Regulation of T cell activation by two conserved phosphotyrosines in Lck and its adapter protein TSAd

Doctoral thesis

by

Paweł Borowicz

Department of Molecular Medicine
Institute of Basic Medical Sciences
Faculty of Medicine
University of Oslo

October 2019

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-575-4

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgment

To whomever will be reading this thesis:

Thank you!

I hope that you enjoy it!

To all the people who helped me during my PhD struggles:

Thank you!

(You know best for what exactly!)

“Everything is connected. Nothing is also connected.”

AND

“I learned a long time ago that things don't always make sense the way people want them to. I accept that, and you're going to have to, too.”

-Dirk Gently

List of papers included

Paper I

A simple and efficient workflow for generation of knock-in mutations in Jurkat T cells using CRISPR/Cas9. P. Borowicz*, H. Chan*, D. Medina, S. Gumpelmair, H. Kjelstrup, A. Spurkland. [Submitted]

*Equally contributed.

Paper II

Tyr^{192} is a critical regulator of lymphocyte-specific tyrosine kinase activity in T cells. P. Borowicz, V. Sundvold-Gjerstad, H. Chan, G. Abrahamsen, H. Kjelstrup, T. A. Nyman, A. Spurkland. [Submitted]

Paper III

A conserved pTyr motif in the unstructured C-terminus regulates the function of the T cell specific adaptor TSAd. P. Borowicz, R. P. Gopalakrishan, H. Chan, H. Kjelstrup, S. Foss, G. A. de Souza, V. Sundvold-Gjerstad, A. Spurkland. [Manuscript]

Table of Contents

Nomenclature and abbreviations list

A – Alanine
AIDS – Acquired immune deficiency syndrome
AIRE – Autoimmune regulator
AKT – Protein kinase B
Ala – Alanine
ALX – Hematopoietic SH2 domain-containing protein
AP-1 – Activator protein 1
APC – Antigen-presenting cell
Arg – Arginine
Asn – Asparagine
Asp – Aspartic acid
AS-PCR – Allele-specific PCR
BAX – Apoptosis regulator BAX
BCR – B-cell receptor
BG – Loop between B-α helix and G-β strand
Blk – B lymphocyte kinase
C1GALTI – Core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1
cAMP – Cyclic adenosine monophosphate
CAR - Chimeric antigen receptor
Cas – CRISPR-associated protein
Cas9n – Cas9 nickase
CCL5 – C-C motif chemokine ligand 5
CD – Cluster of differentiation
cDNA – Complementary DNA
CMC – Critical micelle concentration
CRISPR – Clustered regularly interspaced short palindromic repeats
crRNA – CRISPR RNA
Csk – C-terminal Src kinase
CTLA4 – Cytotoxic T-lymphocyte-associated protein 4
CXCL12 – C-X-C motif chemokine ligand 12
D – Aspartic acid
DNA – Deoxyribonucleic acid
DSCAM – Down syndrome cell adhesion molecule
DSCAML1 – Down syndrome cell adhesion molecule like 1
E – Glutamic acid
EF – Loop between E- and F-β strands
EGFR – Epidermal growth factor receptor
ERK – Extracellular signal–regulated kinase
eSpCas9 – Enhanced specificity *Streptococcus pyogenes* Cas9
F – Phenylalanine
Fgr – Proto-oncogene tyrosine-protein kinase Fgr
FRET – Förster resonance energy transfer
Fyn – Proto-oncogene tyrosine-protein kinase Fyn
Gads – GRB2-related adaptor downstream of Shc
GFP – Green fluorescent protein
Glu – Glutamic acid
GRB2 – Growth factor receptor-bound protein 2
gRNA – Guide RNA
GTPase – Guanosine-5'-triphosphate hydrolase
Hck – Hemopoietic cell kinase
HDR – Homology directed repair
HIV – Human immunodeficiency virus
ICOS – Inducible T-cell Costimulator
IFN-γ – Interferon gamma
IL – Interleukin
IO – Ionomycin
IS – Immunological synapse
ITAM – Immunoreceptor tyrosine-based activation motif
Itk – Interleukin-2-inducible T-cell kinase
JNK – c-Jun N-terminal kinase
JTAg – Jurkat TAg
KO – Knockout
LAD – Lck adaptor protein
LAT – Linker for activation of T cell
Lck – Lymphocyte-specific protein tyrosine kinase
LDS – Lithium dodecyl sulfate
LFA-1 – Lymphocyte function-associated antigen 1
LIME – Lck-interacting transmembrane adapter 1
Lnk – Lymphocyte Adapter Protein
Lyn – Tyrosine-protein kinase Lyn
Lys – Lysine
MAP3K2 – Mitogen-activated protein kinase kinase kinase 2

MAPK – Mitogen-activated protein kinase
MAPK7 – Mitogen-activated protein kinase 7
MHC – Major histocompatibility complex
mRNA – Messenger ribonucleic acid
MS – Mass spectrometry
MSH2 – DNA mismatch repair protein Msh2
mTECs – Medullary thymic epithelial cells
NCK1 – Non-catalytic region of tyrosine kinase adaptor protein 1
NFAT – Nuclear factor of activated T-cells
NF-κB – Nuclear factor kappa-light-chain-enhancer of activated B cells
NHEJ – Non-homologous end joining
NK – Natural killer
N-WASP – Neural Wiskott-Aldrich syndrome protein
PAM – Protospacer adjacent motif
PCR – Polymerase chain reaction
PDGFR – Platelet-derived growth factor receptor
Phe – Phenylalanine
PI3K – Phosphatidylinositol 3-kinase
PKA – Protein kinase A
PKC – Protein kinase C
PLC-γ – Phospholipase C, gamma 1
PMA – Phorbol myristate acetate
PP2 – 4-Amino-5-(4-chlorophenyl)-7-(dimethylethyl)pyrazolo[3,4-d]pyrimidine (Src kinase inhibitor)
pSer – Phosphoserine
PTB – Phosphotyrosine-binding domain
PTEN – Phosphatase and tensin homologue
pThr – Phosphothreonine
PTPN22 – Protein Tyrosine Phosphatase Non-Receptor Type 22
pTyr – Phosphotyrosine
Pyk2 – Protein tyrosine kinase 2 beta
Rho – Ras homolog
RIBP – Rlk/Itk-binding protein
RPSA – 40S ribosomal protein SA
S – Serine
Ser – Serine
SH2 – Src homology 2
SH2D2A – SH2 Domain-Containing Protein 2A
SH3 – Src homology 3
SHC – SH2 domain-containing transforming protein
SHIP – SH2-domain-containing inositol polyphosphate 5' phosphatase
SHP1 – SH2 domain-containing phosphatase-1
SHP2 – SH2 domain-containing phosphatase-2
siRNA – Small interfering RNA
SLP76 – SH2 domain containing leukocyte protein of 76kDa
SNP – Single nucleotide polymorphism
Sos1 – Son of sevenless homolog 1
Src – Proto-oncogene tyrosine-protein kinase Src
ssODN – Single-stranded donor oligonucleotide
SV40 – Simian virus 40
Syk – Spleen tyrosine kinase
T – Threonine
TALEN –Transcription activator-like effector-based nucleases
TCR – T-cell receptor
Tec – Tyrosine-protein kinase Tec
TGFBR – TGF-β receptor
TGF-β – Transforming growth factor beta
Thr – Threonine
TNF-β – Tumor necrosis factor-beta
TP53 – Tumor protein p53
TSAd – T cell specific adaptor protein
Tyr – Tyrosine
VAV1 – Vav Guanine Nucleotide Exchange Factor 1
VE-cadherin – Vascular endothelial cadherin
VEGF – Vascular endothelial cell growth factor
VEGFR2 – VEGF receptor 2
VRAP – VEGF receptor-associated protein
WT – Wild type
Y – Tyrosine
Yes – Proto-oncogene tyrosine-protein kinase Yes
ZAP70 – Zeta-chain-associated protein kinase 70
ZFN – Zinc finger nucleases
Δex7 – deletion of exon 7

Introduction

The adaptive immune system

The immune system cells can detect most of the dangerous or abnormal situations that put the organism at risk. Our common enemies are small, numerous and constantly adjusting to new situations. To overcome attacks from bacteria, fungi, viruses and parasites, our immune system had to develop a method to equally quickly adapt to new invaders. Therefore, evolution equipped us with an adaptive immune system (Flajnik and Kasahara, 2010).

The two components which are the most important parts of the adaptive (acquired) immune system are: antibodies produced by B cells (Fagraeus, 1948) and TCR presented on the surface of T cells (Zinkernagel and Doherty, 1975). B cells, T cells and NK cells are named lymphocytes, as they are the main cells present in the lymph, allowing them to perform a secondary lymphoid organs surveillance. NK cells are considered to be a part of innate immune system, as they receptors cannot acquire specificity toward targets. In contrast, both B cells and T cells can recognize other molecules with very high specificity. Due to that, they are tightly controlled on both the molecular and the cellular level.

The first step in acquiring adaptive immunity is the scavenging mechanism of antigen-presenting cells (APC). The professional APC, which include dendritic cells, are continuously engulfing cellular and molecular debris throughout the body. Whenever they receive a “danger” signal (Matzinger, 1994), they start migrating towards the closest secondary lymphoid organ. In the lymphoid organ, they present the scavenged molecules (antigens) to the T cells (Steinman, 1991). If a T cell recognizes the specific antigen presented by the dendritic cell, it becomes activated and starts to proliferate. There are two major types of T cells:

cytotoxic $CD8^+$ T cells and helper $CD4^+$ T cells, so-named with respect to their main cellular function and surface expression of the TCR co-receptors. A smaller subset of T cells, however crucial in regulating adaptive immune responses, are the regulatory T cells.

