig. 1c), and especially Vγ9/Vδ2 cell populations (Fig. 1d). To determine whether this resulted in a clear shift in Vδ

usage within the total TCRγδ+ T cell population, we then compared the distributions of Vδ1, Vδ2 and other Vδ (non-Vδ1, non-Vδ2) populations in the peripheral blood of all age groups. We found no significant alterations, although the percentage of Vδ2+ cells was decreased in age groups 40-50, 50-60 and 60-70, with a shift towards relatively more Vδ1+ cells (Supplementary Fig. 2c). Overall these data suggest a significant decrease in absolute numbers of TCRγδ+ T cells in elderly age groups, with a parallel decrease in numbers of the most dominant Vγ9/Vδ2 TCRγδ+ T cell population.

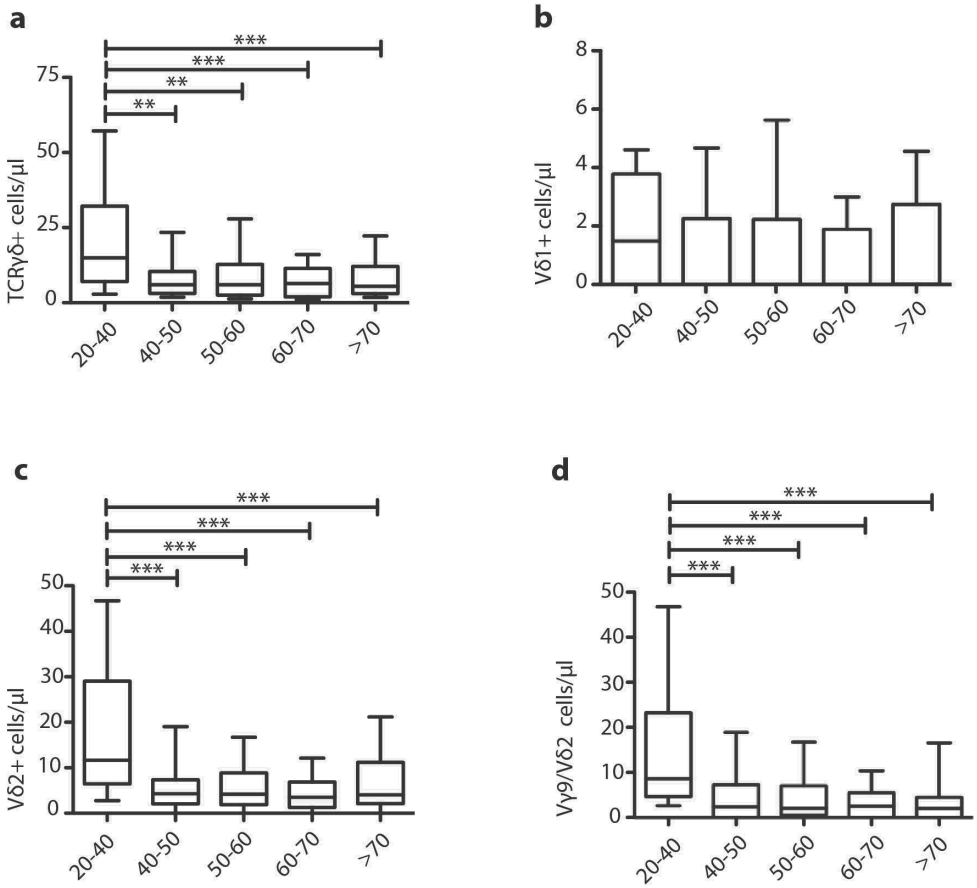


Figure 1. Vγ/Vδ-gene usage in TCRγδ+ T cells in different age groups.
(**a**) Absolute numbers of total TCRγδ+ T cells, (**b**) Vδ1+, (**c**) Vδ2+ and (**d**) Vγ9/Vδ2 TCRγδ+ T cells depicted as 10-90% box-whisker-plots. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test to correct for multiple testing. Significance for the Dunn's test is indicated in the plots: **, $p<0.01$; ***, $p<0.001$.

During ageing the naive TCRγδ+ T cell compartment shrinks and shifts towards a late-differentiated effector phenotype

In analogy to the effects described for CD8+ cytotoxic T lymphocytes (CTL), we next investigated maturation and differentiation of TCRγδ+ T cells. From age 50 onwards slight shifts in the distribution of double-negative (DN), CD4 single-positive (SP), CD8 SP, and double-positive (DP) cells were visible, mostly affecting the predominant DN and CD8 SP compartments (Fig. 2a). TCRαβ+ T cells also showed significant differences in the CD4/CD8 distributions upon ageing (Supplementary Fig. 2d), with a clear shift in the CD4/CD8 ratio towards more CD4+ T cells (Supplementary Fig. 2e), as described before [47-50]. Even though in TCRγδ+ T cells CD4/CD8 ratios have a completely different meaning, given that TCRγδ+ T cells usually do not express CD4 and the CD8αα dimer upon activation [38], we still checked these ratios and did not observe significant changes (Fig. 2b).

Of note, earlier documented changes in CD8+TCRαβ+ CTL maturation were also observed in our cohort, with significant decreases in the naive and significant increases in effector CD8+TCRαβ+ T cell compartments (Supplementary Fig. 2f), whilst CD4+TCRαβ+ T cells did not show similar significant changes in these maturation subsets (Supplementary Fig. 2g). These TCRαβ+ T cell results thus validate our dataset as being representative for investigating immunological ageing of TCRγδ+ T cells. Therefore, next we determined maturation subsets of the DN and CD8+TCRγδ+ T cells. In general, for the DN TCRγδ+ T cell population no significant differences in absolute numbers of naive (CD45RO-CD197+) (Fig. 2c) or central memory (CD45RO+CD197+) cells (Fig. 2d) were found. In contrast, in the effector memory (CD45RO+CD197-) (Fig. 2e) and effector (CD45RO-CD197-) populations (Fig. 2f) numbers decreased significantly, with effector memory cell numbers decreasing already in the age group 40-50 and effector cell numbers decreasing mainly in age groups 50-60 and 60-70. For CD8+TCRγδ+ T cells, only in the effector memory population a significant difference in absolute numbers was observed (Fig. 2h-k). Notably, when further studying relative distributions of these maturation subsets, which in addition to the cell numbers could reflect biologically relevant shifts in subset composition, significant differences were observed for both DN and CD8+TCRγδ+ T cells. These concerned decreases in the naive subset fractions and increases in the effector subset fractions (Fig. 2g, 2l). This was especially true in the oldest age group (>70), although decreasing (naive) and increasing (effector) trends were in fact already visible from age 50 onwards.

As TCRγδ+ effector T cells are known to have a rather late-stage differentiated phenotype [37], we then further focused on early (CD27+CD28+), intermediate (CD27+CD28-) and late (CD27-CD28-) subpopulations. This analysis showed significant decreases in absolute numbers of early and intermediate DN TCRγδ+ effector cells starting from age

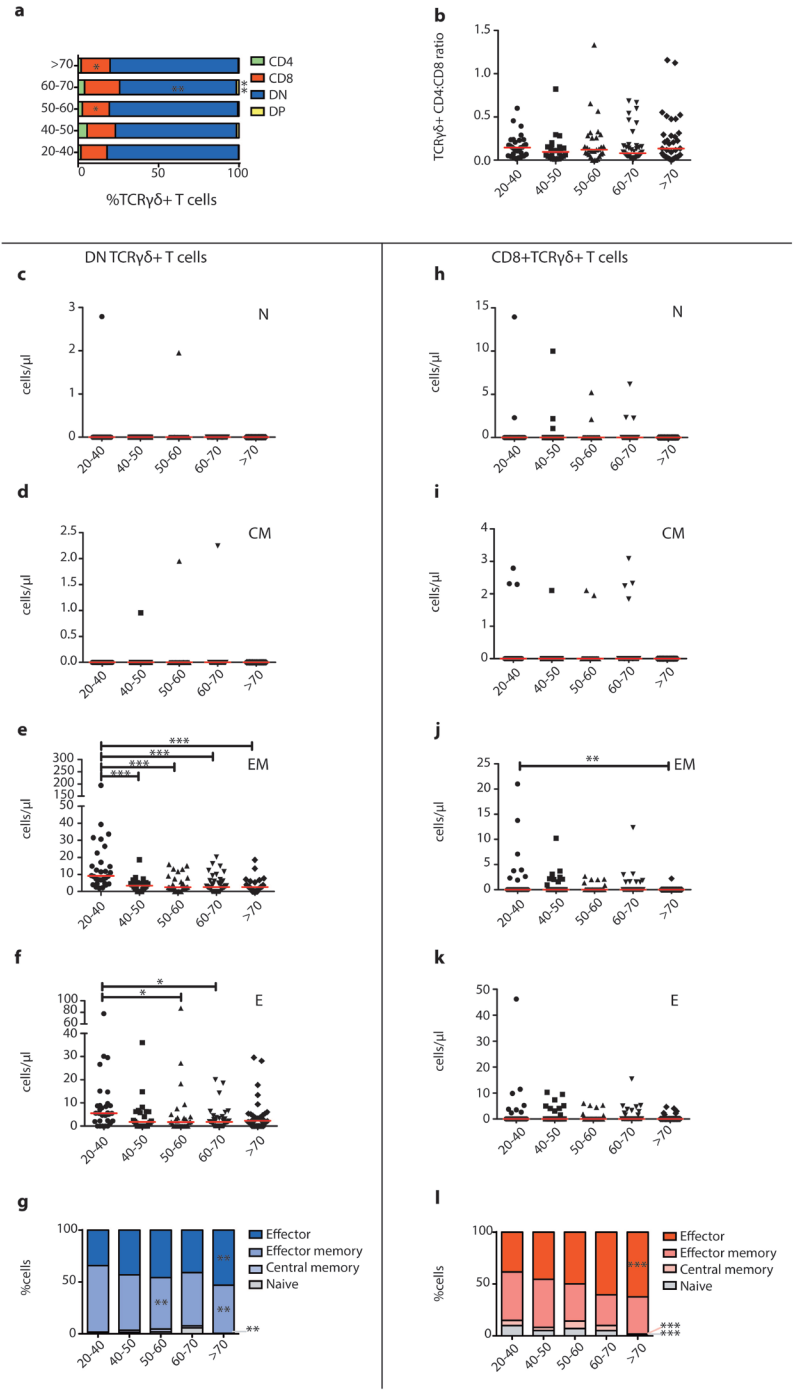

a
>70
60-70
50-60
40-50
20-40
CD4
CD8
DN
DP
0
50
100
%TCRγδ+ T cells
b
TCRγδ+ CD4:CD8 ratio
1.5
1.0
0.5
0.0
20-40
40-50
50-60
60-70
>70
DN TCRγδ+ T cells
CD8+TCRγδ+ T cells
c
N
cells/µl
h
N
cells/µl
d
CM
cells/µl
i
CM
cells/µl
e
EM
cells/µl
j
EM
cells/µl
f
E
cells/µl
k
E
cells/µl
g
%cells
Effector
Effector memory
Central memory
Naive
l
%cells
Effector
Effector memory
Central memory
Naive

Figure 2 (see left page). TCRγδ+ T cell maturation statuses and subset distributions during ageing. (**a**) Relative visualization of CD4 and CD8 single-positive, double-positive (DP, CD4+CD8+) and double-negative (DN, CD4-CD8-) distribution within the total TCRγδ+ T cell compartment between different age groups. (**b**) CD4:CD8 ratios in total TCRγδ+ T cell population. (**c**) Absolute numbers of DN naive (CD45RO-CD197+), (**d**) central memory (CD45RO+CD197+), (**e**) effector memory (CD45RO+CD197-, TemRO) and (**f**) effector (CD45RO-CD197-, TemRA) TCRγδ+ T cells. (**g**) Relative distributions of maturation subsets of DN TCRγδ+ T cells depicted in stacked bar plots. (**h**) Absolute numbers of CD8+ naive, (**i**) central memory, (**j**) effector memory and (**k**) effector CD8+ TCRγδ+ T cells. (**l**) Relative distributions of maturation subsets of CD8+ TCRγδ+ T cells depicted in stacked bar plots. Ratios and absolute numbers are indicated in scatter plots indicated with the median. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test for correction for multiple testing. Significance for the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01, ***, p<0.001.

group 40-50, but not of late-stage differentiated effector cells (Fig. 3a). Given that also the effector memory cells were found to relatively expand in the aged groups (Fig. 2g, 2l), we additionally looked into early, intermediate and late differentiated cells within the effector memory subset. Significant decreases in early and intermediate DN TCRγδ+ effector memory cell numbers were found when all age groups were compared with the 20-40 age control group. Also a significant decrease in the absolute numbers of late-stage differentiated cells was observed, mainly when the oldest age group was compared with the control group (Fig. 3b). CD8+TCRγδ+ cells showed no significant differences in early, intermediate or late stages, except for a decrease in absolute numbers of late-differentiated effector memory cells in the oldest age group (Fig. 3c, 3d).

Since absolute numbers of total TCRγδ+ T cells generally decreased (Fig. 1a), and the maturation subsets displayed clear shifts in distribution (Fig. 2g, 2l), we also investigated the relative shifts of stages within the effector and effector memory cell populations as discussed above. Significant decreases were observed in early and intermediate DN TCRγδ+ effector cell proportions with a concomitant significant increase in percentages of late-stage differentiated cells in especially the >70 age group when all age groups were compared to the control age group (Fig. 4a); this was rather reversed in the effector memory cells where an increase in the proportions of early differentiated cells was observed (Fig. 4b). Furthermore, even though in CD8+TCRγδ+ effector T cells no significant changes in absolute numbers were observed (Fig. 3c), their relative distributions did show significant decreases in especially intermediate effector cells, concurrent with a significant increase in late-stage differentiated cells (Fig. 4c) at age 50-60. Within the CD8+TCRγδ+ effector memory population a significant decrease in the proportions of intermediate differentiated cells was observed (Fig. 4d).

Taken together, we conclude that ageing has a similar effect on TCRγδ+ maturation subsets as was reported for CD8+TCRαβ+ CTL. Despite an overall decrease in TCRγδ+ cell numbers, the relative increase in effector cells and the shift towards a late-stage differentiated phenotype result in stable numbers of the most differentiated effector cell population.

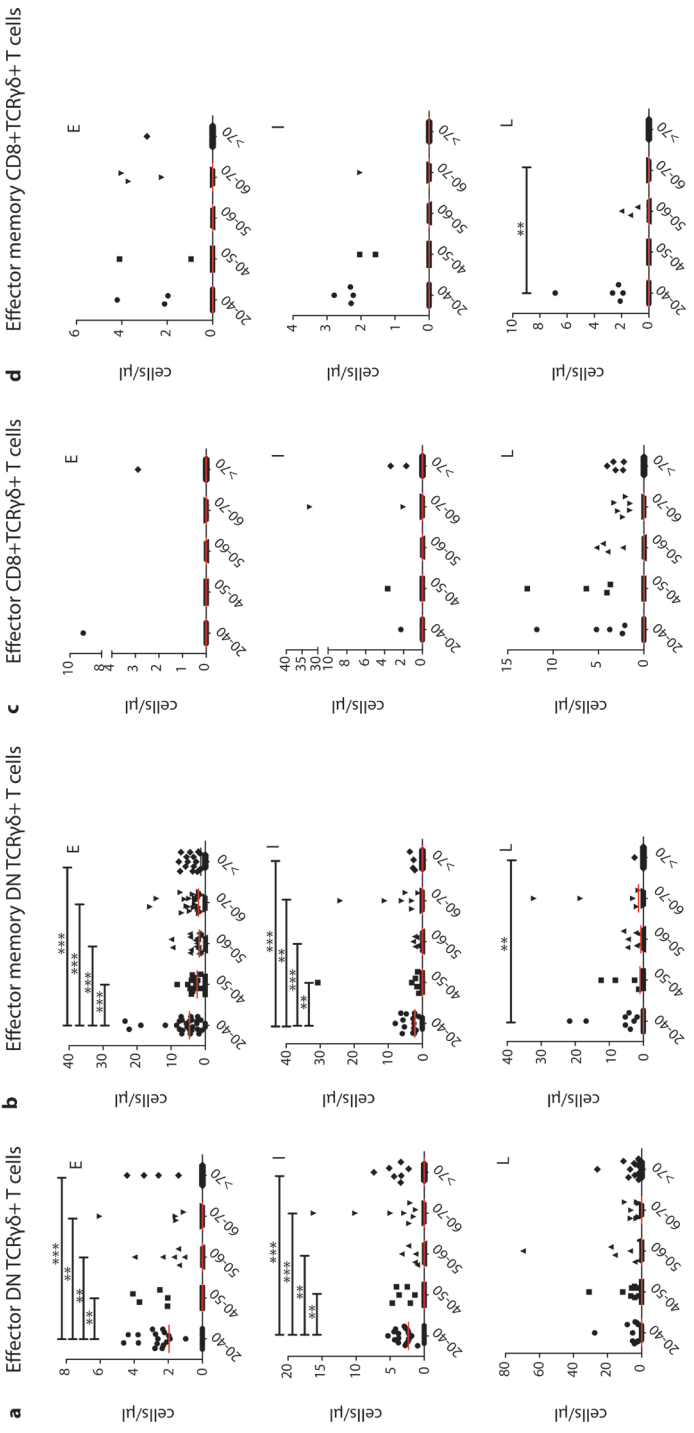

a
Effector DN TCRγδ+ T cells
b
Effector memory DN TCRγδ+ T cells
c
Effector CD8+TCRγδ+ T cells
d
Effector memory CD8+TCRγδ+ T cells
E
I
L
cells/µl
20-40
40-50
50-60
60-70
>70

Figure 3 (see left page). DN and CD8+TCRγδ+ effector and effector memory differentiation stages during ageing.
(**a**) Absolute numbers of early (CD27+CD28+), intermediate (CD27+CD28-) and late (CD27-CD28-) differentiated DN and (**c**) CD8+TCRγδ+ effector (CD45RO-CD197-, TemRA) T cells. (**b**) Absolute numbers of early, intermediate and late differentiated DN and (**d**) CD8+ TCRγδ+ effector memory (CD45RO+CD197-, TemRO) T cells. Early (E), intermediate (I) and late (L) definitions are indicated in the upper right corners of the graphs. Scatter plots are indicated with the median. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test for correction for multiple testing. Significance for the Dunn's test is indicated in the plots: **, p<0.01; ***, p<0.001.

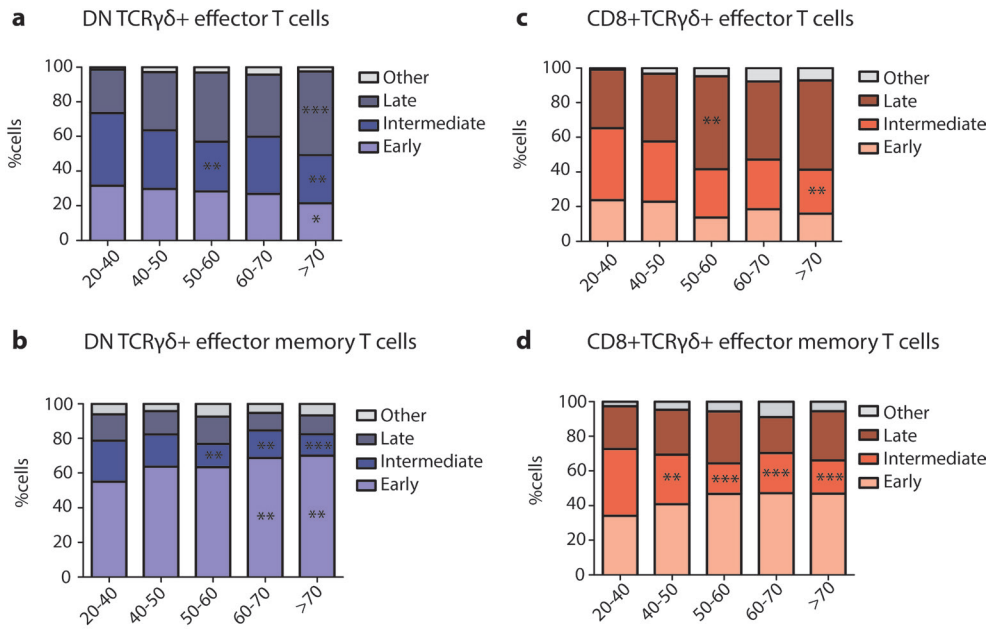


Figure 4. DN and CD8+ TCRγδ+ effector and effector memory differentiation stage distributions during ageing.
(**a**) Relative differentiated subset distributions of percentages of DN and (**c**) CD8+ effector and (**b**) DN and (**d**) CD8+ effector memory TCRγδ+ T cells depicted in stacked bar plots. Significance was tested by a Kruskal-Wallis test, and followed by a post-Dunn's test for correction for multiple testing. Data of different age groups was compared with the control age group. Significance for the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01; ***, p<0.001.

CMV-seropositivity impacts on Vδ-usage at old age

Persistent viruses and especially CMV are known to have major effects on the composition, senescence, and exhaustion of the immune system [13,43,45,46,51]. We therefore studied the potential impact of CMV on immunological ageing of TCRγδ+ T cells. To this end we subdivided our study cohort, according to CMV serology. Furthermore, we also looked at gender as a potential confounding factor for immunological ageing. Although the proportion of CMV-seropositive individuals was higher with age in both males and females, these percentages were not significantly different in any age group

(Supplementary Fig. 3); in fact the CMV seroprevalence of our age groups correlated well with previous reports [52,53]. Recently it was shown that gender and additionally CMV infection were associated with an expansion of late-stage differentiated TCRαβ+ T cell subsets and a reduction of naive, regulatory and CD8+ T cells in especially middle-aged (age category 50-65) males [54]. As we did not observe a gender effect (Supplementary Fig. 4) or specific differences in gender and CMV serology in the middle-aged 50-65 group (data not shown) with respect to TCRγδ+ T cells in our cohort, we further focused our analyses on age and CMV serology only. In order to be able to make clear distinctions, we defined groups of young controls (age 20-40) and elderly individuals (above age 60) (Table 1), and subdivided both groups into seronegative (CMV-) and seropositive (CMV+) subjects.

First, total TCRγδ+ T cell absolute counts were compared between young vs. elderly, and CMV- vs. CMV+ groups, which showed a significantly higher total TCRγδ+ T cell count in young CMV+ individuals (Fig. 5a). In elderly the absolute numbers of total TCRγδ+ T cells were decreased, without a significant additional effect of CMV (Fig. 5a). Of note, TCRαβ+ T cell counts were not significantly affected in these subgroups (Supplementary Fig. 2h). The relative distributions of TCRαβ+ and TCRγδ+ T cells were also determined, showing significant alterations in elderly; particularly in CMV+ elderly the percentage of TCRγδ+ T cells was significantly reduced, with a parallel increase of the TCRαβ+ T cell fraction (Supplementary Fig. 2i). Furthermore, upon evaluation of Vγ/Vδ-usage, changes in absolute numbers of Vδ2+ and Vγ9/Vδ2 populations were largely similar as for total TCRγδ+ T cells, whilst absolute numbers of Vδ1+ cells were increased in both young and old CMV+ individuals (Fig. 5c, 5d). The overall Vδ-usage distribution showed significant changes in the composition in especially CMV-infected elderly, with an increased proportion of Vδ1+ and a decreased proportion of Vδ2+ cells (Supplementary Fig. 2j). The increase in CMV-seroprevalence in elderly (Supplementary Fig. 3) and the relative increase in Vδ1-usage (Supplementary Fig. 2j) in both elderly and young CMV+ individuals correlate with known Vδ1+ cell reactivity to CMV [55].

Collectively, these data suggest that CMV has a profound effect on the numbers of all TCRγδ+ T cells in young infected individuals, and that in elderly the impact of CMV is more distinct TCRγδ+ subgroups showing different Vδ-usage.

The shift towards an effector TCRγδ+ T cell phenotype in elderly is largely explained by CMV infection

Next we also investigated the impact of CMV on maturation phenotypes of TCRγδ+ T cells. Firstly, no significant alterations were seen in the frequencies of the predominant DN and CD8+TCRγδ+ T cell populations (Fig. 6a) or in the CD4/CD8 ratios of TCRγδ+ T cells (Fig. 6b), both in the absence and presence of CMV. In contrast, CD4, CD8, DN, DP

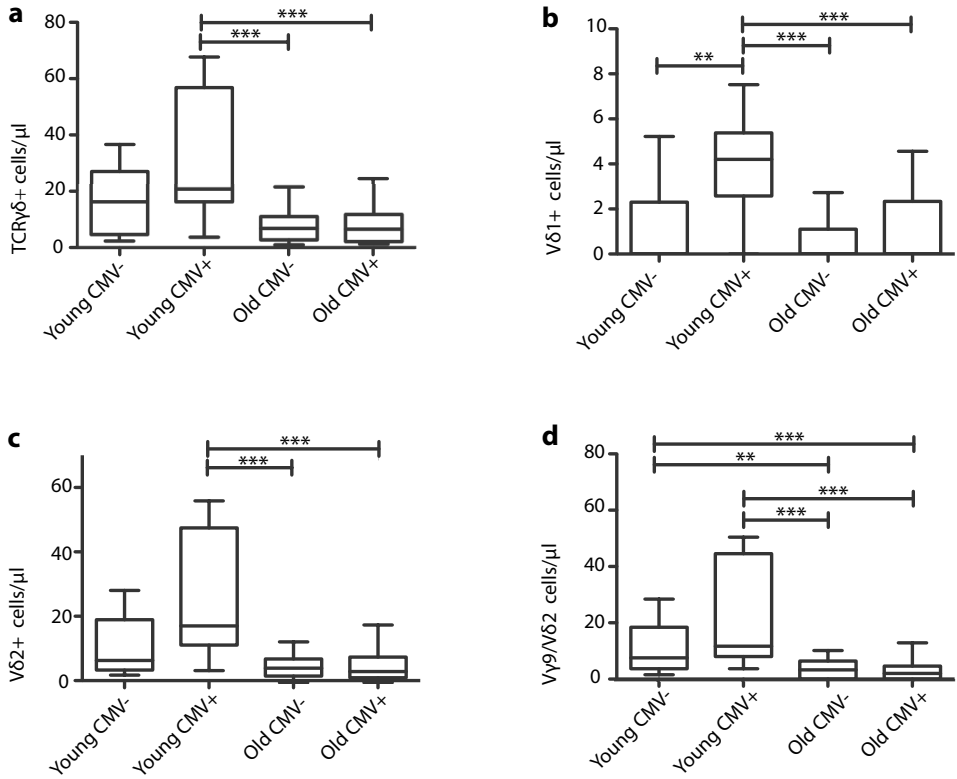


Figure 5. Effect of CMV seropositivity on Vγ/Vδ-usage.
(**a**) Absolute numbers of total TCRγδ+ T cells, (**b**) Vδ1+, (**c**) Vδ2+ and (**d**) Vγ9/Vδ2 TCRγδ+ T cells depicted as 10-90% box-whisker-plots. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test to correct for multiple testing. Significance for the Dunn's test is indicated in the plots: **, p<0.01; ***, p<0.001.

populations in TCRαβ+ T cells were significantly different, mainly in the elderly CMV- group (Supplementary Fig. 2k), with a shift in the CD4/CD8 ratio towards more CD4+ T cells (Supplementary Fig. 2l). In keeping with published data, clear changes were seen in the relative proportions of different maturation stages in the CD4+ and CD8+TCRαβ+ T cells upon CMV (Supplementary Fig. 2m, 2n), thus reinforcing the validity of our cohort for studying TCRγδ+ T cells.

When focusing on the absolute numbers of naive and central memory DN TCRγδ+ cells no significant changes were observed (Fig. 6c, 6d). In contrast, we did find significant changes in effector memory and effector cell absolute counts, with increased numbers being present in especially the young CMV+ group (Fig. 6e, 6f). In the CD8+TCRγδ+ T cell population we did not observe significant changes in the naive and central memory subsets either (Fig. 6h, 6i), whilst again the young CMV+ group showed higher effector

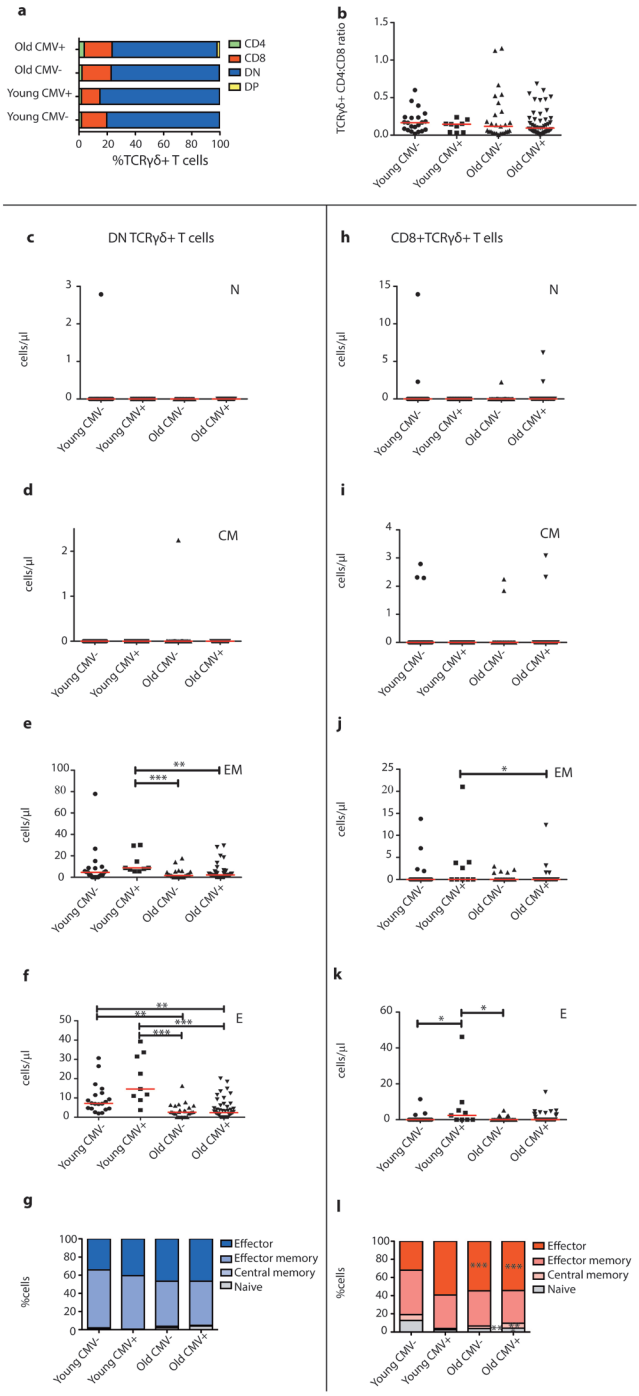

a
Old CMV+
Old CMV-
Young CMV+
Young CMV-
CD4
CD8
DN
DP
0 20 40 60 80 100
%TCRγδ+ T cells
b
TCRγδ+ CD4:CD8 ratio
1.5
1.0
0.5
0.0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
c
DN TCRγδ+ T cells
N
cells/µl
3
2
1
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
h
CD8+TCRγδ+ T ells
N
cells/µl
15
10
5
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
d
CM
cells/µl
2
1
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
i
CM
cells/µl
4
3
2
1
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
e
EM
**
***
cells/µl
100
80
60
40
20
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
j
EM
*
cells/µl
25
20
15
10
5
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
f
E
**
**
***
***
cells/µl
50
40
30
20
10
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
k
E
*
*
cells/µl
60
40
20
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
g
%cells
100
80
60
40
20
0
Effector
Effector memory
Central memory
Naive
Young CMV-
Young CMV+
Old CMV-
Old CMV+
l
%cells
100
80
60
40
20
0
***
***
Effector
Effector memory
Central memory
Naive
Young CMV-
Young CMV+
Old CMV-
Old CMV+

Figure 6 (see left page). Effect of CMV on TCRγδ+ T cell maturation.
(**a**) Relative visualization of CD4 and CD8 single-positive, double-positive (DP, CD4+CD8+) and double-negative (DN, CD4-CD8-) distribution within the total TCRγδ+ T cell compartment between different age and CMV groups. (**b**) CD4:CD8 ratios in total TCRγδ+ T cell population. (**c**) Absolute numbers of DN naive (CD45RO-CD197+), (**d**) central memory (CD45RO+CD197+), (**e**) effector memory (CD45RO+CD197-, TemRO) and (**f**) effector (CD45RO-CD197-, TemRA) TCRγδ+ T cells. (**g**) Relative distributions of maturation subsets of DN TCRγδ+ T cells depicted in stacked bar plots. (**h**) Absolute numbers of CD8+ naive, (**i**) central memory, (**j**) effector memory and (**k**) effector CD8+ TCRγδ+ T cells. (**l**) Relative distributions of maturation subsets of CD8+ TCRγδ+ T cells depicted in stacked bar plots. Ratios and absolute numbers are indicated in scatter plots indicated with the median. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test for correction for multiple testing. Significance for the Dunn's test is indicated in the plots: *, $p<0.05$; **, $p<0.01$, ***, $p<0.001$.

memory and effector cell counts (Fig. 6j, 6k) in keeping with the significant increase in total TCRγδ+ T cells in young CMV+ individuals (Fig. 5a). In order to further investigate the biological impact of both ageing and CMV we then also focused on the relative subset distributions. We did not find significant differences in the subset distribution of DN TCRγδ+ T cells, despite an increasing trend in the effector population (Fig. 6g). However, the relative distributions of CD8+TCRγδ+ T cell maturation subsets did significantly alter. Especially upon the presence of CMV, the percentages of effector cells were increased, with a concomitant decrease in the naive compartment (Fig. 6l). The subset distribution pattern of young CMV+ individuals reflected that of elderly (Fig. 6l).

When looking more in-depth into the differentiation stages within effector and effector memory cells, the absolute numbers of early and intermediate effector DN TCRγδ+ T cells increased, whilst late effector cells showed a significant increase in predominantly the young CMV+ individuals (Fig. 7a). The effector memory DN TCRγδ+ T cells showed increased numbers in all differentiation stages when it comes to ageing in the presence of CMV, although this was not significant for the early differentiated cells (Fig. 7b). In CMV+ elderly the absolute numbers of CD8+TCRγδ+ effector and effector memory cells were significantly higher in almost all differentiation stages, except for intermediate effector memory cells (Fig. 7c, 7d).

As we observed in all differentiation stages an increase in absolute numbers in elderly CMV+ individuals, we then further looked into the relative composition. When evaluating the overall distribution patterns, comparing all groups with each other, young and old CMV- individuals were showing very similar distribution patterns, as well as young and old CMV+ individuals (Fig. 8). In the presence of CMV – in both young and elderly – percentages of DN TCRγδ+ intermediate effector cells decreased, whilst those of late effector cells increased (Fig. 8a). This effect was similar and even more pronounced for the CD8+TCRγδ+ effector population (Fig. 8c). When investigating effector memory subpopulations, analogous to what was seen in elderly age groups, significant increases in absolute numbers and relative distributions were noted, mainly in CMV+ elderly (Fig. 8b, 8d).

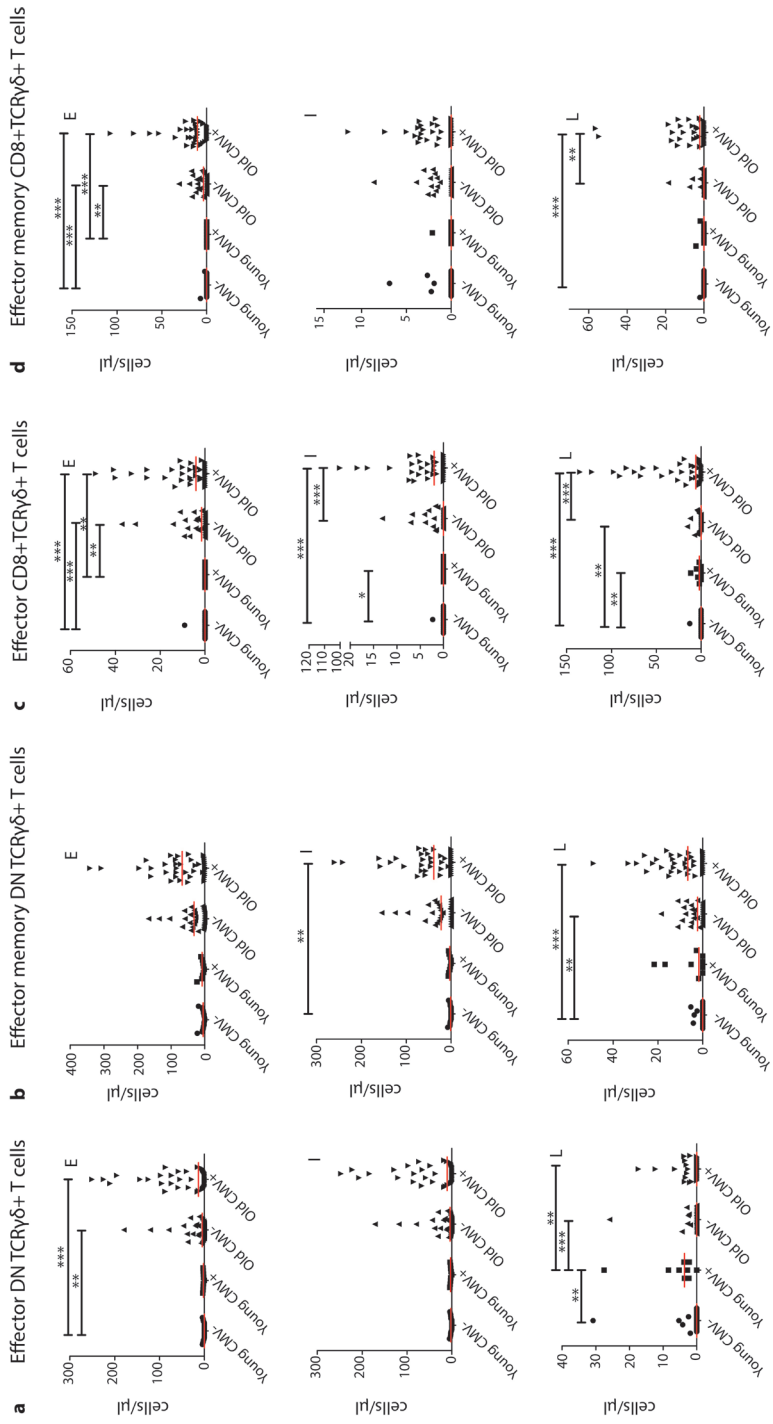

a
Effector DN TCRγδ+ T cells
b
Effector memory DN TCRγδ+ T cells
c
Effector CD8+TCRγδ+ T cells
d
Effector memory CD8+TCRγδ+ T cells
E
I
L
cells/µl
Young CMV-
Young CMV+
Old CMV-
Old CMV+

Figure 7 (see left page). Effect of CMV on effector and effector memory differentiation stages.
(**a**) Absolute numbers of early (CD27+CD28+), intermediate (CD27+CD28-) and late (CD27-CD28-) differentiated DN and (**c**) CD8+TCRγδ+ effector (CD45RO-CD197-, TemRA) T cells. (**b**) Absolute numbers of early, intermediate and late differentiated DN and (**d**) CD8+ TCRγδ+ effector memory (CD45RO+CD197-, TemRO) T cells. Early (E), intermediate (I) and late (L) definitions are indicated in the upper right corners of the graphs. Scatter plots are indicated with the median. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test for correction for multiple testing. Significance for the Dunn's test is indicated in the plots: **, p<0.01; ***, p<0.001.

Altogether these data clearly indicate that the presence of CMV greatly impacts on the absolute numbers of differentiated effector and effector memory populations, as well as induces shifts in maturation subset compositions of TCRγδ+ T cells similar to what is seen in elderly, i.e. a main shift towards effector and effector memory phenotypes, showing more late-staged differentiated effector cells and early-stage differentiated effector memory cells.

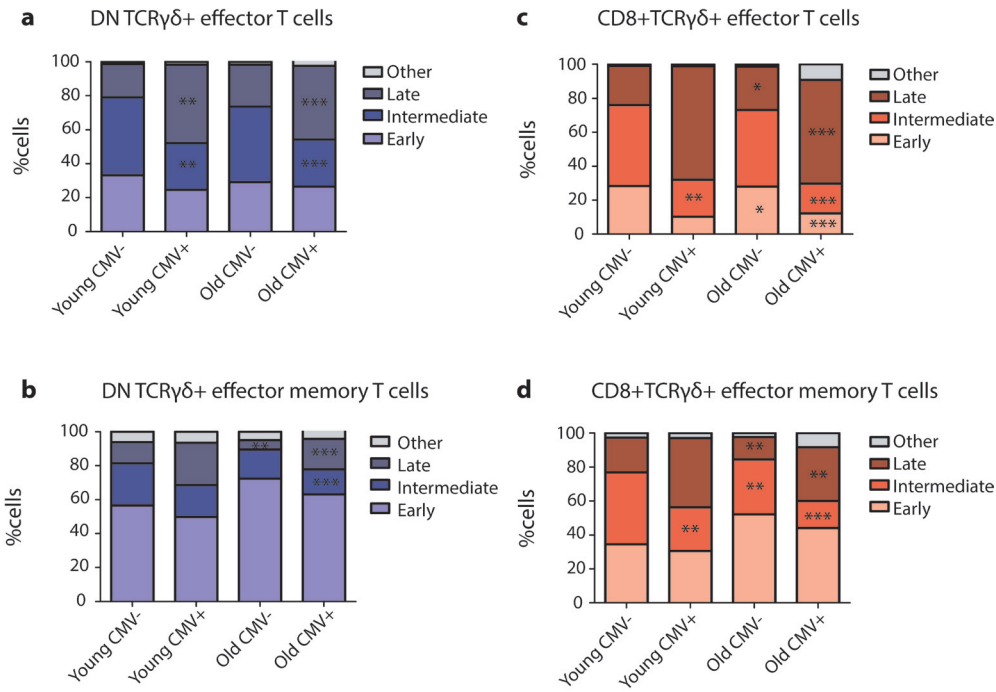


Figure 8. DN and CD8+ TCRγδ+ effector and effector memory differentiation stage distributions during ageing and in absence or presence of CMV.
(**a**) Relative differentiated subset distributions of percentages of DN and (**c**) CD8+ effector and (**b**) DN and (**d**) CD8+ effector memory TCRγδ+ T cells depicted in stacked bar plots. Significant differences between all groups was tested by a Kruskal-Wallis test, and followed by a post-Dunn's test for correction for multiple testing. Significance for the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01; ***, p<0.001.

CMV infection marginally contributes to the increased exhaustion profile of TCRγδ+ T cells in elderly

T cell exhaustion is a phenomenon that is often seen in persistent viral infections like CMV [13,14,56], and that largely shapes the CD8+TCRαβ+ T cell compartment of the immune system [42,46,51]. Within our cohort we also observed increased absolute numbers (Supplementary Fig. 5) and percentages (Supplementary Fig. 6) of CD8+ non-TCRγδ (TCRαβ) T cells expressing exhaustion markers. As the level of exhaustion of TCRγδ+ T cells during ageing and upon the presence of persistent viral infections like CMV has not been properly documented, we investigated absolute numbers of total TCRγδ+ T cells expressing or lacking the senescence and exhaustion associated markers. We observed increased absolute numbers of KLRG1- (Fig.9a), FAS+ (Fig. 9b), CD57+ (Fig. 9d), PD1+ (Fig. 9e) and IL7Rα- (Fig. 9f) TCRγδ+ T cells especially in the context of CMV, in both young and elderly. 2B4+ TCRγδ+ T cells were also increased in case of young CMV+ individuals, although this was not significant (Fig. 9b). Since senescence and exhaustion processes are difficult to separate, and since there are no concrete definitions, we also looked into combinations of different markers. IL7Rα is lost already during early stages of both exhaustion and senescence, and therefore we studied the marker combinations within the IL7Rα- TCRγδ+ T cell population (Fig. 9g-i). We observed only a significant increase in IL7Rα-KLRG1-CD57+ TCRγδ+ T cells, mainly in the young CMV+ population (Fig. 9i).

In view of our findings of increased TCRγδ+ T cell numbers in young but not old CMV+ individuals (Fig. 5a), we considered an exhaustion phenotype in especially elderly and thus looked for the fractions of TCRγδ+ T cells showing exhaustion markers. When analyzing the percentages increasing trends could be appreciated, however there were no significant differences observed, although the variation among younger individuals was higher when compared to elderly, independent of CMV infection (Supplementary Fig. 7).

Overall, these data suggest that immunological ageing does contribute to a more increased exhaustion phenotype of TCRγδ+ T cells, and that CMV plays an additional role.

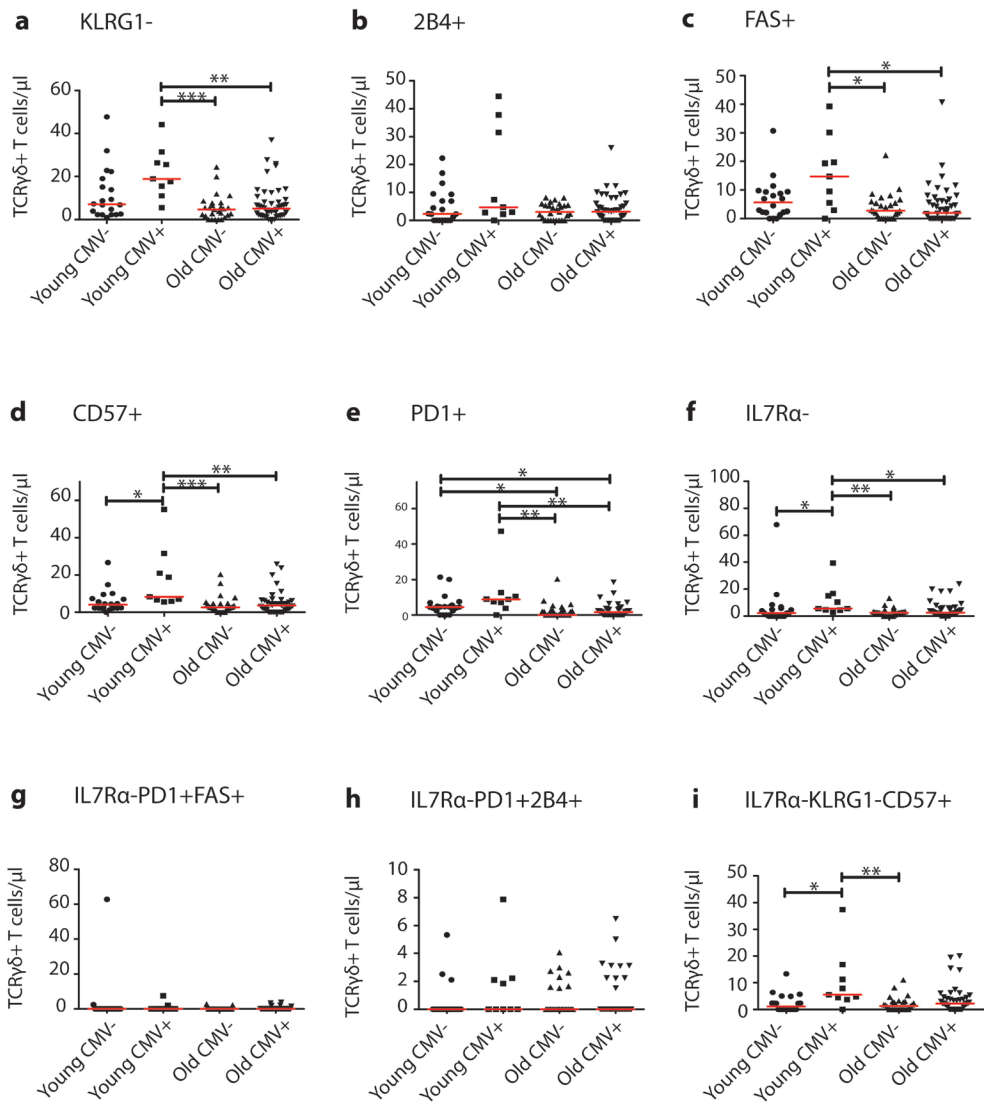


Figure 9. Effect of CMV on exhaustion and senescence profiles of TCRγδ+ T cells.
(**a**) Absolute numbers of TCRγδ+ T cells lacking KLRG1 expression, (**b**) TCRγδ+ T cells expressing 2B4, (**c**) FAS death receptor, (**d**) CD57, (**e**) PD1, (**f**) TCRγδ+ T cells lacking IL7Rα, (**g**) and IL7Rα- TCRγδ+ T cells co-expressing PD1 and FAS, (**h**) PD1 and 2B4, and (**i**) expressing CD57 and lacking KLRG1. Scatter plots are indicated with the median. Significance was tested by a Kruskal-Wallis test, and followed by a post-Dunn's test for correction for multiple testing. Significance for the Dunn's test is indicated in the plots: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

DISCUSSION AND CONCLUSION

With increasing hygiene and improved health care individuals in the Western world become significantly older according to the World Health Organization [57,58]. Immunological ageing, also referred to as immunosenescence, has large impacts on elderly individual's health as evidenced from less efficient responses to infectious agents [11], reduced vaccine responses [5,6], and increased risks of developing cancer due to insufficient anti-tumor activity of the immune system [59]. Immunosenescence has been described at different levels in the immune system with the most well-defined effect of ageing concerning subset changes in CD4+ and CD8+TCRαβ+ T cells [47-50]. T cell exhaustion, characterized by a stepwise loss of effector functions, plays a major role in immunosenescence [13]. This has mainly been investigated in the TCRαβ+ T cell compartment as well, most notably affecting CD8+ T cells. We hypothesized that ageing has similar effects on the TCRγδ+ T cell population, which could largely affect the elderly individuals due to the reduced or aberrant response of TCRγδ+ T cells to pathogens. TCRγδ+ T cells implement different functions from both innate and adaptive immunity by readily responding to antigens in both CD8+ CTL and NK cell like manners [33]. The majority of our data presented here is based on absolute counts, in order to determine direct effects, but in some cases this is complemented with information about biological shifts in the overall composition of TCRγδ+ T cells. We conclude from our data that ageing decreases absolute numbers of total TCRγδ+ T cells, without significantly affecting the absolute numbers of TCRαβ+ T cells, starting at age 40-50, confirming earlier findings [60]. This decrease in total TCRγδ+ T cells mostly affected the most common TCRγδ+ T cell type in the peripheral blood: Vγ9/Vδ2 cells [61]. In contrast, Vδ1+ cells showed a slight increasing trend in absolute numbers during ageing, implicating a potential role of Vδ1+-specific antigens which could maintain this population over time. Immunological ageing is not solely defined by chronological ageing, but also by persistent viruses like CMV [56]. Vδ1+ cells are often CMV-specific [41,55,62], and CMV+ individuals have higher numbers of Vδ1+ cells, as described in earlier studies [40,45]. This could explain the increasing trend in absolute numbers and percentages of Vδ1+ cells as observed in our cohort in elderly individuals. When CMV serology was included in the analysis, we observed a significant increase in absolute counts of total TCRγδ+ T cells in young CMV+ individuals, in line with previous findings [37]. This could indicate a prolonged activation state, even during early phases of latency, as described by van de Berg *et al*., 2010 [63].

This latency of e.g. CMV does not only influence cells bearing specific receptors for its epitopes, it also shapes the immune system in terms of maturation subsets. Ageing alone caused a significant decrease in total TCRγδ+ T cells, and thus decreasing

absolute numbers of effector and effector memory phenotype cells. However, when considering relative TCRγδ+ T cell subset distributions a significant increase in effector and effector memory phenotypes were observed, similar to what has been described for CD8+TCRαβ+ T cell subset distributions [64], suggesting similar underlying ageing processes. The effect of ageing became more evident when CMV serological status was included in the analysis, showing increased absolute numbers of effector and effector memory cells. Again, when evaluating relative subset distributions, the effects of both ageing and CMV persistence became more evident, for both maturation and effector and effector memory differentiation stages. We mainly found decreased proportions of naive, and increased proportions of effector cells, which have a late-stage differentiated profile, as described before [37,46]. In contrast to the effector cells, we observed increased proportions of early-stage differentiated effector memory cells, which could correlate with a more general memory-like response to antigens in elderly upon prior antigen exposure. Furthermore, we could observe an age-independent effect of CMV on the differentiation stages of effector and effector memory TCRγδ+ T cells, which correlates with previous data of Roux *et al*. [45].

Immune ageing is however not only accompanied by shifts from naive to effector cells, since effector cells also progressively lose their function (exhaustion) during the ageing process [13]. We noticed a significantly enlarged TCRγδ+ T cell population expressing exhaustion-related markers in young CMV+ individuals, while CMV serology did not add to T cell exhaustion in elderly. Loss of IL7Rα marked the loss of self-renewal and early stages of T cell exhaustion. We observed a clear increase in absolute counts of IL7Rα-TCRγδ+ T cells in elderly. Also, we observed increased expression of CD57, which is also highly expressed on terminally differentiated CD8+ CTL with high proliferative activity [15], thus marking replicative senescence and susceptibility to activation-induced cell death (AICD) [16]. High expression is also associated with chronic viral infections like CMV [17]. Our data showed significantly increased numbers of CD57+TCRγδ+ T cells, in both young and old CMV+ individuals. Another NK cell marker which is associated with T cell exhaustion is KLRG1, of which high expression marks terminally differentiated or senescent T cells [13,18], whilst severely exhausted CD8+ CTLs are KLRG1-negative [19]. We observed higher numbers of KLRG1-negative cells in CMV-infected individuals. Also, our data showed that exhausted CD8+ CTL cells were present in higher numbers in CMV+ individuals. Combination of loss of IL7Rα, high CD57 and low KLRG1 expression may function as a marker of exhausted cells. Our data confirmed this with an increased cell count of such exhausted cells, especially already in the young CMV+ group. Furthermore, we saw increasing but not significant numbers of TCRγδ+ T cells expressing the NK cell marker 2B4 (CD244), which is normally expressed on memory CD8+ T cells [20,21]. The percentages did show an increasing trend from young to elderly, but this was not as

evident as described earlier [13]. Of note, 2B4 expression is positively associated with Programmed cell death 1 (PD1) in exhausted CD8+ T cells [23], mediating decreasing TCR-mediated proliferation and cytokine production by providing downstream inhibitory signals [24,25]. Our data did show significant increases in PD1+ TCRγδ+ T cells counts, but IL7Rα-PD1+2B4+ TCRγδ+ T cell counts were not altered. IL7Rα- TCRγδ+ T cells co-expressing PD1 with the apoptosis inducer FAS (CD95) did not significantly alter upon ageing and CMV persistence, however, total PD1+ TCRγδ+ T cells did increase upon CMV persistence, although this was only most obvious for the young CMV+ individuals. However, PD1 is also associated with T cell activation, and might be in case of young CMV+ individuals more indicative of a response to CMV, rather than exhaustion.

In order to further investigate the full exhaustion and senescence profile, and to better discriminate between these processes, it would be relevant to extend marker analysis to other exhaustion markers, such as CD160, Tim3 and Lag3 in combination with PD1 and 2B4 [65], and to assess transcription factors that define T cell subsets (such as FoxP3, Blimp1, Eomes, T-bet) [66,67]. Furthermore, functional analyses would be helpful to assess in vitro the proliferative potential, activation status and apoptosis of TCRγδ+ T cells from CMV- and CMV+ individuals at young and old age. Also, studying the epigenetic landscape of exhaustion-, senescence- and activation-related gene profiles of TCRγδ+ T cells could shed more insights on the actual effects of ageing on TCRγδ+ T cells [68].

In summary, we conclude that ageing by itself impacts on TCRγδ+ T cells, leading to a decrease in the absolute counts of total TCRγδ+ T cells and to shifts in maturation and differentiation subsets. Furthermore, CMV has an additional impact on TCRγδ+ T cell receptor usage, maturation subsets, effector differentiation profiles, as recently reviewed by Khairallah *et al.* [69], and ultimately on exhaustion marker profiles. This indicates that TCRγδ+ T cells are subjected to ageing and exhaustion processes in much the same way as CD8+TCRαβ+ CTLs.

ACKNOWLEDGEMENTS

We thank Diana van den Heuvel (LUMC, the Netherlands) for advice on data analysis and statistical support, and Kim Heezen (Erasmus MC, the Netherlands) for technical advice on flow cytometry gating strategies and database building. The study was supported through an unrestricted research grant from Roche. The research for this manuscript was (in part) performed within the framework of the Erasmus Postgraduate School Molecular Medicine.

CONFLICT OF INTEREST

The authors have nothing to disclose.

M.J.K., J.J.M.v.D. and A.W.L. designed the experiments. E.B., A.M.H.B., W.H.A. and J.A.N.V. coordinated the clinical work. M.J.K. and M.Y.v.d.K. executed the laboratory experimental procedures. M.J.K., M.Y.v.d.K. and A.W.L. analysed the data. M.J.K. prepared the Figures. M.J.K. and A.W.L. wrote the manuscript. All authors revised the manuscript critically.

REFERENCES

1. Lee, H.C. & Wei, Y.H. Mitochondria and ageing. *Adv. Exp. Med. Biol.* **942**, 311-327 (2012).
2. Gilbert, S.F. Ageing: the biology of senescence in *Developmental Biology* (ed. 6) Sunderland (MA) (Sinauer Associates, 2000).
3. Sosa, V. *et al.* Oxidative stress and cancer: an overview. *Ageing Res. Rev.* **12(1)**, 3376-390 (2013).
4. Lauri, A., Pompilio, G. & Capogrossi, M.C. The mitochondrial genome in ageing and senescence. *Ageing Res. Rev.* **18**, 1-15 (2014).
5. Weiskopf, D., Weinberger, B. & Grubeck-Loebenstein, B. The ageing of the immune system. *Transpl. Int.* **22**, 1041-1050 (2009).
6. Boraschi, D. *et al*. The gracefully ageing immune system. *Ageing* **5(185)**, 185ps8 (2013).
7. Müller, L., Fülöp, T. & Pawelec, G. Immunosenescence in vertebrates and invertebrates. *Immun. Ageing* **10(12)**, 1-14 (2013).
8. Franceschi, C. Inflammageing as a major characteristic of old people: can it be prevented or cured? *Nutr. Rev.* **65**, S173-S176 (2007).
9. Panda, A. *et al.* Age-associated decrease in TLR function in primary human dendritic cells predicts influenza vaccine response. *J. Immunol.* **184**, 2518-2527 (2010).
10. Gibson, K.L. *et al.* B cell diversity decreases in old age and is correlated with poor health status. *Ageing Cell* **8**, 18-25 (2009).
11. Linton, P.J. & Dorschkind, K. Age-related changes in lymphocyte development and function. *Nat. Rev. Immunol.* **5(20)**, 133-139 (2004).
12. Aspinall, R. & Andrew, P. Thymic involution in ageing. *J. Clin. Immunol*. **20(4)**, 250-256 (2000).
13. Wherry, E.J. T cell exhaustion. *Nat. Rev. Immunol.* **12(6)**, 492-499 (2011).
14. Wherry, E.J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486-499 (2015).
15. Lopez-Vergès, S. *et al.* CD57 defines a functionally distinct population of mature NK-cells in the human CD56dimCD16+ NK-cell subset. *Blood* **116(19)**, 3865-3874 (2010).
16. Focosi, D., Bestagno, B., Burrone, O. & Petrini, M. CD57+ T-lymphocytes and functional immune deficiency. *J. Leukoc. Biol.* **87**, 107-116 (2010).

17. Chong, L.K. *et al*. Proliferation and IL-5 production by CD8hiCD57+ T cells. *Eur. J. Immunol.* **38(4)**, 995-1000 (2008).
18. Rosshart, S. *et al.* Interaction of KLRG1 with E-cadherin: new functional and structural insights. *Eur. J. Immunol*. **38**, 3354-3364 (2008).
19. Wherry, E.J. *et al.* Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27(5)**, 670-684 (2007).
20. McNerney, M.E., Lee, K.M. & Kumar, V. 2B4 (CD244) is a non-MHC binding receptor with multiple functions on natural kicker cells and CD8+ T cells. *Mol. Immunol.* **42**, 489-494 (2005).
21. Schlaphoff, V. *et al.* Dual function of the NK-cell receptor 2B4 (CD244) in the regulation of HCV-specific CD8+ T cells. *PLoS Pathog.* **7(5)**, e1002045 (2011).
22. Yang, B., Wang, X., Jiang, J. & Cheng, X. Involvement of CD244 in regulating CD4+ T cell immunity in patients with active tuberculosis. *PLoS ONE* **8(4)**, e63261 (2013).
23. Bensch, B. *et al.* Coexpression of PD-1, 2B4, CD160 and KLRG1 on exhausted HCV-specific CD8+ T cells is linked to antigen recognition and T cell differentiation. *PLoS Pathog*. **6(6)**, e1000947 (2010).
24. Blank, C. & Mackensen, A. Contribution of the PD-L1/PD-1 pathway to T cell exhaustion: an update in implications on chronic infections and tumor evasion. *Cancer Immunol. Immunother.* **56**, 739-745 (2007).
25. Zhang, J.Y. *et al.* PD-1 up-regulated is correlated with HIV-specific memory CD8+ T cell exhaustion in typical progressors but not in long-term nonprogressors. *Blood* **109**, 4671-4678 (2007).
26. Petrovas, C. *et al.* Differential association of programmed death-1 and CD57 with ex vivo survival of CD8+ T cells in HIV infection. *J. Immunol.* **183**, 1120-1132 (2009).
27. Vranjkovic, A., Crawley, A.M., Gee, K., Kumar, A. & Angel, J.B. IL-7 decreases IL-7 receptor α (CD127) expression and induces the shedding of CD127 by human CD8+ T cells. *Int. Immunol.* **19(12)**, 1329-1339 (2007).
28. Kared, H., Saeed, S., Klein, M.B. & Shoukry, N.H. CD127 expression, exhaustion status and antigen specific proliferation predict sustained virologic response to IFN in HCV/HIV co-infected individuals. *PLoS ONE* **9(7)**, e101441 (2014).
29. Ensslin, A.S. & Formby, B. Comparison of cytolytic and proliferative activities of human γδ and αβ T cells from peripheral blood against various human tumor cell lines. *J. Natl. Cancer Inst.* **83**, 1564-1569 (1991).
30. Jensen, K.D.C. *et al.* Thymic selection determines γδ T cell effector fate: antigen-naive cells make Interleukin-17 and antigen-experienced cells make Interferon γ. *Immunity* **29**, 90-100 (2008).
31. Turchinovich, G. & Hayday, A.C. Skin-1 identifies a common molecular mechanism for the development of interferon-gamma-secreting versus interleukin-17-secreting gamma-delta T cells. *Immunity* **35**, 59-68 (2011).
32. Bonneville, M., O'Brien, R.L. & Born, W.K. γδ T cell effector functions: a blend of innate programming and acquired plasticity. *Nat. Rev. Immunol.* **10**, 467-478 (2010).

33. Vantourout, P. & Hayday, A.C. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13(2)**, 88-100 (2013).
34. Luoma, A.M., Castro, C.D. & Adams, E.J. γδ T cell surveillance via CD1 molecules. *Trends Immunol.* **35(12)**, 613-621 (2014).
35. Godfrey, D.I., Uldrich, A.P., McCluskey, J., Rossjohn, J. & Moody, D.B. The burgeoning family of unconventional T cells. *Nat. Rev. Immunol.* **16(11)**, 1114-1123 (2015).
36. Vasudev, A. *et al.* γ/δ T cell subsets in human aging using the classical α/β T cell model. *J. Leukoc. Biol.* **96(4)**, 647-655 (2014).
37. Wistuba-Hamprecht, K., Haehnel, K., Janssen, N., Demuth, I. & Pawelec, G. Peripheral blood T cell signatures from high-resolution immune phenotyping of γδ and αβ T cells in younger and older subjects in the Berlin Ageing Study II. *Immun. Ageing* **12**, 25 (2015).
38. Sheridan, B.S. & Lefrançois, L. Intraepithelial lymphocytes: to serve and protect. *Curr. Gastroenterol. Rep.* **12(6)**, 513-521 (2010).
39. Garrido, P. *et al.* Monoclonal TCR-Vβ13.1+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis: evidence for an antigen-driven chronic T cell stimulation origin. *Blood* **109**, 4890-4898 (2009).
40. Kabelitz, D., Kalyan, S., Oberg, H.H. & Wesch, D. Human Vδ2 versus non-Vδ2 γδ T cells in antitumor immunity. *Oncoimmunology* **2(3)**, e23304 (2013).
41. Alejenef, A. *et al.* Cytomegalovirus drives Vδ2neg γδ T cell inflation in many healthy virus carriers with increasing age. *Clin. Exp. Immunol.* **176(3)**, 418-428 (2014).
42. Sinclair, J. Human cytomegalovirus: latency and reactivation in the myeloid lineage. *J. Clin. Virol.* **41**, 180-185 (2008).
43. Looney, R.H. *et al.* Role of cytomegalovirus in the T cell changes seen in elderly individuals. *Clin. Immunol.* **90(2)**, 213-219 (1999).
44. Shin, H. & Wherry, E.J. CD8 T cell dysfunction during chronic viral infection. *Curr. Opin. Immunol.* **19**, 408-415 (2007).
45. Roux, A. *et al.* Differential impact of age and cytomegalovirus infection on the γδ T cell compartment. *J. Immunol.* **191**, 1300-1306 (2013).
46. Weltevrede, M., Eilers, R., de Melker, H.E. & van Baarle, D. Cytomegalovirus persistence and T cell immunosenescence in people age fifty and older: a systematic review. *Exp. Geront.* **77**, 87-95 (2016).
47. Yan, J. *et al.* The effect of ageing on human lymphocyte subsets: comparisons of males and females. *Immunity & Ageing* **7:4** (2010).
48. Kennedy, R.B. *et al.* Immunosenescence-related transcriptomic and immunologic changes in older individuals following influenza vaccination. *Front. Immunol.* **7**:450 (2016).
49. Spyridopoulos, I. *et al.* CMV seropositivity and T cell senescence predict increased cardiovascular mortality in octogenarians: results from the Newcastle 85+ study. *Aging Cell* **15(2)**, 389-392 (2016).
50. Johnstone, J. *et al.* T cell phenotypes predictive of fratility and mortality in elderly nursing home residents. *J. Am. Geriatr. Soc.* **65(1)**, 153-159 (2017).

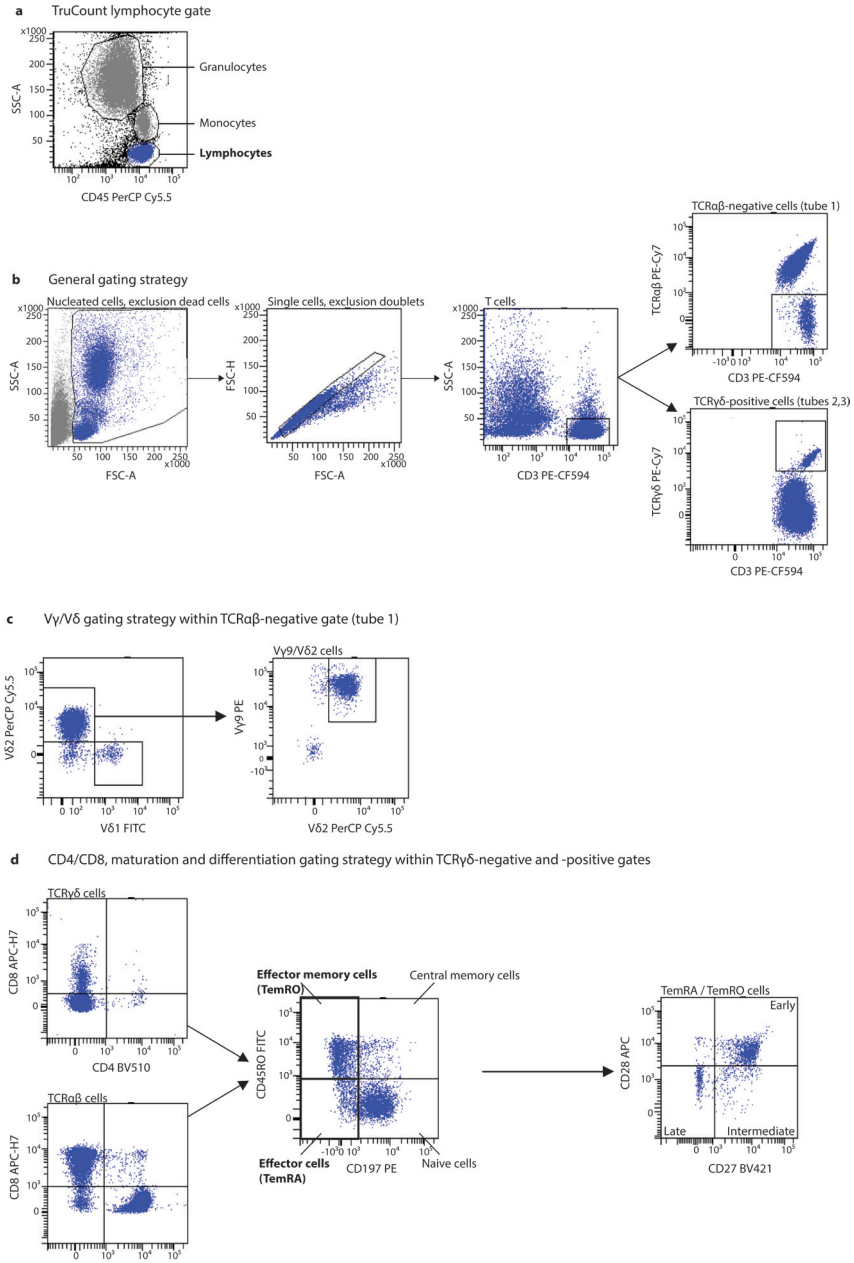
51. Koch, S., Solana, R., Dela Rosa, O. & Pawelec, G. Human cytomegalovirus infection and T cell immunosenescence: a mini review. *Mech. Ageing Dev.* **127**, 538-543 (2006).
52. Korndewal, M.J. *et al.* Cytomegalovirus infection in the Netherlands: Seroprevalence, risk risk factors, and implications. *J. Clin. Virol.* **63**, 53-58 (2015).
53. Jansen, M.A. *et al.* Determinants of ethnic differences in Cytomegalovirus, Epstein-Barr Virus, and Herpes Simplex Virus Type 1 seroprevalence in childhood. *J. Pediatr.* **170**: 126-134 (2016).
54. van der Heiden, M. *et al.* Differential effects of Cytomegalovirus carriage on the immune phenotype of middle-aged males and females. *Sci. Rep.* **6**, 26892; doi: 10.1038/srep26892 (2016).
55. Déchanet, J. *et al.* Implication of γδ T cells in the human immune response to cytomegalovirus. *J. Clin. Invest.* **103**, 1437-1449 (1999).
56. Kahan, S.M., Wherry, E.J. & Zajac, A.J. T cell exhaustion during persistent viral infections. *J. Virol.* **0**, 180-193 (2015).
57. *World Report on Ageing and Health.* World Health Organization 2015.
58. Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J.W. Ageing populations: the challenges ahead. *Lancet Glob. Health* **374(9696)**, 1196-1208 (2009).
59. Gruver, A.L., Hudson, L.L. & Sempowski, G.D. Immunosenescence of ageing. *J. Pathol.* **211(2)**, 144-156 (2007).
60. Colonna-Romano, G. *et al.* Impairment of gamma/delta T-lymphocytes in elderly: implications for immunosenescence. *Exp. Gerontol.* **39**, 1439-1446 (2004).
61. Breit, T.M., Wolvers-Tettero, I.L.M. & van Dongen, J.J.M. Receptor diversity of human T cell receptor γδ expressing cells. *Prog. Histochem. Cytochem.* **26**, 183-193 (1992).
62. Lewis, D. *et al.* Cytomegalovirus infection is associated with expansions of CD8 T cells and highly oligoclonal Vdelta1 gamma/delta T cells in patients treated with Dasatinib for chronic myelogenous leukaemia. *Blood* **124**: 1814 (2014).
63. van de Berg, P.J. *et al.* Human cytomegalovirus induces systemic immune activation characterized by a type 1 cytokine signature. *J. Infect. Dis.* **202**, 690-699 (2010).
64. Czesnikiewicz-Guzik, M. *et al.* T cell subset-specific susceptibility to ageing. *Clin. Immunol.* **127**, 107-118 (2008).
65. Pombo, C., Wherry, E.J., Gostick, E., Price, D.A. & Betts, M.R. Elevated expression of CD160 and 2B4 defines a cytolytic HIV-specific CD8+ T cell population in elite controllers. *J. Infect. Dis*. **212(9)**, 1376-1386 (2015).
66. Buggert, M. *et al.* T-bet and Eomes are differentially linked to the exhausted phenotype of CD8+ T cells in HIV infection. *PLoS Pathog.* **10(7)**, e1004251 (2014).
67. Barathan, M. *et al.* Chronic hepatitis C virus infection triggers spontaneous differential expression of biosignatures associated with T cell exhaustion and apoptosis signaling in peripheral blood mononucleocytes. *Apoptosis* **20(4)**, 466-480 (2015).

68. Schmolka, N., Wencker, M., Hayday, A.C. & Silva-Santos, B. Epigenetic and transcriptional regulation of γδ T cell differentiation: programming cells for responses in time and space. *Semin. Immunol.* **27(1)**, 19-25 (2015).
69. Khairallah, C., Déchanet-Merville, J. & Capone, M. γδ T cell-mediated immunity to cytomegalovirus infection. *Front. Immunol.* **8:** 105 (2017).
70. Van Den Heuvel, D. *et al*. Effects of nongenetic factors on immune cell dynamics in early childhood: the Generation R Study. *J. Allergy. Clin. Immunol.* **S0091-6749 (16)**, 31379-3 (2016).

SUPPLEMENTAL DATA

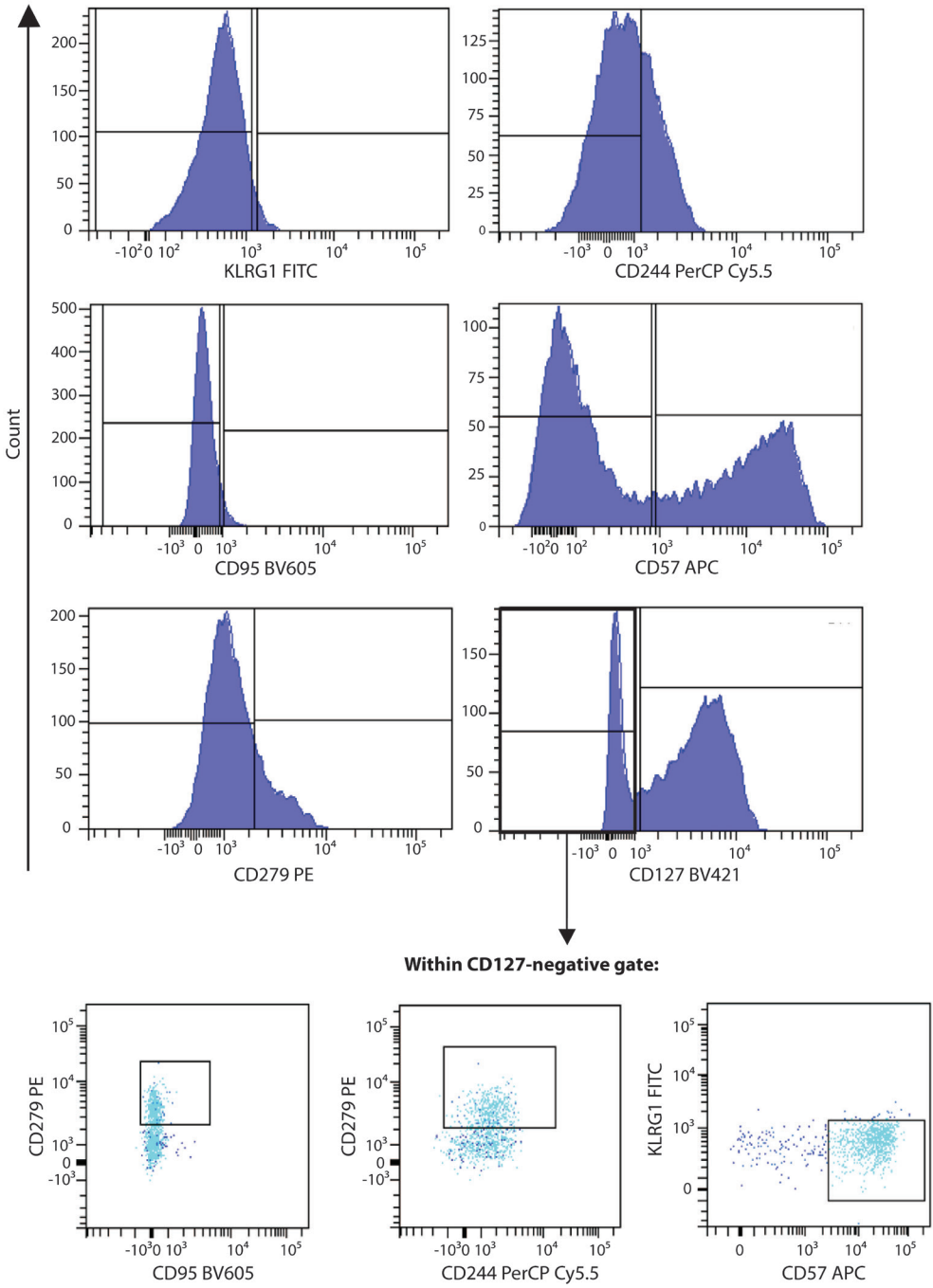
Supplementary Table 1. Antibody details

Tube		Fluorochrome									
		BV421	BV510	BV605	FITC	PerCP-Cy5.5	PE	PE-CF594	PE-Cy7	APC	APC-H7
1	Antibody	CD27	CD4	CD45RA	TCRVδ1	TCRVδ2	TCRVγ9	CD3	TCRαβ	CD25	CD8
	Clone	0323	OKT4	HI100	TS8.2	B6	B3.1	UCHT1	IP26	2A3	SK1
	Manufacturer	BioLegend	BioLegend	BioLegend	Thermo Scientific	BioLegend	BD Biosciences	BD Biosciences	BioLegend	BD Biosciences	BD Biosciences
2	Antibody	CD27	CD4	CD45RA	CD45RO		CD197	CD3	TCRγδ	CD28	CD8
	Clone	0323	OKT4	HI100	UCHL1		3D13	UCHT1	11F2	CD28.2	SK1
	Manufacturer	BioLegend	BioLegend	BioLegend	DAKO		eBiosciences	BD Biosciences	BD Biosciences	BD Biosciences	BD Biosciences
3	Antibody	CD127	CD4	CD95	KLRG1	CD244	CD279	CD3	TCRγδ	CD57	CD8
	Clone	TU27	OKT4	DX2	2F1/KLRG1	C1.7	MIH4	UCHT1	11F2	NK-1	SK1
	Manufacturer	BioLegend	BioLegend	BioLegend	BioLegend	BioLegend	BD Biosciences	BD Biosciences	BD Biosciences	BD Biosciences	BD Biosciences

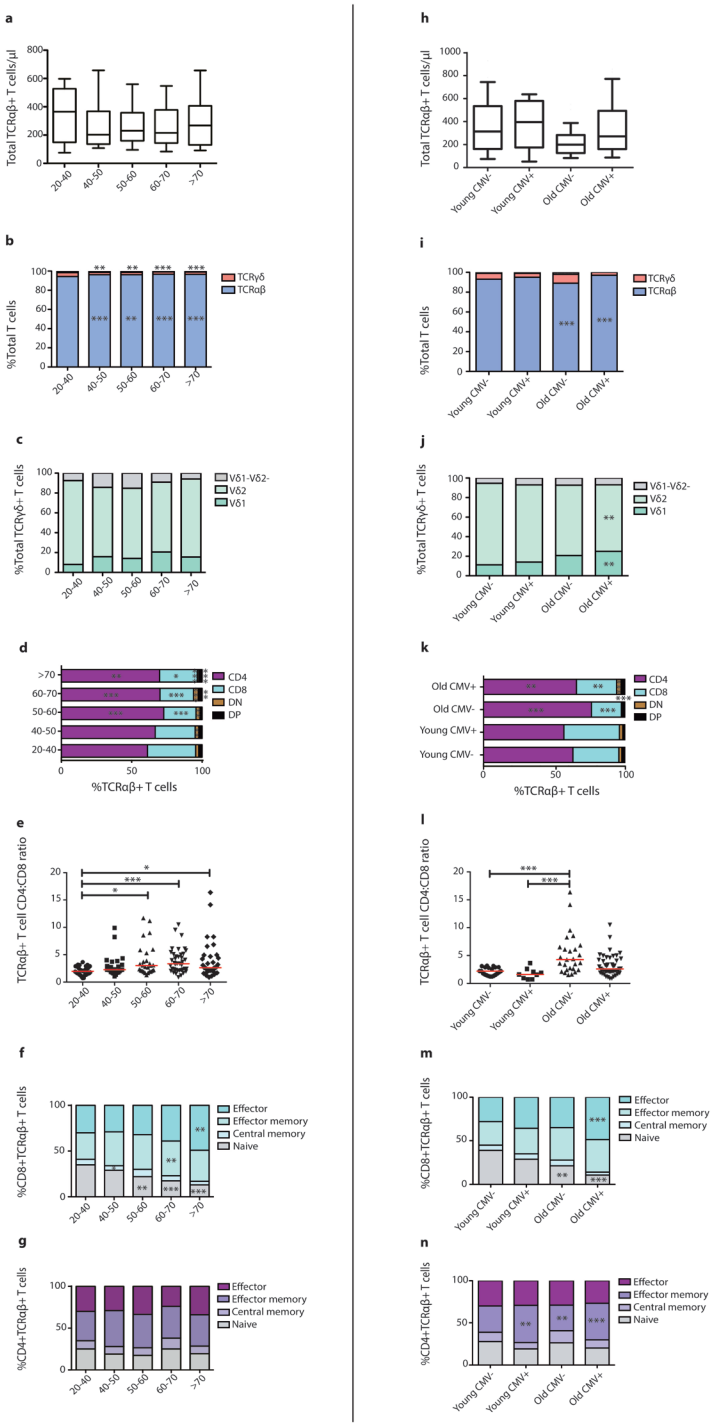


Supplementary Figure 1. Gating strategies for analysis.
(**a**) TruCount gating strategy for determining absolute lymphocyte counts and absolute cell number calculations. (**b**) General gating strategy with the use of TCRαβ antibody in case of Vγ- and Vδ-usage investigation, (**c**) gating strategy using TCRγδ antibody. (**d**) CD4/CD8 usage, maturation and differentiation within effector (TemRA) and effector memory (TemRO) populations. (**e**) Exhaustion profile determination including separate exhaustion marker expressions and exhaustion marker combinations within the CD127- population.