T cells in the adaptive immune system

After activation, cytotoxic T cells migrate towards the "danger" site where the "danger" signal originated and survey local cells. Most cells present fragments of their proteins on the cell surface in the forms of peptides bound to major histocompatibility complex (MHC) class I molecules. If a $CD8^+$ T cell recognizes any of those fragments as their specific antigen, it is most likely a foreign or abnormal protein. For the $CD8^+$ T cell this is a signal to kill the targeted cell through cytotoxic killing. In this way T cells can recognize malfunctioning cells, mutated cancer cells or cells infected with bacteria and viruses (Cerottini et al., 1970).

Meanwhile, the helper $CD4^+$ T cells perform a completely different function (**Fig. 1**). Following activation, they stay in the secondary lymphoid organs producing cytokines and seeking B cells to interact with. B cells also reside in the secondary lymphoid organs, where they have access to constant flow of antigens drained from the tissue. The B-cell receptor (BCR) is similar to the TCR, whereby its specificity to antigen is very high and can recognize only a small number of molecules in their native form. As soon as the BCR binds an antigen, it is internalized and the cargo is digested. The peptides created by the digestion are loaded onto MHC class II molecules and presented on the surface of the B cell (Lanzavecchia, 1985). These complexes are scanned by the activated $CD4^+$ T cells. Whenever the $CD4^+$ T cell recognizes a peptide presented by the MHC class II molecule on the B cell, it starts a process called T-cell dependent activation of B cells. The purpose of this

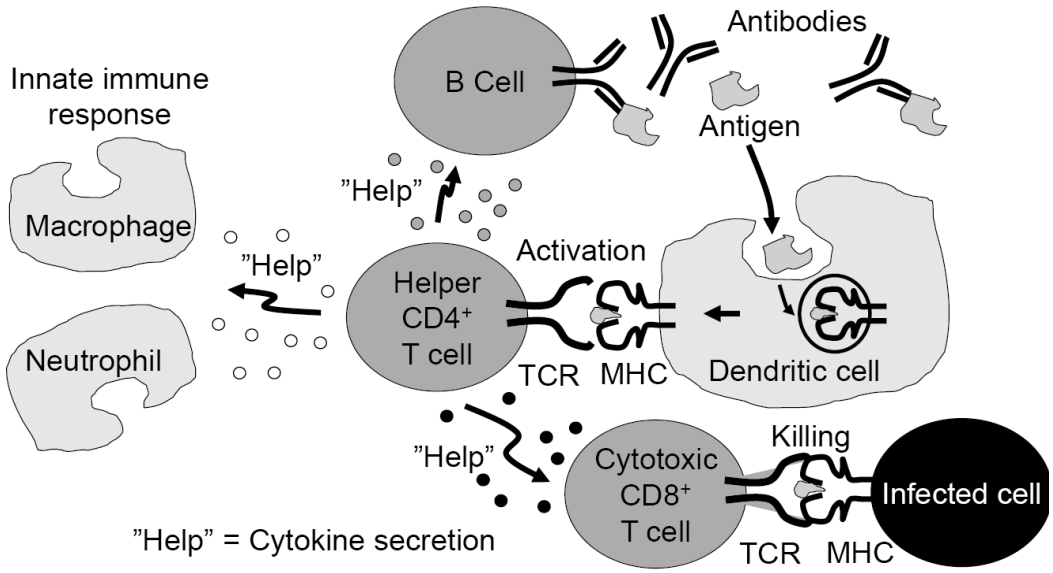


Figure 1. Helper CD4⁺ T cell functions. The main function of helper CD4⁺ T cells is to shape the immune response by TCR-dependent cytokine secretion. Cytokines serve as "help" signals and act as a secondary activation signal for nearly all cells involved in the immune response. For instance, these cytokines can stimulate B cells to produce high affinity antibodies, activate both cytotoxic CD8⁺ T cells as well as innate immune cells.

process is to induce B cells to produce high affinity antibodies and to create long-lasting memory B cells (Crotty, 2015).

The fact that the whole process of producing highly specific antibodies is so complex is beneficial. Even a small mistake in the production of antibodies can lead to autoimmune diseases. This is a state when the immune system deploys abnormal inflammatory responses to healthy parts of the body as a result of mistaking healthy tissue for a threat. The origin of the autoimmune response can vary and usually is unknown. Once it reaches the adaptive immune system response, it becomes chronic and can lead to degradation of body tissues (e.g. thyroid gland in Hashimoto thyroiditis, joints in rheumatoid arthritis, beta cells in pancreas in type 1 diabetes, etc.). Autoreactive lymphocytes are produced all the time, but most of them are removed during development and maturation, or they become anergic due to improper stimulation. Those that escape this regulation, are controlled by a special

subset of T cells, namely regulatory T cells. Regulatory T cells have been linked to the prevention of autoimmune disease development (Sakaguchi et al., 2008). As T cells are responsible for the control over both themselves and B cells, T cell development has to be tightly controlled to prevent the possibility of developing an autoimmune disease. This is termed tolerance.

During maturation of T cells in the thymus central tolerance is established. Firstly, T cells are tested for producing a functional TCR that can interact with MHC class I ($CD8^+$ T cells) or MHC class II ($CD4^+$ T cells) molecules (Klein et al., 2009). Only the T cells whose TCR interacts with MHC molecules are allowed to survive. Secondly, an important step of T cell maturation is the negative selection, when their binding to self-antigens is assessed. As the whole process of maturation is taking place in the thymus, there is a great need for the presence of peptides specific for other tissues. To cope with that problem, the thymus contains medullary thymic epithelial cells (mTECs) equipped with a unique protein called AIRE. AIRE is a transcription factor that allows mTECs' expression of tissue-specific proteins (for example insulin – physiologically produced only by beta cells of the pancreatic isles) (Anderson et al., 2002). All the T cells that interact with self-peptides in the context of MHC class I or II, while being in the thymus medulla, are killed via apoptosis. Thereby, all mature naïve T cells exiting the thymus should be self-tolerant.

The best examples showing the importance of T cells (helper $CD4^+$ T cells in particular) are HIV-infected patients with AIDS. HIV specifically targets $CD4^+$ T cells (Dalgleish et al., 1984), which eventually leads to their depletion. The functional insufficiency of helper $CD4^+$ T cells causes increased susceptibility to cancer development and viral infections, as they are necessary for proper activation of $CD8^+$ T cells. Moreover, people with AIDS are unable to generate new memory B cells or new class-switched antibodies, as helper T cells mediate proper B cell

activation (Shen and Tomaras, 2011). $CD4^+$ T cells orchestrate the whole acquired immune system, thus their depletion results in severe immunodeficiency (AIDS).

T cell receptor

T cell activation depends on more factors than just the TCR affinity alone. First, TCR recognition is MHC restricted (Zinkernagel and Doherty, 1974). This means that a given T cell can recognize an antigen only when presented by a specific MHC variant. If the antigen was presented by another MHC variant, it does not activate that T cell. Secondly, T cells need a co-stimulatory signal from other surface molecules (for example: CD28, ICOS, CD40L, LFA-1) which interact with corresponding ligands/receptors present on the APC (Crotty, 2015). Without co-stimulation, T cells will turn anergic (another mechanism preventing autoimmune responses).

The TCR is built of two different protein chains, which have only short intracellular domains (Wilson and Garcia, 1997). To transduce the extracellular signal into the cell, the TCR forms a complex with four CD3 chains: γ, δ, ε and ζ. CD3 chains contain one or three intracellular motifs called immunoreceptor tyrosine-based activation motifs (ITAM), which can initiate the intracellular signaling cascade. Additionally, a co-receptor specific for MHC stabilizes its binding with TCR. These co-receptors are CD4 and CD8 which bind MHC class II and MHC class I respectively (Klein et al., 2009).

Although many models have been proposed on how the MHC binding initiates the TCR signaling, it is still debated (Chakraborty and Weiss, 2014). One of the hypothesis suggests that the event responsible for initiating the TCR signaling cascade is the release of CD3 chains' ITAMs from the membrane (Xu et al., 2008). Moreover, it seems that this event is more of a mechanical than biochemical nature (Ma and Finkel, 2010). It has been suggested that the pulling force between TCR and peptide

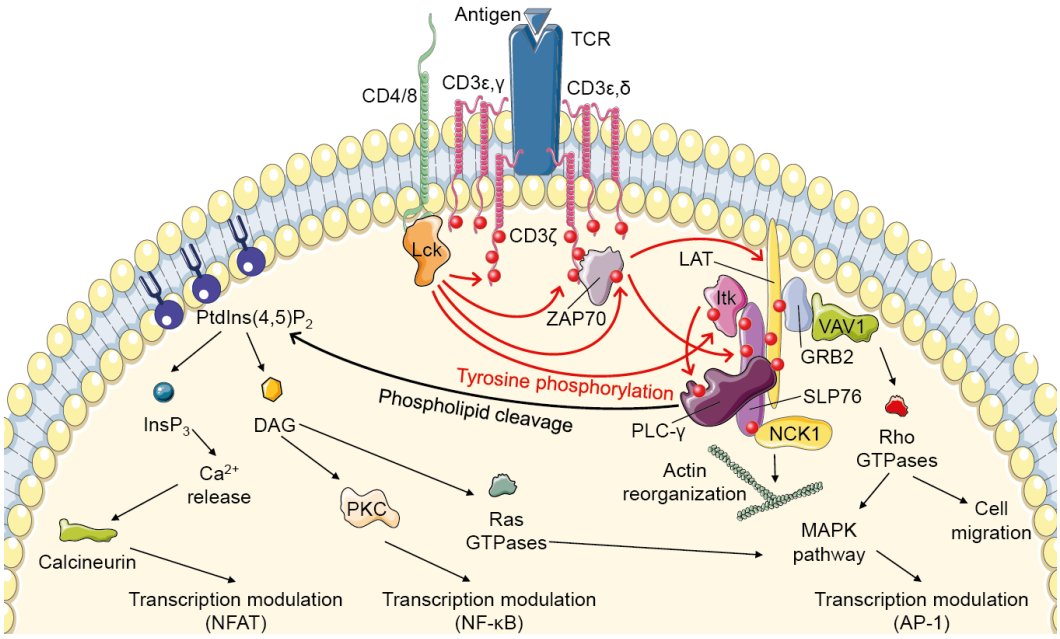


Figure 2. T cell receptor signaling pathway. Upon recognizing an antigen presented by an APC, the T cell receptor releases intracellular chains of co-associated CD3 molecules. The chains can be phosphorylated by Lck and then, serve as a binding platform for ZAP70. These events initiate a phosphorylation cascade, which leads to T cell activation and modulation of certain genes' transcription. See text for more details. $PtdIns(4,5)P_2$ – phosphatidylinositol 4,5-bisphosphate, $InsP_3$ – inositol 1,4,5-trisphosphate, DAG – diacylglycerol.

loaded MHC causes the release of the intracellular domains of the CD3 chains. Exposing ITAMs intracellularly results in their phosphorylation by lymphocyte-specific protein tyrosine kinase (Lck) (Voronova et al., 1984), which initiates the intracellular T cell activation pathway. The presence of Lck in the vicinity of the TCR is guaranteed by its interaction with intracellular part of CD4/8 co-receptors (Veillette et al., 1988) and is stabilized by Lck interaction with lipids in the cell membrane (Yurchak and Sefton, 1995, Yasuda et al., 2000).

T cell receptor signaling pathway

As soon as CD3 ITAMs become exposed in the cytoplasm, Lck may phosphorylate them (**Fig. 2**). Phosphorylated ITAMs serve as a binding site for protein kinase ZAP70 (Wange et al., 1993). ZAP70 recruited to the TCR can be activated by Lck through phosphorylation (Di Bartolo et

al., 1999). In the membrane, ZAP70 propagates the signaling by phosphorylation of LAT (Wang et al., 2010). LAT is an adaptor protein (Weber et al., 1998, Zhang et al., 1998), which operates as a major signaling hub (Sommers et al., 2004). Phosphorylation of LAT recruits other adaptor molecules (GRB2, SLP76 and NCK1), as well as enzymes (Itk, VAV1 and PLC-γ). LAT's main function is to diversify the output from the TCR, which leads to T cell activation at multiple levels.

VAV1 is a guanine nucleotide exchange factor, which translates phosphorylation signaling into GTPase signaling (Barreira et al., 2018). The Rho-family of small GTPases initiate cytoskeleton reorganization, necessary for cell migration, adhesion or division, but it can also activate transcription factor AP-1 through the MAPK pathway. Itk is another kinase in the TCR signaling pathway, which performs phosphorylation of specific phosphotyrosines on LAT (Perez-Villar et al., 2002), SLP76 (Sela et al., 2011), Lck (Granum et al., 2014), and PLC-γ (Andreotti et al., 2010). PLC-γ, activated by Itk, catalyzes generation of inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. These molecules trigger the intracellular release of calcium ions and activate the Ras-family of small GTPases and protein kinase C (PKC). Increased level of intracellular calcium activates the calcineurin pathway, which governs NFAT nuclear import. NFAT is a transcription factor responsible for, among others, IL-2 transcription activation (Shaw et al., 1988). The Ras family of small GTPases regulates, through the MAPK/ERK pathway, the transcription factor AP-1 (Mechta et al., 1997). PKC is responsible for initiation of the pathway that leads to the activation of the NF-κB transcription factor (Ghosh and Baltimore, 1990). All the transcription factors mentioned above are responsible for induction of genes involved in T cells proliferation, differentiation, migration, or cytokine production.

T cell activation

The purpose of multilevel control over the incoming signal received via the TCR is to prevent unnecessary activation of T cells. As mentioned before, T cells with a low threshold for activation can cause autoimmune diseases. The first threshold that the T cell has to cross is to collectively initiate signaling from a certain number of TCRs (Pageon et al., 2016). However, some studies show that this requirement could be overcome by a strong affinity of the TCR to the antigen that would lead to prolonged signaling (Mayya and Dustin, 2016). Therefore, the first threshold can be overcome by either high affinity binding to the antigen or high number of antigens. The second threshold is a consequence of the first one. To sustain the signaling, the T cell forms an immunological synapse (IS) with the APC, the immediate contact area between the T cell and APC. In the IS, various molecules on the T cell surface become organized into separate clusters (Kaizuka et al., 2007). As these molecules are also enzymes participating in TCR signaling, the physical separation on the surface will extend into the intracellular space and modulate signaling (possibly in the form of phase separation (Ditlev et al., 2019, Su et al., 2016)). However, even the prolonged signaling sustained by the IS can be insufficient, if certain molecules in the signaling pathway are blocked or inhibited. Therefore, the second threshold depends on the molecular status of the T cell, which can be changed by external (co-stimulation with receptors, interleukins, cytokines, drugs (Chen and Flies, 2013)) or internal (gene expression, protein regulation (Wirth et al., 2010)) stimuli. The third threshold is again a consequence of the previous. As an end result of the TCR signaling cascade, transcription factors are activated and translocated to the nucleus. Expression of some genes (for example CD69 or CD25) correlates with the amount of specific transcription factor translocated to the nucleus (graded response) (Fuhrmann et al., 2016). However, there are also genes (for example IL-2 and IFN-γ) with a binary response

to the transcription factors – either it happens or not (Fiering et al., 1990). Therefore, the third threshold is the final outcome of the previous steps, which is the amount of transcription factors translocated to the nucleus. The final outcome is the result of the interplay between various gene activators and inhibitors, again modified by different signaling pathways.

One overall purpose of T-cell activation is to upregulate gene expression profiles necessary for T cell effector functions. Activated T cells may produce many interleukins and cytokines, such as: IL-2, IL-4, IL-5, IL-6, IL-9, IL-10, IL-13, IL-17, IFN-γ, and TNF-β (Zhu and Paul, 2010). Moreover, activated T cells can be characterized by expression of the surface proteins including: CD25, CD44, CD69, CD45RO, CD152, and CD95L (Shipkova and Wieland, 2012). There are also intracellular proteins whose expression is induced upon TCR stimulation, such as kinases (Itk (Miller et al., 2004), Tec (Tomlinson et al., 2004), JNK (Weiss et al., 2000)) or adaptor molecules (TSAd (Spurkland et al., 1998), LAT (Peters et al., 2000)). However, the impact of altered expression of the intracellular proteins on TCR signaling of activated T cells is still not well understood.

Lck –Lymphocyte-specific protein tyrosine kinase

Lck belongs to the family of Src tyrosine kinases. Therefore, Lck's structure and regulation are similar to those of Src, Fyn, Yes, Lyn, Hck and Blk. The conserved structure of the Src family includes an N-terminal end (which can be both myristoylated and palmitoylated), an SH3 domain, an SH2 domain, a linker segment, a tyrosine kinase domain, and a C-terminal tail (which can be phosphorylated). Myristoylation (Yasuda et al., 2000) and palmitoylation (Yurchak and Sefton, 1995) of the N-terminal end secure binding to lipid membranes. The SH3 domain binds to proline-rich motifs (Erpel et al., 1995). The SH2 domain can recognize and bind to specific phosphotyrosine motifs

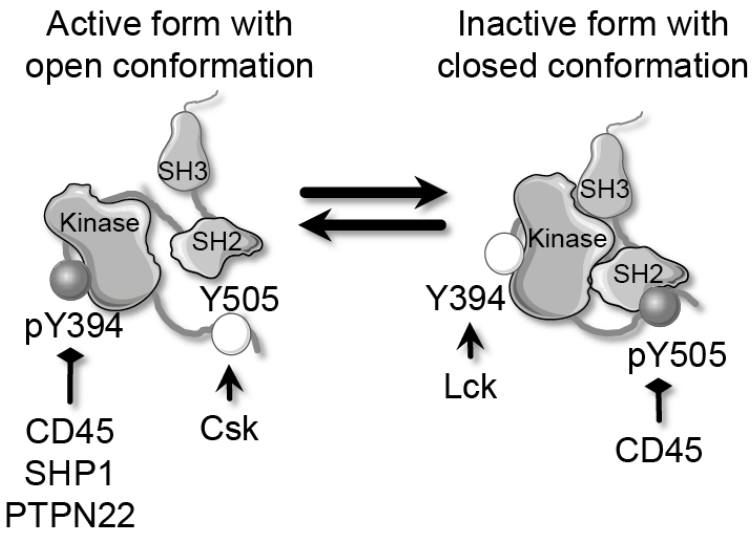


Figure 3. Regulation of Lck activity. Lck is regulated by two phosphotyrosines (Tyr^{394} and Tyr^{505}), which phosphorylation dictates Lck conformation. A number of kinases and phosphatases can affect these phosphosites. Blunt arrows represent dephosphorylation, sharp arrows represent phosphorylation.

(Zhou et al., 1993). Both SH3- and SH2-domains participate in targeting the kinase substrates. The disordered structure of the linker segment facilitates conformational changes of the kinase (Gonfloni et al., 1997). The tyrosine kinase domain has a specificity unique for each Src-family kinase, to assure coherence of the signaling pathway, in which the kinase participates (Shah et al., 2018). The C-terminal tail contains one of the most important regulatory sites of the Src-family kinases – a tyrosine (Tyr) – which in Lck corresponds to Tyr^{505} (Yamaguchi and Hendrickson, 1996).