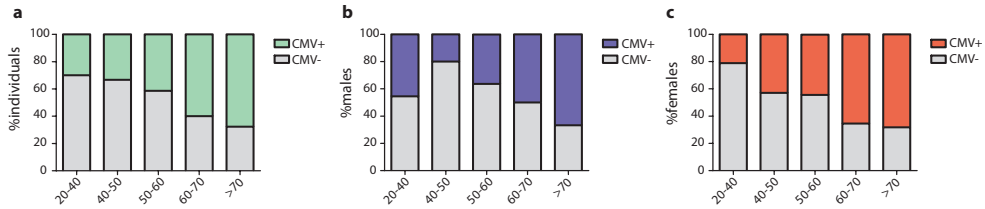
e Exhaustion gating strategy



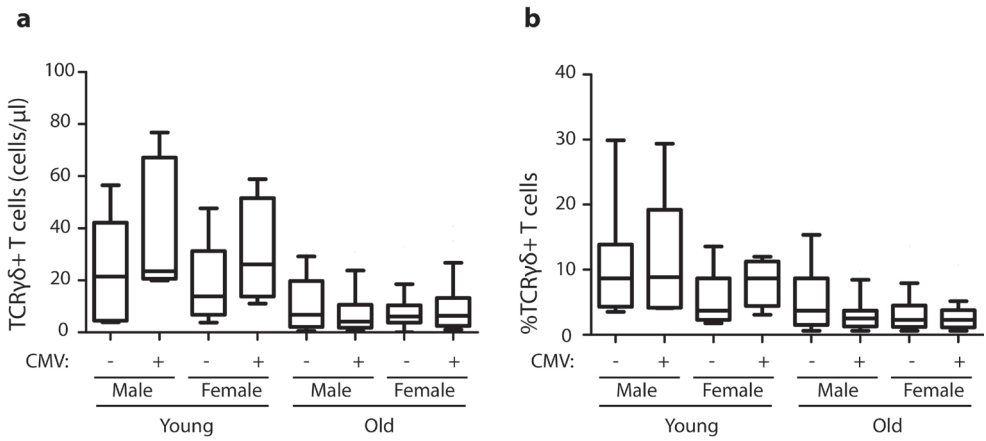
Supplementary Figure 1 continued (see left page). Gating strategies for analysis.
(**a**) TruCount gating strategy for determining absolute lymphocyte counts and absolute cell number calculations. (**b**) General gating strategy with the use of TCRαβ antibody in case of Vγ- and Vδ-usage investigation, (**c**) gating strategy using TCRγδ antibody. (**d**) CD4/CD8 usage, maturation and differentiation within effector (TemRA) and effector memory (TemRO) populations. (**e**) Exhaustion profile determination including separate exhaustion marker expressions and exhaustion marker combinations within the CD127- population.


a
Total TCRαβ+ T cells/μl
800
600
400
200
0
20-40
40-50
50-60
60-70
>70
h
Total TCRαβ+ T cells/μl
1000
800
600
400
200
0
Young CMV-
Young CMV+
Old CMV-
Old CMV+
b
%Total T cells
100
80
60
40
20
0
**
***
TCRγδ
TCRαβ
20-40
40-50
50-60
60-70
>70
i
%Total T cells
TCRγδ
TCRαβ
Young CMV-
Young CMV+
Old CMV-
Old CMV+
c
%Total TCRγδ+ T cells
Vδ1-Vδ2-
Vδ2
Vδ1
20-40
40-50
50-60
60-70
>70
j
%Total TCRγδ+ T cells
Vδ1-Vδ2-
Vδ2
Vδ1
Young CMV-
Young CMV+
Old CMV-
Old CMV+
d
>70
60-70
50-60
40-50
20-40
CD4
CD8
DN
DP
0
50
100
%TCRαβ+ T cells
k
Old CMV+
Old CMV-
Young CMV+
Young CMV-
CD4
CD8
DN
DP
0
50
100
%TCRαβ+ T cells
e
TCRαβ+ T cell CD4:CD8 ratio
20
15
10
5
0
*
***
*
20-40
40-50
50-60
60-70
>70
l
TCRαβ+ T cell CD4:CD8 ratio
***
***
Young CMV-
Young CMV+
Old CMV-
Old CMV+
f
%CD8+TCRαβ+ T cells
100
50
0
Effector
Effector memory
Central memory
Naive
20-40
40-50
50-60
60-70
>70
m
%CD8+TCRαβ+ T cells
Effector
Effector memory
Central memory
Naive
Young CMV-
Young CMV+
Old CMV-
Old CMV+
g
%CD4+TCRαβ+ T cells
100
50
0
Effector
Effector memory
Central memory
Naive
20-40
40-50
50-60
60-70
>70
n
%CD4+TCRαβ+ T cells
Effector
Effector memory
Central memory
Naive
Young CMV-
Young CMV+
Old CMV-
Old CMV+

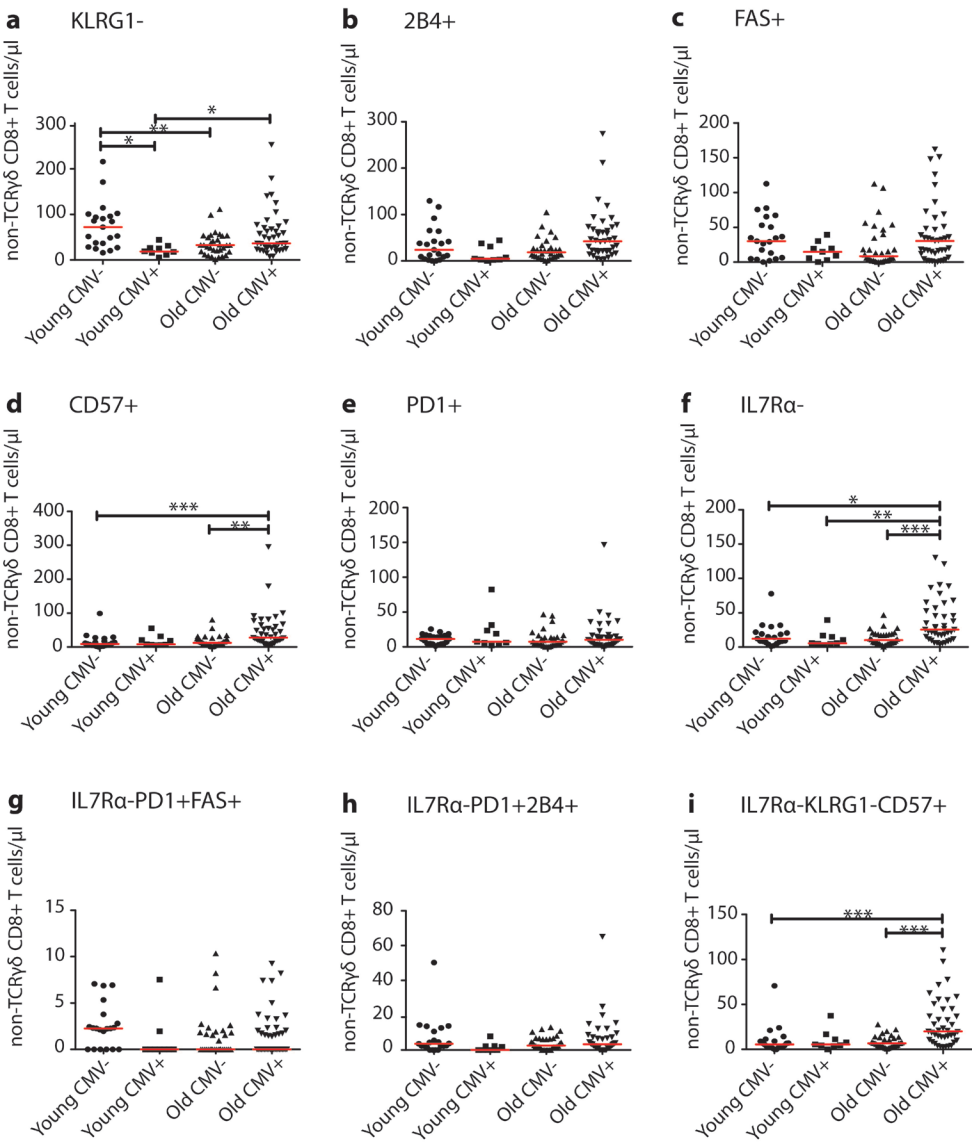
Supplementary Figure 2 (see left page). TCRαβ+ and TCRγδ+ T cell supplementary figure.
(**a,h**) Absolute numbers of TCRαβ+ T cells depicted in 10-90% box-whiskers-plots. (**b,i**) Relative distributions of TCRαβ+ and TCRγδ+ T cells within the total T cell compartment of all groups. (**c,j**) Relative Vδ1+, Vδ2+, and non-Vδ1/non-Vδ2 usage distribution in the peripheral blood total TCRγδ+ T cell compartment. (**d,k**) TCRαβ+ CD4 and CD8 single-positive, double-positive (DP, CD4+CD8+) and double-negative (DN, CD4-CD8-) subset distribution depicted in stacked bar plots. (**e,l**) CD4:CD8 ratios of total TCRαβ+ T cells depicted in scatter plots indicated with the median. (**f,m**) Maturation subset distributions of CD8+TCRαβ+ and (**g,n**) CD4+TCRαβ+ T cells depicted in stacked bar plots. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test. Significance of the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01; ***, p<0.001.



Supplementary Figure 3. CMV serology.
(**a**) Percentages of overall CMV serostatus. (**b**) CMV-positivity and –negativity among males and (**c**) females. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test.

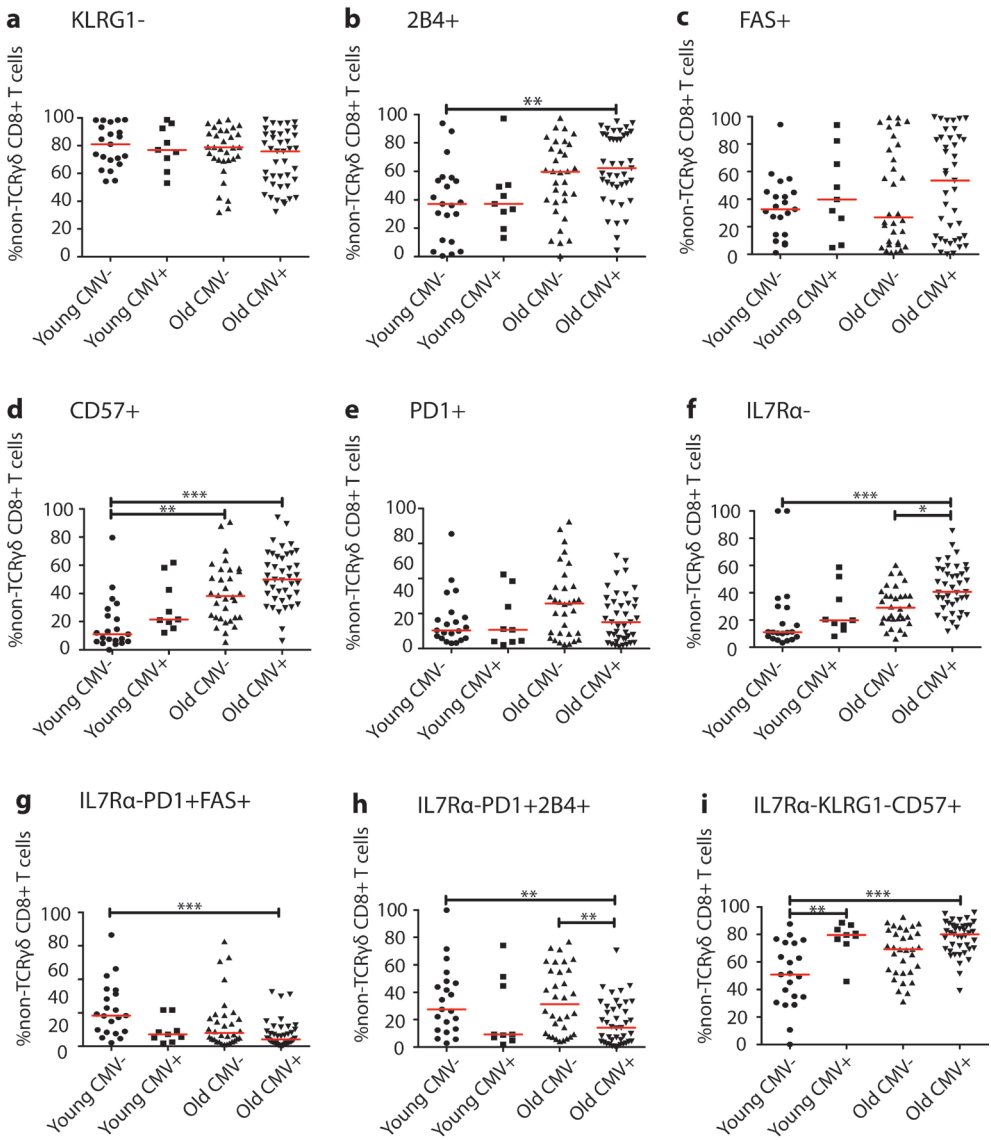


Supplementary Figure 4. Effect of CMV, gender and age on V-gene usage.
Supplementary to Figure 4 and Supplementary Table 4. (**a**) Absolute numbers and (**b**) percentages of total TCRγδ+ T cells in young / elderly CMV- and CMV+ males and females depicted in 10-90% box-whiskers-plots. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test. Significance of the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01; ***, p<0.001.

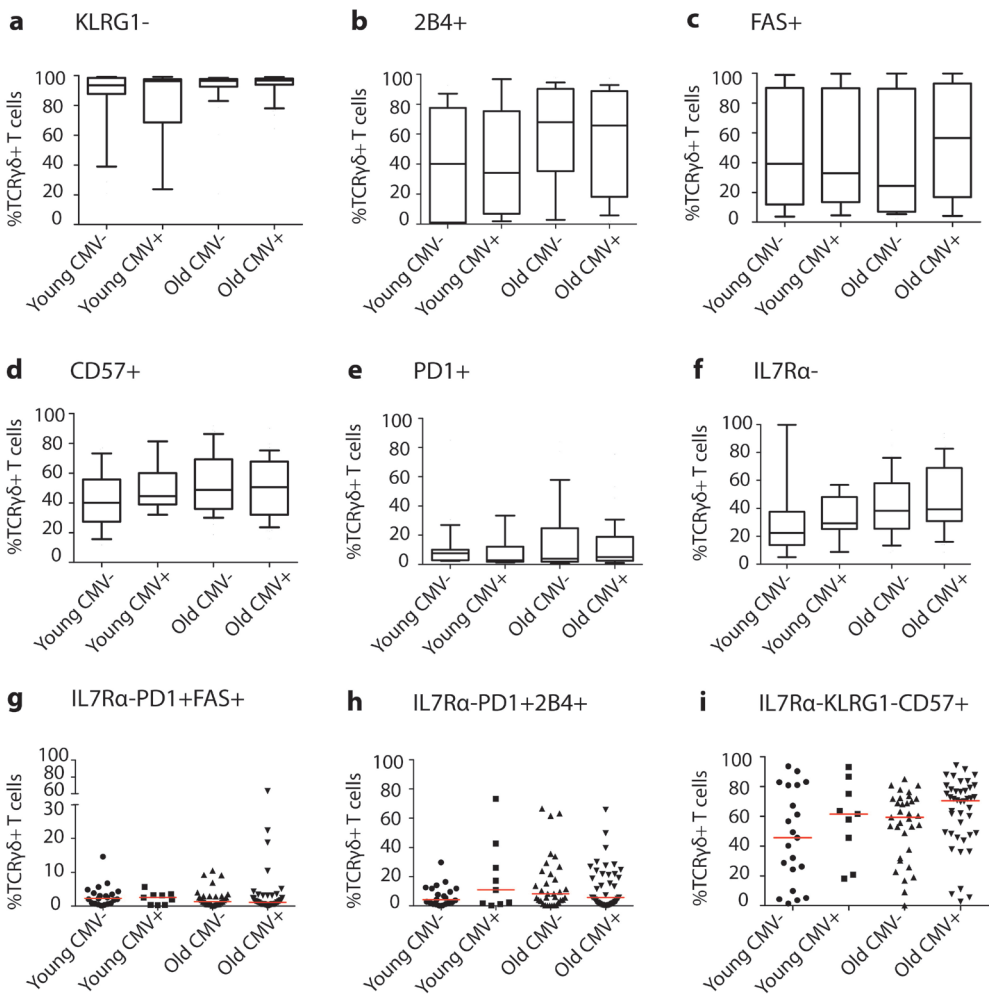


Supplementary Figure 5. Effect of CMV on the exhaustion and senescence of nonTCRγδ (TCRαβ+) CD8+ T cells.

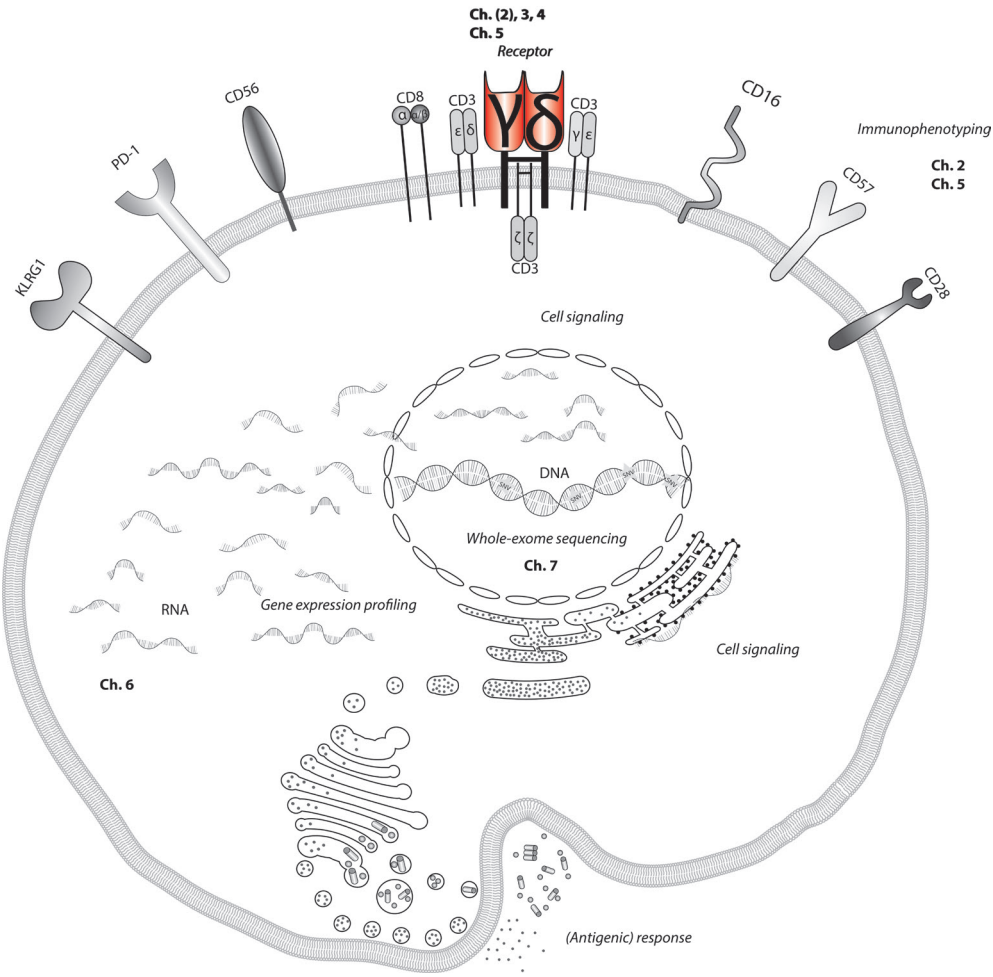
Absolute numbers of TCRαβ+ (TCRγδ-negative T cells) CD8+ T cells (**a**) lacking KLRG1, (**b**) expressing 2B4, (**c**) FAS death receptor, (**d**) CD57, (**e**) PD1, and (**f**) lacking IL7Rα. Absolute numbers of TCRαβ+CD8+ IL7Rα- T cells coexpressing (**g**) PD1 and FAS, (**h**) PD1 and 2B4, (**i**) CD57 and lacking KLRG1. Scatter plots are indicated with the median. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test. Significance of the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01; ***, p<0.001.



Supplementary Figure 6. Relative effects of CMV on the exhaustion marker profile of nonTCRγδ (TCRαβ+) CD8+ T cells.
(**a**) Percentages of KLRG1-, (**b**) 2B4+, (**c**) FAS+, (**d**) CD57+, (**e**) PD1+ and (**f**) IL7Rα- TCRαβ+ (TCRγδ-negative T cells) CD8+ T cells are depicted in 10-90% box-whiskers-plots. (**g**) Percentages of IL7Rα- TCRαβ+ T cells co-expressing PD1 and FAS, (**h**) PD1 and 2B4 and (**i**) lacking KLRG1 with expression of CD57 depicted in scatterplots indicated with the medians. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test. Significance of the Dunn's test is indicated in the plots: *, p<0.05; **, p<0.01; ***, p<0.001.



Supplementary Figure 7. Relative effects of CMV on the exhaustion marker profile of TCRγδ+ T cells. (**a**) Percentages of KLRG1-, (**b**) 2B4+, (**c**) FAS+, (**d**) CD57+, (**e**) PD1+ and (**f**) IL7Rα- TCRγδ+ T cells are depicted in 10-90% box-whiskers-plots. (**g**) Percentages of IL7Rα- TCRγδ+ T cells co-expressing PD1 and FAS, (**h**) PD1 and 2B4 and (**i**) lacking KLRG1 with expression of CD57 depicted in scatterplots indicated with the medians. Significance was tested by a Kruskal-Wallis test, followed by a post-Dunn's test.


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 7
RNA
Gene expression profiling
Cell signaling
Ch. 6
(Antigenic) response

Chapter 3

Optimization of amplicon-based next-generation sequencing methods for characterizing TRG and TRD repertoire diversity

Martine J. Kallemeijn[1], François G. Kavelaars[2], Michèle Y. van der Klift[1], Ingrid L.M. Wolvers-Tettero[1], Peter J.M. Valk[2], Jacques J.M. van Dongen[1] & Anton W. Langerak[1]

[1]: Department of Immunology, Laboratory for Medical Immunology, Erasmus MC, Rotterdam 3000 CA, the Netherlands.
[2]: Department of Hematology, Erasmus MC, Rotterdam 3000 CA, the Netherlands.

Unpublished introductory technical methods manuscript

ABSTRACT

Next-generation sequencing is a novel tool which can be applied for many research purposes; from mapping whole genomes to the repertoire diversity of the immune system. Different companies have established different sequencing technologies, all having their specific traits and trade-offs: deep sequencing often present with sequencing errors and other biases, resulting from either preparation steps or the sequencing technology itself. Several quality internal control mechanisms and analysis algorithms have been developed in order to reduce the sequencing biases, however, in case of repertoire studies the amplification steps induce the major part of sequencing errors. PCR amplification steps are necessary in order to obtain sufficient material for sequencing, but also bring along many factors which could introduce errors: primers, reagents, annealing temperatures, cycle numbers, et cetera. In order to reduce the errors introduced in this amplification step for unbiased TRG / TRD repertoire sequencing we applied stepwise optimization experiments, based on a spike-in method using plasmid DNA. By using samples with known input, we were able to titrate primer concentrations, annealing temperatures and cycle numbers to establish low levels of primer competition and preference, optimized annealing temperatures and preventing PCR plateau phases to maintain receptor diversity. After extensive iterative experiments we were next able to validate the formulated TRG / TRD multiplex PCR assays on genomic DNA and further optimized the assays. Conclusively, we have generated most optimal TRG / TRD multiplex PCR assays for NGS methods with lowest PCR bias and inter-assay variance, which could furthermore become useful in repertoire diversity studies, but also in diagnostic test for diagnosis leukemic clones and for the quantification of minimal residual disease.

INTRODUCTION

Next-generation sequencing (NGS) methods to study cancer mutagenesis, microbiome compositions and antibody and T cell repertoire diversity evolve rapidly into highly advanced sensitive and accurate techniques. A large spectrum of platforms for NGS techniques have become available over time from different companies: Roche 454 sequencing technology [1], Illumina sequencing platforms [2], Applied Biosystems' SOLiD System [3], Life Technologies Ion Torrent [4], Pacific Biosciences PacBio System [5] for long reads (>1 kb), and many more [6-8]. Each system provides its traits and trade-offs in terms of costs, read yield per run and sequencing read length [6,7,9]. Research questions determine often which platform is most suitable. In case of visualizing mutations associated with cancers deep sequencing of the whole genome is most applicable, involving the Illumina HiSeq and TruSeq systems [9]. However, in case of repertoire diversity studies, which targets a specific region in the genome leading to relatively large amplicons, 454 sequencing technology or Illumina MiSeq platforms were most suitable [10,11], since also major analysis pipelines are designed for 454 sequencing technology and Illumina MiSeq datasets [10,12].

All NGS platforms and methodologies bring errors along: from sequencing errors in the sequencing technique itself, to data analysis errors, but also pre-amplification steps in the preparation of samples for NGS cause high degrees of sequencing biases and errors. Usually, this pre-amplification step includes PCR, upon which amplification errors can occur very early. To correct for sequencing errors, a polymerase enzyme with proofreading can be used, but also increasing of the sequencing depth is an option to solve this. The latter implies that every region is covered by a high number of reads, preferably sequenced from both ends (paired-end), upon which data is aligned to a reference genome [9,13]. Furthermore, several data analysis algorithms have been designed to correct for these sequencing errors, or linkage disequilibrium principles are applied [14].

In case of sequencing errors in studying B and T cell repertoire diversities correction is rather difficult, since the junctional regions between the variable (V), diversity (D) and joining (J) genes during VDJ recombination are highly diverse and thus contain many "mutations" according to sequencing analysis software. Next to that, in order to obtain representative data for repertoire diversity, deep sequencing of a given region reduces the diversity of the reads yielded per run. Therefore, also prior PCR steps should be reduced in their cycle numbers, in order to prevent reaching plateau phases. To further include all possible gene combinations which are possibly present in the repertoire of healthy individuals, different primer sets should be included in a multiplex setting.

However, different primers could anneal to similar or different regions. In fact, different primers have different annealing temperatures (Ta), which could result in primer competition. To also reduce the build-in of wrong nucleotides, a Taq polymerase with proofreading can be used in both amplification steps, i.e. initial amplification of the target genomic regions, and consecutive amplification that includes addition of sample identification tags.

In the current study we aimed to optimize multiplex PCR methods for an Illumina-based NGS application in studying the T cell receptor (TCR) repertoire of TCRγδ+ T cells. To do so, we developed two multiplex PCR mixes for the T cell receptor γ (TRG) and T cell receptor δ (TRD) loci and tested these using artificial samples based on a spike-in method. Optimization of the protocol also included the titration of primer concentrations, application of different annealing temperatures and variable PCR cycle numbers.

MATERIALS & METHODS

Study subjects

Cryopreserved thymus material (N=5) was obtained from the biobank from the department of Immunology, Erasmus MC, University Medical Center (Rotterdam, the Netherlands). Thymic lobes were removed upon open heart surgery in individuals under the age of two years and obtained upon written informed consent from parents. Whole thymic material was ruptured and prepared for cryopreservation. Fresh cord blood (N=5) mononuclear cells (CBMNCs) were obtained with CB Collect Bags containing citrate phosphate dextrose (CPD) (Fresenius Kabi, Bad Homburg vor der Höhe, Germany) post-partum or after Caesarian section upon written informed consent in collaboration with the department of Obstetrics, Sophia Children's Hospital, Erasmus MC and the department of Hematology, Erasmus MC. Both thymic and cord blood material were obtained under Medical Ethics Committee approval project number hmPOO2004-003. Healthy adult blood donors (N=5) from Sanquin Blood Bank (Amsterdam, the Netherlands) were included upon written informed consent at the blood bank under project number NVT0012.01. CB- and PBMNCs were isolated by means of Ficoll-Paque (density 1.077 g/ml, Pharmacia, Uppsala, Sweden) density gradient separation. PBMNCs were cryopreserved in Iscove's Modified Dulbecco's Medium (IMDM, Lonza, Basel, Switzerland) with dimethyl sulfoxide and stored in vials at -180 °C until further use, CBMNCs were directly used after obtaining. Materials from all study subjects were anonymized for further use. Studies were conducted in accordance with the principles of the Declaration of Helsinki.

TCRγδ+ T cell isolation and DNA isolation

Fresh cord blood mononuclear cells (CBMNCs) and cryopreserved and thawed thymocytes and healthy control adult peripheral blood mononuclear cells (PBMNCs) were subjected to manual magnetic separation using the Anti-TCRγ/δ MicroBead kit from Miltenyi (Miltenyi Biotech, Bergisch Gladbach, Germany) according to the manufacturer's protocol. Directly after isolation cells were lysed in RLT+ lysis buffer supplemented with 1% beta-mercaptoethanol and further subjected to combined DNA/RNA/miRNA isolation with the QIAGEN DNA/RNA/miRNA AllPrepKit (QIAGEN, Hilden, Germany). DNA concentration and quality evaluation was performed with Nanodrop measurements (Thermo Scientific, Waltham, MA, USA).

Primer design

Primers for cloning and as basis for Illumina sequencing technology were largely based on the BIOMED-2 protocol [15]. Vδ3 primer was redesigned based on target amplicon length and in closer proximity of existing Vδ1 and Vδ2 primers. A Jγ1.2 primer was newly designed, as this was not included in the BIOMED-2 protocol. Primer adaptation for Illumina MiSeq technology involved adding Illumina forward Rd1 adaptor (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse Rd2 adaptor (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3'). Adaptor sequences functioned as templates for indexprimers with sample-specific identification tags in second-step PCR, and for attachment of amplicons to the flowcell for bridge amplification. Primers are summarized in Supplementary Table 1.

Plasmid pool preparation

Single V-J gene combinations were singleplex-PCR amplified and cloned from either DNA from T-ALL patient cell lines or thymus DNA into pGEM T-Easy vector in a 3:1 insert:vector ratio according to the manufacturer's protocol (Promega, Madison, WI, USA). From the TRD locus all possible combinations could be cloned. Even though not all TRG combinations could be obtained two different plasmid pools were generated, at least covering all V- and J-gene primers (Supplementary Table 2).

Two-step PCR amplifications

PCRs were first tested in singleplex and multiplex settings with varying primer concentrations, annealing temperatures and PCR cycle numbers. Each initial PCR mix contained GeneAmp PCR Buffer II (1x), magnesium chloride (2.5 mM), deoxynucleotides (2.0 mM) and AmpliTaqGold (1 U) (Thermo Fischer Scientific). Total forward and reverse primer(s) amounts were 10 pmol. The PCR protocol was largely based on the BIOMED-2 publication (23), with varying annealing temperatures (Tm = 58/59/60/62)

and different numbers of cycles (20 and 25 cycles). In general total forward and reverse primer(s) amount was 10 pmol. Primer concentration adjustment, and optimization of annealing temperatures and number of PCR cycles were based on the results of iterative optimization experiments. Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA), whereafter concentrations were measured with the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fischer Scientific), and the amplicons were adjusted to similar concentrations. The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina) using the KAPA HiFi HotStart PCR Kit (Kapa Biosystems, Wilmington, MA, USA). Second PCR amplicons were evaluated with agarose gel electrophoresis or PicoGreen concentration measurement. Library pool preparation was performed based on the gel image or PicoGreen measurement results. The library pool was further purified with Agencourt AMPure XP beads and normalized from Illumina-based sequencing, according to the manufacturer's protocol (Illumina).

Next-generation sequencing

Next-generation paired-end (2x221 bp) sequencing was performed on the Illumina MiSeq Kit platform (Illumina, San Diego, CA, USA) using the Illumina MiSeq Reagent Kit V3 according to the manufacturer's protocol (Illumina).

Bio-informatic data analysis

Illumina NGS data was obtained in FASTQ format. Paired-end reads were joined with the Pear Paired-End read merger tool [16] in the Erasmus MC Galaxy Server [17]. Joined FASTQ reads were converted into FASTA format [18]. Sequencing annotations were performed with IMmunoGeneTics (IMGT) High V-quest database [19-22]. Data analysis was performed using the Antigen Receptor Galaxy (ARGalaxy) tool [12], which included analysis of V-, D- and J-gene frequencies with bar graphs and heatmaps, CDR3 characteristics (length and amino acid compositions), clonality scores [23] and junction analysis (insertions/deletions, P and N nucleotides). Circoletto plots were generated using the Circos online software (https://www.circos.ca) [24].

Statistical analyses

The Hartigan's Dip Test Statistic for Unimodality [25] was applied to test for normal distribution of the sequencing data using R package version 3.3.0.

RESULTS

Primer target specificity and exclusivity experiments illustrate the need for primer optimization to reduce competition bias

Primers targeting V- and J-genes of TRG / TRD loci were adjusted in view of Illumina MiSeq next-generation sequencing technology with the addition of forward and reverse adaptors (Supplementary Table 1). Since the adaptors are 33 nucleotides in length – and thus longer than the target primers – we first performed singleplex and multiplex PCRs on plasmids to validate primer target specificity and exclusivity. In the singleplex assay 10 pmol forward and 10 pmol reverse primer was used, whereas in the multiplex assay the total amounts of forward and reverse primers were 10 pmol each. Both assays showed that the primer specificity and exclusivity did not alter upon attachment of the forward and reverse adaptors, although some primer combinations also resulted in other V-J rearrangements than the ones expected based on the plasmids used. This was especially true for the TRG assay (Vγ2-Jγ1.1, Vγ8-Jγ1.1 and Vγ9-Jγ2.3) (Table 1), which may be at

Table 1. Evaluation of TRG / TRD primer specificity and exclusivity using plasmid sequences.

Single TRG plasmid	True (%)*	Other (%)**	Single TRD plasmid	True (%)*	Other (%)**
Vγ2-Jγ1.1	4156 (20.58)	16041 (79.42)	Vδ1-Jδ1	16389 (99.82)	29 (0.18)
Vγ2-Jγ2.1	19818 (99.99)	1 (0.01)	Vδ1-Jδ2	10560 (99.82)	19 (0.18)
Vγ2-Jγ1.3	22228 (99.96)	8 (0.04)	Vδ1-Jδ3	16125 (96.08)	658 (3.92)
Vγ2-Jγ2.3	22263 (99.97)	6 (0.03)	Vδ1-Jδ4	19605 (99.89)	22 (0.11)
Vγ3-Jγ1.1	2492 (99.96)	1 (0.04)	Vδ2-Jδ1	23099 (98.14)	438 (1.86)
Vγ3-Jγ2.1	2764 (99.96)	1 (0.04)	Vδ2-Jδ2	11985 (95.74)	533 (4.26)
Vγ3-Jγ1.3	18674 (99.93)	14 (0.07)	Vδ2-Jδ3	10845 (69.38)	4786 (30.62)
Vγ4-Jγ2.1	23235 (99.93)	17 (0.07)	Vδ2-Jδ4	14554 (93.00)	1095 (7.00)
Vγ4-Jγ1.3	25741 (99.95)	14 (0.05)	Vδ3-Jδ1	10845 (99.99)	1 (0.01)
Vγ4-Jγ2.3	22244 (99.95)	12 (0.05)	Vδ3-Jδ2	13168 (99.94)	8 (0.06)
Vγ5-Jγ1.1	17974 (98.42)	289 (1.58)	Vδ3-Jδ3	11874 (99.94)	7 (0.06)
Vγ5-Jγ2.3	22241 (99.96)	10 (0.04)	Vδ3-Jδ4	11187 (99.61)	44 (0.39)
Vγ8-Jγ1.1	9100 (50.55)	8902 (49.45)			
Vγ8-Jγ2.1	18188 (99.86)	26 (0.14)			
Vγ8-Jγ1.3	15088 (99.90)	15 (0.10)			
Vγ8-Jγ2.3	16785 (99.90)	16 (0.10)			
Vγ9-Jγ1.1	18569 (99.79)	39 (0.21)			
Vγ9-Jγ2.1	18594 (99.92)	14 (0.08)			
Vγ9-Jγ1.2	15467 (99.61)	60 (0.39)			
Vγ9-Jγ1.3	18031 (99.94)	10 (0.06)			
Vγ9-Jγ2.3	9581 (53.37)	8371 (46.63)			

*Reads identified as having the correct V-J rearrangement according to the used plasmid.
**Reads identified as having a different V-J rearrangement than one from the plasmid (potentially off-target). Solved with usage of multiple replicates [23].

least partly due to the fact that the Vγ1 Family (Vγ1F) primer covers the Vγ2, Vγ3, Vγ4, Vγ5 and Vγ8 genes. Upon repetitive deep sequencing all correct rearrangements were observed. In order to prevent other rearrangements as indicated in Table 1, multiple replicates were applied. The broader coverage as for the Vγ1F genes also holds for the Jγ primers that cover the Jγ1.1 / Jγ2.1 and Jγ1.3 / Jγ2.3 genes, respectively (TRGJP1 / TRGJP2, and TRGJ1 / TRGJ2 according to IMGT nomenclature [26]). Hereafter Jγ primers are named Jγ1.1/2.1, Jγ1.3/2.3, and Jγ1.2 for the primer recognizing the IMGT TRGJP gene. The first testing round for the multiplex PCR on plasmid pools showed discrepancies in the percentages of expected and observed reads (Fig. 1). All these multiplex

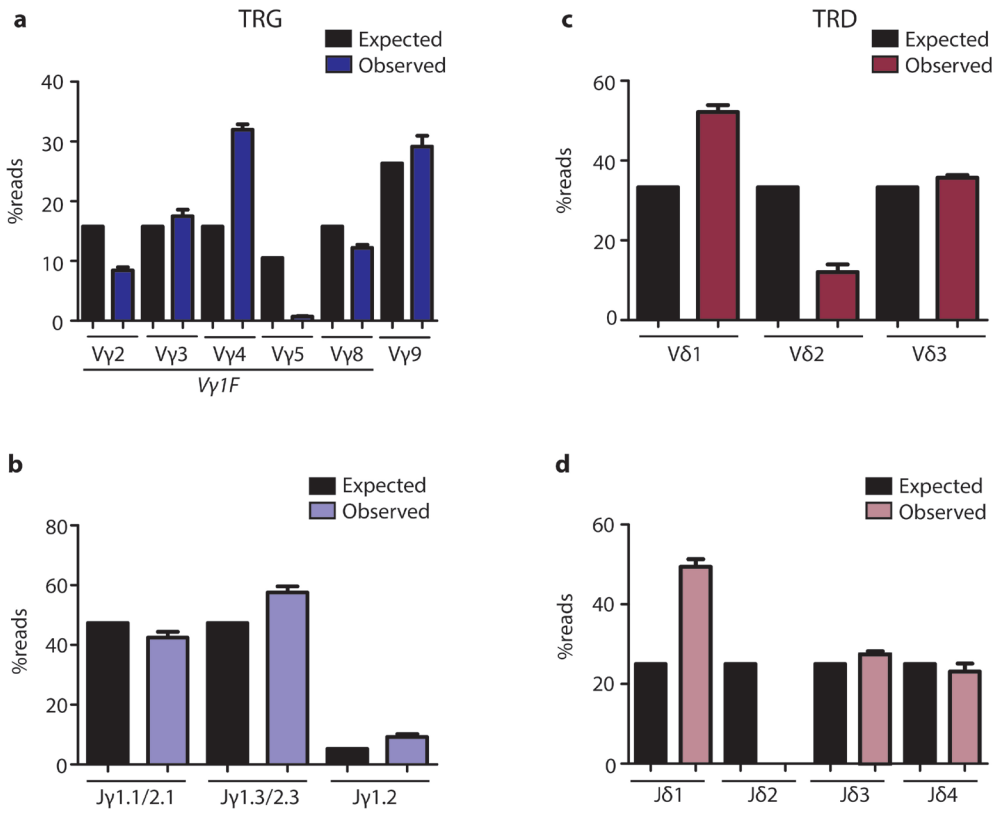


Figure 1. First TRG / TRD multiplex PCR assay on plasmid pools.
Multiplex PCR test on plasmid pools with compositions indicated in Supplementary Table 2, columns A and B, and primer concentrations and assay numbers as summarized in Table 2. %Expected reads were calculated based on the input material of the plasmid pools. %Observed reads represented the percentage of reads identified with the expected V or J gene. TRGV assay with Vγ2/3/4/5/8 genes covered by the Vγ1F primer, and Vγ9 (**a**), TRGJ assay with combined Jγ1.1/2.1 and Jγ1.3/2.3 primers with separate Jγ1.2 primer (**b**). TRDV (**c**) and TRDJ assays (**d**). Bar graphs represent median and SD; replicates per assay N=4.

PCR assays were performed on plasmid pools in four independent replicates. The TRG assay showed differences between expected and observed percentages for some V genes, although correction would be difficult because of the common Vγ1F primer (Fig. 1a). Similarly, the Jγ1.3/2.3 and Jγ1.2 primers also required some correction (Fig. 1b). However, the TRD assay showed most clear discrepancies: Vδ1 and Jδ1 being over-represented, while Vδ2 and Jδ2 were heavily under-represented (Fig. 1c, 1d). Based on these first results we concluded that the primer specificity and exclusivity for the different targets were not affected by the adaptors added for Illumina MiSeq next-generation sequencing technology, but that in the context of multiplex PCR assays primer concentration adjustments were necessary to reduce primer-based amplification bias.

Multiplex PCR fine-tuning with different primer concentrations, annealing temperatures and PCR cycle numbers

The discrepancies as found in the first round of testing could be explained by differences in primer concentration, primer binding efficiency, CG-nucleotide enrichment, Ta's and PCR cycle numbers. Of note, the Illumina forward and reverse adaptors each have a GC-composition of 17/33 and 16/33 nucleotides respectively. Furthermore, the Vγ9, Jδ1 and Jδ3 primers showed the highest %GC content (>50%) (Supplementary Table 2), whereas the Vγ9, Jγ1.2, Vδ3 and Jδ3 primers contained particular GC-stretches (lengths of 4, 5, 4, 6 nt, respectively). As this could highly influence the most optimal Ta of the primers (Supplementary Table 2), and thus impact on performance of the multiplex assay, we then performed primer concentration titration experiments and also varied with Ta's and PCR cycle numbers (Table 2). Since it would be difficult to adjust for gene discrepancies caused by the common Vγ1F primer, we repeated the experiment with similar primer concentrations as used in the initial experiment, which included the first multiplex PCR assay on plasmid pools (Table 2, EXP1/2 column). We did adjust the Jγ primer concentrations in an attempt to correct the discrepancies as observed in the first round. Since the TRD assay showed considerable differences between the %expected and %observed reads, we also adjusted TRD primer concentrations accordingly and formulated four different mixes (Table 2, EXP3 column). Along with the different primer concentrations, we tested several Ta's (58, 60 and 62 ^{0}C), which correlated with the calculated Ta's as indicated in Supplementary Table 1. Furthermore, we tested a reduced PCR cycle number (20 vs. 25 cycles). Each condition, consisting of a specific primer mixture, Ta and PCR cycle number combination, was tested once in order to obtain insights in what the different conditions would possibly lead to (Fig. 2). While using the same mix as in the previous experiment, the %observed values of Vγ1F primers were close to the %expected values (Fig. 2a), as indicated with the dashed

Table 2. Overview of PCR assay fine-tuning experiments.

Variables	EXP2	EXP3				EXP4		EXP5	EXP6		EXP7/8
Material used*											
	TRG 1 and TRD plasmid pools	TRG 2 and TRD plasmid pools				TRG 2 and TRD plasmid pools		gDNA Thy/CB/PB	TRG 2 and TRD plasmid pools		gDNA Thy/CB/PB
Primers (pmol / µl)											
TRG primer mix		G1				G1	G2	G2	G3	G4	G2
TRGV1F	5	5				5	6	6	6.5	7	6
TRGV9	5	5				5	4	4	3.5	3	4
TRGJP1/P2	3.3	4				4	4	4	4.5	5	4
TRGJ1/J2	3.3	2				2	2	2	1.5	1	2
TRGJP	3.3	4				4	4	4	4	4	4
TRD primer mix		D1	D2	D3	D4	D5	D6	D5	D7	D8	D7
TRDV1	3.3	2	2	2	2	3	3	3	2.5	2	2.5
TRDV2	3.3	6	5	6	5	4	4	4	4.5	5	4.5
TRDV3	3.3	2	3	2	3	3	3	3	3	3	3
TRDJ1	2.5	1.25	1	1	1.25	2	2.5	2	2	2	2
TRDJ2	2.5	3.75	4	4	3.75	2.5	2.5	2.5	2	2	2
TRDJ3	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
TRDJ4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	PCR program variables										
Ta (^{0}C)	60	58, 60, 62				58, 59, 60		TRG: 58 TRD: 59	TRG: 58 TRD: 59		TRG: 58 TRD: 59
#Cycles	25	20, 25				20, 25		TRG: 25 TRD: 20	TRG: 25 TRD: 25		TRG: 25 TRD: 25

* Used material for PCR assays either consisted of plasmid pools according to composition summarized in Supplementary Table 2, or of genomic DNA isolated from total TCRγδ+ T cells from thymus material (Thy), neonatal cord blood (CB) and adult peripheral blood (PB) (all N=5).

lines, except for Vγ8 and Vγ9. Different Ta's gave variable results, and the number of cycles also heavily influenced the results. Jγ1.2 %expected and %observed values were in close proximity, with a slight Jγ1.3/2.3 dominance, and also showing varying effects of both Ta and cycle numbers (Fig. 2b). Since in the previous experiment the Vδ2 and Jδ2 genes were highly under-represented, four different mixes were developed with increased Vδ2 and Jδ2 primer concentrations in order to adjust for these under-representations. In all four mixes Vδ2 and Jδ2 were now over-represented, indicating the adjustments were too high. Mixes D2 and D4 showed %expected and %observed values that were more closely together, especially Vδ3 gene was represented well in these mixes (Fig. 2c). Similar results were observed for the Jδ genes, with over-representation of Jδ2 genes. The Jδ3 and Jδ4 %observed and %expected values correlated best (Fig. 2d). Only a small effect of Ta was observed, showing better results at Ta=60. There were few differences in the results due to different PCR cycle numbers; although in some cases 20 cycles gave results closer to %expected values (Fig. 2c, 2d). Based on the results of this first fine-tuning experiment further adjustments in primer concentrations and Ta's were deemed necessary. Furthermore, as the PCR cycle number seemed to

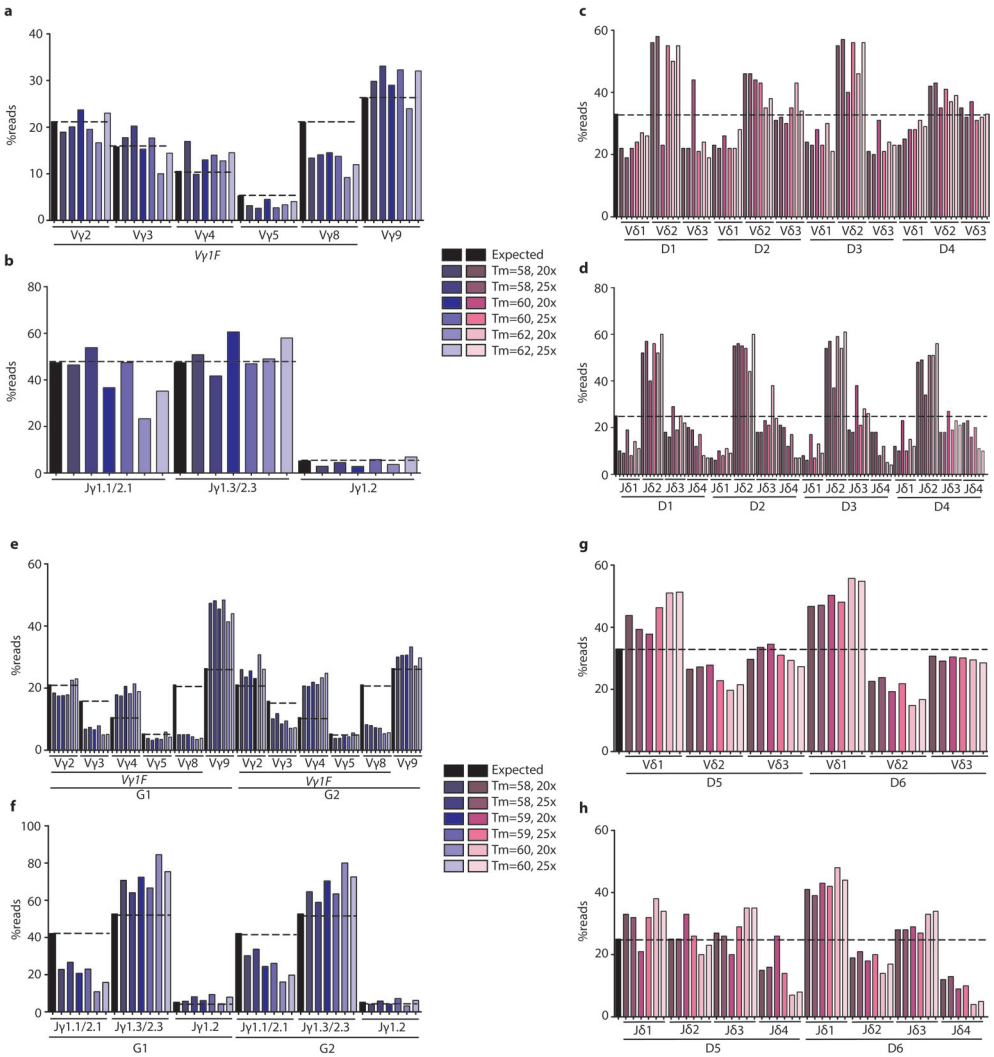


Figure 2. TRG / TRD primer titration and PCR optimization experiments.
Multiplex PCR tests on plasmid pools with compositions as indicated in Supplementary Table 2 and primer concentrations and PCR assay layout as described in Table 2. Different annealing temperatures (Ta) of 58, 60 and 62 ^{0}a (**a-d**) and 58, 59 and 60 ^{0}C (**e-h**), and two different cycle numbers (20x versus 25x) were tested. The different annealing temperatures and cycle numbers are indicated in the legend with different colors. %Expected indicated with black bar graph and dashed line in graphs. Varying %Expected values for TRGV1F genes due to differential plasmid pool compositions. Each mix, in combination with Ta and cycle number, tested once.

have influence on %expected and %observed values, it also still needed further testing. Based on these results we therefore formulated a next optimization experiment, again including different primer concentrations, different Ta's (58, 59 and 60 ^{0}C) and variable (20, 25) PCR cycles (Fig. 2e-h, Table 2 EXP4 column). With the adjustment of TRG

forward primers we generally observed high variability in the %observed values of the Vγ1F genes, but mix G2, with lower Vγ9 and higher Vγ1F primer concentrations, partly reduced the overrepresentation of Vγ9 genes (Fig. 2e). Both TRG mixes contained similar Jγ primer concentrations, and showed largely similar results, albeit with some inter-assay variability (Fig. 2f). For the TRG assay 25 cycles and Ta=58 ^{0}C showed most optimal results. Higher Ta's were associated with increased primer competition of Jγ1.3/2.3 over Jγ1.1/2.1 (Fig. 2f). Since in the initial experiment the TRD assay showed over-representations of Vδ2 and Jδ2 genes, two new mixes were formulated. As in the previous experiment Ta's 58 and 60 ^{0}C showed the best results, different Ta's (58, 59 and 60 ^{0}C) were tested in combination with 20 vs. 25 cycles. Vδ2 primer adjustment indeed resulted in decreased Vδ2 gene detection, with mainly mix D5 showing %observed values closest to %expected values (Fig. 2g). This mix also lead to more %expected values in Jδ genes (Fig. 2h). For the TRD assay Ta=59 ^{0}C showed the best correlations between %expected and %observed values, especially with 20 cycles (Fig. 2g, 2h). Based on these fine-tuning experiments on plasmid pools the most optimal assays (TRG assay with primer concentration mix G2, Ta=58 ^{0}C, 25 cycles; TRD assay with primer concentration mix D5, Ta=59 ^{0}C, 20 cycles; Table 2) were defined for further biological validation studies on genomic DNA.

Initial assay validation on genomic DNA and further fine-tuning experiments result in the most optimal TRG / TRD multiplex PCR assays for NGS purposes

Primary PCR fine-tuning experiments were performed on plasmid DNA, consisting of a 3015 bp vector, and TRD or TRG inserts of 200-400 bp. Even though the use of plasmid DNA is valuable for technical titration experiments, it does not reflect genomic DNA (gDNA) in which the region of interest is more difficult to target with specific primers. Therefore in the next experiment gDNA of total TCRγδ+ T cells was used to further test the most optimal assay as defined in the plasmid fine-tuning experiments. Initially four different thymus (Thy), cord blood (CB) and adult peripheral blood (PB) samples were used to validate the TRG / TRD assays on gDNA. First results showed low TRD absolute read numbers, while the TRG assay showed high numbers, especially for the Thy samples (Supplementary Table 3). The low TRD read numbers could be explained by the lower number of PCR cycles (20) as compared to the TRG assay (25 cycles). Based on the low numbers of productive reads two samples were excluded from further TRD assay analyses (samples Thy11-03 and CB4, Supplementary Table 3). In the Thy samples the frequency of unproductive TRG reads was higher than that of productive TRG sequences, whereas CB and PB samples showed more productive than unproductive TRG rearrangements (Fig. 3a). The naive immature and non-antigen

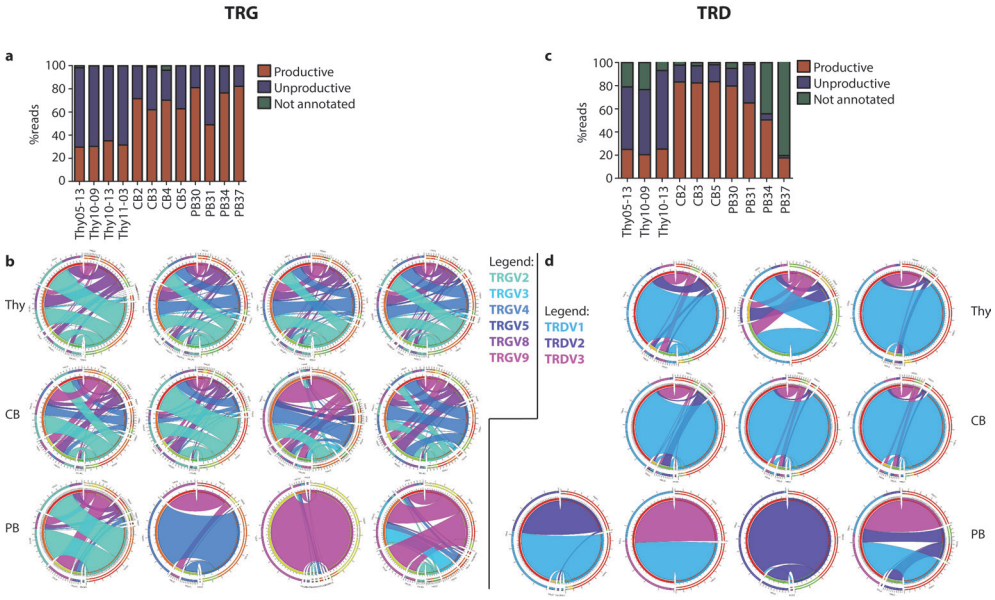


Figure 3. Initial biological validation experiments using the most optimal TRG and TRD PCR assays as based on plasmid pools.
TRG mix G2 with Ta=58.0 and 25 cycles, TRD mix D5 with Ta=59.0 and 20 cycles validation on genomic DNA from total TCRγδ+ T cells from thymus (Thy), cord blood (CB) and adult peripheral blood (PB) samples (N=4 each sample). IMGT annotation resulting in productive, unproductive or not annotated reads (**a, c**). V-J rearrangements depicted as Circoletto plots (www.circos.ca), with Thy, CB and PB samples in the following order (left-right): Thy05-13, Thy10-09, Thy10-13, Thy11-03; CB2, CB3, CB4, CB5; PB30, PB31, PB34, PB37 (**b**) and Thy05-13, Thy10-09, Thy10-13; CB2, CB3, CB5; PB30, PB31, PB34, PB37 (**d**).

selected stage of the thymocytes was also reflected in the diversity of the rearrangements; also, the naive mature TCRγδ+ T cells from CB samples still showed a high diversity. Actually the Thy and CB samples showed similar patterns of rearrangements and diversity, while the patterns observed in the PB samples were donor-specific with skewing towards Vγ9-Jγ1.2 rearrangements (Fig. 3b). The TRD assay yielded significant lower read numbers (Supplementary Table 3), which heavily influenced the data. Nonetheless, the Thy samples again contained more unproductive than productive reads, which was reversed in the CB and PB samples (Fig. 3c). Also, all Vδ genes were observed in the naive immature and mature Thy and CB TCRγδ+ T cells, albeit with a clear Vδ1-Jδ1 dominance. The Thy and CB samples showed similar patterns of rearrangement distributions, in contrast to the more skewed profiles in the adult PB samples, with donor-specific patterns, and Vδ2 dominance (Fig. 3d). Based on the initial biological validation of the TRG / TRD multiplex PCR assays as optimized using plasmid DNA tests, the TRG assay looked promising. TRG read numbers were high and data were in line with previous literature [27,28], albeit that Vγ9 and

Jγ1.3/2.3 were still over-represented, and some Vγ1F genes were under-represented. In contrast, the TRD assay appeared to still require more optimization, due to low read yields using genomic DNA. We therefore formulated two novel TRG mixes in order to further optimize the assay (Table 2, column EXP6). Furthermore, as the TRD assay still showed discrepancies in Vδ1 and Vδ2 %expected vs, %observed values, we formulated two novel TRD mixes as well. Furthermore, in order to obtain sufficient TRD amplicon for sequencing the cycle number was increased to 25. The Ta's were kept at Ta=58 ^{0}C (TRG) and Ta=59 ^{0}C (TRD) based on the results of the plasmid fine-tuning experiments. These additional primer titration experiments (Table 2, columns EXP6), performed in quadruplicates according to Boyd et al. [23] now showed under-representation of Vγ9 (Fig. 4a) and Jγ1.3/2.3 (Fig. 4b) with mix G3, and even more so with mix G4 (Fig. 4c, 4d). Mix G2, which was included in this experiment as well, actually showed the most optimal results (Fig. 4e, 4f). The TRD assay was rather optimal following the final primer concentration adaptations, without significant differences between %expected and %observed of Vδ and Jδ genes (Fig. 4g, 4h). Further Vδ2 primer compensation in mix D8 did not show an obvious difference, but rather created over-representation of Vδ1 and under-representation of Vδ3 genes (Fig. 4g). Jδ primer concentrations did not differ between mixes D7 and D8, and still some differences between %expected and %observed values of all Jδ genes were found (Fig. 4h).

Based on these final fine-tuning experiments we finally chose for the most optimal TRG / TRD multiplex PCR assays, consisting of mix G2 with Ta=58^{0}C and 25 cycles for TRG, and mix D7 with Ta=59^{0}C and 25 cycles for TRD (Table 2).

Sequencing analysis of total TCRγδ+ T cells isolated from different compartments confirmed the most optimal PCR assays

After several rounds of plasmid pool validations, which – despite some level of inter-assay variance – showed high correlations between %expected and %observed values, we then used these most optimal assay designs for evaluating repertoire diversity of total TCRγδ+ T cells from different T cell compartments including thymus, cord blood and peripheral blood. All samples showed high sequencing read yields (except for CB10 TRD assay) and high sequencing reproducibility (Table 3). Thy samples showed relatively more unproductive than productive sequences, while CB and PB samples showed more productive than unproductive rearrangements for both TRG and TRD (Fig. 5a). The low sequencing yield of the TRD assay of CB10 has also lead to higher percentages of sequences that could not be annotated with IMGT. Productive sequences were used for further analysis. The level of diversity could not only be detected through the different V-J combinations, but also through the identification of unique (productive) reads: uniqueness was defined based on V-J combination and amino acid CDR3 (aaCDR3) level. Thy

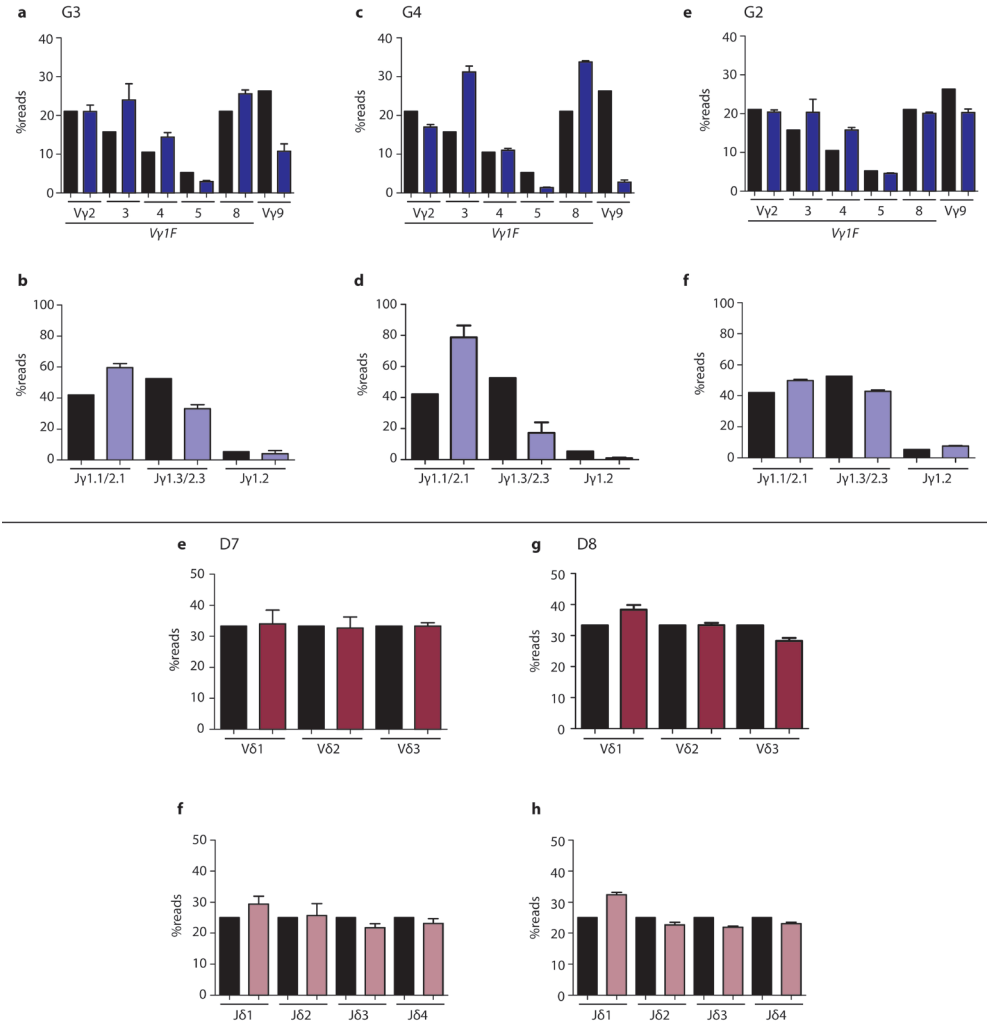


Figure 4. Final PCR fine-tuning experiments on plasmid pool DNA leading to most optimal assays. Further attempts for PCR assay optimization to maximally reduce primer bias using multiplex PCR on plasmid pools with compositions indicated in Supplementary Table 1, and primer concentrations and assay numbers as summarized in Table 2. TRGV results (**a, c, e**) and TRGJ results (**b, d, f**) of mixes G3 (**a, b**), G4 (**c, d**) and G2 as comparison (**e, f**). TRDV results (**e, g**) and TRDJ results (**f, h**) of mixes D7 (**e, f**) and D8 (**g, h**).Ta for TRG was set at 58.0, for TRD 59.0^{0}C. Cycle number for both TRG and TRD PCR assays was 25. Black bars indicate %Expected, colored bars represent the %Observed values. Number of replicates per assay is 4.

samples showed the highest percentage of unique reads among productive sequences, PB samples the lowest percentage, especially the TRG locus (Fig. 5b). Productive rearrangements were visualized with stacked bar graphs, showing a high diversity in all Thy samples, a generally high diversity - albeit with some skewing - in CB samples, and a high level of skewing and selection in PB samples (Fig. 5c, left panel). Productive TRD

Table 3. NGS read for TRG and TRD rearrangements using TCRγδ+ T cells derived from different T cell compartments.

Sample	TRG assay				TRD assay			
	Productive	*Unproductive*	*No results*	*Total*	*Productive*	*Unproductive*	*No results*	*Total*
Thy04-10								
Replicate #1	17903	28695	173	46771	20943	14980	202	36125
Replicate #2	14902	24687	167	39756	28570	20915	262	49747
Replicate #3	15658	26457	150	42265	11812	9125	280	20217
Replicate #4	16499	26141	145	42785	16261	11240	246	27747
Thy05-13								
Replicate #1	12430	25865	107	38402	9539	12367	1524	23430
Replicate #2	12238	25274	123	37635	11468	15023	59	26550
Replicate #3	12984	27084	128	40196	6862	9520	66	16448
Replicate #4	13350	28861	110	42321	17461	24528	130	42119
Thy10-09								
Replicate #1	12444	26303	115	38862	10949	19156	37	30142
Replicate #2	11542	24104	95	35741	8787	16874	22	25683
Replicate #3	11341	23200	148	34689	9251	14584	39	23874
Replicate #4	12357	25981	142	38480	6768	12084	29	18881
Thy10-13								
Replicate #1	13189	28229	145	41563	6626	14600	147	21373
Replicate #2	11628	27805	146	39579	7346	16789	68	24203
Replicate #3	13094	31104	145	44343	7368	17302	32	24702
Replicate #4	12866	29929	151	42946	9639	21421	31	31091
Thy11-03								
Replicate #1	12532	29057	176	41765	9560	29921	140	39621
Replicate #2	14429	32963	177	47569	10111	34270	124	44505
Replicate #3	12520	30037	150	42707	8474	25590	83	34147
Replicate #4	13987	31798	197	45982	9371	27828	176	37375
CB2								
Replicate #1	27620	13722	193	41535	29168	5705	38	34911
Replicate #2	28743	13585	236	42564	32125	7156	73	39354
Replicate #3	25894	12059	257	38210	31973	5994	121	38088
Replicate #4	22356	10416	287	33059	33605	6328	47	39980
CB3								
Replicate #1	24851	14430	239	39520	43227	10422	79	53728
Replicate #2	25620	15258	282	41160	41965	9605	54	51624
Replicate #3	26622	14696	376	41694	39769	9472	105	49346
Replicate #4	24241	13912	282	38435	39786	9461	63	49310
CB4								
Replicate #1	24642	11082	181	35905	26260	5062	110	31432
Replicate #2	22939	13013	111	36063	29235	5513	60	34808
Replicate #3	8758	4454	96	13308	33926	5936	72	39934
Replicate #4	4160	2209	137	6506	41914	7699	77	49690

rearrangements generally showed less diversity due to fewer Vδ-Jδ combinations, but the diversity of Thy and CB samples was reasonably broad, despite the Vδ1 predominance. In contrast, PB samples showed a higher level of skewing and selection for even Vδ2 (Fig. 5c, right panel). The Thy and CB samples showed low levels of inter-individual variance, whereas PB samples showed donor-specific patterns (Fig. 5c). Finally, diversity was analyzed by means of CDR3-length distributions, showing clear Gaussian

(Table 3 continued).

Sample	TRG assay				TRD assay			
	Productive	*Unproductive*	*No results*	*Total*	*Productive*	*Unproductive*	*No results*	*Total*
CB9								
Replicate #1	171028	43102	211	214341	39958	7725	79	47762
Replicate #2	32031	8674	148	40853	47669	8292	92	56053
Replicate #3	31340	8391	134	39865	41036	7394	97	48527
Replicate #4	32539	9196	110	41845	45641	9187	70	54898
CB10								
Replicate #1	23436	10791	134	34361	1325	246	277	1848
Replicate #2	21050	11497	63	32610	66	21	297	384
Replicate #3	11948	6283	70	18301	2090	334	249	2673
Replicate #4	16641	9001	116	25758	117	47	271	435
PB30								
Replicate #1	29267	7677	81	37025	46652	6440	236	53328
Replicate #2	28932	8094	83	37109	45057	6414	388	51859
Replicate #3	31701	8559	71	40331	37594	6008	389	43991
Replicate #4	31767	9169	97	41033	41228	5805	256	47289
PB31								
Replicate #1	22892	16048	57	38997	27634	17058	334	45026
Replicate #2	20998	15695	64	36757	29187	18539	280	48006
Replicate #3	21033	15202	51	36286	30062	19045	308	49415
Replicate #4	21820	16230	64	38114	32780	19342	338	52442
PB34								
Replicate #1	27376	7992	60	35428	48444	6459	328	55231
Replicate #2	28079	8059	51	36189	55831	7352	432	63615
Replicate #3	28952	8214	46	37212	44545	5033	335	49913
Replicate #4	30121	9090	46	39257	45685	5893	290	51868
PB37								
Replicate #1	27949	5154	74	33177	36637	4872	117	41626
Replicate #2	29989	5258	69	35316	33303	4627	103	38033
Replicate #3	26138	4956	56	31150	32631	4403	102	37136
Replicate #4	27601	5024	53	32678	40518	5323	119	45961
PB50								
Replicate #1	29151	5366	49	34566	45766	7713	105	53584
Replicate #2	26634	4814	62	31510	50506	7403	121	58030
Replicate #3	26690	4415	54	31159	37819	5576	111	43506
Replicate #4	27289	4949	60	32298	47334	6539	142	54015

distributions in Thy and CB samples in both the TRG and TRD assays, with some level of selection in CB TCRγδ+ T cells, and high level of skewing in PB samples (Fig. 5d).

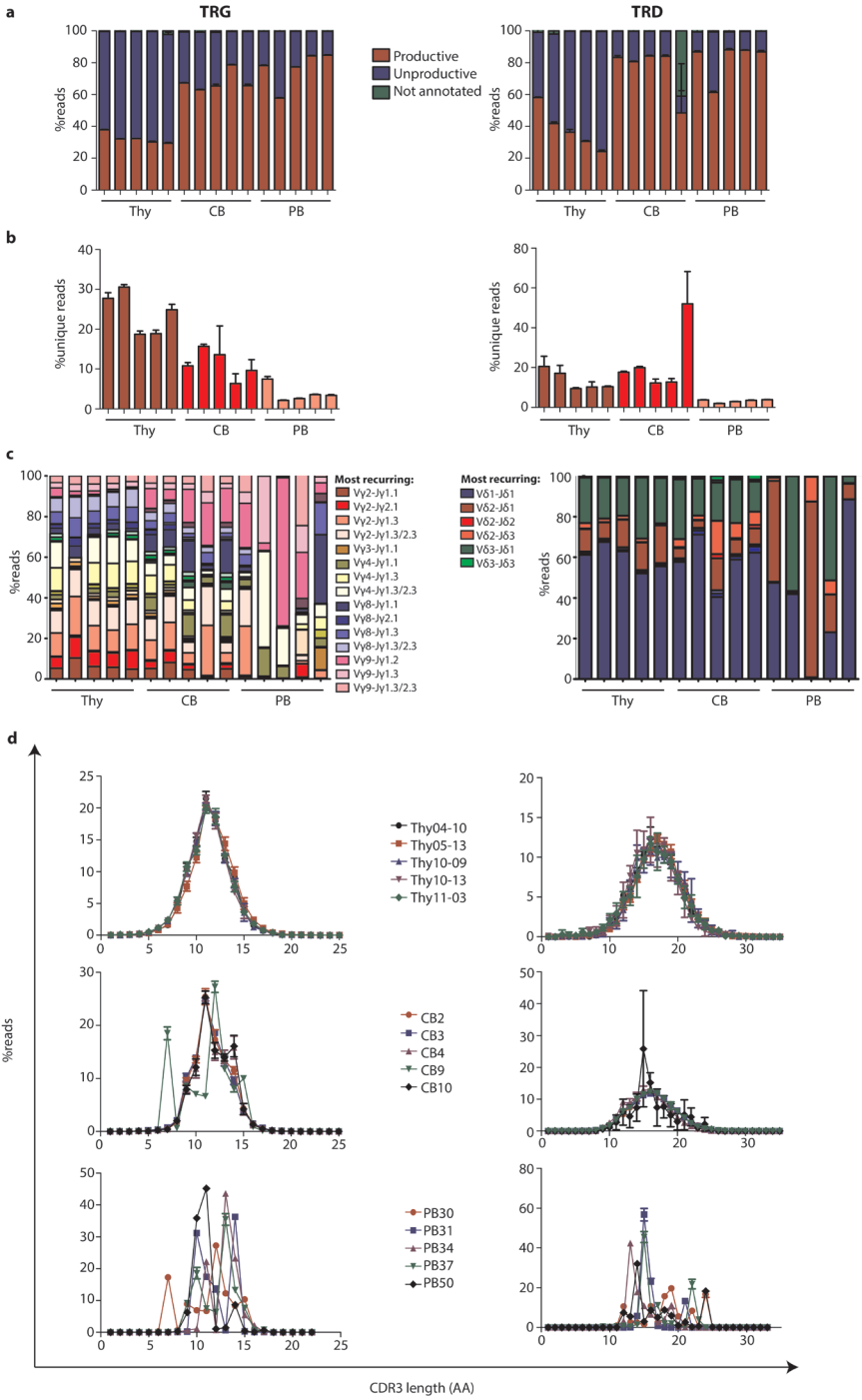

a
TRG
TRD
Productive
Unproductive
Not annotated
%reads
Thy
CB
PB
b
%unique reads
c
Most recurring:
Vγ2-Jγ1.1
Vγ2-Jγ2.1
Vγ2-Jγ1.3
Vγ2-Jγ1.3/2.3
Vγ3-Jγ1.1
Vγ4-Jγ1.1
Vγ4-Jγ1.3
Vγ4-Jγ1.3/2.3
Vγ8-Jγ1.1
Vγ8-Jγ2.1
Vγ8-Jγ1.3
Vγ8-Jγ1.3/2.3
Vγ9-Jγ1.2
Vγ9-Jγ1.3
Vγ9-Jγ1.3/2.3
Most recurring:
Vδ1-Jδ1
Vδ2-Jδ1
Vδ2-Jδ2
Vδ2-Jδ3
Vδ3-Jδ1
Vδ3-Jδ3
d
%reads
Thy04-10
Thy05-13
Thy10-09
Thy10-13
Thy11-03
CB2
CB3
CB4
CB9
CB10
PB30
PB31
PB34
PB37
PB50
CDR3 length (AA)

Figure 5 (see left page). Compartment analysis and assay validation on genomic DNA from total TCRγδ+ T cells from thymus, cord blood and adult peripheral blood.
Distribution of productive sequences, unproductive sequences and sequences which could not be annotated through IMGT analysis (no results). IMGT results with annotations productive, unproductive or not annotated (**a**). Frequencies of unique reads within total productive reads (**b**). V-J rearrangements depicted in stacked bar graphs; legends indicate most recurring rearrangements in bar graphs (**c**). All bar graphs indicated with median and SD error bars. CDR3-length distributions (in amino acids) indicated with median frequencies of reads of a certain length with SD error bars (two-sided) (**d**).

DISCUSSION

Recent advances in immune repertoire sequencing methods have increased exponentially over the past 10-15 years with the development of next-generation DNA / RNA sequencing methods and high-throughput sequencing platforms. For many years, Sanger sequencing [29] served as the gold standard for sequencing in all fields [30], but due to a high demand for a faster, higher, cheaper and more accurate sequencing yield in immune repertoire studies several different platforms have been developed, [31-34]. Currently, most sequencing methods rely on cyclic-array sequencing, involving the sequencing of dense DNA strands by iterative cycles of enzymatic manipulation. The data collection is based on imaging, by means of pyrosequencing (Roche 454, Illumina) [35] or chemical gradients (IonTorrent). In our study we used Illumina MiSeq paired-end sequencing to study the build-up of the TRG and TRD immune repertoire.

Applying NGS to immune repertoire studies is highly prone to errors at different levels in the total workflow, starting from experimental design, to sample preparation, to PCR amplification including primer bias and PCR cycling temperatures, to finally sequencing, imaging and data analysis: numerous reviews have been devoted to the identification and possible solutions for such sequencing errors [36-39]. Quality of the input DNA can be achieved by quick isolations without the use of formalin fixative; in order to prevent preferential amplification 5' rapid amplification of cDNA ends (5'RACE) primers can be used, and in order to reduce further experimental errors the sequencing depth can be increased and replicates can be applied. Here in the current study we attempted to address the most important problems foreseen with immune repertoire sequencing by quick isolation of DNA without the use of formalin fixative, titrations of primer concentrations and fine-tuning of PCR cycling temperatures. Previously, the use of artificial DNA as spike-ins has been applied as a quantitative resource against genetic variation measurement versus sequencing errors [40]. In their study these authors created artificial human DNA based on the human genome (Hg38) which they inserted into vectors and therefore clear distinction between human genome and artificial DNA could be performed. The use of these artificial DNAs is particularly useful for constituting internal

controls that enable quantification of artifacts and variability that occurs during sample preparation, processing and comparing different samples.

In our study we applied a similar approach by using cloned TRG and TRD rearrangements in vectors to carefully quantify the output with varying primer concentrations and PCR cycling temperatures in order to fully optimize TRG and TRD assays to assess quantitative features of the TRG and TRD repertoire. However, even with the use of technical replicates, which involves the repeated processing, sequencing and analysis of the exactly the same sample, and biological replicates, using multiple biological samples from the same input DNA under similar conditions [36,37], still some level of inter-assay variance was observed. This phenomenon was also observed earlier by Deveson *et al*. [40]. These variances could have arisen from intrinsic sequence-specific biases during e.g. sample preparation or library preparation, the sequencing procedure and alignment to the IMGT database [19-22]. Also, such variance could be originating from aliquot variability and pipetting variation [40]. Nevertheless, the majority of the variability was reproducible and especially in case of the genes in the Vγ1F inevitable since one primer covers the Vγ2, Vγ3, Vγ4, Vγ5 and Vγ8 genes. Despite the inter-assay variability and the coverage of multiple TRGV genes by one common primer, the results from our analysis of TCRγδ+ T cells from different compartments were quite in line with previously reported results [27,28]. Results on the biological samples were also in accordance with the hypothesis that the thymus and cord blood compartments show a relatively highly diverse repertoire, with generally low inter-sample variability, whereas selection and skewing would be most apparent in adult peripheral blood samples, with clear donor-specific patterns.

Not only should PCR assay aspects of immune repertoire sequencing methods be addressed when it comes to reducing PCR and sequencing biases and errors. Experimental design could be highly underestimated when it comes to the material from which the target DNA is isolated, but also the number of replicates that are required for optimal data analysis when using quantitative features of the sequence data [23,36,37]. Therefore we used in screening experiments single samples, while in validation experiments at least 4 replicates per sample were included. Furthermore it is important to formulate different mixes, perform the PCR programs in different machines, even when using the same primer concentration combinations, annealing temperatures and cycle numbers (technical replicates), and in different types of samples (plasmid DNA versus genomic DNA, including genomic DNA samples from different body compartments; biological replicates).

After performing the sequencing experiments, adequate data analysis is vital in order to reduce the number of false positive and false negative variants that are being identified. In case of immune repertoire sequencing the determination of variants can

be challenging, since it can be difficult to distinguish between falsely annotated variants and biologically relevant variants [37]. Here we used Illumina MiSeq paired-end sequencing, upon which directly after sequencing quality controls are performed and paired-end sequences are merged, which increases the overlapping region quality. In order to determine diversity in the TRG and TRD loci the CDR3-region, which is determined through V-J combination and the junctional region at the amino acid or nucleotide level, is of most interest. Since this is the overlapping region, which is of most importance in immune repertoire studies, we could in this way exclude low quality sequences as well: reads that could not be aligned, have a high probability of containing sequencing or PCR errors, and are thus excluded. Furthermore, the ARGalaxy pipeline analysis [12] excludes possible scarce clones by only focusing on unique sequences, based on V-J combinations and amino acid or nucleotide CDR3 regions, as also an important manner to exclude falsely identified sequences. Additionally, ARGalaxy pipeline also contains a feature to calculate so-called clonality scores, based on identical sequences that are present in all technical and/or biological replicates, dependent on the study design. In this way, falsely identified sequences should be removed as well, presenting immune repertoire sequencing data as much filtered as possible, which optimize the available data set for evaluation of quantitative features [12].

Other approaches to even further exclude PCR and sequencing errors and biases could be to perform single cell sequencing, preferably on mRNA/cDNA level to determine the actual productive receptor as being expressed on the sorted cell. However, this still is a more time consuming and labor intensive approach, especially due to the required sorting procedures. In order to obtain sufficient cells and sequences as being representative for the individual, at least a few thousand cells should be sorted. For analyzing TRG and TRD repertoires this could be challenging due to the scarcity of these cells in the healthy adult peripheral blood [41]. In order to obtain sufficient numbers of these cells, especially in case of studying different subsets such as naive, effector or memory cells, large amounts of blood are thus needed, leading to extensive sorting protocols. Sorting procedures might influence DNA quality, but definitely influence mRNA quality for later cDNA-based repertoire evaluation. However, the greatest advantage would be that from one single cell the combined TRG / TRD chains could be determined, and that the full receptor can thus be visualized which has further implications for insight in normal development, repertoire shaping, and ultimately in infections and malignancies.

In conclusion, through extensive plasmid pool validations and additional assay adjustments with respect to primer concentrations, annealing temperatures and cycle numbers upon tests with genomic DNA, the most optimal TRG and TRD multiplex PCR assays for NGS methods were defined, which also showed compartment analysis results correlating with previous literature. In this way we have formulated the most optimal

multiplex PCR assays for TRG / TRD repertoire analysis with the lowest PCR bias and inter-assay variance as possible. Further applications could therefore be used in diagnostic tests, in for instance the diagnosis of leukemic clones, but also in the quantitative detection of minimal residual disease [43,44].

ACKNOWLEDGEMENTS

The research for this manuscript was (in part) performed within the framework of the Erasmus Postgraduate School Molecular Medicine.

REFERENCES

1. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
2. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59 (2008).
3. McKernan, K.J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527-1541 (2009).
4. Cebula, A. *et al.* Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota. *Nature* **497**, 258-262 (2013).
5. Ee, R., Lim, Y.L., Yin, W.F. & Chan, K.G. De novo assembly of the quorum-sensing Pandoraea sp. strain RB-44 complete genome sequence using PacBio single-molecule real-time sequencing technology. *Genome Accounc.* **2**, e00245-14 (2014).
6. Bahassi, E.M. & Stambrook, P.J. Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis* **29(5)**, 3303-3310 (2014).
7. Nederbragt, L. Developments in NGS, figshare. https://doi.org/10.6084/m9.figshare.100940.v9 (2016).
8. Mardis, E.R. DNA sequencing technologies: 2006-2016. *Nat. Protoc.* **12(2)**, 213-218 (2017).
9. Sims, D., Sudbery, I., Ilott, N.E., Heger, A. & Ponting, C.P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121-132 (2014).
10. Moorhouse, M.J. *et al.* ImmunoGlobulin Galaxy (IGGalaxy) for simple determination and quantification of immunoglobulin heavy chain rearrangements from NGS. *BMC Immunol.* **15**, 59 (2014).
11. Gudmundsdottir, J. *et al.* T-cell receptor sequencing reveals decreased diversity 18 years after early thymectomy. *J. Allergy Clin. Immunol.* **S0091-6749(17)**, 31344-31351 (2017).

12. Ijspeert, H., van Schouwenburg, P.A., van Zessen, D., Pico-Knijnenburg, I., Stubbs, A.P. & van der Burg, M. Antigen Receptor Galaxy: a user-friendly, web-based tool for analysis and visualization of T and B cell receptor repertoire data. *J. Immunol.* **198(10)**, 4156-4165 (2017).
13. Schneeberger, K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* **15**, 662-676 (2014).
14. Shpak, M., Ni, Y., Lu, J. & Müller, P. Variance in estimated pairwise genetic distance under high versus low coverage sequencing: the contribution of linkage disequilibrium. *Theor. Popul. Biol.* **117**, 51-63 (2017).
15. Van Dongen, J.J.M. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* **17**, 2257-2317 (2003).
16. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30(5)**, 614-620 (2013).
17. Moorhouse, M.J. *et al.* ImmunoGlobulin galaxy (IGGalaxy) for simple determination and quantitation of immunoglobulin heavy chain rearrangements from NGS. *BMC Immunology* **15**, 59 (2014).
18. Blankenberg, D., Gordon, A., von Kuster, G., Coraor, N., Taylor, J. & Nekrutenko, A. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26(14)**, 1783-1785 (2010).
19. Alamyar, E., Giudicelli, V., Shuo, Li., Duroux, P. & Lefranc, M-P. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8**, 1:2 (2012).
20. Alamyar, E., Duroux, P., Lefranc, M.P. & Giudicelli, V. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* **882**, 569-604 (2012).
21. Li, S. *et al.* IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat. Commun.* **4**, 2333 (2013).
22. Giudicelli, V., Duroux, P., Lavioe, A., Aouinti, S., Lefranc, M-P. & Kossida, S. From IMGT-ONTOLOGY to IMGT/HighVQUEST for NGS immunoglobulin (IG) and T cell receptor (TR) repertoires in autoimmune and infectious diseases. *Autoimmun. Infec. Dis.* **1(1)** (2015).
23. Boyd, S.D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1(12)**, 12ra23 (2009).
24. Krzywinksi, M.I. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19(9)**, 1639-1645 (2009).
25. Hartigan, J.A. & Hartigan, P.M. The Dip Test of Unimodality. *Ann. Stat.* **13**, 70-84 (1985).
26. Lefranc, M.P. & Lefranc, G. The T cell receptor factsbook. Academic Press. ISBN 0-12-441352-8 (2001).
27. Breit, T.M., Wolvers-Tettero, I.L.M. & van Dongen, J.J.M. Receptor diversity of human T cell receptor γδ expressing cells. *Prog. Histochem. Cytochem.* **26**, 182-183 (1992).
28. Sandberg, Y. *et al.* TCRγδ+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRγδ+ T cells. *Leukemia* **20**, 505-513 (2006).