Lck is a protein specifically expressed in lymphocytes, particularly in T cells and NK cells. Removing Lck from T cells essentially abrogates TCR signaling (Straus and Weiss, 1992). Lck knockout (KO) mice have a significant, but incomplete block of T cell development (Molina et al., 1992). If Lck KO is combined with Fyn KO, there is a complete block of T cell development.

Lck regulation and function is intrinsically linked with other proteins specific for T cells (**Fig. 3**). The canonical model of Lck regulation involves two phosphotyrosines (pTyr), pTyr^{394} located in the activation loop and pTyr^{505} located in the C-terminal tail. Phosphorylated Tyr^{505} binds to the Lck SH2-domain, which forces Lck to adopt a closed conformation (Boggon and Eck, 2004). However, dephosphorylation of Tyr^{505} is insufficient for Lck to become catalytically active (Philipsen et al., 2017). Phosphorylation of Tyr^{394} is required for stabilization of the open conformation of Lck (Boggon and Eck, 2004, Yamaguchi and Hendrickson, 1996, Philipsen et al., 2017).

Two main enzymes govern Lck phosphorylation: tyrosine protein kinase Csk and protein tyrosine phosphatase CD45 (Palacios and Weiss, 2004). Csk is responsible for phosphorylation of Tyr^{505}, while CD45 dephosphorylates this site. Tyr^{394} is phosphorylated through transphosphorylation, by another Lck molecule (Yamaguchi and Hendrickson, 1996). A number of phosphatases have been held responsible for dephosphorylation of Lck Tyr^{394}: SHP1 (Chiang and Sefton, 2001), PTPN22 (Hasegawa et al., 2004, Wu et al., 2006) and CD45 (McNeill et al., 2007). Although Lck pTyr^{394}/Tyr^{505} is considered to be the main active form and Lck Tyr^{394}/pTyr^{505} is considered to be the main inactive form of Lck, the existence of other variants is not excluded (Philipsen et al., 2017). Moreover, there is no consensus in the field, whether Lck pTyr^{394}/Tyr^{505} exists prior to TCR stimulation (Nika et al., 2010), whether it is consequently generated (Burkhardt et al., 1994), or whether it is subsequently increased (Ballek et al., 2015, Stirnweiss et al., 2013).

Lck SH2 domain

The Lck SH2 domain is part of a fundamental mechanism involved in all phosphotyrosine-based signaling pathways (Lim and Pawson, 2010). The SH2 domains have highly conserved structures which consist of a

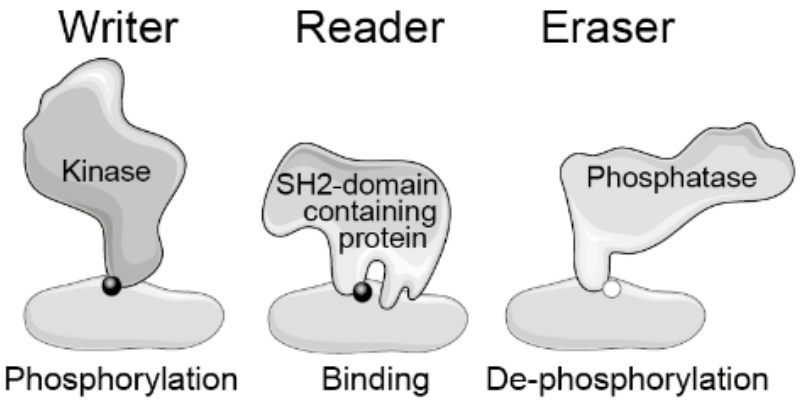


Figure 4. The role of SH2 domain in phosphotyrosine signaling-based pathways. Phosphotyrosine signaling is based on kinases' ability to phosphorylate certain tyrosines ("writing"). These phosphotyrosines can be recognized by an SH2-domain containing protein ("reading"), which can perform its own specific functions or serve as a binding platform for other proteins. The signaling is disrupted by phosphatases which dephosphorylate the tyrosines ("erasing"). Phosphorylated tyrosines are depicted by a filled circle and desphosphorylated tyrosines are depicted with an open circle.

hydrophobic β-sheet flanked by two α-helices (Eck et al., 1994). Additionally, two flexible loops (EF and BG), which are flanking the binding groove of the SH2 domain, define the domain's specificity (Liu et al., 2019, Kaneko et al., 2010).

While kinases are "writers", responsible for signal propagation, and phosphatases are "erasers", which dispose of the unnecessary signal, SH2 domains are referred to as "readers" (**Fig. 4**). These domains evolved to recognize pTyr motifs with high specificity. Their role in phosphotyrosine signaling cascades is to act as direct "readers" of phosphorylation events. There are three main functions of SH2 domains present in kinases:

- Control of the molecule's conformation (for example: the Lck SH2 domain binding to Lck $pTyr^{505}$ results in the closed conformation of Lck (Boggon and Eck, 2004)).
- Recruitment of the substrates and stimulation of processive phosphorylation (for example: the Lck SH2 domain facilitates the

processive phosphorylation of phosphosites within the TCR complex (Lewis et al., 1997)).

- Promotion of transphosphorylation (for example: SH2 domains of ZAP70 bind phosphorylated ITAMs (Wange et al., 1993) which exposes the molecule for a phosphorylation by another ZAP70 (Di Bartolo et al., 1999))

Binding of pTyr^{505} to the Lck SH2 domain is relatively weak and can be easily outcompeted by other targets (Nika et al., 2007). Therefore, the Lck SH2 domain can actively participate in positioning the molecule in the vicinity of kinase substrates. However, its binding partners are not well characterized. Many phosphopeptides have been attributed as potential targets of the Lck SH2 domain, but only some of them were validated on the protein level in cellular systems (**Appendix Tab. 1**). Validated proteins include ZAP70 (Granum et al., 2014, Duplay et al., 1994, Pelosi et al., 1999, Straus et al., 1996, Yamasaki et al., 1996), CD3ε (Lewis et al., 1997), CD45 (Courtney et al., 2017, Ng et al., 1996), LIME (Brdickova et al., 2003, Hur et al., 2003), TSAd (Granum et al., 2008, Granum et al., 2014, Marti et al., 2006), SHP1 (Granum et al., 2014, Štefanová et al., 2003), Syk (Couture et al., 1996a), Pyk2 (Granum et al., 2014), Itk (Granum et al., 2014), Lnk (Li et al., 2000), LAT (Granum et al., 2014) and Lck itself (Lee-Fruman et al., 1996).

The Lck SH2 domain itself harbors another important phosphotyrosine that participates in Lck regulation (Couture et al., 1996b, Courtney et al., 2017). Phosphorylation of Tyr^{192} was recently found to inhibit CD45 interaction with Lck. As a result, dephosphorylation of Tyr^{505} is blocked and Lck preserves a closed conformation (Courtney et al., 2017). This phenotype is similar to CD45 KO T cells (McNeill et al., 2007).

Modification of the SH2 domain through phosphorylation is not uncommon – Tyr^{192} is a conserved residue in the Src kinase family (i. e. Src Tyr^{216}). In both Src (Stover et al., 1996) and Fyn (Smida et al., 2007),

phosphorylation of this residue increased kinase activity. There are also examples of other proteins containing an SH2 domain that can be post-translationally modified causing a change in protein behavior. For instance, the SH2 domain of the adaptor protein Tensin-3, after being phosphorylated, gets increased affinity towards its targets (Qian et al., 2009).

It has also been shown that phosphorylation of Lck Tyr^{192} changes the specificity of the Lck SH2 domain (Couture et al., 1996b, Granum et al., 2014). Potentially, this may diverge Lck signaling to another pathway in the T cells, through higher specificity toward given substrates. Syk (Couture et al., 1996b), ZAP70 (Sjölin-Goodfellow et al., 2015, Couture et al., 1996b) and Itk (Granum et al., 2014) have been shown to phosphorylate Lck Tyr^{192}. As all of these are downstream of Lck in the TCR signaling pathway, there is a possibility for the existence of negative feedback loops. To fully understand the influence of Lck Tyr^{192} on TCR signaling, it is necessary to establish a physiologically relevant system. Only endogenously expressed Lck, which contains an altered SH2 domain, can ultimately reveal how the Lck SH2 domain specificity affects TCR signaling pathway.

TSAd – T cell specific adaptor protein

TSAd, a protein encoded by the *SH2D2A* gene, was first identified in 1998 (Spurkland et al., 1998). It was soon found to interact with Lck, thus its alternate designation is Lck adaptor protein (LAD) (Choi et al., 1999). It was also named after other interaction partners: Rlk/Itk-binding protein (RIBP) (Rajagopal et al., 1999) and VEGF receptor-associated protein (VRAP) (Wu et al., 2000). TSAd has been mainly studied in the context of TCR and VEGF signaling pathways. This may be due to the TSAd expression pattern, as it is expressed only in lymphocytes (specifically in T cells (Spurkland et al., 1998) and NK cells

(Nejad et al., 2004)), endothelial (Wu et al., 2000) and epithelial cells (Kolltveit et al., 2010).

TSAd has two particular domains: an SH2 domain and a long, disordered C-terminal proline-rich tail containing four tyrosines. Five alternative splicing variants of TSAd exist. Especially interesting is variant 5, which omits exon 7 and, as a result, does not contain the proline-rich region and the tyrosines (Granum et al., 2006). In T cells, TCR (Sundvold et al., 2000, Spurkland et al., 1998) and cAMP signaling (Dai et al., 2004) regulate TSAd expression. cAMP strongly induces TSAd mRNA expression in primary T cells. Additionally, protein kinase A (cAMP-dependent protein kinase) activity was required for TCR-dependent induction of TSAd expression (Kolltveit et al., 2008). Another potent activator of TSAd expression is a combination of phorbol myristate acetate (PMA) and Ionomycin (IO), both in T cells and NK cells (Moussa et al., 2015). PMA activates PKC directly, while IO directly induces Ca^{2+} release from the endoplasmic reticulum (Chatila et al., 1989). Although it is an artificial system which bypasses ITAM phosphorylation, it is an efficient method of initiating the gene transcription, induced by aforementioned signaling pathways.

A polymorphism in the promoter region of *SH2D2A*, resulting in a shorter promoter sequence, has been associated with increased susceptibility to multiple sclerosis (Dai et al., 2001), juvenile rheumatoid arthritis (Smerdel et al., 2004), chronic inflammatory demyelinating polyradiculoneuropathy (Notturno et al., 2008) and Sjögren's Syndrome (Taylor et al., 2017). T cells homozygous for shorter variants of the promoter region displayed lower levels of TSAd upon TCR stimulation (Dai et al., 2001). *SH2D2A* contains at least 63 known single nucleotide polymorphisms (SNPs). A SNP resulting in serine to asparagine substitution at amino acid position 52 increased susceptibility to multiple sclerosis (Lorentzen et al., 2008) and ovarian cancer (Kaplun et al., 2012). Furthermore, asparagine in position 52 increased TSAd

transcription and its interaction with Lck as measured by the yeast β-galactosidase reporter assay (Kaplun et al., 2012).

The initial report of *SH2D2A*-deficient mice having a mild autoimmune phenotype (Drappa et al., 2003) has not been confirmed by our group (Berge et al., 2012). However, a number of phenotypic characteristics were discovered, when the mice were challenged in various models. Tumor growth in TSAd KO mice is slower, possibly due to reduced vascularization (Matsumoto et al., 2005). Additionally, they are slightly more resistant to myeloma development than TSAd wild-type (WT) mice in a myeloma-specific TCR-transgenic model (Berge et al., 2012). TSAd KO mice are not responding to VEGF-induced vascular permeability. VEGF stimulation does not disrupt VE-cadherin junctions in the endothelial cell lacking TSAd (Sun et al., 2012). Challenged with viral infections, TSAd-deficient mice displayed reduced clearance of murine cytomegalovirus in the spleen (Moussa et al., 2015). Murine *SH2D2A*$^{-/-}$ $CD4^+$ T cells have impaired polarization of multiple molecules involved in TCR signaling in the IS (Abrahamsen et al., 2018). Moreover, TSAd KO mice had an accelerated rejection of heart transplants in the MHC class II-mismatched model and they were resistant to graft-prolonging therapy of costimulatory blockade in the fully mismatched model (Wedel et al., 2019). Although the majority of studies pinpoint TSAd regulatory role in T cells, the exact mechanism of action is still under consideration.

TSAd in the intracellular cell signaling

TSAd has been reported to control the opening of adherens junctions of the endothelium and consequently, vascular permeabilization (Sun et al., 2012, Li et al., 2016, Matsumoto et al., 2005). Upon VEGF binding to VEGFR2, the receptor dimerizes and initiates downstream signaling which engages tyrosine protein kinases. The SH2 domain of TSAd recognizes $pTyr^{951}$ (Tyr^{949} in mice) on the intracellular tail of VEGFR2. Through that binding, TSAd bridges VEGFR2 with Src kinase, which can

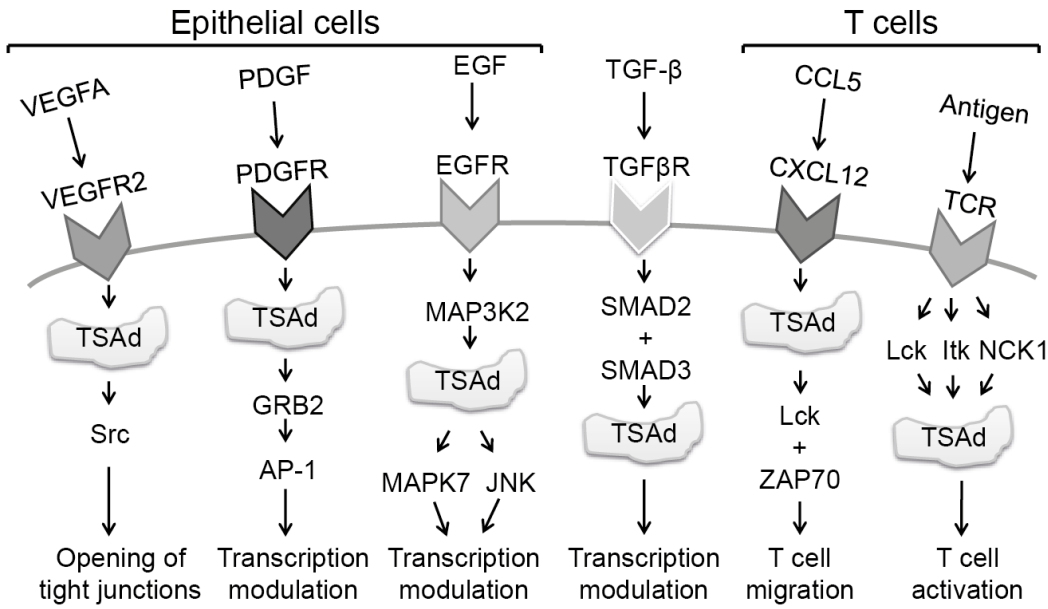


Figure 5. TSAd in the intracellular cell signaling. TSAd has been implied to participate in a number of cell signaling pathways. Some of them, which were studied the most, are illustrated in the figure. See text for more details.

bind to the TSAd proline-rich region. As a result of membrane translocalization, Src kinase phosphorylates the VE-cadherin's intracellular domain, which leads to opening of adherens junctions (Sun et al., 2012). VEGFR2 Y949F mutant mice have significantly reduced vascular permeabilization, which blocks molecular extravasation, oedema and metastatic cancer spread (Li et al., 2016).

While TSAd has been studied much more extensively in TCR signaling, there is no consensus so far on its role in T cells. TSAd interaction with Lck has been investigated the most. Lck can bind with its SH3 domain to the proline rich region of TSAd (Granum et al., 2006). Additionally, Lck can phosphorylate multiple tyrosines on TSAd (Tyr^{280}, Tyr^{290} and Tyr^{305}) and bind to them via its SH2 domain (Granum et al., 2008). It has also been shown that presence of TSAd promotes diffusion of the Lck clusters in the membrane (Kapoor-Kaushik et al., 2016). TSAd influences phosphorylation of various molecules of the TCR pathway. For example, TSAd promotes phosphorylation of Itk Tyr^{511} (Berge et al., 2010) and Lck Tyr^{192} (Granum et al., 2014), but downregulates phosphorylation of LAT,

SLP76, PLC-γ, ZAP70 and CD3ζ (Marti et al., 2006, Sundvold et al., 2000).

As TSAd is an adaptor protein without any enzymatic activity, its only function is to bind and bring together other molecules (for example kinases with its substrates). Lck, Src and VEGFR are not the only established TSAd interaction partners (**Fig. 5**). Itk can also bind to the proline-rich region of TSAd via its SH3 domain (Berge et al., 2010, Andersen et al., 2019). In mice, TSAd has been shown to become phosphorylated upon platelet-derived growth factor receptor (PDGFR) activation in bronchial epithelial cells (Park et al., 2001). That event leads to TSAd association with PDGFR and GRB2, another adaptor molecule. Also in mice, TSAd was proven to bind MAP3K2 and colocalize with it in the IS (Sun et al., 2001). Upon epidermal growth factor receptor (EGFR) stimulation, the TSAd-MAP3K2 interaction is responsible for activation of MAPK7 and JNK, most likely by facilitating their phosphorylation by Src (Yao et al., 2010, Sun et al., 2003). TSAd can bind with its SH2 domain to Smad2 and Smad3, which are involved in TGF-β receptor (TGFBR) signaling transduction (Richard et al., 2006). Upon stimulation with CXCL12 and CCL5 chemokines, TSAd brings together the β-subunit of G protein-coupled chemokine receptor, Lck and ZAP70, which is necessary for the activation of the latter molecule (Park et al., 2007). The TSAd SH2 domain and the proline-rich region mediates binding to the laminin binding receptor (RPSA), which promotes T cell migration (Park et al., 2009). The SH2 domain of NCK1 can recognize TSAd $pTyr^{280}$ and $pTyr^{305}$, while its SH3 domain can bind to the TSAd proline-rich region (Hem et al., 2015). That binding facilitates NCK1 interaction with Lck and SLP76, which potentially promotes actin polymerization in T cells. Lastly, TSAd can possibly interact with DSCAM and DSCAML1 in neurons, recognizing their phosphotyrosine motifs, although the physiological relevance is unknown (Sachse et al., 2019).

We have previously shown a surprising connection between Lck pTyr^{192} and TSAd pTyr^{290}: phosphorylation of the Lck SH2 domain on Tyr^{192} (as mimicked by Lck Y192E mutation) changed specificity and increased affinity toward TSAd pTyr^{290} peptides (Granum et al., 2014). Therefore, the original aim of the project was to define the interactome of the Lck SH2 domain with Tyr^{192} mutated into a glutamic acid and to define the binding partners of TSAd pTyr^{290}. However, during the project it was revealed that the role of Lck pTyr^{192} is much more complex than originally thought and that TSAd pTyr^{290} does not necessary serve as a binding site.