29. Sanger, F., Nicklen, S. & Couldson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467 (1977).
30. Metzker, M.L. Emerging technologies in DNA sequencing. *Genome Res.* **15**, 1767-1776 (2005).
31. Marziali, A. & Akeson, M. New DNA sequencing methods. *Annu. Rev. Biomed. Eng.* **3**, 195-223 (2001).
32. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135-1145 (2008).
33. Metzker, M.L. Sequencing technologies – the next generation. *Nat. Rev. Gen.* **11**, 31-46 (2010).
34. Pickrell, W.O., Rees, M.I. & Chung, S-K. Next generation sequencing methodologies – an overview. *Adv. Protein Chem. Struct. Biol.* Volume 89, ISSN 1876-1923 (2012).
35. Ahmadian, A., Ehn, M. & Hober, S. Pyrosequencing: history, biochemistry and future. *Clin. Chim. Acta.* **363**, 83-94 (2006).
36. Robasky, K., Lewis, N.E. & Church, G.M. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Gen.* **15**, 56-62 (2014).
37. Friedensohn, S., Khan, T.A. & Reddy, S.T. Advanced methodologies in high-throughput sequencing of immune repertoires. *Trends Biotechnol.* **35(3)**, 203-214 (2017).
38. Benichou, J., Ben-Hamo, R., Louzoun, Y. & Efroni, S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* **135**, 183-191 (2011).
39. Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H. & Quake, S.R. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32(2)**, 158-168 (2014).
40. Deveson, I.W. *et al.* Representing genetic variation with synthetic DNA standards. *Nat. Methods* **13(9)**, 784-791 (2016).
41. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13(2)**, 88-100 (2013).
42. Mori, A., Deola, S., Xumerle, L., Mijatovic, V., Malerba, G. & Monsurrò, V. Next generation sequencing: new tools in immunology and hematology. *Blood Res.* **48**, 242-249 (2013).
43. Langerak, A.W. *et al.* High-throughput immunogenetics for clinical and research applications in immunohematology: potential and challenges. *J. Immunol.* **198(10)**, 3765-3774 (2017).
44. Wren, D. *et al.* Comprehensive translocation and clonality detection in lymphoproliferative disorders by next-generation sequencing. *Haematologica* **102(2)**, e57-e60 (2017).

SUPPLEMENTAL DATA

Supplementary Table 1. Primers for cloning and Illumina MiSeq sequencing technology.

Primer name	%GC content of target primer (excl. Rd1/Rd2 adaptors)†	Most optimal Ta target primer (^{0}C)•	Complete fusion primer for Illumina MiSeq analysis‡
Vγ1F	47	58.7	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GGTTGTGTTGGAATCAGGAGTCA**
Vγ9	52	57.8	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CGGCACTGTCAGAAAGGAATC**
Jγ1.1/2.1	39	53.4	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**AGTTACTATGAGCTTAGTCCCTT**
Jγ1.3/2.3	50	58.2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**GTGTTGTTCCACTGCCAAAGAG**
Jγ1.2	50	58.2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**TAAGCTTTGTTCCGGGACCA**
Vδ1	38	53.8	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ATGCAAAAAGTGGTCGCTATT**
Vδ2	39	52.8	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ATACCGAGAAAAGGACATCTATG**
Vδ3	36	55.7	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TTTGTCTTTTATGGGGATAACAGCA**
Jδ1	57	60.3	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**GTTCCACAGTCACACGGGTTC**
Jδ2	48	55.3	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**GTTCCACGATGAGTTGTGTTC**
Jδ3	67	64.8	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**CTCACGGGGCTCCACGAAGAG**
Jδ4	48	54.8	TCGCGAGTTAATGCAACGATCGTCGAAATTCGC**TTGTACCTCCAGATAGGTTCC**

All primers are given in 5'-3' format. Target sequences are indicated in bold.
† %GC content of target primers only. %GC content of Rd1 and Rd2 adaptors are 17/33 (52%) and 16/33 (48%) respectively.
• Most optimal annealing temperatures of target primer (excluding Rd1/Rd2 adaptor overhang) according to ThermoFischer Ta calculator.
‡ Primers used for first PCR assay in Illumina MiSeq experiment protocol.

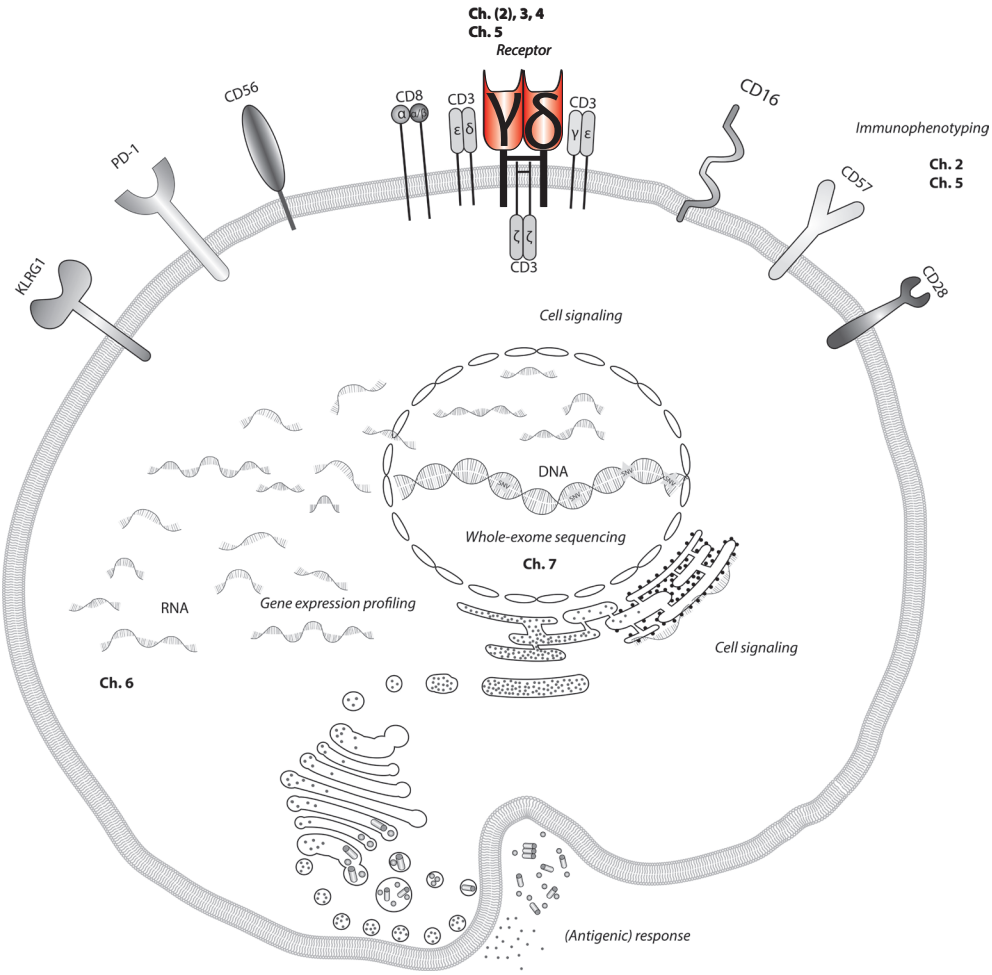
Supplementary Table 2. Plasmid pool compositions for TRG / TRD primer titration and PCR fine-tuning experiments.

TRD plasmid pool	TRG plasmid pool 1	TRG plasmid pool 2
Vδ1 – Jδ1	Vγ2 – Jγ1.1	Vγ2-Jγ2.1
Vδ1 – Jδ2	Vγ2 – Jγ2.1	Vγ2-Jγ1.3
Vδ1 – Jδ3	Vγ2 – Jγ1.3	Vγ2-Jγ2.3
Vδ1 – Jδ4	Vγ2 – Jγ2.3	Vγ2-Jγ1.3/2.3
Vδ2 – Jδ1	Vγ3 – Jγ1.1	Vγ3-Jγ1.1
Vδ2 – Jδ2	Vγ3 – Jγ2.1	Vγ3-Jγ2.1
Vδ2 – Jδ3	Vγ3 – Jγ1.3	Vγ3-Jγ1.3
Vδ2 – Jδ4	Vγ4 – Jγ2.1	Vγ4-Jγ2.1
Vδ3 – Jδ1	Vγ4 – Jγ1.3	Vγ4-Jγ1.3
Vδ3 – Jδ2	Vγ4 – Jγ2.3	Vγ5-Jγ1.3
Vδ3 – Jδ3	Vγ5 – Jγ1.1	Vγ8-Jγ1.1
Vδ3 – Jδ4	Vγ5 – Jγ2.3	Vγ8-Jγ2.1
	Vγ8 – Jγ1.1	Vγ8-Jγ1.3
	Vγ8 – Jγ2.1	Vγ8-Jγ2.3
	Vγ8 – Jγ1.3	Vγ9-Jγ1.1
	Vγ8 – Jγ2.3	Vγ9-Jγ2.1
	Vγ9 – Jγ1.1	Vγ9-Jγ1.2
	Vγ9 – Jγ2.1	Vγ9-Jγ1.3
	Vγ9 – Jγ1.2	Vγ9-Jγ2.3
	Vγ9 – Jγ1.3	
	Vγ9 – Jγ2.3	

Supplementary Table 3. Absolute TRG and TRD read numbers in the initial biological validation experiment.

Sample*	Joined reads	IMGT annotated	Productive reads	Unique reads**
TRG				
Thy05-13	45564	44768	13493	3941
Thy10-09	40584	40534	12324	3423
Thy10-13	100771	100234	35344	4546
Thy11-03	45802	45720	14454	3428
CB2	4977	4968	3555	1321
CB3	11798	11663	7303	2780
CB4	12326	11835	8649	1108
CB5	5482	5483	3448	1999
PB30	32005	31976	25945	1051
PB31	13609	13584	6653	167
PB34	16464	16356	12568	392
PB37	15105	15108	12425	558
TRD				
Thy05-13	1126	890	282	200
Thy10-09	194	124	36	18
Thy10-13	274	213	58	57
Thy11-03	6269	-	-	-
CB2	139	128	109	108
CB3	155	138	117	116
CB4	3730	-	-	-
CB5	431	394	336	322
PB30	477	417	351	75
PB31	1974	1828	1210	35
PB34	3676	1988	1799	135
PB37	6794	1292	1162	128

*Thymus samples obtained from cryobank Erasmus MC, Rotterdam, cord blood and peripheral blood freshly obtained from healthy donors. Sample numbering according to anonymous sample annotations in the database.
Paired-end sequencing reads joined using PEAR [16], and analyzed using IMGT annotation database [19-22].
**Uniqueness of reads defined by V-J recombination and amino acid CDR3 region using ARGalaxy [12].


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 7
RNA
Gene expression profiling
Cell signaling
Ch. 6
(Antigenic) response

Chapter 4

Next-generation sequencing analysis of the human TCRγδ+ T cell repertoire reveals differential Vγ- and Vδ-usage in memory populations upon ageing

Martine J. Kallemeijn[1], François G. Kavelaars[2], Michèle Y. van der Klift[1], Ingrid L.M. Wolvers–Tettero[1], Peter J.M. Valk[2], Jacques J.M. van Dongen[1+], Anton W. Langerak[1]

[1]Laboratory for Medical Immunology, Department of Immunology, Erasmus University Medical Center, Rotterdam, the Netherlands;
[2]Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands
[+]Current address: Dept. of Immunohematology and Blood Transfusion (IHB), LUMC, Leiden, the Netherlands.

Published in Front. Immunol. ***9****,* 448 (2018)

ABSTRACT

Immunological ageing remodels the immune system at several levels. This has been documented in particular for the T cell receptor (TCR)αβ+ T cell compartment, showing reduced naive T cell outputs and an accumulation of terminally differentiated clonally expanding effector T cells, leading to increased proneness to autoimmunity and cancer development at older age. Even though TCRαβ+ and TCRγδ+ T cells follow similar paths of development involving V(D)J-recombination of TCR genes in the thymus, TCRγδ+ T cells tend to be more subjected to peripheral selection rather than central selection. Age-related chronic (antigenic) stimulation could contribute to TCRγδ+ T cell large granular lymphocyte (T-LGL) leukemia pathogenesis. However, the role of ageing in shaping of the peripheral TRG / TRD repertoire remains largely elusive. NGS analysis methods were optimized using plasmid vector DNA-samples based on a spike-in method for accurate TRG / TRD receptor diversity quantification, resulting in optimally defined primer concentrations, annealing temperatures and cycle numbers. Thymic and cord blood samples were used to study TRG / TRD repertoire diversity during TCRγδ+ T cell ontogeny, showing a broad and diverse repertoire with Gaussian CDR3-length distributions, in contrast to the more skewed repertoire in mature circulating total TCRγδ+ T cells in adult peripheral blood (PB). During ageing the naive repertoire maintained its diversity with Gaussian CDR3-length distributions, while in the central and effector memory populations a clear shift from Vγ9/Vδ2 dominance in young to Vγ2/Vδ1 dominance in elderly was observed. Together with less clear Gaussian CDR3-length distributions, this is highly suggestive of differentially heavily selected repertoires. Despite the age-related shift from Vγ9/Vδ2 to Vγ2/Vδ1 no clear ageing-effect was observed on the Vδ2 invariant T nucleotide and canonical Vγ9-Jγ1.2 selection determinants. When zooming in on specific TRG and TRD clonotypes, pathogen-related clonotypes such as for CMV were identified in a few donors without a significant ageing-effect, while *M.tuberculosis*-specific clonotypes were absent in the repertoire of healthy individuals. Interestingly, TCRγδ+ T-LGL leukemia-related receptor chains were identified in effector subsets of elderly individuals. This could correlate with the concept that TCRγδ+ T-LGL leukemia is an ageing-associated disease, originating from normal antigen-experienced effector cells that are present in peripheral blood of elderly.

Keywords: TCRγδ+, development, ageing, repertoire, next-generation sequencing.

INTRODUCTION

Immunological ageing, also referred to as immunosenescence, is a complex phenomenon consisting of senescence and exhaustion processes, which are characterized by different functional and marker expression profiles [1,2]. Immunosenescence acts on different levels in the immune system e.g. reduced antigen-specific responses [3], thymic shrinkage, and a significantly reduced naive T cell output [3-5], convergence of the innate and adaptive immunity [6], and ultimately T cell exhaustion [1]. Immunosenescence is believed to play a major role in shaping of the antigen receptor repertoire of T cells.

T cells develop in the thymus, where they undergo commitment, rearrangement, selection and maturation processes. The main event during T cell development is the rearrangements of the variable (V), diversity (D) and joining (J) genes of the T cell receptor (TR) loci, in order to establish a large diversity of antigen receptors [7,8]. Two main types of T cells are generated; firstly, TCRγδ+ thymocytes, through early TR delta and gamma (TRD, TRG) rearrangements, then followed by TCRαβ+ thymocytes upon TR beta and alpha (TRB, TRA) rearrangements [8]. TCRαβ+ thymocytes undergo positive selection through TCR signaling to subsequently mature into functional T cells, followed by negative selection in order to eliminate self-reactive T cell precursors [9]. In contrast, TCRγδ+ thymocytes do not undergo positive and/or negative selection in the thymus [10], but extrathymic development and peripheral (antigenic) selection of TCRγδ+ T cells have been described [11].

TCRγδ+ T cells appear to be the first functional population of circulating T lymphocytes in both murine and human PB (reviewed in [12]). In the human fetal and neonatal situation these functional circulating TCRγδ+ T cells mainly concern Vδ1+ cells. Readily after birth and during further development to adulthood a switch occurs in the circulating TCRγδ+ T cell population with the number of Vδ1+ cells decreasing and Vγ9/Vδ2 cells becoming the predominant TCRγδ+ T cell types [13]. This process is believed to be the result of peripheral antigenic selection, exemplified by the presence of an invariant T nucleotide in the majority of the selected Vδ2 – Jδ1 rearrangements [13-15]. Furthermore, pathogens or other antigens providing epitopes that could stimulate and select TCRγδ+ T cell types have been described: *Mycobacterium tuberculosis* has been found to be a major stimulator of Vγ9/Vδ2 cells in both infected lungs and PB [16], whereas non-Vγ9/Vδ1 cells are known to be stimulated by viruses, such as cytomegalovirus (CMV) [17,18] and Epstein-Barr virus (EBV) [19]. TCRγδ+ T cells do not only recognize antigens via their receptor, but they also respond to lipid antigens presented on CD1d-molecules, and that are associated with stress, inflammation and cancer (reviewed by [20]). Most TCRγδ+ T cells recognizing these CD1d-lipid antigen complexes are Vδ1 or Vδ3 cells, commonly located in the gut [21]. TCRγδ+ T cells can

also recognize butyrophilins (BTN), tumor-antigens, endothelial antigens, antigen-presenting cells (APCs) and Toll-like Receptors (TLRs) (reviewed by [22]), all of which are postulated to contribute to shaping of the TCRγδ+ T cell repertoire.

TCRγδ+ T cell recognition and selection has been mostly described in the context of the developing immune system from fetus to neonate and adulthood, but – contrary to the TCRαβ+ T cell repertoire – effects of ageing on the TCRγδ+ T cell repertoire have not been extensively addressed. Since it has been found that TCRγδ+ T cells follow the classical ageing model as found in mainly CD8+TCRαβ+ T cells [23], we hypothesized that the naive mature TCRγδ+ T cell repertoire depicts a broad spectrum of rearrangements, which will show a more skewed pattern during further development from neonates to young adults and eventually elderly individuals. Here, we investigated the developing and ageing TRG / TRD repertoire in TCRγδ+ T cell subsets, using optimized experimental next-generation sequencing (NGS) procedure to minimize technical biases of PCR-based methods. Our data show subset- and donor-specific TRG / TRD repertoires, suggestive of selection, with significant differences in the combinatorial repertoire in especially memory populations between young and elderly individuals. Furthermore, when looking closer into TRG / TRD clonotypes, TCRγδ+ T-LGL leukemia receptor chains could be traced in especially the effector subsets of elderly individuals, suggesting that TCRγδ+ T-LGL leukemia cells originate from the normal healthy antigen-experienced TCRγδ+ T cells.

MATERIALS & METHODS

Subjects and materials

Blood from healthy blood donors from Sanquin Blood Bank (Amsterdam, The Netherlands) in the age ranges 20-35 (young adults, N=11) and 56-70 (elderly, N=12) was used upon written informed consent at the blood bank (project number NVT0012.01) and anonymized for further use. The maximum age to donate blood is 70 years. Healthy neonatal cord blood was obtained postpartum or after Caesarian section through collaboration and upon written informed consent at the departments of Obstetrics and Hematology. Cord blood was drawn using CB Collect bags containing citrate phosphate dextrose (CPD) solution as anticoagulant. Thymic lobes were removed upon heart surgery in individuals under the age of two years upon written informed consent from parents. Both cord blood and thymus material was obtained under Medical Ethics Committee approval (project number hmPOO2004-003). Whole thymic material was sliced and prepared prior to cryopreservation. Peripheral blood mononuclear cells (PBMCs) and cord blood mononuclear cells (CBMCs) were obtained through Ficoll

density gradient separation. Isolated PBMCs, CBMCs, and thymocytes were cryopreserved in Iscove's Modified Dulbecco's Medium (IMDM, Lonza, Basel, Switzerland) with dimethyl sulfoxide and stored in vials at -180 °C until further use. All studies were conducted in accordance with the principles of the Declaration of Helsinki.

Cell sorting

Cryopreserved material was thawed and sorted using CD3, CD45, TCRαβ, TCRγδ, CD45RA, CD45RO, CD27 and CD197 antibodies (Supplementary Table 1) to obtain TCRγδ+ naive (CD45RA+CD27+CD197+), central memory (CD45RA-CD45RO+CD27+CD197+), effector memory (TemRO population defined as CD45RA-CD45RO+CD27-CD197-), and effector (TemRA population, CD45RA+CD27-CD197-) T cells (Supplementary Fig. 1). Cell sorting was performed with FACS Aria I and III instruments (BD Biosciences, San Jose, CA, USA).

DNA isolation

Following isolation, cells were lysed and subjected to DNA isolation using the DNA/RNA/miRNA AllPrepKit according to the manufacturer's protocol (Qiagen, Hilden, Germany). DNA concentration and quality (A260/A280 absorption ratio) were determined by Nanodrop measurements (Thermo Fischer Scientific, Waltham, MA, USA).

Primer design

Primers for cloning and Illumina-based sequencing were largely based on those reported in BIOMED-2 assays [24]. The Vδ3 primer was redesigned to better fit amplicon length of PCR products generated with the existing Vδ1 and Vδ2 primers. The Jγ1.2 primer was newly designed, as this primer was not included in the BIOMED-2 TRG assay. Vγ1F and Jγ1.3/2.3 primers were adjusted compared with the BIOMED-2 protocol (Supplementary Table 2). Primers were adapted for Illumina-based sequencing by adding Illumina forward (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3') overhang adaptor sequences to the respective primers. The second PCR, by means of these overhang adaptor sequences, attaches sample-specific dual indices for sample identification and Illumina sequencing adaptors using primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina, San Diego, CA, USA).

Plasmid pool preparation

Primer validation and titration was done using plasmid vectors with cloned TRD and TRG gene rearrangements. All possible V-J gene combinations were PCR amplified and cloned from immature T cell lines [25] and thymus DNA into the pGEM T-Easy vector in

a 3:1 insert:vector ratio according to the manufacturer's protocol (Promega, Madison, WI, USA). Composition of the plasmid pools is summarized in Supplementary Table 3.

Assay optimization experiments

PCRs were first tested in singleplex and multiplex settings with varying primer concentrations, annealing temperatures and PCR cycle numbers. Each initial PCR mix contained GeneAmp PCR Buffer II (1x), magnesium chloride (2.5 mM), deoxynucleotides (2.0 mM) and AmpliTaqGold (1 U) (Thermo Fischer Scientific). Total forward and reverse primer(s) amounts were 10 pmol. The PCR protocol was largely based on the BIOMED-2 publication [24], with varying annealing temperatures (Tm = 58/59/60/62) and different numbers of cycles (20 and 25 cycles). In general total forward and reverse primer(s) amount was 10 pmol. Primer concentration adjustment, and optimization of annealing temperatures and number of PCR cycles were based on the results of iterative optimization experiments as summarized in Supplementary Table 4 and Figures 2 and 4 from Chapter 3.

Amplicon preparation

Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA), whereafter concentrations were measured with the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fischer Scientific), and the amplicons were adjusted to similar concentrations. The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina) using the KAPA HiFi HotStart PCR Kit (Kapa Biosystems, Wilmington, MA, USA). Second PCR amplicons were evaluated with agarose gel electrophoresis or PicoGreen concentration measurement. Library pool preparation was performed based on the gel image or PicoGreen measurement results. The library pool was further purified with Agencourt AMPure XP beads and normalized from Illumina-based sequencing, according to the manufacturer's protocol (Illumina).

Next-generation sequencing

Paired-end next-generation sequencing (2x221 bp) was performed on the MiSeq platform (Illumina, San Diego, CA, USA) with the use of an Illumina MiSeq Reagent Kit V3, according to the manufacturer's protocol (Illumina).

Bio-informatic data analysis

Illumina NGS data was obtained in FASTQ format. Paired-end reads were combined using the FASTQ-join tool in the Erasmus MC Galaxy Server [26], with the use of usegalaxy.org [27-29] converted from FASTQ to FASTA with the converter tool [30].

Sequencing annotations were made via the IMGT High V-quest database [31-34]. Calculation of the clonality score for multiple replicates was based on the algorithm described by Boyd et al. [35]. Clonal type definition was based on V and J gene usage and CDR3-region at the nucleotide level. Rearrangements were visualized using Circoletto plots (www.circos.ca, [36]). CDR3 amino acid compositions were visualized using WebLogo online tool (www.weblogo.berkeley.edu, [37,38])

Statistical analysis

Data was checked for normal distributions using the Hartigan's Dip Test Statistic for Unimodality package [39-41] in R version 3.4.1 [42]. All statistical analyses were performed with Prism 5 (GraphPad, La Jolla, CA, USA).

RESULTS

Multiplex PCR assay fine-tuning leads to an optimized, bias-free NGS assay for reliable quantification of the TRG / TRD repertoire

The multiplex PCR assay to be analyzed by NGS was optimized and fine-tuned for more accurate quantification and receptor diversity analysis of the TRG and TRD loci using a diverse set of artificial DNA spike-in samples and primer concentration titration experiments (Supplementary Table 4 and 5, Fig. 2 and 4 from Chapter 3). Each artificial DNA sample, represented by plasmid vector DNA, contained a mixture of known V(D)J rearrangements, cloned from either immature T cell lines or thymus DNA in equimolar proportions (Supplementary Table 3). After several rounds of fine-tuning (Fig. 2 and 4 from Chapter 3), and repeated technical validation with plasmid spike-in pools we established the most optimal PCR conditions for both TRG (Fig. 1a) and TRD (Fig. 1b) multiplex assays in view of unbiased NGS data. Slight differences between observed and expected read frequencies are due to random chance introduced by PCR and inevitable inter-assay variation, which were reduced to minimal using approx. four replicates for each sample, which included differently pipetted mixes to reduce pipetting bias and the use of four different PCR machines to reduce machine-dependent bias. These optimization experiments resulted in variable primer concentrations and defined annealing temperatures and cycle numbers for the TRG and TRD multiplex PCR reactions (Supplementary Table 5).

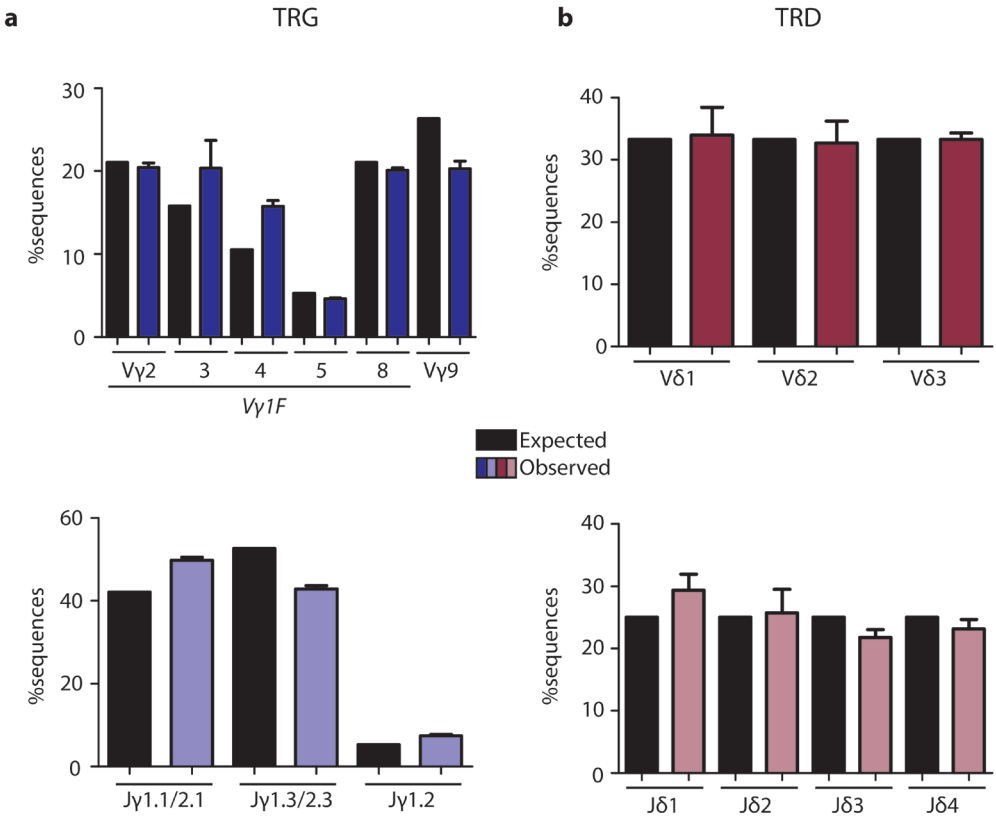


Figure 1. Technical optimization of NGS assays for TRG / TRD loci.
Multiplex PCR assays were optimized with balanced primer concentrations, annealing temperatures and cycle numbers as summarized in Supplementary Table 3. Plasmid pools were used as spike-in samples to determine the percentage expected sequences per V and J gene vs. the observed percentage after sequencing (Supplementary Table 2). Expected percentages are indicated in black bars, observed percentages are indicated in colored bars. TRG assays showed high overlap between frequencies of expected and observed sequences for Vγ and Jγ (**a**) genes, with some variation due to single primers covering multiple genes (Vγ1F covering Vγ2-8, Jγ1.1/2.1 covering Jγ1.1 and Jγ2.1, and Jγ1.3/2.3 covering Jγ1.3 and Jγ2.3). TRD assays showed nearly similar percentages of expected and observed sequences for Vδ and Jδ (**b**) genes. Error bars represent SD of PCR replicates (N=3/4).

The TRG / TRD repertoire is diverse in immature thymus and cord blood, and more skewed in mature circulating TCRγδ+ T cells

In order to determine changes in TCRγδ+ T cell repertoire in healthy individuals we first investigated TRG / TRD repertoire diversity during ontogeny using purified TCRγδ+ T cells from different compartments i.e. thymus (Thy) and neonatal cord blood (CB). In addition we sequenced the total mature TCRγδ+ T cell population of healthy adult peripheral blood (PB) samples. TRG rearrangements in Thy and CB samples were highly diverse (Fig. 2a, upper 2 rows), and the inter-sample variation of especially Thy samples

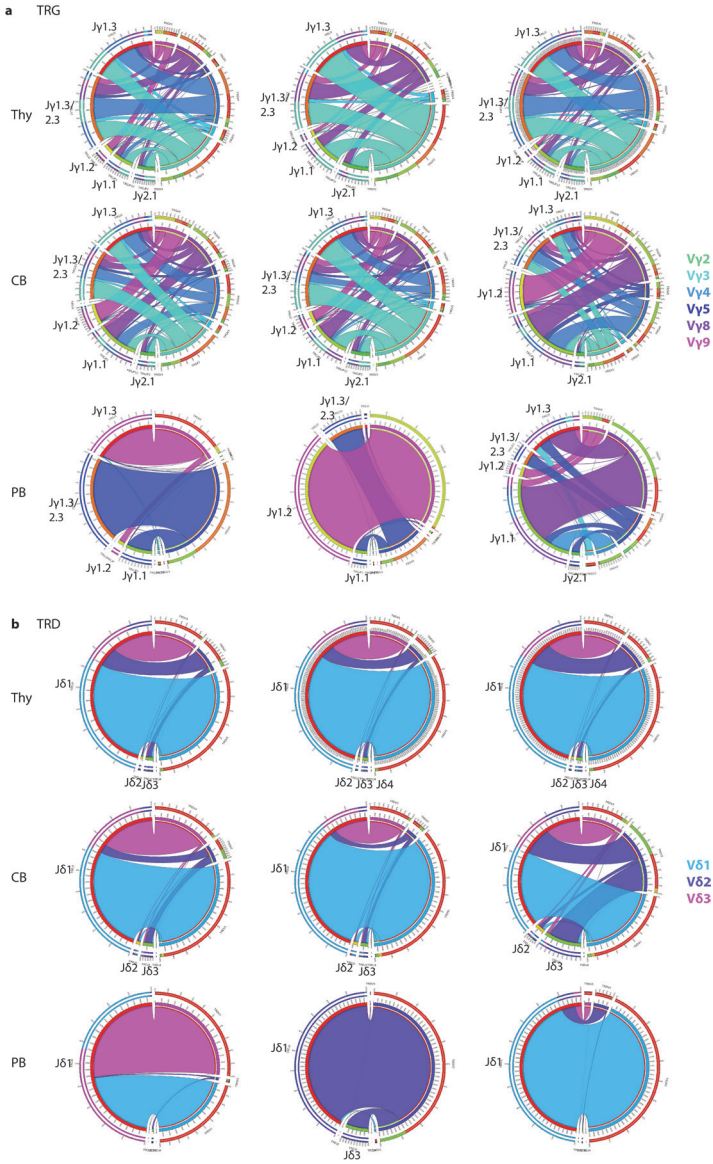


Figure 2. Circoletto visualization of the TRG / TRD repertoire during ontogeny.
Optimized multiplex PCR NGS assays were applied on total TCRγδ+ T cells sorted from thymus (Thy), neonatal cord blood (CB) and adult peripheral blood (PB). TRG assays showed high repertoire diversity in both Thy and CB samples, with low inter-individual variation, while adult PB samples showed individual-specific repertoire patterns with less receptor diversity (**a**). TRD assays showed high dominance of Vδ1 (light blue bars), which was also observed in CB samples, both with low inter-sample variation. Adult PB samples showed donor-specific patterns with sometimes skewing towards Vδ2 and even Vδ3 (**b**). Three representative samples of each samples type are visualized: Thy04-10, Thy05-13, Thy10-13, CB2, CB3, CB4, PB30, PB31 and PB50. Plots were made using the Circoletto online software tool (www.circos.ca, [35]). Each band represents a V-J rearrangement, with colors based on V-gene usage.

was low, in keeping with the non-selected character of the TCRγδ+ T cells in these compartments. These findings were in strong contrast to PB samples, which showed a high level of skewing and predominance of certain receptors (including Vγ9-Jγ1.2 sequences) were observed (Fig. 2a, bottom row), albeit with high inter-individual differences, illustrating the dominant role of (antigenic) selection. TRD diversity was less apparent, although in Thy and CB samples (Fig. 2b, upper 2 rows) all three predominant Vδ-genes were identified. Again, inter-sample variation was low, illustrating the non-selected character of Thy and CB cells. Inter-sample variation was more evident for the PB samples (Fig. 2b, bottom row), with predominance of Vδ2 usage, but also Vδ3 usage in some cases, reflecting different types of (antigenic) selection between individuals.

Collectively, these data confirmed our hypothesis of a broad and diverse TRG / TRD repertoire in the immature Thy and CB samples and a more skewed TRG / TRD repertoire in mature circulating TCRγδ+ T cells in adults, thereby validating our optimized multiplex PCR-based TRG / TRD NGS assays.

Upon ageing memory TCRγδ+ T cells show shifts in V-gene usage, whereas naive and effector populations do not

As it has become evident that ageing plays a major role in shaping the elderly immune system [43], we next evaluated the role of ageing on the combinatorial TRG / TRD repertoire. To this end, we sorted TCRγδ+ T cells from healthy young (N=11; age range 20-35) and elderly (N=12; age range 56-70) individuals into four subsets: naive (CD45RA+CD45RO-CD27+CD197+), central memory (CD45RA-CD45RO+CD27+CD197+), effector memory (TemRO; CD45RA-CD45RO+CD27-CD197-) and effector (TemRA; CD45RA+CD45RO-CD27-CD197-) TCRγδ+ T cells. Subset distributions of young and elderly individuals (Supplementary Fig. 2) correlated with those from our previous ageing study, as it is known that TCRγδ+ T cells have little CCR7 expression and thus low absolute and relative numbers of naive and central memory cells [44]. Even though the spectrum of V-J combinations for both TRG and TRD varied in a donor-specific way between individuals (Supplementary Fig. 3), the overall TRG/ TRD combinatorial diversity appeared to be mostly determined by differences in Vγ / Vδ usage rather than Jγ / Jδ gene usage.

Naive TCRγδ+ T cells of both young and elderly individuals showed a relatively diverse TRG repertoire, which was in strong contrast to (central and effector) memory TCRγδ+ T cells that showed dominant Vγ9 gene usage. Effector TCRγδ+ T cells of both age groups were more diverse again. Of note, significant differences between young and elderly could be observed in mainly the memory populations, as reflected by a significant increase of Vγ2-usage in central memory TCRγδ+ T cells in elderly, as well as significantly increased Vγ2-8-usage and significantly decreased Vγ9 gene usage in effector memory cells of elderly (Fig. 3a).

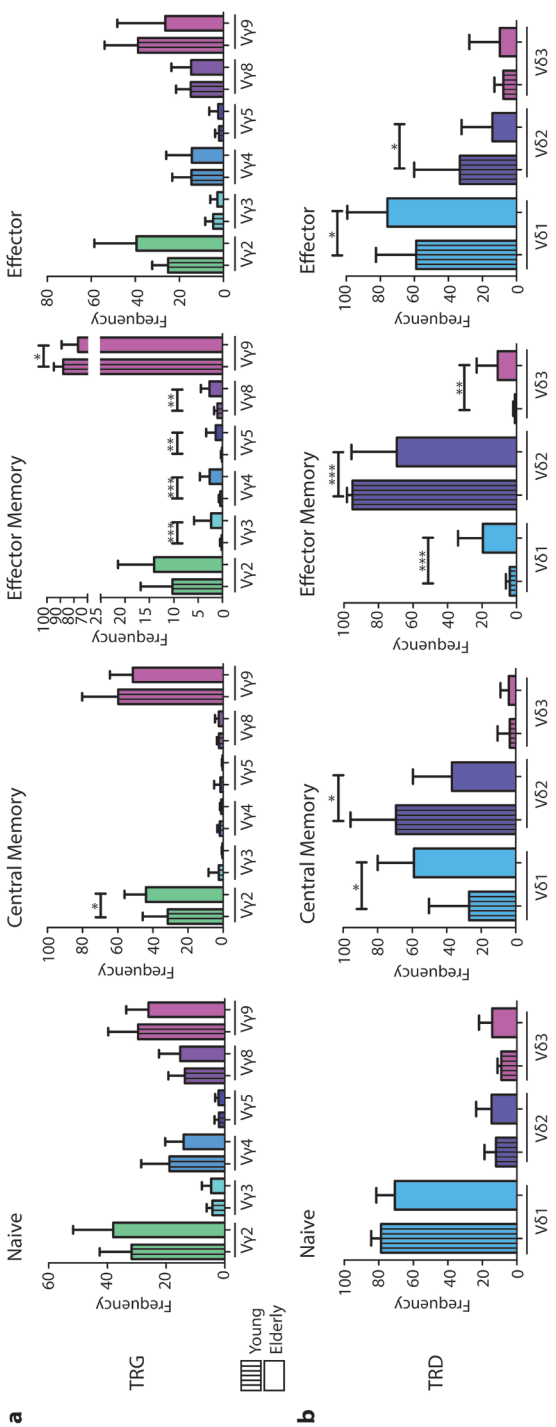


Figure 3. V gene diversity in the combinatorial TRG / TRD repertoire of different TCRγδ+ T cell subsets from young and elderly individuals. V-gene usage per subset visualized in bar graphs indicating a diverse repertoire in naive and effector subsets of both young and elderly, with skewing towards Vγ2 and Vγ9 in memory populations (**a**). Naive populations showed diversity in Vδ usage with Vδ1 dominance in young and elderly. Young individuals showed relatively higher Vδ2 usage in especially the memory and effector subsets, whereas elderly individuals showed clear Vδ1 dominance in all subsets, except for effector memory cells (**b**). Median V-gene frequencies of productive sequences with SD bars were indicated in bar graphs for young (N=11) and elderly (N=12) individuals. Color legends are indicated in the figure. Statistical significance was tested using the Mann Whitney U test. Level of significance is indicated in the plots: * , p<0.05; ** , p<0.01; *** , p<0.0001.

When comparing TRD combinatorial profiles in the different subsets between young and elderly individuals, significantly increased Vδ1 and significantly decreased Vδ2 gene usage was observed in memory populations of elderly individuals. This effect was also observed in the effector population. In the effector memory cells of elderly Vδ3 gene usage was also significantly increased (Fig. 3b).

Overall, these data show clear differences in the TRG / TRD combinatorial repertoire between naive TCRγδ+ T cells on the one hand and especially memory TCRγδ+ T cells on the other hand. Notably, the clear dominance of Vγ9 and Vδ2 usage in memory and effector TCRγδ+ T cells in young individuals was less prominent in elderly individuals, who on average showed significant shifts towards more Vγ2 and Vδ1 gene usage in addition to Vγ9 and Vδ2. Most significant differences between young and elderly were identified in central and effector memory populations.

The TRG / TRD junctional region repertoire shows signs of selection in memory and effector cell populations of both young and old individuals

For a more detailed view of the TRG / TRD repertoire we then studied CDR3-regions, which reflect the most relevant antigen-binding part of the antigen receptors. These CDR3-length distributions are indicative of the junctional repertoire. TRG / TRD CDR3-length distributions of Thy TCRγδ+ T cells showed Gaussian profiles, just like the TRD CDR3-length distributions of CB TCRγδ+ T cells; TRG CDR3-lengths of CB TCRγδ+ T cells showed less clear Gaussian distributions and more prominent peaks, probably reflecting low-level selection (Supplementary Fig. 4). The effect of selection became even more evident in adult individuals; naive TCRγδ+ T cells showed mostly Gaussian CDR3 profiles, in contrast to memory and effector TCRγδ+ T cells of young individuals, which showed dominant peaks for both the TRG and TRD CDR3-regions (Fig. 4a, 4b). Elderly individuals did not show clear Gaussian profiles, and even prominent peaks in all subsets, thus reflecting a more heavily selected repertoire (Fig. 4a, 4b). The average TRG and TRD CDR3-lengths were not markedly different between young and elderly individuals.

TRG canonical and TRD invariant T selection determinants are detectable in normal TCRγδ+ T cells but do not increase upon ageing

During development selection of TCRγδ+ T cells is known to be associated with so-called selection determinants, which represent molecular fingerprints in the CDR3-regions of TRG and TRD chains. In circulating TCRγδ+ T cells a high frequency of Vγ9 – Jγ1.2 recombinations with preferential joining at the GCA sequence has been noted (Fig. 5a). We therefore studied this so-called canonical Vγ9 – Jγ1.2 rearrangement, characterized by a defined CDR3-length and amino acid composition (Fig. 5a), in different subsets of young and elderly healthy controls. Approximately 10-20% of all

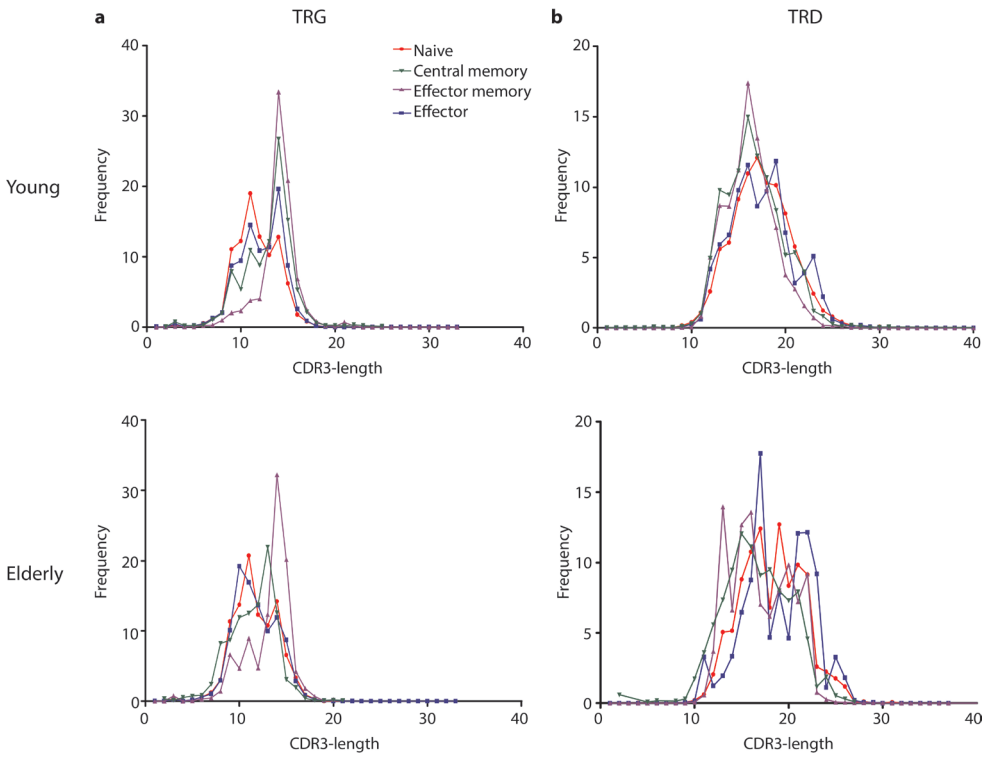


Figure 4. CDR3-length distributions of TCRγδ+ T cell subsets from young and elderly individuals. Frequencies of CDR3-lengths from different TCRγδ+ T cell subsets are summarized per TCR receptor chain and age group. Naive subsets (red lines) showed distributions resembling Gaussian profiles, while memory subsets showed dominant peaks suggestive of selection and receptor skewing (green and purple lines). TRG CDR3-lengths showed similar distributions between young and elderly (**a**), TRD CDR3-lengths in young individuals showed a Gaussian profile in the naive subset and more skewing in memory and effector subsets, whereas in elderly individuals all subsets showed dominant peaks (**b**). Mean frequencies per subset were indicated for young (N=11) and elderly (N=12) individuals. Data normality was tested using the diptest package in R.

productive Vγ9 – Jγ1.2 rearrangements contained the canonical sequence (Fig. 5b). The frequencies of canonical Vγ9 – Jγ1.2 sequences did not clearly differ between different subsets in young and elderly (Fig. 5b). In TCRγδ-receptors the canonical Vγ9 – Jγ1.2 chain is frequently combined with a Vδ2-derived chain, especially resulting from Vδ2 – Jδ1 recombination. These Vδ2 – Jδ1 rearrangements often contain a so-called invariant T nucleotide, a selection determinant at the relative second position of the first codon of the junctional region (Fig. 5c), translating into Leucine (L), Valine (V) or Isoleucine (I) amino acids at that first codon in the junction. The invariant T was observed in all individuals (Fig. 5d, outer grey circles), and resulted in L, V or I amino acids at this position (Fig. 5d, inner blue pie charts). The invariant T was present at higher frequency

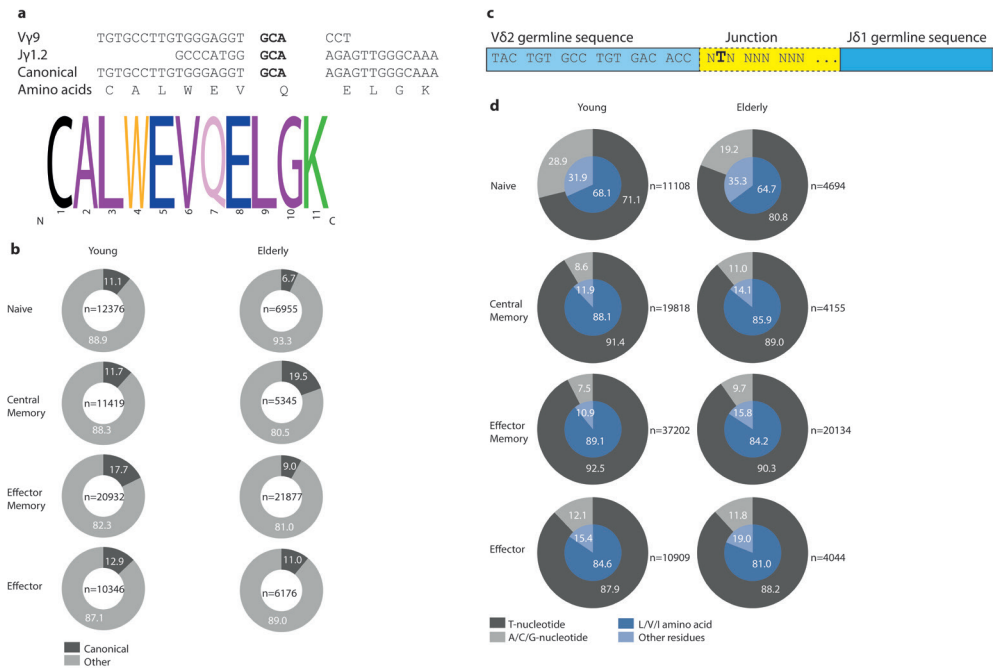


Figure 5. Canonical Vγ9 – Jγ1.2 sequences and Vδ2 – Jδ1 invariant T selection determinant in young and elderly healthy individuals.
GCA sequence in bold mediates preferential recombination of Vγ9 with Jγ1.2, resulting in specific CDR3 length and amino acid composition, adapted from [61] (**a**). Frequencies of all productive Vγ9 – Jγ1.2 sequences containing the canonical sequence per age group and subset marked in dark grey. In light grey Vγ9 – Jγ1.2 rearrangements with other, non-canonical sequences are indicated. Numbers of total Vγ9 – Jγ1.2 sequences are indicated in center of plots (**b**). The invariant T selection determinant is located at the relative second position of the first codon in the junction of Vδ2 – Jδ1 rearrangements. Adapted from [13] (**c**). Frequencies of invariant T selection determinant in young and elderly subsets (dark grey part outer ring), leading to L, V or I amino acid residues (inner blue pie chart, dark blue part. Absolute number of unique productive Vδ2 – Jδ1 sequences indicated next to plots (**d**).

in memory and effector subsets compared to naive TCRγδ+ T cell subsets. On average, invariant T frequencies per subset did not differ much between young and elderly individuals, although the percentage of invariant T-containing sequences at the nucleotide level of naive TCRγδ+ T cells of young individuals was clearly lower than that of elderly naive TCRγδ+ T cells (Fig. 5d).

Taken together, the most common selection determinants described in TCRγδ+ T cells (i.e. the Vγ9 – Jγ1.2 canonical sequence and the Vδ2 – Jδ1 invariant T nucleotide) were readily identified in different TCRγδ+ T cell subsets in our healthy control cohort, albeit that frequencies did not clearly differ between young and elderly.

Analysis of TRG / TRD clonotypes shows the presence of TCRγδ+ T-LGL leukemia related clonotypes in especially effector cells of elderly

In view of TCRγδ+ T cell selection processes, we then studied the possible recurrence of specific TRG / TRD clonotypes in the repertoire of young and elderly individuals, as a sign of activated TCRγδ+ T cell clones. To this end multiple replicates (N=3) of each TCRγδ+ T cell subset were studied in independent PCR reactions and the number of so-called coincident sequences was determined [35,45]. In all subsets, both young and elderly, the frequency of clonotype sequences found in only one of the replicates was the highest, while the frequencies of coincidences found in 2 or 3 replicates were relatively low for both TRG and TRD (Supplementary Fig. 4). When comparing young and elderly, small shifts leading to higher numbers of coincidences in 2 or 3 replicates were seen in the latter (Supplementary Fig. 5). We then only focused on the coincidences present in all three replicates, since these sequences best reflect the individuals' repertoire selection. Especially in the effector memory population absolute numbers of sequences found in all three replicates were higher, while in naive subsets from both young and elderly these numbers were lower, except for a few cases (Supplementary Table 5).

To understand if the recurrence of clonotypes would be associated with particular infections, we next evaluated receptor clonotypes linked to pathogens such as *Mycobacterium tuberculosis* [16,46] and herpesviruses such as cytomegalovirus (CMV) [18]. Whereas in our healthy controls no *M.tuberculosis*-specific clonotypes could be identified, CMV-specific TRG or TRD clonotypes were found in most controls, and in one case even a complete CMV-specific TCRγδ receptor could be identified (data not shown). However, there were no evident differences between young and elderly individuals.

Finally, as leukemic TCRγδ+ T cells can be associated with specific clonotypes, we retrospectively reviewed our TCRγδ+ T-LGL leukemia database of clonal TRG / TRD sequences [13,46] and searched for these LGL clonotypes in the normal TCRγδ+ T cell repertoire of young and elderly. Interestingly, two TCRγδ+ T-LGL leukemia-associated TRG and TRD clonotypes were found in four older individuals and in in one young individual (Table 1). The Vδ3 – Jδ1 receptor as identified in TCRγδ+ T-LGL leukemia case 12-098 was identified in one young individual (naive subset, 26-year-old female), and in three older individuals (naive subset, 56-year-old male; effector subset, 69-year-old female and 68-year-old male) (Table 1). The TCRγδ+ T-LGL leukemia related receptor from case 10-200 was found twice in older individuals (naive subset, 56-year-old male; effector subset, 70-year-old male) (Table 1). Although the numbers are low, the fact that two TCRγδ+ T-LGL leukemia related receptors could specifically be identified in effector cells of elderly would support the idea that TCRγδ+ T-LGL leukemia cells originate from the normal TCRγδ repertoire, especially from antigen-experienced TCRγδ+ T cells of individuals of older age [13,47,48].

Table 1. Complete TCRγδ+ T-LGL leukemia receptor clonotypes identified in the repertoire of healthy young and elderly individuals.

Sample information				LGL sample	TRG receptor chain		TRD receptor chain	
Donor	Sex	Age	Subset		V-J rearrangement	CDR3 composition	V-J rearrangement	CDR3 composition
Young individuals								
B49	F	26	Naive	LGL 12-098	Vδ3 – Jδ1	CAFSSLTGGYKEYTDKLIF	Vγ9 – Jγ1.3	CALWEVPNYKKLF
Elderly individuals								
B41	M	56	Naive	LGL 10-200	Vδ2 – Jδ1	CACDTVGDRDTDKLIF	Vγ9 – Jγ1.3	CALWEVQYYKKLF
				LGL 12-098	Vδ3 – Jδ1	CAFSSLTGGYKEYTDKLIF	Vγ9 – Jγ1.3	CALWEVPNYKKLF
B51	M	70	Effector	LGL 10-200	Vδ2 – Jδ1	CACDTVGDRDTDKLIF	Vγ9 – Jγ1.3	CALWEVQYYKKLF
B44	F	69	Effector	LGL 12-098	Vδ3 – Jδ1	CAFSSLTGGYKEYTDKLIF	Vγ9 – Jγ1.3	CALWEVPNYKKLF
B60	M	68	Effector	LGL 12-098	Vδ3 – Jδ1	CAFSSLTGGYKEYTDKLIF	Vγ9 – Jγ1.3	CALWEVPNYKKLF

DISCUSSION

Ageing of the immune system has become increasingly important due to increased hygiene and higher life expectancies in the Western World [3,5,49]. Immunosenescence plays an additional role in shaping the immune repertoire. Shaping of the immune system during ontogeny and upon ageing relies on continuous antigenic exposures, varying from pathogens to cellular stress. In the current study we showed that (antigenic) selection starts during early ontogeny in the thymus and cord blood samples and continues in circulating TCRγδ+ T cells in young and elderly individuals. While maintaining diversity in the naive subsets, the effect of ageing is most significant in memory subsets with strong receptor skewing, and effector subsets.

Following technical optimization of multiplex PCR assays for NGS analysis we demonstrated highly diverse TRG, but Vδ1-skewed TCRγδ+ T cell repertoires in precursor TCRγδ+ T cells from thymus and cord blood, with low inter-sample variation. This was in clear contrast to circulating mature TCRγδ+ T cells that showed Vγ9/Vδ2 receptor skewing with high inter-sample variation and donor-specific patterns. As we recently showed significant effects of ageing on maturation profiles of TCRγδ+ T cells [48] we investigated the immune repertoire composition of different TCRγδ+ T cell subsets including naive, central, effector memory and effector cells. Even though the naive TCRγδ+ T cell population shrinks upon ageing [3,5,48], its diversity – being the primary source for mounting immune responses – was maintained in elderly individuals. To date, only one study documented the maintenance of the naive CD4+TCRαβ+ T cell repertoire until the age of 70, after which the repertoire profoundly declined [50] (reviewed by [51]). These results are in keeping with our findings, although our cohort consisted of elderly until the age of 70, since 70 is the maximum age to donate blood at our national blood bank. It would however be interesting to also study healthy individuals >70 years

of age, but high volumes of blood are needed to obtain sufficient numbers of naive TCRγδ+ T cells complicate such studies. Furthermore, low cell numbers form limitations in studying the repertoire due to possibly low levels of input DNA. To overcome such limitations and to directly link overall receptor usage for both TRG and TRD loci e.g. single molecule-based assays could be applied. However, these assays require extensive optimization and validation experiments as well.

Age-related differences were most evident in central and effector memory populations: Vγ9 usage was highly important in young individuals, while a shift towards Vγ2, but also to other Vγ1-family genes was observed in effector memory TCRγδ+ T cells of elderly. The significant increase in Vγ2 usage in elderly was accompanied by a significant decrease in Vδ2 and increase in Vδ1 usage, collectively indicating a shift from Vγ9/Vδ2 specificity in young to Vγ2/Vδ1 in elderly. These findings might suggest differences in antigenic selection, or might be due to underlying clonal expansion in these populations [52]. Furthermore, CMV is known to elicit Vδ1+ TCRγδ+ T cell-specific responses [17]. Additionally, we have recently demonstrated the effect of CMV on the TCRγδ+ T cell immune system, through increasing Vδ1+ TCRγδ+ T cells in elderly carrying CMV [48], and it has been shown that latent CMV carriage is related to the expansion of CMV specific T cells [53]. By zooming in on the antigen-binding part, the CDR3 region, we could identify CMV specific CDR3 regions in a few donors, albeit without a significant effect of age. This could reflect high anti-CMV responses mounted by Vδ1+ TCRγδ+ T cells, although such responses were mainly observed in renal allograft recipients and not in healthy controls [17]. Given that CMV infects mainly fibroblasts and epithelial cells [54], and that the majority of Vδ1+ TCRγδ+ T cells reside in epithelial and mucosal tissues [55-57] these findings could indicate that healthy individuals have a local, rather than circulatory, protection by Vδ1+ TCRγδ+ T cells against CMV.

Investigating the TRG / TRD repertoires of tissue-residing TCRγδ+ T cells is not only relevant for CMV-specific responses, but also for other local antigens contributing to the TCRγδ+ T cell repertoire. Although the ability of TCRγδ+ T cells moving in and out of tissues has not been proven yet, sequencing TCRγδ+ T cells from different tissues could provide insight in both residing, migrating and circulating properties, as well as in development of local immune repertoires and additional functions and specificities of TCRγδ+ T cells.

We also examined other dominant TRG / TRD clonotypes, such as for *M. tuberculosis*, since TCRγδ+ T cells are known to elicit strong responses [16], but these were not identified. This could be related to the recruitment of our donors (mostly of Caucasian descent) via the national Dutch blood bank, and the fact that donors are tested prior to blood donation. Furthermore, open tuberculosis is not endemic in the Netherlands.

Curiously, TRG / TRD clonotypes derived from complete TCRγδ+ T-LGL leukemia receptors were identified in the healthy effector subset repertoire. These findings confirmed earlier correlations identified between TCRγδ+ T-LGL leukemia cells and healthy effector TCRγδ+ T cells [48] and highlight the fact that TCRγδ+ T-LGL is typically a disease that arises in effector cells of elderly. This also correlates with a recent study from Davey et al., in which they have shown that the Vδ1 population in CB is unfocused, while in adult PB clonal expansions could be found, which have direct differentiation from naive into effector phenotypes, together with the down-regulation of CD27. In contrast, Vδ2 cells maintained the invariant TCR expression observed from birth to adulthood. Linking these findings to our results, together with the effect of CMV upon ageing, could both give a suggestion for the increasing Vδ1 usage in elderly, possibly providing a source of clonopathies [58].

The application of our optimized multiplex PCR assays for NGS analysis could also be applicable for other disease states, such as TCRγδ+ T cell lymphomas, or treatments, such as bone marrow transplantation (BMTx). TCRγδ+ T cells have been described to reconstitute in increased numbers after BMTx in acute leukemia patients [59]. It would be interesting to investigate to what extent the TRG / TRD repertoire has changed upon BMTx, and how the TCRγδ+ T cell compartment regenerates. Also, circulating TCRγδ+ T cells have been described in metastatic melanomas, in which it would be interesting to distinguish pro- and anti-tumor specific TCRγδ+ T cells [60], the latter particularly in view of anti-tumor effects [61,62].

In summary, using an optimized NGS assay we identified specific TRG / TRD repertoires during ontogeny and upon ageing. Despite strong individual-specific repertoire compositions, significant differences in Vγ and Vδ gene usage were identified upon ageing in specifically memory TCRγδ+ T cell subsets. These age-dependent effects caused shifts in Vγ9/Vδ2 in young, to Vγ2/Vδ1 dominance in elderly. Additionally, TRG / TRD clonotypes related to TCRγδ+ T-LGL leukemia were identified in normal effector TCRγδ+ T cells of especially elderly individuals, strongly suggesting that TCRγδ+ T-LGL leukemia originate from normal circulating, antigen-experienced effector TCRγδ+ T cells.

ACKNOWLEDGEMENTS

We are grateful to Mr. S.J.W. Bartol and Mrs. H. Charif – Bouallouch for help with cell sorting experiments, and to Mr. A. Eggink for organizing cord blood samples.

AUTHOR CONTRIBUTIONS STATEMENT

MJK, JD and AL designed the experiments. MJK and AL wrote the manuscript. MJK, FK, MYK and IW performed the experiments. MJK and AL analyzed the data and prepared the figures. PV, JD and AL supervised the project. FK and PV revised the manuscript. All authors read the manuscript carefully.

REFERENCES

1. Wherry, E.J. T cell exhaustion. *Nat. Immunol.* **12(6),** 492-499 (2011).
2. Wherry, E.J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15(48),** 486-499 (2015).
3. Boraschi, D. *et al.* The gracefully ageing immune system. *Sci. Transl. Med.* **5(185),** 185ps8 (2013).
4. Aspinall, R. & Andrew, P. Thymic involution in ageing. *J. Clin. Immunol.* **20(4),** 250-256 (2000).
5. Weiskopf, D., Weinberger, B. & Grubeck-Loebenstein, B. The ageing of the immune system. *Transpl. Int.* **22,** 1041-1050 (2009).
6. Pereira, B.I. & Akbar, A.N. Convergence of innate and adaptive immunity during human ageing. *Front. Immunol.* **7,** 445 (2016).
7. Oettinger, M.A., Schatz, D.G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2 adjacent genes that synergistically activate V(D)J recombination. *Science* **248(4962),** 1517 – 1523 (1990).
8. Dik, W.A. *et al.* New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201(11),** 1715-1723 (2005).
9. Fu, G., Rybakin, V., Brzostek, J., Paster, W., Acuto, O. & Gascoigne, N.R.J. Fine-tuning T cell receptor signaling to control T cell development. *Trends Immunol.* **35(7),** 311-318 (2014).
10. Van Dongen, J.J.M., Comans-Bitter, W.M., Wolvers-Tettero & I.L.M., Borst, J. Development of human T lymphocytes and their thymus-dependency. *Thymus.* **16(3-4),** 207-234 (1990).
11. Parker, C.M. *et al.* Evidence for extrathymic changes in the T cell receptor gamma/delta repertoire. *J. Exp. Med* . **171(5),** 1597-1612 (1990).
12. Prinz, I., Silva-Santos, B. & Pennington, D.J. Functional development of γδ T cells. *Eur. J. Immunol.* **43(8),** 1988-1994 (2013).
13. Sandberg, Y. *et al.* TCRγδ+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRγδ+ T cells. *Leukemia.* **20,** 505-513 (2006).
14. Davodeau, F., Peyrat, M.A., Hallet, M.M., Houde, I., Vie, H. & Bonneville, M. Peripheral selection of antigen receptor junctional features in a major human γδ subset. *Eur. J. Immunol.* **23,** 804-808 (1993).
15. Breit, T.M., Wolvers-Tettero, I.L.M. & van Dongen, J.J.M. Unique selection determinant in polyclonal Vδ2 – Jδ1 junctional regions of human peripheral γδ T lymphocytes. *J. Immunol.* **152,** 2860-2864 (1994).

16. Xi, X., Han, X., Li, L. & Zhao Z. Gammadelta T cells response to *Mycobacterium tuberculosis* in pulmonary tuberculosis patients using preponderant complementary determinant region 3 sequence. *Indian J. Med. Res.* **134,** 356-361 (2011).
17. Déchanet, J. *et al.* Implication of γδ T cells in the human immune response to cytomegalovirus. *J. Clin. Invest.* **103,** 1437-1449 (1999).
18. Vermijlen, D. *et al.* Human cytomegalovirus elicits fetal γδ T cell responses in utero. *J. Exp. Med.* **207(4),** 807-821 (2010).
19. Fujishima, N. *et al.* Skewed T cell receptor repertoire of Vδ1+ γδ T lymphocytes after human allogeneic haematopoietic stem cell transplantation and the potential role for Epstein-Barr virus-infected B cells in clonal restriction. *Clin. Exp. Immunol.* **149,** 70-79 (2007).
20. De Jong, A. Activation of human T cells by CD1 and self-lipids. *Immunol. Rev.* **267(1),** 16-29 (2015).
21. Uldrich, A.P. *et al.* CD1d-lipid antigen recognition by the γδ TCR. *Nat. Immunol.* **14(11),** 1137-1145 (2013).
22. Kabelitz, D., Lettau, M. & Janssen, O. Immunosurveillance by human γδ T lymphocytes: the emerging role of butyrophilins. *F100Res.* **6(F1000 Faculty Rev),** 782 (2017).
23. Vasudev, A. *et al.* γ/δ T cell subsets in human ageing use the classical α/β T cell model. *J. Leukoc. Biol.* **8,** 647-655 (2014).
24. Van Dongen, J.J.M. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia*. **17,** 2257-2317 (2003).
25. Sandberg, Y. *et al.* Human T cell lines with well-defined T cell receptor gene rearrangements as controls for the BIOMED-2 multiplex polymerase chain reaction tubes. *Leukemia* **21,** 230-237 (2007).
26. Moorhouse, M.J. *et al.* ImmunoGlobulin galaxy (IGGalaxy) for simple determination and quantitation of immunoglobulin heavy chain rearrangements from NGS. *BMC Immunol.* **15,** 59 (2014).
27. Goecks, J., Nekrutenko, A., Taylor, J. & The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11 (8),** R86 (2010).
28. Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* **19,** unit 19.10.1-21 (2010).
29. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15 (10),** 1451-1455 (2005).
30. Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics.* **26 (14)**: 1783-1785 (2010).
31. Alamyar, E., Giudicelli, V., Shuo, L., Duroux, P. & Lefranc, M.-P. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8(1),** 26 (2012).

32. Alamyar, E., Duroux, P., Lefranc, M.P. & Giudicelli, V. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* **882,** 569-604 (2012).
33. Li, S. *et al.* IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next-generation repertoire immunoprofiling. *Nat. Commun.* **4,** 2333 (2013).
34. Giudicelli, V., Duroux, P., Lavoie, A., Aouinti, S., Lefranc, M.P. & Kossida, S. From IMGT-ONTOLOGY to IMGT/HighVQUEST for NGS Immunoglobulin (IG) and T cell receptor (TR) repertoires in autoimmune and infectious diseases. *Autoimmun. Infec. Dis.* **1(1)** (2015).
35. Boyd, S.D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1(12),** 12ra23 (2009).
36. Krzywinksi, M.I. *et al.* Circos: an information aesthetic for comparative genomic. *Genome Res.* **19,** 1639-1645 (2009).
37. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188-1190 (2004).
38. Schneider, T.D. & Stephens, R.M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097-6100 (1990).
39. Hartigan, P.M. Algorithm AS 217: computation of the dip statistic to test for unimodality. *J. Appl. Stat.* **34,** 320-325 (1995).
40. Hartigan, J.A. & Hartigan, P.M. The dip test of unimodality. *Ann. Stat.* **13,** 70-84 (1985).
41. Mächler, M. Diptest 0.25-1. http://www.r-project.org/ (2003).
42. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org (2016).
43. Liston, A., Carr, E.J. & Linterman, M.A. Shaping variation in the human immune system. *Trends Immunol.* **36(10),** 637-646 (2016).
44. Kallemeijn, M.J. *et al.* Ageing and latent CMV infection impact on maturation, differentiation and exhaustion profiles of T cell receptor gammadelta T cells. *Sci. Rep.* **7,** 5509 (2017).
45. IJspeert, H. *et al.* Antigen Receptor Galaxy: a user-friendly, web-based tool for analysis and visualization of T and B cell receptor repertoire data. *J. Immunol*. **198,** 4156-4165 (2017).
46. Spencer, C.T., Abate, G., Blazevic, A. & Hoft, D.F. Only a subset of phosphoantigen-responsive gamma9delta2 T cells mediate protective tuberculosis immunity. *J. Immunol.* **181,** 4471-4484 (2008).
47. Sandberg, Y. *et al.* Lack of common TCRA and TCRB clonotypes in CD8(+)/TCRαβ(+) T cell large granular lymphocyte leukemia: a review on the role of antigenic selection in the immunopathogenesis of CD8(+) T-LGL. *Blood Cancer J.* **10;4,** e172 (2014).
48. Kallemeijn, M.J. *et al.* Dysregulated signaling, proliferation and apoptosis impact on the pathogenesis of TCRγδ+ T cell large granular lymphocyte leukemia. *PLoS One.* **12(4),** e0175670 (2017).
49. World Health Organization. *World Report on Ageing and Health.* Luxembourg (2015).
50. Naylor, K. *et al.* The influence of age on T cell generation and TCR diversity. *J. Immunol.* **174,** 7446-7452 (2005).

51. Appay, V. & Sauce, D. Naive T cells: the crux of cellular immune ageing? *Exp. Gerontol.* **54,** 90-93 (2014).
52. Yoshida, K. *et al.* Ageing-related changes in human T cell repertoire over 20 years delineated by deep sequencing of peripheral T cell receptors. *Exp. Gerontol.* **96,** 29-37 (2017).
53. Jackson, S.E., Sedikides, G.X., Okecha, G., Poole, E.L., Sinclair, J.H. & Wills, M.R. Latent cytomegalovirus (CMV) infection does not detrimentally alter T cell responses in the healthy old, but increased latent CMV carriage is related to expanded CMV-specific T cells. *Front. Immunol.* **8,** 733 (2017).
54. Merani, S., Pawelec, G., Kuchel, G.A. & McElhaney, J.E. Impact of ageing and Cytomegalovirus on immunological response to Influenza vaccination and infection. *Front. Immunol.* **8,** 784 (2017).
55. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13(2),** 88-100 (2013).
56. Chien, Y.H., Meyer, C. & Bonneville, M. γδ T cells: first line of defense and beyond. *Annu. Rev. Immunol.* **32,** 121-155 (2014).
57. Nielsen, M.M., Witherden, D.A. & Havran, W.L. γδ T cells in homeostasis and host defence of epithelial barrier tissues. *Nat. Rev. Immunol.* **17(12),** 733-745 (2017).
58. Davey, M.S. *et al.* Clonal selection in the human Vδ1 T cell repertoire indicates γδ TCR-dependent adaptive immune surveillance. *Nat. Commun.* **8:** 14760 (2017).
59. Godder, K.T. *et al.* Long term disease-free survival in acute leukemia patients recovering with increased gammadelta T cells after partially mismatched related donor bone marrow transplantation. *Bone Marrow Transplant.* **39(12),** 751-757 (2007).
60. Wistuba-Hamprecht, K. *et al.* Phenotypic characterization and prognostic impact of circulating γδ and αβ T cells in metastatic malignant melanoma. *Int. J. Cancer* **138(3),** 698-704 (2016).
61. Correia, D.V. *et al.* Highly active microbial phosphoantigen induces rapid yet sustained MEK/Erk- and PI-3K/Akt-mediated signal transduction in anti-tumor human gammadelta T cells. *PLoS One* **4(5),** e5657 (2009).
62. Lafont, V. *et al.* Plasticity of γδ T cells: impact on the anti-tumor response. *Front. Immunol.* **5,** 662 (2014).
63. Breit, T.M., Wolvers-Tettero, I.L.M. & van Dongen, J.J.M. Receptor diversity of human TcR-γδ expressing cells. *Prog. Histochem. Cytochem.* **26,** 182-193 (1994).

SUPPLEMENTAL DATA

Supplementary Table 1. Antibody panel used for sorting TCRγδ+ T cell fractions.

	Fluorochrome							
	BV421	**PO**	**FITC**	**PE**	**PE-CF594**	**PE-Cy7**	**APC**	**APC-H7**
Antibody	**CD27**	**CD45**	**CD45RO**	**CD197**	**CD3**	**TCRγδ**	**TCRαβ**	**CD45RA**
Clone	O323	HI30	UCHL1	3D13	UCHT1	11F2	IP26	HI100
Manufacturer	BioLegend	Invitrogen	DAKO	e-Biosciences	BD Biosciences	BD Biosciences	e-Biosciences	BD Biosciences

Supplementary Table 2. Plasmid spike-in sample compositions for primer titration experiments.

TRD plasmid pool		**TRG plasmid pool**	
Rearrangement	**Source***	**Rearrangement**	**Source***
Vδ1 – Jδ1	ARR	Vγ2-Jγ2.1	Karpas299
Vδ1 – Jδ2	Thymus	Vγ2-Jγ1.3	T005
Vδ1 – Jδ3	Thymus	Vγ2-Jγ2.3	T136
Vδ1 – Jδ4	Thymus	Vγ2-Jγ1.3/2.3	T005
Vδ2 – Jδ1	T103	Vγ3-Jγ1.1	T004
Vδ2 – Jδ2	T036	Vγ3-Jγ2.1	T086
Vδ2 – Jδ3	T142	Vγ3-Jγ1.3	Thymus
Vδ2 – Jδ4	Thymus	Vγ4-Jγ2.1	T095
Vδ3 – Jδ1	T106	Vγ4-Jγ1.3	T068
Vδ3 – Jδ2	T006	Vγ5-Jγ1.3	T018
Vδ3 – Jδ3	Thymus	Vγ8-Jγ1.1	Hut78
Vδ3 – Jδ4	Thymus	Vγ8-Jγ2.1	Molt16
		Vγ8-Jγ1.3	T109
		Vγ8-Jγ2.3	T001
		Vγ9-Jγ1.1	ARR
		Vγ9-Jγ2.1	T013
		Vγ9-Jγ1.2	T167
		Vγ9-Jγ1.3	T018
		Vγ9-Jγ2.3	T106

* Rearrangements were obtained from cell lines (ARR, Molt16, Hut78) [13], thymus material or primary T-ALL samples (T).

Supplementary Table 3. Optimized protocol for multiplex PCR assays for TRG/TRD NGS experiments.

TRG		TRD	
Primer	**Concentration (pmol/µl)**	**Primer**	**Concentration (pmol/µl)**
Vγ1F	6	Vδ1	2.5
Vγ9	4	Vδ2	4.5
Jγ1.1/2.1	4	Vδ3	3
Jγ1.3/2.3	2	Jδ1	2
Jγ1.2	4	Jδ2	2
		Jδ3	2.5
		Jδ4	2.5
Tm	58.0	Tm	59.0
Cycle number	25	Cycle number	20/25*

*After TRD biological validation with genomic DNA, the cycle number was adjusted from 20 to 25 cycles to obtain sufficient DNA amplicon for sequencing.

Supplementary Table 4. Donor characteristics and cell numbers of sorted cells.

Donor characteristics			Absolute cell numbers of different TCRγδ+ T-cell subsets*				Frequencies of different TCRγδ+ T-cell subsets**			
Donor number	**Sex**	**Age**	**Naive**	**Central memory**	**Effector memory**	**Effector**	**Naive**	**Central memory**	**Effector memory**	**Effector**
Young (age 20-35)										
B49	F	26	28263	72280	31931	5380	0.30	0.10	26.30	73.30
B50	M	34	17483	150390	14431	1393414	0.05	0.20	9.70	90.05
B53	F	31	28569	80054	924065	790997	0.90	0.30	8.60	90.20
B54	M	30	22959	77880	345713	127702	11.30	3.60	7.20	77.90
B55	M	29	26138	6097	9986	22064	9.90	0.70	9.20	80.20
B58	F	21	22633	17276	112810	16535	27.70	8.90	36.80	26.60
B59	M	28	107919	101833	550244	689371	27.00	16.60	45.60	10.80
B63	F	29	35225	2369	507075	438441	0.90	0.50	19.00	79.60
B67	M	28	9622	1765	100966	68890	1.00	1.80	37.00	60.20
B73	F	20	52185	8722	248533	163600	1.30	4.30	44.40	50.00
B74	M	23	43721	301724	136523	435786	1.45	5.35	44.10	49.10
Elderly (age 56-70)										
B41	M	56	9740	5425	57313	172030	2.20	0.60	27.70	69.50
B42	M	61	2670	3490	46100	119277	1.40	0.30	18.40	79.90
B43	F	56	29390	18095	36855	452425	0.20	0.10	69.50	30.20
B44	F	69	2579	6420	28052	315285	0.20	0.90	30.00	68.90
B45	M	58	7023	12282	164705	137987	0.80	3.40	42.90	52.90
B51	M	70	1950	6443	121493	275430	0.30	0.05	11.20	88.45
B52	F	60	1996	6560	56438	40149	3.70	1.50	33.60	61.20
B60	M	68	95780	331290	18342	420358	1.60	2.10	17.10	79.20
B65	M	62	190507	65911	32914	81127	12.10	6.40	33.60	47.90
B68	M	60	3665	8450	31728	261805	0.60	0.10	4.30	95.00
B69	M	56	18059	4259	179862	2003909	4.10	0.01	39.00	56.89
B70	M	67	6400	1080	33077	23905	1.60	1.60	49.60	47.20

* Subset definitions: naive, CD45RA+CD45RO-CD27+CD197+; central memory, CD45RA-CD45RO+CD27+CD197+; effector memory, CD45RA-CD45RO+CD27-CD197- (TemRO); effector, CD45RA+CD45RO-CD27-CD197- (TemRO).
** Frequencies of naive, central memory, effector memory and effector cells within total TCRγδ+ T cell populations.

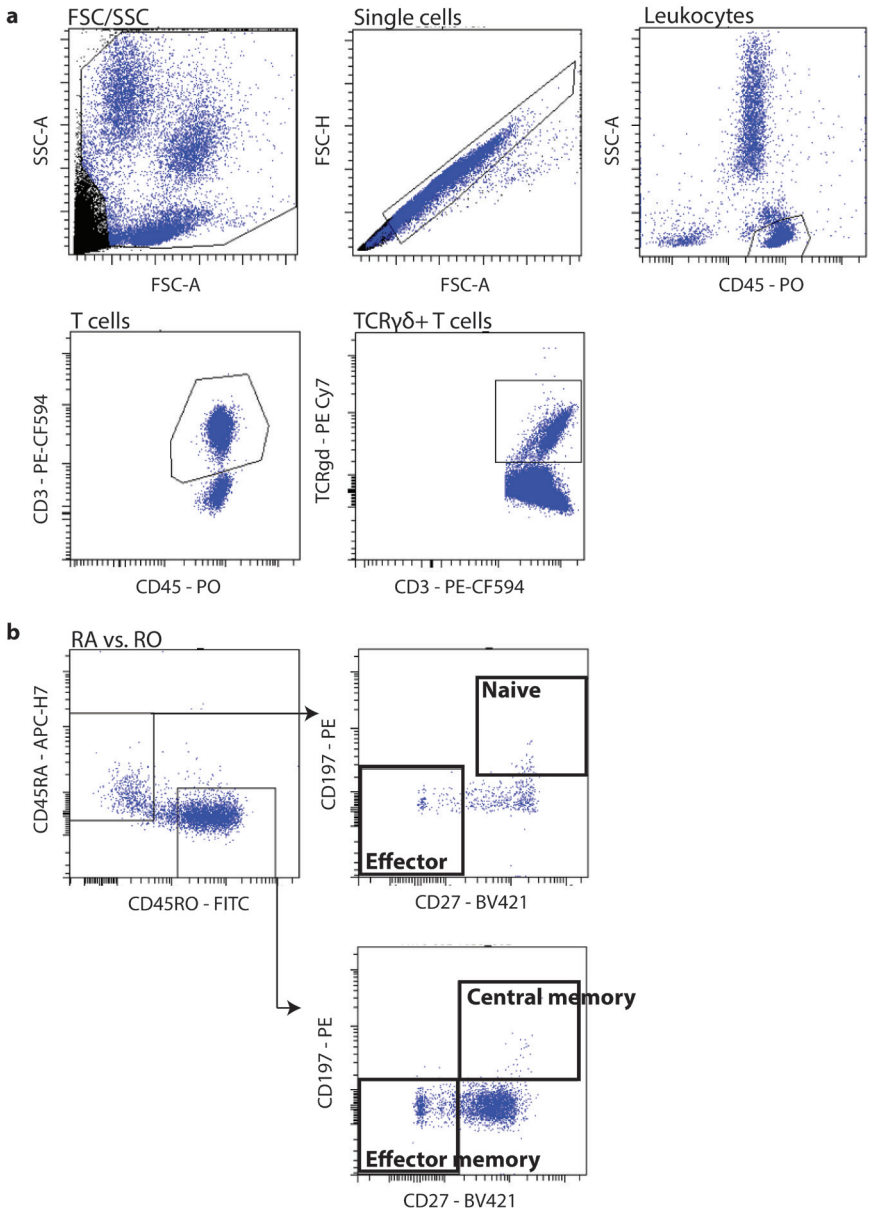
Supplementary Table 5. Clonality scores and number of coincidences.

Sample	Naive*		Effector*		Effector memory*		Sample	Naive*		Effector*		Effector memory*	
	Clonality score	Coincidence #3***	Clonality score	Coincidence #3***	Clonality score	Coincidence #3***		Clonality score	Coincidence #3***	Clonality score	Coincidence #3***	Clonality score	Coincidence #3***
TRG – Young individuals							*TRD – Young individuals*						
B49	2.00e-04	15	4.22e-05	16	2.49e-05	172	B49	1.25e-06	24	8.98e-05	3	1.58e-04	156
B50	9.37e-05	37	2.22e-04	433	8.27e-04	107	B50	1.30e-04	594	.45e-04	66	2.14e-04	260
B53	2.91e-05	2	3.07e-04	261	1.06e-03	86	B53	9.63e-05	1	3.32e-04	361	1.59e-04	356
B54	6.33e-05	1888	1.81e-04	407	2.89e-04	295	B54	3.26e-05	36	2.09e-04	454	1.04e-04	339
B55	4.71e-05	34	9.64e-05	79	6.03e-04	35	B55	3.33e-05	1	1.08e-04	646	1.63e-04	100
B58	6.33e-05	464	6.40e-05	13	1.70e-04	281	B58	3.46e-04	41	2.22e-05	2	7.74e-05	491
B59	1.46e-05	10	1.01e-04	35	2.04e-04	595	B59	3.32e-05	1	6.42e-05	52	1.10e-04	919
B63	2.47e-04	777	1.81e-04	31	2.65e-04	253	B63	2.13e-04	729	1.53e-04	573	1.15e-04	390
B67	2.23e-04	10	2.51e-04	667	2.93e-04	259	B67	9.57e-05	3353	2.84e-04	617	1.38e-04	342
B73	1.23e-04	152	3.00e-04	83	4.46e-04	221	B73	5.50e-04	15	6.55e-04	81	6.80e-04	165
B74	5.13e-01	402	2.30e-04	227	2.42e-04	295	B74	3.72e-05	20	2.67e-04	342	1.19e-04	441
TRG – Elderly individuals							*TRD – Elderly individuals*						
B41	2.34e-04	24	3.41e-04	378	2.32e-03	20	B41	1.08e-05	21	5.95e-05	6	4.60e-03	6
B42	1.54e-04	45	3.87e-04	208	2.25e-04	514	B42	3.47e-04	43	1.71e-04	304	4.19e-04	173
B43	2.27e-04	61	3.44e-04	160	5.57e-03	13	B43	1.79e-04	181	1.11e-03	169	3.86e-03	5
B44	1.61e-04	11	1.85e-04	112	9.78e-04	232	B44	7.09e-05	1378	1.40e-03	119	1.84e-04	166
B45	1.36e-04	38	2.39e-04	368	1.26e-04	629	B45	2.49e-04	116	9.21e-04	126	1.15e-04	456
B51	2.76e-04	270	1.50e-03	190	2.11e-04	532	B51	3.12e-04	146	9.33e-04	274	2.68e-04	318
B52	2.23e-04	459	3.98e-04	1	2.50e-04	356	B52	1.80e-04	345	3.94e-04	416	3.48e-04	201
B60	7.14e-04	110	8.33e-04	57	5.70e-04	198	B60	1.39e-04	14	1.14e-03	143	8.19e-04	44
B65	3.91e-04	461	1.52e-04	29	3.11e-04	108	B65	2.01e-04	1122	4.03e-04	355	2.81e-04	66
B68	1.57e-04	68	5.81e-04	189	3.59e-04	243	B68	1.41e-04	12	6.43e-04	182	4.90e-04	73
B69	1.76e-04	701	1.60e-04	823	3.49e-04	301	B69	6.21e-04	172	5.62e-04	154	2.84e-04	73
B70	2.66e-04	610	2.90e-04	719	6.72e-04	62	B70	1.59e-04	59	1.71e-04	528	6.42e-04	247

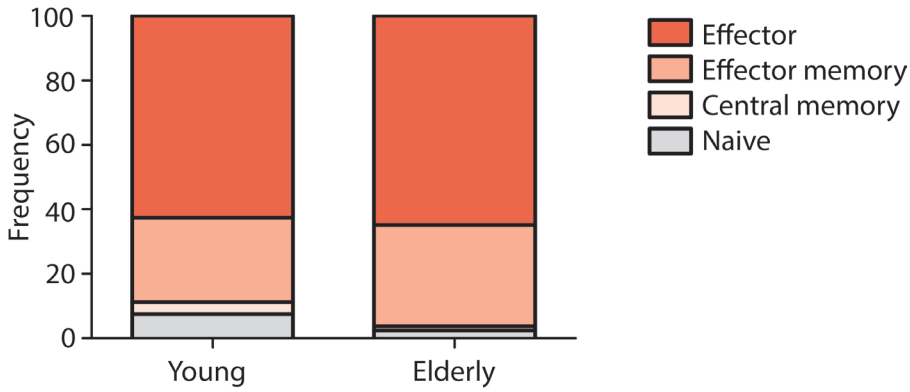
*From subsets naive, effector and effector memory 3 replicates performed; subset central memory was sequenced once due to low quantities of DNA material.