Aims of the project

The overall aim of the project was to decipher the role of conserved non-classical phosphosites located in Lck (pTyr^{192}) and its adapter protein TSAd (pTyr^{290}). Therefore, the project had the following objectives:

1. To develop, using CRISPR/Cas9 genome editing, cell lines with the endogenous expression of the mutated variants of Lck (Y192F and Y192E) and TSAd (Y290F and Y290E).

2. To analyze the role of Lck pTyr^{192} in the TCR signaling pathway, as well as in the Lck regulation, using stably mutated cell lines, various pull-downs, mass spectrometry and other biochemical assays.

3. To analyze the role of pTyr^{290} in the TCR signaling pathway, as well as in the TSAd regulation, using stably mutated cell lines, various pull-downs, mass spectrometry and other biochemical assays.

Methodological considerations

Cells

The primary choice of cells in the work presented in this thesis were Jurkat TAg (JTAg) cells (Northrop et al., 1993). JTAg cells were derived from Jurkat cells and were made to constitutively express SV40 large T antigen, which enhances expression of SV40 origin-containing vectors. The Jurkat cell line was first described in 1980, as a transformed human T-cell line derived from childhood T acute lymphoblastic leukemia, which was a particularly potent producer of IL-2 upon phytohaemagglutinin stimulation (Gillis and Watson, 1980). The Jurkat cell line soon became a model of resting T cells, which could be activated upon various stimulations (anti-TCR, PMA/IO) (Manger et al., 1985). Over the next 20 years, many mutants were generated from the initial cell line, essentially defining our current knowledge of the TCR signaling pathway (Abraham and Weiss, 2004). However, with growing knowledge of the signaling molecules in T cells, several issues of the Jurkat cells were also revealed. Jurkat cells does not express Syk kinase (Fargnoli et al., 1995), CTLA4 (Lindsten et al., 1993), SHIP phosphatase nor PTEN (Freeburn et al., 2002). Lack of PTEN (a tumor suppressor commonly lost in cancer (Chen et al., 2005)) causes hyperactivity of phosphatidylinositol 3-kinase (PI3K). As a result, Jurkat cells have elevated activity of the protein serine-threonine kinase AKT and Itk, since these enzymes are recruited to the cell membrane upon PI3K signaling (Shan et al., 2000). Moreover, genome-wide sequencing has confirmed that Jurkat cells are mainly tetraploidal, which means that most of their genes have four copies. Additionally, some genes have mutations linked to clinical syndromes (*TP53, MSH2*), or resulting in protein KO (*BAX, C1GALT1*) (Gioia et al., 2018).

Despite all these defects, Jurkat cells remain a solid model system for T cell receptor studies (Abraham and Weiss, 2004). The alternative to

Jurkat cells are primary T cells, either human or mice. Mice models are in general more physiologically relevant than cell lines, but generation of new transgenic or KO strains is time-consuming and requires significant expenses. Moreover, some mutations or gene KOs are lethal, and others result in a developmental block or defects in T cells. This was partially resolved with introduction of conditional KO strains, but even these systems are not perfect. Nonetheless, even if the murine immune system is considered to closely resemble the human immune system, certain details differ. ZAP70 is for example necessary for development of $CD4^+$ T cells in mice (Negishi et al., 1995), while it is not essential in humans (Arpaia et al., 1994). Additionally, working with both human and mice primary T cells is ethically more challenging than performing experiments on cell lines.

The versatility of the Jurkat cells increased with the development of small interfering RNA (siRNA) technology, which has, within the last few years, been supplemented with the CRISPR/Cas9 genome editing technology. Both methods can be used for high-throughput screenings of knock-down (Yeung et al., 2009) or KO gene libraries (Datlinger et al., 2017), and therefore, they require a potent cell source. Although, siRNA allows to knock-down the expression of defined proteins (Sharp, 1999), its efficiency depends on the number of siRNA molecules per cell. As a result, its efficiency varies from cell to cell and this dilutes the phenotype of the bulk of cells. Additionally, siRNA knock-down can also cause off-target effects (Echeverri et al., 2006). CRISPR gRNA libraries quickly replaced siRNA libraries, as they could provide clean KO (Cong et al., 2013).

To sum it up, Jurkat cells remain the major model system in biochemical studies of TCR signaling. They are a relatively cheap and a quick source of cells, due to their robustness and high proliferative rate. With new technologies emerging (siRNA and CRISPR/Cas9) their potential has increased even more. In this thesis, Jurkat cells were extensively used

for the generation of cell lines expressing mutated versions of Lck and TSAd. As most of the experiments were performed with biochemical methods, the robustness of the Jurkat cells model was a requirement.

CRISPR/Cas genome editing

Clustered regularly interspaced short palindromic repeats (CRISPR) together with CRISPR-associated proteins (Cas) were initially discovered as a prokaryotic immune system. In 2007, scientists showed that bacteria and archaea use the CRISPR/Cas system in an adaptive manner to defend against foreign DNA (for example of viral origin) (Barrangou et al., 2007). First, Cas1 and Cas2 proteins acquire foreign DNA material by recognizing in their sequences' protospacer adjacent motifs (PAM). PAM is a 2-6 base pair DNA sequence. In the most broadly used CRISPR/Cas system, derived from *Streptococcus pyogenes,* the PAM site is 5'-NGG-3' (Mojica et al., 2009). The foreign sequence following the PAM site is cut out and incorporated into the CRISPR locus as a "spacer". Next, the CRISPR locus is transcribed as an RNA which is subsequently cleaved into pieces of CRISPR RNA (crRNA). crRNA contains each spacer's sequence with a part of the repeat sequence present in the CRISPR locus. The crRNA associates with Cas proteins, forming a ribonucleotide complex (Brouns et al., 2008). This complex detects foreign DNA, which has to contain the PAM site followed by the specific spacer. Nuclease activity of the Cas protein is used to cut DNA bound to the complex.

This bacterial process has been adapted and exploited into a simple, but powerful gene editing method. The system requires only two things to be delivered into cells: Cas9 protein and guide RNA (gRNA). gRNA is a designable and simplified version of the crRNA. The whole system is highly flexible, with many possible alternative uses. For example, Cas9 molecules can be delivered as a protein, as ribonucleocomplexes, as a Cas9-coding RNA, or as a Cas9-coding DNA, with multiple available

transfection methods (viral vectors, electroporation, lipofection, etc.). Moreover, there are many various versions of Cas9 protein in use (Ribeiro et al., 2018): fluorescently labelled (for screening purposes), Cas9n (nicking only one strand of DNA) (Ran et al., 2013a), eSpCas9 (with enhanced target specificity) (Slaymaker et al., 2016), chemically-inducible Cas9 (Liu et al., 2016), or base-editing Cas9 (Komor et al., 2016).

However, the most important part of the system remains the chosen gRNA and PAM site. gRNA is the spacer, from the bacterial CRISPR/Cas system. Designing it for targeting the specific gene, although simple in principle, it is more complicated than anticipated. The canonical PAM site (recognized by predominantly used Cas9 from the *Streptococcus pyogenes*) is prevalent enough to provide a high flexibility in gene targeting (in humans "GG" frequency is 5.21%, which should give a "GG" every 42 nucleotides (Read, 2009)). Nevertheless, gene targeting proved to be much more difficult, as not all the PAM sites result in the Cas9-mediated DNA cut. To overcome the problem, multiple *in silico* techniques have been developed and has become publicly available as an online software (Chuai et al., 2017). These gRNA design tools typically use experimental data to predict gRNA on-target efficacy with machine learning models. Although, they are taking into account multiple parameters and their predictability increases with each generation, it is still hard to explain why some gRNAs fail to target. Among the proposed reasons are differences between cell types/organisms, nucleotide preference, GC content, microhomology sequence patterns, gRNAs' secondary structure, chromatin accessibility and epigenetic features (Chuai et al., 2017).

Although specific targeting of the Cas9 machinery is the main key to successful genome editing, the main concern of the scientific community are off-target effects. These are a major drawback of any CRISPR/Cas9-related clinical therapy trial. Although the initial reports about

CRISPR/Cas9 unspecific targeting were alarming (Fu et al., 2013, Lin et al., 2014), scientists are now more careful with the estimations (Tan et al., 2015). Again, they point to gRNA as the most crucial part of CRISPR/Cas9 specificity (Young et al., 2019). Nevertheless, the off-target effects remain a subject of intensive research, for example, in development of novel off-target prediction tools (Chuai et al., 2017), or improved CRISPR/Cas9 systems, such as paired nickases Cas9n (Ran et al., 2013a) or enhanced specificity eSpCas9 (Slaymaker et al., 2016).

The most unpredictable feature of CRISPR/Cas9 system is the probabilistic nature of DNA-repair mechanisms. When scientists introduce knock-in mutations, they utilize the mechanism of homology directed repair (HDR), which requires a homologous piece of DNA to recreate the original sequence (or introduce a mutated sequence) (Ran et al., 2013b). However, a much more prevalent endogenous repair mechanism is non-homologous end joining (NHEJ), which utilizes short homologous DNA sequences present at the site of DNA break (Moore and Haber, 1996). NHEJ can perfectly recreate the original sequence, if the DNA break resulted in "sticky ends". Otherwise, the NHEJ mechanism will create a homologous match by trimming the single DNA strands. When the aim is to generate KO cells, the latter outcome is preferred. Random deletions and insertions are the source of premature stop codons, which terminate gene transcription prematurely, producing a non-functional protein or none at all.