**Absolute number of sequences identified in all three replicates.

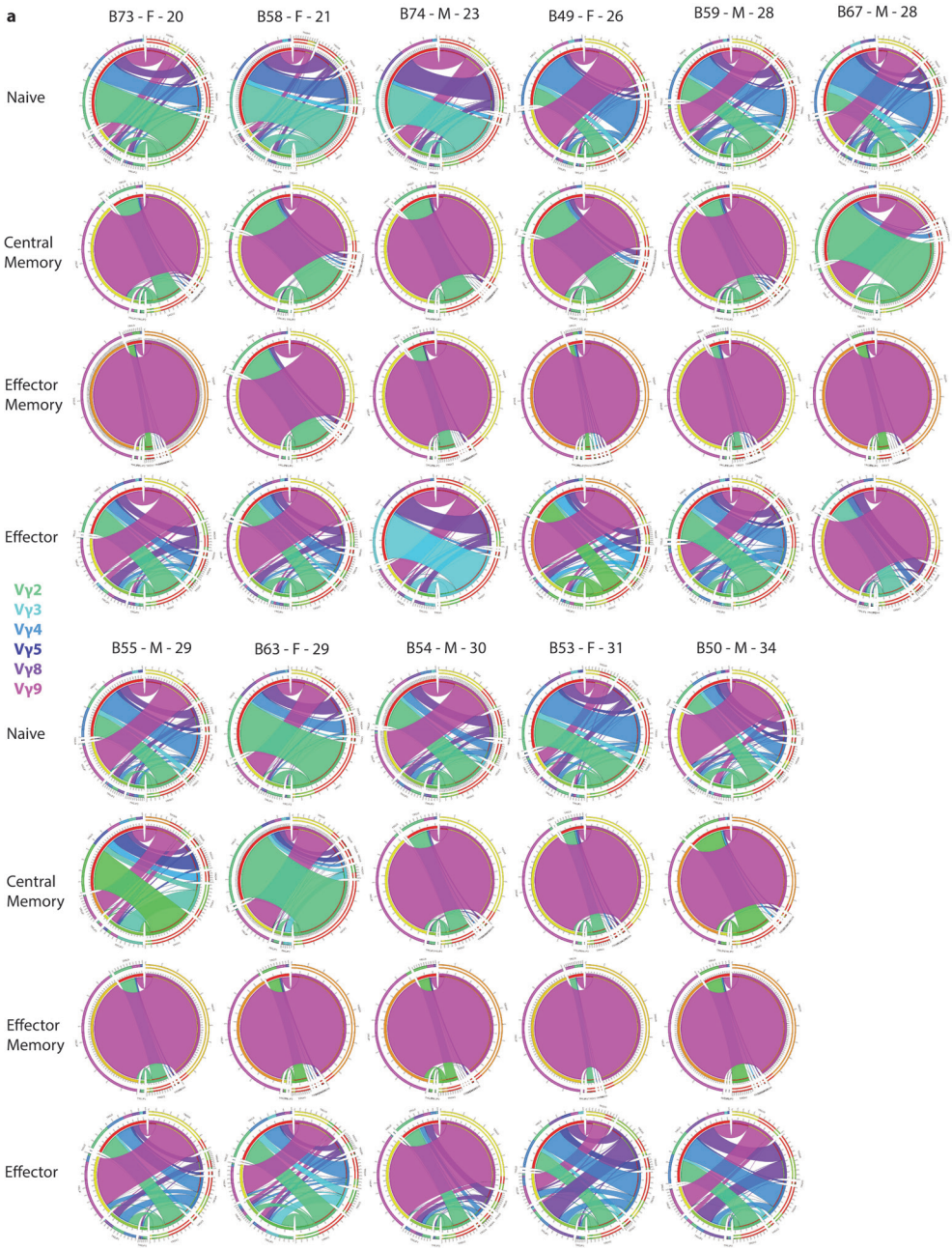
Calculations based on the algorithm of Boyd et al. [34].



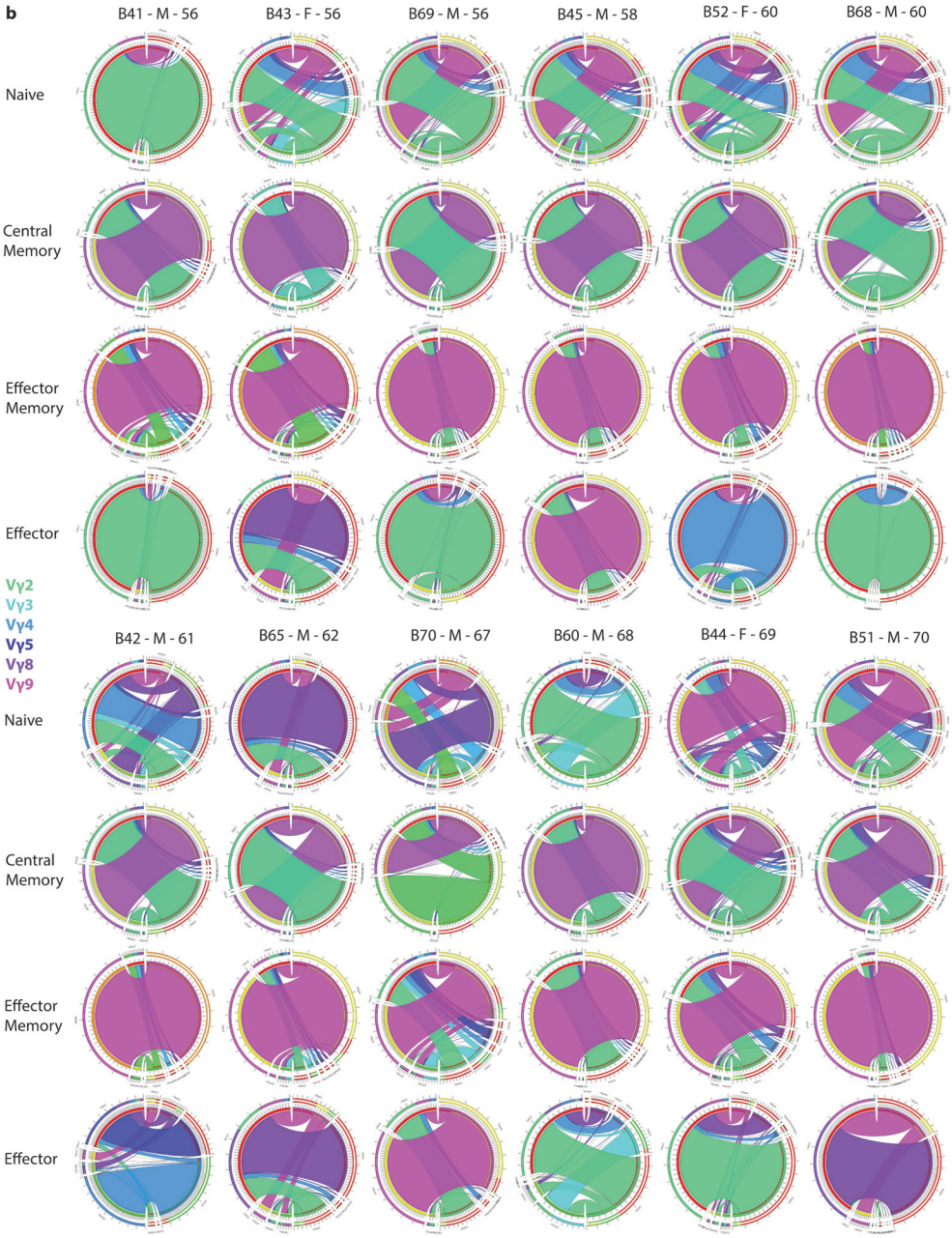
Supplementary Figure 1. Gating strategies for sorting experiments of TCRγδ+ T cell subsets.
General gating strategies for determining viable cells using FSC/SSC, single cells using FSC-H/FSC-A, leukocytes using CD45, T cells (CD3+/CD45+) and TCRγδ+ T cells (TCRγδ+TCRαβ-) (**a**). Further determination of CD45 splice variants CD45RA vs. CD45RO to sort within CD45RA+CD45RO- for naive (CD27+CD197+) and effector (CD27-CD197-) TCRγδ+ T cells, and within CD45RA-CD45RO+ for central memory (CD27+CD197+) and effector memory (CD27-CD197-) TCRγδ+ T cells (**b**). Antibodies used for these experiments are summarized in Supplementary Table 1.

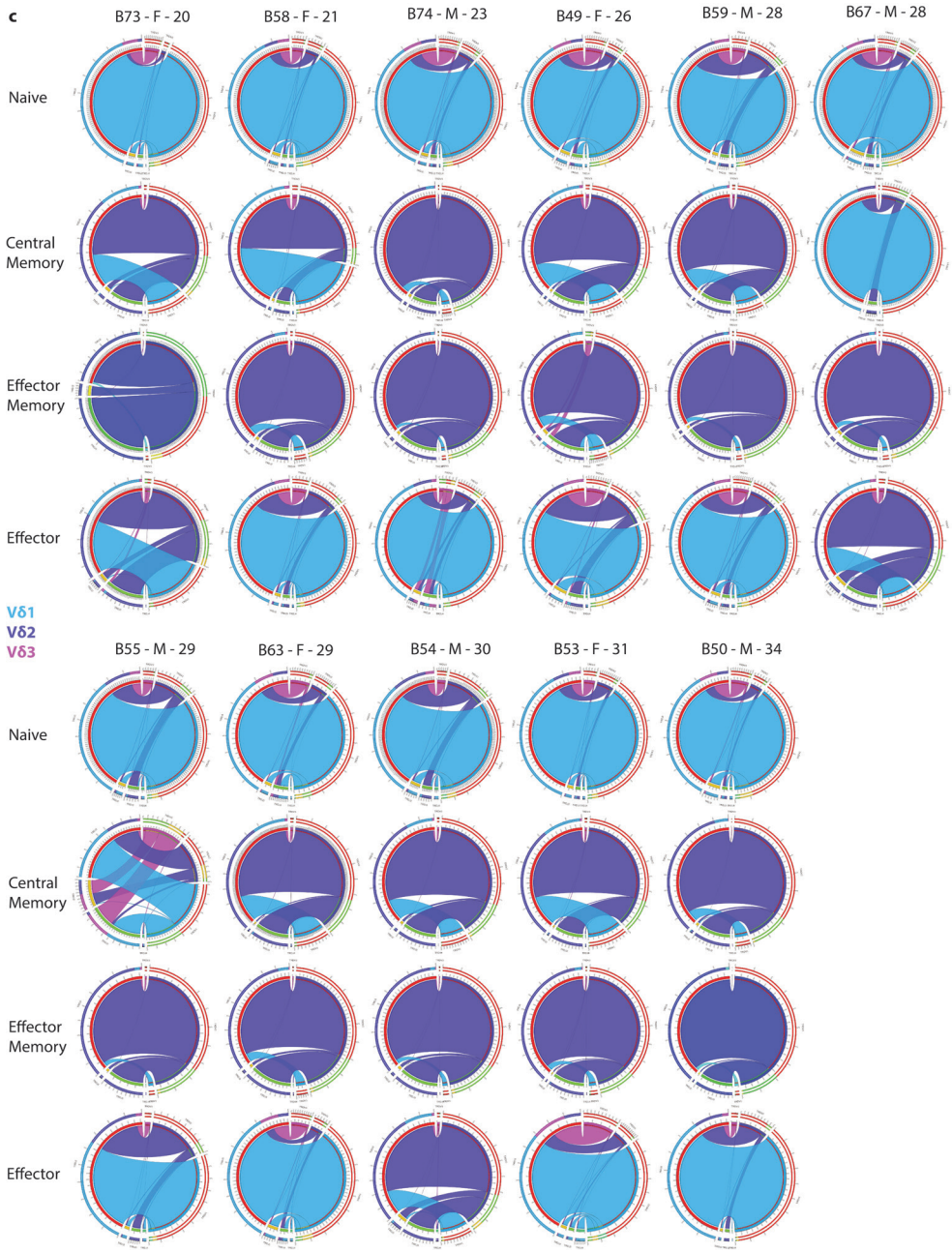


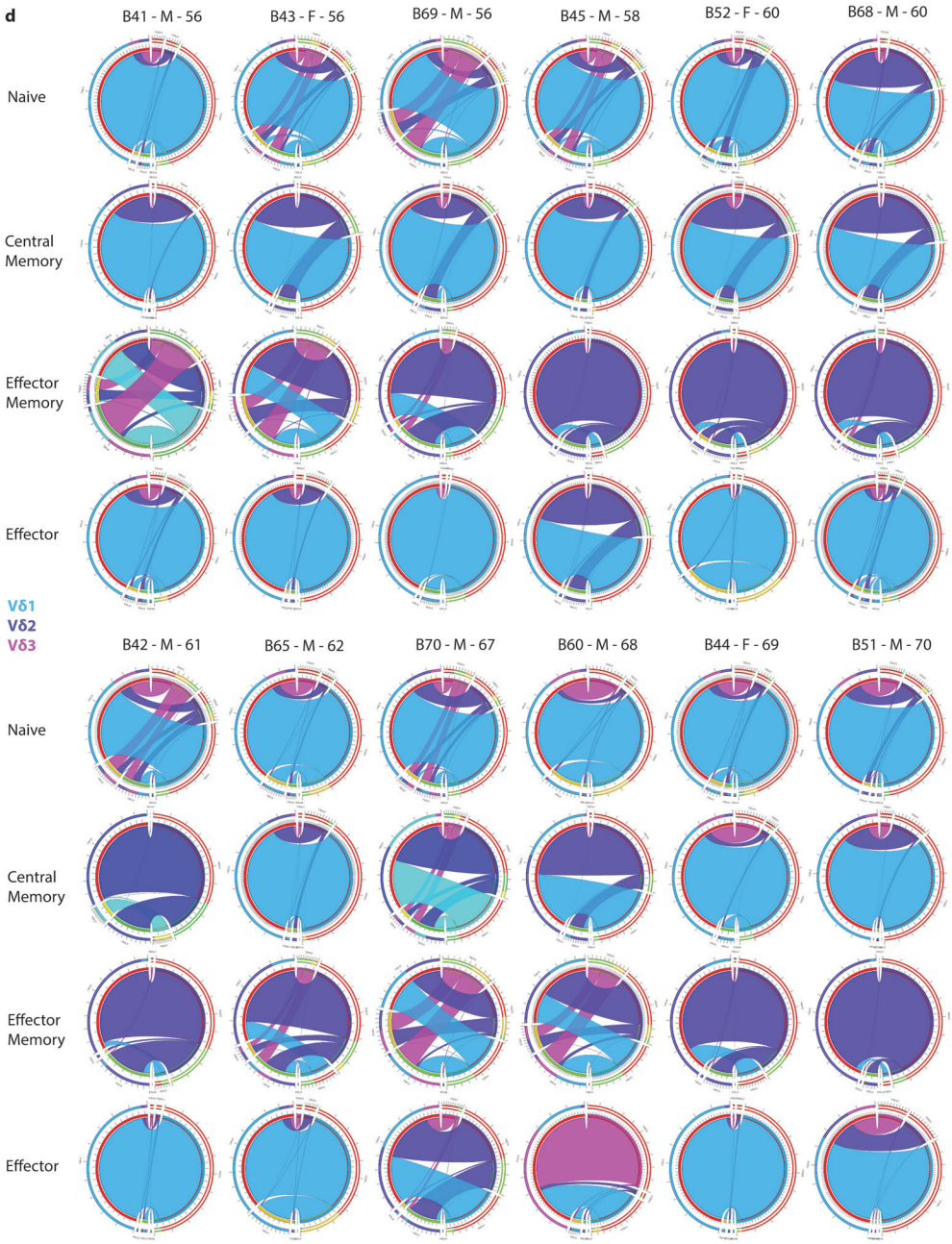
Supplementary Figure 2. Relative TCRγδ+ T cell subset distributions in young and elderly individuals. Mean frequencies of naive (CD45RA+CD45RO-CD27+CD197+), central memory (CD45RA-CD45RO+CD27+CD197+), effector memory (CD45RA-CD45RO+CD27-CD197-) and effector (CD45RA+CD45RO-CD27-CD197-) TCRγδ+ T cells as analyzed during sorting experiments. Absolute cell numbers and frequencies are summarized per donor in Supplementary Table 6.

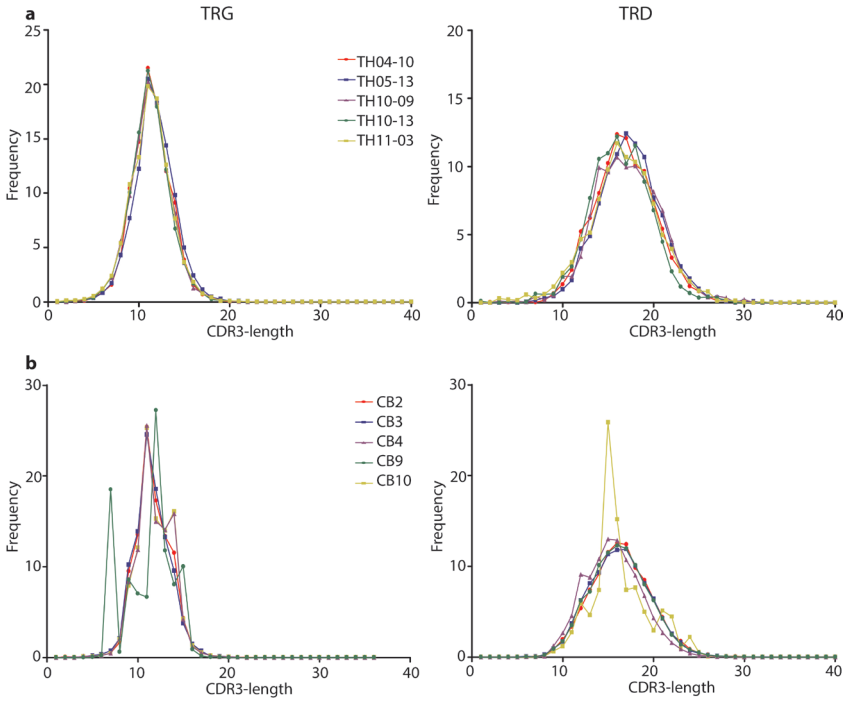


Supplementary Figure 3 (see also the next three pages). Circoletto visualizations of naive, central memory, effector memory and effector subsets of young and elderly individuals.
V-J distributions of TRG young (**a**), TRG elderly (**b**), and TRD young (**c**) and TRD elderly (**d**) rearrangements per donor and subset of young and elderly individuals. Figures were made using the Circoletto online software tool (www.circos.ca, [35]). Each band represents a V-J rearrangement, with different colors reflecting different V gene usage.

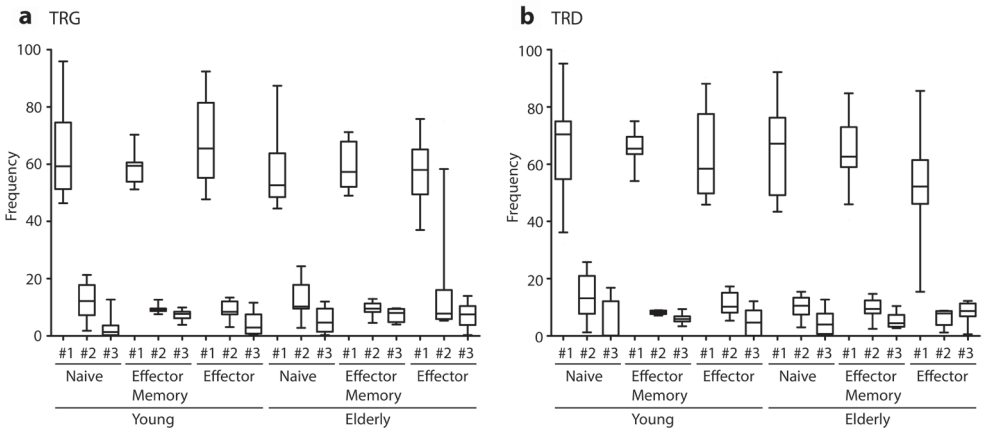

b
B41 - M - 56
B43 - F - 56
B69 - M - 56
B45 - M - 58
B52 - F - 60
B68 - M - 60
Naive
Central Memory
Effector Memory
Effector
Vγ2
Vγ3
Vγ4
Vγ5
Vγ8
Vγ9
B42 - M - 61
B65 - M - 62
B70 - M - 67
B60 - M - 68
B44 - F - 69
B51 - M - 70
Naive
Central Memory
Effector Memory
Effector


c
B73 - F - 20
B58 - F - 21
B74 - M - 23
B49 - F - 26
B59 - M - 28
B67 - M - 28
Naive
Central Memory
Effector Memory
Effector
Vδ1
Vδ2
Vδ3
B55 - M - 29
B63 - F - 29
B54 - M - 30
B53 - F - 31
B50 - M - 34
Naive
Central Memory
Effector Memory
Effector


d
B41 - M - 56
B43 - F - 56
B69 - M - 56
B45 - M - 58
B52 - F - 60
B68 - M - 60
Naive
Central Memory
Effector Memory
Effector
Vδ1
Vδ2
Vδ3
B42 - M - 61
B65 - M - 62
B70 - M - 67
B60 - M - 68
B44 - F - 69
B51 - M - 70
Naive
Central Memory
Effector Memory
Effector



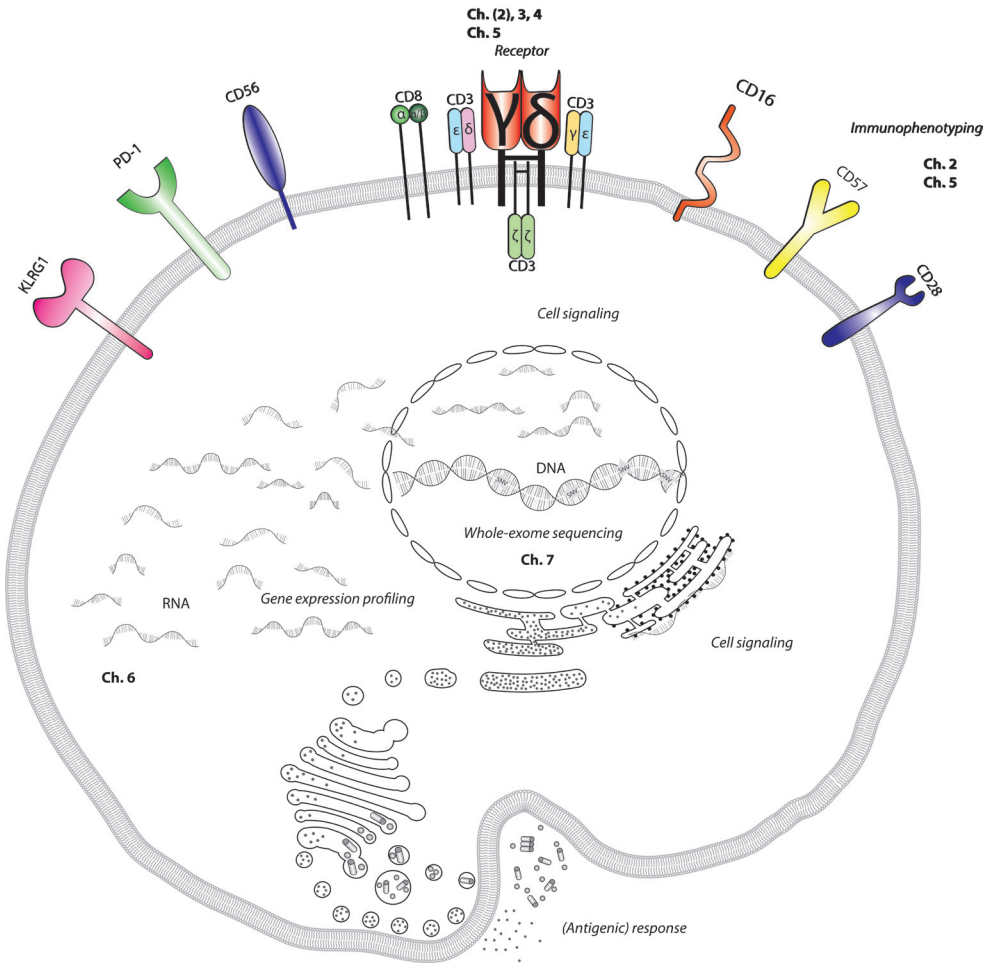
Supplementary Figure 4. Frequencies of coincidence sequences in 1, 2 or 3 PCR replicates per subset and age group.
Box-whiskers plots (range 10-90%) depicting numbers of sequences present in either one (#1), two (#2) or three (#3) PCR replicates within all unique sequences based on V-J combination and CDR3 region on nucleotide level. Error bars indicate the SD.



Supplementary Figure 5. TRG and TRD CDR3-length distributions in thymus (Thy) and cord blood (CB) samples.
Mean frequencies of four replicates per Thy and CB sample are depicted for TRG and TRD CDR3-regions (**a**). CB samples showed more prominent peaks in TRG CDR3-lengths, while TRD CDR3-regions showed relatively normal distributions (**b**). Number of replicates per sample, N=3/4.

Part II

Aberrant TCRγδ+ T cells


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 7
RNA
Gene expression profiling
Cell signaling
Ch. 6
(Antigenic) response

Chapter 5

Lack of common TRA and TRB clonotypes in CD8+TCRαβ+ T cell large granular lymphocyte leukemia: a review on the role of antigenic selection in the immunopathogenesis of CD8+TCRαβ+ T-LGL leukemia

Yorick Sandberg[1], Martine J. Kallemeijn[1], Wim A. Dik[1], Dennis Tielemans[1], Ingrid L.M. Wolvers-Tettero[1], Ellen J. van Gastel-Mol[1], Tomek Szczepanski[1,2], Youri Pol[1], Nikos Darzentas[3], Jacques J.M. van Dongen[1] & Anton W. Langerak[1]

[1]: Department of Immunology, Laboratory for Medical Immunology, Erasmus MC, Rotterdam 3000 CA, the Netherlands.
[2]: Department of Pediatric Hematology and Oncology, Medical University of Silesia, Zabrze, Poland.
[3]: Institute of Agrobiotechnology, Center for Research and Technology Hellas, Thessaloniki, Greece.

Blood Cancer J. **4**, e172 (2014)

ABSTRACT

Clonal CD8+ T cell receptor (TCR)αβ+ T cell large granular lymphocyte (T-LGL) proliferations constitute the most common subtype of T-LGL leukemia. Although the etiology of T-LGL leukemia is largely unknown, it has been hypothesized that chronic antigenic stimulation contributes to the pathogenesis of this disorder. In the present study, we explored the association between expanded TCR-Vβ and TCR-Vα clonotypes in a cohort of 26 CD8+TCRαβ+ T-LGL leukemia patients, in conjunction with the HLA-ABC genotype, to find indications for common antigenic stimuli. In addition, we applied purpose-built sophisticated computational tools for an in-depth evaluation of clustering of TCRβ (TCRB) complementarity determining region 3 (CDR3) amino-acid LGL clonotypes. We observed a lack of clear TCRA and TCRB CDR3 homology in CD8+TCRαβ+ T-LGL, with only low level similarity between small numbers of cases. This is in strong contrast to the homology that is seen in CD4+TCRαβ+ T-LGL and TCRγδ+ T-LGL and thus underlines the idea that the LGL types have different etiopathogenesis. The heterogeneity of clonal CD8+TCRαβ+ T-LGL proliferations might in fact suggest that multiple pathogens or autoantigens are involved.

Keyword(s): T-LGL, TCRB, TCRA, CDR3, HLA, antigenic stimulation.

INTRODUCTION

Large granular lymphocyte (LGL) proliferations are derived from normal cytotoxic LGL cells, which comprise 10–15% of peripheral blood (PB) mononuclear cells (MNCs) [1–3]. The majority of the normal LGL cells (85%) are of NK cell origin, and a minority is derived from mature (post-thymic) T lymphocytes. Lymphoproliferations of LGLs range from activated polyclonal expansions to clinically overt leukemias. T cell LGL (T-LGL) leukemia is the most common subtype, representing ~85% of all LGL leukemia cases diagnosed in western countries.

T-LGL leukemia is a rare and heterogeneous disorder and about one-third of patients is asymptomatic at diagnosis. The main clinical manifestations are related to chronic neutropenia and/or anemia [3–7]. There is a frequent association with a wide variety of autoimmune diseases (33%) and other malignancies (13%) [8]. Its diagnosis is based on a persistent (>6 months) morphologically and/or immunophenotypically increased clonal CD3+CD57+ LGL population in PB, usually >2 x 10^9/l, though a lower count (range 0.4–2 x 10^9/l) may also be compatible with a diagnosis of T-LGL leukemia [9,10]. T-LGL leukemias can be divided into three groups on the basis of their immunophenotypical and molecular characteristics: CD8+, CD4+ and T cell receptor (TCR)γδ+ T-LGL. Monoclonal CD8+TCRαβ+ T-LGL leukemia forms the largest subgroup (80–90%) of monoclonal T-LGL lymphoproliferative disorders [11]. It presents in elderly individuals (mean age 60 years) and generally has an indolent clinical course [3,12]. However, cases with a more aggressive clinical course that are associated with a CD3+CD8+CD56+CD57- phenotype have been reported as well [13,14]. CD3+CD4+TCRαβ+ T-LGL leukemia and CD3+TCRγδ+ T-LGL leukemia are far less common (<5 and 5–10%, respectively), but a considerable number of cases from both disease entities have recently been described in detail [15,16].

Clonality assessment via PCR-based studies of TCR genes is essential to discriminate true T-LGL leukemia from other reactive proliferations. It should be stressed that the finding of clonality does not necessarily imply malignancy in this disease, since most cases are indolent and do not require therapy. Therefore, patients are often diagnosed as having T cell clonopathy of undetermined significance [1,4,17].

Although the etiology of T-LGL leukemia is still largely unknown, it has been hypothesized that chronic antigenic stimulation contributes to the pathogenesis of this disorder. This is in line with the activation-associated effector phenotype and the skewed TCR expression pattern found in T-LGL leukemia. An exaggerated response to immunodominant autoantigens or viral/bacterial antigens might be the initial step in the development of this disorder [18–21]. Recently, the T cell repertoire has been demonstrated to be dynamic in a large proportion (37%) of T-LGL leukemia patients, a phenomenon referred

to as "clonal drift" [22]. This supports the hypothesis that extreme clonal evolution is the result of a polarized reactive process. On top of that, secondary molecular events are assumed to be required to establish the full leukemic phenotype of the chronically antigen stimulated T-LGL population. Those events especially lead to dysregulated apoptosis and constitutively activation of multiple cell survival pathways [11,12,23–26].

To further substantiate the potential involvement of a common antigen in driving development of clonal T-LGL proliferations, the complementarity determining region 3 (CDR3) sequences of the rearranged TCR genes are being analyzed. The CDR3 region of the TCR molecule has the highest antigenic specificity and directly binds to the antigenic peptide presented in the context of HLA [27]. Garrido *et al.* [28] demonstrated strikingly similar motifs in CDR3 TCR-Vβ13 sequences in 42% of CD4+TCRαβ+ T-LGL leukemia cases and a clear association with the HLA-DR*0701 genotype. Interestingly, highly similar CDR3 sequences could also be detected in TCRγ (TCRG) and TCRδ (TCRD) genes in nearly half of patients diagnosed with TCRγδ+ T-LGL leukemia, supporting a common antigen-driven origin of this disorder [16].

In CD8+TCRαβ+ T-LGL leukemia non-random clonal selection has been suggested [18,29], even though no consistent single structural homologous motif could be detected in CDR3 sequences of TCRβ (TCRB) genes, The seeming lack of such identical TCR specificities could however also be explained by the diverse HLA background of these patients. Furthermore, the TCRα chain might have an important role next to the TCRβ chain, especially in the initial phase of high-affinity clonal TCR selection [30]. However, the CDR3 regions of the TCRα (TCRA) genes have not been extensively studied in CD8+TCRαβ+ T-LGL leukemia.

In the present study, we therefore explored the existence of a potential association between CDR3 sequences of both the TCRA and TCRB clonotypes in a cohort of 26 patients diagnosed with CD8+TCRαβ+ T-LGL leukemia in conjunction with the HLA genotype. In addition, we applied purpose-built sophisticated computational tools, specifically developed for sequence pattern discovery in CDR3 amino-acid sequences to evaluate clustering of a large number of TCRB CDR3 amino-acid clonotypes of CD8+TCRαβ+ T-LGL leukemia patients.

MATERIALS & METHODS

Patients and cell samples

PB and/or bone marrow (BM) samples from 26 patients with CD8+TCRαβ+ T-LGL leukemia were obtained. The diagnosis of T-LGL leukemia was established by clinical and laboratory parameters as defined previously [7,9,31]. Patients with a persistent (>6

months) and increased (>1 x 10^9/l) monoclonal CD3+CD8+TCRαβ+ T-LGL proliferation in PB were included. All patient samples were obtained according to the Helsinki declaration following guidelines of the Medical Ethics Committee of Erasmus MC, University Medical Center (Rotterdam, The Netherlands). PB/BM MNCs were isolated by Ficoll-Paque (density: 1.077 g/ml; Pharmacia, Uppsala, Sweden) centrifugation and were used for DNA isolation and RNA isolation. Immunophenotyping was performed on whole PB or BM samples and occasionally on MNC fractions. Cytomorphological May-Grünwald-Giemsa staining of PB smears was used for morphological evaluation of LGLs. HLA genotyping for HLA-ABC was performed by Luminex-based SSOP-PCR techniques (One Lambda Inc., Canoga Park, CA, USA).

Immunophenotypical analysis

Cells were analyzed for membrane expression using a routine panel of monoclonal antibodies, including CD2, CD3, CD4, CD5, CD7, CD8, CD16, CD56, CD57, anti-TCRαβ (BMA031 and WT31), anti-TCRγδ (11F2) and anti- HLA-DR. Immunofluorescence stainings were performed as described [32] and evaluated on a FACSCalibur or FACSCanto II (BD Biosciences, San Jose, CA, USA) flow cytometer. Data analysis was performed using CellQuest and Paint-A-Gate Pro software (BD Biosciences). The PB and/or BM samples were studied in more detail for Vβ domain expression to quantify the contribution of each Vβ family to the CD8+ lymphocyte population. To this end, flow-cytometric analysis was performed using the IO Test Beta Mark kit (Beckman Coulter, Brea, CA, USA) as described [32]. Samples in which the Vβ restriction of the expansion could not be identified by flow cytometry were analyzed by TCRB RT–PCR as described [33].

DNA and RNA isolation and cDNA synthesis

High-molecular weight DNA from fresh or frozen PB MNCs was extracted using a phenol-chloroform extraction-based protocol, followed by ethanol precipitation and resolution in Tris-EDTA buffer [34]. In a subset of cases, DNA was isolated using the GenElute Mammalian Genomic DNA miniprep kit (Sigma-Aldrich, St Louis, MO, USA) according to the manufacturer's protocol. Total RNA was extracted from fresh or frozen PB and/or BM MNCs from patients and reverse transcribed into cDNA as previously described [35]. cDNA quality was checked using ABL as a control gene.

TCRA and TCRB gene rearrangement analysis

For TCRA gene rearrangement analysis, cDNA was amplified using newly developed TCRA primers: one constant region reverse primer (Cα) and 54 different Vα family-specific forward primers distributed over 5 different multiplex tubes; each of these multiplex contained 10 (TCRA tube B) or 11 (TCRA tubes A, C, D and E) Vα primers

in combination with the Cα primer (Supplementary Table 1). In each 50 ml PCR, 2 ml of cDNA, 10 pmol of 50 and 30 oligonucleotide primers, 3 mmol/l MgCl2, 0.2 mmol/l dNTP, 5 ml 10x buffer II, and 1–2 U AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA, USA) were used. TCRB gene rearrangement analysis was performed according to the BIOMED-2 multiplex PCR protocol [36]. BIOMED-2 multiplex PCR kits were obtained from InVivoScribe Technologies (San Diego, CA, USA; www.invivoscribe.com). Amplification reactions were performed in an automated thermocycler (model ABI 2700; Applied Biosystems).

Sequence analysis

After PCR amplification of TCRA and TCRB gene rearrangements, products were subjected to heteroduplex analysis [37]. Products found to be monoclonal in heteroduplex analysis were directly sequenced except for cases with more than one clonal product. In such cases, homoduplexes were excised from the polyacrylamide gel and DNA was eluted before sequencing. Sequencing was performed on the ABI 3100 or 3130xl Genetic Analyzers (Applied Biosystems), using the dye terminator cycle sequencing kit and AmpliTaqFS DNA polymerase (Applied Biosystems). Assignment of Vβ, Dβ, Jβ, Vα and Jα genes and reading frames of the involved TCRB and TCRA gene rearrangements was done using the IMGT database (www.imgt.org) [38].

In silico analysis and data visualization

V-J combinatorial diversity was visualized using Circoletto, an online visualization tool based on Circos (bat.ina.certh.gr/tools/circoletto/) [39]. The collected TCRB CDR3 amino-acid diversity was further analyzed using the TEIRESIAS algorithm, a computational tool developed by the Bioinformatics and Pattern Discovery group at the IBM Computational Biology Center, as described previously [40]. This algorithm uses a motif-based clustering approach with predefined thresholds for amino-acid identity and similarity, CDR3 length differences and offsets for sequence motifs within CDR3 sequences. TCRB CDR3 amino-acid patterns of different subsets were visualized using WebLogo (weblogo.berkeley.edu/). Each logo consists of multiple stacks of symbols, one stack for each position of the sequence. CDR3 is shown based on IMGT position definitions.

RESULTS

Clinical and hematological features are heterogeneous in CD8+TCRαβ+ T-LGL leukemia

The most relevant clinical and hematological findings at diagnosis of the 26 CD8+TCRαβ+ T-LGL leukemia patients enrolled in this study are summarized in Table 1. The median age was 58 years (range 31–86 years) and there was no male or female predominance. Out of the 26 patients, 21 (81%) were symptomatic at presentation. Nine of the 26 patients (35%) had an episode of bacterial infection or B symptoms (fever, night sweats and weight loss). Most frequent presentations concerned neutropenia and/or anemia (62%), whereas some T-LGL leukemias presented with neutropenia plus thrombocytopenia (12%). Thrombocytopenia with coexistent anemia was found in just one case (4%), splenomegaly in two (8%) and lymphadenopathy also in only one (4%). Examination of PB smears showed an increased number of LGLs with abundant cytoplasm containing azurophilic granules in virtually all analyzed cases.

An associated disease was found in 11 cases (42%) (Table 1). In our cohort, 7 patients (27%) had a co-existent autoimmune disorder. The most common autoimmune manifestation was rheumatoid arthritis, which was diagnosed in 3 patients (12%). Co-existence of a malignancy was found in five cases (18%), three of which showed a second hematological malignancy (12%). To compare the relative frequencies of associated cytopenias, autoimmune disorders and malignancies in all three types of T-LGL leukemia, we analyzed the clinical data of our CD8+ T-LGL cohort (N=26) and the data of 56 CD8+ T-LGL leukemia patients described by Wlodarski *et al*. [29] and compared those with the clinical features of 36 patients with CD4+ T-LGL leukemia [28] and our cohort of 44 published [16] and 19 novel TCRγδ+ T-LGL leukemia patients. On average, similar clinical features were observed between CD8+TCRαβ+ T-LGL leukemia and TCRγδ+ T-LGL leukemia. Notably, the frequency of cytopenias and autoimmune disorders appeared to be much lower in CD4+TCRαβ+ T-LGL leukemia as compared with other types (Fig. 1).

At closing of the study, the median follow-up of the patients was 34 months (range 6–122 months). Two-thirds of patients required therapy with one or more agents (Table 1). The therapeutic strategy was largely aimed at improving cytopenias and included erythrocyte transfusions and various immunosuppressive drugs. We observed one disease-related death in these 26 patients (case 86-041) which is in line with the generally indolent clinical course of this disease.

Typical LGL immunophenotype is seen in all CD8+TCRαβ+ T-LGL

T-LGL cells of all 26 cases showed membrane co-expression of CD3, CD8 and TCRαβ molecules. The majority of leukemic LGLs expressed CD2 (100%), CD5 (77%) and CD7

Table 1. Characteristics, clinical presentation and immunophenotype of 26 patients diagnosed with CD8+TCRαβ+ T-LGL leukemia.

Case no.	Sample	Age, sex	Immunophenotype	Main clinical presentations[a]	Associated disease	LGL clone size (x 10/l)	Therapy
1	86-041	58,F	CD2/3/8/5/7	Anemia	PRCA	21.6	RBC transfusion, chlorambucil
2	96-013	58,F	CD2/3/8/16/56/57	Anemia/neutropenia	-	13.0	Unknown
3	96-043	73,F	CD2/3/8/5/7/16/56/57	Neutropenia/thrombocytopenia	Oligoarthritis	9.9	MTX
4	92-050	50,M	CD2/3/8/5/7/16neg/56neg/57/HLADR	Neutropenia	RA(RF+), DLBCL	1.9	None
5	98-126	73,M	CD2/3/8/5/7/16/56/57/HLADR	Anemia/thrombocytopenia	HCC	1.0	Chlorambucil
6	92-024	55,F	CD2/3/8/5/7/16/56neg/57/HLADR	Neutropenia/anemia	AIHA, ITP, bronchus carcinoma	7.4	CSA
7	93-027	53,M	CD2/3/8/5/7/16neg/56neg/57/HLADR	Anemia	AIHA	2.0	Unknown
8	96-067	39,F	CD2/3/8/5/16/57	Neutropenia	-	5.9	None
9	97-064	71,M	CD2/3/8/16neg/56neg/57/HLADRneg	Neutropenia	-	1.2	Unknown
10	99-100	49,M	CD2/3/8/5/7/16/56neg/57/HLADRneg	Anemia	-	5.5	Unknown
11	98-194	38,F	CD2/3/8/5/7/16/56/57neg/HLADR	Neutropenia	-	6.6	None
12	05-060	70,M	CD2/3/8/5/7/16neg/56neg/57/HLADR	Neutropenia/thrombocytopenia	-	3.1	Unknown
13	05-100	31,F	CD2/3/8/5/7/57	Neutropenia/anemia	-	2.0	Corticosteroids, CSA
14	05-191	73,M	CD2/3/8/5/7/16neg/56neg/57/HLADR	Neutropenia/recurrent infections	Chronic NK-LGL leukemia	4.4	MTX
15	03-030	41,F	CD2/3/8/5/7neg/57/HLADR	Anemia/B symptoms	PRCA	2.5	CSA
16	03-086	39,F	CD2/3/8/5/7/16/56neg/57	Fatigue	M.Hodgkin	2.0	None
17	00-113	72,M	CD2/3/8/5neg/7/16neg/56neg/57/HLADRneg	Neutropenia/B symptoms	-	3.5	None
18	93-054	62,M	CD2/3/8/7/16neg/57	Anemia/splenomegaly/B symptoms	Chronic osteomyelitis, melanoma	4.4	Chlorambucil, RBC transfusion
19	02-047	76,F	CD2/3/8/5/7/57/HLADR	B symptoms	-	12.0	None
20	91-004	86,F	CD2/3/8/5/7/16neg/56neg	Neutropenia/anemia/lymphadenopathy	-	3.8	None
21	05-281	59,M	CD2/3/8/5/7neg/16neg/56/57/HLADRneg	B symptoms	-	2.7	None
22	06-026	67,M	CD2neg/3/8/5/7/16/56neg/57/HLADRneg	Neutropenia/B symptoms/thrombocytopenia	(Oligo)RA	2.0	HCQ
23	06-038	69,M	CD2/3/8/5neg/7/16neg/56neg/57/HLADR	Neutropenia	RA	3.2	MTX
24	06-127	45,F	CD2/3/8/5/7/16neg/56/57/HLADR	Neutropenia	-	3.7	None
25	06-131	61,F	CD2/3/8/5/7/16neg/56/57/HLADR	Recurrent infections	-	1.2	None
26	06-246	47,F	CD2/3/8/5/7/16/56/57/HLADR	Neutropenia/splenomegaly	-	10.2	CSA

Abbreviations: AIHA, autoimmune hemolytic anemia; CSA, cyclosporine A; DLBCL, diffuse large B-cell lymphoma; HCC, hepatocellular carcinoma; HCQ, hydroxychloroquine; ITP, idiopathic thrombocytopenic purpura; MTX, methotrexate; PRCA, pure red cell aplasia; RA, rheumatoid arthritis; RBC, red blood cell; RF, rheumatoid factor; TCR, T-cell receptor; T-LGL, T-cell large granular lymphocyte.

[a]Neutropenia was defined as absolute neutrophil count (ANC) <1.5 x 10^9 neutrophils/l; anemia was defined as hemoglobin level <10 g/dl; thrombocytopenia was defined as platelet count <150 x 10^9/l.

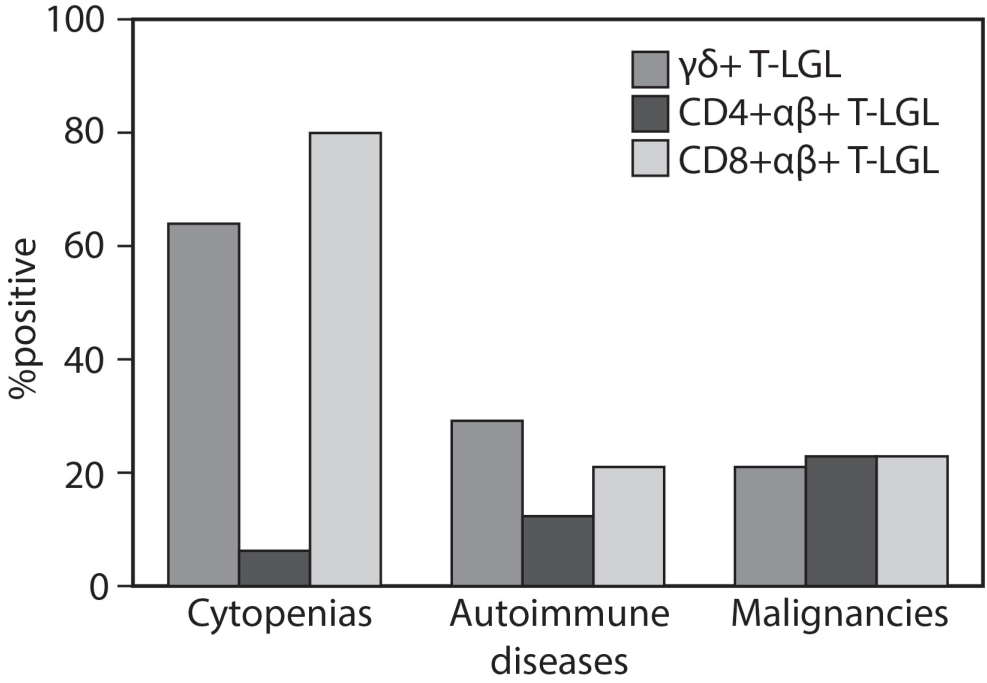


Figure 1. Relative frequency (%) of associated cytopenias, auto-immune diseases and malignancies in CD8+TCRαβ+, CD4+TCRαβ+ and TCRγδ+ T-LGL leukemia.

(81%). Furthermore, the T-LGL cells of all cases showed expression of one or more markers like CD16, CD56 or CD57 that have typically been associated with LGL and that reflect the antigen-experienced nature of the cells (Table 1). Out of 22 cases 11 were CD16 positive (50%), whereas CD56 expression was seen in 7 out of 20 cases analyzed (35%). Heterogeneous expression of CD57 could be demonstrated in 23 out of 24 evaluable cases (96%). Thus, these CD8+TCRαβ+ T-LGLs show the typical effector T cell phenotype.

TCRA and TCRB combinatorial diversity greatly differs between CD8+TCRαβ+ T-LGL and CD4+TCRαβ+ T-LGL

To evaluate whether CD8+TCRαβ+ T-LGL shows signs of antigen stimulation in their antigen receptors, we aimed for a comprehensive analysis of both the TCRα and the TCRβ chain in parallel to the HLA genotype.

First, we studied the clonotypic TCRB repertoire using specific anti-TCRVβ domain MoAbs. Dominant TCR-Vβ reactivity was observed in 22 out of 26 cases (Table 2). All

cases, including the four cases without detectable TCRVβ expression, demonstrated clonal in-frame TCRB gene rearrangements in multiplex PCR and/ or RT–PCR analysis (Table 3). Results of sequence analysis of Vβ-Jβ gene rearrangements and Vβ protein/ mRNA expression were concordant in all cases. On the basis of these results, a slight predominance of Vβ2, Vβ5 and Vβ12 was noted. Jβ2 genes were used more frequently than Jβ1 genes (62 vs. 38%), with the Jβ2.1 proportion being highest (15%). This Jβ2 predominance is in line with the non-random Jβ gene distribution as it is known from mature polyclonal PB TCRαβ+ T cells of healthy individuals [41].

To study the clonotypic TCRA repertoire, we developed a novel multiplex RT-PCR assay. Using this assay, clonal TCRA gene rearrangements could be demonstrated in all 22 cases that were analyzed (Table 2). In the remaining four cases no RNA could be isolated due to lack of material. Similar to Vβ usage, there was also no common Vα gene usage among the 22 patients studied. Genes from the Vα19, Vα8 and Vα12 families were used most frequently, being expressed in three cases each. Jα gene usage was also highly diverse (Table 4).

When visualizing the TCRA and TCRB combinatorial diversity using the Circoletto tool, both the Vα-Jα and Vβ-Jβ diversities were indeed largely random in CD8+TCRαβ+ T-LGL (Fig. 2a, 2b). This is in strong contrast to the non-random distribution of Vβ-Jβ combinations as seen in monoclonal CD4+TCRαβ+ T-LGL lymphocytosis patients [28] (Fig. 2c); unfortunately, from these CD4+TCRαβ+ T-LGLs no Vα-Jα combinatorial data are available for a direct comparison of the TCRA combinatorial diversity. For a complete picture on the combinatorial diversity of all three T-LGL subtypes, we also analyzed the Vγ-Jγ and Vδ-Jδ combinations in our cohort of 63 TCRγδ+ LGL patients using Circoletto. Similar to the diversity in CD4+TCRαβ+ T-LGL, but unlike CD8+TCRαβ+ T-LGL, the combinatorial diversity in TCRG and TCRD appeared to be non-random as well (Fig. 2d, 2e). It should be noted that the lower number of V genes might already impact on the more limited combinatorial diversity of these two loci.

Finally, we evaluated the HLA-ABC genotype of the CD8+TCRαβ+ T-LGL patients. No clear predominance of HLA-A, -B or -C alleles or combinations thereof was observed (Table 2). When evaluating the expanded Vβ and Vα families in conjunction with the HLA-ABC alleles, also no clear association was observed between a particular TCR-Vα/ Vβ specificity and the involved HLA genotype.

Collectively, these data show a clear heterogeneity in the combinatorial diversity of the TCR clonotypes in CD8+TCRαβ+ T-LGL, which does not seem to be directly linked to the HLA genotype. This heterogeneity clearly differs from the more homogeneous patterns seen in TCRγδ+ T-LGL and especially CD4+TCRαβ+ T-LGL.

Table 2. HLA genotype and Vα/Vβ usage in CD8+TCRαβ+ T-LGL leukemia patients.

Case no.	Sample	LGL expansion	Expanded TCR Vβ family[a]	Expanded TCR Vα family[b]	HLA-A	HLA-B	HLA-C
1	86-041	Oligoclonal	Vβ1	Vα19	02, 11	35, 50	04, 06
2	96-013	Oligoclonal	Vβ12	Vα19	02, 68	40, 44	03, 07
3	96-043	Monoclonal	Vβ22	Vα29	01, 25	07, 08	03, 07
4	92-050	Oligoclonal	Vβ23	Vα17	02, 02	13, 15	03, 12
5	98-126	Oligoclonal	Vβ2	Vα19	ND	ND	ND
6	92-024	Monoclonal	Vβ7.2	Vα12.2	01, 24	08, 40	03, 07
7	93-027	Oligoclonal	Vβ8.1/8.2	Vα6	02, 02	40, 40	03, 03
8	96-067	Oligoclonal	No reactivity (Vβ6 in PCR)	Vα26	02, 31	27, 58	02, 07
9	97-064	Oligoclonal	Vβ13	Vα35	24, 29	45, 49	06, 07
10	99-100	Monoclonal	Vβ17	Vα23	01, 02	08, 41	07, 17
11	98-194	Biclonal	No reactivity (Vβ6/12 in PCR)	Vα6/21	02, 03	07, 27	01, 07
12	05-060	Monoclonal	Vβ17	Vα30/26	03, 32	41, 55	03, 17
13	05-100	Monoclonal	Vβ16	Vα12.2	02, 03	18, 51	05, 14
14	05-191	Monoclonal	Vβ5	Vα1	02, 02	15, 35	03, 04
15	03-030	Biclonal	No reactivity (Vβ6/16 in PCR)	ND	02, 02	08, 08	07, 07
16	03-086	Monoclonal	Vβ14	Vα3	02, 03	08, 35	04, 07
17	00-113	Monoclonal	Vβ8.2	Vα8	01, 26	38, 44	05, 12
18	93-054	Oligoclonal	Vβ12/15	Vα12.3	01, 02	15, 57	03, 06
19	02-047	Monoclonal	Vβ7.2	Vα8	ND	ND	ND
20	91-004	Biclonal	Vβ5/6	Vα41	ND	ND	ND
21	05-281	Monoclonal	Vβ2	Vα1.2	ND	ND	ND
22	06-026	Monoclonal	No reactivity (Vβ24 in PCR)	Vα29	ND	ND	ND
23	06-038	Monoclonal	Vβ2	Vα12.1	01, 24	08, 18	05, 07
24	06-127	Monoclonal	Vβ13	ND	01, 11	27, 35	02, 04
25	06-131	Monoclonal	Vβ12	ND	ND	ND	ND
26	06-246	Monoclonal	Vβ5.1	ND	ND	ND	ND

Abbreviations: ND, not done; TCR, T-cell receptor; T-LGL, T-cell large granular lymphocyte.

[a]Dominant TCR Vβ family usage defined by immunophenotyping and molecular analysis. [b]Dominant TCR Vα family usage defined by molecular analysis.

Table 3. Amino acid sequences of CDR3 motifs of in-frame TCRB rearranged alleles in patients with CD8+TCRαβ+ T-LGL leukemia.

Case no.	Sample	Rearrangement	V					N-(D)-N														J				
1	86-041	Vβ1-Jβ2.1	C	A	S	S		L	S	G	R	A	L									N	E	Q	F	F
2	96-013	Vβ12.1-Jβ2.2	C	A	I	S	E	G	S	G	P											G	E	L	F	F
3	96-043	Vβ22-Jβ2.6	C	A	S			G	G	D	R	G	T							G	A	N	V	L	T	F
4	92-050	Vβ23-Jβ1.5	C	A	S	S	L	G	G	R	Y									S	N	Q	P	Q	H	F
5	98-126	Vβ2-Jβ1.3	C	S	A			S	L	G	G	R	P	T	I	A					G	N	T	I	Y	F
6	92-024	Vβ7.2-Jβ1.5	C	A	S	S	Q	D	V	R	P	P	P	E	D	R	P	Y		S	N	Q	P	Q	H	F
7	93-027	Vβ8.2-Jβ2.5	C	A	S	S	L	G	T	G	G	M										E	T	Q	Y	F
8	96-067	Vβ6.2-Jβ2.1	C	A	S	S		L	A	H										S	Y	N	E	Q	F	F
9	97-064	Vβ13.3-Jβ2.7	C					G	S	L	G	Q	G	A	W							Y	E	Q	Y	F
10	99-100	Vβ17-Jβ2.7	C	A	S			S	P	E	S	L	F								S	Y	E	Q	Y	F
11	98-194	Vβ6.4-Jβ1.1 Vβ12.3-Jβ2.3	C C	A A	S	S		R C	S Y	F Q	S P	P G	S L	L D	D	T			S L	S P	L R	F A	V D	E T	A Q	F Y	F F
12	05-060	Vβ17-Jβ2.7	C	A	S	S		I	F	R	G	N											E	Q	Y	F
13	05-100	Vβ16-Jβ2.5	C	A	S	S		P	V	G	A	Y	P	K								E	T	Q	Y	F
14	05-191	Vβ5.1-Jβ1.4	C	A	S	S	L	A	K	G	K	G	A							T	N	E	K	L	F	F
15	03-030	ND																								
16	03-086	Vβ14-Jβ2.1	C	A	S			N	N	R	G	S									Y	N	E	Q	F	F
17	00-113	Vβ8.2-Jβ1.5	C	A	S	S		W	G	G											N	Q	P	Q	H	F
18	93-054	Vβ12.1-Jβ2.3	C	A	I	S		G	R	L	A	G	G	R	T					S	T	D	T	Q	Y	F
19	02-047	Vβ7.2-Jβ2.2	C	A	S			G	G	V	G	G										G	E	L	F	F
20	91-004	Vβ5.6-Jβ2.1 Vβ6.5-Jβ1.1	C C	A A	S S	S S		L F	D S	N P	Y	T	R	P							Y	N	E E	Q A	F F	F F
21	05-281	Vβ2-Jβ1.4	C	A				V	P	T	G	R									N	E	K	L	F	F
22	06-026	Vβ24-Jβ2.7	C	A	T	S		R	D	L	L	T										Y	E	Q	Y	F
23	06-038	Vβ2-Jβ2.5	C	S	A			R	N	G	P	N	Y								Q	E	T	Q	Y	F
24	06-127	Vβ13.6-Jβ1.5	C	A	S	S		Y	G	S	P	L	D	I	D	S	A	I	S				P	Q	H	F
25	06-131	Vβ12.2-Jβ1.2	C	A	S	S		P	K	G												Y	G	Y	T	F
26	06-246	Vβ5.1-Jβ1.2	C	A	S	S	L	G	S	G												Y	G	Y	T	F

Abbreviations: CDR3, complementarity determining region 3; ND, not done; TCRB, T-cell receptor β; T-LGL, T-cell large granular lymphocyte.

A L I V G (P): neutral side chain; D E: acidic side chain; S T: aliphatic side chain; N Q: amide side chain; K R H: basic side chain; M: sulfur-containing side chain; F Y W: cyclic side chain.

Table 4. Amino acid sequences of CDR3 motifs of in-frame TCRA rearranged alleles in patients with CD8+TCRαβ+ T-LGL leukemia.

Case no.	Sample	Rearrangement	V					N-(D)-N								J										
1	86-041	Vα19-Jα49	C	A	L	S	E	S	G												G	N	Q	F	Y	F
2	96-013	Vα19-Jα26	C	A	L	S	E	G	S	R	F									Y	G	Q	N	F	V	F
3	96-043	Vα29-Jα52	C					G	R	V											A	G	G	T	S	F
4	92-050	Vα17-Jα20	C	A				T	L	S									S	N	D	Y	K	L	S	F
5	98-126	Vα19-Jα37	C	A	L	S	E	A	E	G	S								S	N	T	G	K	L	I	F
6	92-024	Vα12.2-Jα53	C	A				V	T									G	G	S	N	Y	K	L	T	F
7	93-027	Vα6-Jα21	C					V	G												F	N	K	F	Y	F
8	96-067	Vα26.2-Jα20	C	I				P	S	P	S									N	D	Y	K	L	S	F
9	97-064	Vα35-Jα49	C	A				G	F											T	G	N	Q	F	Y	F
10	99-100	Vα23-Jα52	C	A				A	P	V							G	G	T	S	Y	G	K	L	T	F
11	98-194	Vα8.6-Jα56 Vα21-Jα57	C C	A A	V V	S		L K										T	G G	A G	N S	S E	K K	L L	T V	F F
12	05-060	Vα26.2-Jα24 Vα30-Jα44	C C	I G	L	R	D	V T	E P	S G	W N	G	K	F	Q				G	T	A	S	K	L	T	F F
13	05-100	Vα12.2-Jα23	C	A	V															Q	G	G	K	L	I	F
14	05-191	Vα1.1-Jα10	C	A				V	D	P	G	L	A	A						G	G	N	K	L	T	F
15	03-030	ND																								
16	03-086	Vα3-Jα36	C	A				S	D									Q	T	G	A	N	N	L	F	F
17	00-113	Vα8.1-Jα28	C	A	V			M								Y	S	G	A	G	S	Y	Q	L	T	F
18	93-054	Vα12.3-Jα9	C	A	M	S		A	V	M	R										G	F	K	T	I	F
19	02-047	Vα8.1-Jα39	C	A	V			M	S	D										A	G	N	M	L	T	F
20	91-004	Vα41-Jα48	C	A	V			N																		F
21	05-281	Vα1-2-Jα33	C	A				T	L										D	S	N	Y	Q	L	I	W
22	06-026	Vα29-Jα45	C	A	A			K	G	F													G	L	T	F
23	06-038	Vα12.1-Jα34	C	V	V			K										S	Y	N	T	D	K	L	I	F
24	06-127	ND																								
25	06-131	ND																								
26	06-246	ND																								

Abbreviations: CDR3, complementarity determining region 3; ND, not done; TCRA, T cell receptor α; T-LGL, T cell large granular lymphocyte.
A L I V G (P): neutral side chain; D E: acidic side chain; S T: aliphatic side chain; N Q: amide side chain; K R H: basic side chain; M: sulfur-containing side chain; F Y W: cyclic side chain.

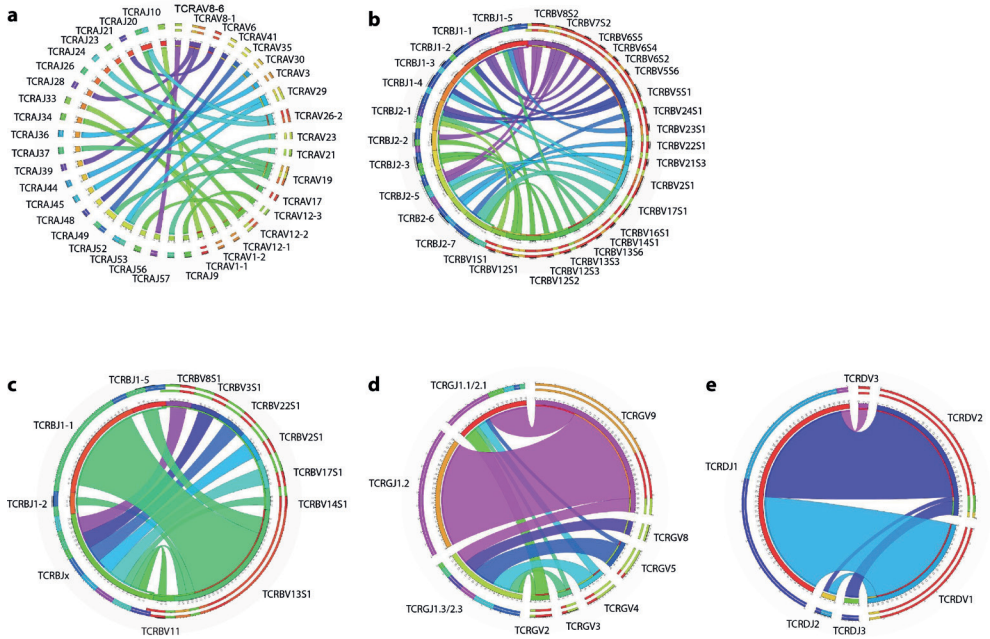


Figure 2. Frequencies of V-J pairing in TCRαβ+ T-LGL and TCRγδ+ T-LGL leukemia.
Highly diverse Vα-Jα (**a**) and Vβ-Jβ (**b**) combinations are seen in our cohort of 26 patients with CD8+ T-LGL leukemia, while the Vβ-Jβ pairing is clearly non-random in monoclonal CD4+ T-LGL proliferations (**c**). Also in TCRγδ+ T-LGL leukemia limited combinatorial diversity of TCRG and TCR genes (**d** and **e**, respectively) is seen. (Blue to purple rectangular bands) J genes and (red to cyan rectangular bands) V genes. The width of the bands is proportional to the number of times the V and J genes are connected. This figure was generated using the Circos software package [47].

Lack of common TCRA and TCRB CDR3 motifs in CD8+TCRαβ+T-LGL

Given the lack of skewing of TCRA and TCRB combinatorial diversity, we then explored the possibility of a more subtle TCR homology in CD8+TCRαβ+T-LGL leukemia patients. To this end, we analyzed TCRA and TCRB junctional diversity by studying CDR3 sequences in more detail. A total of 24 T-LGL TCRA CDR3 clonotypes from 22 patients and 27 T-LGL TCRB CDR3 clonotypes from 25 patients were evaluated, but the TCRA and TCRB CDR3 motifs of the immunodominant T cell clones did not show clearly identical sequences (Tables 3 and 4).

To exclude the possibility that the number of evaluable TCRB CDR3 sequences was limiting the possibility to find clear similarities, additional CDR3 sequences that had been published in the literature were included for further analysis. In this way, 81 additional TCRB CDR3 motifs of a large series of 56 CD8+TCRαβ+ T-LGL leukemia patients [29] could be evaluated. Similar to our cohort, in a proportion of patients more than one immunodominant clone was found, suggesting that some T-LGL proliferations are

biclonal. In this extended data set, the Vβ-Jβ combinatorial diversity appeared to be equally heterogeneous (Supplementary Fig. 1) as in our own cohort.

For a more comprehensive evaluation of CDR3 motifs, detailed *in silico* analysis was performed on the 108 combined TCRB CDR3 sequences in parallel to 14 TCRB CDR3 sequences of earlier described CD4+ T-LGL [28]. By applying a recently described sequence motif-based clustering methodology [40] using thresholds of 50% amino-acid identity and 70% similarity between any two CDR3 sequences, 13 out of 14 CD4+ T-LGL displayed a highly homogeneous and similar TCR with clear similarities in length and amino-acid positions in the CDR3 sequence logo (Fig. 3a). The similarity of the CDR3 sequence logo was even more impressive when concentrating on a higher level cluster of 11 CD4+ T-LGL cases that are all characterized by TCRVβ13.1-Jβ1.1 rearrangements (Fig. 3b). This is in line with the proposed CMV antigen-driven selection in the pathogenesis of this type of T-LGL. Interestingly, the two CD4+ T-LGL that were slightly different from the other 11 CD4+ T-LGLs based on Jβ1.5 usage did show low level clustering with two CD8+ T-LGLs, as evidenced from the common Jβ1.5 and Vβ8/Vβ13 gene usage, and the CDR3 sequence logo (Fig. 3c). Finally, some clustering was seen between four other CD8+ T-LGL cases that

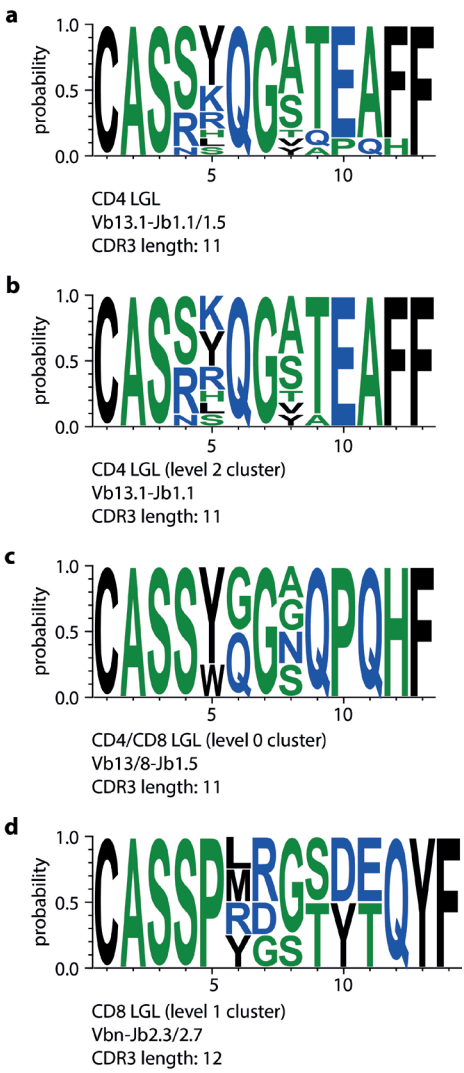


Figure 3. Sequence logos of selected subsets in T-LGL leukemia.
(**a**) Subset 1 comprises 13 Vβ13.1-Jβ1.1/Jβ1.5 gene rearrangements in CD4+ T-LGL leukemia cases, characterized by pronounced overall similarity. (**b**) The largest high-level subset in the present study is present in subset 2, comprising clonally expanded CD4+ T-LGL showing Vβ13.1-Jβ1.1. (**c**) Low level clustering was seen in the Vβ8/Vβ13-Jβ1.5 gene rearrangements in two CD4+ LGL and two CD8+ LGL cases. (**d**) Some higher level clustering is present in Vβn.-Jβ2.3/2.7 gene rearrangements in four CD8+ T-LGL leukemia cases. The height of the symbols within the stack indicates the relative frequency of each amino acid at that position. Amino-acid position is according to the IMGT numbering for the V domain. This figure was generated using WebLogo (weblogo.berkeley.edu/).

were characterized by Jβ2.3/2.7 usage and a CDR3 length of 12 (Fig. 3d). Unfortunately not enough TCRA sequences were available for a meaningful *in silico* TCRA CDR3 analysis.

Collectively the *in silico* analyses illustrate that, in strong contrast to CD4+ T-LGL, CD8+ T-LGL do not show clear and consistent signs of TCR homology that would reflect involvement of a common antigen.

DISCUSSION AND CONCLUSION

Molecular analysis of the TCR repertoire can be a powerful tool in the study of T cell responses to pathogens and in autoimmune diseases. Thus, analysis of the TCR expression pattern in patients with T-LGL leukemia might provide insight into the pathogenesis of this disorder. Similar TCR clonotypes between T-LGL clones of different patients would in that respect be suggestive of a common antigenic stimulus underlying the pathogenesis of this disorder. Furthermore, it has been suggested that the cytopenias associated with T-LGL leukemia would be the result of highly specific recognition and killing of individual hematopoietic cell lineages by T-LGL clones.

In the present study, we identified and characterized TCR clonotypes in a group of CD8+TCRαβ+ T-LGL leukemia patients. We could not detect specific predominant Vβ family usage in our CD8+TCRαβ+ T-LGL cohort. Immunophenotypical analysis of TCR Vα expression has so far only been explored in a minority of TCRαβ+ T-LGL leukemia cases. Likewise, complete sequencing of TCRA gene rearrangements has only been performed in few (N=5) T-LGL leukemia patients, thus far not showing any signs of common Vα or Jα gene usage [42]. In our cohort of 22 patients, we could not detect preferential Vα or Jα gene usage; moreover, TCRA CDR3 sequence analysis did not show a common amino-acid motif between the various patients, which is thus in line with the TCRB results.

In silico analysis using a recent purpose-built bioinformatics method did not identify common TCRB CDR3 amino-acid sequences in a large cohort of CD8+ T-LGL leukemia patients. This is in strong contrast to monoclonal CD4+Vβ13.1+ T-LGL proliferations, in which virtually all published TCRB CDR3 sequences can be assigned to one cluster with unique characteristics. Our current results therefore underline the distinct pathogenesis between the CD4+ and CD8+ T-LGL disease entities.

On the basis of our findings, we did not find straightforward evidence for common (super)antigen involvement in the pathogenesis of CD8+TCRαβ+ T-LGL leukemia, since no immunodominant clones with identical or highly similar TCRA and TCRB CDR3 amino-acid sequences could be identified in a majority of patients. In the study of Wlodarski *et al*. [29] identical expanded clonotypes were found in only 2 out of 56 patients. However, clonotypes specific for malignant clones were not encountered to a

great extent in 172 clones from healthy individuals. We did not encounter the 108 clonotypes of the 82 evaluated leukemic LGL cases in CD8+ T cells of healthy individuals (data not shown), although high-throughput analysis of the TCRB repertoire of CD8+ T cells would be needed to draw firm conclusions. Recently, deep sequencing of the T cell repertoire in healthy controls and CD8+ T-LGL leukemia has been performed and confirms that T-LGL clonotypes are not present in the general public and are therefore private to the disease [43].

The heterogeneity of TCR clonotypes in clonal CD8+TCRαβ+ T-LGL proliferations is partially explained by the presence of biclonal LGL proliferations and clonal switching. Both have been described in T-LGL leukemia and are suggestive of pervasive antigenic drive [22]. In addition, Clemente *et al*. [43] recently demonstrated that individual T-LGL clones were present at basal levels in almost all other T-LGL leukemia cases. This suggests the presence of an as yet undefined mechanism whereby certain clonotypes may predispose an individual toward the extreme monoclonal expansions commonly found in T-LGL leukemia [43]. Another explanation for the seemingly low level of similarity between the dominant clonotypes in our study might be the large variability in HLA genotype in our patient series. Peptide binding is affected through the physical amino acid properties and the tertiary CDR3 structure. Therefore, the linear amino acid homology comparisons might not be the most appropriate method to identify common motifs in TCR molecules.

Altogether this may suggest that CD8+TCRαβ+ T-LGL clones could have evolved in a stepwise manner from an initial polyclonal/oligoclonal immune response directed against multiple (auto)antigenic targets. Following an exaggerated immune response, secondary molecular events would then lead to global deregulation of cell proliferation and survival of one or a few clonotypes. Among the survival signaling pathways, the Janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway has been associated with LGL transformation [26]. Most recently, the role of STAT family genes (STAT3 and STAT5b) in the pathogenesis of T-LGL leukemia has been emphasized [44]. Activating somatic mutations in STAT3 were found in up to 40% of T-LGL leukemia patients and it has been suggested that mutational analysis of STAT3 might distinguish true T-LGL leukemia cases from clonally skewed reactive processes [11,45,46].

In summary, no clear indications for common TCRA or TCRB CDR3 motifs were found in our CD8+TCRαβ+ T-LGL cohort. When clonotypes of our 26 patients were cross-referenced against the previously reported clonotypic database of 56 T-LGL leukemia patients, we could only identify homologous clonotypes between a limited number of patients. This is in contrast to the shared clonotypes as seen in CD4+TCRαβ+ and TCRγδ+ T-LGL leukemia and might point to a more random clonal selection in TCRαβ+ T-LGL leukemia. The heterogeneity of clonal CD8+TCRαβ+ T-LGL proliferations might

in fact suggest that multiple pathogens or autoantigens are involved. Additional studies taking into account the triad of HLA genotype, peptide-groove binding and TCR specificity are needed to precisely define the impact of (auto)-antigen stimulation in the pathogenesis of CD8+TCRαβ+ T-LGL leukemia.

ACKNOWLEDGEMENTS

We would like to thank Dr Kirsten van Lom (Department of Hematology, Erasmus MC) for cytomorphological analysis, Dr Kees Sintnicolaas (Sanquin Blood Bank, South West region Rotterdam) for performing the HLA typing, Mr Edwin de Haas (Department of Immunology, Erasmus MC) for high-speed sorting of cell populations, Mr Edwin Florencia for help with sequence analysis of TCRB CDR3 motifs, and Mrs Marieke Comans-Bitter and Mrs Sandra de Bruin-Versteeg for help in preparing the figures.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Langerak, A.W., Sandberg, Y. & van Dongen, J.J. Spectrum of T-large granular lymphocyte lymphoproliferations: ranging from expanded activated effector T cells to T cell leukaemia. *Br. J. Haematol.* **123**, 561-562 (2003).
2. Loughran Jr, T.P. Clonal diseases of large granular lymphocytes. *Blood* **82**, 1-14 (1993).
3. Sokol, L. & Loughran Jr, T.P. Large granular lymphocyte leukemia. *Oncologist* **11**, 263-273 (2006).
4. Dhodapkar, M.V., Li, C.Y., Lust, J.A., Tefferi, A. & Phyliky R.L. Clinical spectrum of clonal proliferations of T-large granular lymphocytes: a T cell clonopathy of undetermined significance? *Blood* **84**, 1620-1627 (1994).
5. Lamy, T. & Loughran Jr, T.P. Current concepts: large granular lymphocyte leukemia. *Blood Rev.* **13**, 230-240 (1999).
6. Lamy, T. & Loughran Jr, T.P. Clinical features of large granular lymphocyte leukemia. *Semin. Hematol.* **40**, 185-195 (2003).
7. Rose, M.G. & Berliner, N. T cell large granular lymphocyte leukemia and related disorders. *Oncologist* **9**, 247-258 (2004).

8. Bareau, B. *et al.* Analysis of a French cohort of patients with large granular lymphocyte leukemia: a report on 229 cases. *Haematologica* **95**, 1534-1541 (2010).
9. Lamy, T. & Loughran Jr, T.P. How I treat LGL leukemia. *Blood* **117**, 2764-2774 (2011).
10. Semenzato, G., Zambello, R., Starkebaum, G., Oshimi, K. & Loughran Jr, T.P. The lymphoproliferative disease of granular lymphocytes: updated criteria for diagnosis. *Blood* **89**, 256-260 (1997).
11. Jerez, A. *et al.* STAT3 mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T cell large granular lymphocyte leukemia. *Blood* **120**, 3048-3057 (2012).
12. Dearden, C. Large granular lymphocytic leukaemia pathogenesis and management. *Br. J. Haematol.* **152**, 273-583 (2011).
13. Alekshun, T.J., Tao, J. & Sokol, L. Aggressive T cell large granular lymphocyte leukemia: a case report and review of the literature. *Am. J. Hematol.* **82**, 481-485 (2007).
14. Gentile, T.C. *et al.* CD3+CD56+ aggressive variant of large granular lymphocyte leukemia. *Blood* **84**, 2315-2321 (1994).
15. Lima, M. *et al.* TCRalphabeta+/CD4+ large granular lymphocytosis: a new clonal T cell lymphoproliferative disorder. *Am. J. Pathol.* **163**, 763-771 (2003).
16. Sandberg, Y. *et al.* TCRgammadelta+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRgammadelta+ T cells. *Leukemia* **20**, 505-513 (2006).
17. Sabnani, I. & Tsang, P. Are clonal T cell large granular lymphocytes to blame for unexplained haematological abnormalities? *Br. J. Haematol.* **136**, 30-37 (2007).
18. O'Keefe, C.L. *et al.* Molecular analysis of TCR clonotypes in LGL: a clonal model for polyclonal responses. *J. Immunol.* **172**, 1960-1969 (2004).
19. Sokol, L., Agrawal, D. & Loughran Jr, T.P. Characterization of HTLV envelope seroreactivity in large granular lymphocyte leukemia. *Leuk. Res.* **29**, 381-387 (2005).
20. Starkebaum, G. *et al.* Serum reactivity to human T cell leukaemia/lymphoma virus type I proteins in patients with large granular lymphocytic leukaemia. *Lancet* **1**, 596-599 (1987).
21. Zambello, R. *et al.* Analysis of the T cell receptor in the lymphoproliferative disease of granular lymphocytes: superantigen activation of clonal CD3+ granular lymphocytes. *Cancer Res.* **55**, 6140-6145 (1995).
22. Clemente, M.J. *et al.* Clonal drift demonstrates unexpected dynamics of the T cell repertoire in T-large granular lymphocyte leukemia. *Blood* **118**, 4384-4393 (2011).
23. Lamy, T., Liu, J.H., Landowski, T.H., Dalton, W.S. & Loughran Jr, T.P. Dysregulation of CD95/CD95 ligand-apoptotic pathway in CD3(+) large granular lymphocyte leukemia. *Blood* **92**, 4771-4777 (1998).
24. Liu, J.H. *et al.* Blockade of Fas-dependent apoptosis by soluble Fas in LGL leukemia. *Blood* **100**, 1449-1453 (2002).
25. Yang, J. *et al.* Antigen activation and impaired Fas-induced death-inducing signaling complex formation in T-large-granular lymphocyte leukemia. *Blood* **111**, 1610-1616 (2008).
26. Teramo, A. *et al.* Intrinsic and extrinsic mechanisms contribute to maintain the JAK/STAT pathway aberrantly activated in T-type large granular lymphocyte leukemia. *Blood* **121**, 3843-3854 S1 (2013).

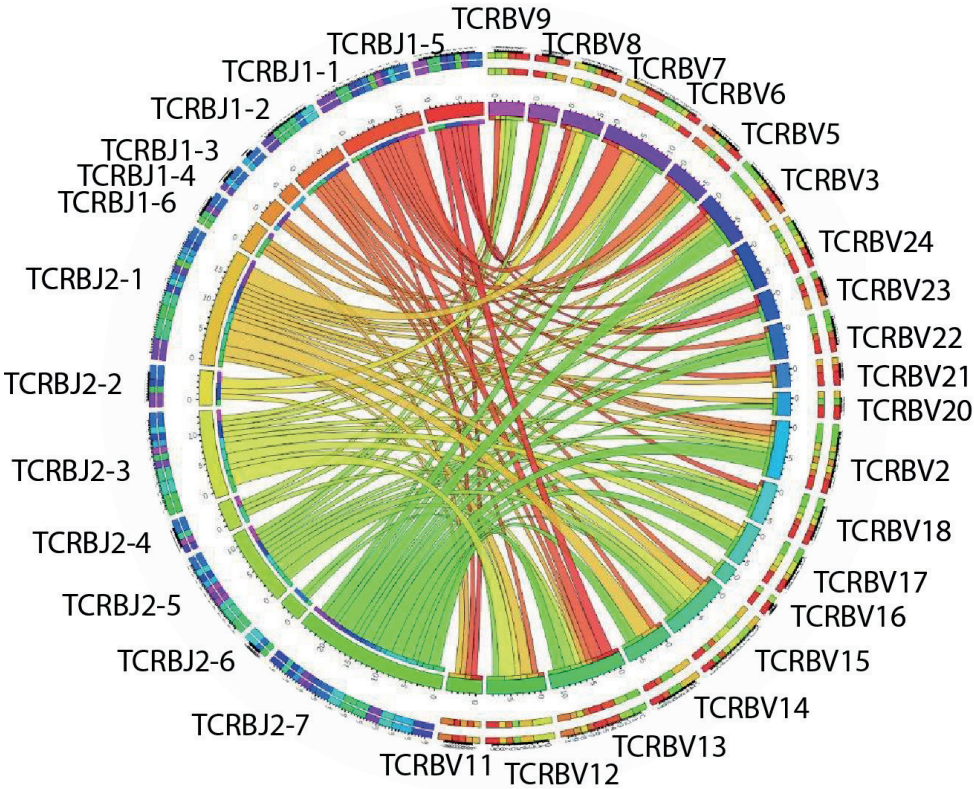
27. Borg, N.A. *et al.* The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. *Nat. Immunol.* **6**, 171-180 (2005).
28. Garrido, P. *et al.* Monoclonal TCR-Vbeta13.1+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis: evidence for an antigen-driven chronic T cell stimulation origin. *Blood* **109**, 4890-4898 (2007).
29. Wlodarski, M.W. *et al.* Pathologic clonal cytotoxic T cell responses: nonrandom nature of the T cell-receptor restriction in large granular lymphocyte leukemia. *Blood* **106**, 2769-2780 (2005).
30. Kjer-Nielsen, L. *et al.* A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* **18**, 53-64 (2003).
31. Herling, M., Khoury, J.D., Washington, L.T., Duvic, M., Keating, M.J. & Jones D. A systematic approach to diagnosis of mature T cell leukemias reveals heterogeneity among WHO categories. *Blood* **104**, 328-335 (2004).
32. Van den Beemd, R. *et al.* Flow cytometric analysis of the Vbeta repertoire in healthy controls. *Cytometry* **40**, 336-345 (2000).
33. Langerak, A.W. *et al.* Molecular and flow cytometric analysis of the Vbeta repertoire for clonality assessment in mature TCRalphabeta T cell proliferations. *Blood* **98**, 165-173 (2001).
34. Van Dongen, J.J. & Wolvers-Tettero, I.L. Analysis of immunoglobulin and T cell receptor genes. Part I: Basic and technical aspects. *Clin. Chim. Acta.* **198**, 1-91 (1991).
35. Beillard, E. *et al.* Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) – a Europe against cancer program. *Leukemia* **17**, 2474-2486 (2003).
36. Van Dongen, J.J. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMWED-2 Concerted Action BMH4-CT-98-3936. *Leukemia* **17**, 2257-2317 (2003).
37. Langerak, A.W., Szczepanski, T., van der Burg, M., Wolvers-Tettero, I.L. & van Dongen, J.J. Heteroduplex PCR analysis of rearranged T cell receptor genes for clonality assessment in suspect T cell proliferations. *Leukemia* **11**, 2192-2199 (1997).
38. Breit, T.M. & van Dongen, J.J. Unravelling human T cell receptor junctional region sequences. *Thymus* **22**, 177-199 (1994).
39. Darzentas, N. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* **26**, 2620-2621 (2010).
40. Darzentas, N. *et al.* A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia* **24**, 125-132 (2010).
41. Freeman, J.D., Warren, R.L., Webb, J.R., Nelson, B.H. & Holt, R.A. Profiling the T cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* **19**, 1817-1824 (2009).
42. Kasten-Sportes, C., Zaknoen, S., Steis, R.G., Chan, W.C., Winton, E.F. & Waldmann, T.A. T cell receptor gene rearrangement in T cell large granular leukocyte leukemia: preferential V alpha but diverse J alpha usage in one of five patients. *Blood* **83**, 767-775 (1994).

43. Clemente, M.J. *et al.* Deep sequencing of the T cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. *Blood* **122**, 4077-4085 (2013).
44. Rajala, H.L. *et al.* Discovery of somatic STAT5b mutations in large granular lymhocytic leukemia. *Blood* **121**, 4541-4550 (2013).
45. Koskela, H.L. *et al.* Somatic STAT3 mutations in large granular lymphocytic leukemia. *N. Engl. J. Med.* **366**, 1905-1913 (2012).
46. Ohgami, R.S., Ma, L., Merker, J.D., Martinez, B., Zehnder, J.L. & Arber, D.A. STAT3 mutations are frequent in CD30+ T cell lymphomas and T cell large granular lymphocytic leukemia. *Leukemia* **27**, 2244-2247 (2013).
47. Krzywinksi, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639-1645 (2009).

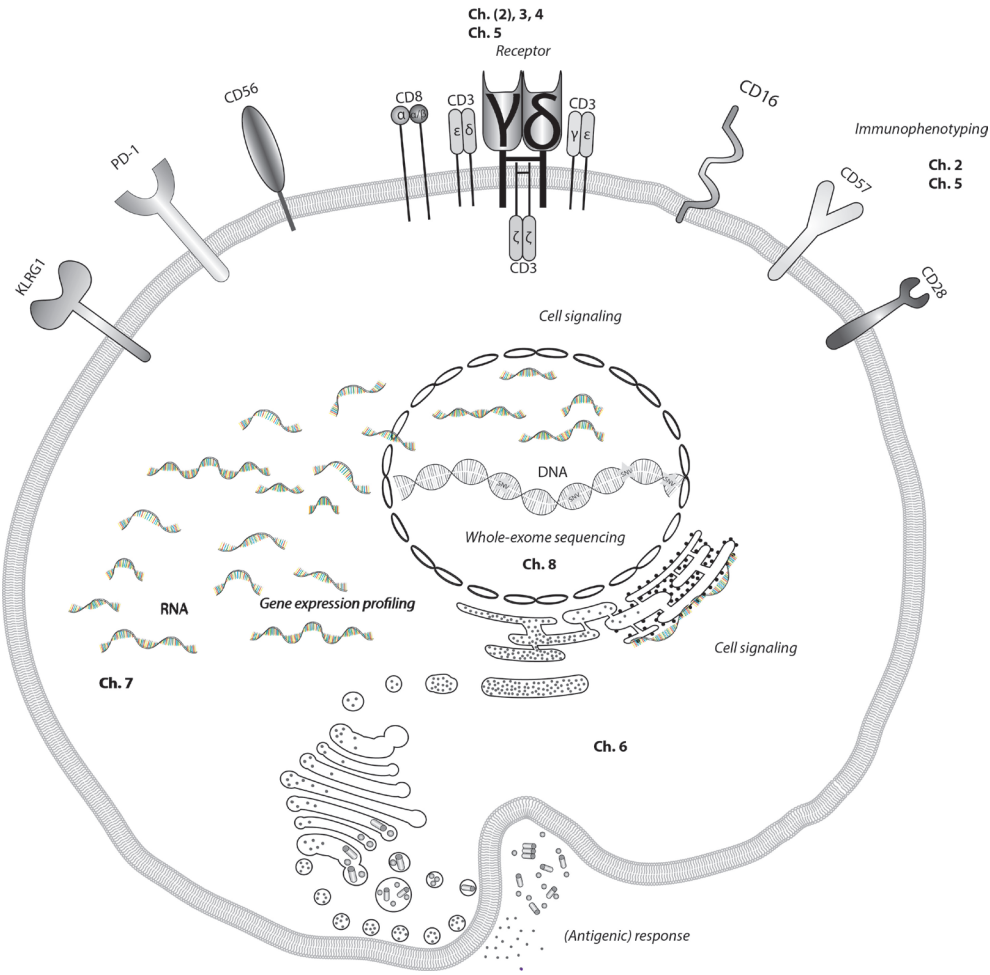
SUPPLEMENTAL DATA

Supplementary Table 1. Composition of TCRA multiplex PCR tubes.

Tube A		**Size of PCR product**			
Vα1-1	GAGCCATTGTCCAGATAAACTG	232	30	Cα	GGTACACGGCAGGGTCAG
Vα8-4	GGAGCCCTGGTTCTGCTG	239			
Vα8-7	AAGAAGCCCCTCTGGAACTG	247			
Vα12-1	CCAGAGGGAGCCACTGTC	235			
Vα12-2	GAGCCATTGCCTCTCTCAACT	231			
Vα12-3	GGGAGCCATTGTTTCTCTCA	233			
Vα30	GGGGAAGATGCTGTCATCAA	231			
Vα35	GGAGAAGATGTCTCCATGAACT	231			
Vα39	CATGCAGGAGGGAAAAAACT	242			
Vα40	GGAGGGAGCATCTGTGACT	225			
Vα41	GCCCAGGAAGGAGAATTTATCA	234			
Tube B					
Vα3	GTTGCTGAAGGGAATCCTCT	250	30	Cα	GGTACACGGCAGGGTCAG
Vα4	AGAAGTGAACATAACCTGTAGCCA	240			
Vα9-2	AAACTGCACGTACACAGCCA	219			
Vα13-1	GAGACAGCGCTGTTATCAAGTG	234			
Vα16	TTTAAAGGGGCCCCAGTG	233			
Vα17	AGGAGGGTGAAAATGCCA	237			
Vα18	CCAGTTACCCTCCCTGAGAG	251			
Vα31	GTGAGACCGTGAAACTGGACT	243			
Vα33	AGGCAGAAAGGAGTAGCTGTGA	243			
Vα38-2	TCTGTGCAGGAGGCAGAGA	256			
Tube C					
Vα5	ACAGCTCCGTTATAAACTGCAC	231	30	Cα	GGTACACGGCAGGGTCAG
Vα8-1	GCCTCACTGGAGTTGGGAT	236			
Vα22	GGAGGGAGCCAATTCCAC	230			
Vα24	GGGAGACAGCACCAATTTCA	233			
Vα26-2	GAGCCTGTTCACTTGCCTTG	231			
Vα27	GGGAGAAAATCTCACTGTGTACT	230			
Vα29	GCGTCCAGGAAGGAAGAATTT	246			
Vα32	GAAATGGCCGTTATTAATGACA	235			
Vα36	GGGAGACACCGTAACTCTCAAT	233			
Vα37	GGAAGGTGACAGCGTCACA	238			
Vα38-1	GCAGGAGGCAGAGACTGTG	252			
Tube D					
Vα2	GGAGCTGTGGTGGAAATCT	221	30	Cα	GGTACACGGCAGGGTCAG
Vα6	ACATTCAGGAGGGTAAAACGG	246			
Vα7	CGTTGCCTCCATGAGCTG	224			
Vα8-3	GGAGCCTCACTGGAGTTGAGA	239			
Vα8-5	GAAGGAGCCTCACTGGAGTT	315			
Vα8-6	AAGCCCCTGTGGAGCTG	238			



Supplementary Figure 1. Frequencies of Vβ-Jβ pairing in TCRαβ+ T-LGL leukemia patients.
Highly diverse combinations are seen in a cohort of 82 patients with $CD8^+$ T-LGL leukemia, representing 108 clonotypes.


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 8
RNA
Gene expression profiling
Cell signaling
Ch. 7
Ch. 6
(Antigenic) response

Chapter 6

Dysregulated signaling, proliferation and apoptosis impact on the pathogenesis of TCRγδ+ T cell large granular lymphocyte leukemia

Martine J. Kallemeijn[1], Dick de Ridder[2], Joyce Schilperoord-Vermeulen[1], Michèle Y. van der Klift[1], Yorick Sandberg[1], Jacques J.M. van Dongen[1], Anton W. Langerak[1]

[1]Laboratory of Medical Immunology, department of Immunology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; [2]Department of Bioinformatics, Wageningen University, Wageningen, The Netherlands.

PLOS ONE **12(4)**, e0175670 (2017)

ABSTRACT

TCRγδ+ T-LGL leukemia is a rare form of chronic mature T cell disorders in elderly, which is generally characterized by a persistently enlarged CD3+CD57+TCRγδ+ large granular lymphocyte population in the peripheral blood with a monoclonal phenotype. Clinically, the disease is heterogeneous, most patients being largely asymptomatic, although neutropenia, fatigue and B symptoms and underlying diseases such as autoimmune diseases or malignancies are also often observed. The etiology of TCRγδ+ T-LGL proliferations is largely unknown. Here, we aimed to investigate underlying molecular mechanisms of these rare proliferations by performing gene expression profiling of TCRγδ+ T-LGL versus normal TCRγδ+ T cell subsets.

From our initial microarray dataset we observed that TCRγδ+ T-LGL leukemia forms a separate group when compared with different healthy control TCRγδ+ T cell subsets, correlating best with the healthy TemRA subset. The lowest correlation was seen with the naive subset. Based on specific comparison between healthy control cells and TCRγδ+ T-LGL leukemia cells we observed up-regulation of survival, proliferation and hematopoietic system related genes, with a remarkable down-regulation of apoptotic pathway genes. RQ-PCR validation of important genes representative for the dataset, including apoptosis (XIAP, CASP1, BCLAF1 and CFLAR), proliferation/development (ID3) and inflammation (CD28, CCR7, CX3CR1 and IFNG) processes largely confirmed the dysregulation in proliferation and apoptosis.

Based on these expression data we conclude that TCRγδ+ T-LGL leukemia is likely the result of an underlying aberrant molecular mechanisms leading to increased proliferation and reduced apoptosis.

Keyword(s): TCRγδ+ T cells, T-LGL, leukemia, gene expression profiling, apoptosis.

INTRODUCTION

T cell large granular lymphocytic (T-LGL) leukemia is a heterogeneous chronic mature T cell neoplasia, which is recognized as a separate hematological disorder according to the World Health Organization (WHO) classification [1]. T-LGL leukemia originates from normal LGL cells which comprise 10-15% of peripheral blood mononuclear cells (PBMCs) [2] and can be subdivided into two major groups based on the type of T cell receptor (TCR): TCRαβ or TCRγδ. The majority of T-LGL leukemia involves the TCRαβ+ CD8+ variant (80-90%), while only a small part of the T-LGL leukemia has a TCRαβ+CD4+ (1-5%) or TCRγδ+ phenotype (5%) [3,4]. TCRγδ+ T-LGL leukemia is a chronic and heterogeneous disorder, which in fact comprises a spectrum – from lymphoproliferative disease to leukemia – and generally shows an indolent disease course, affecting mostly elderly patients with an average age of 60 years [5-7]. Approximately one-third of the patients is asymptomatic at diagnosis [8,9]. Most clinical features concern neutropenia, recurrent bacterial infections and B symptoms that are associated with chronic leukemia. Typically, TCRγδ+ T-LGL leukemia is highly associated with cytopenia, autoimmune diseases such as rheumatoid arthritis, and malignancies varying from other hematological cancers to solid tumors [4]. The diagnosis is based on a persistent (>6 months) monoclonal CD3+CD57+TCRγδ+ LGL population (>0.4 x 10^9/L) in the peripheral blood (PB) and/or bone marrow (BM), confirmed by flow cytometry or cell morphology [9]. Furthermore, TCRγδ+ T-LGL leukemia is mostly CD4-negative, partly CD8-positive, and approximately 50% show CD16 and CD56 expression [6].

Despite recent advances for T-LGL leukemia in general, the disease etiology of TCRγδ+ T-LGL leukemia remains largely unknown. It has been hypothesized that chronic (antigenic) stimulation would play a major role in the development of the proliferation, as has actually been shown for the TCRαβ+CD4+ T-LGL leukemia type [10]. Also, in 2006 Sandberg *et al*. identified the presence of the so-called invariant T selection determinant leading to a conserved amino acid at the relative second position of the CDR3 region in Vδ2 – Jδ1 rearrangements in a large subgroup of TCRγδ+ T-LGL proliferations, indicating that these cells are antigen-experienced. Furthermore, an activation-associated effector phenotype and a skewed and dynamic TCR repertoire – also referred to as clonal drift [11] – have been found in T-LGL proliferations, again suggestive of a role for antigens, but the exact antigen or the type of antigens have so far not been elucidated [3,12]. Even though the involved antigen(s) are unknown and might be variable, and the disease is clinically heterogeneous, we hypothesized that TCRγδ+ T-LGL leukemia patients have common underlying molecular defects that contribute to the leukemogenesis. In order to obtain more in-depth insights into the mechanisms driving these TCRγδ+ T-LGL

proliferations we therefore performed gene expression profiling analysis in purified TCRγδ+ T-LGL leukemia samples.

Here we present data that TCRγδ+ T-LGL leukemia cells originate from the most common antigen-experienced TCRγδ+ T cell population in the adult peripheral blood, and that they have undergone a transformation leading to an imbalance in proliferation and apoptosis, eventually contributing to the TCRγδ+ T-LGL leukemia pathogenesis.

MATERIALS AND METHODS

Patients and healthy controls

The database files from the department of Immunology, Erasmus MC, University Medical Center (Rotterdam, The Netherlands) were retrospectively reviewed for cases with a proven mature persistent (>6 months) TCRγδ+ T cell proliferation in PB and/or BM based on a combination of clinical, histological (HE sections), cytomorphological (May-Grünwald-Giemsa staining), laboratory, immunophenotypical (including for the majority of cases most of the following markers: CD3, TCRγδ, CD4, CD8, CD16, CD56, CD57, CD45RA, CD45RO, CD27, CD197) and molecular data (monoclonal TRG and TRD gene rearrangements) [13,14]. Other TCRγδ+ T cell proliferation diseases such as hepatosplenic lymphomas were excluded based on clinical, cytomorphological and/or histopathological data. From 10 TCRγδ+ T-LGL leukemia cases frozen cell material was available for inclusion in the current study. The immunophenotypical and molecular analyses were performed on PB and/or BM samples which were erythrocyte-lysed using FACS lysing solution (BD Biosciences, San Jose, CA, USA). PB- or BMMC were isolated by means of Ficoll-Paque (density 1.077 g/ml, Pharmacia, Uppsala, Sweden). As controls, healthy blood donors from Sanquin Blood Bank (Amsterdam, The Netherlands) were included upon informed consent, which were anonymized for further use. PBMCs were obtained with Ficoll density gradient separation and cryopreserved in Iscove's Modified Dulbecco's Medium (IMDM, Lonza, Basel, Switzerland) with dimethyl sulfoxide and stored in vials at -180 °C until further use.

Use of all samples for the study was approved by the Erasmus MC Medical Ethics Committee (MEC-2015-617). Studies were conducted in accordance with the principles of the Declaration of Helsinki.

Immunophenotyping and cell sorting

Control PB samples were immunophenotyped to examine the composition of TCRγδ+ T cell subsets based on membrane marker expression, including CD3, CD4, CD8, CD19, CD27, CD45RA, CD45RO, CD27, CD197, TCRαβ, TCRVδ1 and TCRVδ2 (Supplementary

Table 1). Measurements were performed with a FACS Canto II or FACS LSR Fortessa flow cytometer (BD Biosciences). The immunophenotype of the LGL proliferations was already determined during diagnostic work-up, and was therefore available in the database.

For sorting experiments cryopreserved PB- and/or BMMCs were thawed and sorted. Sorting was performed specifically on the tumor cells from the patient samples, using CD3, CD45, TCRαβ, TCRγδ, TCRVδ1 and TCRVδ2 markers.

Healthy control samples were sorted into five different CD3+TCRγδ+ T cell subsets based on gating strategies in Supplementary Fig. 1: total TCRVδ1, total TCRVδ2, effector memory (=TemRO population, CD45RO+CD27-CD197-), effector (=TemRA population, CD45RA+CD27-CD197-) subsets and a naive, non-antigen stimulated control (CD45RA+CD27+CD197+) subset. Sorting was performed with a FACS Aria I or III (BD Biosciences).

RNA isolation, cDNA synthesis

After sorting the cells were lysed and subjected to combined DNA/ RNA isolation with the QIAGEN DNA/RNA/miRNA AllPrepKit (QIAGEN, Hilden, Germany).

For the gene expression profiling experiments RNA from the sorted healthy control TCRγδ+ T cell subsets was pooled in order to obtain higher amounts of RNA and to create pooled healthy control subset samples; N=3 for Vδ1 and Vδ2 subsets, N=8 for naive, TemRO and TemRA subsets due to lower sorting yields.

For RQ-PCR tests cDNA was synthesized from isolated RNA as well as cryopreserved RNA with reverse transcriptase Superscript II (Invitrogen Life Technologies, Waltham, MA, USA), 10x CA buffer (0.2 M Tris pH 8.3, 0.5M KCl), dNTP (GE Healthcare, Cleveland, OH, USA), dithiotreitol (Invitrogen Life Technologies), $MgCl_2$ (Applied Biosystems Life Technologies, Waltham, MA, USA), recombinant RNAsin (Promega, Fitchburg, WI, USA) and random primers (Invitrogen Life Technologies).

Gene expression profiling

Isolated RNA from sorted patient tumor cells and the 5 control TCRγδ+ T cell subsets was further amplified, reverse transcribed into cDNA, purified, fragmented, biotinylated and hybridized to Affymetrix HG-U133 Plus 2.0 GeneChip arrays (containing 54,675 probe sets) according to the Affymetrix GeneChip 3' IVT Express Kit user manual (Affymetrix, Santa Clara, CA, USA), as described previously [15-17]. Robust multi-array average (RMA) background removal, compensation for systematic technical differences, quantile normalization and probe set summary were performed [18]. Unsupervised correlation and clustering plots were made based on selected probe sets that showed signal, i.e. for which the median absolute deviation (MAD) from the median exceeded a

certain threshold *Thr* on a $\log_2$ scale. To further assess differential expression, a number of supervised analyses were applied: fold change calculation (FC), analysis of variance (ANOVA) [19] and significance analysis of microarrays (SAM) were applied [20]. All comparisons concerned disease cases versus normal control cases. Data were analyzed through the use of the Database for Annotation, Visualization, and Integrated Discovery (DAVID Database) [21,22], and QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN, Redwood City, USA; www.qiagen.com/ingenuity).

Real-time quantitative PCR (RQ-PCR)

Assays were designed with the Roche Universal Probe Library (Roche, Basel, Switzerland) (Supplementary Table 2). RQ-PCR experiments were performed with TaqMan Universal PCR master mix (2x) (Applied Biosystems) on the StepOnePlus instrument (Thermo Fisher, Waltham, MA, USA). Ct values of disease samples were normalized to the Ct value of the ABL housekeeping gene [23] and normalized Ct values of healthy control samples (ΔΔCt method).

Statistical analysis

Supervised statistical analyses for the gene expression profiling data were performed using analysis of variance (ANOVA), with significance cutoff at $p<0.05$ and as threshold of up- and/or down-regulation ≥2 fold change. Significance of microarray analysis (SAM) with significance cutoff <0.05 was included in the analysis as comparison. Bonferroni and Benjamini-Hochberg p-value adjustment were applied for multiple testing correction and comparison of different statistical tests on the data set.

RESULTS

Heterogeneity in clinical presentation, associated diseases and immunophenotype and –genotype among TCRγδ+ T-LGL leukemia patients

In all ten patients a proven TCRγδ+ T-LGL leukemia was identified, based on immunophenotype, immunogenotype, high leukocyte count, and high percentage of TCRγδ+ T-LGL cells compared with reference values of PB TCRγδ+ T cells (<5% of all lymphocytes in general) [6]. All cases were diagnosed as TCRγδ+ T-LGL leukemia, while other causes of TCRγδ+ T cell proliferation, such as hepatosplenic lymphoma, were excluded. There was a slight male predominance among the patients (seven males, three females). Most patients showed chronic leukemia-associated symptoms, such as fever and cytopenia, and underlying (auto)immune diseases such as rheumatoid arthritis, Graves' disease and

uveitis (Table 1). In general, immunophenotyping showed CD3-positivity, CD4-negativity and in some cases CD8-positivity, together with variable expression of markers that have been associated with LGL cells such as CD16, CD56 and/or CD57, and frequent CD45RA positivity in combination with CD27 and/or CD197 negativity, pointing towards TemRA and TemRO phenotypes (Table 1). For some patients Vγ and Vδ-usage was determined with flow cytometry as well, which correlated with the immunogenotyping data (Table 1). Among the patients half showed Vδ1 – Jδ1 rearrangements, whereas the other half showed Vδ2 – Jδ1 rearrangements. Most patients had a Vγ9 – Jγ1.1/2.1 or Jγ1.3/2.3 rearrangement, leading to an overall receptor predominance of Vγ9/Vδ1 and Vγ9/Vδ2. Most patients showed a monoclonal TCRγδ+ T-LGL cell population at the receptor level, whereas two patients showed a somewhat oligoclonal profile albeit with a dominant clone, possibly indicating the presence of additional non-aberrant TCRγδ+ T cells or subclones (Table 1).

Unsupervised clustering of TCRγδ+ T-LGL leukemia cases shows highest resemblance to normal effector TCRγδ+ T cells

Despite some level of phenotypical heterogeneity between patients, we next evaluated the transcriptomes of 10 TCRγδ+ T-LGL leukemia patients versus 5 healthy control TCRγδ+ T cell subsets in order to obtain more insight into the immunobiology of the disease and to possibly identify common downstream pathogenic events in TCRγδ+ T-LGL leukemia. As from literature it is known that TCRγδ+ T-LGL leukemias are antigen-experienced [4,6], TemRO and TemRA TCRγδ+ subsets were used as healthy control counterparts for the TCRγδ+ T-LGL cells. As a non-antigen stimulated control the naive TCRγδ+ population was included. In order to get a general view from gene expression profiling, first a correlation analysis was performed based on selected probe sets that showed significant variation between all microarrays (with a median absolute deviation from the median (MAD) of at least *Thr* = 0.7). In the heatmap TCRγδ+ T-LGL leukemia cases and healthy control TCRγδ+ T cell subsets were on purpose represented separately, as they represent distinct groups. The heatmap showed a clear distinction between TCRγδ+ T-LGL leukemia cases and healthy control TCRγδ+ T cell subsets, suggesting that the TCRγδ+ T-LGL leukemia cases indeed form a separate group with a distinct transcriptome (Fig. 1a). The healthy control subset that correlated best with the TCRγδ+ T-LGL leukemia cases was the TemRA TCRγδ+ T cell subset, and to a lesser extent the TemRO subset, in line with our patient phenotype data (Table 1), whereas the naive TCRγδ+ T cell subsets showed the lowest correlation, as was expected (Fig. 1b). Hierarchical clustering confirmed that 8/10 TCRγδ+ T-LGL leukemia cases indeed formed a separate group from the healthy control subsets, whilst two cases, LGL087 and LGL113, grouped together with TemRA, TemRO and TCRVδ2 subsets (Fig. 1c). The same

Table 1. TCRγδ+ T-LGL leukemia patient characteristics and immunogenotypic features.

	Phenotype						Genotype						
							TRD rearrangement‡			TRG rearrangement‡			
Patient	**Sex**	**Age at diagnosis (years)**	**Clinical presentation and associated disease**	**Immunophenotype**	**Tumor load (%leukocytes)**	**Absolute LGL count (10^9/ml)**	**TRDV**	**TRDD**	**TRDJ**	**TRGV**	**TRGJ+**	**Overall receptor**	**Clonalityx**
LGL056	F	40	Anemia, neutropenia, M. Graves	CD3+/CD4-/CD8+/CD16+/ CD56+/ CD57+/ CD45RA+/CD45RO-/Vγ9+/Vδ1+	41	3.6	1*01	2*01/ 3*01	1*01	9*01	1*02	Vγ9/Vδ1	Monoclonal
LGL057	M	46	Fever, positive Mantoux test	CD3+/CD4+/CD8+/CD56+/Vγ9+/Vδ2+	22.7	0.8-1.2	2*01/ 02/03	3*01	1*01	9*01	P1*01	Vγ9/Vδ2	Monoclonal
LGL058	M	38	Lymphadenopathy, uveitis, sarcoidosis	CD3+/CD4/CD16+/CD56+/CD57+/CD45RA+/CD27-/Vδ2+	6.9	3.4	2*01/ 02/03	3*01	1*01	9*01	1*02	Vγ9/Vδ2	Monoclonal
LGL063	M	56	Unknown	CD3+/CD4-/CD8+/CD16+/CD56+/CD57+/CD27-/CD197-/Vδ1+	23	1.4	1*01	2*01/ 3*01	1*01	9*01	P1*01	Vγ9/Vδ1	Oligoclonal
LGL064	M	76	Anemia, thrombocytopenia, hepatosplenomegaly, rheumatoid arthritis	CD3+/CD4+/CD8+/CD16+/CD56+/CD57-/CD45RA+/CD45RO+/CD27+/Vδ1+/Vδ2-	24	1.7	1*01	2*01/ 3*01	1*01	4*01	1*01/02/ 2*01	Vγ4/Vδ1	Monoclonal
LGL083	F	30	Unknown	CD3+/CD4-/CD8+/CD16+/CD57+/CD45RA+/CD27-/CD197-/Vγ9-/Vδ2+	8	0.3-2.4	2*01/ 02/03	3*01	1*01	8*01	2*01	Vγ8/Vδ2	Oligoclonal
LGL087	F	74	Unknown	CD3+/CD4-/CD8+/CD56+/Vδ1+	28	n.d.	1*01	2*01/ 3*01	1*01	2*02	1*01	Vγ2/Vδ1	Monoclonal
LGL088	M	54	Unknown	CD3+/CD4-/CD8-/CD16-/CD56+/CD57+/CD45RA+/CD27-/CD197-/Vδ1+	26	0.8	1*01	3*01	1*01	2*01	P2*01	Vγ2/Vδ1	Monoclonal
LGL089	M	54	Unknown	CD3+/CD4-/CD8+/CD16+/CD56+/CD57+/CD45RA+/CD27-/CD197-/Vγ9+/Vδ2+	10	4.3	2*02	2*01/ 3*01	1*01	9*01	P1*01	Vγ9/Vδ2	Monoclonal
LGL113	M	70	Unknown	CD3+/CD4-/CD8-/CD16-/CD56-/CD57-/CD45RA+/CD45RO+/CD27-/CD197-/Vγ9+/Vδ2+	51	6	2*01/ 02/03	2*01/ 3*01	1*01	9*01	1*01/02/ 2*01	Vγ9/Vδ2	Monoclonal

‡Productive and in-frame rearrangements are shown. N.d., not determined.
+TRGJ annotations according to the IMGT nomenclature [24]. TRGJP1: Jγ1.1, TRGJP2: Jγ2.1, TRGJ1: Jγ1.3, TRGJ2: Jγ2.3. The canonical TRGJP (Jγ1.2) was not included in the diagnostic work-up with IVS TRG multiplex PCR (Invivoscribe).
xBased on GeneScan and heteroduplex analysis after multiplex PCR.

clustering was observed when a principal component analysis (PCA) was performed (Fig. 1d). One case was studied in duplo (LGL058, Fig. 1) to assess reproducibility of the arrays, which was indeed high. Overall the TCRγδ+ T-LGL leukemia cases thus mostly form a separate group based on their gene expression profile, showing closest correlation with the healthy normal TemRA TCRγδ+ T cell subset.

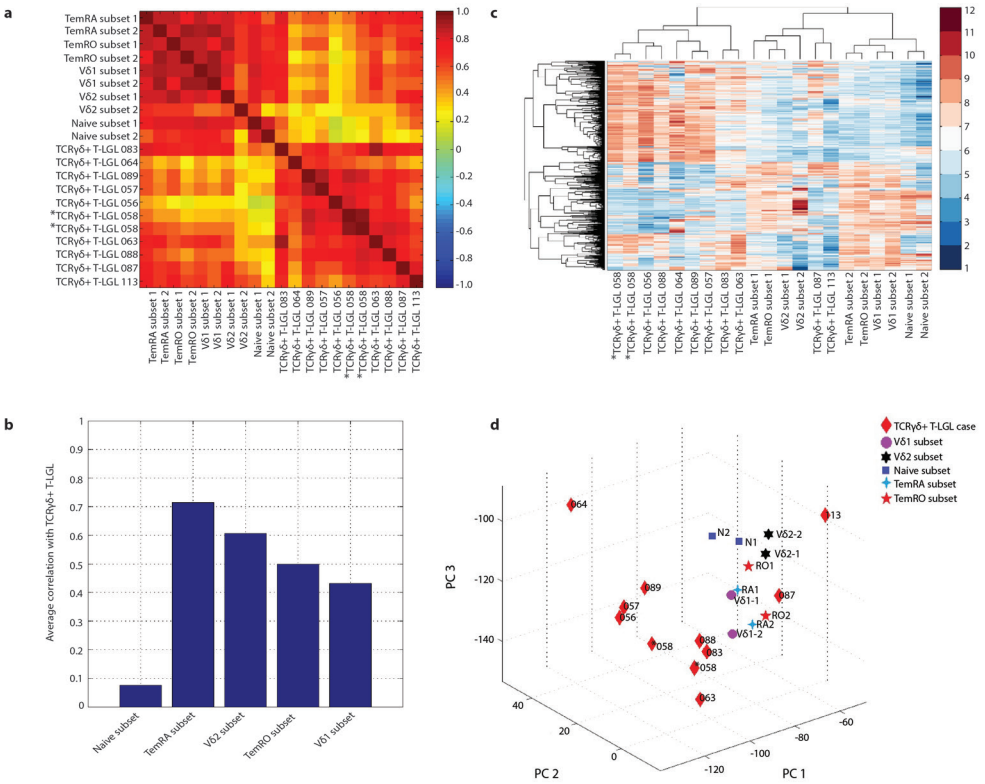


Figure 1. Unsupervised gene expression data analysis.
(**a**) Heatmap analysis shows clear correlations between TCRγδ+ T-LGL proliferation cases, and lower correlations to the healthy control TCRγδ+ T cell subsets, (**b**) the highest being to the TemRA TCRγδ+ T cell subset. (**c**) The clustergram shows that the TCRγδ+ T-LGL proliferation cases indeed form a distinct group, except for two cases which cluster closer to healthy control TemRA, TemRO and TCRVδ2 TCRγδ+ T cell subsets. (**d**) Principal component analysis (PCA) in three-dimensional (3D) graph shows similar clustering results. *Thr*=0.7; 1336 probe sets. Case LGL058 indicated in italics was studied in duplo to assess reproducibility of the assay.

Supervised analysis showed a dysregulated balance in proliferation and apoptosis

Following unsupervised clustering analysis, specific comparisons were performed between all TCRγδ+ T-LGL leukemia cases vs. naive, TemRO and TemRA healthy control TCRγδ+ T cell subsets. Two statistical comparisons, significance analysis of microarrays (SAM) and two-way analysis of variance (ANOVA) were applied that yielded varying numbers of differentially expressed probe sets (Supplementary Table 3). Stringent statistical filtering using SAM showed low numbers of mainly down-regulated genes in the comparison between TCRγδ+ T-LGL leukemia cases and TemRO and TemRA subsets. When the comparison was performed with the naive subset, the numbers of differentially expressed probe sets found by both statistical analysis methods were high. Both methods confirmed the relatively high similarity of TCRγδ+ T-LGL leukemia cases to the healthy TemRO and TemRA subsets, and the lower similarity to naive TCRγδ+ T cells as also observed in the unsupervised clustering. Since the SAM analysis showed low numbers of differentially expressed probe sets, and in order to reduce the risk of missing data, we therefore used the two-way ANOVA set for further supervised analysis. Top up- and down-regulated genes in TCRγδ+ T-LGL leukemia cells vs. the best correlating subset, the TemRA TCRγδ+ T cells, are depicted in Supplementary Table 4. Most up-regulated genes appeared to be involved in the inflammatory response and response to bacteria: *LYZ*, *S100A8*, *S100A9*, *ALOX5* and *LTB* according to the DAVID database. The top down-regulated genes were mostly associated with the process of transcription; these genes included amongst others *GTF2H3*, *CTBP2* and *ZNF260*.

In order to identify genes associated with biologically more relevant pathways and processes, the probe sets that were significantly differentially expressed with a fold change of at least 2 (both up- and down-regulated) in TCRγδ+ T-LGL leukemia cells when compared to the healthy TemRA TCRγδ+ T cell subset were then further analyzed with the DAVID database [21,22]. Gene Ontology biological process analysis showed processes involved in the regulation of the immune system and response, T cell activation, and response to stress to be affected (Table 2). Enrichment analysis of KEGG pathways in DAVID showed that differentially expressed genes were largely involved in the hematopoietic cell lineage, osteoclast differentiation, rheumatoid arthritis disease and NFκB signaling, although after correction for multiple testing none of these were found to be significant. Nevertheless, similar processes, such as hematopoietic system and inflammation, were identified (Table 3). Differentially expressed probe sets in the comparison between TCRγδ+ T-LGL leukemia cells vs. TemRO cells showed similar processes to be affected (Supplementary Table 5). Taken together, both annotation enrichment analyses (Gene Ontology, KEGG) indicated that proliferation, cell survival signaling and inflammatory processes were affected.

Table 2. Gene Ontology biological process annotation analysis of TCRγδ+ T-LGL leukemia cases versus healthy TCRγδ+ TemRA cells in DAVID.

Term	Gene count*	Genes	p-value	Bonferroni**	Benjamini**
Immune response	138	A.o.: BTK, CCL4, **CCR7**, CXCL16, CXCL8, CXCR5, **CX3CR1**, CLEC7A, CD14, CD1D, CD226, **CD28**, CD36, CD3D, CD58, CD79A, CD86, CD8B, CRK, FYN, FAS, FCER genes , JAK3, MYB, NLRP3, RAB29, REL, S100 genes, TAB2, **XIAP**, XAF1, AIF1, APOBEC genes, BTN3A3, CAMK4, CR1, CR2, C5AR1, CFP, CYBB, DLL1, EGR1, FCN1, FOXP1, HSP genes, HAVCR2, IFIT, IRF, IRAK2, IL23A, IL27RA, IL6R, KLRD1, LILRA genes, LY86, LY96, LEF1, LTB, HLA-DRA, HLA-DRB4, MAPK1, MAP3K1, NCF, NFKB, PRF1, PIK3R1, PF4, PARP9, STAT1, TLR2, TFEB, TRIM genes, TNFSF13B, TNFSF8	3.9E-19	2.5E-15	2.5E-15
Defense response	129	A.o.: ALS2, ANKHD, APOBEC genes, BTK, **CCL4**, CCR7, CXCL16, **CX3CR1**, CLEC7A, CD14, CD1D, CD226, **CD28**, CD36, CD58, CD86, CD8B, DDX3X, DDX60, FYN, FAS, FCER genes, IFI30, JAK3, KLF4, NLRP3, PRDM1, REL, S100 genes, TAB2, **XIAP**, XAF1, AIF1, ALOX5, CAMK4, **CASP1**, CSF3R, CR1, CR2, C5AR1, CFD, CFP, CST3, CYBB, EGR1, FCN1, FOXP1, FPR1, GFI1, HSP genes, HAVCR2, IG genes, IFNGR2, **IFNG**, IFIT genes, IRF genes, IL23A, IL27RA, IL6R, KLRD1, LILRA genes, LY genes, LYZ, HLA-DRA, HLA-DRB4, NCF genes, NFKB genes, PRF1, STAT1, TLR2	5.3E-13	3.4E-9	1.1E-9
Cell activation	87	A.o.: ABAT, ADAM10, BTK, **CCR7**, CXCL8, CXCR5, **CX3CR1**, CLEC7A, CD1D, CD2, CD226, **CD28**, CD3D, CD5, CD79A, CD86, CD8B, LIG4, FYN, FAS, FCER genes, JAK3, MAFB, MYB, NLRP3, SOX4, CAMK4, CR2, CST3, DLL1, EGR1, FOXP1, HAVCR2, IRS2, **IFNG**, IRF1, IL23A, IL27RA, IL6ST, LEF1, HLA-DRA, HLA-DRB4, PRF1, PIK3R1, PF4, SLAMF1, TLR2, TNFSF13B, TNFSF8, VCL	4.9E-12	3.2E-8	6.3E-9
Regulation of immune system process	116	A.o.: ADAM10, BTK, CCL4, **CCR7**, CXCL8, CLEC7A, CD14, CD1D, CD2, CD226, **CD28**, CD36, CD3D, CD5, CD79A, CD86, CD8B, CRK, DDX60, FYN, FANCL, FAS, FCER genes, JAK3, MAFB, MYB, RAB29, TAB2, **XIAP**, AIF1, CSF3R, CR1, CR2, CFD, CFP, DLL1, FCN1, FOXP1, FPR1, HSP genes, HAVCR2, IFNGR2, **IFNG**, IFIT1, IRF1, IRAK2, IL23A, IL27RA, IL6R, KLRD1, LILRB genes, LEF1, HLA-DRA, HLA-DRB4, MAPK1, MAP3K1, NFKB1, PIK3R1, PARP9, SLAMF1, TLR2, TNFSF13B, VEGFA	2.5E-11	1.6E-7	2.3E-8
Leukocyte activation	74	A.o.: ADAM10, BTK, **CCR7**, CXCL8, CXCR5, **CX3CR1**, CLEC7A, CD1D, CD2, CD226, **CD28**, CD3D, CD5, CD79A, CD86, CD8B, LIG4, FYN, FAS, JAK3, MAFB, MYB, RAB29, AIF1, SOX4, DLL1, FOXP1, HAVCR2, IRS2, **IFNG**, IRF1, IL23A, IL27RA, IL6ST, LEF1, HLA-DRA, HLA-DRB4, PRF1, PIK3R1,	3.9E-11	2.5E-7	3.2E-8
Response to stress	238	A.o.: ABAT, ALS2, APC, APOBEC genes, BTK, BCLAF1, CCL4, **CCR7**, CXCL16, CXCL8, **CX3CR1**, CLEC7A, CFLAR, CD14, CD1D, CD226, **CD28**, CD36, CD58, CD86, CD8B, DDX3X, DDX60, LIG4, POLH, FYN, FANCL, FANCM, FAS, JAK3, KLF4, MAFF, NLRP3, PRDM1, REL, S100 genes, SOX4, TIAM1, TAB2, **XIAP**, XAF1, ALOX5, CASP1, CDC7, CSF3R, CR1, CR2, C5AR1, CFD, CFP, EGR1, FCN1, FOXP1, FPR1, HSP genes, HAVCR2, **ID3**, IFNGR2, **IFNG**, IFIT genes, IRF1, IRF2, IL23A, IL27RA, IL6ST, LILRA genes, LYZ, HLA-DRA, HLA-DRB4, NCF genes, NFKB genes, PARP9, STAT1, SLAMF1, PIK3R1, PRF1, USP genes, VEGFA, TNFSF8	6.0E-11	3.8E-7	44.3E-8
Cytokine production	65	A.o.: BTK, **CCR7**, CD14, CD2, CD226, **CD28**, CD36, CD58, CD86, DDX3X, DDX60, JAK3, KLF4, MAF, NLRP3, REL, S100 genes, TIA1, CAMK4, **CASP1**, CYBB, DLL1, EGR1, FCN1, FOXP1, GBP1, HSP genes, HAVCR2, IFI16, **IFNG**, IRF1, IL23A, IL27RA, IL6R, IL6ST, LEF1, LY96, LTB, NFKB1, NFKB2, HLA-DRB4, PF4, SLAMF1, TLR2	2.3E-10	1.5E-6	1.5E-7
T cell activation	50	A.o.: **CCR7**, CLEC7A, CD1D, CD2, **CD28**, CD3D, CD5, CD85, CD8B, LIG4,FYN, FAS, FCER1G, JAK3, MAFB, MYB, NLRP3, PRDM1, RAB29, SOX4, AIF1, BATF1, CAMK4, EGR1, FOXP1, HAVCR2, ITPKB, **IFNG**, IFR1, IL23A, IL6ST, LEF1, HLA-DRA, HLA-DRB4, PIK3R1, PRNP, RHOH, SLAMF1, TNFSF13B, TNFSF8	3.6E-9	2.3E-5	1.2E-6

*Total 1024 differentially expressed genes in LGL versus TemRA dataset with FC=2 both up- and down-regulated, p<0.05 (ANOVA), of which 805 were annotated by DAVID using Affymetrix Human Genome U133 Plus 2.0 array as background and selecting Homo Sapiens as species.

**Adjusted p-value based on Bonferroni and Benjamini-Hochberg correction for multiple testing.

Genes also identified through IPA analyses, which are further validated with RQ-PCR are indicated in bold.

For further interpretation and visualization of differentially expressed genes, Ingenuity Pathway Analysis (IPA) was used. First, TCRγδ+ T-LGL leukemia cases were compared to healthy TemRA TCRγδ+ T cells, showing a number of processes which were up-regulated in the TCRγδ+ T-LGL leukemia cases, such as hematological development and function. However, cell death and survival processes appeared to be significantly down-regulated (Fig. 2a). Again, similar results were obtained after analyzing the

Table 3. KEGG enrichment pathway analysis of TCRγδ+ T-LGL leukemia cases versus healthy TCRγδ+ TemRA cells in DAVID.

Term	Gene count	Genes	p-value	Bonferroni*	Benjamini*
Hematopoietic cell lineage	16	CD14, CD1d, CD2, CD36, CD3d, CD5, CD8b, ANPEP, CSF3R, CR1, CR2, ITGA6, IL6R, HLA-DRA, HLA-DRB, MS4A1	1.3E-5	3.3E-3	3.3E-3
Osteoclast differentiation	19	BTK, FYN, TAB2, CAMK4, CYBB, IFNGR2, **IFNG**, LILRA/LILRB genes, NCF1, NCF2, NFKB1, NFKB2, PIK3R1, SIRPA, STAT1	9.9E-5	2.6E-2	1.3E-2
Rheumatoid arthritis	13	ATP6V genes, CXCL8, **CD28**, CD86, **IFNG**, IL23A, LTB, HLA-DRA, HLA-DRB4, TLR2, TNFSF13B, VEGFA	8.1E-4	1.9E-1	5.2E-2
NF-kappa B signaling pathway	13	BTK, CCL4, CXCL8, **CFLAR**, CD14, TAB2, **XIAP**, LY96, LTB, NFKB1, NFKB2, PIAS4, TNFSF13B	1.0E-2	2.3E-1	5.1E-2

*Adjusted p-value based on Bonferroni and Benjamini-Hochberg corrections for multiple testing.
Genes also identified through IPA analyses, which are further validated with RQ-PCR, are indicated in bold.

differentially expressed probe sets between TCRγδ+ T-LGL leukemia cases and the healthy TemRO subset (Fig. 2b), confirming comparable correlation of both subsets with the TCRγδ+ T-LGL leukemias as described from the unsupervised analysis (Fig. 1).

After identifying affected processes we focused more in-depth on specific genes. Genes involved in apoptosis, such as *BCLAF1*, *CASP1*, *XIAP* and *CFLAR,* appeared to be the most relevant candidate genes. The apoptosis-inducing genes *CASP1* and *CFLAR* were down-regulated, as well as the apoptosis inhibitor *XIAP*, while *BCLAF1* was up-regulated. Furthermore, transcription factors such as *ID3*, *KLF4*, *LEF1* and *SOX4* in the up-regulated cell survival process were all found to be up-regulated. Finally, altered expressions of *CD28*, *CCR7*, *CX3CR1*, *IFNG*, *LTB* and *PRF1* were all contributing to the inflammatory profile of the TCRγδ+ T-LGL leukemias.

Collectively, data from all analysis methods (DAVID Gene Ontology, DAVID KEGG, IPA) thus showed that proliferation and cell survival signaling were up-regulated, while apoptosis was down-regulated in TCRγδ+ T-LGL leukemia cells when compared to normal TCRγδ+ T cell subsets.

RQ-PCR validation indicates possible signature genes within a highly heterogeneous disease profile

Next, representative candidate genes, which were found in the comparison between TCRγδ+ T-LGL leukemia cells and healthy TemRA and TemRO TCRγδ+ T cells in both IPA and DAVID analysis, and which appeared biologically relevant for the disease, were selected. These included genes involved in cell death and survival processes (*BCLAF1*, *CFLAR*, *CASP1,* and *XIAP*), genes associated with proliferation-driving transcription or

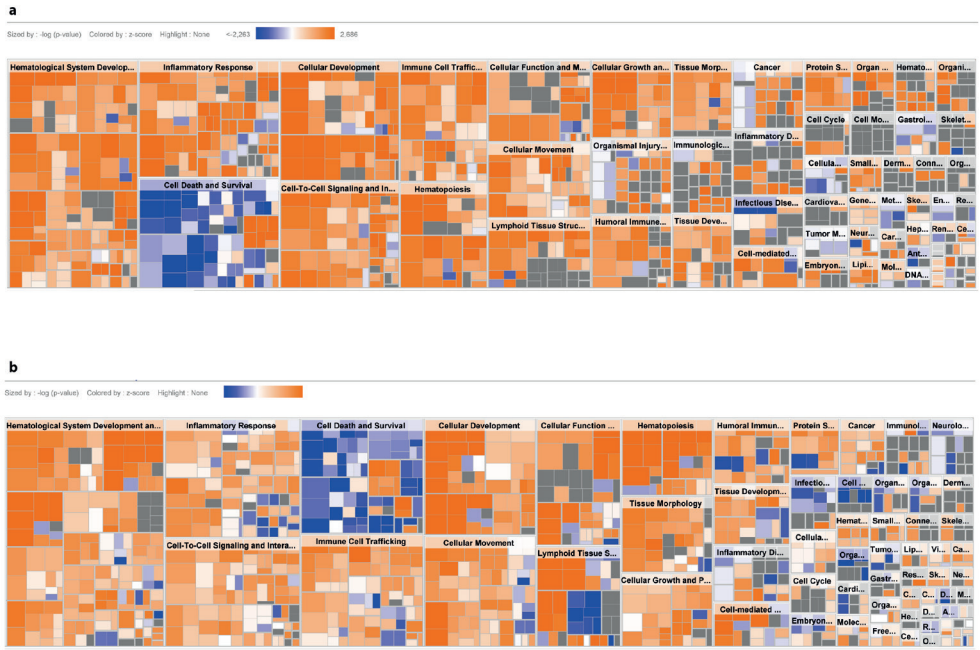


Figure 2. Functional annotation of genes differentially expressed between TCRγδ+ T-LGL leukemia cases and healthy control TCRγδ+ T cell subsets.
(**a**) Comparison of TCRγδ+ T-LGL leukemia cases versus healthy control TemRA and (**b**) TemRO TCRγδ+ T cell subsets showing similar patterns with down-regulation of apoptosis and up-regulation of cancer-related processes. Sizing based on significance level (ANOVA), up- and down-regulation based on z-scoring. Plots were generated in QIAGEN's Ingenuity Pathway Analysis.

cancer (*KLF4*, *LEF1*, *SOX4*, *ID3*), and genes associated with inflammation (*CD28*, *CCR7*, *CX3CR1* and *IFNG*). Differences in gene expression levels were analyzed by means of real-time quantitative (RQ)-PCR using Universal Probe Library reagents. Sorted healthy control TemRA TCRγδ+ T cells were compared with the sorted TCRγδ+ T-LGL leukemia cells from patients (both from patients included in the microarray analysis, and from new patients). In order to validate the fold changes obtained from microarray analysis, expression ratios between patients and healthy controls were calculated from the RQ-PCR data. All selected apoptosis-related genes from our microarray analysis, were confirmed to be differentially expressed in TCRγδ+ T-LGL leukemia cases when compared to healthy controls; *XIAP, CASP1, BCLAF1* were all validated to be over two-fold higher, and especially *CFLAR* to an even higher extent than observed in the microarrays (Fig. 3). Transcription factor *ID3* could also be validated in RQ-PCR; the ratio between TCRγδ+ T-LGL leukemia patients and healthy controls was even higher than that observed in the microarrays. Other transcription factors identified from the microarray analysis (*KLF4*, *LEF1* and *SOX4*) could not be confirmed by RQ-PCR validation (Supplementary

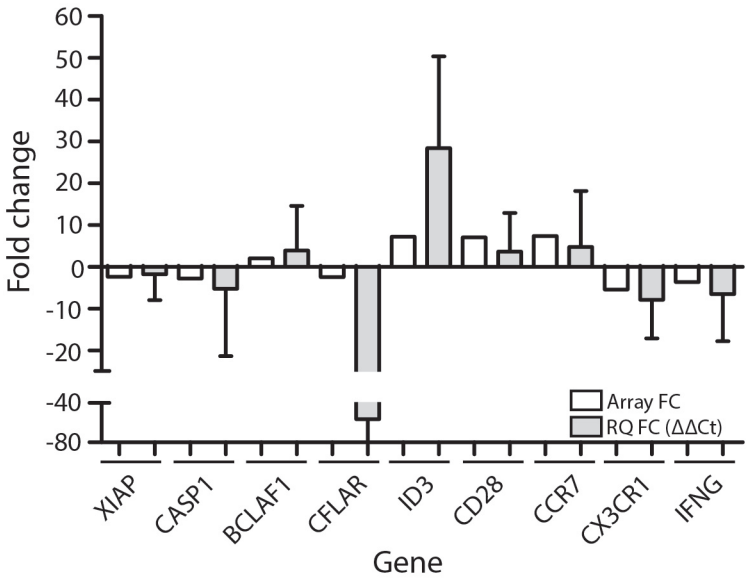


Figure 3. RQ-PCR validation of most representative genes.
Fold changes of representative genes identified through gene expression profiling and by RQ-PCR. Relative expression of genes found differentially expressed using microarrays was first normalized to ABL housekeeping gene (ΔCt). Average ΔCt values from healthy controls (N=6) were used to calculate patient (N=10) to healthy control ratios per patient to obtain fold change values (ΔΔCt). White bars depict fold change values obtained from microarray data, grey bars depict average ΔΔCt values after RQ-PCR validation. For RQ-PCR validation 4 patients from the original microarray dataset were used and 6 novel patients. Mean expression fold change values are indicated with the standard deviation.

Table 6). *CD28*, *CCR7* and *CX3CR1* were validated as well, albeit with a lower fold change than in the microarrays. *IFNG* was validated with an approximately two-fold higher fold change in RQ-PCR than in the microarrays (Fig. 3). Of note, considerable heterogeneity in RQ-PCR levels was observed within the samples used for validation (Supplementary Table 6), thus reflecting the heterogeneity of the disease.

Taken together, we were able to identify possible candidate genes to distinguish aberrant from normal TCRγδ+ T cells, although the known largely heterogeneous profile of the disease remains noticeable, which may partly mask differences between leukemic TCRγδ+ T-LGL cells and normal TCRγδ+ T cell subsets.

DISCUSSION AND CONCLUSION

T cell large granular lymphocyte disorders form a large group of heterogeneous mature T cell neoplasms, of which the TCRγδ+ T cell variant is rare and most poorly understood. TCRγδ+ T-LGL leukemia is known to have a slowly progressive indolent

disease course, mostly affecting elderly individuals, while showing similar (initial) clinical presentations such as neutropenia, anemia and pancytopenia [4,6,25]. Most of our patients displayed malignancy, autoimmunity or cytopenia as underlying or associated diseases, as described previously [26], but there was no consistent pattern. Furthermore, immunophenotypic features did not correlate with the symptomatic status or underlying diseases. Underlying (chronic) diseases could play a role in the pathogenesis of TCRγδ+ T-LGL leukemia, possibly as a (chronic) stimulus that causes the TCRγδ+ T cells to expand, in line with the idea that TCRγδ+ T-LGL leukemia develops from dominant normal TCRγδ+ T cell populations in the adult peripheral blood. However, underlying molecular mechanisms in TCRγδ+ T-LGL leukemogenesis are largely unknown, and a common underlying molecular defect has not been described earlier.

TCRγδ+ T-LGL leukemias are characterized by skewed receptor expression and an antigen stimulation-associated activation status, but are also considered to harbour particular changes in gene expression. In our current study we aimed to address this by means of gene expression profiling, which revealed that the TCRγδ+ T-LGL leukemia cells displayed different gene expression profiles when compared to healthy, polyclonally expanded TCRγδ+ T cells. Initial unsupervised clustering based on microarray data from aberrant TCRγδ+ T cells and different healthy TCRγδ+ T cell subsets, showed a clear distinction between the TCRγδ+ T-LGL leukemia and its healthy counterparts. Even though TCRγδ+ T-LGL leukemia forms a separate group, with the lowest correlation with the naive subset as expected, the correlation with the TemRA subset was quite high, suggesting that it is derived from the most predominant and antigen-experienced type of TCRγδ+ T cell populations in adult peripheral blood, in line with earlier published data [27]. When focusing more in-depth on differences in comparison with normal TemRA TCRγδ+ T cells it became clear that particular biological functions, processes and genes were affected. Most relevant genes involved processes such as proliferation, stimulation and apoptosis, with an increase in proliferation and cell survival, and a decrease in apoptosis. One clear example concerned transcription factor ID3, which is involved in proliferation and haematological development, a.o. by enhancing TCRγδ+ T cell development [28]. Up-regulation in TCRγδ+ T-LGL leukemia cases could be confirmed with RQ-PCR. As opposed to proliferation genes, some apoptosis-inducing genes such as CASP1 and CFLAR were down-regulated. Caspase-1 is an apoptosis-related cysteine peptidase, which is involved in inflammation and apoptosis by proteolysis of pro-inflammatory cytokines and activating other caspases and pro-apoptotic proteins [29,30]. CFLAR has been described in multiple studies as being important in the development of cancer, and is therefore also being described as a target for therapy (reviewed by Fulda in 2013 [31]); it induces apoptosis [32], but is also a key role player in autophagy and necroptosis [33]. XIAP, which normally regulates and inhibits apoptosis [34] was also shown

to be down-regulated, resulting in a net effect of inhibited apoptosis. The up-regulated apoptosis-related gene BCLAF1 induces apoptosis and represses transcription [35]. Our RQ-PCR validation data also showed up-regulation of BCLAF1, and down-regulation of CFLAR and CASP1. All apoptosis-related genes could be confirmed with RQ-PCR, thus showing consistency of these shared aberrancies in different TCRγδ+ T-LGL leukemia patients. Interestingly, our data on TCRγδ+ T-LGL leukemia are thus completely in line with earlier studies, showing signaling towards survival through positive regulation of T cell receptor signaling and an enhanced immune response, rather than towards induction of apoptosis [36,37]. Additionally, the supervised analysis yielded aberrancies in the normal functioning of the TCRγδ+ T-LGL lymphoproliferative cells, with more skewing towards activation and inflammation given the up-regulation of CD28 and CCR7. IFNG on the other hand was down-regulated, implying that normal functioning of TCRγδ+ T cells through IFN-γ production upon activation during infection [38], is lost. These expression levels were also confirmed on fold change level by means of RQ-PCR. Furthermore, communication of the TCRγδ+ T-LGL leukemia cells with other immune cells was affected as reflected by the down-regulation of chemokine CX3CR1.

In CD8+ TCRαβ+ T-LGL leukemia altered signaling through STATs has been implicated in the leukemogenesis, based on the frequent occurrence in mutations in the STAT3 and/or STAT5b genes [39,40]. In our cohort of 10 TCRγδ+ T-LGL leukemia cases analyzed with gene expression profiling we checked expression levels of STAT3 and STAT5b genes but did not find significant alterations as compared to healthy control TCRγδ+ T cell subsets. This might suggest that STAT3 and STAT5b are less clearly implicated in the pathogenesis of TCRγδ+ T-LGL leukemia than CD8+TCRαβ+ T-LGL and/or NK-LGL leukemia. Furthermore, it has been shown previously that the survival of leukemic T-LGL cells is rather through STAT3-independent signaling [36].

Notably, not all genes could be confirmed with the same fold change in microarrays and RQ-PCR. High heterogeneity in expression levels within both healthy controls and patients was observed. This could be due to general differences between individuals, displaying different gene expression profiles, ranging from ones with high proliferative activity of TemRA TCRγδ+ T cells, to ones with low activity. Therefore, a more extended investigation into the aberrant apoptotic profile of a higher number of TCRγδ+ T-LGL leukemia patients is warranted. Furthermore, stimulation of the TCRγδ+ T-LGL leukemia cells, or in vitro blocking of the apoptotic genes identified and validated in our study, should provide more insights in the activation profile or the activity against for instance other blood cells, which could potentially explain the accompanying cytopenia as seen in these patients. Also, this could shed light on the underlying stimulations of the TCRγδ+ T cells required for proliferation. Additionally, novel techniques such as RNA-sequencing

might be helpful in creating a broader perspective on the disease, including more information about possible splice variants and single nucleotide variants.

Overall, our current study provides more insight in the pathogenesis of TCRγδ+ T-LGL leukemia by showing a disturbed balance in proliferation and apoptosis, but also in immune and inflammatory responses and normal functioning of the TCRγδ+ T cells. TCRγδ+ T-LGL leukemia cells originate from antigen-experienced normal TemRA/TemRO TCRγδ+ T cells in adult peripheral blood, but these TCRγδ+ T-LGL cells have undergone a shift in the proliferation-apoptosis balance, towards increased proliferation and survival.

ACKNOWLEDGEMENTS

The research for this manuscript was (in part) performed within the framework of the Erasmus Postgraduate School Molecular Medicine.

CONFLICT OF INTEREST

The authors have nothing to declare.

REFERENCES

1. Harris, N.L. *et al.* World Health Organisation classification of tumours of haematopoietic and lymphoid tissues. Lyon: IARC press (2001).
2. Rose, M.G. & Berliner, N. T cell large granular lymphocyte leukemia and related disorders. *Oncologist* **9**, 247 – 258 (2004).
3. Sokol, L. & Loughran, T.P. Large granular lymphocyte leukemia. *Oncologist* **11**, 263 – 273 (2006).
4. Sandberg, Y. *et al.* Lack of common TCRA and TCRB clonotypes in CD8+/TCRαβ+ T cell large granular lymphocyte leukemia: a review on the role of antigenic selection in the immunopathogenesis of CD8+ T-LGL. *Blood Cancer J.* **4**, e172 (2014).
5. Langerak, A.W., Sandberg, Y. & van Dongen, J.J.M. Spectrum of T-large granular lymphocyte lymphoproliferations: ranging from expanded activated effector T cells to T cell leukaemia. *Br. J. Haematol.* **123**, 561 – 562 (2003).
6. Sandberg, Y. *et al.* TCRγδ+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRγδ+ T cells. *Leukemia* **20**, 505 – 513 (2006).

7. Lamy, T. & Loughran Jr, T.P. Clinical features of large granular lymphocyte leukemia. *Semin. Hematol.* **40**, 185 – 195 (2003).
8. Lamy, T., Loughran, T.P. Current concepts: large granular lymphocyte leukemia. *Blood*. *Rev*. **13**, 230 – 240 (1999).
9. Semenzato, G., Zambello, R., Starkebaum, G., Oshimi, K. & Loughran, T.P. The lymphoproliferative disease of granular lymphocytes: updated criteria for diagnosis. *Blood* **89**, 256 – 260 (1997).
10. Garrido, P. *et al.* Monoclonal TCR-Vβ13.1+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis: evidence for an antigen-driven chronic T cell stimulation origin. *Blood* **109**, 4890-4898 (2007).
11. Clemente, M.J. *et al.* Clonal drift demonstrates unexpected dynamics of the T cell repertoire in T-large granular lymphocyte leukemia. *Blood* **118(16)**, 4384 – 4393 (2011).
12. Wlodarski, M.W. *et al.* Phenotypic differences between healthy effector CTL and leukemic LGL cells support the notion of antigen-triggered clonal transformation in T-LGL leukemia. *J. Leukoc. Biol.* **83(3)**, 589 – 601 (2007).
13. Van Dongen, J.J.M. *et al.* EuroFlow antibody panels for standardized *n*-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* **26**, 1908-1975 (2012).
14. Van Dongen, J.J.M. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* **17**, 2257-2317 (2003).
15. Staal, F.J.T. *et al.* DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers. *Leukemia* **17**, 1324 – 1332 (2003).
16. Nodland, S.E. *et al.* IL-7R expression and IL-7 signaling confer a distinct phenotype on developing human B-lineage cells. *Blood* **118(8)**, 2116 – 2127 (2011).
17. Van Zelm, M.C. *et al.* Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. *J. Immunol.* **175 (9)**, 5912 – 5922 (2005).
18. Irizarry, R.A., Hobbs, B., Collin, F. & Speed, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4 (2)**, 249 – 264 (2003).
19. Dik, W.A. *et al.* New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201(11)**, 1715 – 1723 (2005).
20. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA.* **98 (9)**, 5116 – 5121 (2001).
21. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systemic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc*. **4 (1)**, 44 – 57 (2009).
22. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37 (1)**, 1 – 13 (2009).
23. Beillard, E. *et al.* Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) – a Europe against cancer program. *Leukemia* **17(12)**, 2474 – 2486 (2003).

24. Lefranc, M.P. *et al.* IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Res.* **37**, 1006-1012 (2009).
25. Shaw, G.R. & Naik, V.S. The γδ variant of T cell large granular lymphocyte leukemia is very similar to the common αβ type: report of two cases. *J. Hematopathol.* **1 (2)**, 139 – 143 (2008).
26. Pérez Sánchez, I. *et al.* Chronic clonal proliferative disease of gamma-delta (γδ) T-cells in a patient with rheumatoid arthritis and neutropenia: lack of the morphology and immunophenotype of large granular lymphocytes. *Leukemia Lymphoma* **45**, 1935 – 1937 (2004).
27. Bonneville, M., O'Brien, R.L. & Born, W.K. γδ T cell effector functions: a blend of innate programming and acquired plasticity. *Nat. Rev. Immunol.* **10**, 467-478 (2010).
28. Alonzo, E.S. *et al.* Development of promyelocytic zinc finger and ThPOK-expressing innate γδ T cells is controlled by strength of TCR signalling and Id3. *J. Immunol.* **184**, 1268-1279 (2010).
29. Kim, Y.R., Kim, K.M., Yoo, N.J. & Lee, S.H. Mutational analysis of CASP1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 14 genes in gastrointestinal stromal tumours. *Hum. Pathol.* **40**, 868-871 (2009).
30. Kotas, M.E. *et al.* Role of caspase-1 in regulation of triglyceride metabolism. *Proc. Natl. Acad. Sci.* **110**, 4810-4815 (2013).
31. Fulda, S. Targeting c-FLICE-like inhibitory protein (CFLAR) in cancer. *Expert Opin. Ther. Targets* **17(2)**, 195-201 (2013).
32. Chandrasekaran, Y., McKee, C.M., Ye, Y. & Richburg, J.H. Influence of TRP53 status on FAS membrane localization, CFLAR (c-FLIP) ubiquitinylation, and sensitivity of GC-2spd (ts) cells to undergo FAS-mediated apoptosis. *Biol. Reprod.* **74(3),** 560-568 (2006).
33. He, M.X. & He, Y.W. CFLAR/c-FLIPL: a star in the autophagy, apoptosis and necroptosis alliance. *Autophagy* **9(5)**, 791-793 (2013).
34. Holcik, M., Gibson, H. & Korneluk, R.G. XIAP: apoptotic brake and promising therapeutic target. *Apoptosis* **6(4)**, 253-261 (2001).
35. Sarras, H., Alizadeh Azami, S. & McPherson, J.P. In search of a function for BCLAF1. *Sci. World J.* **10**, 1450-1461 (2010).
36. Zhang, R. *et al.* Network model of survival signaling in large granular lymphocyte leukemia. *Proc. Natl. Acad. Sci.* **105(42)**, 16308-16313 (2008).
37. Shah, M.V. *et al.* Molecular profiling of LGL leukemia reveals role of sphingolipid signaling in survival of cytotoxic lymphocytes. *Blood* **112**, 770-781 (2008).
38. Skeen, M.J. & Ziegler, H.K. Activation of gamma delta T cells for production of IFN-gamma is mediated by bacteria via macrophage-derived cytokines IL-1 and IL-12. *J. Immunol.* **154(11)**, 5832-5841 (1995).
39. Koskela, H.L.M. *et al.* Somatic STAT3 mutations in large granular lymphocyte leukemia. *N. Engl. J. Med.* **366**, 1905-1913 (2012).
40. Rajala, H.L.M. *et al.* Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. *Blood* **121**, 4541-4550 (2013).

SUPPLEMENTAL DATA

Supplementary Table 1. Antibody details for FACS-based cell sorting experiments.

Tube		Fluorochrome								Used for / applied on
		PB	**PO**	**FITC**	**PerCP-Cy5.5**	**PE**	**PE-Cy7**	**APC**	**APC-H7**	
1	**Antibody**		**CD45**	**TCRVδ1**	**CD3**	**TCRVδ2**	**CD19**	**TCRαβ**		Healthy control Vδ1/Vδ2 subsets;
	Clone		HI30	TS8.2	SK7	B6	SJ25C1	IP26		Patient material, sorting on
	Manufacturer		Invitrogen	Thermo Scientific	BD Biosciences	BD Biosciences	BD Biosciences	e-Bioscience		tumor-specific cells
2	**Antibody**	**CD3**	**CD45**	**TCRαβ**	**CD27**	**CD197**	**CD45RO**	**CD45RA**	**CD19**	Healthy control, Tn, Tcm, TemRO
	Clone	UCHT1	HI30	WT31	L128	3D13	UCHL1	HI100	SJ25C1	and TemRA subsets*
	Manufacturer	BD Biosciences	Invitrogen	BD Biosciences	BD Biosciences	e-Bioscience	BD Biosciences	BD Biosciences	BD Biosciences	

*Maturation subsets naive (Tn), central memory (Tcm), TemRO and TemRA cells. Naive cells are defined as CD45RA+CD45RO-CD197+CD27+, central memory cells as CD45RA-CD45RO+CD197+CD27+, TemRO cells as CD45RA-CD45RO+CD197-CD27- and TemRA cells as CD45RA+CD45RO-CD197-CD27-.

Supplementary Table 2. Primers and probes from Roche Universal Probe Library for RQ-PCR design.

Gene	Forward	Reverse*	Probe**
BCLAF1	AGTCTAGGGGCCGTTCCTC	TCCCAGTCTTTGCAGTTTCC	22
CASP1	TCACTGCTTCGGACATGACT	GCTGTCAGAGGTCTTGTGCTC	53
CCR7	GCTCAAGACCATGACCGATAC	CAGAAGGGAAGGGTCAGGA	87
CD28	CTAAGCCCTTTTGGGTGCT	CCAGAAAATAATAAAGGCCACTG	22
CFLAR	CTCAGGAACCCTCACCTTGT	CAGATTTATCCAAATCCTCACCA	53
CX3CR1	CAGTGACAGAAAACTTTGAGTACGA	AGACCACGATGTCCCCAATA	30
FAS	ATGGCCAATTCTGCCATAAG	TGACTGTGCAGTCCCTAGCTT	65
ID3	CCTGTCGGAACGCAGTCT	ATGTCGTCCAGCAAGCTCA	73
IFNG	GGAAAGAGGAGAGTGACAGAAAA	TTGGATGCTCTGGTCATCTTTA	21
KLF4	GGGAGAAGACACTGCGTCA	GGAAGCACTGGGGGAAGT	52
LEF1	CCAAACAAGGCATGTCCA	CCGGAGACAAGGGATAAAAAG	88
LTB	CTTCTCTGGTGACCTTGTTGC	CCTGATCCTGGGGCACTA	76
PRF1	GAGGGTGTGGACGTGACC	CAGGAACCTTTGTGTGTCCA	12
SOX4	CAACGCCGAGATCTCCA	GGATCTTGTCGCTGTCTTTGA	11
XIAP	TGGTATCCAGAATGGTCAGTACA	TGGCCTGTCTAAGGCAAAAT	38

*Reverse complementary primers.
**Probe numbers according to the Roche Universal Probe Library.

Supplementary Table 3. Differentially expressed probe sets between TCRγδ+ T-LGL leukemia cases and healthy TCRγδ+ T cell subsets after different supervised statistical analyses.

TCRγδ+ T-LGL leukemia vs. subset	Statistical level	Total differentially expressed probe sets	Up-regulated	Down-regulated
Naive	ANOVA 0.05*	8016	3037	4979
	SAM 0.05**	16199	13027	3172
Effector	ANOVA 0.05	1633	681	952
	SAM 0.05	298	298	0
Effector memory	ANOVA 0.05	2653	1222	1431
	SAM 0.05	513	506	7

*Two-way, multiple factor analysis of varaince (ANOVA), $p<0.05$.
**Significance analysis of microarrays (SAM), $p<0.05$.

Supplementary Table 4. Top up- and down-regulated genes after supervised analysis.

Gene symbol	Gene name	Adjusted p-value*	Fold Change**
Top 25 up-regulated genes in TCRγδ+ T-LGL leukemia cases relative to effector cells			
LYZ	Lysozyme	0.000	36.170
S100A8	S100 calcium binding protein A8	0.000	26.823
VCAN	Versican	0.000	19.917
S100A9	S100 calcium binding protein A9	0.000	18.571
RCAN3	RCAN family member 3	0.000	18.062
FCN1	Ficolin 1	0.000	16.518
G0S2	Go/G1 switch 2	0.000	15.844
NELL2	Neural EGFL like 2	0.000	15.712
CSTA	Cystatin A	0.000	14.029
SERPINB6	Serpin family B member 6	0.000	13.542
SOX4	SRY (sex determining region Y) box 4	0.000	13.277
MAFB	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog	0.000	12.790
CPVL	Carboxypeptidase, vitellogenic like	0.000	11.861
IRS2	Insulin receptor substrate 2	0.000	11.489
ALOX5	Arachidonate 5-lipoxygenase	0.000	11.421
GPR183	G protein-coupled receptor 183	0.000	10.627
MAL	Mal T cell differentiation protein	0.000	9.505
TGFB1	Transforming growth factor beta induced	0.000	9.118
CD36	CD36 molecule	0.000	8.746
CLEC7A	C-type lectin domain family 7 member A	0.000	8.647
PLBD1	Phospholipase B domain containing 1	0.000	8.415
SCML1	Sex comb on midleg-like 1 (Drosophila)	0.000	8.397
LTB	Lymphotoxin beta	0.000	7.338
CCR7	C-C motif chemokine receptor 7	0.000	7.377
AIF1	Allograft inflammatory factor 1	0.000	7.290
Top 25 down-regulated genes in TCRγδ+ T-LGL leukemia cases relative to effector cells			
ENPP5	Ectonucleotide pyrophosphatase/phosphodiesterase 5	0.000	-6.938
HS3ST3B1	Heparan sulfate-glycosamine 3-sulfotransferase 3B1	0.000	-6.191
ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	0.000	-6.137
ETNK1	Ethanolamine kinase 1	0.000	-5.674
IFIT3	Interferon induced protein with tetratricopeptide repeats 3	0.000	-5.458
CX3CR1	C-X3-C motif chemokine receptor 1	0.000	-5.443
PCMT1	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	0.000	-4.934
SNTB2	Syntrophin beta 2	0.000	-4.769
ELOVL6	ELOVL fatty acid elongase 6	0.000	-4.399
PRR5L	Proline rich 4 like	0.000	-4.390
CD226	CD226 molecule	0.000	-4.369
PRF1	Perforin 1	0.000	-4.238
ZNF260	Zinc finger protein 260	0.000	-4.094
MAN1A1	Mannosidase alpha class 1A member 1	0.000	-4.087
DYNLL1	Dynein light chain LC8-type 1	0.000	-4.061
TOMM5	Translocase of outer mitochondrial membrane 5	0.000	-3.989
CTBP2	C-terminal binding protein 2	0.000	-3.938
G3BP1	G3BP stress granule assembly factor 1	0.000	-3.918
KATNBL1	Katanin regulatory subunit B1 like 1	0.000	-3.913
RASGEF1A	RasGEF domain family member 1A	0.000	-3.896
GTF2H5	General transcription factor IIH subunit 5	0.000	-3.853
SKA2	Spindle and kinetochore associated complex subunit 2	0.000	-3.825
PEX3	Peroxisomal biogenesis factor 3	0.000	-3.815
ERMP1	Endoplasmic reticulum metallopeptidase 1	0.000	-3.807
ARFIP1	ADP ribosylation factor interacting protein	0.000	-3.781

Top 25 up- and down-regulated genes after supervised analysis on significance level of ANOVA $p<0.05$.

*The adjusted p-value is after Benjamini-Hochberg correction for multiple testing.

**The Fold Change is based on probe set intensity in LGL cases versus healthy control after normalization and statistical and multiple testing.

Supplementary Table 5. Gene Ontology biological processes and KEGG enrichment pathway analysis of TCRγδ+ T-LGL leukemia cases versus healthy TCRγδ+ TemRO cells in DAVID.

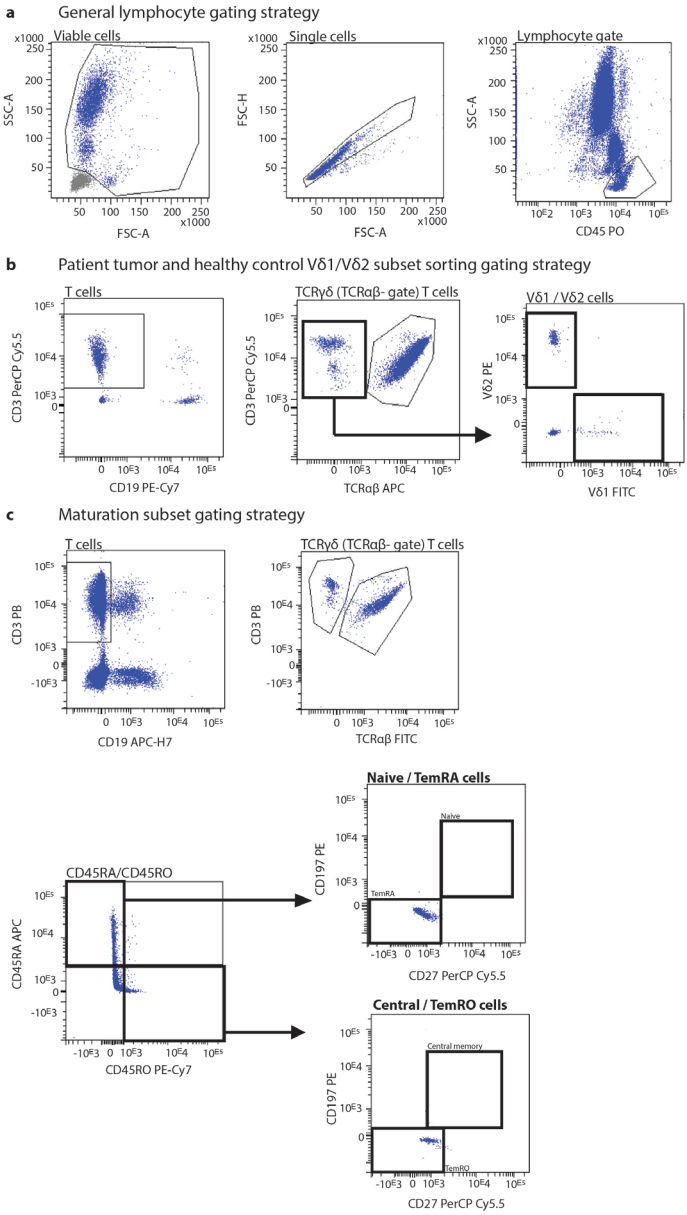
Gene Ontology biological processes					
Term	**Gene count***	**Genes**	**p-value**	**Bonferroni****	**Benjamini****
Adaptive immune response	29	CLEC10A, CD244, CD6, CD86, CD8B, DCLRE1C, GPR183, KLRC1-KLRK1, PRDM1, SH2D1A, SLAMF7, TRAT1, ALCAM, ADGRE1, CRACR2A, CAMK4, CTLA4, EOMES, HAVCR2, IGHM, **IFNG**, LILRA2/B2/B3, LAX1, LAMP3, PIK3CG, PAG1, TFEB	9.8E-8	3.7E-4	3.7E-4
Neutrophil chemotaxis	17	CCL3L3, CCL4, CXCL8, CXCR2, FCER1F, S100A12/8/9, CSF3R, C5AR1, **IFNG**, JAML, PIK3CG, PDE4B/D, PPBP, VAV3	1.2E-6	4.3E-3	2.2E-3
Innate immune response	50	ANKHD1, APOBEC3A_B, CLEC10A/7A, CD14, CD1D, CD244, CD6, DDX3X, DDX58/60, FGR, FCER1G, KLRC4-KLRK1, NLRP3, PRDM1, REL, S100A12/8/9, SH2D1A, SRPK2, APP, APOBEC3F/G, CR1, CYBB, HAVCR2, IGHM, IFI116, IL23A, KIR3DS1, KIR2DS1/2/5, KLRD1, LILRA5, LY86/96, MID2, NCF1/2, NFKB2, PTX3, PIK3CG, RNF135, SLAMF1, SUSD4, TLR2	9.2E-6	3.4E-2	1.1E-2
Cellular defense response	15	CCR6, CXCR2, **CX3CR1**, SH2D1A, TRAT, C5AR1, KLRC3/4, LILRB2, LY96, MNDA, NCR1, NCF1/2, PRF1	1.5E-5	5.4E-2	1.4E-2
Cellular response to DNA damage stimulus	32	APC, BCL6, FANCL, MRE11, NEK4, SETD7, SLF2, SHPRH, SMARCAL1, **XIAP**, ARMT1, CHD2, CDKN1A, CTLA4, CTX3L, FOXO1, MCM8, MAPK1, MNDA, NFATC, PLK3, PSEN1, SETX, SGK1, SMC6, UBR5, USP28/47, MYC, VAV3, ZBTB38, ZMAT3	1.9E-5	6.9E-2	1.4E-2
Immune response	48	CCL4, CCR6, **CCR7**, CXCL8, CXCR5, CD1D, CD27, CD36, CD86. CD8B, FCGR3A/B, GPR183, LY75-CD302, PRELID1, CTSC, C5AR1, CFP, CTLA4, FTH1, GZMH, ICOS, **IFNG**, IFITM2, IL27RA, KIR3DL and KIR2DL genes, LILRB2, LST1, LAX1, LTB, MARCH1, MBP, NCF1, OSM, PF4, PPBP, PRKRA, TLR2, TCF7, TNFSF13B/8, ZEB1	2.2E-5	8.0E-2	1.4E-2
Apoptotic process	56	DHCR2, **BCLAF1**, **CFLAR**, CD14, CFFB, FAIM, GRAMD4, NCKAP1, NLRP3, NME6, PRELID1, POLR2G, S100A8/9, TIAM1, TIA1, **XIAP**, XAF1, APAF1, ARRB1, AHR, BEX2, **CASP1**, CTSC, C5AR1, CST3, CSRNP1, CIAPIN1, DAXX, DYNLL, ECE1, EFNA5, FEM1B, FOXO1, GZMB/H, GADD45A/B, ING4, **IFNG**, IRF, LY86, MAL, MAPK1, MAP3K1, NCF1, PPID, PRF1, PLK3, PPP2R2B, RPS6KA1, SGK1, STAT1, TLR2, TGFBR1, TNFSF8	1.5E-3	1.0E0	4.4E-1
KEGG enrichment pathway analysis					
Term	**Gene count***	**Genes**	**p-value**	**Bonferroni****	**Benjamini****
Antigen processing and presentation	19	CD8B, IFI30, HSP90AB1, HSPA6, **IFNG**, KIR3DL and KIR2DL/DS genes, KLRC3/4, KLRD1, NFYB, RFXAP	2.2E-6	5.9E-4	5.9E-4
Natural killer cell mediated cytotoxicity	25	CD244, FCER1G, FCGR3A/B, KLRC4-KLRK1, SH2D1A, GZMB, IFNGR2, **IFNG**, KIR2DL and KIR2DS genes, KLRD1, MAPK1, NCR1, NFATC2, PRF1, PIK3CG, PIK3R1, RAC2, VAV3	5.0E-6	1.3E-3	6.7E-4
Osteoclast differentiation	22	FCGR3A/B, CAMK4, CYBB, IFNGR2, **IFNG**, LILRA/B genes, MAPK1, NCF1/2/4, NFKB2, NFATC2, PIK3CG, PIK3R1, SIRPA, STAT1, SOCS3, TGFBR1	7.5E-4	1.8E-1	6.5E-2
Rheumatoid arthritis	15	ATPase genes, CCL3L3, CXC8, **CD28**, CD86, CTLA4, **IFNG**, IL23A, LTB, TLR2, TNFSF13B, VEGFA	3.1E-3	5.6E-1	1.9E-1

*Total 1686 differentially expressed genes in LGL versus TemRO dataset with FC=2 both up- and down-regulated, p<0.05 (ANOVA), of which 1563 were annotated by DAVID using Affymetrix Human Genome U133 Plus 2.0 array as background and selecting Homo Sapiens as species.
**Adjusted p-value based on Bonferroni and Benjamini-Hochberg correction for multiple testing.
Genes also identified through DAVID LGL versus effector subset and IPA analysis, which are further validated with RQ-PCR are indicated in bold.

Supplementary Table 6. Median (interquartile range) mRNA expressions, median ΔΔCt and microarray fold changes of genes used for RQ-PCR validation.