However, each KO clone, despite being generated in a similar manner, usually has a completely different genotype. As long as Cas9, paired with gRNA, is present in the cell, the target DNA sequence is repeatedly cut, until NHEJ removes or mutates the PAM site. Such a strong selection pressure gives a high rate of successful mutations, but the resulting DNA sequence remains random. Recently, scientists noticed that the outcome, to a certain extent, can be predicted. Using machine learning and knowledge about the NHEJ mechanism, an online tool has been

developed – inDelphi (Shen et al., 2018). The algorithm can predict what the most probable DNA sequence outcome will be after using a certain gRNA. Although, it is definitely a clever step toward reducing the randomness of CRISPR/Cas9 genome editing, the final outcome remains stochastic.

Besides off-target effects, another potential risk of CRISPR genome editing is that each allele may be differentially repaired. Although CRISPR genome editing targets the genes with unprecedented precision, there is always a chance of generating heterozygous or hemizygous clones. Combining it with the possibility of off-target effects and the unpredictability of DNA-repair mechanisms, the outcome is hard to foresee. Moreover, the source cells are always heterogeneous. Cell lines are constantly mutating as their molecular brakes are removed, while primary cells have various stages of development/activation/senescence. Therefore, in this thesis all the generated mutants were validated by treating them as individuals (as in animal research). Any observed phenotype was validated in a number of independently generated clones, to exclude the influence of clonal heterogeneity.

Despite all its flaws, CRISPR/Cas9 is the most simple, user-friendly, affordable, fast, robust, accessible and precise genome editing technique currently in use. The increasing number of studies utilizing this technique, including results presented in this thesis, testify to this statement.

Phosphomimetics

Studying transient modifications of endogenously expressed proteins is not a trivial task. For example, there are only three commonly used methods measuring phosphorylation events: radiolabelling, antibody-based techniques and mass spectrometry. All of them have their own limitations and none are perfect. Measuring phosphorylation in living

cells is even more problematic and it is mostly based on genetically encoded, fluorescently labelled biosensors (Förster resonance energy transfer (FRET)-based in particular) (HS Lu et al., 2012). A biosensor structure is highly artificial and can disrupt biologically-relevant protein interactions. Another popular method of monitoring phosphorylation-based interactions is by using phosphoamino acid mimics (HS Lu et al., 2012). However, these unnatural amino acids are synthetic, semi-synthetic or incorporated via nonsense suppression, which can significantly affect physiological behavior of proteins harboring such amino acids.

Therefore, a lot of studies use natural phosphomimetics – amino acid substitutions that resemble in size and charge phosphorylated serine (Ser/S), threonine (Thr/T) or tyrosine (Tyr/Y). The most popular are glutamic (Glu/E) and aspartic (Asp/D) acids, which quite closely mirror phosphorylated serines and threonines (Maciejewski et al., 1995). It has even been shown that Glu and Asp may be evolutionary precursors of pSer, pThr and pTyr (Pearlman et al., 2011). However, substitution of pTyr with Glu is widely criticized, as the amino acids are similar only in the negative charge. Moreover, substitution of pTyr with alanine (Ala/A) quite often results in similar phenotype as Glu, despite Ala being the second simplest amino acid. On the positive side, phenylalanine (Phe/F) is a great mimic of constantly dephosphorylated Tyr, as they share a similar aromatic structure.

The most important advantage of using natural phosphomimetics is that they can be physiologically incorporated into a protein structure, following the proteins natural transcription and translation. Therefore, they can be used to generate genetically modified animal models. Even if Glu is not a perfect mimic of pTyr, it is still widely used in the functional studies of phosphotyrosinylated proteins. However, the Glu substitution should always be validated with comparison to Phe or Ala (Anthis et al., 2009), to exclude the possibility that Glu is actually

mimicking a not phosphorylated Tyr. In this thesis, all Glu substitutions were compared with Phe mutations.

Immunoprecipitation and pull-down assays

Although immunoprecipitation and various pull-down assays of similar concept are very old techniques, as reviewed in *Current Protocols* (Bonifacino et al., 2016), they have never become trivial. The variety of options and adjustable parameters make them flexible and difficult at the same time. As the protocol and basic rules for the immunoprecipitation and the pull-down assays are the same, all the following considerations apply to both methods, unless specified differently.

The first step that requires precise insight are the lysis conditions of the cellular sample. The binding epitope on the protein can differ drastically between denaturating and nondenaturating lysis conditions. In most cases it is essential to perform the pull-down in nondenaturating conditions, as this is the only way for most of the protein-protein interactions to be preserved. The basic concept behind the lysis of cells is to use a detergent strong enough to dissolve cellular membranes and release the proteins of interest into the solution. By increasing the detergent's strength (defined by critical micelle concentration - CMC) one can achieve a fractionation of cellular content, subsequently releasing proteins from different cellular compartments (Baghirova et al., 2015). However, believing that a lysis buffer affects only lipid membranes within the cell is too simplistic. Detergents and ionic strength of lysis buffers are not only disrupting the cell membranes, but may also affect protein complexes and their intramolecular interactions. Moreover, it is not known how various lysis conditions affect liquid-like, phase separated nanoclusters, which are based on low affinity protein-protein interactions (Ditlev et al., 2019, Su et al., 2016). Therefore, even if the best conditions for retrieving the protein of interest will be defined,

it does not necessarily mean that the conditions are optimal to preserve interactions with other proteins from the complex. The majority of experiments presented in this thesis was done using a combination of 0.1% LDS and 1% Triton in the cell lysis buffer. This rather harsh lysis condition releases even proteins associated with cellular membranes into the solution, allowing to screen for most interaction partners in one experiment.

Another issue that requires proper considerations is unspecific binding. Any beads, covered with antibodies or other molecules will, to a certain degree, bind to unspecific proteins. A common tactic to solve that problem is to perform preclearing of the lysate (Bonifacino et al., 2016). Either this is done with uncoated beads or beads covered with an irrelevant antibody (for example an isotype control which does not bind anything specifically). However, unspecific binding is also generated by the specific antibody used for immunoprecipitation. Therefore, the preclearing step can be omitted if a proper control is used, in other words a sample which does not contain the specific target. Such a control can show the degree of unspecific binding in the experiment. It is especially important for co-immunoprecipitation, as such a control can define the background threshold of co-precipitated targets. The co-immunoprecipitation experiments presented in this thesis were performed using a preclearing step with empty beads. KO cell lines of the target proteins were used as a negative control whenever it was relevant.

One of the major challenges of immunoprecipitation is the specificity of the antibody. Antibody production is a stochastic process where the researcher has very limited influence on the actual end product. Each monoclonal antibody recognizes only a certain epitope, which does not have to be present on all possible forms of the target protein. Therefore, each antibody should be thoroughly validated. However, the protein forms present in the samples are dependent on the sample separation. As a consequence, it is nearly impossible to predict the actually

precipitated molecular targets. The antibody can interact with different post-translationally modified protein forms, with different protein isoforms created by alternative splicing, with various protein degradation products, with protein paralogs, with different conformational states of the protein, or with various protein-protein complexes. In tagged proteins pull-down there is additionally a risk that the tag itself is affecting the physiological interactions of the protein. As the assessment of the real antibody interactions within each experiment would involve a lot of advanced methods, it is necessary to overcome the problem through proper control design. Again, the most helpful control is the one which does not contain the antibody's target. That control can easily define the antibody's unspecific targets, which are unrelated to the protein of interest. Alternatively, the immunoprecipitation experiment can be performed in parallel with two different monoclonal antibodies targeting the same protein. As mentioned in the previous paragraph, the first option was the default approach used in this thesis.

The bait molecule itself is a commonly underappreciated problem in pull-down assays. It is usually present in the final sample in amounts higher than any other molecule, which can cause serious problems during analysis procedures. In western blotting, the bait can give cross-reactivity with the primary or secondary antibody. Due to bait molecule abundancy, its unspecific interaction may give a very pronounced band or a very strong background signal on the membrane. Additionally, if the bait molecule's major band migrates within the range of the analyzed gel, it will thoroughly cover the area, as a consequence, blocking less distinctive signals through spatial constraints. In this thesis, to avoid such problems, beads covered with the bait molecule were directly run on the gel and used as an additional control. Such a control analyzed separately can reveal all the unspecific signal caused by the presence of the bait in the sample. Additionally, in antibody-based pull-downs, light-chain specific secondary antibodies were used to decrease the

background caused by the presence of the antibody bound bait in the sample.

The best alternative for studying protein-protein interactions, are microscopic methods combined with assays giving signal only if the proteins of interest are in close proximity. The best-known method of this kind is FRET. However, the largest disadvantage of imaging-based methods is that these techniques do not allow for unbiased screening. With these techniques, each protein-protein interaction has to be tested with separately designed constructs. Design of such constructs is demanding and there is a high risk of affecting the physiological interactions with the bulky fluorescent/molecular labels. For those reasons, and especially the inability to easily adapt the method for screening purposes, microscopic methods were not used to study protein-protein interactions in this thesis.

Mass spectrometry

Mass spectrometry (MS) is an analytical method, which allows to identify molecules in the sample through measurements of the mass-to-charge ratio of ions. Protein samples are fragmented with digesting enzymes, purified and fractionated with liquid chromatography. The fractionated peptides are subsequently ionized, sorted by the mass analyzer according to their mass-to-charge ratio and sent to the detector, which records the abundance of each ion. Generated mass spectra are compared with reference sequences from the UniProt human database to identify presence of proteins.

MS performed in this thesis is based on immunoprecipitated/pull-downed samples. MS results are completely dependent on the sample preparation, hence they are affected by the same bias and issues as described in the previous chapter. For example, MS is not compatible with all detergents used for cell lysis, so its choice is crucial. However, both LDS and Triton are tolerated by the MS, so the lysis methodology

applied in this thesis did not require changes for the purpose of MS. Other than this, there are a few other important differences between analyzing samples with western blotting and MS.