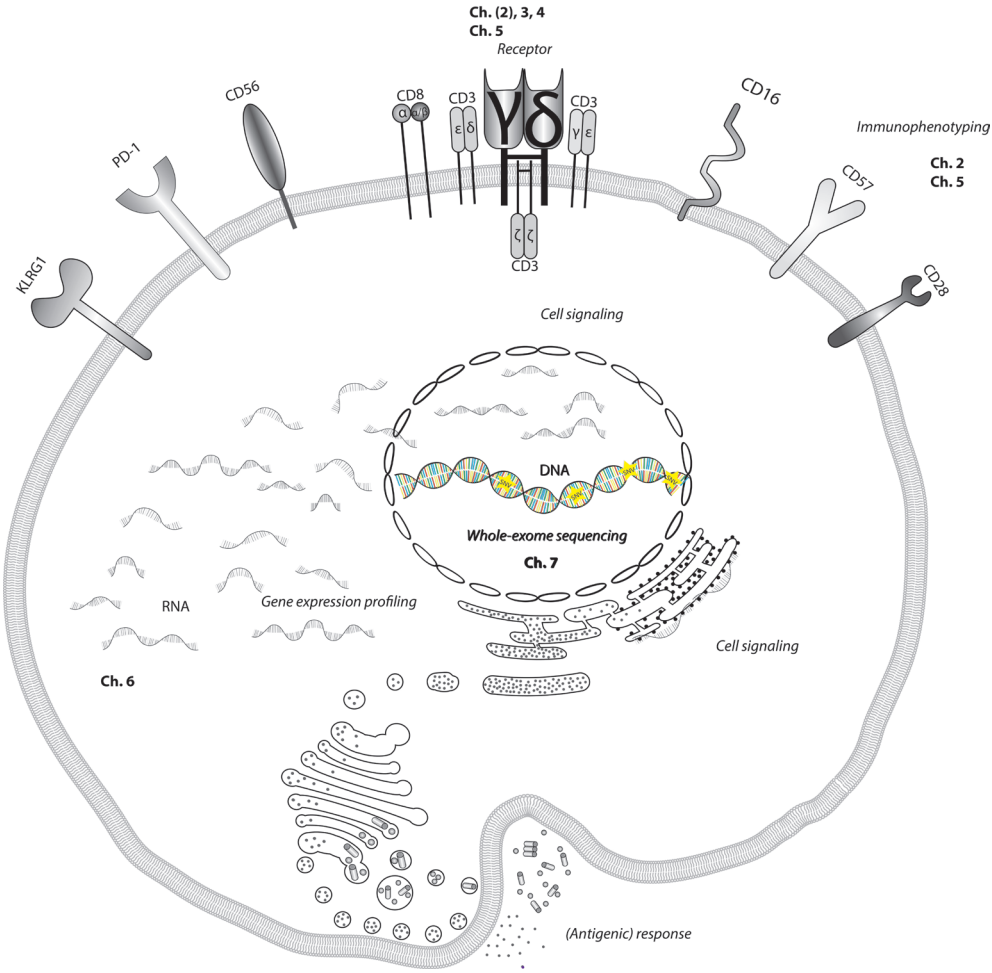
Gene	Healthy controls*	TCRγδ+ T-LGL leukemia patients*	Median LGL ΔΔCt value**	Microarray LGL FC value***
		Validated genes		
XIAP	3.360 (1.336 – 8.128)	2.698 (0.5617 – 12.02)	-2.489	-2.416
CASP1	3.137 (2.056 – 5.116)	3.060 (1.252 – 6.208)	-5.269	-2.783
BCLAF1	55.75 (41.62 – 71.92)	187.7 (10.38 – 600.1)	3.867	2.039
CFLAR	4.452 (4.231 – 4.672)	0.08167 (0.05796 – 0.2283)	-65.767	-2.449
ID3	0.006839 (0.001403 – 0.3605)	2.784 (1.346 – 5.998)	28.410	7.228
CD28	2.700 (1.414 – 5.076)	1.073 (0.7529 – 5.330)	3.283	7.012
CCR7	1.197 (0.3527 – 3.101)	0.5301 (0.07773 – 6.065)	4.372	7.377
CX3CR1	52.88 (15.34 – 90.41)	10.05 (1.092 – 16.23)	-2.044	-5.433
IFNG	0.5165 (0.3107 – 0.8488)	0.4938 (0.1770 – 0.8534)	-6.531	-3.682
		Non-validated genes		
FAS	0.5521 (0.3813 – 0.6345)	1.598 (0.03642 – 2.156)	3.059	-2.215
KLF4	2.083 (1.085 – 10.84)	0.2712 (0.05194 – 0.7014)	-9.363	6.813
LEF1	2.764 (0.9663 – 12.62)	0.4080 (0.2412 – 1.563)	-4.125	16.365
SOX4	0.2511 (0.04589 – 0.5471)	0.1589 (0.03550 – 0.3438)	-32.236	13.277
LTB	5.418 (3.768 – 20.24)	2.839 (0.4984 – 7.019)	-31.265	7.388
PRF1	60.35 (33.89 – 84.32)	125.0 (30.92 – 205.1)	2.090	-4.238

*Relative mRNA expression after *ABL* housekeeping gene correction.
**Median ΔΔCt value of TCRγδ+ T-LGL leukemia patients after correction with average ΔCt values from healthy control samples.
***Fold changes according to supervised TCRγδ+ T-LGL leukemia cells vs. healthy TemRA TCRγδ+ T cells comparison.



Supplementary Figure 1. Gating strategies for FACS-based sorting experiments for sorting patient and healthy control material.

(**a**) General gating strategy to exclude debris and doublets, and to define lymphocytes based on CD45. (**b**) Further gating strategy for sorting patient material and TCRVδ1 and TCRVδ2 healthy subsets, and (**c**) gating strategy for sorting maturation subsets naive (CD45RA+CD45RO-CD197+CD27+), central memory (CD45RA-CD45RO+CD197+CD27+), TemRO (CD45RA-CD45RO+CD197-CD27-) and TemRA (CD45RA+CD45RO-CD197-CD27-).


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 7
RNA
Gene expression profiling
Cell signaling
Ch. 6
(Antigenic) response

Chapter 7

Whole exome sequencing reveals novel insights in the pathogenesis of TCRγδ+ T cell large granular lymphocyte leukemia

Martine J. Kallemeijn[1], Jeroen G.J. van Rooij[2], Robert Kraaij[2], Jacques J.M. van Dongen[1], Anton W. Langerak[1]

[1]Department of Immunology, Laboratory for Medical Immunology, Erasmus MC, Rotterdam 3000 CA, the Netherlands;
[2]Department of Internal Medicine, HuGeF Human Genome sequencing Facility, Erasmus MC, Rotterdam 3000 CA, the Netherlands.

Manuscript in preparation

ABSTRACT

TCRγδ+ T-LGL leukemia is a rare type of mature chronic T cell lymphocytic disorder in the peripheral blood. Previous studies have shown that these cells originate from antigen-experienced terminally differentiated effector TCRγδ+ T cells with a disturbance in normal signaling, and most importantly a dysregulated balance between proliferation and apoptosis. As the clinical disease phenotype is rather heterogeneous, this raises questions about a possible common pathogenic mechanism in TCRγδ+ T-LGL leukemia. Recently, the presence of somatic *STAT3* and *STAT5B* mutations in in has provided insights in the pathogenesis of both NK and CD8+TCRαβ+ T-LGL leukemia, but this has not been confirmed in TCRγδ+ T-LGL leukemia cells. Here, we applied whole exome sequencing on sorted tumor cell populations of patients with proven TCRγδ+ T-LGL leukemia, and in parallel we sorted healthy monocytes to be used as internal control to correct for germline associated variants. Our data showed a rather heterogeneous profile of possible pathogenic gene variants, with affected genes being involved in transcription, immune cell activation and lymphangiogenesis. Gene variants were validated via Sanger-based sequencing and were evaluated in the context of gene expression data. A selected gene set showed homozygous variants, that were present in all patient tumor cells, but absent in healthy monocytes, and which could act as common factor in the pathogenesis of TCRγδ+ T-LGL leukemia.

Keywords: TCRγδ+ T cells, T-LGL, leukemia, whole exome sequencing.

INTRODUCTION

T cell large granular lymphocyte (T-LGL) leukemia is a heterogeneous chronic mature T cell neoplasia, which is recognized as a separate hematological disorder according to the World Health Organization (WHO) classification [1]. T-LGL leukemia originates from normal LGL cells which comprise 10-15% of peripheral blood mononuclear cells (PBMNCs) [2] and can be subdivided into two major groups based on the type of T cell receptor (TCR): TCRαβ or TCRγδ. The majority of the T-LGL leukemias involve the CD8+TCRαβ+ variant (80-90%) [3], whereas only a small part has a CD4+TCRαβ+ (1-5%) or TCRγδ+ phenotype (5%) [4]. T-LGL leukemia is part of a spectrum, ranging from lymphoproliferation to leukemia [5-7]. T-LGL leukemia generally shows an indolent disease course, with approximately 33% of patients being asymptomatic at diagnosis [7,8], and affects mostly elderly patients with an average age of 60 years. Most prominent clinical features involve neutropenia, recurrent bacterial infections and typical B symptoms which are generally associated with chronic leukemia.

Typically, TCRγδ+ T-LGL leukemia shows high association with cytopenia (thrombocytopenia, neutropenia), autoimmune diseases such as rheumatoid arthritis, and malignancies varying from other hematological cancers to solid tumors [4]. Diagnosis is based on a persistent (>6 months) monoclonal CD3+CD57+TCRγδ+ LGL cell population (>0.4 x 10^9/L), established by flow cytometry and/or cell morphology [9]. Furthermore variable expression of CD8, CD16 and CD56 is associated with TCRγδ+ T-LGL leukemia as well [6]. Monoclonality can also be assessed with PCR-based assays [10].

Despite recent advances the LGL disease etiology remains largely unknown. It has been hypothesized that chronic (antigenic) stimulation plays a major role in the development of the proliferation [6]. An activation-associated effector phenotype and a (time-dependent) skewed TCR expression pattern – also referred to as clonal drift [11] – has been observed in LGL, which could be driven by autoantigens or pathogens [3,12]. Both the clinical characteristics and the observed clonal drifts illustrate the high heterogeneity of the disease. Furthermore, in our previous gene expression profiling study we observed that the TCRγδ+ T-LGL cells correlated best with antigen-experienced effector cells. Transcriptome data of the LGL cells disclosed an anti-apoptotic and pro-survival phenotype [13].

Some years ago CD8+TCRαβ+ T- and NK-LGL leukemias were analyzed by means of whole exome sequencing, revealing shared somatic mutations in the *STAT3* and *STAT5B* genes [14,15]. Four somatic mutational hotspots were identified in the Src Homology 2 (SH2) domain of *STAT3* [14], and two in the *STAT5B* gene, also in the SH2 domain [15]. These mutations were shown to affect expression of downstream genes in the STAT3 and

STAT5B pathways, including *IFNGR2*, *BCL2L1* and *JAK2*. From these studies it became clear that the different mutations are associated with different clinical phenotypes and prognoses, thus partially explaining the heterogeneity of the disease, despite common morphological and cellular features. In TCRγδ+ T-LGL leukemia the presence of these *STAT3* and *STAT5B* mutations has not been documented yet. Of note, approximately 33% of patients with TCRγδ+ T cell lymphoma has been associated with a *STAT5B* mutation [16]. Even though differences at the cytogenetic level have been associated with diagnosis and prognosis [17,18], other mutations associated with TCRγδ+ T-LGL leukemia have not yet been described. In the current study we aimed to perform whole exome sequencing analysis on tumor cells from ten patients with a proven TCRγδ+ T cell leukemia. We also included monocyte DNA of these cases to filter out germline variants present in hematopoietic cells and to establish true tumor-associated variants and/or mutations. Here we report on mutations in genes involved in cell cycle progression and nuclear repression (*NCOR2*), cancer development (*HAVCR1*, *CREB3L2*), immune cell activation (*HAVCR1*, *KLRF1*), metastasis (*VEGFC*) and small molecule synthesis (*NUDT11*), which were commonly associated with TCRγδ+ T-LGL leukemia with minor allele frequency MAF=1.000 within the dataset, and which could contribute to the pathogenesis. Additionally, through transcription factor target enrichment analysis we could identify transcription factors that could be driving factors in TCRγδ+ T-LGL leukemogenesis, based on the appearance of protein-altering variants in tumor cases only.

MATERIALS & METHODS

Patient samples

Patient database files from the department of Immunology, Erasmus MC (Rotterdam, The Netherlands) were reviewed retrospectively for selection of cases with a proven persistent (>6 months) mature TCRγδ+ T cell leukemia in peripheral blood (PB) and/or bone marrow (BM) based on combined clinical, histological (HE sections), cytomorphological (May-Grünwald-Giemsa staining), immunophenotypical (CD3+CD56+CD57+TCRγδ+TCRαβ-) and molecular data. Isolation of PB and / or BM mononuclear cells (MNCs) had been performed by means of Ficoll-Paque (density 1.077 g/ml, Pharmacia, Uppsala, Sweden) density gradient separation and cells were stored in Iscove's Modified Dulbecco's Medium (IMDM, Lonza, Basel, Switzerland) containing 10% dimethyl sulfoxide (DMSO) (Sigma-Aldrich, Saint Louis, MI, USA) at -180 °C until further use. Our retrospective search resulted in 10 TCRγδ+ T-LGL leukemia patients of whom 6 patient samples had sufficient material to also include control monocyte populations for paired analysis (Supplementary Table 1). All samples had been obtained

upon informed consent and the study was approved by the Erasmus MC Medical Ethics Committee (MEC-2015-617). All studies were conducted in accordance with the principles of the Declaration of Helsinki.

Cell sorting

For sorting experiments cryopreserved PB- or BM-MNCs were thawed. Based on the immunophenotype of the particular TCRγδ+ T-LGL leukemia as determined during diagnostic work-up, tumor cells were sorted using CD45, CD3, TCRαβ, TCRVδ1 and TCRVδ2 markers. In addition, from the same sample monocytes were sorted as control cell population, using CD45, CD3, CD19 and CD14 markers (Supplementary Table 2). Sorting experiments were performed on FACS Aria I and III sorting instruments (BD Biosciences). Employed gating strategies are depicted in Supplementary Fig. 1.

DNA/RNA isolation

After sorting the cells were lysed and subjected to combined DNA / RNA isolation using the Qiagen DNA/RNA/miRNA AllPrepKit (Qiagen, Hilden, Germany). DNA concentration was measured with the NanoDrop 1000 spectrophotometer (Thermo Fischer Scientific, Waltham, MA, USA), and the A260/A280 ratio was determined to check DNA quality for further whole exome sequencing.

Exome capture and sequencing

Isolated DNA (1 μg) from sorted patient tumor cells and monocytes was subjected to exome capture and library preparation with the NimbleGen SeqCap EZ v2 (Roche, Basel, Switzerland). Exome sequencing was performed with the Illumina TruSeq V3 2x100 PE sequencing protocol (Illumina, San Diego, CA, USA).

Mutation annotation and analysis pipeline

Paired-end reads were aligned to the hg19 (GRCh37) reference, built using the Burrows-Wheeler Aligner (BWA) [19]. The Genome Analysis Toolkit (GATK) version 3.4 [20] was used to remove duplicates (Samtools), realign insertions and deletions (indels), recalibrate the base quality score and to remove duplicate reads. Furthermore, single nucleotide variants (SNVs) and indels were called via standard default parameters. The ANNOVAR tool [21] was used to annotate mutations based on RefSeq (Release 70). To analyze SNVs and indels present in tumor material only we excluded all variants found in monocyte control samples, using either pooled tumor samples vs. pooled monocyte samples or paired tumor and monocyte samples on a per patient basis. Genes with highest density scores based on the Rotterdam Study [22] were excluded. Severity scores of missense variants were calculated using the LJB prediction database, as incorporated

in the ANNOVAR analysis pipeline, and a database for nonsynonymous SNVs' functional predictions (dbNSFP), version ljb2B [23], which included SIFT [24], PolyPhen2 [25], MutationTaster [26], MutationAssessor [27,28], likelihood ratio test (LRT), Functional Analysis through Hidden Markov Models (FATHMM) [29], and conservation prediction methods such as GERP [30] and PhyloP [31] tools. For other functional predictions associated with diseases and metabolism pathways the COSMIC database [32] and Reactome Knowledgebase [33,34] were applied.

Functional dataset analyses

Further data filtering was performed with R Studio [35; www.rstudio.com] and TIBCO Spotfire (TIBCO Software Inc., Palo Alto, CA, USA). Data interpretation and analysis was performed with WebGestalt [36,37; bioinf.vanderbilt.edu/webgestalt], Database for Annotation, Visualization, and Integrated Discovery (DAVID Database) [38,39], and QIAGEN's Ingenuity Pathway Analysis (IPA) [40; QIAGEN, Redwood City, USA; www.qiagen.com/ingenuity]. WebGestalt gene ID mapping was based on Entrez gene IDs [41]. SNP rs IDs were based on the NCBI dbSNP database [42].

Validation of mutations

Candidate variants were selected based on different filtering methods. DNA from samples included in the whole exome sequencing study cohort (N=11) as well as additional DNA samples from other TCRγδ+ (N=33, total TCRγδ+ T-LGL patient number N=44) and CD8+TCRαβ+ (N=19) T-LGL patients were used to validate the presence of the mutation or variant via Sanger sequencing using primers as listed in Supplementary Table 3. Primers were developed using Oligo 6.41 software (Molecular Biology Insights, Cascade, CO, USA), Primer Express 3.0 software (Thermo Fischer Scientific) and Primer3 software [43,44]. Sequencing was performed using the Big Dye Terminator (BDT) enzyme (Life Technologies, Carlsbad, CA, USA) and the ABI Prism 3130 sequencer (Applied Biosystems, Foster City, CA, USA). Potential effects of variants on gene expression were measured with real-time quantitative (RQ)-PCR. RQ-PCR assays containing primers and probes were designed using the Roche Universal Probe Library (Roche, Basel, Switzerland) (Supplementary Table 4). RQ-PCR experiments were performed with 2x TaqMan Universal Mastermix (Applied Biosystems) on the StepOnePlus instrument (Thermo Fischer Scientific). Ct values of tumor samples were normalized to the Ct value of the ABL housekeeping gene (ΔCt) [45]. Fold changes were calculated based on the ΔCt values ($2^{-\Delta Ct}$) and used for calculation of the LGL cells to control cells ratio (ΔΔCt).

Statistical analyses

WebGestalt GO and KEGG enrichment analyses were statistically analyzed with the hypergeometric method (default settings). Multiple test adjustments were based on the Benjamini-Hochberg correction. Level of significance was set at 0.05. The minimum number of genes for a category was set at 10 for the comparison of all tumor samples vs. all monocyte samples, and set at 2 for the paired-wise tumor vs. monocyte analysis on a per patient basis (default settings due to lower number of differential variants). DAVID database OMIM disease functional annotation related genes indicated with Cohen's kappa score, were calculated by the percentage for chance of corrected observed agreement divided by the percentage for chance of corrected potential agreement to rate the inter-rater agreement for qualitative (categorical) subjects. The kappa score is indicated between 1 (highest, best agreement) and 0 (lowest, no agreement). IPA calculations for p-values interpreting functions, processes and pathways in IPA Knowledgebase, Global Functional Analysis (GFA) and Global Canonical Pathways (GCP) were performed with the right-tailed Fisher Exact Test. The p-value is therefore a measure of the likelihood that the association between the gene set and processes or canonical pathways is due to random chance. The smaller the p-value the less likely the association is considered to be random, and thus the more significant.

RESULTS

Whole exome sequencing of TCRγδ+ T-LGL leukemia cells shows a high frequency of nonsynonymous missense variants

First, sequencing data were subjected to genome alignment, variant calling and quality control checks. As all patients were from Caucasian descent, the expected number of single nucleotide polymorphisms (SNPs) from the whole exome sequencing (WES) dataset would be approximately 30,000 per sample, and the expected number of indels around 3,000. Furthermore, coverage should reach >50X, contamination should be <5% (0.05) and transition/transversion (TiTv) ratio should reach around 2.0-2.1, all according to recent 1000Genomes human whole genome sequencing studies [46]. All samples, except for the monocyte sample from patient LGL092 – which was therefore excluded from further analysis – fulfilled this initial quality control check (Supplementary Table 5).

Based on total comparison of all TCRγδ+ T-LGL leukemia samples versus all monocyte control samples, to exclude germline hematopoietic variants, most variants were found in the intronic and exonic regions, although variants were also found in other regions, such as splicing, intergenic, 5-prime (5') untranslated region (UTR), 3'-UTR and

noncoding RNA (ncRNA) (Fig. 1). Only exonic variants could be annotated, of which most were nonsynonymous single nucleotide variants (SNVs), indicating protein alterations. Other SNVs included synonymous (nonsense) SNVs, SNVs causing stop codons (stop gain) and causing loss of stop codons (stop loss), or (non-)frameshift substitutions, and insertions or deletions (Fig. 1). These results indicate that TCRγδ+ T-LGL leukemia cells have a high frequency of missense mutations underlying their pathogenesis.

Protein-changing variants affect pathways in metabolism, signaling, proliferation and apoptosis based on oncogenic transcription factors targeting affected genes

After selecting variants i) only present in tumor cells (N=18449), ii) being exonic (N=7663), iii) excluding synonymous and unknown variants (N=4662), iv) excluding genes with the highest SNP density in general (N=3882, based on results Rotterdam Study, [22]) and v) selecting variants of which all alleles were called in all samples, a total of 3645 variants in 2733 remaining genes were further analyzed in WebGestalt and DAVID databases. WebGestalt's Gene Ontology (GO) slim classifications resulted in summaries of variants which affected genes associated with biological processes (Fig. 2a), molecular functions (Fig. 2b) and cellular components (Fig. 2c). Numbers of genes in each process, function or component are indicated above the bars in the graphs, with most affected genes being involved in "metabolism", "signaling" and "nuclear activities". Upon GO enrichment analysis biological processes involved in "cell adhesion", molecular functions involved in "molecule binding", "catalytic activity" and "receptor signaling", and cellular components such as "plasma membrane", "cell-cell junction" and "cytoskeleton were found to be significantly affected (Table 1, Supplementary Fig. 2). When focusing more in-depth by means of KEGG enrichment analyses, pathways involved in "metabolism", "lysosome", "signaling" (e.g. Calcium, cytokines, phosphatidylinositol, Notch), "cancer pathways" (cancer in general, prostate cancer, colorectal cancer) and "apoptosis" appeared significantly affected (Table 1).

To investigate which transcription factors were driving the expression of these genes, transcription factor target (TFT) enrichment analysis was performed. The majority of the genes was regulated by typical T cell-, T cell signaling- and proliferation-related transcription factors such as E12, NFAT, PAX4 and STAT5B. Furthermore, many of the variant genes appeared to be regulated by (proto-)oncogenes involved in solid tumors and hematopoietic malignancies, including FOXO4, LEF1, SP1, HNF3, ETS2, MEIS1, MYC and AML1 (Table 1).

From six patients we obtained both tumor DNA and control monocyte DNA on which WES could be performed (Supplementary Tables 1 and 5). After separate paired analyses of these samples (i.e. tumor versus monocyte WES data for each patient), the numbers of

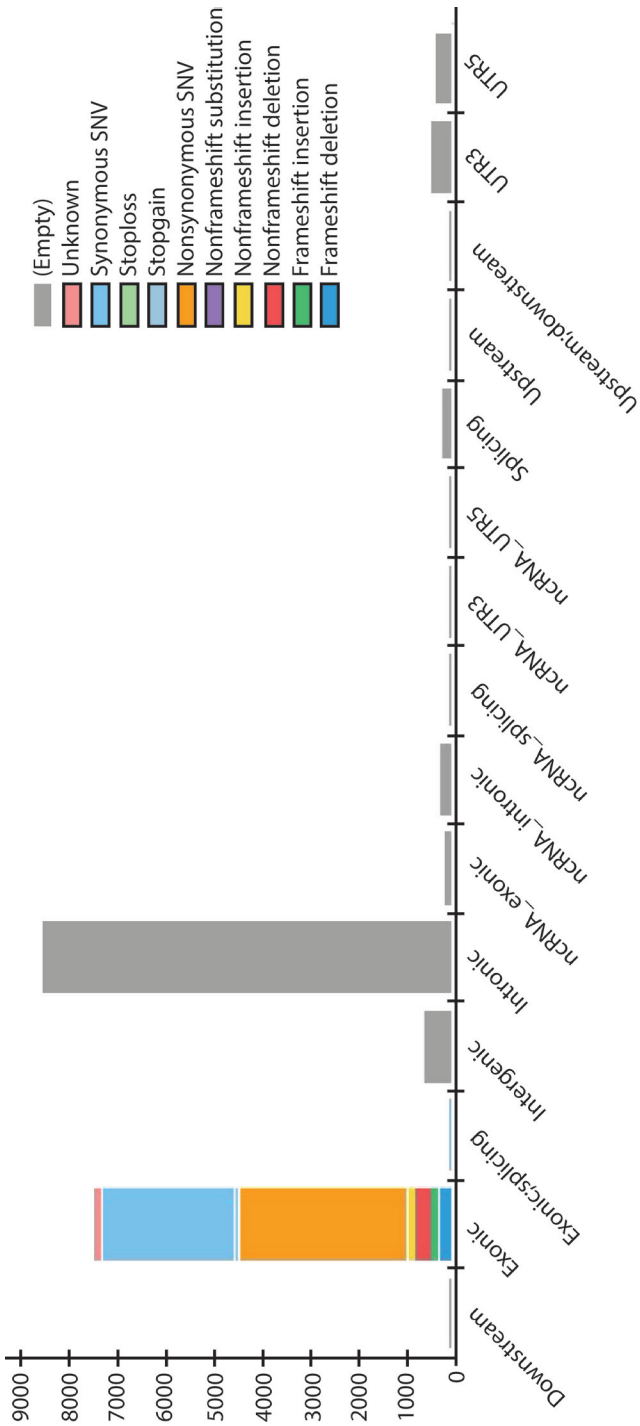


Figure 1. TCRγδ+ T-LGL leukemia variant distribution.
Distribution of all TCRγδ+ T-LGL leukemia-only related variants is shown in bar graphs (adapted from TIBCO Spotfire software). Most variants are located in introns and exons. Exonic variant distribution is depicted in colors, the majority of which is nonsynonymous.

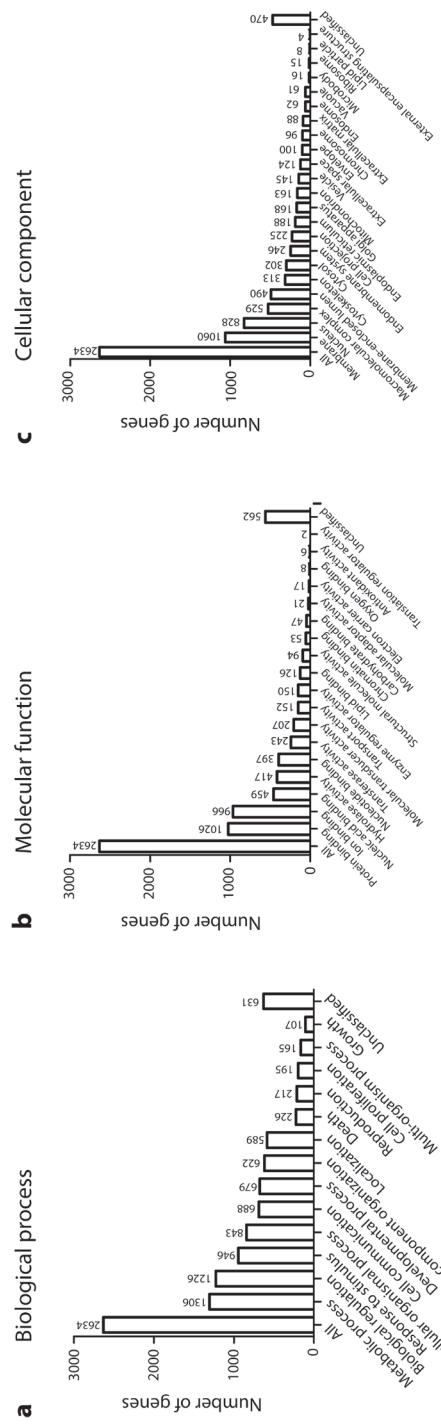


Figure 2. GO slim classifications in WebGestalt.
All TCRγδ+ T-LGL leukemia-only exonic, protein-altering variants (excluding synonymous or unknown variants and genes with a generally high SNP density score) are depicted in GO slim classification bar graphs adapted from WebGestalt's analysis. Of note, many genes contained multiple variants, but here only genes are shown. Bar graphs depict genes bearing variants affecting biological processes (**a**), molecular functions (**b**) and cellular components (**c**). Numbers of genes per process, function or organelle are indicated above the corresponding bars. Total number of genes was 2,733, of which 2,634 were unambiguously mapped to 2,634 Entrez gene IDs. The remaining 99 genes were mapped to multiple IDs, or could not be mapped to any Entrez ID. Of note, genes can be overlapping between different biological processes, molecular functions and cellular components.

Table 1. GO, KEGG and TFT enrichment analysis in WebGestalt

GO							
Biological process		**#Genes**				**Adjusted P***	
Cell adhesion		218				8.66E-07	
Molecular function		**#Genes**				**Adjusted P**	
Calcium ion binding		153				1.96E-05	
Nucleotide binding		427				1.21E-02	
Peptidase activity		122				3.00E-03	
Ras guanyl-nucleotide exchange factor activity		30				1.16E-02	
Rho guanyl-nucleotide exchange factor activity		24				3.40E-03	
Cellular component		**#Genes**				**Adjusted P**	
Intrinsic to plasma membrane		237				6.90E-03	
Lysosomal lumen		20				4.10E-02	
Microtubule organizing center		99				1.82E-02	
Cell-cell junction		62				3.40E-02	
KEGG							
Pathway	**#Genes**	**Statistics****					
		C	**O**	**E**	**R**	**RawP**	**AdjP**
Lysosome	34	121	34	8.19	4.15	5.40E-13	2.11E-11
Calcium signaling pathway	32	177	32	11.98	2.67	3.23E-07	4.20E-06
Pathways in cancer	43	326	43	22.06	1.95	2.25E-05	0.0001
Notch signaling pathway	11	47	11	3.18	3.46	0.0002	0.0010
Inositol phosphate metabolism	12	57	12	3.86	3.11	0.0004	0.0016
Apoptosis	14	87	14	5.89	2.38	0.0021	0.0053
Jak-STAT signaling pathway	17	155	17	10.49	1.62	0.0336	0.0452
TFT							
Transcription factor	**#Genes**	**Statistics**					
		C	**O**	**E**	**R**	**RawP**	**AdjP**
E12	361	2450	361	165.77	2.18	2.16E-46	1.20E-43
NFAT	278	1871	278	126.59	2.20	3.50E-36	3.25E-34
LEF1	280	1939	280	131.19	2.13	2.99E-34	2.38E-32
MYC	144	1015	144	68.68	2.10	2.54E-17	6.15E-16
SMAD3	42	235	42	15.90	2.64	7.24E-09	6.20E-08

Selected GO biological processes, molecular functions, cellular components, KEGG pathways and transcription factor targeting (TFT) driving genes with variants. Table adapted from WebGestalt analysis.
*Adjusted P value after Benjamini-Hochberg correction for multiple testing.
**Statistics column divided into six variables. C: the number of reference genes in the category; O: the number of genes in the gene set and also in the category; E: the expected number in the category; R: ration of enrichtment. The rawP is the resulting p-value from Benjamini-Hochberg multiple testing correction. Most relevant pathways are indicated in bold. Table adapted from WebGestalt analysis.

genes bearing tumor-only, exonic, protein-changing variants were lower than in case of all variants of the pooled tumor and the pooled monocyte samples (Supplementary Table 6). Yet, GO slim summarizing classifications showed similar bar graphs, with processes, functions and cellular components in similar orders and with similar gene numbers (Supplementary Fig. 3). In-depth GO, KEGG and TFT enrichment analysis showed the same processes, pathways and transcription factors as observed in the overall analysis (Supplementary Table 6).

Collectively, these results show that TCRγδ+ T-LGL leukemia cells contained variants affecting metabolism, signaling pathways, proliferation and apoptosis, with well-known oncogenic transcription factors as potential main drivers.

DAVID database and Ingenuity Pathway Analysis confirm affected pathways in DNA replication, damage response, and normal signaling

Since the paired analyses of the patients' tumor vs. monocyte samples resulted in low numbers of genes with tumor-specific exonic, protein-altering variants, we then further used the overall dataset in which all tumor samples were compared with monocyte control samples. For further insights in possible affected pathways we validated the WebGestalt results with DAVID database [38,39] and Qiagen's Ingenuity Pathway Analysis (IPA) [40]. Using DAVID database GO analysis the biological processes, molecular functions and cellular components that were identified were similar to what was observed in WebGestalt analysis (Table 2). However, KEGG analyses showed fewer associated pathways, and most were not significant anymore after multiple testing adjustments. Clusters of orthologous group analysis (COG) of ontology mainly showed involvement of the "cytoskeleton" as well as "DNA replication", "recombination" and "repair", as also predicted by WebGestalt analysis (Table 2).

In order to obtain more visual insights in affected pathways, additional IPA analysis was then applied. From all genes bearing tumor-specific exonic, protein-altering variants (n=2733) a total of 2706 genes were mapped and ready for analysis in IPA, which revealed as top disease and biological function "cancer", varying from solid tumors to hematologic malignancies (Fig. 3a). Additional canonical pathway analysis revealed that affected genes were involved in DNA double-strand break repair by homologous recombination and base-excision repair (BER) pathways, but also "bupropion degradation" and "granzyme B signaling" were found to be affected (Fig. 3b).

Together these analyses imply that variants causing defects in DNA damage repair pathways might play an important role in TCRγδ+ T-LGL leukemogenesis, whereas altered or defective granzyme B signaling would also be associated with dysregulated functioning of TCRγδ+ T-LGL leukemia cells.

Stringent selection with MAF <= 0.500 revealed candidate genes involved in DNA replication, transcription, DNA damage repair and immune signaling pathways

Following extensive filtering based on tumor-specific, exonic variants, as described above, we aimed to further reduce the total list of 3645 remaining variants in 2733 genes to a smaller number of the most relevant genes and variants potentially involved in the pathogenesis of TCRγδ+ T-LGL leukemia. We therefore looked into the minor allele

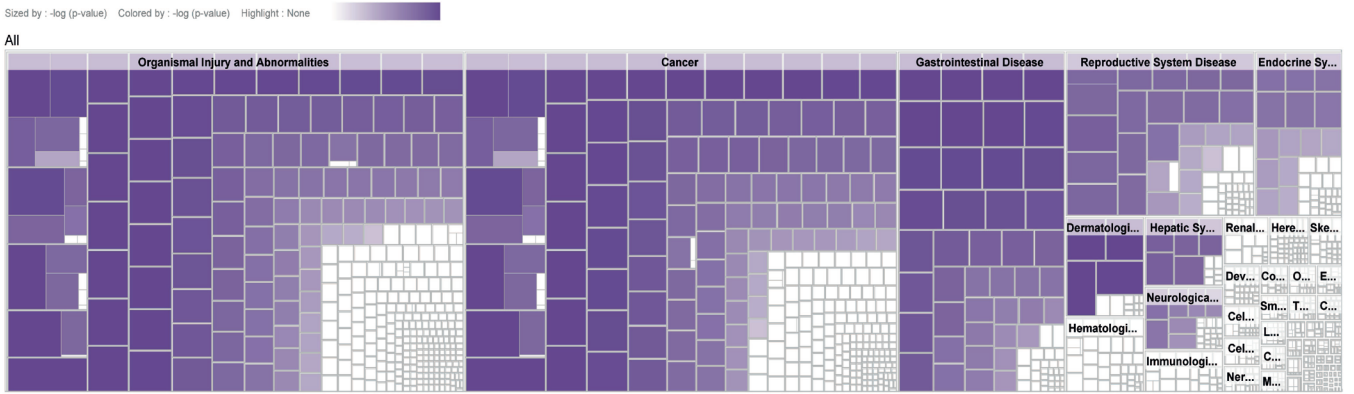
Table 2. GO, KEGG and COG ontology analysis using DAVID database

TERM	#GENES	GENES (a.o.)	ADJUSTED P*
GO – biological processes			
Biological adhesion	331	ADAM genes, APC, CCR9, CX3CR1, CD5, CD8B, FAT genes, GPR183, JAK2, KLF4, NLRC3, RASAL3, SMAD3/6, ACTN genes, CDH genes, ITGA7, ITGB4/5, ICAM1, IL12A/B/RB1, IL23R, IL27, NOTCH1, PTPN genes, RREB1, REL, SIGLEC genes, TNFSF9	1.6E-04
Anatomical structure development	919	HMGCR, ADAM genes, ATM, BCL9/10, CX3CR1, CREBBP, DDX11, POL genes, E2F4/7, JAK1/2, KLF4/10/13/15, MYCBPAP, RAB genes, SOX1/6/11/17, TBX2/5/15, TNFRSF genes, TRADD, WNT10A, DVL1, HOX genes, ITGA7, ITGB4/5, IL12, JAG1, NOTCH1/2, NCOA, NCOR2, RREB1, RUNX2	1.5E-04
Cell projection assembly	93	APC, BAG4, E2F4, HRAS, RABGAP1, RAB23, DNAH genes, MYO genes	1.7E-02
Cell-matrix adhesion	48	ADAM genes, SMAD3, ACTN1/3, BCAM, CSF1, ITGA7, ITGB4/5, MADCAM1, SIGLEC1	6.6E-02
GO – molecular functions			
ATP binding	296	ATM, ABC genes, DDX genes, HELB, LIG1/3, POLQ, NLR genes, RAD51B, RAD54L2, CHD genes, MYO genes, PIK3CA, PIK3C2G, TTL genes	1.0E-05
Calcium ion binding	160	BNIP2, NOX5, S100A genes, ACTN genes, MMP genes, NOTCH1/2, RYR1/3	1.3E-05
Scavenger receptor activity	18	CD163L1, CD5, DMBT1, ENPP1/3, LGALS3BP, LOXL3, PRG4, SCARF1, SSC5D, STAB1/2, SUSD2, TMPRSS3/13/15, TINAG, VTN	5.1E-02
Rho guanyl-nucleotide exchange factor activity	24	AKAP13, ALS2, BCR, EPS8L2, FGDS, MCF2L2, ARHGEF genes, TIAM2, DOCK1/11, DNMBP, ECT2L, ISTN2, PREX1, PLEKHG2/3/5, SPATA13, VAV3	5.8E-02
GO – cellular components			
Cytoskeleton	402	AKAP genes, ALS2, APC, ATM, BCL10, BRCA1, DDX11/60, GLI3, JAK1/2, NLRC4, ARHGAP genes, S100A14, TRADD, ANK2/3, CEP genes, CCDC genes, DVL1, MAP2/6/10, MYO genes, NFATC4	2.7E-08
Spindle	67	ATM, BNIP2, DDX11, NLRC4, ASPM, CETN1, CEP genes, NUMA1, NUMBP2, TUB genes	2.2E-02
Cytoplasm	1644	HMGCR2, AKAP genes, ADAM genes, BCL9/10, BBX, CLEC genes, CREBBP, HELB, LIG1/3, POL genes, JAK1/2, MKNK1/2, MEFV, NOX5, NLRC genes, PIF1, RAS genes, RELA, TBX5, TRADD, UBXN11/8, ANKRD genes, APOBEC genes, DTX3L, DVL1, GBA3, MRPL genes, MAPK7, MAP3K1/4/14, NFAT genes, NCOA genes, TTN, USP genes	5.1E-02
KEGG pathways			
ABC transporters	18	ABC genes, CFTR	6.7E-03
Lysosome	34	ABCA2, ATP6V0A2, ATP6AP1, 6M2A, NAGLU, SGSH, NPC1, ACP5, AP genes, ARSA/B, CTS genes, GALNS, GALC, HEXA, MCOLN, PSALP1, SORT1, SMPD1	4.6E-03
ECM-receptor interaction	22	ITG genes, LAM genes, RELN, SV genes, TNN, VTN VWF	3.4E-01 NS
Calcium signaling pathway	34	GNAL, CACN genes, ITPR1, ITPKB, PLCD3, PLCE1, PZRX genes	8.4E-01 NS
COG Ontology			
Cytoskeleton	16	ASPM, CNN, DNAH genes, DYNC2H1, MYO genes, TUBE1	1.1E-03
DNA replication, recombination and repair	27	CDC42BPA, DDX58/60, HELB, POLQ, FANCM, PIF1, UPF1, ANKRD26, DLG5, IFIH1, LAM genes, MLH3, MSH2/3, MUTYH, NBPF genes, NUDT12, SYNE1, TOP1, UTRN	3.2E-03

Analysis of same gene list as used for WebGestalt analysis (genes bearing variants only found in tumor sample,s exonic and protein-changing) in DAVID database GO, KEGG and COG ontology.
*Adjusted p-value upon Benjamini-Hochberg correction for multiple testing. Default settings used for clustering, GO, KEGG and COG ontology analysis.

frequency (MAF) of the alternative alleles (variants), as observed in our tumor dataset. We set our MAF cut-off at >=0.500, meaning that the protein-altering variant was present in at least half of all called alleles (i.e. at least heterozygous in all cases and/or homozygous in part of the cases), which resulted in 121 variants in 99 genes. Further analysis of this remaining gene list in IPA again showed "cancer" (86/99 genes associated) as the highest and most significant hit (Table 3). Involved cancers ranged from hematological malignancies to solid tumors, of mainly the thyroid gland, brain and gut, but also involved skin cancers such as melanoma. Most genes were associated with these diseases (>50/99 genes associated), and defects in the *CREB3L2*, *HAVCR1*, *NCOR2*, and *VEGFC* genes were commonly shared between the different types of cancer. We further investigated these genes and discovered that all of our patients were homozygous for

a Heatmap top diseases and bio functions



b Canonical pathways

Pathway	#Genes	Genes	Ratio	P-value
DNA double-strand break repair by homologous recombination	6	ATM, BRCA1, GEN1, LIG1, POLA1, RAD52	6/14 (0.429)	4.76E-03
Chondroitin sulfate degradation	6	ARSB, GALNS, GM2A, HEXA, HYAL4, MGEA5	6/16 (0.375)	1.02E-02
Protein kinase A signaling	65	ADCY genes, AKAP genes, ANAPC1/5, APEX1, CAMK2B, CNG genes, CREB3L4, CREBBP, FLN genes, GLI3, ITPR1, MYH4, MYL genes, NFAT5, NFATC1/4, NFKBIE, PDE genes, PLCE1, PLCH2, PRKD1, PTCH1, PTPN genes, RELA, RYR1/3, SIRPA, SMAD3, TCF7L1, TTN, TULP2	65/392 (0.166)	1.14E-02
BER pathway	5	APEX1, LIG1, LIG3, PARP1, POLG	5/12 (0.417)	1.15E-02

Figure 3. Functional classification of affected processes and canonical pathways in IPA.
All TCRγδ+ T-LGL leukemia-only exonic, protein-altering variants (excluding synonymous or nonsense variants and genes with a generally high SNP density score) are shown based on Ingenuity Pathway Analysis (IPA) (**a**). Color coding is based on p-values, which in IPA are calculated using the right-tailed Fisher exact test, based on the association with a function or pathway in the IPA Knowledgebase. This measures the likelihood that the association between a gene set and a given process or (canonical) pathway is due to random chance (i.e. the smaller the p-value, the less likely that the observed association is random, and the more significant the association is). Most important genes described in the affected pathways (**b**). The p-values indicate the relation between the affected gene and pathway based on the Fischer exact test calculated by IPA.

Table 3. IPA analysis of a gene list with a MAF cut-off at 0.500

Top diseases / biological functions	P-value range	Ratio (%)
Cancer*	4.86E-02 – 2.74E-51	86/99 (86.6%)
Neurological disease	4.32E-02 – 2.74E-51	61/99 (61.6%)
Organismal injuries and abnormalities	4.86E-02 – 2.74E-51	87/99 (87.8%)
Endocrine system disorders	2.43E-02 – 2.85E-07	47/99 (47.4%)
Gastrointestinal disease	4.90E-02 – 5.83E-06	76/99 (76.7%)

*Cancer types included colon cancer, brain cancer, squamous cell carcinoma, thymic carcinoid tumor.

the alternative allele of *NCOR2* and *VEGFC* (MAF=1.000), meaning that all tumor cells from our leukemic patients shared this nonsynonymous variant, which was not present in their monocyte control sample. We then checked the original gene list whether there were more variants with MAF=1.000, which showed 18 other homozygous gene variants in only the leukemic tumor cells in all patients (Supplementary Table 7). Among these genes were the genes as identified from IPA (Table 4), but also some other candidate genes of potential interest. One of these concerned the *KLRF1* gene, encoding an NK cell-related receptor that is expressed on TCRγδ+ T cells [47]. Also striking was the complete deletion of the whole *NUDT11* gene on the X-chromosome (Supplementary Table 7), which is an important phosphohydrolase involved in the inositol phosphate metabolism pathway [33,34,48].

Gene expression analysis of candidate variant genes identified in TCRγδ+ T-LGL leukemia

Since we, previously performed gene expression profiling of TCRγδ+ T-LGL leukemia cells versus normal TCRγδ+ T cell effector and effector memory cells [13], we then checked the expression levels of the above described genes (Table 4). Even though not all genes could be traced within the gene expression profiling data set, occasionally other gene family members or closely related genes did show altered expression levels compared to healthy cells. The *NCOR, HAVCR2*, *KLRF1*, *NUDT6*, *NUDT9*, and *CREB*-related genes were all found to be down-regulated (Table 4). Additionally, since we repeatedly observed the same transcription factors being associated with the gene set upon transcription factor target enrichment analysis, we also looked into expression levels of the *E12, LEF1, P53, TCF1, STAT3* and *STAT5B* transcription factors (Supplementary Table 6, highlighted transcription factors).

Altogether, these results suggest that defects in DNA replication and transcription (*NCOR2, CREB3L2, TCF3, TCF7, LEF1*), DNA damage repair (*TP53*) and normal signaling

Table 4. Candidate genes overlapping between different cancer types after analysis and filtering on leukemia-only, exonic alleles with MAF>=0.500

Gene	Disease and process association	p-value	Variant type*	AF**	DNA^	Protein$	rs ID#	Prediction@	GEP LGL vs. effector**	GEP LGL vs. effector memory**
NCOR2	Thymic carcinoid tumor, arrest in cell cycle progression of lymphoma cells	4.4E-03	Non-fs ins	1.000	5487_5488ins AGCAGCGGC	1830delinsSSGG	rs143952466	COSM 1359887 Large intestine, Breast	NCOR1 -2.292 NCOA3 +1.786 NCOA5 -1.706	NCOA5 -1.637 NCOR1 -2.318
VEGFC	Metastasis, movement of malignant tumor	1.3E-02	Fs del	1.000	1256_1258del TCA	419_420delS	rs5864401	-	VEGFA +2.854	VEGFA +3.851
HAVCR1	Melanoma, Skin cancer, Atopic disease	9.5E-05 1.3E-05 4.4E-03	Non-fs del	0.636	473_478delTTGG AACAGTCGTCA	158_163delLEQSS	rs14102387	COSM 1619912 Liver	HAVCR2 -2.363	HAVCR2 -3.997
CREB3L2	Liver lesion, Liver cancer	2.6E-07 4.6E-04	Non-fs del	0.727	299_301delTGG	100_101delW	rs66593747	COSM 1579673 CNS	CREB5 +1.910 CREBZF -2.203	CREBZF -2.558 CREBL2 -2.554
KLRF1	Lipid metabolism, molecular transport, small molecule biochemistry	2.4E-02	Non-fs del	1.000	223_225delTGT	75_75delC	rs33911869	-	KLRD1 -2.888	KLRF1 -7.750
NUDT11	D-myo-inositol-(1,4,5,6)-Tetrakisphosphate biosynthesis	2.4E-02	Fs del	1.000	Whole gene	Whole gene	-	COSM 127214 CNS, upper aerodigestive tract, breast, skin, prostate, urinary tract	NUDT19 -1.778 NUDT9 -1.958	NUDT4 +1.576 NUDT6 -2.233 NUDT9 -1.762

*Variant type non-frameshift (fs) insertion (ins) or deletion (del).

**Allele frequency as found in the total patient dataset among tumor-specific variants.

^Variant as identified at DNA level.

$Effect of variant on amino acid and protein level.

#rs ID based on dbSNP version 138 [42].

@Effect of variant predicted with COSMIC database [32].

**Gene expression level (GEP) of the LGL cases versus sorted effector and effector memory TCRγδ+ T cells from healthy donors as published in [13].

pathways (*KLRF1, VEGFC, HAVCR1, NUDT11, STAT3, STAT5B*) both at the genetic and mRNA expression level could be involved in TCRγδ+ T-LGL leukemogenesis.

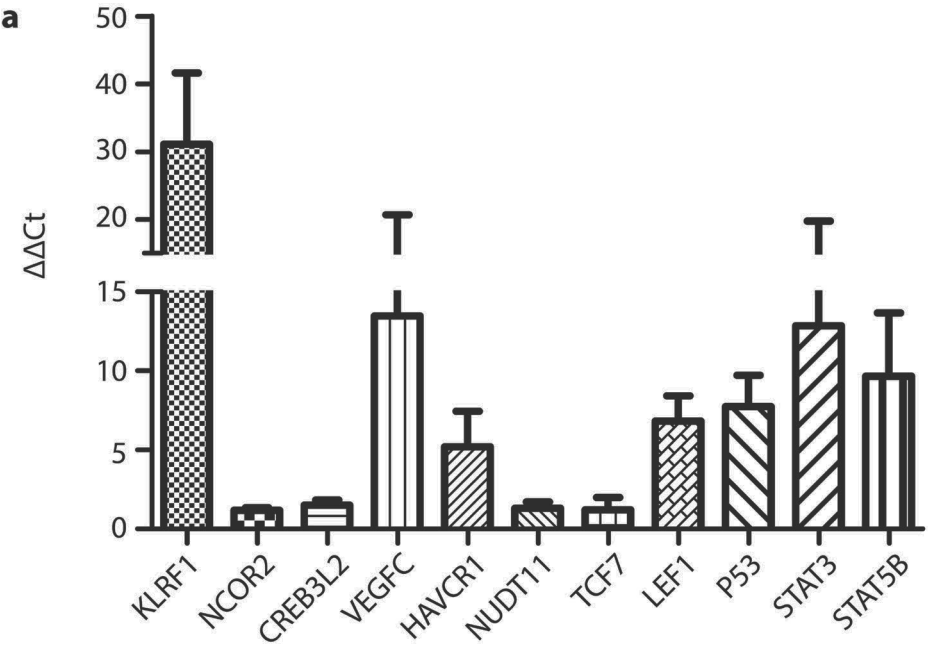
Evaluation of transcript levels of candidate variant genes in TCRαβ+ T-LGL leukemia type samples

As our previous gene expression profiling data from TCRγδ+ T-LGL leukemia patients provided a validation dataset for the identified candidate variants, cDNA RQ-PCR analysis of clinically and phenotypically similar CD8+TCRαβ+ T-LGL leukemia patient cells could provide further insight into the role of these candidate genes in this disease type. In the RQ-PCR analysis we also included the factors as identified from transcription factor enrichment analysis (Supplementary Table 6), using a total of 9 healthy controls and 36 CD8+TCRαβ+ T-LGL leukemia samples. Gene expression was adjusted using the *ABL* housekeeping gene (ΔCt), and calculated as fold change (FC). Relative mRNA expression in patients was corrected with the FC of the corresponding genes as identified from healthy control cells (ΔΔCt). *KLRF1* (>30-fold), and *VEGFC* (>10-fold) were found to be up-regulated in LGL patients when compared to the healthy control T cells (Fig. 4). Furthermore, even though *STAT3* and *STAT5B* did not contain protein-altering variants, their expression profiles did seem to be altered, possibly driving other oncogenic genes and processes as well, as identified before [14,15]. *HAVCR1*, *LEF1* and *TP53* were all up-regulated, albeit to a lesser extent. Expression levels of *NCOR2*, *CREB3L2*, *NUDT11* and *TCF7* did not clearly differ compared with healthy control T cells (Fig. 4).

Sanger sequencing confirms VEGFC and HAVCR1 gene variants in both TCRγδ+ and CD8+TCRαβ+ T-LGL leukemia, whereas NUDT11 deletion is only seen in TCRγδ+ T-LGL leukemia

Based on the above analyses in different databases, we selected the most important candidate genes that were present in almost all leukemia patient samples, but not in the control monocyte samples (Table 4), including genes encoding transcription factors that could act as main driving factors for the mutation-bearing genes. In order to validate the presence of these variants we developed primers for Sanger sequencing (Supplementary Table 3). Sanger sequencing-based validation also included CD4+TCRαβ+, CD8+TCRαβ+ and TCRγδ+ T cell DNA from sorted cells of five healthy donors. Furthermore, we also checked the presence of these variants in CD8+TCRαβ+ T-LGL leukemia, in view of the clear clinical overlap between these diseases and the intrinsic functional similarities of their cell types of origin. The *VEGFC* variant was found in 100% of the CD8+TCRαβ+ T-LGL leukemia patients, and in 75% of our TCRγδ+ T-LGL leukemia cohort. Also the *HAVCR1* variant was found in high percentages of both CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia cases (Fig. 5). However,

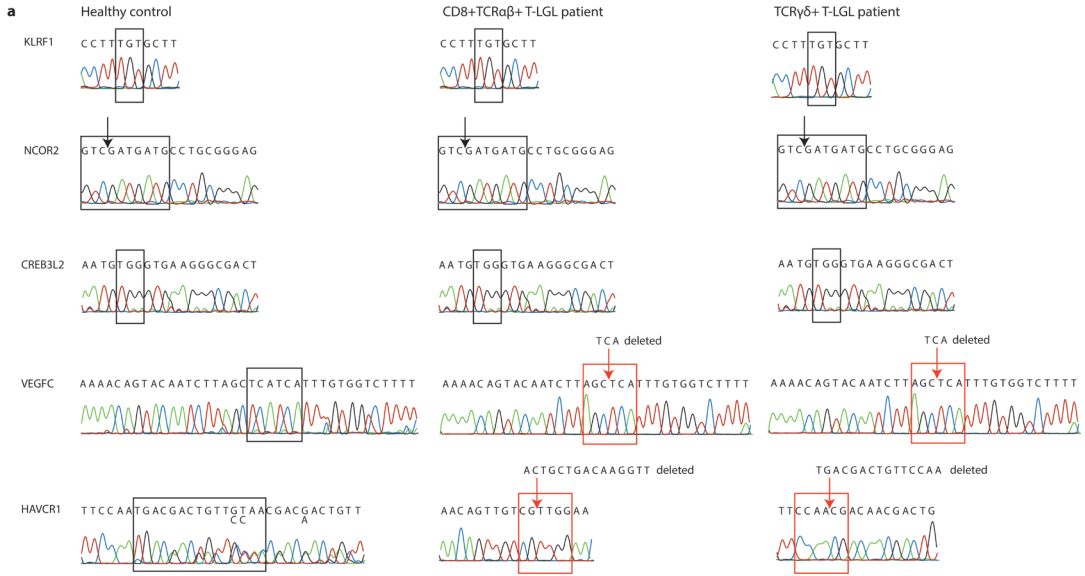
a



b

Gene	Mean FC HC	Mean FC PT	Mean ΔΔCt (PT vs. HC)
KLRF1	0.001066	0.03315	31.11174
NCOR2	4.25025	4.9903064	1.174119
CREB3L2	1.76358	2.6313927	1.492068
VEGFC	0.0893634	1.2032619	13.46482
HAVCR1	0.1329596	0.6907585	5.195253
NUDT11	0.2353806	0.3096345	1.315463
TCF7	21.336403	25.376824	1.189367
LEF1	0.3070765	2.0953332	6.82349
P53	0.0782046	0.6064637	7.754832
STAT3	4.1951724	53.885386	12.84462
STAT5B	2.397618	23.153356	9.656816

Figure 4. RQ-PCR validation of the effect of the identified variants in candidate genes.
Relative mRNA expression of genes containing protein-altering variants and transcription factors as identified through transcription factor target enrichment analysis, following normalization to the *ABL* housekeeping gene (=ΔCt). ΔCt values were calculated into fold change values. Average fold changes from healthy controls (N=9) were used to calculate patient (N=36) to healthy control ratios translating into fold change values (=ΔΔCt) (**a**). Bar graphs are indicated with mean and SD (**b**). Mean absolute values of the fold change gene expression for genes shown in (**a**). Abbreviations used: FC, fold change; HC, healthy control; PT, patient.



b Incidence of variants in HC, CD8+ PT & GD PT (table)

Gene	Healthy CD4+TCRαβ+ T cells with variant (%)*	Healthy CD8+TCRαβ+ T cells with variant (%)*	Healthy TCRγδ+ T cells with variant (%)*	CD8+TCRαβ+ T-LGL patient cells with variant (%)	TCRγδ+ T-LGL patient cells with variant (%)
KLRF1	0/5 (0%)	0/5 (0%)	0/5 (0%)	0/19 (0%)	0/44 (0%)
NCOR2	0/5 (0%)	0/5 (0%)	0/5 (0%)	0/19 (0%)	0/44 (0%)
CREB3L2	0/5 (0%)	0/5 (0%)	0/5 (0%)	0/19 (0%)	0/44 (0%)
VEGFC	1/5 (20%)	4/5 (80%)	5/5 (100%)	19/19 (100%)	34/44 (77%)
HAVCR1	1/5 (20%)	4/5 (80%)	2/5 (40%)	14/19 (74%)	33/44 (75%)
NUDT11**	0/5 (0%)	0/5 (0%)	0/5 (0%)	0/19 (0%)	24/44 (54%)

*CD4+TCRαβ+, CD8+TCRαβ+ and TCRγδ+ T cells sorted from 5 healthy donors.
**Variant is absence of complete NUDT11 gene.
Absolute numbers and frequencies of patient and healthy control cells containing, the variants.

Figure 5. Sanger sequencing validation of identified variants in candidate genes.
Representative peak patterns of wildtype and/or variant sequences in healthy control T cell fractions and sorted CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia cell patient DNA. Black boxes indicate the template DNA, red boxes indicate the variant. Arrows pinpoint the site of insertion or deletion. Nucleotide letters below consensus sequence indicate the original nucleotides at that position (**a**). Frequencies of cases showing presence of the respective variants in DNA from CD4+ TCRαβ+ (N=5), CD8+TCRαβ+ (N=5) and TCRγδ+ (N=5) healthy control T cells versus DNA from CD8+TCRαβ+ (N=19) and TCRγδ+ (N=44) T-LGL leukemia patients (**b**).

both variants were also identified in healthy control CD8+TCRαβ+ and TCRγδ+ T cells (Fig. 5). The variants identified in the *KLRF1*, *NCOR2* and *CREB3L2* genes could not be confirmed via Sanger sequencing (Fig. 5). The *NUDT11* sequence could not be identified in 24 out of 44 TCRγδ+ T-LGL leukemia samples, but the *NUDT11* gene was normally present in all healthy controls and CD8+TCRαβ+ T-LGL leukemia cells (Fig. 5).

DISCUSSION

Recently, Lamy *et al.* reviewed the LGL leukemia pathogenesis and treatment developments [49], highlighting the contribution of *STAT* gain-of-function mutations, which were first described in 2012-2013 [14,15] and which are currently included in the latest WHO classification of chronic leukemias [1]. Another possible mechanism involved in the pathogenesis of LGL leukemia could be Fas/FasL apoptosis resistance [50,51], related to up-regulated Ras activity and thus down-regulated Fas-mediated apoptosis [52]. Furthermore, dysregulation in the PI3K/Akt and NFκB pathways has also been described and reviewed by Lamy *et al.,* which was mostly associated with altered gene and protein expression levels. Despite all these recent advances, pathogenic mechanisms have only been described for the CD4+TCRαβ+, CD8+TCRαβ+ T-LGL and NK-LGL leukemia types, whereas the TCRγδ+ variant and its possible genetic aberrancies were largely neglected. Despite overlapping clinical presentations with CD8+TCRαβ+ T-LGL leukemia, the pathogenesis of TCRγδ+ T-LGL leukemia therefor remains largely elusive, making the TCRγδ+ T cell variant of chronic T-LGL leukemia the most poorly understood form of this disease.

In our previous study we identified differential gene expression profiles between normal TCRγδ+ T cells and aberrant TCRγδ+ T-LGL leukemia cells at the level of dysregulated signaling, disturbed proliferation and apoptosis, and increased survival [13,53]. Here we attempted to investigate the underlying genetic aberrancies contributing to TCRγδ+ T-LGL leukemia by means of whole exome sequencing technology. In order to increase the relevance of our findings for leukemia cells, we included healthy monocyte cell populations from our patients as normal hematopoietic control cells, to exclude possible germline or individual-specific polymorphic variants. With this approach we were able to identify homozygous variants that are present in all leukemia samples, but not in control monocyte cell populations. Using an in-depth analysis including several distinct databases, we initially identified 99 candidate genes, which could be narrowed down to six major candidate genes that are affected in a statistically significant way.

In our top list of candidate genes involved in TCRγδ+ T-LGL leukemogenesis, the gene encoding the *CREB3L2* transcription factor was found to contain a homozygous

non-frameshift deletion. *CREB3L2* targets (proto-) oncogenes such as *cFOS*, but also neuropeptides, and is known to be associated with various types of cancer according to the COSMIC database [cancer.sanger.ac.uk, 32]. Gene expression profiling data only provided information on other *CREB* isoforms, but using RQ-PCR we could show a slight up-regulation of this gene. This could indicate a slight enhancement in driving (proto-) oncogene activation, thus contributing to the leukemogenesis. On the other hand, the *NCOR2* repressive transcriptional regulator for e.g. *cFOS*, but also other oncogenes [54,55], contained a non-frameshift insertion leading to similar expression levels as observed in healthy controls. This nuclear corepressor is involved in transcriptional co-regulation through recruitment of histone deacetylases, and therefore in down-regulation of gene expression. The *NCOR2* gene has been described earlier to be involved in human colon and rectal carcinomas [56]. Although, direct (functional) associations have not been described, a slight up-regulation of *CREB3L2* transcription together with a relatively normal *NCOR2* expression level of could contribute to enhanced (proto-) oncogene activation and thus to the proliferative capacity of leukemic cells. In spite of the *NCOR2* and *CREB3L2* variants being identified by WES, they could not be validated by means of Sanger sequencing in CD8+TCRαβ+ or TCRγδ+ T-LGL leukemia samples. This could be explained by admixture of normal cells in the leukemia cell population, since there is no specific T-LGL leukemia marker that allows sorting the leukemia cells to purity. Besides, WES technology provides a higher resolution than Sanger sequencing.

Other candidate genes from our top list of gene variants are involved in immune cell activation and cytotoxicity and cytokine production. The *HAVCR1* gene, also known as *TIM-1* or *KIM-1*, is involved in kidney injury and plays a critical role in the regulation of immune cell activation in response to virus infections [57]. Furthermore, IPA analysis showed associations of this gene with the development of melanoma and skin cancer, indicating an important role of its gene product in skin physiology. This indicates the cells are activated and ready to respond to antigens or other stimuli. Activated TCRγδ+ T cells are also known to express NK cell like receptors [58], especially in the context of tumor immunity [59] and virus infections such as EBV [60]. The *HAVCR1* variant was validated to be present in T-LGL leukemia patients, although it was also seen in healthy CD8+TCRαβ+ and TCRγδ+ T cells, but not in CD4+TCRαβ+ T cells. CD4+TCRαβ+ T cells might follow a different activation to T-LGL leukemogenesis [61]. KLRF1 is also an NK cell marker, and in particular an inhibitory molecule involved in NK cell and monocyte activation, but also in other lymphocyte subsets [62]. KLRF1 has also been found to be associated with melanoma cell proliferation [63]. The *KLRF1* gene displayed a dramatic difference in gene expression between healthy control and LGL leukemia cells, indicating that up-regulation of an inhibitory molecule might lead to dysregulated normal functioning and regulation of the immune response, which

could be driving the cytopenias that are clinically associated with T-LGL leukemia [4,6]. However, also the *KLRF1* variant could not be validated with Sanger sequencing, again possibly due to admixture of non-leukemic cells and the sequencing depth. Furthermore, from our analyses we also observed frameshift deletions in *VEGFC* and *NUDT11* genes. These variants potentially cause more downstream problems due to the complete change in protein through these frameshift deletions. *VEGFC* is known to be important in lymphangiogenesis [64] and plays major roles in this process in gastric cancer [65] and metastatic breast cancer [66]. Our study showed an increase in the *VEGFC* gene expression level in LGL leukemia samples when compared to healthy control cells, suggestive of altered lymphangiogenic and metastatic effects. The frameshift deletion was also validated with Sanger sequencing in both leukemic and healthy control CD8+TCRαβ+ and TCRγδ+ T cells, but not in healthy control CD4+TCRαβ+ T cells, thus induced high increases in *VEGFC* gene expression, contributing to the leukemic character. Finally, also the *NUDT11* gene contained a frameshift deletion; which in fact was a deletion of the entire gene. We also observed relatively similar gene expression as observed from healthy controls and patients. Normally, the *NUDT11* gene is associated with phosphohydrolase activity, playing a role in signal transduction according to GeneCards database [67]. Particularly in TCRγδ+ T-LGL leukemia patients this gene was absent, indicating that it might underlie processes specifically for TCRγδ+ T-LGL leukemia cells. However, further evaluation of the exact role of this gene requires more functional studies in which the gene is for example silenced by means CRISPR-Cas9 technology.

Besides genes containing protein-altering variants, further enrichment analysis using WebGestalt also identified specific transcription factors that could act upstream of affected genes. These included the transcription factors TCF7, LEF1, TP53, and the well-known STAT3 and STAT5B factors. TCF7 and LEF1 are the main downstream signal transduction molecules in the Wnt signaling pathway, up-regulation of which is associated with colon carcinoma [68]. TP53 is a well-known important transcription factor involved in tumor suppression mechanisms [69]. Finally, the "central hub" in LGL proliferations STAT3 was also identified as main driving transcription factor of our gene list [49]. However, we did not observe the described mutations in *STAT3* and *STAT5B* genes as identified before in CD8+TCRαβ+ T-LGL leukemia [14,15], indicating that these mutations are probably not as common in TCRγδ+ T-LGL leukemia, as they are in the CD8+TCRαβ+ variant. However, since we were limited in the number of samples available for our current WES analysis, we cannot fully exclude the possibility that *STAT* mutations are present in TCRγδ+ T-LGL leukemia variants, but only at a low frequency. In order to prove this, a larger cohort of patients is needed, and preferably the specific leukemia cell populations need to be sorted to purity to rule out any other role of remaining normal

TCRγδ+ T cells in the sample. Since there is no clear LGL leukemia-related marker [4], it remains difficult to distinguish the remaining normal TCRγδ+ T cells from aberrant leukemic cells. Future experiments should include more LGL leukemia patients from which the leukemia cells need to be sorted based on a combination of markers also including a.o. CD56 and CD57, in order to perform WES, Sanger sequencing and RQ-PCR validations. Furthermore, extra cells should be preserved for functional assays, to fully understand the potential of the variants identified.

In summary, so far our data show leukemia-specific homozygous (non-)frameshift insertions and deletions in six genes involved in transcription, immune cell activation and metastasis. These genes mainly exhibited increased gene expression levels when compared to healthy controls without having the gene variants. These variants could point to underlying molecular and/or genetic defects in the development of TCRγδ+ T-LGL leukemia specifically, but perhaps also CD8+TCRαβ+ T-LGL leukemia. However, further (functional) studies are required to disclose the impact of these gene defects on T-LGL leukemogenesis.

ACKNOWLEDGEMENTS

This work was performed as part of a PhD project in the MolMed postgraduate school. AWL was financially supported by an unrestricted research grant from Roche-Genentech.

REFERENCES

1. Swerdlow, S.H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127(20)**, 2375-2390 (2016).
2. Rose, M.G. & Berliner, N. T-cell large granular lymphocyte leukemia and related disorders. *The Oncologist* **9**, 247-258 (2004).
3. Sokol, L. & Loughran, T.P. Large granular lymphocyte leukemia. *The Oncologist* **11**, 263-273 (2006).
4. Sandberg, Y. *et al.* Lack of common TCRA and TCRB clonotypes in CD8+/TCRαβ+ T-cell large granular lymphocyte leukemia: a review on the role of antigenic selection in the immunopathogenesis of CD8+ T-LGL. *Blood Cancer J.* **4**, e172 (2014).
5. Langerak, A.W., Sandberg. Y. & van Dongen, J.J.M. Spectrum of T-large granular lymphocyte lymphoproliferations: ranging from expanded activated effector T cells to T-cell leukaemia. *Br. J. Haematol.* **123**, 561-562 (2003).
6. Sandberg, Y. *et al.* TCRγδ+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRγδ+ T-cells. *Leukemia* **20**, 505-513 (2006).

7. Lamy, T. & Loughran, T.P. Clinical features of large granular lymphocyte leukemia. *Semin. Hematol.* **40**, 185-195 (2003).
8. Lamy, T. & Loughran, T.P. Current concepts: large granular lymphocyte leukemia. *Blood Rev.* **13**, 230-240 (1999).
9. Semenzato, G., Zambello, R., Starkebaum, G., Oshimi, K. & Loughran, T.P. The lymphoproliferative disease of granular lymphocytes: updated criteria for diagnosis. *Blood* **89**, 256-260 (1997).
10. Van Dongen, J.J.M. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* **17**, 2257-2317 (2003).
11. Clemente, M.J. *et al.* Clonal drift demonstrates unexpected dynamics of the T-cell repertoire in T-large granular lymphocyte leukemia. *Blood* **118 (16)**, 4384-4393 (2011).
12. Wlodarski, M.W. *et al.* Phenotypic differences between healthy effector CTL and leukemic LGL cells support the notion of antigen-triggered clonal transformation in T-LGL leukemia. *J. Leukoc. Biol.* **83 (3)**, 589-601 (2007).
13. Kallemeijn, M.J. *et al.* Dysregulated signaling, proliferation and apoptosis impact on the pathogenesis of TCRγδ+ T cell large granular lymphocyte leukemia. *PLoS ONE* **12(4)**, e0175670 (2017).
14. Koskela, H.L.M. *et al.* Somatic STAT3 mutations in large granular lymphocyte leukemia. *N. Engl. J. Med.* **366**, 1905 – 1913 (2012).
15. Rajala, H.L.M. *et al.* Discovery of somatic STAT5B mutations in large granular lymphocytic leukemia. *Blood* **121**, 4541 – 4550 (2013).
16. Küçük, C. *et al.* Activating mutations of STAT5B and STAT3 in lymphomas derived from γδ-T or NK cells. *N. Comms.* **6**, 6025 (2015).
17. Yabe, M. *et al.* Distinguishing between hepatosplenic T-cell lymphoma and γδ T-cell large granular lymphocytic leukemia: a clinicopathologic, immunophenotypic, and molecular analysis. *Am. J. Surg. Pathol.* **41(1)**, 82-93 (2017).
18. Chen, Y.H. *et al.* Clinical, morphologic, immunophenotypic, and molecular cytogenetic assessment of CD4-/CD8-γδ T-cell large granular lymphocytic leukemia. *Am. J. Clin. Pathol.* **136(2)**, 289-299 (2011).
19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25(14)**, 1754-1760 (2009).
20. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303 (2010).
21. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38(16)**, e164 (2010).
22. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.* **30**, 661-708 (2015).
23. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Hum. Mut.* **32**, 894-899 (2011).

24. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073-1081 (2009).
25. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7(4)**, 248-249 (2010).
26. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D.. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11(4)**, 361-362 (2014).
27. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 1- 14 (2011).
28. Reva, B.A., Antipin, Y.A. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol*. **8(11)**, R232 (2007).
29. Shibab, H.A. *et al.* Predicting the functional, molecular and phenotypic consequences of amino acid substitutions using Hidden Markov Models. *Hum. Mut.* **34**, 57-65 (2013).
30. Cooper, G.M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods* **7**, 250- 251 (2010).
31. Cooper, G.M., *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15(7)**, 901-913 (2005).
32. Forbes, S.A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45(1)**, 777-783 (2016).
33. Milacic, M. *et al.* Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)* **4(4)**, 1180-1211 (2012).
34. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res.* **44(1)**, 481-487 (2016).
35. Rstudio Team. Rstudio: Integrated Development for R. RStudio, Inc., Boston, MA.
36. Zhang, B., Kirov, S.A. & Snoddy, J.R. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, 741-748 (2005).
37. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based Gene SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, 77-83 (2013).
38. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systemic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* **4(1)**, 44-57 (2009).
39. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37(1)**, 1-13 (2009).
40. Krämer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30(4)**, 523-530 (2014).
41. Maglott, D., Ostell, J., Pruitt, K.D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. **35**, 26-31 (2007).
42. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29(1)**, 308-311 (2001).
43. Untergasser, A. *et al.* Primer3 – new capabilities and interfaces. *Nucleic Acids Res.* **40(15)**, e115 (2012).

44. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23(10)**, 1289-1291 (2007).
45. Beillard, E. *et al.* Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) – a Europe against cancer program. *Leukemia* **17(12)**, 2474-2486 (2003).
46. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
47. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13(2)**, 88-100 (2013).
48. Hidaka, K. *et al.* An adjacent pair of human NUDT genes on chromosome X is preferentially expressed in testis and encodes two new isoforms of diphosphoinositol polyphosphate phosphohydrolase. *J. Biol. Chem.* **277(36)**, 32730-32738 (2002).
49. Lamy, T., Moignet, A. & Loughran, T.P. Jr. LGL leukemia: from pathogenesis to treatment. *Blood* **129(9)**, 1082-1094 (2017).
50. Liu, J.H. *et al.* Blockade of Fas-dependent apoptosis by soluble Fas in LGL leukemia. *Blood* **100(4)**, 1449-1453 (2002).
51. Yang, J. *et al.* Antigen activation and impaired Fas-induced death-inducing signaling complex formation in T-large-granular lymphocyte leukemia. *Blood* **111(3)**, 1610-1616 (2008).
52. LeBlanc, F., Zhang, D., Liu, X. & Loughran, T.P. Jr. Large granular lymphocyte leukemia: from dysregulated pathways to therapeutic targets. *Future Oncol.* **8(7)**, 787-801 (2012).
53. Zhang, R. *et al.* Network model of survival signaling in large granular lymphocyte leukemia. *Proc. Natl. Acad. Sci. USA* **105(42)**, 16308-16313 (2008).
54. Wong, C.W. & Privalsky, M.L. Components of the SMRT corepressor complex exhibit distinctive interactions with the POZ domain oncoproteins PLZF, PLXF-RARalpha and BCL-6. *J. Biol. Chem.* **273(42)**, 27695-27702 (1998).
55. Huynh, K.D., Fischle, W., Verdin, E. & Bardwell, V.J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev*. **14(14)**, 1810-1823 (2000).
56. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487(1707)**, 330-337 (2012).
57. Tami, C. *et al.* Immunoglobulin A (IgA) is a natural ligand of hepatitis A virus cellular receptor 1 (HAVCR1), and the association of IgA with HAVCR1 enhances virus-receptor interactions. *J. Virol.* **81(7)**, 3437-3446 (2007).
58. De Libero, G. Control of gammadelta T cells by NK receptors. *Microbes Infect.* **1(3)**, 263-267 (1999).
59. Moris, A., Rothenfusser, S., Meuer, E., Hangretinger, R. & Fisch, P. Role of γδ T cells in tumor immunity and their control by NK receptors. *Microbes Infect.* **1(3)**, 227-234 (1999).
60. Zhang, Y., Ohyasaki, J.H., Shimizu, N. & Ohyasaki, K. Aberrant expression of NK cell receptors in Epstein-Barr virus-positive γδ T-cell lymphoproliferative disorders. *Hematology* **15(1)**, 43-47 (2010).

61. Garrido, P. *et al.* Monoclonal TCR-Vbeta13.1+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis: evidence for an antigen-driven chronic T-cell stimulation origin. *Blood* **109**, 4890-4898 (2007).
62. Roda-Navarro, P., Arce, I., Renedo, M., Montgomery, K., Kucherlapati, R. & Fernández-Ruiz, E. Human KLRF1, a novel member of the killer cell lectin-like receptor gene family: molecular characterization, genomic structure, physical mapping to the NK gene complex and expression analysis. *Eur. J. Immunol.* **30**, 568-576 (2000).
63. Thies, A., Nugel, D., Pfüller, U., Moll, I. & Schumacher, U. Influence of mistletoe lectins and cytokines induced by them on cell proliferation of human melanoma cells in vitro. *Toxicology* **207(1)**, 105-116 (2005).
64. Jeltsch, M. *et al.* Hyperplasia of lymphatic vessels in VEGF-C transgenic mice. *Science* **276(5317)**, 1423-1425 (1997).
65. Yonemura, Y. *et al.* Role of VEGF-C and VEGF-D in lymphangiogenesis in gastric cancer. *Int. J. Clin. Oncol.* **10(5)**, 318-327 (2005).
66. Skobe, M. *et al.* Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nat. Med*. **7(2)**, 192-198 (2001).
67. Safran, M. *et al.* GeneCards Version 3: the human genome integrator. *Database (Oxford)* **2010**, 1-16 (2010).
68. Hrckulak, D., Kolar, M., Strnad, H. & Korinek, V. TCF/LEF transcription factors: an update from the internet resources. *Cancers (Basel)* **8(7)**, 70 (2016).
69. Fischer, M., Steiner, L. & Engeland, K. The transcription factor p53: not a repressor, solely an activator. *Cell Cycle* **13(19)**, 3037-3058 (2014).

SUPPLEMENTAL DATA

Supplementary Table 1. Patient sample characteristics

Patient	Gender	Age at diagnosis (years)	Immunophenotype	Tumor load (%leukocytes)	TCR phenotype‡	Samples*
LGL058	M	38	CD3/CD4neg/CD2/CD5neg/CD7neg/CD56/CD57	6.9	Vγ9/Vδ2	Paired
LGL063	M	53	CD3/4neg/8/2/5neg/7/27neg/Vδ1	23	Vγ9/Vδ1	Paired
LGL067	F	70	CD3/4neg/8/5/7/57/27neg/Vγ9/Vδ2	52	Vγ9/Vδ2	Paired
LGL091	F	80	CD3/4neg/8/2/5/7/56/57/27neg/Vδ2	34	Vγ9/Vδ2	Tumor only
LGL092	F	55	CD3/4neg/8neg/2/5/7/Vδ2	23	Vγ9/Vδ2	Tumor only
LGL095	M	73	CD3/4neg/8/2/5neg/7/Vγ9/Vδ1neg/Vδ2neg	65	Vγ9/Vδ3	Paired
LGL113	M	69	CD3/4neg/8neg/2/5/7/27neg/Vγ9/Vδ2	25	Vγ9/Vδ2	Paired
LGL125	F	41	CD3/4neg/8/2/5/7/56neg/ Vδ1	12	Vγ9/Vδ1	Tumor only
LGL129	F	55	CD3/4/8/2/5/7/56/ Vγ9/Vδ1	ND	Vγ9/Vδ1	Tumor only
LGL132	F	65	CD3/4neg/8/2/5/7/56/57/27neg/Vδ2	ND	Vγ9/Vδ1	Paired

‡ TCR phenotype based on GeneScan and heteroduplex analysis after multiplex PCR

* Paired samples (tumor and monocyte DNA) or tumor samples only.

ND: not determined.

Supplementary Table 2. Antibody details for sorting experiments.

	PO	FITC	PerCP Cy5.5	PE	PE Cy7	APC	APC-H7
Antibody	**CD45**	**TCRVδ1**	**CD3**	**TCRVδ2**	**CD19**	**TCRαβ**	**CD14**
Clone	HI30	TS8.2	SK7	B6	SJ25C1	IP26	MO-P9
Manufacturer	Invitrogen	Thermo Scientific	BD Biosciences	BD Biosciences	BD Biosciences	e-Bioscience	BD Biosciences
Use	5 µl undiluted	5 µl undiluted	10 µl undiluted	5 µl 1:16 diluted	5 µl undiluted	5 µl undiluted	5 µl undiluted

Supplementary Table 3. Sanger sequencing primers for variant validation.

Gene	Forward 5'-3'	Reverse 5'-3'
KLRF1	ACTCAGTATGAGGACACTGGAGATCTAAA	TTGAGACATACCTGTCTGGTCTGC
NCOR2	GTCCTCGCTGGCACTCAAC	ACAGGCAGGTGTGGCACTT
CREB3L2	GAGCCTCGGGCCCAGT	CGTCATTGAAGCTGTCACTGGT
VEGFC	AAGAAGTGTGTCGTTGTGTCCCTT	TTCCATAATAGAAAATCGATGAACTGG
HAVCR1	ACTGTTCCAACGACAACGACTGT	TTGTCGTTGGAATGCTCATTGT
NUDT11	AAGTCGAAGATGCCATCAAGGT	CCCAGCTTTAGTTTCTCCAGATATTC

Gene primers were developed using Oligo 6.41 software (Molecular Biology Insights, Cascade, CO, USA), Primer Express 3.0 software (Thermo Fischer Scientific) and Primer3 software [43,44].

Supplementary Table 4. Primers and probs for RQ-PCR-based gene expression analysis

Gene	Forward 5'-3'	Reverse 5'-3'	Probe number
KLRF1	TGCCAAAAAGGAAGTTGTTCA	TTTCTTCTTGTGCCATTATTCAT	#47
NCOR2	GACCGCCTTGCCTACCTC	ACCTCCTGGGGAGAGTGG	#23
CREB3L2	CCTGAATGATCCTTTCCTCTCA	GCTGTGCTCAGCCTGGAT	#10
VEGFC	GCCCCAAACCAGTAACAATC	CTGTAAACATCCAGTTTAGACATGC	#4
HAVCR1	TGAGATTTAAGACTTGATCAGATACCA	GCCTGAAGGAAAATGAGCAG	#22
NUDT11	CGAGGACGAGGTCCTGTTAG	ACTCCCGCTTCTTCGTACAC	#63
TCF3	TTTTGTTGGGACCAGTGACC	GACAGGGTGTCCAGGAGGT	#1
LEF1	ACAGCGGAGCGGAGATTAC	TTGAAGGGGATCATCTCGTC	#10
TP53	CATTTTCACCCCACCCTTC	ACACAGGTGGCAGCAAAGTT	#7
TCF7	CCTTGATGCTAGGTTCTGGTG	CAGCTCCTGCTTCCCTGA	#3
STAT3	ACCTAGGGCGAGGGTTCA	CCTAAGGCCATGAACTTGACA	#50
STAT5B	GTCCTGGCTCACTTTGCAGT	CTCAGGGAAGCCACACTCAT	#69

Assays were designed using the Roche Universal Probe Library (Roche, Basel, Switzerland).

Supplementary Table 5. Sequencing run specs.

Sample	Tissue	Reads	Aligned reads	Coverage*	Contamination**	Variants#	TiTv ratio^	Indels$
LGL058	Monocytes	80365027	80216360	94.44	0.00494	35249	2.51	3463
	Tumor	56262604	56130666	68.16	0.0046	35050	2.46	3299
LGL063	Monocytes	75476397	75277789	89	0.00609	34580	2.51	3411
	Tumor	84478020	84360055	102.64	0.00533	34020	2.59	3280
LGL067	Monocytes	78270051	78047033	92.7	0.00635	34678	2.52	3363
	Tumor	53002768	52928866	125.16	0.02768	34142	2.49	3118
LGL091	Tumor	64678132	64530000	78.73	0.00473	34418	2.52	3214
LGL092	Monocytes+	2864525	2832750	3.06	0.04431	61225	1.77	3067
	Tumor	100392994	100200675	125.16	0.02768	34324	2.62	3263
LGL095	Monocytes	72387446	72649522	88.47	0.004543	34706	2.46	3276
	Tumor	51006338	50878656	62.15	0.01165	34038	2.52	3137
LGL113	Monocytes	71078380	70928901	85.39	0.00691	34627	2.52	3255
	Tumor	57280966	57147886	70.68	0.0061	34216	2.57	3150
LGL125	Tumor	52202900	52080131	65.21	0.00483	34780	2.44	3220
LGL129	Tumor	61645640	61482001	79.16	0.00713	34499	2.49	3287
LGL132	Monocytes	60111638	59945116	71.07	0.00515	34050	2.46	3229
	Tumor	6524452	65135621	78.68	0.00466	34261	2.44	3226

* Coverage cut-off at 50X. Sequencing is performed with 100X sequencing depth, default cut-off after sequencing set at 50X.
** PCR contamination cut-off at <5%.
# Number of variants per Caucasian sample approx. 30,000.
^ Transition/transversion ratio cut-off approx. 2.0-2.1.
$ Number of insertions and deletions per Caucasian sample approx. 3,000.
+ Sample excluded based on aberrant sequencing results (read number, coverage and TiTv ratio all too low; number of variants above average).

Supplementary Table 6. GO, KEGG and TFT enrichment analysis of paired tumor-monocyte samples.

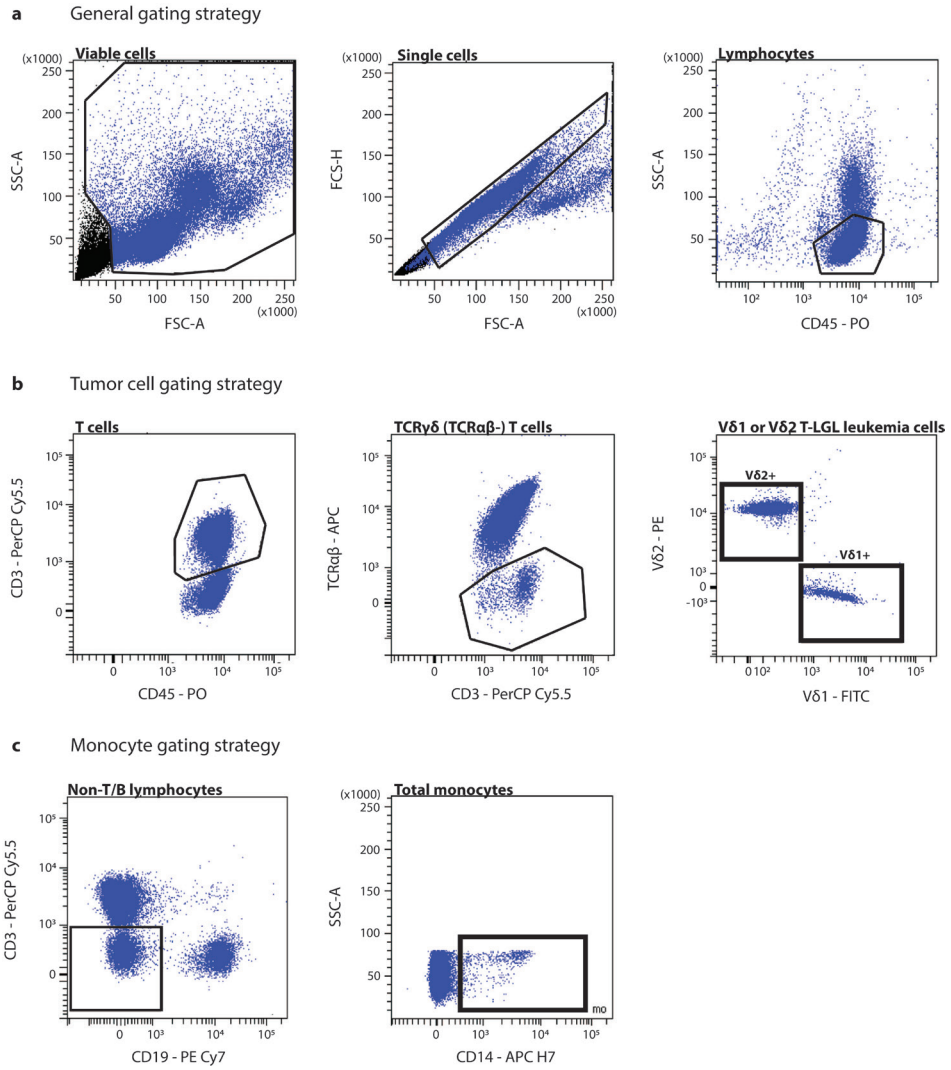
Sample	Case only+	#genes	Mapped *	GO									KEGG			TFT		
				BP	N$	adjP&	MF	N	adjP	CC	N	adjP	Pathway	N	adjP*	TF	N	adjP
LGL058	638	452	432	O-glycan processing	8	0.025	Hydrolase activity Actin binding Ca^{2+} binding Molecular transducer Signal receptor activity GPCR activity	9 18 28 53 46 34	0.032 0.032 0.032 0.032 0.032 0.032	Golgi lumen	8	0.045	Inositol phosphate metab. DNA replication Phosphatidylinositol system ABC transporters Ca^{2+} signaling Metabolic pathways	4 3 4 3 6 20	0.039 0.048 0.049 0.049 0.049 0.049	**E12** FOXO4 SREBP1 GATA HOXA4 MAZ	52 40 6 10 10 38	0.000 0.007 0.018 0.018 0.018 0.040
LGL063	649	448	430	No results	-	-	No results	-	-	No results	-	-	ALS Protein digestion and absorption	5 5	0.005 0.021	**E12** HOXA4 NFAT MAZ **LEF1** **STAT5B** **P53**	50 11 37 40 36 11 9	0.000 0.013 0.013 0.013 0.020 0.021 0.028
LGL067	557	410	392	Protein O-linked glycosylation	9	0.0067	GPCR activity	33	0.027	No results	-	-	Inositol phosphate metabolism	4	0.042	**E12** **LEF1** MAZ **TCF7** FOXO4 NFAT CMYB	50 37 41 11 36 34 8	4.4e-05 0.003 0.003 0.004 0.005 0.005 0.035
LGL095	708	520	498	Protein O-linked glycosylation	9	0.0428	Oxidoreductase activity Inhibitory MHC-I receptor activity	5 3	0.0333 0.0101	Clathrin complex Golgi lumen Intermediate filament cytoskeleton	2 9 15	0.0432 0.0247 0.0432	No results	-	-	**E12** NFAT MAZ FOXO4 SRF **LEF1** STAT	66 44 49 42 10 27 9	8.3e-08 6.72e-05 0.0018 0.0126 0.0168 0.0289 0.0468
LGL113	675	487	470	No results	-	-	Oxidoreductase activity GPCR activity	5 38	0.0223 0.0099	No results	-	-	No results	-	-	**E12** MAZ NFAT FOXO4 **LEF1** **STAT3** CMYB	65 48 40 40 24 6 8	1.82e-08 0.0006 0.0026 0.0088 0.0360 0.0426 0.0443
LGL132	805	593	568	No results	-	-	No results	-	-	Intermediate filament cytoskeleton	19	0.0037	Alanine, aspartate and glutamate metabolism	4	0.0324	**E12** **LEF1** NFAT MAZ MYB STAT1 **STAT5B**	68 54 46 53 12 5 11	3.42e-06 3.26e-05 0.0023 0.0023 0.0062 0.0252 0.0477

Paired-wise analysis of tumor-monocyte cells per patient to prevent false-negative variants which could be a variant in monocytes in one patient, but a specific variant in a tumor sample of one another patient.
+Filtered on tumor-only, exonic, and all except synonymous or unknown variants. Some variants are located in the same gene.
*Number of genes mapped to Entrez gene IDs in WebGestalt analysis.
$Number of genes associated with the Biological process, Molecular function, and Cellular component in GO enrichment, in KEGG pathways and as transcription factor target (TFT).
&Benjamini-Hochberg method adjusted P value for correction for multiple testing.
Highlighted transcription factors in bold were selected for gene expression level validation with RQ-PCR.

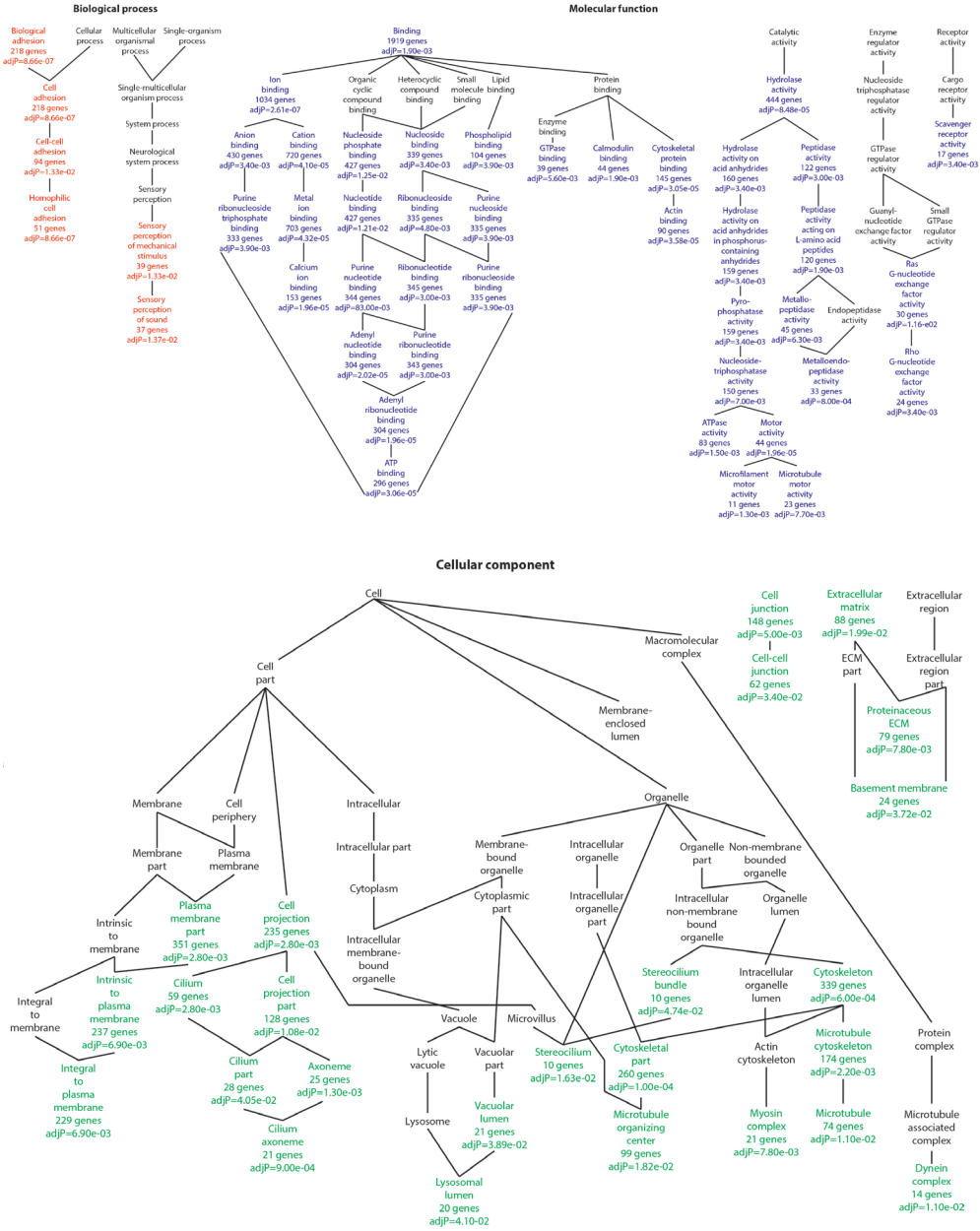
Supplementary Table 7. Top gene list with MAF=1.000.

Gene	Variant type	DNA change	Amino acid change	rs ID*
ZNF2	Non-frameshift deletion	228_230delGCG	76_77delA	rs10578519
FMNL2	Non-frameshift insertion	1671_1671insCCA	P557delinsPP	-
MUC13	Non-frameshift insertion	184_185insCTT	F62delinsSF	rs10630030
VEGFC	Non-frameshift deletion	1256_1258delTCA	419_420delS	rs5864401
CYFIP2	Frameshift insertion	280dupC	I93fs	rs5872508
SRRM3	Frameshift insertion	1815dupC	R605fs	rs75544239
DNHD1	Non-frameshift insertion	5726_5727insTGCCCTACTGCA	A1909delinsAALLH	-
KLRF1	Non-frameshift deletion	223_225delTGT	75_75delC	rs33911869
NCOR2	Non-frameshift insertion	5487_5488insAGCAGCGGC	G1830delinsSSGG	rs143952466
REC8	Non-frameshift insertion	681_682insGAA	E227delinsEE	rs10690822
SARM1	Frameshift insertion	544_545insGC	S182fs	-
SARM1	Frameshift deletion	549_550delGT	G183fs	-
GREB1L	Non-frameshift insertion	5582_5583insTCT	D1861delinsDL	rs34960489
RTTN	Non-frameshift insertion	727_728insGAG	D243delinsGD	rs74941514
VSIG10L	Frameshift insertion	2576dupC	A859fs	rs11402251
SON	Frameshift insertion	1320dupA	G2412fs	rs34377180
SON	Frameshift deletion	1332delA	R2416fs	rs34373121
NUDT11	Frameshift deletion	Whole gene	Whole gene	-

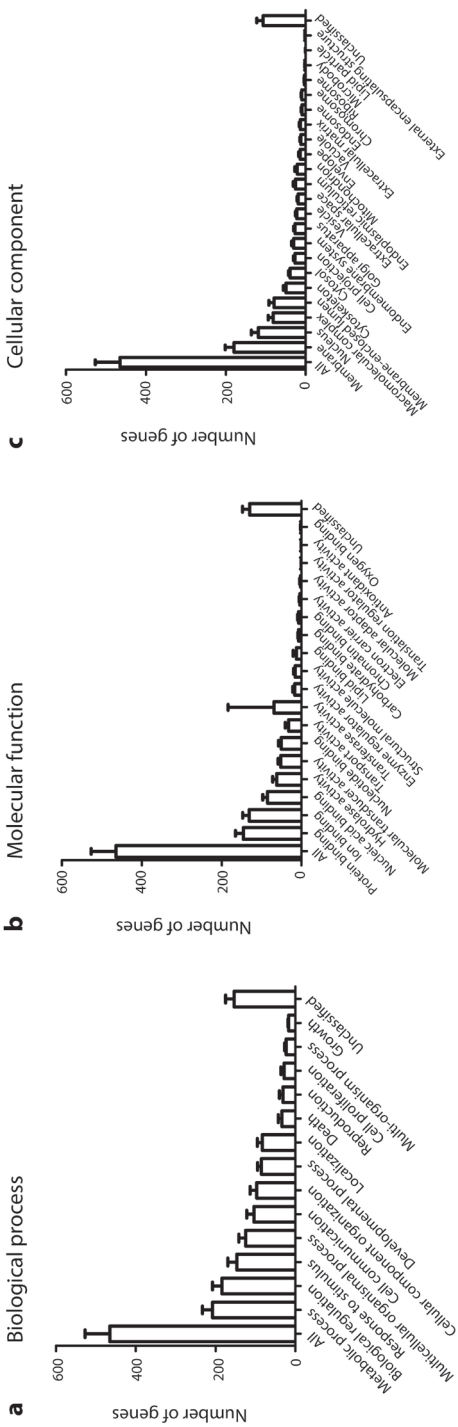
*rs ID based on dbSNP138 database [42].



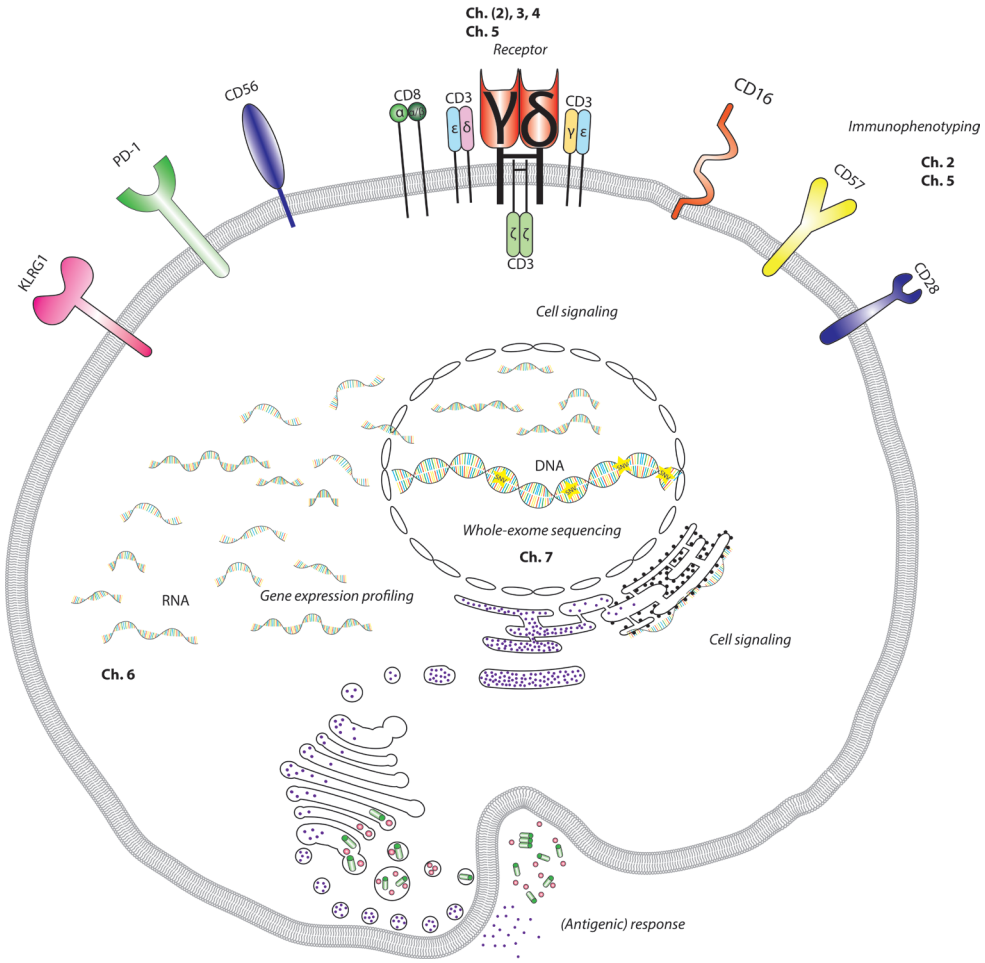
Supplementary Figure 1. Gating strategies to obtain TCRγδ+ T-LGL leukemia and monocyte cell populations.
General gating strategy to define viable cells, single cells and lymphocytes (**a**). Gating strategy for sorting leukemic tumor cells, further defined by Vδ-usage (**b**). Gating strategy for sorting control monocytes within the same sample (**c**).



Supplementary Figure 2. GO enrichment graph.
GO enrichment analysis results depicted in tree graph (adapted from WebGestalt). Biological processes, molecular functions and cellular components affected by the genes bearing variants are indicated in red. Number of genes and statistical significance indicated below the named process, function or component. Adjusted P-value is based on Benjamini-Hochberg adjustment for multiple testing correction.



Supplementary Figure 3. GO slim classification graphs from paired tumor-monocyte analyses.
From six patients paired TCRγδ+ T-LGL leukemic cells and control monocytes were sorted for WES analysis. After filtering out monocyte-related variants from TCRγδ+ T-LGL leukemia cell DNA and excluding high SNP density genes, the gene list was analyzed for GO slim summary classifications. Bar graphs are indicated with mean and SD.


Ch. (2), 3, 4
Ch. 5
Receptor
CD56
CD8
CD3
CD3
CD16
Immunophenotyping
Ch. 2
Ch. 5
PD-1
CD57
KLRG1
CD28
CD3
Cell signaling
DNA
Whole-exome sequencing
Ch. 7
RNA
Gene expression profiling
Cell signaling
Ch. 6
(Antigenic) response

Chapter 8

General Discussion

BACKGROUND

TCRγδ+ T-LGL leukemia is a chronic, rare and heterogeneous disorder, affecting mostly elderly individuals of average age above 60 [1,2, Chapter 1 this thesis]. The etiology of the disease still remains largely elusive, due to the low incidence of the disease, the frequent lack of symptoms and a highly heterogeneous disease spectrum [1,3, Chapter 5 this thesis]. In this thesis we aimed to identify underlying (molecular) pathways and mechanisms that might contribute to the pathogenesis of this disease. Currently it has been hypothesized that chronic and/or continuous stimulation, due to signals from the (micro-)environment of TCRγδ+ T cells, underlies proliferation, which ultimately results in a monoclonal population [4]. Since it is a chronic disease affecting mostly elderly individuals above age 60, continuous stimulation has often been linked to ageing processes [5,6]. On the other hand, cell-intrinsic mechanisms could contribute to the aberrant profile of TCRγδ+ T cells, as cancers are known to result from underlying genetic aberrancies both on DNA and RNA level. In this thesis we aimed to investigate the underlying pathology of TCRγδ+ T-LGL leukemia by focusing on both cell-intrinsic and external mechanisms, using multiple approaches.

CELL-INTRINSIC MECHANISMS

Genetic aberrancies as the foundation of leukemia

Cancer results from genetic aberrancies that have occurred in healthy cells, eventually leading to uncontrolled proliferation [7]. Most cancers develop by acquisition of multiple mutations over time [8], while 5-10% of all cancers have a predisposition in the germline sequence [9], as originally described in 1971 by Alfred Knudson in his "two-hit" hypothesis [10,11]. In this model hereditary cancer syndromes, e.g. childhood cancers (0-19 years of age), would already have their onset or first hit in the germline, [12], which includes amongst others mutations in DNA repair genes [13], or specific BRCA mutations in ovarian and breast cancer [14,15]. In these hereditary cases at least one additional mutation can cause neoplastic development, whereas sporadic cancers and more chronic type of cancers develop spontaneously or with ageing after acquiring two (or more) somatic mutations (Fig. 1).

In the hematopoietic system several forms of chronic cancer exist, which involve different (immune) cell types: chronic myeloid leukemia (CML; myeloid cells), chronic lymphocytic leukemia (CLL; B cells) [16], and also the less common large granular lymphocyte leukemia (LGL; T and NK cells) [17]. As underlying pathogenic mechanisms also somatic mutations have been described, such as mutations in the *TP53*, *ATM*, *EGR2*,

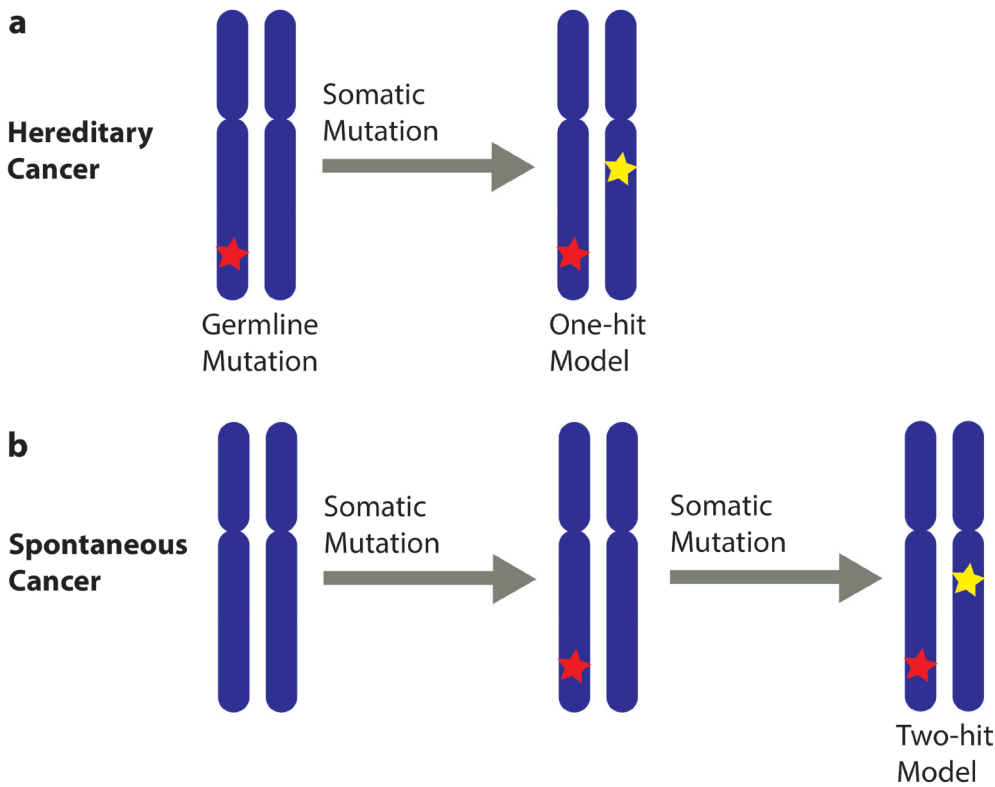


Figure 1. Two-hit hypothesis in carcinogenesis.
Alfred Knudson proposed his two-hit model of oncogenesis [10] based on acquisition of two genetic aberrancies that are required to initiate neoplastic development. In hereditary cancers a germline mutation is already present (**a**), which requires one additional somatic mutation, while spontaneous cancers (**b**) require two (or more) genetic events prior to neoplastic transformation. Figure adapted from Jozwiak *et al.* 2008 [11].

BRAF, *RPS15*, *IKZF3* genes in CLL [18,19] (reviewed by [20]), or the *BCR-ABL1* translocation in CML [21]. T-LGL leukemia however has only recently been associated with somatic *STAT3* and *STAT5b* mutations [22,23], thereby unifying this chronic leukemia with NK-LGL in which such mutations have also been described [24]. Of these, the *STAT3* mutations were identified most frequently in LGL leukemia patients (11-70%). Initially, the somatic *STAT3* mutations were mainly located in the SH2 domain, but more recently also gain-of-function mutations outside this domain were observed [25]. *STAT* mutations are currently highlighted in the latest World Health Organization classification of LGL leukemias [3,26] as diagnostic and therapeutic targets [27]. Of note, *STAT* mutations have only been described for CD8+TCRαβ+ T-LGL leukemia. An association of TCRγδ+ T-LGL leukemia and *STAT3* and/or *STAT5b* mutations has not been reported yet, although activating mutations have been found in lymphomas of the TCRγδ+ T cell type [28]. Due

to recent (technical) advances in sequencing methodologies more and more mutations are disclosed that underlie these chronic leukemias, and possibly explain their pathogenesis and/or etiology.

Novel genetic aberrancies potentially associated with TCRγδ+ T-LGL leukemia development

TCRγδ+ T-LGL leukemia is associated with clonal drifts and shifts, leading to differences in dominant clones over time [29]. A clear common underlying mechanism has not been described yet. Based on earlier whole exome sequencing (WES) studies, as performed in NK- and CD8+TCRαβ+ T-LGL leukemias and which mainly revealed *STAT3* gain-of-function mutations, we aimed for WES analysis using TCRγδ+ T-LGL leukemic cells sorted from patients, in parallel with sorted monocytes as internal control to distinguish tumor-specific mutations from individual-specific single nucleotide variants (SNV) [Chapter 7, this thesis]. Based on our WES data, we could define and validate six candidate genes containing homozygous variants that were present in patient leukemic cells, but absent in healthy monocyte cells (Table 1; Fig. 2).

Firstly, we observed an insert in the gene encoding the CREB3L2 transcription factor, which is involved in the regulation of transcription of (proto-) oncogenes such as c-Fos and c-Jun that are part of the AP-1 early response transcription factors [30] and that targets a.o. c-Fos for auto-regulation [31] (Fig. 2). Another target of CREB3L2 is E2F, which is driving expression of genes in tissue regeneration, proliferation [32], cancer cell growth and tumor invasiveness [33].

Transcriptional regulator NCOR2 also contained a homozygous insertion. This gene is associated with histone deacetylase recruitment and thus with down-regulation of gene expression [34]. A different variant in the *NCOR2* gene has been associated with Tamoxifen resistance in breast cancer [35] and early tumor recurrence [36]. At the protein level NCOR2 can show aberrant cytoplasmic occurrence; nuclear exclusion of NCOR2 been associated with colorectal cancers, thus supporting the role for NCOR2 in tumorigenesis [37]. This could lead to blocked repression of target genes such as E2F, Fos and also CREBBP [38] (Fig. 2). Other associations have been described with melanoma [39] and leukemogenesis in general [40]. As TCRγδ+ T cells are known to reside in epithelial tissues of the gut and skin, genes related to colorectal and skin cancer could also impact on the function and anatomical position of TCRγδ+ T cells [41]. The alteration of both CREB3L2 and NCOR2 thus could have detrimental downstream effects on the development of clones in mainly TCRγδ+ T-LGL leukemia, but it should be noted that altered expression levels were observed in CD8+TCRαβ+ T-LGL leukemia [Chapter 7, this thesis].

Table 1. Variant-containing and up- or down-regulated genes.

Gene	Dataset (WES/GEP/both)	Variant*	Fold change**	
			RQ**	Array (GEP)**
Up-regulated genes				
ID3	GEP	-	+28.410	7.228
KLRF1	WES	Non-fs del	+31.112	Related gene: KLRD1 -2.888
STAT3	Both	TFT	+12.845	-
STAT5B	Both	TFT	+9.657	-
IL6R	*Both indirectly*	*Found as central hub in regulating all genes*	-	-
VEGFC	WES	Non-fs del	+13.465	Related gene: VEGFA +2.854
BCLAF1	GEP	-	3.867	2.039
CCR7	GEP	-	4.372	7.377
CD28	GEP	-	3.283	7.012
NCOR2	WES	Non-fs del	+3.767	Related genes: NCOR1 -2.292 NCOA3 +1.786 NCOA5 -1.706
TP53	WES	TFT	+7.755	-
LEF1§	Both	TFT	-4.125	16.365
CREB3L2	WES	-	+1.492	-
Down-regulated genes				
CFLAR	GEP	-	-65.767	-2.449
XIAP	GEP	-	-2.489	-2.416
CASP1	GEP	-	-5.269	-2.783
CX3CR1	GEP	-	-2.044	-5.433
FAS§	GEP	-	+3.059	-2.215
IFNG	GEP	-	-6.531	-3.682

*Variant information only available from WES dataset. Non frameshift (fs) deletions (del), or identified upon transcription factor target analysis (TFT) in WebGestalt database [85,86].
**Fold change either from array data (Chapter 6) or RQ-PCR validation experiment (both datasets). Fold change calculated with ΔΔCt method.
§Genes that were not fully validated using RQ-PCR technique (see Chapter 6).
-: not available.

Since CREB3L2 contained a gain-of-function variant, this would lead to increased expression of proto-oncogenes *c-Fos* and *c-Jun*, which have also been implicated in invasive tumor growth [33]. Interestingly, our WES study showed a more direct relation to invasive tumor growth with a frameshift deletion in the *VEGFC* gene. VEGFC is an important factor in lymphangiogenesis due to its native expression in epithelial cells [42], and has been described to be associated with metastatic gastric [43] and breast cancer [44]. The frameshift deletion that we found in our study ws associated with

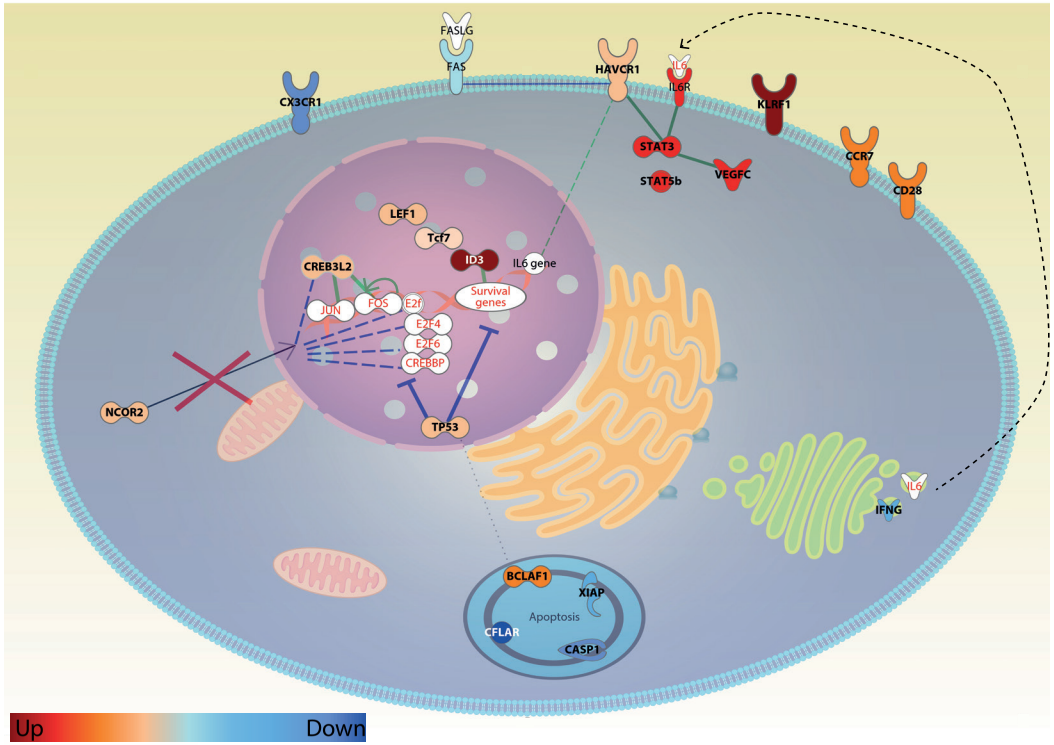


Figure 2. IPA-based schematic network of WES and GEP candidate genes to be involved in TCRγδ+ T-LGL leukemia pathogenesis.
Candidate genes from both datasets are indicated in bold, associated molecules and genes in regular font type. Oncogenes and other genes playing major roles in cancer development are indicated in red font. Up- and down-regulation are indicated with red and blue color intensities respectively indicated in gradient bar in figure: up-regulated genes are colored red gradients based on their level of up-regulation; dark red strong, lighter red/orange less strong or weakly up-regulated. Down-regulated genes are colored in a similar gradient in blue: dark blue strong, light blue weak down-regulation. Green lines between molecules indicate a positive regulation, while blue lines show negative regulations and black lines show movements. Solid lines represent direct interactions, dotted lines indicate indirect interactions between molecules and/or genes. Apoptosis-related genes are clustered together as a complex. Figure is created using My Pathway Design tool in Qiagen's IPA [57].

enhanced gene expression in patients (Fig. 2). Given the fact that TCRγδ+ T cells can reside as intra-epithelial lymphocytes (IEL) in mucosal and epithelial tissues, the elevated expression of VEGFC in especially leukemic cells [Chapter 7 this thesis] and the association with epithelial esophageal squamous cell carcinoma [45] and melanoma [46] are intriguing. However, to further gain insights in the origin of the leukemic cells in TCRγδ+ T-LGL leukemia, more extensive validation studies on these gene aberrancies would be required, including functional assays to determine the effect of the variants on transformation and neoplastic development.

Genes related to immune cell activation, cytotoxicity and cytokine production such as *HAVCR1* and *KLRF1* also contained damaging variants. HAVCR1 is known to be important in the regulation of immune cell activation in virus infection [47] (Fig. 2). Additionally, the role of HAVCR1 in cancer progression was recently defined through elevated IL-6 production, which in turn activated the STAT3 pathway and increased HIF-1α production in various tumors, and thus increased angiogenesis and cancer metastasis [48,49]. Notably, T-LGL leukemia clones have also been found to highly express the *HAVCR1* gene [50]. Together with the variant in the *KLRF1* gene and the increased expression of this gene that is known to encode an inhibitory NK cell like receptor involved in tumor immunity [51,52], this could be associated with dysregulated functioning of TCRγδ+ T cells in TCRγδ+ T-LGL leukemia, resulting in increased cytotoxicity towards other blood cells and causing cytopenias [1]. Such dysregulated immune responses might also explain the association between TCRγδ+ T-LGL leukemia and underlying autoimmune diseases and other (hematological) malignancies [1; Chapter 5, this thesis].

Altered gene expression profiles associated with TCRγδ+ T cell leukemia development

Upon extended WES transcription factor target (TFT) enrichment analysis major transcription factors in hematopoietic development, proliferation, survival and immune cell activity/signaling (TCF7, LEF1, TP53, STAT3 and STAT5b) were found to be the main drivers of genes containing genetic variants [Chapter 7, this thesis] (Fig. 2). Expression levels of these transcription factors were found to be up-regulated in TCRγδ+ and CD8+TCRαβ+ T-LGL leukemia cells, thereby indicating a possible role in driving T-LGL leukemia pathogenesis. Notably, even though the *STAT3* and *STAT5b* genes did not bear exonic mutations in our TCRγδ+ T-LGL leukemia WES dataset, they did appear as main driving factors in aberrant transcription of target genes. Moreover, gene expression analysis of both STAT3 and STAT5b showed clear increases in TCRγδ+ T-LGL leukemia cells, thus further supporting a role for STAT factors in TCRγδ+ T-LGL leukemia as well [Chapter 7, this thesis].

In the broader gene expression profiling (GEP) data set the *ID3* transcription factor gene was found to be highly up-regulated in TCRγδ+ T-LGL leukemia cells. ID3 has been described to be clearly up-regulated and associated with high resistance to spontaneous apoptosis and chemotherapy in B cell CLL [53-55] and has been associated with cell cycle progression, differentiation and tumor invasiveness [56]. Furthermore, ID3 is important in cell survival, as evidenced from Ingenuity Pathway Analysis (IPA) [57] (Fig. 2). The role of ID3 in TCRγδ+ T cells has been addressed in transgenic mouse models showing a disruption in TCRαβ+ but not in TCRγδ+ T cell development [58], while in absence of ID3 T cell precursors failed to adopt the TCRγδ fate [59], suggesting TCRγδ+ T cell dependency on ID3 expression [60]. Given that mouse and human TCRγδ+ T cells are structurally and functionally different [61,62] it is difficult to compare both systems in order to understand TCRγδ+ T cell biology and aberrant GEP in human diseases such as TCRγδ+ T-LGL leukemia. Nevertheless, these findings would underline the importance of ID3 in survival, which could explain the major up-regulation of this gene.

Our GEP analysis further revealed significantly altered expression of genes involved in proliferation, apoptosis and normal immune cell activation [Chapter 6, this thesis]. Due to the chronic disease character and the underlying autoimmune diseases in T-LGL leukemia patients, it was hypothesized that T-LGL leukemia cells might be resistant to FAS-mediated apoptosis [63]. Normal LGL cells are programmed to undergo apoptosis after fulfilling effector functions (known as activation induced cell death, AICD) [64], a feature that appears to be lost in LGL leukemia [65]. Furthermore, the CD95/CD95 ligand apoptosis resistance was found to be related to aggressive disease showing resistance to chemotherapy. However, GEP of TCRγδ+ T-LGL leukemia cells did not show significantly altered genes in the CD95/CD95 ligand pathway, but other genes involved in apoptosis induction were actually down-regulated [Chapter 6, this thesis] (Fig. 2). Clinically, TCRγδ+ T-LGL leukemia is not an aggressive disease [1], which might explain why aberrancies in the CD95/CD95 ligand-apoptotic pathway are less likely to play a major role in its pathogenesis. In TCRγδ+ T-LGL leukemia down-regulation of the Caspase-8 and FADD like apoptosis regulator (*CFLAR*) gene was observed. To date, only one variant in the enhancer region of CFLAR was found to be linked with the development of breast cancer, but without direct functional association [66]. Even though this genetic variant unfortunately could not be evaluated in our WES study due to its location outside the exome, CFLAR might be an interesting candidate for further functional studies, especially since it has been found to be associated with breast cancer and altered expression levels in TCRγδ+ T-LGL leukemia.

Another increasingly important gene in the cancer field is the Bcl-2-associated transcription factor-1 (*BCLAF1*) gene. It has been identified as tumor suppressor gene in leukemias and lymphomas in particular, acting through promotion of TP53 activation

in DNA damage repair (DDR) [67,68], but also as oncogene in colorectal cancer [69]. This controversial role of BCLAF1 in different cancers strongly underlines the necessity to further investigate functional consequences, but a complicating factor is that complete *BCLAF1* knock-outs are lethal [70]. In TCRγδ+ T-LGL leukemia *BCLAF1* gene expression levels were up-regulated, possibly suggesting an apoptosis-inhibitory role, based on apoptosis-resistance findings associated with T-LGL leukemia pathogenesis [63]. Together with the observed Caspase-1 down-regulation in TCRγδ+ T-LGL cells, this would strongly suggest that further apoptosis profiling of these leukemic cells might provide essential insights on the full apoptosis profile (Fig. 2). Application of a multiplex ligation-dependent amplification (MLPA) procedure using specific inflammation and apoptosis probe sets [71,72] should help to establish a more comprehensive expression profile of apoptosis- and inflammation-related genes.

Finally, clinical symptoms such as cytopenias point to altered immune functioning and autoimmune characteristics in TCRγδ+ T-LGL leukemia. This idea is supported by the WES and GEP datasets, which both showed altered expression of genes associated with enhanced immune responses, i.e. increased inflammation through up-regulation of CD28 and CCR7, while infection-specific interferon (IFN)-γ production was down-regulated (Fig. 2). The ability to communicate with other cells could be lost through CX3CR1 chemokine receptor down-regulation (Fig. 2). This fractalkine receptor is associated with adhesion and migration of cells that are armed with perforin and granzyme B to act in a cytotoxic response at the site of infection [73,74]. Furthermore, when cells lack CX3CR1, their migration into tissues would be reduced, which could be linked to the increased numbers of circulating leukemic TCRγδ+ T cells seen in these patients.

Whole exome sequencing and gene expression profiling cross-validation of aberrations in TCRγδ+ T-LGL leukemia

As discussed above, the WES and GEP methods both revealed candidate genes that are affected (i.e. altered and/or differentially expressed) in TCRγδ+ T cell leukemia cells. Notably, each of these datasets can serve for validation purposes for the other dataset. WES data can provide clues for differential gene expression of especially downstream target genes, while GEP data can illustrate the effect of genetic variants on expression of a particular gene. In fact, WES and GEP datasets can be analyzed in parallel using IPA Comparison Analysis, which integrates both datasets, thus providing gene interconnections with variants identified from WES, and significantly differentially expressed genes from GEP [57]. When we performed this direct IPA comparison analysis we observed significant alterations in “T and B cell signaling in rheumatoid arthritis (RA)”, “nitric oxide (NO)” and “reactive oxygen species (ROS) production” pathways in the GEP dataset only, whereas the canonical pathways “xenobiotic metabolism signaling”, “triggering receptor

expressed on myeloid cells 1 (TREM1) signaling", "B cell activating factor signaling" and "liver X receptor/retinoid X receptor (LXR/RXR) activation" were all significantly affected according to both datasets using the Fisher Exact test (Fig. 3a, indicated in red).

Each of the significantly affected canonical pathways showed affected genes in both WES and GEP datasets upon Ingenuity Target Explorer analysis [75]. These pathways and their specific genes have been described to play roles in pathogenic events, particularly in immunopathogenesis and carcinogenesis (Fig. 3b). An important pathway in the development and propagation of cancer is to gain resistance to therapy through e.g. altered xenobiotic metabolism. Of note, not only genes involved in actual metabolism, but also genes involved in active removal of chemotherapeutic agents were found (ABC receptors). This could suggest that TCRγδ+ T-LGL leukemia cells have undergone additional transformation in the recognition and metabolism of foreign molecules [41]. Some patients received treatment, either for an underlying autoimmune disease or malignancy (Table 2).

Next to these overlapping genes and affected canonical pathways, both datasets showed distinct sets of affected genes which could be explained by the difference between DNA (WES) and RNA (GEP) levels and methods of investigation (Table 1, Fig. 3). The up-regulation of the xenobiotic metabolism pathway could actually be the result of such treatment. Altered metabolic signaling was also observed in the LXR/RXR metabolic signaling pathway. This pathway has been shown to contain SERPINA1, a possible biomarker in e.g. thyroid and cutaneous squamous cell carcinoma [76]. LXR/RXR signaling is also involved in immune regulation with antimicrobial effects. The TREM1 signaling pathway is involved in the regulation of inflammation, but rather at the level of cell-cell and cell-extracellular matrix interactions. Aberrant TREM1 signaling has been shown to be associated with immune-suppression and thus with tumor-favorable environments, an effect that could also be achieved through altered B cell signaling. TCRγδ+ T cells could be connected with such immune regulation pathways based on their ability to recognize a wide variety of antigens including microbes, due to their anatomical positions, and their ability to communicate with a large number of cell and tissue types [41]. Strikingly, all pathways showed overlapping genes, including NFKB1 and 2, which form a central hub in immune regulation [77]. Furthermore, the combined WES / GEP analysis also confirmed that STAT3 forms a central feature in LGL pathogenesis. This also holds true for PIK3CA; even though no directly damaging mutations were observed in these genes, there seemed to be an altered expression, possibly due to other upstream genes (Fig. 3b). Datasets showed distinct sets of affected genes, which could be explained by the difference between DNA (WES) and RNA (GEP) levels and methods of investigation. Linking genetic aberrancies with variable gene expression through expression quantitative trait loci (eQTL) analysis allows to identify underlying biological mechanisms of

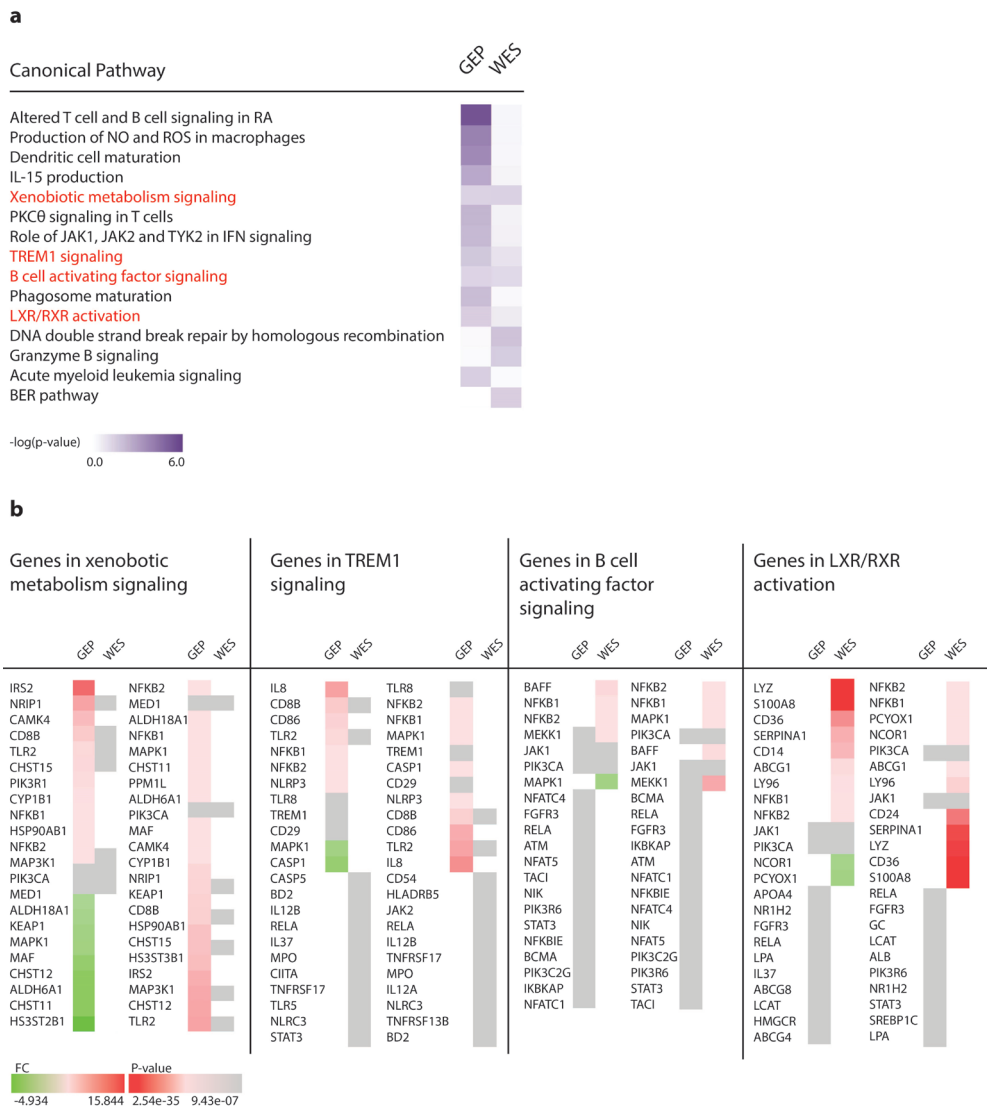


Figure 3. IPA comparison analysis of GEP and WES datasets.
Shown are –log(p-value) scores of canonical pathways affected according to both datasets. Pathways of interest that are affected equally according to both datasets are highlighted in red (**a**). Input datasets concern the supervised GEP analyses of T-LGL leukemia cells vs. effector TCRγδ+ T cells [Chapter 6 this thesis] and the WES dataset of N=2733 affected genes after filtering and exclusion of high density genes [Chapter 7 this thesis]. The Table summarizes genes that were affected according to both GEP and WES analysis in IPA using Fischer Exact test [57] and Ingenuity Target Explorer [Ingenuity Target Explorer, QIAGEN, Inc., targetexplorer.ingenuity.com, 75]. Genes overlapping between pathways are indicated in bold (**b**). p-values are calculated with IPA [57, QIAGEN, Inc.] with the Fisher Exact test. Grey boxes indicate lowest p-value (0.05). White boxes indicate that the gene was not found to be significantly affected in the corresponding dataset.

Table 2. Patient characteristics.

Patient	Sex	Age at diagnosis (years)	Clinical presentation and associated disease	Immunophenotype	Tumor load (%leukocytes)	Absolute LGL count (10^9/L)	Overall receptor	Clonality	Treatment
LGL056	F	40	Anemia, neutropenia, M. Graves	CD3+/CD4-/CD8+/CD16+/ CD56+/ CD57+/ CD45RA+/CD45RO-/Vγ9+/Vδ1+	41	3.6	Vγ9/Vδ1	Monoclonal	No treatment (Sep 2014)
LGL057*	M	46	Fever, positive Mantoux test	CD3+/CD4+/CD8+/CD56+/Vγ9+/Vδ2+	22.7	0.8-1.2	Vγ9/Vδ2	Monoclonal	No treatment (Nov 2017)
LGL058*	M	38	Lymphadenopathy, uveitis, sarcoidosis	CD3+/CD4/CD16+/CD56+/CD57+/CD45RA+/CD27-/Vδ2+	6.9	3.4	Vγ9/Vδ2	Monoclonal	No treatment (Mar 2016)
LGL063*	M	56	Heart transplantation	CD3+/CD4- /CD8+/CD16+/CD56+/CD57+/CD27-/CD197-/Vδ1+	23	1.4	Vγ9/Vδ1	Oligoclonal	No treatment (Sep 2015)
LGL064	M	76	Anemia, thrombocytopenia, hepatosplenomegaly, rheumatoid arthritis	CD3+/CD4+/CD8+/CD16+/CD56+/CD57-/CD45RA+/CD45RO+/CD27+/Vδ1+/Vδ2-	24	1.7	Vγ4/Vδ1	Monoclonal	Unknown (referral case)
LGL067	F	70	Granulocytopenia	CD3+/CD4-/CD8p/CD16-/CD57p/CD27-/CD45RA+/Vγ9+/Vδ1+			Vγ9/Vδ1	Monoclonal	MTX, Cyclosporin (Mar 2013)
LGL083	F	30	Unknown	CD3+/CD4-/CD8+/CD16+/CD57+/CD45RA+/CD27-/CD197-/Vγ9-/Vδ2+	8	0.3-2.4	Vγ8/Vδ2	Oligoclonal	Unknown (referral case)
LGL087	F	74	Unknown	CD3+/CD4-/CD8+/CD56+/Vδ1+	28	ND	Vγ2/Vδ1	Monoclonal	Unknown (referral case)
LGL088	M	54	Unknown	CD3+/CD4-/CD8-/CD16-/CD56+/CD57+/CD45RA+/CD27-/CD197-/Vδ1+	26	0.8	Vγ2/Vδ1	Monoclonal	Unknown (referral case)
LGL089	M	54	Unknown	CD3+/CD4-/CD8+/CD16+/CD56+/CD57+/CD45RA+/CD27-/CD197-/Vγ9+/Vδ2+	10	4.3	Vγ9/Vδ2	Monoclonal	Unknown (referral case)
LGL091	F	80	Unknown	CD3+/CD4-/CD8w/CD16+/CD56p/CD57+/CD27-/Vδ2+			Vγ9/Vδ2	Monoclonal	Unknown (referral case)
LGL092	F	55	Unknown	CD3+/CD4-/CD8-/CD25-/CD26-			Vγ9/Vδ2	Monoclonal	Unknown (referral case)
LGL095	M	74	Unknown	CD3+/CD4-/CD8w/CD45RA+/CD45RO-/Vγ9+			Vγ9/Vδ3	Monoclonal	Unknown (referral case)
LGL113*	M	70	Unknown	CD3+/CD4-/CD8-/CD16-/CD56-/CD57-/CD45RA+/CD45RO+/CD27-/CD197-/Vγ9+/Vδ2+	51	6	Vγ9/Vδ2	Monoclonal	Unknown (referral case)
LGL125	F	45	Unknown	CD3+/CD4-/CD8p/CD56-/CD197-/CD45RA+			Vγ9/Vδ1	Monoclonal	Unknown
LGL129	F	55	Asymptomatic; associated RA	CD3+/CD4+/CD8+/CD56+/CD197-/CD45RO+/Vγ9+/Vδ1+			Vγ9/Vδ1	Monoclonal	Unknown (referral case)
LGL132	F	65	Unknown	Unknown			Vγ9/Vδ1	Monoclonal	Unknown

* Patients included in both WES and GEP datasets.
ND, not determined.

the gene variant and expression level association [78]. By using RNA-sequencing direct associations can be made, since paired-end RNA-sequencing data is able to provide insights in both the genetic variants (qualitative) and the level of expression (quantitative genetic information). Due to disease heterogeneity in T-LGL leukemia patients in general, both cis- and trans-eQTL information can provide in-depth information on the mechanisms driving this heterogeneity. Cis-eQTLs are located on the same chromosome as the variant gene, regulating the transcription of the target gene, giving allele-specific expression and patient-specific and T-LGL leukemia specific patterns, whereas trans-eQTLs are located anywhere in the genome, regulating target gene transcription through modification of the activity and abundance of other proteins or enzymes regulating the target gene, which does not discriminate between alleles of which the data is useful for identifying common underlying biological mechanisms [79]. Whole genome sequencing (WGS) can also provide cis- and trans-eQTL information, but since information on gene

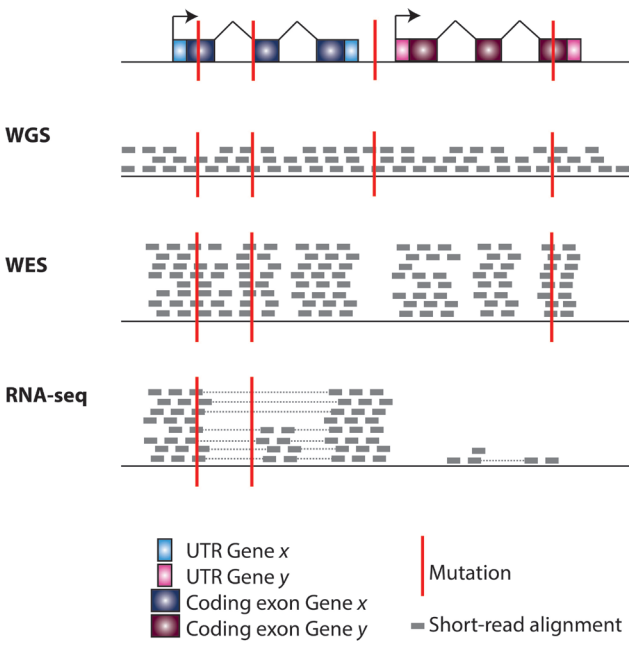


Figure 4. Schematic overview of different NGS methods (WGS vs. WES vs. RNA-seq) for mutation detection.
Overview of NGS techniques targeting mutations located in exons and intergenic regions. The number of short-read alignments marked in grey represents the sequencing depth of the techniques. WGS covers all areas of the genome, whereas WES mainly covers the coding regions (exons), and RNA-seq also provides expression-related data. Dotted lines between RNA-sequencing reads represent the "align-then-assemble" methods of RNA-sequencing data alignment: first short RNA-sequencing reads are aligned to the reference genome, taking splicing events into account, then transcripts are reconstructed from spliced alignments [82]. Gene X exemplifies an expressed gene, and gene Y a non-expressed gene. Figure adapted from Schneeberger, 2014 [81].

expression levels is lacking [80-82], other approaches such as RNA-sequencing should follow the WGS analysis (Fig. 4).

These techniques are highly useful for addressing research questions, together with functional assays based on CRISPR (clustered regularly interspaced palindromic repeat)-Cas9 technology [83] to further investigate down-stream functional and mechanistic effects involving these pathways on T-LGL leukemia development. Besides, owing to recent major advances, such NGS techniques also show high potential in diagnostic and prognostic settings in clinical care. NGS techniques provide high quality, fast and robust data, upon which data analysis algorithms tackle the limitations of the complex features of the sequencing results [84]. Using only small amounts of materials, standardized NGS protocols in multicenter settings contribute to easy and quick diagnosis and prognosis, thereby increasing feasibility and reducing invasive investigations in patients [85].

Future research should be focused on functionally validating the WES and GEP candidate genes to define their role in cell-intrinsic molecular mechanisms that contribute to LGL leukemogenesis. In this context, not only genetic aberrations should be considered, but also epigenetic changes (e.g. non-coding RNAs, methylation profiles etc.) will have to be included.

EXTERNAL STIMULATION AND ACTIVATION PROCESSES

Besides the presence of variants (WES) and altered expression levels (GEP) in genes involved in proliferation, apoptosis, survival and immune cell functioning, GEP also shows a high transcriptome similarity between TCRγδ+ T-LGL leukemia cells with (terminally differentiated) effector TCRγδ+ T cells was observed. As TCRγδ+ T-LGL leukemia typically presents in elderly individuals, additional processes such as immunological ageing, continuous stimulation, exhaustion and senescence (all possibly related to persistence of CMV or other pathogens), which are acting on normal TCRγδ+ T cells could contribute to the TCRγδ+ T-LGL leukemia pathogenesis.

Stimulation of normal TCRγδ+ T cells and their involvement in tumor immunity

TCRγδ+ T cells are well-known in establishing anti-tumor responses, and are proposed as immunotherapy through the direct killing of tumor cells [86]. Vδ1+ cells generally recognize hematological malignancies [87], whereas the more common circulating Vγ9+/Vδ2+ cells recognize and kill diverse solid tumors [88]. Already in 1990 human TCRγδ+ T cells were found to recognize and lyse Daudi (Burkitt lymphoma) cells [89]. In 1995 it was found that TCRγδ+ T cells respond to microbial products, such as

phosphates. Metabolites produced by e.g. mycobacteria and tumor cells include isopentenyl and prenyl pyrophosphates, both derivatives of poly-isoprenoid synthesis [90], or intermediates of the mevalonate synthesis pathway, a pathway that is elevated in tumor cells due to high cell turnover [91]. Also, patients with bone resorption receiving pamidronate displayed a significant increase in CD3+TCRγδ+ T cells in the PB, suggesting that this aminobisphosphonate could be a stimulating agent as well. Furthermore, all patients experienced an acute-phase reaction upon administration of pamidronate, which led to fever and a flu-like response, indicative of high IFN-γ release [92]. Further evaluation of this side-effect showed that TCRγδ+ T cells elicited anti-plasma cell activities in multiple myeloma after stimulation with e.g. pamidronate [93,94].

Watanabe *et al.* demonstrated that the TCRγδ+ T cell expansion and cytotoxic response against leukemia cells, from both cell lines and fresh material, through aminobisphosphonate treatment was indeed mediated by type-I IFN responses [95] (Fig. 5a). In 2008 these authors presented data from a clinical trial, using *in vitro* generated tumor-specific patient-derived zoledronate-activated TCRγδ+ T cells. Despite some variation in expansion and response of these TCRγδ+ T cells, cytotoxicity against patient myeloma and lymphoma cells could be measured [96]. Currently, TCRγδ+ T cells are being tested in the clinic, but this type of immunotherapy still requires improvements, as TCRγδ+ T cells can easily undergo apoptosis upon continuous stimulation. Furthermore, many patients developed side effects, as activated TCRγδ+ T cells are known to be highly pro-inflammatory. While favorable for killing the tumor cells, this pro-inflammatory effect is at the same time unfavorable for the patients themselves [97]. Also, different patients responded differently, indicating that other mechanisms in tumor immunity could play a role in skewing TCRγδ+ T cell responses, such as differences in the involved TCRγδ+ T cell subsets.

Importantly, an ambiguous role of TCRγδ+ T cells in tumor immunity has been recognized, as TCRγδ+ T cells recognize and respond to tumor cells, but have also been associated with tumorigenesis. The latter occurs through TGF-β production, which inhibits CTLs, and through IL-17 production, which leads to neutrophil-related inhibition of CTLs and thus to promotion of tumorigenesis and metastasis [98] (Fig. 5b).

Although an increased proportion of Vγ9/Vδ2 cells, acting as tumor-infiltrating lymphocytes (TILs), has generally been associated with good prognoses [99], tumor-promoting Vδ1+ cells were also found. These have been associated with poor prognoses in breast and colon cancer, since they suppress the immune response by negatively regulating dendritic cells. Pro-tumor activities of Vδ1+ T cells were often found to be correlated with IL-17 [100], rather than IFN-γ production [101] (Fig. 5).

TCRγδ+ T cells are not only able to exert anti- and pro-tumor activities, in rare cases they can also transform into cancer cells themselves, e.g. into TCRγδ+ T-LGL leukemia

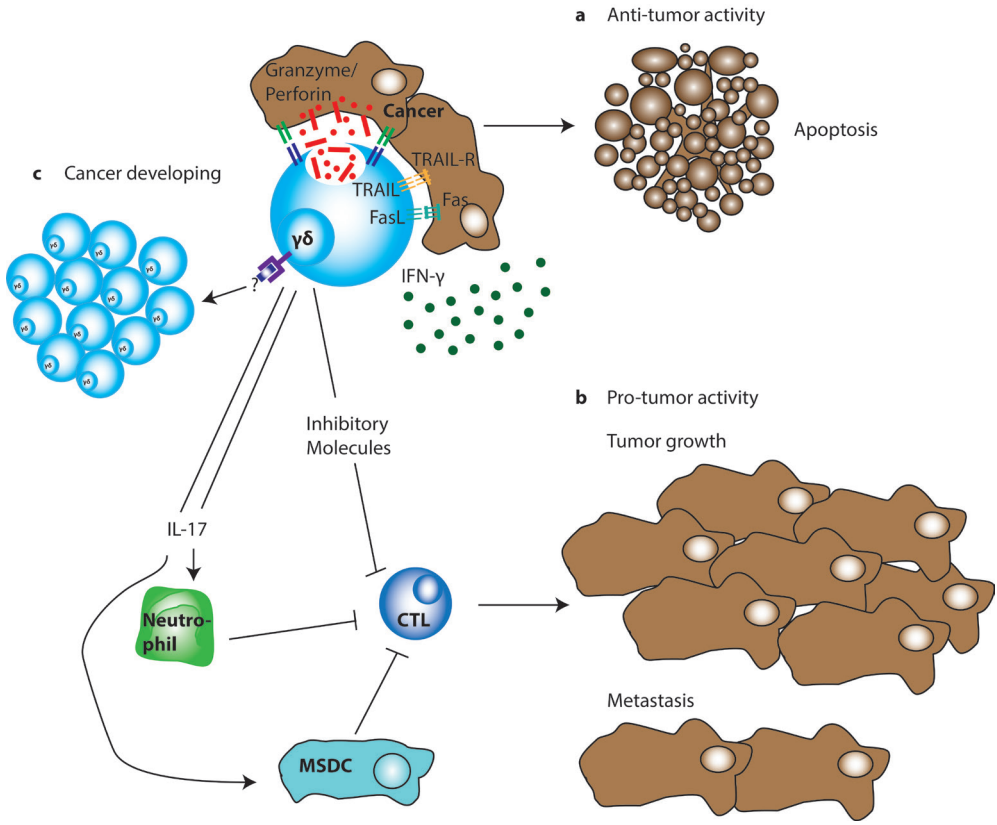


Figure 5. Multiple roles of TCRγδ+ T cells in tumor immunity and cancer development.
Schematic overview of the ambiguous roles of TCRγδ+ T cells in tumor immunity and cancer development. TCRγδ+ T cells can express inhibitory molecules and death receptors directed to tumor cells, which together with IFN-γ, perforins and granzymes trigger tumor cells to undergo apoptosis (**a**). TCRγδ+ T cells are able to express inhibitory molecules for CTLs, thereby blocking induction of tumor cell death; TCRγδ+ T cells also produce IL-17 that activates neutrophils and myeloid derived suppressor cells (MDSC), which in turn also inhibit CTLs (**b**). This contributes to pro-tumor environment and leads subsequently to metastasis. In case of chronic or continuous (antigenic) stimulation TCRγδ+ T cells can accumulate and develop into cancer cells themselves, as is the case in TCRγδ+ T-LGL leukemia (**c**). Figure adapted from Chitadze *et al.* 2017 [98].

[1, Chapter 5, this thesis] (Fig. 5c). T-LGL leukemia patients frequently display underlying autoimmune diseases or malignancies which are hypothesized to be a driving factor for leukemia development. In contrast to CD4+TCRαβ+ T-LGL leukemia where the disease etiology is strongly linked to CMV infection [102], in TCRγδ+ and CD8+TCRαβ+ T-LGL leukemia specific antigens, epitopes or stimuli are virtually unknown.

Effect of ageing on (terminally differentiated effector) TCRγδ+ T cells

As TCRγδ+ T-LGL leukemia cells show characteristics of antigen-experienced cells [1], it can be hypothesized that the LGL leukemic cells have an effector cell origin [Chapters 4 and 5, this thesis]. Indeed, our unsupervised GEP data showed a high correlation between TCRγδ+ T-LGL leukemia cells and healthy TCRγδ+ effector T cells [Chapter 6, this thesis]. During ageing, a phenomenon of accumulating terminally differentiated effector cells has been described [103, Chapter 2, this thesis], and an increase in both late-stage effector TCRγδ+ T cells, senescent and exhausted cells was observed in healthy elderly. Flow cytometric analysis of PB from healthy individuals aged 20 – 95 years showed a progressive decrease in TCRγδ+ T cell absolute numbers, including a relative shift in Vδ-usage; upon ageing Vδ1+ cells increased, whereas the most common Vγ9/Vδ2 cell type in PB decreased. Furthermore, we found that ageing (along with CMV infection; see later) affected the relative distributions of naive, central memory, effector memory and effector subsets towards late-stage effector phenotypes, with a profound increase in the expression of senescence markers [Chapter 2, this thesis]. These shifts in elderly towards effector and CD57+KLRG1+ subsets were especially observed in the presence of CMV [Chapter 2, this thesis].

Although exhausted T cells are generally observed during chronic viral infections, exhausted cells could also be identified in healthy individuals [Chapter 2, this thesis]. During chronic viral infections T cells are heavily subjected to continuous antigenic stimulation, upon which the cells become anergic. This results in T cell exhaustion, and eventually apoptosis. The exhaustion process can be reversed by means of PD-L1 blockade, as was shown for HIV-infections [104]. As there is no functional proof that T cells become exhausted in the context of the beta-herpesvirus CMV (reviewed by [105]), CMV infection and persistence is associated more with immunosenescence than exhaustion [106].

Effect of CMV infection on development and responsiveness of normal TCRγδ+ T cells

CMV is known to have major effects on the T cell population, especially with come of age [107]. CMV latency is established in hematopoietic stem cells, monocytes, macrophages, dendritic cells, and epithelial cells [108]. Despite CD8+ CTL responses upon acute infection, viral shedding continues, thus preventing the contraction phase of the CMV-specific CD8+ CTLs and resulting in “memory inflation” [109]. Furthermore, CMV is able to elicit immune evasion by modulating antigen presentation and producing miRNAs to up-regulate inhibitory molecules in the host cell in order to survive the host’s immune response [110,111]. Several factors have been investigated for their contribution to CMV-induced immunosenescence; these include the initial inoculum upon first

infection, the inflation of CMV-specific T cell responses, and the balance in functional effector T cells [112]. More recently, it has also been shown that CMV can evade the immune system by interfering with antigen loading onto MHC-I [113-115] and MHC-II [116] molecules.

The CMV immune evasion tactics as described above mostly apply to TCRαβ+ T cells and their corresponding antigen-presenting cells (APCs). TCRγδ+ T cells on the other hand, do not necessarily need MHC-molecules presented by APCs to recognize antigens. In fact, TCRγδ+ T cells are able to recognize also freely floating non-peptide antigens [117-119]. They can readily recognize a wide variety of antigens via their TCRs, upon which they can also directly respond in a CTL and NK-cell like manner, eliciting quick immune responses and thus forming a bridge between innate and adaptive immunity [41]. Also MHC-molecules themselves, especially the non-classical T22 MHC molecule [120,121] (reviewed by [119]), form antigens for TCRγδ+ T cells [122,123], linking them to NK cells and to roles in autoimmunity and viral recognition (Fig. 3 of the Prologue, page number 16).

Responsiveness of TCRγδ+ T cells to CMV has first been identified in 1999 in kidney allograft recipients, who displayed major expansions of TCRγδ+ T cells upon CMV reactivation and/or infection [124]. These expansions were thought to be a consequence rather than a cause of the CMV infection due to immunosuppressive therapy [125]. This was concluded from experiments showing that TCRγδ+ T cells could recognize and respond to viral epitopes *in vitro* [126]. Déchanet *et al.* were able to directly correlate the less common Vδ1+ and Vδ3+ T cells in PB to CMV responses and observed an increased long-lasting TCRγδ+ T cell expansion in the PB of kidney allograft recipients [127]. These cells expressed activation markers, specific TCRγ and δ chains with restricted junctional diversity, thus responding to CMV epitopes *in vitro* [127]. This increase in particularly Vδ1+ cells was also observed in healthy CMV-positive elderly [Chapter 2, this thesis], again implying that CMV-specific cells at least partly reside in the Vδ1+ cell population. Other studies have shown an association between Vδ1+ cells and CMV infection as well, but additionally showed that CMV also affects the total TCRγδ+ T cell compartment [128-130; Chapter 2, this thesis]. *In vitro* studies revealed that circulating CMV-specific Vδ2-negative – and thus Vδ1+ or Vδ3+ cells – produced large amounts of TNF-α and IFN-γ upon co-culture with CMV-infected cells or IgG-opsonized CMV. Next to that, TCRγδ+ T cells were also cytotoxic against CMV-infected cells through production of perforin and granzyme B (reviewed by [131]).

CMV recognition by TCRγδ+ T cells can occur at different levels, since CMV is able to inhibit immune cells through NKG2D-ligand expression on infected cells [132], and through altering the MHC-I and MHC-II trafficking as described above. Since TCRγδ+ T cells do not necessarily need MHC-molecules to recognize freely floating antigens,

CMV epitopes can still be recognized via the TCRγδ-receptor. Specific CDR3 clonotypes [133] in Vγ8/Vδ1 cells have been found to be associated with the recognition of CMV by TCRγδ+ T cells [134]. Following repetitive technical optimization experiments to reduce potential PCR-biases [Chapter 3, this thesis], we observed during ageing in healthy individuals, dominance towards Vγ2 or Vγ8 plus Vδ1 usage memory cells, which was in strong contrast to the Vγ9/Vδ2 dominance in young individuals. However, only in a few individuals the presence of both CMV-specific TRG and TRD clonotypes could be shown [Chapter 4, this thesis]. These findings could indicate that active, circulating CMV-specific TCRγδ+ T cells are only present in individuals with active CMV-infection, e.g. in patients undergoing immunosuppressive therapy during transplantation [127,131,135]. Alternatively, this occurs only locally, since CMV infects mainly lymphocytes and epithelial cells, a location where TCRγδ+ T cells are also present [41,88]. Given that TCRγδ+ T cells can respond to a wide variety of antigens, it is very likely that TCRγδ+ T cells also recognize CMV at different levels, as described above. More insights in the recognition of CMV by TCRγδ+ T cells could also provide valuable knowledge when applying TCRγδ+ T cells for immunotherapy in other viral infections such as other herpes viruses (HSV, EBV), hepatitis viruses (HBV, HCV), or even HIV. However, when using TCRγδ+ T cells in immunotherapy (either for cancer, or for (chronic/latent) viral infections) senescence and exhaustion phenomena should be considered, in view of chronic stimulation.

Immunosenescence impacts on TCRγδ+ T cell function and repertoire

To date, no functional signs of immune exhaustion have been associated with persistent CMV infections. Since CMV is a latent virus, it is highly important for the virus that the host cell survives. As the ultimate outcome of exhaustion is apoptosis [136], this would be unfavorable for viruses. Hence, the effect of CMV on the immune system is more apparent in the context of immune suppression, e.g. following organ transplantation or upon immunosuppressive therapy. CMV has also been associated with lymphoproliferative diseases since it is known to elicit "memory inflation", and thus causing a dramatic increase in effector memory cells without these cells undergoing the contraction phase [109]. Furthermore, CD4+TCRαβ+ T-LGL leukemia pathogenesis has been linked to CMV infection [102], underlining the role of CMV as a continuous antigen in the development of this rare type of LGL leukemia. Immunosenescence *per se* is associated with co-morbidity and mortality at different levels of the human immune system. Age-related changes in immunity do not only concern reduced antigen-specific responses and vaccine responses, but also involve an imbalance in effector and regulatory T cells, possibly contributing to autoimmune diseases, varying from diabetes mellitus to psoriasis, and even neurodegenerative diseases [137]. This increased level of inflammation has been associated with the secretion-associated senescence profile or "inflammageing" [104,138].

Inflammageing is a phenomenon characterized by chronically activated immune cells, initially only macrophages ("macrophageing"). Interestingly, also in inflammageing a two-hit hypothesis has been proposed, i.e. involving genetic and environmental factors. The first hit is provided by the genetic background, which determines susceptibility to age-related disease development, whereas the second – environmental – hit is necessary to develop overt age-related disease, morbidity and mortality [138,139].

Immune ageing and inflammageing could be at the basis of CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia as these cells display high correlation with terminally differentiated effector cells, a subset which is also increasingly present in individuals upon ageing. Moreover, these leukemia cells display mixed exhaustion and senescence phenotypes in terms of surface markers, cytotoxicity and AICD resistance features (see below). Notably, CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia patients often display underlying autoimmune diseases and cytopenias, which are both also more frequent upon ageing.

As discussed above, immune ageing plays a major role in shaping the immune system [5], which thus logically also impacts on the immune repertoire. (Long-term) exposure to particular antigens could provide selective pressure for immune cells bearing specific antigen-specific receptors (immune repertoire). Humans become exposed to many different environmental factors throughout their life, but these exposures might differ between individuals and at different ages. Immune repertoire shaping already starts at a very early life stage, namely in the womb and during early T cell development, which is evident from consequences of serious Rubella [140] and HIV [141] infections that occur during pregnancy. Therefore it has been hypothesized that both T cell development and antigenic exposure play a major role in shaping the TCRγδ immune repertoire [1]. This is indeed apparent from both phenotypic and genotypic characteristics [Chapters 2 and 4, this thesis]. The phenotype of thymic and cord blood TCRγδ+ T cells has been investigated by Sandberg *et al.* in 2006, showing dominance of Vδ1 expression in thymic and cord blood material, with a shift towards Vγ9/Vδ2 expression after birth and during adulthood. In Chapter 2 we observed progressive decreases of absolute and relative numbers of total TCRγδ+ T cells, which mainly affected the most common TCRγδ+ T cell population in adult peripheral blood, i.e. Vγ9/Vδ2+ cells. Strikingly, upon ageing Vδ1 cells were relatively more frequent, especially in CMV-seropositive elderly individuals, suggesting a role for CMV in shaping the TCRγδ+ T cell immune system towards Vδ1-expressing cells [124,127].

Additionally, NGS-based TCR gamma (TRG) / TCR delta (TRD) repertoire analysis was performed in TCRγδ+ T cells from thymus, cord blood, young adult PB, and elderly PB samples [Chapter 4, this thesis]. One of the main findings was that we generally observed maintenance of the naive diversity during ageing. In contrast to the lack of changes in the naive repertoire in both young and elderly, major shifts in

memory population compositions were observed, with a clear Vγ9/Vδ2 dominance in young individuals (aged 20-34) to a Vγ2/Vδ1 dominance in elderly (aged 56-70). This increasing Vδ1 dominance in elderly could -at least partly- be due to CMV persistence, although no clear TRG / TRD clonotype co-existence was identified with proven CMV specificity. Nevertheless, it would be interesting to further investigate the major epitopes recognized by these different memory populations, as this could provide insights in what could possibly drive the further shaping of the TRG / TRD repertoire, and could give suggestions on potential antigens driving LGL pathogenesis. In order to confirm the origin of aberrant TCRγδ+ T cells in T-LGL leukemia, TCRγδ+ T-LGL-related clonotypes as identified from the dominant LGL clones in patients were traced back in the repertoire of healthy young and elderly individuals. Strikingly, some TCRγδ+ T-LGL-related clonotypes were indeed found in the healthy repertoire, especially in the effector subsets of healthy elderly [Chapter 4, this thesis]. Though preliminary and restricted, this would indeed fit the hypothesis that TCRγδ+ T-LGL leukemia cells do originate from the normal, antigen-experienced, circulating effector TCRγδ+ T cells that have ultimately undergone transformation towards more survival and proliferation, rather than apoptosis, and towards an inflammatory state. To further investigate this, single cell sequencing could be applied, to truly identify the complete TRG/TRD receptor combination in a single cell. This would however require single-cell sorting of large numbers of TCRγδ+ T cells (and thus large volumes of blood) to generate a representative picture of the TRG / TRD repertoire. Nevertheless, further studying the TRG / TRD immune repertoire remains highly interesting, as there are many relevant aspects to study with an increasing availability of high-throughput techniques. As the starting material could pose serious limitations and challenges to address the relevant questions, the material to be investigated and the way of investigation should be carefully considered.

Functional alterations in TCRγδ+ and TCRαβ+ T-LGL leukemia cells seem to be very similar

Immunological ageing in an important biological phenomenon characterized by two major processes: senescence and exhaustion. Senescence affects the immune system at different levels, varying from reduced antigen-specific responses [5] to thymic involution and subsequently reduced naive T cell outputs [5,6,142], whereas exhaustion reflects a diminished immune response due to chronic (antigenic) T cell stimulation [136]. Both processes are characterized by specific surface membrane markers, cytokine profiles and changes in proliferative capacity, as summarized in Figure 6 in the General Introduction of this thesis, page number 33. In short, senescent cells are characterized by a high level of DNA damage due to aged DDR mechanisms and increased ROS activity, decreased telomere length, decreased proliferation, but polyfunctional cytokine

production and cytotoxicity. Strikingly, senescent cells have the ability to "multitask" in the sense that they produce TNF-α and IFN-γ as well as granzyme B. Furthermore, senescent cells are less likely to undergo apoptosis and thus seem to have gained apoptosis resistance. On the other hand, exhausted cells also obtain DNA damage by similar mechanisms as senescent cells, but exhausted cells do not proliferate, do not renew telomeres, and do not produce (any) cytokines. In fact, these cells are highly sensitive to apoptosis (reviewed by [143]). Major differences between senescent and exhausted cells lie in (antigenic) stimulation: senescent cells are still able to respond to antigens and/or stimuli, while exhausted cells are saturated and irresponsive.

T-LGL leukemic cells display variable senescence-related markers, but it is unclear why they do not undergo apoptosis, assuming that they are possibly subjected to chronic (antigenic) selection [1, Chapters 5 and 6, this thesis]. This raises the question to what extent T-LGL leukemic cells are senescent or even exhausted. To address this, we first optimized stimulation, proliferation and apoptosis assays for T-LGL leukemic cells. Based on TCR-mediated stimulation using αCD3/αCD28 stimuli, in the presence and/or absence of IL-2, IL-15 and zoledronic acid (the latter being a phospho-antigen stimulus for TCRγδ+ T cells), the level of senescence and exhaustion with respect to cell surface and intracellular marker expression, cytokine production, survival and transcription factor expression was investigated in a small series of T-LGL leukemia samples.

When healthy and tumor T cells from patients were screened for senescence and/or exhaustion marker expression, both CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia patients showed increased percentages of CD244+CD57+ cells (Fig. 6a). Apart from being an NK cell marker, CD57 is also a marker for T cell replicative senescence and LGL leukemia [1,144], and therefore expected to be highly expressed in LGL patients but less so in healthy subsets. CD244 (2B4) is also an NK cell marker, which is more associated with exhaustion, especially in combination with CD57 and PD-1 expression in triple positive cells [136]. Even though CD244+CD57+ cells were increased in both T-LGL leukemia types, only in CD8+TCRαβ+ T-LGL leukemia a significant increase in the triple positive cells was found (Fig. 6a), suggesting that CD8+TCRαβ+ leukemic T cells are slightly more affected by exhaustion and senescence than TCRγδ+ leukemic T cells, which is in line with our earlier data [Chapter 2, this thesis].

Next, we determined the proliferative capacities of T-LGL cells. Even though LGL cells are expressing CD57, and thus are showing signs of replicative senescence, it was obvious that these cells could still proliferate to some extent upon TCR-dependent stimulation (Fig. 6b, upper panels). From direct comparison with healthy cells it was clear that especially the CD8+TCRαβ+ T-LGL leukemia cells lost their replicative capacity after two generations, whereas TCRγδ+ leukemic T cells were capable of undergoing more divisions than CD8+TCRαβ+ leukemic T cells (Fig. 6b, lower panels). Since especially TCRγδ+ T

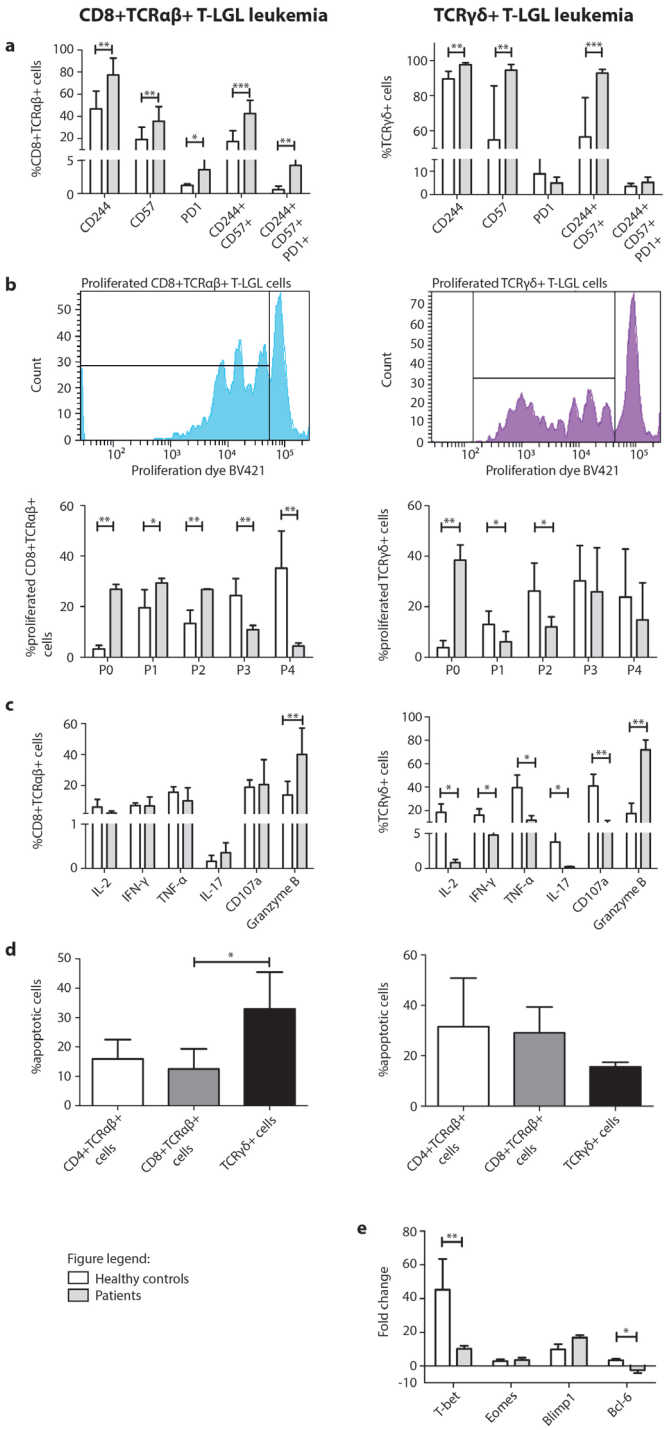

CD8+TCRαβ+ T-LGL leukemia
TCRγδ+ T-LGL leukemia
a
%CD8+TCRαβ+ cells
%TCRγδ+ cells
CD244
CD57
PD1
CD244+ CD57+
CD244+ CD57+ PD1+
b
Proliferated CD8+TCRαβ+ T-LGL cells
Proliferated TCRγδ+ T-LGL cells
Count
Proliferation dye BV421
%proliferated CD8+TCRαβ+ cells
%proliferated TCRγδ+ cells
P0
P1
P2
P3
P4
c
%CD8+TCRαβ+ cells
%TCRγδ+ cells
IL-2
IFN-γ
TNF-α
IL-17
CD107a
Granzyme B
d
%apoptotic cells
CD4+TCRαβ+ cells
CD8+TCRαβ+ cells
TCRγδ+ cells
Figure legend:
Healthy controls
Patients
e
Fold change
T-bet
Eomes
Blimp1
Bcl-6

Figure 6 (see left page). Evaluation of senescence and/or exhaustion features in CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia cells upon stimulation, proliferation and apoptosis.
Data from functional assays represent CD8+TCRαβ+ T-LGL leukemia cells (left column) and TCRγδ+ T-LGL leukemia cells (right column). Relative numbers of T-LGL leukemia cells expressing surface exhaustion/ senescence markers (**a**). Proliferation levels as determined based on histogram plots and dilution of proliferation dye and as quantified data (bar graphs). The number of divisions is indicated as P0-P4 (**b**). Cytokine production and degranulation marker measurement CD107a (level of cytotoxicity) based on intracellular cytokine staining (**c**). Levels of apoptosis induction in CD4+TCRαβ+, CD8+TCRαβ+ and TCRγδ+ T cell populations in the two patient groups upon AICD following TCR-mediated stimulation (**d**). Relative mRNA transcription factor expression indicated as fold change upon normalization to ABL housekeeping gene (ΔCt) in healthy control versus patient samples (**e**). All data visualized with bar graphs and SD error bars indicate median values, whereas histogram plots (**b**) indicate representative experiments. Statistical significance of the apoptosis assay was tested using the Kruskal Wallis test followed by a post-Dunn's test. Other statistical significances were tested using the Mann Whitney U test. Level of significance indicated in the plots: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

cells still showed proliferative capacity, this could reflect a level of exhaustion that can be saved upon proper stimuli (136). However, the fact that both TCRγδ+ and CD8+TCRαβ+ T-LGL leukemic cells did not replicate as good as healthy cells, also correlates with the chronic disease phenotype, which is less associated with highly proliferating cells than for example acute leukemia [Chapter 5, this thesis].

Next, we evaluated cytokine and cytotoxicity profiles in the context of exhaustion and senescence. Senescent cells are known to be polyfunctional, which implies the production of multiple cytokines [145], while exhausted cells lose this feature during the first steps of exhaustion (reviewed in [136]). CD8+TCRαβ+ T-LGL leukemia cells generally showed normal levels of cytokines when compared to healthy CD8+TCRαβ+ T cells; in fact, granzyme B production was even significantly increased (Fig. 6c). However, TCRγδ+ T-LGL leukemia cells did show significant decreases of all cytokines and the degranulation marker CD107a, whereas granzyme B was also significantly increased in TCRγδ+ T-LGL leukemia cells (Fig. 6c). The increased granzyme B levels may explain the cytopenia symptoms that T-LGL patients frequently present with [Chapter 5, this thesis]. High cytokine levels and also high levels of cytotoxicity might result in autoimmunity against other blood cells, and thus in cytopenias. Approximately 20% of the T-LGL leukemia patients indeed have underlying or associated autoimmune diseases, which could result from the high cytotoxic state (Fig. 6c) [Chapter 5, this thesis].

Activated CTLs undergo activation induced cell death (AICD) [146]. When the level of apoptotic cells upon TCR stimulation (18 hours) was assessed in T-LGL leukemia patient samples, in the CD8+TCRαβ+ T-LGL leukemia samples a significant difference was observed between leukemic CD8+TCRαβ+ and normal TCRγδ+ T cells. While the fraction of apoptotic CD8+TCRαβ+ leukemic T cells was as low as for normal CD4+TCRαβ+ T cells, TCRγδ+ T cells in CD8+TCRαβ+ T-LGL leukemia patients did undergo apoptosis (Fig. 6d). In TCRγδ+ T-LGL leukemia patients no significant differences were observed, although a decreasing, but non-significant, effect was seen with respect to the percentages of

apoptotic normal CD4+TCRαβ+ and CD8+TCRαβ+ cells, and aberrant TCRγδ+ T-LGL leukemia cells (Fig. 6d). These preliminary data would imply that both the CD8+TCRαβ+ leukemia T cells and the TCRγδ+ leukemia T cells are less prone to undergo AICD, which is in line with the disease phenotype. However, as only a limited number of patient samples was available for these experiments, the AICD assays should be repeated with a sufficient sample size, in order to confirm our preliminary findings.

Finally, specific transcription factors have been associated with exhaustion and survival processes. In fact, high levels of Eomesodermin (*Eomes*) are known to be associated with more exhausted cells (with cytotoxic activity) [147], while cells with high *T-bet* expression are less exhausted and also less cytotoxic [148]. As high *T-bet* expression is also associated with repression of *PD-1*, exhausted cells logically show high *Eomes* levels and low *T-bet* levels [149]. Exhaustion is not only controlled by transcription factors *Eomes* and *T-bet*, but also by *Blimp1* (*PRDM1*) and *Bcl-6* (reviewed by [150]). Up-regulation of transcription repressor *Blimp1* [151], and down-regulation of survival-related transcription factor *Bcl-6* are both linked with exhaustion, due to the resulting loss of survival [148]. Our preliminary RQ-PCR analysis of TCRγδ+ T-LGL leukemia cells showed differences in transcription factor expression when compared with healthy TCRγδ+ T cells under similar stimulatory conditions. *T-bet* expression was down-regulated in patients, while *Eomes* expression remained unchanged. *Blimp1* expression was slightly up-regulated, though not significant, and *Bcl-6* was highly down-regulated, reflecting drastic alterations in gene expression (Fig. 6e). Altogether, the transcription factor expression levels in the leukemic TCRγδ+ T cells correlated most with exhaustion, due to the down-regulation of both *T-bet* and *Bcl-6*, and the slight up-regulation of *Blimp1*. So far we have not produced RQ-PCR data for CD8+TCRαβ+ T-LGL leukemia cells.

In conclusion, based on the preliminary data from our functional analyses in T-LGL leukemia cells we identified increased exhaustion and senescence surface marker expression, especially of CD57 and CD244, in combination with PD1, which would correlate with an exhaustion state. Despite the high CD57 expression, which is also associated with replicative senescence, T-LGL leukemia cells were able to proliferate. This seems contradictory, but it should be stressed that a state of exhaustion can be rescued upon proper stimulation. Furthermore, T-LGL leukemia cells showed an ability to produce high levels of granzyme B, which might correlate with cytopenias that are associated with LGL leukemia disease. Granzyme B production could be involved in autoimmunity against other blood cells, possibly underlying the autoinflammatory or autoimmune diseases and flu-like symptoms seen in those patients [1, Chapter 5, this thesis]. This cytotoxicity profile might also reflect a state of low level inflammation, termed inflammageing [143], which is considered a hallmark of senescence and which fits the clinical presentation of most LGL leukemias at older age [1, Chapter 5, this thesis]. Finally, T-LGL

leukemia cells were less prone to AICD, although the altered transcription factor profiles still partially associated with exhaustion. It should be stressed that these data are preliminary and given the heterogeneity between different T-LGL leukemia patients, a larger series of both CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia patient samples should provide more insights on the functional properties of T-LGL leukemia cells in general. In this context it should also be noted that a better definition and sorting of the leukemic population based on unique markers, would be instrumental to fully disclose the functional properties of these leukemic LGL cells.

T-LGL leukemia cells and immune dysregulation

As noted above, T-LGL leukemia shows typical clinical characteristics beyond leukemia, which include autoimmune and other immune dysregulation features that are overlapping with other diseases. For example, in Felty syndrome (FS) patients the associated RA, neutropenia and (hepato)splenomegaly [152], together with the presence of oligo-/monoclonal LGL cells, and the recently reported high frequency of somatic *STAT3* and *STAT5b* mutations [153], all indicate an overlap between clonal T-LGL leukemic proliferations and FS. Not only in case of FS, but also in other (auto) immune dysregulation diseases LGL clones were identified, thus reflecting a common (auto)immune pathogenesis in which *STAT3* mutations could be relevant. This raises the question whether T-LGL leukemia is a consequence of autoimmune diseases due to continuous stimulation by (auto)antigens, or whether the *STAT3* mutated T-LGL leukemia cells might actually cause dysregulation of the immune system, leading to e.g. autoimmune disease [154]. In this respect it is relevant to remark that T-LGL leukemia mostly develops in elderly individuals, while autoimmune diseases sometimes have an early age onset, even though most appear later in life (from age 20-40 onwards or even >80) [155,156]. Based on this, it would be tempting to conclude that T-LGL leukemias would therefore be a consequence of the continuous exposure to autoantigens and subsequent damage-associated epitopes. However, this could work both ways, i.e. upon autoimmune disease development autoantigens could trigger immune cells including LGL cells, whereas proliferating LGL cells that develop into clones could also mediate (auto)immune dysregulation and cytopenias [1, Chapter 5 this thesis]. To further address "the chicken or the egg" question, patients diagnosed at a young age with for example RA would have to be followed over time to monitor the LGL cell numbers, clonality, and responsiveness.

CONCLUSION

Based on results presented in this thesis, it has become clear that TCRγδ+ T cells are subjected to ageing, immuneageing and inflammageing processes, in a way similar to other T cell types. As a consequence, differences in subset composition and distribution were apparent upon ageing with reduced naive and central memory cells, and increased effector memory, effector and senescent CD57+KLRG1+ cells. In these subset composition alterations CMV also appeared to play a major role. Differences in TCRγδ+ T cell compositions have also been observed in our molecular studies, as performed on naive, central memory, effector memory and effector TCRγδ+ T cell subsets. The dominance of Vγ9/Vδ2 memory cells at younger age shifted to Vγ2/Vδ1 and Vγ8/Vδ1 memory cell dominance in elderly individuals, most probably illustrating the impact of peripheral antigenic selection throughout life. In this process, also a role for CMV could be envisioned. Notably, normal elderly effector TCRγδ+ T cells showed TCRγδ+ T-LGL leukemia-related clonotypes as part of their TRG/TRD repertoire. Even though the TCRγδ+ T cells bearing these LGL receptors are not necessarily malignantly transformed from the beginning, they might have undergone changes upon ageing leading to an excessive imbalance towards proliferation, survival and inflammation. These underlying changes might result from the process of "inflammageing", one individual being more susceptible than the other. However, underlying mutations could also contribute to these transformations. All these aberrations contribute to immune dysregulation, with the production of cytokines, perforins and granzymes, eventually also resulting in an uncontrolled immune responses to self, thus explaining the underlying autoimmune diseases and cytopenias as observed in T-LGL leukemia patients. Conversely, the underlying autoimmune diseases and malignancies could also provide stimuli for the T-LGL leukemia cells in general, and perhaps especially for TCRγδ+ T cells, since they are able to recognize a wide variety of antigens, with or without MHC-molecules, peptides and non-peptides, self and non-self. Despite this senescent profile of T-LGL leukemia cells, they did not lose their replicative potential and thus will continue to proliferate, thereby contributing to chronic leukemia formation (Fig. 7).

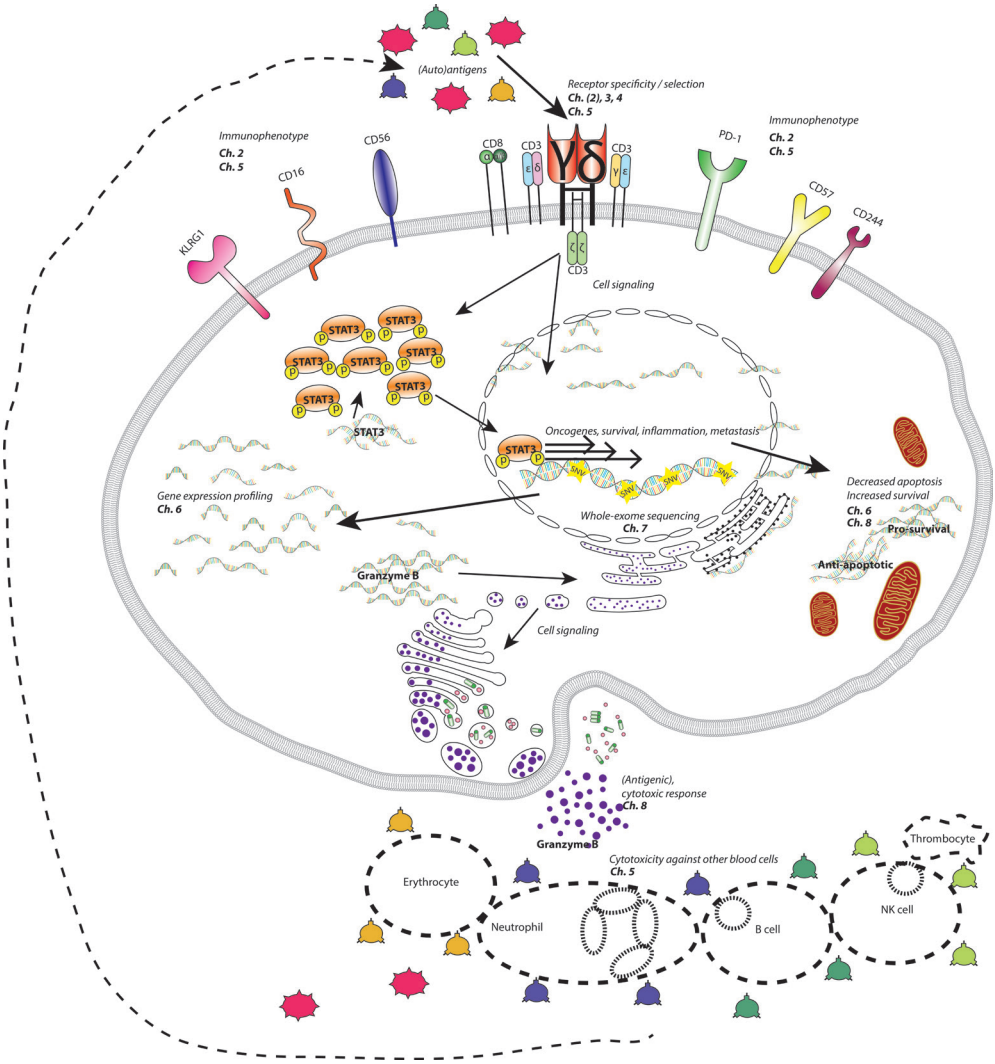


Figure 7. Summary of this thesis.
Upon stimulation from an unknown source TCRγδ+ T cells become activated, possibly involving up-regulated STAT3 signaling. Through persistent activation genetic aberrancies can occur, resulting in transcription of genes involved in survival and inflammation, anti-apoptosis genes and oncogenes. This results in elevated STAT3 expression, granzyme B production and decreased apoptosis with subsequent increased survival. The high inflammatory state of the T-LGL leukemia cells with elevated granzyme B possibly results in an immune response, also against other blood cells, leading to cytopenias. Epitopes and self-antigens can subsequently act as stimuli for TCRγδ+ T cells, potentially resulting in amplification of the process through a positive feedback loop.

REFERENCES

1. Sandberg, Y. *et al.* TCRγδ+ large granular lymphocyte leukemias reflect the spectrum of normal antigen-selected TCRγδ+ T cells. *Leukemia* **20**, 505-513 (2006).
2. Dinmohamed, A.G., Brink, M., Visser, O. & Jongen-Lavrencic, M. Population-based analyses among 184 patients diagnosed with large granular lymphocyte leukemia in the Netherlands between 2001 and 2013. *Leukemia* **30(6)**, 1449-1451 (2016).
3. Lamy, T., Moignet, A. & Loughran, T.P. Jr. LGL leukemia: from pathogenesis to treatment. *Blood* **129(9)**, 1082-1094 (2017).
4. Langerak, A.W., Sandberg, Y. & van Dongen, J.J. Spectrum of T-large granular lymphocyte lymphoproliferations: ranging from expanded activated effector cells to T-cell leukaemia. *Br. J. Haematol.* **123(3),** 561-562 (2003).
5. Boraschi, D. *et al.* The gracefully aging immune system. *Sci. Transl. Med.* **5(185)**, 185ps8 (2013).
6. Weiskopf, D., Weinberger, B. & Grubeck-Loebenstein, B. The aging of the immune system. *Transplant. Internat.* **22**, 1041-1050 (2009).
7. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458(7239),** 719-724 (2009).
8. Baez-Ortega, A. & Gori, K. Computational approaches for discovery of mutational signatures in cancer. *Brief. Bioinform.* 1-12 (2017).
9. Romero-Laorden, N. & Castro, E. Inherited mutations in DNA repair genes and cancer risk. *Curr. Probl. Cancer* **41(4)**, 251-264 (2017).
10. Knudson, A.G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Nat. Acad. Sci. USA* **68(4)**, 820-823 (1971).
11. Jozwiak, J., Jozwiak, S. & Wlodarski, P. Possible mechanisms of disease development in tuberous sclerosis. *Lancet Oncol.* **9**, 73-79 (2008).
12. Philips, S.M. *et al.* Survivors of childhood cancer in the United States: prevalence and burden of morbidity. *Cancer Epidemiol. Biomarkers Prev.* **24(4)**, 653-663 (2015).
13. Lindor, N.M. *et al.* Recommendations for the care of individuals with an inherited predisposition to the Lynch syndrome: a systematic review. *JAMA* **296(12)**, 1507-1517 (2006).
14. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).
15. Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q-12-13. *Science* **265**, 2088-2090 (1994).
16. Freireich, E.J. *et al.* The hematologic malignancies. Leukemia, lymphoma, and myeloma. *Cancer* **54**, 2741-2750 (1984).
17. Loughran, T.P. Jr. *et al.* Leukemia of large granular lymphocytes: association with clonal chromosomal abnormalities and autoimmune neutropenia, thrombocytopenia, and hemolytic anemia. *Ann. Intern. Med.* **102(2)**, 169-175 (1985).

18. Damm, F. *et al.* Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov.* **4(9),** 1088-1101 (2014).
19. Landau, D.A. *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525-530 (2015).
20. Kipps, T.J. *et al.* Chronic lymphocytic leukemia (CLL) is a malignancy of CD5+ B cells that is characterized by the accumulation of small, mature-appearing lymphocytes in the blood. *Nat. Rev. Dis. Primers* **3**, 16096 (2017).
21. Rowley, J.D. Chromosomal patterns in myelocytic leukemia. *N. Engl. J. Med.* **289**, 220 – 221 (1973).
22. Koskela, H.L. *et al.* Somatic STAT3 mutations in large granular lymphocytic leukemia. *N. Engl. J. Med.* **366(20)**, 1905-1913 (2012).
23. Rajala, H.L. *et al.* Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. *Blood* **121(22)**, 4541-4550 (2013).
24. Jerez, A. *et al.* STAT3 mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T-cell large granular lymphocyte leukemia. *Blood* **120(15)**, 3048-3057 (2012).
25. Andersson, E. *et al.* Activating somatic mutations outside the SH2-domain of STAT3 in LGL leukemia. *Leukemia* **30(5)**, 1204-1208 (2016).
26. Swerdlow, S.H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127(20)**, 2375-2390 (2016).
27. Rajala, H.L. *et al.* Uncovering the pathogenesis of large granular lymphocytic leukemia-novel STAT3 and STAT5b mutations. *Ann. Med.* **46(3)**, 114-122 (2014).
28. Küçük, C. *et al.* Activating mutations of STAT5B and STAT3 in lymphomas derived from γδ-T or NK cells. *Nat. Commun.* **6**, 6025 (2015).
29. Clemente, M.J. *et al.* Clonal drift demonstrates unexpected dynamics of the T cell repertoire in T-large granular lymphocyte leukemia. *Blood* **118(16)**, 4384-4393 (2011).
30. Chiu, R., Boyle, W.J., Meek, J., Smeal, T., Hunter, T. & Karin, M. The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell* **54**, 541-552 (1988).
31. Sassone-Corsi, P., Sisson, J. & Verma, I.M. Transcriptional autoregulation of the proto-oncogene fos. *Nature* **334(6810)**, 314-319 (1988).
32. Angel, P., Szabowski, A. & Schorpp-Kistner, M. Function and regulation of AP-1 subunits in skin physiology and pathology. *Oncogene* **20(19)**, 2413-2423 (2001).
33. Shen, Q. *et al.* The AP-1 transcription factor regulates breast cancer cell growth via cyclins and E2F factors. *Oncogene* **27**, 366-377 (2008).
34. Hudson, G.M., Watson, P.J., Fairall, L., Jamieson, A.G. & Schwabe, J.W. Insights into the recruitment of class IIa histone deacetylases (HDACs) to the SMRT/NCoR transcriptional repression complex. *J. Biol. Chem.* **290(29)**, 18237-18244 (2015).
35. Zhang, L. *et al.* SpliceArray profiling of breast cancer reveals a novel variant of NCOR2/SMRT that is associated with Tamoxifen resistance and control of ERα transcriptional activity. *Cancer Res.* **71(1)**, 246-255 (2013).

36. Smith, C.L. *et al.* Elevated nuclear expression of the SMRT corepressor in breast cancer is associated with earlier tumor recurrence. *Breast Cancer Res. Treat.* **136(1)**, 253-265 (2012).
37. Fernández-Majada, V. *et al.* Aberrant cytoplasmic localization of N-CoR in colorectal tumors. *Cell Cycle* **6(14)**, 1748-1752 (2007).
38. Tacutu, R. *et al.* Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.* **41**, 1027-1033 (2013).
39. Zhang, W. *et al.* Functional variants in Notch pathway genes NCOR2, NCSTN and MAML2 predict survival of patients with cutaneous melanoma. *Cancer Epidemiol. Biomarkers Prev.* **24(7)**, 1101-1110 (2015).
40. Mengeling, B.J., Phan, T.Q., Goodson, M.L. & Privalsky, M.L. Aberrant corepressor interactions implicated in PML-RARα and PLZF-RARα leukemogenesis reflect an altered recruitment and release of specific NCoR and SMRT splice variants. *J. Biol. Chem.* **286(6)**, 4236-4247 (2011).
41. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat. Rev. Immunol.* **13(2)**, 88-100 (2013).
42. Jeltsch, M. *et al.* Hyperplasia of lymphatic vessels in VEGF-C transgenic mice. *Science* **276(5317)**, 1423-1425 (1997).
43. Yonemura, Y. *et al.* Role of VEGF-C and VEGF-D in lymphangiogenesis in gastric cancer. *Int. J. Clin. Oncol.* **10(5)**, 318-327 (2005).
44. Skobe, M. *et al.* Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nat. Med.* **7(2)**, 192-198 (2001).
45. Liu, J. *et al.* Associations between the expression of MTA1 and VEGF-C in esophageal squamous cell carcinoma with lymph angiogenesis and lymph node metastasis. *Oncol. Lett.* **14(3)**, 3275-3281 (2017).
46. Fankhauser, M. *et al.* Tumor lymphangiogenesis promotes T cell infiltration and potentiates immunotherapy in melanoma. *Sci. Transl. Med.* **9(407)**, pii: eaal4710 (2017).
47. Tami, C. *et al.* Immunoglobulin A (IgA) is a natural ligand of hepatitis A virus cellular receptor 1 (HAVCR1) and the association of IgA with HAVCR1 enhances virus-receptor interactions. *J. Virol.* **81(7)**, 3437-3446 (2007).
48. Telford, E.J., Jiang, W.G. & Martin, T.A. HAVcR-1 involvement in cancer progression. *Histol. Histopathol.* **32(2)**, 121-128 (2017).
49. Cuadros, T. *et al.* HAVCR/KIM-1 activates the IL-6/STAT-3 pathway in clear cell renal cell carcinoma and determines tumor progression and patient outcome. *Cancer Res.* **74(5)**, 1416-1428 (2014).
50. Wlodarski, M.W. *et al.* Phenotypic differences between healthy effector CTL and leukemic LGL cells support the notion of antigen-triggered clonal transformation in T-LGL leukemia. *J. Leukoc. Biol.* **83(3)**, 589-601 (2008).
51. De Libero, G. Control of gammadelta T cells by NK receptors. *Microbes Infect.* **1(3)**, 263-267 (1999).
52. Moris, A., Rothenfusser, S., Meuer, E., Hangretinger, R. & Fisch, P. Role of γδ T cells in tumor immunity and their control by NK receptors. *Microbes Infect.* **1(3)**, 227-234 (1999).

53. Zheng, Z., Venkatapathy, S., Rao G. & Harrington, C.A. Expression profiling of B cell chronic lymphocytic leukemia suggests deficient CD1-mediated immunity, polarized cytokine response, altered adhesion and increased intracellular protein transport and processing of leukemic cells. *Leukemia* **16(12)**, 2429-2437 (2002).
54. Jelinek, D.F. *et al.* Identification of a global gene expression signature of B-chronic lymphocytic leukemia. *Mol. Cancer Res.* **1(5)**, 346-361 (2003).
55. Weiler, S., Ademokun, J.A. & Norton, J.D. ID helix-loop-helix proteins as determinants of cell survival in B-cell chronic lymphocytic leukemia cells *in vitro. Mol. Cancer* **14(1)**, 30 (2015).
56. Norton, J.D. ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. *J. Cell. Sci.* **113(Pt 22)**, 3897-3905 (2000).
57. Krämer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30(4)**, 523-530 (2014).
58. Blom, B. *et al.* Disruption of αβ but not γδ T cell development by overexpression of the helix-loop-helix protein Id3 in committed T cell progenitors. *EMBO J.* **18**, 2793-2802 (1999).
59. Lauritsen, J.P.H. *et al.* Differential induction of Id3 signals lineage divergence, Notch-independent differentiation, and functional maturation of γδ T cells. *Immunity* **31(4)**, 565-575 (2009).
60. Kreslavsky T., Gleimer, M., Garbe, A.I. & von Boehmer, H. αβ versus γδ fate choice: counting the T-cell lineages at the branch point. *Immunol. Rev.* **238(1)**, 169-181 (2010).
61. Siegers, G.M. *et al.* Different composition of the human and the mouse gammadelta T cell receptor explains different phenotypes of CD3gamma and CD3delta immunodeficiencies. *J. Exp. Med.* **204(11)**, 2537-2544 (2007).
62. Pang, D.J., Neves, J.F., Sumaria, N. & Pennington, D.J. Understanding the complexity of γδ T-cell subsets in mouse and humans. *Immunology* **136(3)**, 283-290 (2012).
63. Lamy, T., Liu, J.H., Landowski, T.H., Dalton, W.S. & Loughran, T.P. Jr. Dysregulation of CD95/CD95 ligand-apoptotic pathway in CD3(+) large granular lymphocyte leukemia. *Blood* **92(12)**, 4771-4777 (1998).
64. Watters, R.J., Liu, X. & Loughran, T.P. Jr. T-cell and natural killer-cell large granular lymphocyte leukemia neoplasias. *Leuk. Lymphoma* **52(12)**, 2217-2225 (2011).
65. Zhang, R., Shah, M.V. & Loughran, T.P. Jr. The root of many evils: indolent large granular lymphocyte leukemia and associated disorders. *Hematol. Oncol.* **28(3)**, 105-117 (2010).
66. Barrdahl, M. *et al.* Gene-environment interactions involving functional variants: results from the breast cancer association consortium. *Int. J. Cancer* **141(9)**, 1830-1840 (2017).
67. Sarras, H., Alizadeh Azami, S. & McPherson, J.P. In search of a function for BCLAF1. *Scientific World-Journal* **10**, 1450-1461 (2010).
68. Kasof, G.M., Goyal, L. & White, E. Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins. *Mol. Cell. Biol.* **19(6),** 4390-4404 (1999).
69. Zhou, X. *et al.* BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells. *Nat. Commun.* **5(5)**, 4581 (2014).

70. McPherson, J.P. *et al.* Essential role for Bclaf1 in lung development and immune system function. *Cell Death Differ.* **16(2)**, 331-339 (2009).
71. Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F. & Pals, G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30(12)**, e57 (2002).
72. Eldering, E. *et al.* Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signaling pathways. *Nucleic Acids Res.* **31(23)**, e153 (2003).
73. Imai, T. *et al.* Identification and molecular characterization of fractalkine receptor CX3CR1, which mediates both leukocyte migration and adhesion. *Cell* **91**, 521 (1997).
74. Nishimura, M. *et al.* Dual functions of fractalkine/CX3C ligand 1 in trafficking of perforin+/granzyme B+ cytotoxic effector lymphocytes that are defined by CX3CR1 expression. *J. Immunol.* **168(12)**, 6173-6180 (2002).
75. Tian, Y. Epigenetic regulation of pregnane X receptor activity. *Drug Metab. Rev.* **45(2)**, 166-172
76. Chan, H.J., Haiqing, L., Zheng, L., Yuan, Y-C., Mortimer, J. & Chen, S. SERPINA1 is a direct estrogen receptor target gene and a predictor of survival in breast cancer patients. *Oncotarget* **6(28)**, 25815-25827 (2015).
77. Pires, B.R.B., Silva, R.C.M.C., Ferreira, G.M. & Abdelhay, E. NF-kappaB: two sides of the same coin. *Genes (Basel)* **9(1)**, E24 (2018).
78. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44(10)**, 1084-1089 (2012).
79. Hu, Y-J., Sun, W., Tzeng, J-Y. & Perou, C.M. Proper use of allele-specific expression improves statistical power for *cis*-eQTL mapping with RNA-seq data. *J. Am. Stat. Assoc.* **110(511)**, 962-974 (2015).
80. Sims, D., Sudbery, I., Ilott, N.E., Heger, A. & Ponting, C.P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121-132 (2014).
81. Schneeberger, K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* **15**, 662-676 (2014).
82. Haas, B.J. & Zody, M.C. Advancing RNA-Seq analysis. *Nat. Biotechnol.* **28(5)**, 421-423 (2010). (In Fig.4 Legend)
83. Doudna, J.A. & Charpentier, E. Genome editing, The new frontier of genome engineering with CRISPR-Cas9. *Science* **346(6213)**, 1258096 (2014).
84. Yohe, S. & Thyagarajan, B. Review of clinical next-generation sequencing. *Arch. Pathol. Lab. Med.* **141(11)**, 1544-1557 (2017).
85. Matthijs, G. *et al.* Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* **24**, 2-5 (2016).
86. Xiang, Z. & Tu, W. Dual face of Vγ9Vδ2-T cells in tumor immunology: anti- versus pro-tumoral activities. *Front. Immunol.* **8**, 1041 (2017).

87. Poggi, A. & Zocchi, M.R. Gammadelta T lymphocytes as a first line of immune defense: old and new ways of antigen recognition and implications for cancer immunotherapy. *Front. Immunol.* **5**, 575 (2014).
88. Chien, Y.H., Meyer, C. & Bonneville, M. Gammadelta T cells: first line of defense and beyond. *Annu. Rev. Immunol.* **32**, 121-155 (2014).
89. Fisch, P. *et al.* Recognition by human V gamma 9 / V delta 2 T cells of a GroEL homolog on Daudi Burkitt's lymphoma cells. *Science* **250(4985)**, 1269-1273 (1990).
90. Tanaka, Y., Morita, C.T., Tanaka, Y., Nieves, E., Brenner, M.B. & Bloom, B.R. Natural and synthetic non-peptide antigens recognized by human gamma delta T cells. *Nature* **375(6526)**, 155-158 (1995).
91. Gober, H-J., Kistowska, M., Angman, L., Jenö, P., Mori, L. & De Libero, G. Human T cell receptor γδ cells recognize endogenous mevalonate metabolites in tumor cells. *J. Exp. Med.* **197(2)**, 163-168 (2003).
92. Kunzmann, V., Bauer, E. & Wilhelm, M. Gamma/delta T-cell stimulation by pamidronate. *N. Engl. J. Med.* **340(9)**, 737-738 (1999).
93. Kunzmann, V., Bauer, E., Feurle, J., Weissinger, F., Tony, H.P. & Wilhelm, M. Stimulation of gammadelta T cells by aminobisphosphonates and induction of antiplasma cell activity in multiple myeloma. *Blood* **96(2)**, 384-392 (2000).
94. Mariani, S. *et al.* Effector γδ T cells and tumor cells as immune targets of zoledronic acid in multiple myeloma. *Leukemia* **19**, 664-670 (2005).
95. Watanabe, N. *et al.* Type I IFN-mediated enhancement of anti-leukemic cytotoxicity of γδ T cells expanded from peripheral blood cells by stimulation with zoledronate. *Cytother.* **8(2)**, 118-129 (2006).
96. Saitoh, A. *et al.* Anti-tumor cytotoxicity of γδ T cells expanded from peripheral blood cells of patients with myeloma and lymphoma. *Med. Oncol.* **25**, 137-147 (2008).
97. Zou, C., Zhao, P., Xiao, Z., Han, X., Fu, F. & Fu, L. γδ T cells in cancer immunotherapy. *Oncotarget* **8(5)**, 8900-8909 (2017).
98. Chitadze, G., Oberg, H.H., Wesch, D. & Kabelitz, D. The ambiguous role of γδ T lymphocytes in antitumor immunity. *Trends Immunol.* **38(9)**, 668-678 (2017).
99. Gentles, A.J. *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* **21**, 938-945 (2015).
100. Murugaiyan, G. & Saha, B. Protumor vs. antitumor functions of IL-17. *J. Immunol.* **183(7)**, 4169-4175 (2009).
101. Lafont, V. *et al.* Plasticity of γδ T cells: impact on the anti-tumor response. *Front. Immunol.* **5**, 622 (2014).
102. Garrido, P. *et al.* Monoclonal TCR-Vbeta13.1+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis: evidence for an antigen-driven chronic T-cell stimulation origin. *Blood* **109**, 4890-4898 (2007).
103. Yoshida, K. *et al.* Aging-related changes in human T-cell repertoire over 20 years delineated by deep sequencing of peripheral T-cell clones. *Exp. Gerontol.* **96**, 29-37 (2017).
104. Xu, W. & Larbi, A. Markers of T cell senescence in humans. *Int. J. Mol. Sci.* **18(8),** 1742 (2017).

105. Nikolich-Zugich, J. & van Lier, R.A.W. Cytomegalovirus (CMV) research in immune senescence comes of age: overview of the 6th international workshop on CMV and immunosenescence. *GeroScience* **39(3)**, 245-249 (2017).
106. Weltevrede, M., Eilers, R., de Melker, H.E. & van Baarle, D. Cytomegalovirus persistence and T-cell immunosenescence in people aged fifty and older: a systematic review. *Exp. Gerontol.* **77**, 87-95 (2016).
107. Looney, R.J. *et al.* Role of cytomegalovirus in the T cell changes seen in elderly individuals. *Clin. Immunol.* **90(2)**, 213-219 (1999).
108. Goodrum, F. Human cytomegalovirus latency: approaching the Gordian knot. *Annu. Rev. Virol.* **3(1)**, 333-357 (2016).
109. Karrer, U. *et al.* Memory inflation: continuous accumulation of antiviral CD8+ T cells over time. *J. Immunol.* **170(4)**, 2022-2029 (2003).
110. Lilley, B.N. & Ploegh, H.L. Viral modulation of antigen presentation: manipulation of cellular targets in the ER and beyond. *Immunol. Rev.* **207**, 126-144 (2005).
111. Stern-Ginossar, N. *et al.* Host immune system gene targeting by a viral miRNA. *Science* **317(5836)**, 376-381 (2007).
112. Jackson, S.E. *et al.* CMV immune evasion and manipulation of the immune system with aging. *GeroScience* **39(3)**, 273-291 (2017).
113. Grant, B.D. & Donaldson, J.G. Pathways and mechanisms of endocytic recycling. *Nat. Rev. Mol. Cell. Biol.* **10(9)**, 597-608 (2009).
114. Zagorac, G.B., Mahmutefendic, H., Tomas, M.I., Kucic, N., Le Bouteiller, P. & Lucin, P. Early endosomal rerouting of major histocompatibility class I conformers. *J. Cell. Physiol.* **227(7)**, 2953-2964 (2012).
115. Lucin, P., Mahmutefendic, H., Blagojevic Zagorac, G. & Ilic Tomas, M. Cytomegalovirus immune evasion by perturbation of endosomal trafficking. *Cell. Mol. Immunol.* **12(2)**, 154-169 (2015).
116. Neefjes, J., Jongsma, M.L., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11(12)**, 823-836 (2011).
117. Porcelli, S.A., Morita, C.T. & Modlin, R.L. T-cell recognition of non-peptide antigens. *Curr. Opin. Immunol.* **8(4)**, 510-516 (1996).
118. Kabelitz, D., Glatzel, A. & Wesch, D. Antigen recognition by human gammadelta T lymphocytes. *Int. Arch. Allergy Immunol.* **122(1)**, 1-7 (2000).
119. Born, W.K., Aydintug, M.K. & O'Brien, R.L. Diversity of γδ T-cell antigens. *Cell. Mol. Immunol.* **10(1)**, 13-20 (2013).
120. Wingren, C., Crowley, M.P., Degano, M., Chien, Y. & Wilson, I.A. Crystal structure of a gammadelta T cell receptor ligand T22: a truncates MHC-like fold. *Science* **287(5451)**, 310-314 (2000).
121. Adams, E.J., Chien, Y.H. & Garcia K.C. Structure of a gammadelta T cell receptor in complex with the nonclassical MHC T22. *Science* **308(5719)**, 227-231 (2005).
122. Matis, L.A., Cron, R. & Bluestone, J.A. Major histocompatibility complex-linked specificity of gamma delta receptor-bearing T lymphocytes. *Nature* **330(6145)**, 262-264 (1987).

123. Matis, L.A., Fry, A.M., Cron, R.Q., Cotterman, M.M., Dick, R.F. & Bluestone, J.A. Structure and specificity of a class II MHC alloreactive gamma delta T cell receptor heterodimer. *Science* **245(4919)**, 746-749 (1989).
124. Déchanet, J. *et al.* Major expansion of gammadelta T lymphocytes following cytomegalovirus infection in kidney allograft recipients. *J. Infect. Dis.* **179(1)**, 1-8 (1999).
125. Snydman, D.R., Rubin, R.H. & Werner, B.G. New developments in cytomegalovirus prevention and management. *Am. J. Kidney Dis.* **21**, 217-228 (1993).
126. Bukowski, J.F., Morita, C.T. & Brenner, M.B. Recognition and destruction of virus-infected cells by human γδ CTL. *J. Immunol.* **153**, 5133-5140 (1994).
127. Déchanet, J. *et al.* Implication of γδ T cells in the human immune response to cytomegalovirus. *J. Clin. Invest.* **103(10)**, 1437-1449 (1999).
128. Roux, A. *et al.* Differential impact of age and cytomegalovirus infection on the γδ T cell compartment. *J. Immunol.* **191(3)**, 1300-1306 (2013).
129. Wistuba-Hamprecht, K., Frasca, D., Blomberg, B., Pawelec, G. & Derhovanessian, E. Age-associated alterations in γδ T-cells are present predominantly in individuals infected with cytomegalovirus. *Immun. Ageing* **10(1)**, 26 (2013).
130. Wistuba-Hamprecht, K., Haehnel, K., Janssen, N., Demuth, I. & Pawelec, G. Peripheral blood T-cell signatures from high-resolution immune phenotyping of γδ and αβ T-cells in younger and older subjects in the Berlin Aging Study II. *Immun. Ageing* **12**, 25 (2015).
131. Couzi, L., Pitard, V., Moreau, J.F., Merville, P. & Déchanet-Merville, J. Direct and indirect effects of cytomegalovirus-induced γδ T cells after kidney transplantation. *Front. Immunol.* **6**, 3 (2015).
132. Slavuljica, I., Krmpotic, A. & Jonjic, S. Manipulation of NKG2D ligands by cytomegaloviruses: impact on innate and adaptive immune response. *Front. Immunol.* **2**, 85 (2011).
133. Yassai, M.B., Naumov, Y.N., Naumova, E.N. & Gorski, J. A clonotype nomenclature for T cell receptors. *Immunogenet.* **61(7),** 493-502 (2009).
134. Vermijlen, D. *et al.* Human cytomegalovirus elicits fetal γδ T cell responses in utero. *J. Exp. Med.* **207(4)**, 807-821 (2010).
135. Godder, K.T. *et al.* Long term disease-free survival in acute leukemia patients recovering with increased gammadelta T cells after partially mismatched related donor bone marrow transplantation. *Bone Marrow Transplant.* **39(12)**, 751-757 (2007).
136. Wherry, E.J. T-cell exhaustion. *Nat. Immunol.* **12(6)**, 492-499 (2011).
137. Michaud, M. *et al.* Proinflammatory cytokines, ageing, and age-related diseases. *J. Am. Med. Dir. Assoc.* **14(12)**, 877-882 (2013).
138. Franceschi, C. *et al.* Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* **908**, 244-254 (2000).
139. Franceschi, C. *et al.* Immunobiography and the heterogeneity of immune responses in the elderly: a focus on inflammaging and trained immunity. *Front. Immunol.* **8**, 982 (2017).

140. Bouthry, E., Picone, O., Hamdi, G., Grangeot-Keros, L., Ayoubi, J.M. & Vauloup-Fellous, C. Rubella and pregnancy: diagnosis, management and outcomes. *Prenat. Diagn.* **34(13)**, 1246-1253 (2014).
141. Van den Heuvel, D. *et al.* Persistent subclinical immune defects in HIV-1-infected children treated with antiretroviral therapy. *AIDS* **29(14)**, 1745-1756 (2015).
142. Aspinall, R. & Andrew, P. Thymic involution in aging. *J. Clin. Immunol.* **20(4)**, 250-256 (2000).
143. Pereira, B.I. & Akbar, A. Convergence of innate and adaptive immunity during human aging. *Front. Immunol.* **7**, 445 (2016).
144. Lamy, T. & Loughran, T.P. Jr. How I treat LGL leukemia. *Blood* **117(10)**, 2764-2774 (2011).
145. Chattopadhyay, P.K. & Roederer, M. Good cell, bad cell: flow cytometry reveals T-cell subsets important in HIV disease. *Cytometry A.* **77(7)**, 614-622 (2010).
146. Krammer, P.H., Arnold, R. & Lavrik, I.N. Life and death in peripheral T cells. *Nat. Rev. Immunol.* **7**, 532-542 (2007).
147. Paley, M.A. *et al.* Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science* **338**, 1220-1225 (2012).
148. Nayar, R. *et al.* IRF4 regulates the ratio of T-Bet to Eomesodermin in CD8+ T cells responding to persistent LCMV infection. *PLoS One* **10(12)**, e0144826 (2015).
149. Kao, C. *et al.* Transcription factor T-bet represses expression of the inhibitory receptor PD-1 and sustains virus-specific CD8+ T cell responses during chronic infection. *Nat. Immunol.* **12(7)**, 663-671 (2011).
150. Wherry, E.J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15(8)**, 486-499 (2015).
151. Shin, H. *et al.* A role for the transcriptional repressor Blimp-1 in CD8(+) T cell exhaustion during chronic viral infection. *Immunity* **31(2)**, 309-320 (2009).
152. Balint, G.P. & Balint, P.V. Felty's Syndrome. *Best Practice Res. Clin, Rheumatol.* **18(5)**, 631-645 (2004).
153. Savola, P. *et al.* Somatic STAT3 mutations in the Felty syndrome: an implication for a common pathogenesis with large granular lymphocyte leukemia. *Haematologica* **103(2)**, 304-312 (2018).
154. Langerak, A.W. & Assmann, L.J.L.C. Large granular lymphocyte cells and immune dysregulation diseases – the chicken or the egg? *Haematologica* **103(2)**, 193-194 (2018).
155. Beeson, P.B. Age and sex associations of 40 autoimmune diseases. *Am. J. Med.* **96(5)**, 457-462 (1994).
156. Negoescu, A. & Ostör, A.J. Early recognition improves prognosis in elderly onset RA. *Practitioner* **258(1767)**, 11-14 (2014).

Addendum

ABBREVIATIONS

%GC	Percentage of Guanine and Cytosine nucleotides
2B4	Natural killer cell receptor 2B4, CD244
5'RACE	5' rapid amplification of cDNA ends
ABL	Abelson murine leukemia viral oncogene homolog 1, housekeeping gene
AICD	Activation-induced cell death
AIHA	Autoimmune hemolytic anemia
AIRE	Autoimmune regulator
ALDH	Aldehyde dehydrogenase
AML	Acute myeloid leukemia
AML1	Acute myeloid leukemia 1 protein or Runt-related transcription factor 1 (RUNX1)
ANNOVAR	Annotate variation
ANOVA	Analysis of variance
APC	Antigen-presenting cell
BAFF	B cell activating factor (TNFSF13B)
BCL2L1	B cell lymphoma 2 (BCL2) like 1
BCLAF1	BCL2-associated transcription factor 1
BDT	Big dye terminator
BER	Base excision repair
BLIMP1	Transcription factor encoded by PR domain zinc finger protein 1 (PRDM1) gene
BM	Bone marrow
BMTx	Bone marrow transplantation
BRCA	Breast cancer susceptibility protein
BTN	Butyrophilin
BWA	Burrows-Wheeler Aligner
C	Constant
CASP1	Caspase-1
CB	Cord blood
CCR	C-C chemokine receptor type
CCR7	C-C chemokine receptor type 7 (CD197)
CD	Cluster of differentiation
cDNA	Copy- or complement DNA
CDR3	Complementarity determining region 3
CFLAR	Caspase-8 and Fas-associated protein with death domain (FADD) like apoptosis regulator
cis-eQTL	Expression quantitative trait locus/loci on the same (cis) chromosome
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
CMV	Cytomegalovirus
COG	Clusters of orthologous group analysis
COSMIC	Catalogue of somatic mutations in cancer
CREB3L2	cAMP response element-binding protein 3-like 2
CRISPR	Clustered regularly interspaced short palindromic repeats
CSA	Cyclosporine A
Ct	Threshold cycle
cTEC	Cortical thymic epithelial cell
CTL	Cytotoxic T lymphocyte
CX3CR1	CX3C chemokine receptor 1
CXCL	C-X-C chemokine ligand
CYP	Cytochrome P enzyme
D	Diversity
DAVID	Database for annotation, visualization and integrated discovery
dbNSFP	Database for nonsynonymous SNVs' functional predictions
DC	Dendritic cell
DDR	DNA damage response
DETC	Dendritic epidermal T cell
DLBCL	Diffuse large B cell lymphoma

DN	Double negative
DNA	Deoxyribonucleic acid
DP	Double positive
E12	TCF3 transcription factor
EBV	Epstein-Barr virus
EOMES	Eomesodermin
ETS2	Erythroblastosis virus E26 oncogene (transcription factor)
FACS	Fluorescence-activated cell sorting
FATHMM	Functional analysis through hidden Markov Models
FC	Fold change
FSC	Forward scatter
FOXP3	Forkhead box P3
FOXO4	Forkhead box O4
GATK	Genome Analysis Toolkit
GCP	Global canonical pathways
gDNA	genomic DNA
GERP	Genomic evolutionary rate profiling
GEP	Gene expression profiling
GFA	Global functional analysis
GO	Gene Ontology
GPA	Granulomatosis polyangiitis
HAVCR1	Hepatitis A virus cellular receptor 1 (TIM-1)
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCL	Hairy cell leukemia
HCQ	Hydroxychloroquine
HCV	Hepatitis C virus
HE	Hematoxylin-eosin staining
Hg38	Human genome version 38
HIF-1α	Hypoxia-inducible factor 1-alpha heterodimer subunit
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HM-BPP	Hydroxymethyl-but-2-enyl-pyrophosphate
HMG	High mobility box
HNF3	Hepatocyte nuclear factor (FOXA1)
HSV	Herpes simplex virus
HTLV-1	Human T cell lymphotropic virus type 1
ID3	Inhibitor of DNA binding 3
IEL	Intra-epithelial lymphocyte
IFN	Interferon
IFNG	Interferon gamma gene
IFNGR2	Interferon gamma receptor 2 gene
Ig	Immunoglobulin
IL	Interleukin
IL7Rα	Interleukin 7 receptor alpha subunit
IMGT	ImMunoGeneTics database
Indel	Insertion/deletion
IPA	Ingenuity Pathway Analysis
IRS2	Insulin receptor substrate 2
ISP	Immature single positive
ITP	Idiopathic thrombocytopenic purpura
J	Joining
JAK	Janus kinase
KEGG	Kyoto encyclopedia of genes and genomes
KIM1	Kidney injury molecule 1
KLF4	Kruppel-like factor 4
KLRF1	Killer cell lectin-like receptor subfamily F member 1
KLRG1	Killer cell lectin-like receptor subfamily G member 1
LAG3	Lymphocyte-activation gene 3

LEF1	Lymphoid enhancer binding factor 1
LGL	Large granular lymphocyte
ljb2B	Liu, Jian & Boerwinkle algorithm version 2B
LN	Lymph node
lncRNA	Long noncoding RNA
LRT	Likelihood ratio test
LTB	Lymphotoxin β
LXR/RXR	Liver X receptor / retinoid X receptor
MAD	Median absolute deviation
MAF	Minor allele frequency
MAPK	Mitogen-activated protein kinase
MEC	Medical Ethics Committee
MDSC	Myeloid-derived suppressor cell
MEIS1	Meis homeobox 1
MHC	Major histocompatibility complex
MLPA	Multiplex ligation-dependent probe amplification
MNC	Mononuclear cell
mRNA	messenger RNA
miRNA	micro RNA
mTEC	Medullary thymic epithelial cell
MTX	Methotrexate
nBP	Aminobisphosphonate
NCOA	Nuclear receptor coactivator
NCOR2	Nuclear receptor corepressor 2
ncRNA	noncoding RNA
NF1	Neurofibromin 1 oncogene
NFAT	Nuclear factor of activated T cells
NFKB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next-generation sequencing
NK cell	Natural killer cell
NKG2D	Natural killer group 2D
NO	Nitric oxide
NUDT11	Nudix hydrolase 11
NWO	Normaal waarden ouderen
OMIM	Online Mendelian inheritance in man
P53	TP53, tumor protein 53
PAX4	Paired box gene 4
PB	Peripheral blood
PBM(N)C	Peripheral blood mononuclear cell
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD1	Programmed cell death 1
PDL1	Programmed cell death ligand 1
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (p110α)
PRCA	Pure red cell aplasia
PRF1	Perforin 1
PXR	Pregnane X receptor
RA	Rheumatoid arthritis
RAG	Recombination-activating gene
RBC	Red blood cell
RMA	robust multi-array average
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing technology
ROS	Reactive oxygen species
RQ/RT-PCR	Realtime quantitative polymerase chain reaction
RSS	Recombination signal sequence
RT-MLPA	Reverse transcriptase multiplex ligation-dependent probe amplification
S1P	Sphingosine-1-phosphate
S1PR1	Sphingosine-1-phosphate receptor 1

SAM	Significance analysis of microarrays
SH2	Src homology domain 2
siRNA	silencing RNA
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOX4	SRY-box 4
SP1	Specificity protein 1 transcription factor
spp.	Species (plural)
SSC	Side scatter
STAT	Signal transducer and activator of transcription
Ta	Annealing temperature
TACI	Transmembrane activator and CAML interactor
TAP	Transporter associated with antigen processing
T-BET	T-box transcription factor (Tbx21)
TCF1	TCF7 in humans
TCR	T cell receptor
TCUS	T cell clonopathy of undetermined significance
TdT	Terminal deoxynucleotidyl transferase
TemRA	CD45RA+CD45RO- effector T cell
TemRO	CD45RA-CD45RO+ effector memory T cell
TF	Transcription factor
TFT	Transcription factor target
Thr	Threshold
Thy	Thymus
TIL	Tumor infiltrating lymphocyte
Ti/Tv	Transition (A↔G, C↔T) / transversion (A↔C, A↔T, G↔C, G↔T)
TIM3	T cell immunoglobulin and mucin-domain containing 3, also known as HAVCR2
T-LGL	T cell large granular lymphocyte
TLR	Toll-like receptor
TNF-α	Tumor necrosis factor alpha
TRA	T cell receptor alpha (α)
trans-eQTL	Expression quantitative trait locus/loci on a different (trans) chromosome
TRB	T cell receptor beta (β)
TRD	T cell receptor delta (δ)
TREM1	Triggering receptor expressed on myeloid cells 1
TRG	T cell receptor gamma (γ)
UTR	Untranslated region
V	Variable
VEGFA	Vascular endothelial growth factor A
VEGFC	Vascular endothelial growth factor C
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
XIAP	X-linked inhibitor of apoptosis

SUMMARY

T cell receptor (TCR) γδ+ T cells form a distinct T cell type, which follows a different path of development and exerts different types of immune responses to a wide variety of peptide and non-peptide antigens. Just like any other immune cell, TCRγδ+ T cells are highly likely to be subjected to the effect of ageing. Ageing has been considered to be increasingly important, regarding the higher life expectancy in the Western World. For many immune cell types it has been well described how ageing affects composition, functionality and the possible gain of aberrancies of immune cells. Remarkably, it is still largely unknown how ageing impacts on TCRγδ+ T cells. As TCRγδ+ T cells show a high level of similarity to CD8+TCRαβ+ cytotoxic T lymphocytes, regarding their responses, we hypothesized that TCRγδ+ T cells show a similar impact of ageing as CD8+TCRαβ+ T cells, characterized by reduced responses, decreased (absolute) numbers and altered subset compositions.

In our immunophenotyping study in **Chapter 2** we observed that upon ageing the absolute numbers of total TCRγδ+ T cells decreased drastically, especially the most common population in the peripheral blood: Vγ9/Vδ2 cells. Furthermore, subset distributions shifted towards effector memory and effector cells, while the fractions of especially naive and central memory cell populations decreased. These effects were enhanced upon persistence of CMV, a virus which is known for its effects on the T cell compartment. An additional CMV effect was observed, leading to a reduction of Vγ9/Vδ2 cells, and therefore a relative increase of Vδ1+, mostly late-stage effector cells.

In order to further investigate the effect of ageing and the change in TCRγδ usage we applied next generation sequencing to study the repertoire of TCRγδ+ T cells from different ages (starting from developmental stages in thymus and cord blood to circulating cells in young and elderly individuals) and from different subsets (including immature, naive mature, central and effector memory, and effector cells). Prior to these investigations we first optimized TCR gamma (TRG) and delta (TRD) NGS assays in **Chapter 3** to reduce the effect of technical biases. Following titration of primer concentrations, establishing the most optimal annealing temperatures to reduce primer competition, and reduction of cycle numbers to prevent plateau effects in the PCR reaction, we were able to generate results that are representative for the actual repertoire composition. Following this optimization, we could conclude from our biological study, as presented in **Chapter 4**, that naive TCRγδ+ T cells maintain a highly diverse repertoire during development and ageing, as seen from thymus and cord blood TCRγδ+ T cells and circulating TCRγδ+ T cells from young and elderly individuals. The most clear changes upon ageing were observed in memory cell populations: young individuals showed a Vγ9/Vδ2 dominance in circulating TCRγδ+ T cells, while elderly individuals showed more a

Vγ2/Vδ1 dominance. These immunogenotyping data thus confirmed our immunophenotyping data, especially with regards to the effect of both ageing and CMV persistence. Based on both the immunophenotyping and –genotyping datasets, we thus could show that TCRγδ+ T cells are subjected to ageing in a way that is very similar to CD8+TCRαβ+ T cells.

In elderly individuals chronic leukemias and lymphomas may develop, potentially as a consequence of ageing processes. These chronic lymphoid malignancies mainly involve B cells, but T cells can also transform into chronic leukemias and/or lymphomas. One of these is the T cell large granular lymphocyte (T-LGL) leukemia, which is a relatively rare and heterogeneous disorder consisting of three variants based on the T cell type that is involved in the disease, i.e. CD4+TCRαβ+, CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia. As the etiology for CD4+TCRαβ+ T-LGL leukemia has been extensively described before, we reviewed existing knowledge on the TCRγδ+ and CD8+TCRαβ+ T-LGL leukemia variants in **Chapter 5**. Both T-LGL leukemias mostly involve elderly individuals (above age 60), and have a rather indolent disease course, but a direct cause of both types of leukemia has not yet been identified. Most patients display associated cytopenias, autoimmune diseases and other malignancies, which could be considered as cause or consequence of the disease. Based on the obvious lack of identity or even homology in TCR usage between patients, involvement of a common (auto)antigen in T-LGL leukemia seems less likely.

Based on the fact that TCRγδ+ T-LGL leukemia is a chronic disorder affecting mostly elderly individuals, we hypothesized that the initial cause might lie in a continuous stimulation of these cells, which would then have to be followed by genetic aberrations. In **Chapter 6** we therefore investigated, whether underlying molecular mechanisms such as signaling pathways, could be aberrant and thus contributing to leukemia development. To this end, we performed gene expression profiling studies and observed a high correlation between the TCRγδ+ T-LGL leukemia transcriptome and that of healthy effector and effector memory TCRγδ+ T cells, but not of other healthy TCRγδ+ T cell subsets. More in-depth comparison between effector TCRγδ+ T cells and TCRγδ+ T-LGL leukemia cells revealed an altered expression of genes involved in increased proliferation, reduced apoptosis and aberrant immune cell responses in the latter. This could not only explain the enhanced proliferative capacity and additional apoptosis resistance of these leukemic cells, but also enhanced immune responses that can possibly linked to the autoimmune diseases and cytopenias that are seen in these patients.

Altered gene expression profiles might result from external factors, causing overstimulation of cells, thus resulting in stress and altered gene transcription, but might also be caused by intrinsic events involving genetic aberrancies and the loss or gain of function of genes. Nowadays, advances in sequencing technologies allow determining

the full genome of cancer cells, thus providing tools to identify all possible underlying genetic variants. By applying one of these novel technologies, whole exome sequencing (WES), on both healthy control cells and TCRγδ+ T-LGL leukemia cells from the same patients, we were able to determine leukemia-specific gene variants as described in **Chapter 7**. After extensive filtering, we identified six candidate genes which could contribute to the aberrant character of TCRγδ+ T-LGL leukemia cells. Gene variants concerned (non-)frameshift insertions and deletions, which could contribute to altered transcription, immune activation and responses, and metastasis. In validation experiments we then identified some of the variants to be present in CD8+TCRαβ+ T-LGL leukemia cells as well, thereby further linking the two T-LGL leukemia types. Future studies should be directed at functionally validating these gene variants to fully link the genetic aberrancies with altered gene transcription and protein levels, and altered function with respect to proliferation, apoptosis resistance, and/or cytotoxicity against other (blood) cells.

In the **General Discussion** extrinsic factors and cell-intrinsic aspects of TCRγδ+ T-LGL leukemia cells are integrated and critically discussed. Additionally, preliminary data on senescence and exhaustion profiles of T-LGL leukemia cells are presented, which may also be relevant for LGL leukemia disease pathology. Finally, the link between LGL leukemia cells and immune dysregulations such as autoimmune diseases is discussed in view of "the chicken or the egg" question.

In summary, the studies described in this thesis illustrate the complex interactions of genetic aberrations, altered gene expression levels, functional properties, receptor and surface marker expression during TCRγδ+ T cell ontogeny, peripheral (antigenic) selection, ageing and leukemogenesis. The research performed within the framework of this thesis does not only form a basis for better understanding of both normal and aberrant TCRγδ+ T cell development, but might eventually also impact on better diagnosis, prognosis, treatment and prevention possibilities. Finally, this thesis underlines the important role of TCRγδ+ T cells in normal physiology, but also highlights their potential to transform into leukemia; TCRγδ+ T cells should not be ignored or underestimated.

SAMENVATTING

T cellen, een belangrijk type afweercellen, kunnen worden onderverdeeld in twee groepen op basis van hun T cel receptor (TCR), nl. TCR alfa beta (TCRαβ) en TCR gamma delta (TCRγδ) positieve T cellen. Deze twee typen T cellen volgen eigen ontwikkelingspaden en zorgen voor verschillende afweerreacties tegen verschillende soorten (eiwit)moleculen, ook wel antigenen genoemd. TCRαβ+ T cellen – de meest voorkomende T cellen in het bloed – herkennen peptide antigenen gepresenteerd in de context van zogenaamde *major histocompatibility complex* (MHC) moleculen op het oppervlak van cellen. TCRγδ+ T cellen daarentegen hebben deze beperking niet en kunnen ook andersoortige cel-gebonden of zelfs vrije antigenen herkennen. Net als alle andere cellen zijn ook TCRγδ+ T cellen onderhevig aan het verouderingsproces. Veroudering wordt in toenemende mate een belangrijk onderwerp in de wetenschap, mede dankzij de toegenomen levensverwachting in de Westerse wereld. Voor veel afweercellen is het effect van veroudering duidelijk beschreven; veroudering zorgt voor een veranderde samenstelling van de afweercellen, de cellen hebben een veranderde functionaliteit en activiteit, en ze kunnen genetische afwijkingen verkrijgen. Het is echter nog vrij onduidelijk wat er met TCRγδ+ T cellen gebeurt tijdens veroudering. Hoewel er duidelijke verschillen zijn tussen TCRαβ+ en TCRγδ+ T cellen, worden ook diverse overeenkomsten gezien tussen CD8+TCRαβ+ cytotoxische T cellen en TCRγδ+ T cellen, m.n. in hun reactie tegen antigeen. Op basis hiervan veronderstelden wij dat TCRγδ+ T cellen eenzelfde manier van veroudering ondergaan als CD8+TCRαβ+ cytotoxische T cellen, nl. afwijkende afweerreacties, verlaagde aantallen en veranderde subset samenstellingen.

In onze immunofenotyperingsstudies in **Hoofdstuk 2**, waarin wij op basis van oppervlaktemoleculen cellen in zogenaamde populaties en subsets indelen, zien wij dat tijdens veroudering het absolute aantal TCRγδ+ T cellen in het bloed drastisch daalt en dat dit met name effect heeft op de meest voorkomende TCRγδ+ T cel populatie in het bloed, de Vγ9/Vδ2 cellen. Verder verandert de verdeling in subsets richting effector geheugen en effector cellen, terwijl de fracties van naieve en centraal geheugen cellen afnemen. Deze effecten worden versterkt in ouderen wanneer er sprake is van blijvende aanwezigheid van CMV, een virus dat bekend staat om zijn indrukwekkende effecten op het T cel compartiment van het afweersysteem. In geval van CMV positiviteit wordt een drastische daling van Vγ9/Vδ2 cellen gezien, en daarbij een relatieve stijging van Vδ1+ cellen, die bovendien vrijwel allemaal terminale effector cellen betreffen.

Om de verouderingseffecten op TCRγδ+ T cellen verder te onderzoeken hebben wij daarnaast gebruik gemaakt van *next generation sequencing* (NGS) technologie. Met deze techniek is het mogelijk een bepaald stuk van het genoom van heel veel cellen tegelijkertijd in kaart te brengen. Wij hebben deze NGS technologie toegepast om het T

cel receptor repertoire, ofwel de samenstelling van T cel receptoren, van TCRγδ+ T cellen in verschillende subsets te onderzoeken. Hierbij kijken wij vanaf de ontwikkeling in de thymus en navelstrengbloed, tot circulerende cellen in jong volwassenen en ouderen, en maken onderscheid tussen onrijp/rijp, naief, centraal en effector geheugen, en effector cellen. Om een zo duidelijk en nauwkeurig mogelijk beeld te krijgen van de samenstelling van het repertoire hebben wij eerst de TCR gamma (TRG) en TCR delta (TRD) testen geoptimaliseerd (**Hoofdstuk 3**), om daarmee het effect van technische afwijkingen in de procedure zoveel mogelijk te beperken. Na deze optimalisatie experimenten hebben wij gezien dat naieve TCRγδ+ T cellen hun brede diversiteit behouden, zowel tijdens ontwikkeling in de thymus en navelstrengbloed, als in jong volwassenen en tijdens veroudering (**Hoofdstuk 4**). De meest duidelijke ouderdom-gerelateerde veranderingen worden dan ook gezien in de geheugencellen. Geheugencellen in jonge individuen laten voornamelijk Vγ9/Vδ2 dominantie zien, terwijl deze in ouderen juist meer een Vγ2/Vδ1 dominantie vertonen (**Hoofdstuk 4**). Deze data bevestigen tevens het beeld dat we hadden gekregen vanuit de immunofenotyperingsstudies m.b.t. de effecten van zowel veroudering als CMV positiviteit (**Hoofdstuk 2**). Op basis van de immunofenotyperingsstudies en TCR repertoire data concluderen wij dat TCRγδ+ T cellen op een vergelijkbare manier onderworpen worden aan een verouderingsproces als CD8+TCRαβ+ T cellen.

Tijdens veroudering kunnen tumoren ontstaan, waaronder ook chronische leukemieën (vorm van bloedkanker) en lymfomen (lymfeklierkanker). Deze tumoren van het afweersysteem ontstaan vaak vanuit zogenaamde B cellen, maar ook T cellen kunnen transformeren tot leukemie of lymfoom. Één daarvan is de T cel *large granular lymphoyte* (T-LGL) leukemie, welke een relatief zeldzame ziekte is en bestaat uit drie varianten op basis van het T cel type, nl. CD4+TCRαβ+, CD8+TCRαβ+ en TCRγδ+ T-LGL leukemie. Wij hebben ons voornamelijk gericht op de TCRγδ+ en CD8+TCRαβ+ varianten van de ziekte, welke veel overeenkomsten vertonen (**Hoofdstuk 5**). Beide typen worden met name gezien in oudere individuen (leeftijd vanaf 60 jaar) en kennen een vrij rustig ziektebeloop. De meeste T-LGL leukemiepatiënten laten een tekort aan bepaalde bloedcellen zien (zogenaamde cytopenieën), wat tot bloedarmoede, verlaagde afweer en bloeduitstortingen kan leiden. Verder hebben deze patiënten vaak al bestaande onderliggende (afweer)ziekten in de vorm van autoimmuunziekten of andere tumoren. In tegenstelling tot de CD4+TCRαβ+ variant waarin in verschillende patiënten vaak een zelfde TCR op het oppervlakte wordt gezien en waarin CMV als aanjager is beschreven, worden geen overeenkomsten gevonden tussen de TCR moleculen bij de andere twee T-LGL leukemie varianten (**Hoofdstuk 5**). Dit betekent dat in de TCRγδ+ en CD8+TCRαβ+ T-LGL tumoren, de ongecontroleerde celdeling waarschijnlijk niet door één gemeenschappelijk (auto-) antigeen wordt veroorzaakt, maar dat meerdere antigenen betrokken kunnen zijn.

Omdat TCRγδ+ T-LGL leukemie een chronische ziekte is, die voornamelijk wordt gediagnosticeerd bij ouderen, veronderstelden wij dat continue en langdurige stimulatie van de T cellen een rol speelt in het ontstaan van de leukemie via het ontwikkelen van genetische afwijkingen en een verhoogde celdeling. In **Hoofdstuk 6** hebben wij onderzocht of er welke afwijkende, onderliggende moleculaire mechanismen mogelijk ten grondslag liggen aan de ontwikkeling van TCRγδ+ T-LGL leukemie. Hiertoe hebben wij zogenaamde genexpressie studies uitgevoerd, waarbij wij een zeer duidelijke overeenkomst tussen de TCRγδ+ T-LGL leukemiecellen en gezonde TCRγδ+ effector en effector geheugencellen zagen. Bij diepgaander onderzoek naar de verschillen tussen de gezonde en leukemiecellen blijken bepaalde genen een afwijkende expressie te vertonen, wat zorgt voor toegenomen celdeling, verminderde celdood en afwijkende afweerreacties. Dit afwijkende patroon van expressie lijkt niet alleen te duiden op mechanismen die bij de ontwikkeling van de leukemie zijn betrokken, maar kan mogelijk ook de relatie met de autoimmuunziekten en cytopenieën in patiënten verklaren.

De veranderde genexpressiepatronen, zoals we die hebben gezien in **Hoofdstuk 6**, kunnen worden veroorzaakt door externe factoren, zoals overmatige prikkeling van cellen en stress. Maar intrinsieke veranderingen in de cellen, zoals genetische afwijkingen, kunnen er ook toe leiden dat genen meer of minder actief kunnen worden. Dankzij recente ontwikkelingen in de sequencing technologie zijn wij in staat het complete genoom van kankercellen in kaart te brengen. Door één van deze nieuwe technologieën, *whole exome sequencing* (WES), toe te passen, hebben wij T-LGL leukemie-specifieke genetische varianten kunnen identificeren (**Hoofdstuk 7**). Na uitgebreide analysestappen hebben wij nu een lijst van zes kandidaatgenen, die een directe correlatie lijken te vertonen met TCRγδ+ T-LGL leukemie. In deze genen zijn kleine toegevoegde of verwijderde stukken DNA te zien, wat leidt tot verschuivingen in de genetische code. Uiteindelijk kunnen de verschuivingen in de kandidaatgenen leiden tot veranderde genexpressie, en daarmee bijdragen aan activatie van afweerreacties en uitzaaiingen. In toekomstige experimenten zal de directe relatie tussen deze genetische afwijkingen en de daadwerkelijke veranderingen in celdeling, resistentie tegen celdood en de toxiciteit tegen andere (bloed)cellen verder vastgesteld moeten worden.

In de **Algemene Discussie** worden alle mogelijke externe en cel-intrinsieke aspecten van TCRγδ+ T-LGL leukemie samengevat en kritisch besproken. Daarnaast presenteren wij nieuwe data over aspecten van veroudering en uitputting van T-LGL leukemiecellen, waar eveneens sleutels kunnen liggen om LGL leukemievorming beter te gaan begrijpen. Ook proberen wij een verbinding te maken tussen LGL leukemie en immuundysregulaties in de vorm van autoimmuunziekten, waarbij ook de vraag van “de kip en het ei” ter sprake komt.

Samenvattend, de studies zoals gepresenteerd in dit proefschrift beschrijven de complexe interacties van genetische afwijkingen, veranderde genexpressieniveaus, functionele eigenschappen, expressie van TCR en andere oppervlaktemoleculen tijdens TCRγδ+ T celontogenie, perifere (antigene) selectie, veroudering en leukemogenese. Het onderzoek in deze thesis vormt niet alleen een basis voor het beter begrijpen van zowel normale als afwijkende TCRγδ+ T cellen, maar draagt mogelijk ook bij aan nieuwe inzichten die uiteindelijk moeten leiden tot betere diagnosestelling, prognose, behandeling en/of preventie. Ten slotte benadrukt dit proefschrift het belang van TCRγδ+ T cellen in de normale fysiologie, maar tevens hun potentie om een leukemie te vormen: "*negeer ze niet, want ze kunnen bijten*".

DANKWOORD

Dr. A.W. Langerak, beste Ton, al sinds het begin van de Master in 2011 hadden we een connectie: CLL. Maar jij had wat anders voor mij in gedachten, wat later ook veel leuker bleek te zijn dan CLL: gammadelta T cellen. Tijdens mijn stage, maar ook tijdens de PhD, heb je mij heel veel vrijheid gegeven. Dit was soms moeilijk, maar uiteindelijk is dit alleen maar goed geweest. Je geduld, adviezen, begeleiding, maar vooral je grote benaderbaarheid heb ik zeer gewaardeerd. Bedankt hiervoor, maar ook voor dat je altijd in mij hebt geloofd en deze kans hebt gegeven.

Prof.dr. J.J.M. van Dongen, beste Jacques, bedankt dat ik destijds mocht beginnen als masterstudent en uiteindelijk het onderzoek zelfs mocht voortzetten als PhD student. Deze afgelopen tijd heb ik als heel prettig ervaren, omdat ik werd vrijgelaten in de ontwikkeling tot onderzoeker. Ik ben dankbaar dat ik deze kans heb gekregen en dat ik een mooi proefschrift heb mogen en kunnen afleveren.

Kleine commissieleden prof.dr. A.M.H. Boots, prof.dr. E.F. Eldering, dr. J.N. Samsom: hartelijk dank voor jullie waardevolle beoordeling. Dankzij jullie is het een compleet boekje. In het bijzonder wil ik Mieke bedanken voor de mooie samenwerkingen. Grote commissieleden, bedankt dat jullie als opponenten wilden deelnemen aan mijn verdediging: het is mij een eer.

Lieve MIDjes, al sinds mijn Master-labrotatie november 2011 kwam ik in een warm nest terecht. Bedankt hiervoor, en voor al het lief en leed wat wij met elkaar hebben gedeeld. Joyce, bedankt voor je master-stage-begeleiding, gezelligheid en steun; door jouw gepush ben ik mooi PhD student bij de MIDjes geworden. Kim (Kimmie), bedankt voor de gezellige uurtjes, zowel op het lab, als in een sushitent of op de fiets. Ik heb enorme bewondering voor je sterkte en doorzettingsvermogen. Alice, bedankt voor je “senior” steun tijdens mijn PhD project. Ook al zijn we weg, het contact blijft hetzelfde. Ellen, bedankt voor je nuchterheid en onbezorgdheid. Tamara (de Zeeuwsche taele is ’t mooist van oallemaele), Irene, Brigitte, Fatemeh, bedankt voor de dikke pret in onze groep. Ruud, je paste precies in de MID-groep. Bedankt voor je zwarte humor en gezelligheid zo vroeg in de ochtend. Michèle, Jorn, mijn (oud-)studenten: wat heb ik het met jullie getroffen. Ik ben trots op jullie. Michèle, toen je je afstudeerpresentatie had gedaan, slaakten we beiden een zucht van opluchting: nu kunnen we samen gaan shoppen. Menig etentje en stedentripje is voorbij gekomen. Bedankt voor alles èn dat je mijn paranimf wilde zijn! Jorn, februari 2017 begon ik de druk van de PhD en de afronding te voelen. Na overleg met Ton, die toevallig een sollicitatie van een student uit Wageningen had gekregen, lieten we je last-minute langskomen. We besloten een gok op je te wagen, en zodra in de eerste paar weken bleek dat je goed was, smeedden we een plannetje om je te houden. Uiteindelijk is dit een hele goede gok geweest, en ga je verder als PhD student.

Bedankt voor je inzet, motivatie, en je vreselijke humor. Ingrid, lieve Ing, wat een vechter ben jij geweest. Bedankt voor je hulp en wijsheid. Ik mis je.

Lieve Kamergenootjes: Diana, Diaantje: jij was er altijd voor mij als beginnende PhD'er. Je had genoeg voor je kiezen gekregen, maar daar kwam je sterker dan voorheen uit. Ik ben blij dat je nu je plekje hebt gevonden. Marjolein, ik zag jou vaker dan wie dan ook, omdat we namelijk ook vaak een (hotel)kamer deelden op congressen. Het was altijd een dolle boel met theetjes, "SB", etc. Heel veel succes met alles! Anne, ik heb nog nooit iemand ontmoet die zowel computer-, programmeer-, statistiek-gek, als humoristisch is. Je bent al een heel eind op weg. Heel veel succes met alles, en blijf lekker wie je bent.

Grote dank aan alle mede-PhD'ers en postdocs, voor de geweldige etentjes en retreats: Jamie, Liza, Fabian (mein verrückter deutscher Freund), Jorn, Britt, Christina, Chris, Jenna, Hanna, Pauline, Malou, Willem Jan, Wouter, Iris, Dew, Prayer, Marieke, Prisca, Wendy, Wida, Marlieke, and all the others that I forgot to mention: thanks!

Tijdens een PhD-project komt steun van allerlei kanten. Daarom wil ik graag de LLD analisten, Mirjam & PID, en alle andere analisten bedanken: voor reagentia, technieken of data kon ik altijd bij jullie terecht. Monique, eventjes lekker bij jou kletsen was zo'n heerlijke afleiding en oppepper, bedankt! François van de Hematologie, bedankt voor je enorme inzet en het meedenken over ons project/experimenten, paper, etc. Ik vond het een fijne en vruchtbare samenwerking! Lieve Gellof, Marie Joan, Sascha, Bibi, Erna en Daniëlle, bedankt voor jullie hulp en gezelligheid. Daniëlle, bedankt voor het ineenzetten van het boekje. Je dacht aan zoveel dingen, waaraan ik nooit had gedacht. Bedankt voor de fijne praatjes, welke ook serieus konden zijn.

Lieve Denise, al sinds 2008 zijn wij vriendinnen. Alles is met jou bespreekbaar, en ik kan altijd bij je terecht: dát is echte vriendschap. Het is fijn om te weten dat je altijd voor mij klaarstaat (en ik uiteraard voor jou!). Ashley, samen met de MIDjes waren we een hecht groepje, ook toen je een Master en PhD ging doen. Bedankt voor de gezellige uurtjes en goede gesprekken, en veel geluk met alles. Ik ben blij dat ik jou heb leren kennen! Astari, Tari, I remember the exact day we've met during the Master (Wed. August 17th, 2011). From that day on we were friends. It's amazing to have so many things in common with a person from a totally different continent, culture and religion. I am happy this never changed. Many thanks for your friendship. Bianca, As, ik had nooit gedacht dat ik zulke vriendinnen zou overhouden aan een bijbaan tijdens mijn opleidingen. Die periode bij jullie op het secretariaat is zeer sterk bij mij blijven hangen. Bedankt voor alle lunchloopjes en ontspanningen, gekkigheid, Rotterdamse uitspraken en levenslessen!

Lieve Wouter, mijn grote broer, voorbeeld en paranimf: op 8 april 2014 mocht ik jouw paranimf zijn bij jouw verdediging "Activity-based probes for retaining β-glucosidases:

novel tools for research and diagnostics". Jij bent mij altijd voor geweest, dus kon ik heel veel van je leren. Wout, fantastisch dat je grant na grant binnensleept en een eigen lab hebt aan Imperial College London. Echt een broer waar ik trots op mag zijn! Marusela, my dear sister-in-law. I really enjoyed your company, our lunch dates when you worked at Erasmus, your support in my career development. Many thanks.

Lieve omie, het is altijd fijn eventjes met je te bellen; een goed moment om te worden herinnerd dat er ook andere dingen belangrijk zijn. Bedankt voor je steun, liefs, en het branden van kaarsen. Lieve opie, opeens was je er niet meer, en ik zie je nog steeds midden in de Schepenenlaan ons uitzwaaien. Je fameuze uitspraken klinken vaak in mijn hoofd. Lieve opie, Wouter en ik hebben altijd zoveel aan je gehad. Bedankt voor alles!

Lieve Regeertjes en Köhlertjes, jullie hebben mij vanaf het eerste moment opgenomen in de familie. Niets is te gek, en het is altijd gezellig bij jullie. Alle weekendjes bij jullie of in Limburg of Nijmegen waren erg welkome ontspanningen. Jullie hebben ons altijd gesteund, bedankt.

Lief, wat rest mij nog te zeggen? Als er iemand was die mij begreep, of mij nog meer begreep dan ikzelf, dan was jij het wel. We hoeven geen gesprek meer te voeren, want óf één blik volstaat, óf we kunnen elkaars antwoorden invullen – inclusief gezichtsuitdrukkingen en reacties. Jouw steun heeft mij tot dingen gebracht, die ik soms in de eerste instantie niet zou durven. We hebben veel lief en leed gedeeld, en er gaat vast nog een hele hoop komen. Maar één ding is zeker: jij zal er zijn.

Lieve pep en mama. Bedankt voor het engelengeduld en de betrokkenheid. Mama, bedankt voor de heerlijke daagjes winkelen, kletsen, lachen, handwerkjes: gewoon even "niks moet". Pep, bedankt voor de HD toertjes, hoofd leegmaken, nieuwe energie opdoen. Jullie eeuwig durende motivatie, medeleven, steun en liefde valt niet onder woorden te brengen: geen enkele uitdrukking volstaat mijn dankbaarheid. Jullie zijn de beste.

CURRICULUM VITAE

Personal information

Surname: Kallemeijn
Name: Martine Joanna
Date of birth: 22 December 1989
Place of birth: Vlissingen, the Netherlands
Nationality: Dutch

Current Position

2018 **Technical Application Specialist Life Sciences**
Roche Diagnostics Deutschland GmbH
Molecular Solutions, EMEA Customer Support Center Mannheim

Education

2013 – 2017 **PhD study in Molecular Immunology, Laboratory for Medical Immunology, Department of Immunology, Erasmus MC, Rotterdam, the Netherlands**
Normal and aberrant TCRγδ+ T cells and T cell large granular lymphocyte leukemia
Promotor: Prof.dr. J.J.M. van Dongen
Copromotor: Dr. A.W. Langerak

2011 – 2013 **Master of Science in Infection and Immunity**
MolMed Postgraduate School, Erasmus University Erasmus MC, Rotterdam, the Netherlands

2008 – 2011 **Bachelor of Science in Biomedical Sciences**
Amsterdam Medical Center – University of Amsterdam, the Netherlands

2002 – 2008 **Gymnasium**
Pre-university education, Nature & Health, Christelijk Scholengemeenschap Walcheren, Middelburg, the Netherlands

Research experience

2012 – 2013 **Second Master Internship**
Aberrant TCRγδ+ T cells in T-LGL lymphoproliferations originate from the normal TCRγδ+ T cell population in adult peripheral blood
Department of Immunology, Erasmus MC, Rotterdam
Supervisors: Dr. A.W. Langerak and Drs. J. Schilperoord – Vermeulen

2011 – 2012	**First Master Internship** *Normal versus aberrant TCRγδ+ T cells: what makes the difference?* Department of Immunology, Erasmus MC, Rotterdam Supervisors: Dr. A.W. Langerak and Drs. J. Schilperoord – Vermeulen
2011	**Bachelor Graduation Internship** *The effect of IL-21 on B cell chronic lymphocytic leukemia* Department of Experimental Immunology, AMC-UvA, Amsterdam Supervisors: Prof.dr. E.F. Eldering and Dr. M.F. Pascutti
2010	**Literature Study Physiology** *The effect of zero gravity on bone resorption in astronauts* Department of Physiology, AMC-UvA, Amsterdam Supervisor: Prof.dr. J.H. Ravesloot

Personal Awards

2012/2013	**Royal Dutch Academy of Sciences (KNAW)** Academy Assistantship for promising research talent in Immunology + Travel Grant
Various dates	**Dutch Society of Immunology (NVVI) Travel Grants** Joint BSI/NVVI Congress, Liverpool, UK International Gammadelta T cell Congress, London, UK European Congress of Immunology, Vienna, Austria

PHD PORTFOLIO

PhD student:	Martine J. Kallemeijn
Department:	Laboratory for Medical Immunology, department of Immunology
PhD period:	September 2013 – December 2017
Promotor:	Prof.dr. J.J.M. van Dongen
Copromotor:	Dr. A.W. Langerak
Research school:	MolMed Postgraduate School

PhD Courses

2017	Laboratory Animal Science
2016	Research Integrity
2016	BioBusiness Summerschool
2016	Research Management
2016	GEP & R
2015	SNP course
2015	R
2015	CLC Genomic Workbench and Ingenuity Variant Analysis
2015	Ensembl
2014	Advanced Immunology
2014	Galaxy
2014	BD Flow Cytometry (basic and advanced)
2014	Basic and advanced Excel
2013	Radiation Hygiene Levels 5A/5B
2012	Molecular Diagnostics
2012	Adobe Photoshop, Illustrator & InDesign
2012	Molecular Medicine
2012	Microscopic Imaging Analysis Techniques
2011 – 2012	Statistics & Survival Analysis
2011	NIHES Study Design
2011	Biomedical Research Techniques
2011	Human Genetics
2011	SPSS

(Inter)national Conferences

2016	Joint BSI/NVVI Congress, Liverpool, UK: poster presentation
2016	Gammadelta T cell Congress, London, UK: poster presentation
2015	NVVI Winter Seminar, Noordwijkerhout, the Netherlands: poster presentation

2015	European Congress of Immunology, Vienna, Austria: poster presentation
2015	Dutch Society of Hematology Congress: oral presentation in special in-focus session
2014	NVVI Winter Seminar, Society's 50th Anniversary Congress, Efteling, Kaatsheuvel, the Netherlands: oral presentation
2013	NVVI Winter Seminar, Noordwijkerhout, the Netherlands: oral presentation

Seminars and Workshops

2013 – 2017	Annual MolMed days
2013 – 2017	Medical Immunology, Rotterdam

Other

2013 – 2017	Journal Club department of Immunology
2013 – 2017	Seminars and minisymposia department of Immunology and department of Hematology
2015 – 2017	Laboratory Medical Immunology (LMI) monthly lab meetings

Teaching

2014 – 2017	Histology practical classes for first and second year medical students
2014 – 2015	Supervision of HLO Bachelor thesis, Hogeschool Rotterdam
2015 – 2016	Supervision of HLO Bachelor thesis, Hogeschool Leiden
2017	Supervision of Master thesis, Wageningen University

Extracurricular activities, committees

2014 – 2017	SMBWO certificate (exam grade 8)
2015 – 2017	Biosafety committee department of Immunology
2015 – 2016	Party committee department of Immunology
2015 – 2016	NTI Course Deutsch als Fremdsprache
2014 – 2015	PhD committee department of Immunology

Membership

2013 – 2017	NVVI member
2011 – 2017	Member of MolMed Postgraduate School

PUBLICATIONS

Accepted

Kallemeijn, M.J., Kavelaars, F.G., van der Klift, M.Y., Wolvers – Tettero, I.L.M., Valk, P.J.M., van Dongen, J.J.M. & Langerak, A.W. Next generation sequencing analysis of the human TCRγδ+ T-cell repertoire upon aging. *Front. Immunol.* **9**, 448 (2018).

Kallemeijn, M.J., de Ridder, D., Schilperoord – Vermeulen, J., van der Klift, M.Y., Sandberg, Y., van Dongen, J.J.M. & Langerak, A.W. Dysregulated signaling, proliferation and apoptosis impact on the pathogenesis of TCRγδ+ T cell large granular lymphocyte leukemia. *PLoS One* **12(4)**, e0175670 (2017).

Kallemeijn, M.J., Boots, A.M.H., van der Klift, M.Y., Brouwer, E., Abdulahad, W.H., Verhaar, J.A.N., van Dongen, J.J.M. & Langerak, A.W. Ageing and latent CMV infection impact on maturation, differentiation and exhaustion profiles of T-cell receptor gammmadelta T-cells. *Sci. Rep.* **7**, 5509 (2017).

Sandberg, Y., **Kallemeijn, M.J.**, Dik, W.A., Tielemans, D., Wolvers – Tettero, I.L.M., van Gastel – Mol, E.J., Szczepanski, T., Pol, Y., Darzentas, N., van Dongen, J.J.M. & Langerak, A.W. Lack of common TCRA and TCRB clonotypes in CD8+/TCRαβ+ T-cell large granular lymphocyte leukemia: a review on the role of antigenic selection in the immunopathogenesis of CD8+ T-LGL. *Blood Cancer J.* **4**, e172 (2014).

Manuscripts in preparation

Kallemeijn, M.J., van Rooij, J.G.J., Kraaij, R., van Dongen, J.J.M. & Langerak, A.W. Whole exome sequencing reveals novel insights in the pathogenesis of TCRγδ+ T cell large granular lymphocyte leukemia: altered transcriptional regulation, immune cell activity and invasive growth. *Manuscript in preparation.*

Kallemeijn, M.J., Assmann, J.L.J. & Langerak, A.W. Exhaustion and senescence in CD8+TCRαβ+ and TCRγδ+ T-LGL leukemia cells. *Manuscript in preparation.*

...et Emergo.