With improved MS performance, the bait abundancy in the MS sample is usually no longer as disturbing as the bait presence in the western blot membrane. The bait peptides may still hide other, less numerous peptides, but due to the different nature of the analyzing technique, it should not affect the overall identification of proteins.

The biggest advantage of MS is an unbiased identification of all of the most abundant proteins in the sample. Using western blot for the same purpose, would require a tremendous amount of work and the use of a few hundred antibodies. On the other hand, the antibody-based detection by western blot depends on the antibody sensitivity, which is usually much higher than the sensitivity of MS. This is especially apparent when antibodies against post-translationally modified sites are used. The MS needs a specific enrichment for a certain molecule or post-translational modification to be able to detect the peptides containing the modification. But similarly, as with unbiased detection of proteins, MS enables unbiased detection of multiple (possibly unknown) post-translationally modified sites on the protein of interest. In this thesis, all the results generated by the MS, whenever possible, were confirmed by western blot.

The largest advantage of MS, which is to detect unbiasedly the most abundant proteins in the sample, is at the same time its biggest disadvantage. The most abundant proteins in the sample, if it is of cellular origin, are the housekeeping proteins. For cell lines, biasedness can be further increased due to cancerous transformation, what led to the development of a database called CRAPome (Mellacheruvu et al., 2013), where the most abundant proteins for each cell line, under certain conditions of sample preparation, were defined. To determine if the

presence of the housekeeping protein in the sample is relevant, it is necessary to have a proper control to define the background binding. In this thesis, all samples were used to control each other. The significance of each protein or protein modification detected by MS, was defined by the comparison with another sample. The consecutive difference between samples defined the specificity of the results.

Summary of results:

Paper I – A simple and efficient workflow for generation of knock-in mutations in Jurkat T cells using CRISPR/Cas9

In this paper we reported our experience and methodological considerations for the use of CRISPR/Cas9 genome editing to generate knock-in mutated cell lines. We aimed to mutate two different genes on multiple loci in Jurkat TAg cells. We succeeded in mutations of Lck Tyr^{192} into Phe^{192} or Glu^{192}, Lck Tyr^{505} into Phe^{505}, and TSAd Tyr^{290} into Phe^{290} or Glu^{290}. However, we were not able to substitute TSAd Arg^{120} for Lys^{120}, neither TSAd Asn^{52} for Ser^{52}.

Our protocol consists of several basic steps:

1. Subcloning of the original cell line to obtain monoclonal founder cell lines.

2. Transfection of cells with the gRNA- and Cas9-encoding plasmid, a GFP-encoding plasmid and a single-stranded donor oligonucleotide (ssODN) repair sequence.

3. Bulk cell analysis to confirm the mutation occurrence and subsequent limiting dilution to separate mutated clones from the bulk.

4. Allele-specific PCR (AS-PCR) screening to identify mutated clones.

5. Mutants validation with western blotting, flow cytometry and DNA sequencing.

We underlined in the paper the importance of initial cell line subcloning, because it will decrease the heterogeneity of generated clones. It is especially important for the work with Jurkat cells, as their surface CD3 expression is heterogeneous within the initial population. Another

important matter is the optimization of the AS-PCR in detection of mutants, which is not always a simple task.

Next, we explained our choice of transfection setup. We discussed how to test the setup efficiency with PCR and flow cytometry prior to clonal dilution. We emphasized the need for a thorough screening of generated mutants, as there are multiple possible defects that can be overlooked. Western blotting can show if the protein products are the same in size as the WT protein or produced new alternative variants. Flow cytometry can show if the clones are truly homogenous. While sequencing can confirm if the mutation was correctly introduced on both chromosomes and if no additional mutations were introduced.

Finally we showed that on average, our protocol can generate around 2% of homozygous knock-in mutants. We also discuss why some of our target sites failed to undergo CRISPR/Cas9 genome editing. One of the possibilities are differences in the histone methylation (and therefore, the chromatin accessibility).

To sum it up, paper I presented a simple protocol for the generation of CRISPR/Cas9-induced knock-in cell lines, which can be applied in any basic molecular biology laboratory.

Paper II – Tyr^{192} is a critical regulator of lymphocyte-specific tyrosine kinase activity in T cells

In this paper, we explored the role of Lck Tyr^{192} in a model, where mutated variants of Lck Tyr^{192} were endogenously expressed. We used the cell lines described in paper I which stably express Lck Phe^{192} and Glu^{192}. These mutations are supposed to mimic constantly unphosphorylated Tyr and constantly phosphorylated Tyr, respectively. We aimed to answer the question if pTyr^{192}, located in the Lck SH2

domain, can change its specificity and, as a consequence, change the substrate repertoire of Lck.

However, we observed that the phenotype of Lck Glu^{192} did not allow us to achieve our goal. We showed that Lck Glu^{192} does not respond properly to TCR activation, as it did not induce CD69 expression, had reduced NFAT translocation and had reduced initial phosphorylation of early signaling molecules. The reason for such a phenotype was a concomitant hyperphosphorylation of Lck Tyr^{505}, which left the majority of Lck in the closed conformation. Therefore, the observation that Lck Glu^{192} did not bind Lck SH2 domain targets properly, was not necessarily valid, as this mutant could not initiate phosphorylation of Lck SH2 domain binding partners.

To overcome this problem, we generated another set of CRISPR/Cas9 genome-edited cell lines, where we combined Lck Tyr^{192} mutations with Lck Phe^{505} substitution, creating double mutants. These Lck mutants were constantly in the open conformation, as they lost the ability to be phosphorylated on Lck Tyr^{505}. To our knowledge, we have generated the most active Lck mutants so far reported. Both Lck mutants, Phe^{192}/Phe^{505} and Glu^{192}/Phe^{505}, had a several times increased initial phosphorylation of early signaling molecules upon TCR stimulation. Although, Lck Glu^{192}/Phe^{505} was generating more phosphorylation, there was no significant difference between both mutants in binding Lck SH2 domain target proteins. The hyperactive phenotypes could be explained by increased phosphorylation of Lck Tyr^{394} in both double mutants.

Our results support the recent finding that Lck $pTyr^{192}$ abrogates Lck interaction with CD45 (Courtney et al., 2017). As Lck Glu^{192} cannot be dephosphorylated on Tyr^{505} any longer, the closed-conformation Lck accumulates, negating the cellular response to TCR stimulation. When we combined Lck Glu^{192} with Phe^{505}, we observed the accumulation of $pTyr^{394}$, which is consistent with CD45 being a regulator also of this Lck

phosphosite. However, upon TCR stimulation, the amount of pTyr^{394} increased, which suggested that not all Lck was constantly phosphorylated on Tyr^{394}. To examine how Glu^{192} affects dephosphorylation of Lck pTyr^{394}, we performed kinetics experiments. Upon TCR-stimulation Lck pTyr^{394} increased, to fall back to initial value after 20 min, and to altogether vanish after 60 min of TCR-stimulation. The kinetics were the same for all three mutants, which suggest that Lck pTyr^{394} dephosphorylation, initiated by TCR signaling, is Lck pTyr^{192}-independent (and therefore, most likely also CD45-independent).

Finally, to elucidate why Lck Phe^{192} also displayed increased Lck pTyr^{394}, we performed Lck SH2 domain pull-down experiments, to compare binding ability of Lck SH2 WT, Lck SH2 Phe^{192} and Lck SH2 Glu^{192}. The data showed that Lck SH2 Phe^{192} binding capacity of the known substrates was the same as WT, while Lck SH2 Glu^{192} binding capacity was significantly reduced. This suggests that stable SH2 domain specificity affects Lck pTyr^{394}. Increased stability of SH2 domain binding to the substrates, promotes Lck Tyr^{394} transphosphorylation, and therefore, increases its activity.

Overall, the data in paper II showed that mutation of Lck Tyr^{192} can unlock Lck's enzymatic potential and reveal novel details about its regulation.

Paper III – A conserved pTyr motif in the unstructured C-terminus regulates the function of the T cell specific adaptor TSAd

In this paper, we aimed to discover the role of the conserved Tyr^{290} in the function of the Lck adapter protein TSAd. This tyrosine, and its surrounding amino acid motif, is conserved in amniotes, which suggests that it plays an important role. However, it is not a conventional SH2 or phosphotyrosine-binding (PTB) domain recognition motif.

To define the interactome of pTyr^{290}, we performed an MS experiment to identify proteins pulled down by TSAd Tyr^{290}, pTyr^{290} and pTyr^{305} peptides. TSAd pTyr^{290} peptides were not preferentially bound by any SH2 or PTB domain-containing proteins in comparison to TSAd Tyr^{290}. Moreover, there were no proteins which preferred pTyr^{290} over Tyr^{290}. As a control, we compared protein binding of pTyr^{290} to that of pTyr^{305}, which contains a conventional SH2 domain binding motif. We did not identify any convincing TSAd pTyr^{290} novel binding partners, although we found a group of SH2 domain-containing proteins which specifically preferred TSAd pTyr^{305}, i.e. PI3K, PLC-γ, Src kinases (Lck, Src, Fyn) and SHP2. We confirmed these MS results by western blotting.

As we were not able to identify the role of TSAd pTyr^{290} with peptide pull-downs, we used CRISPR/Cas9 genome editing to generate cell lines stably expressing TSAd Phe^{290} or TSAd Glu^{290}. The first mutant should mimic a constantly unphosphorylated Tyr^{290}, while the latter one should mimic a constantly phosphorylated Tyr^{290}. As a by-product of the procedure, we also generated cell lines expressing TSAd Δex7 – a TSAd variant which does not contain the proline-rich region and tyrosines described in this paper.

The mutated versions of TSAd behaved in the western blot similarly to the TSAd versions which they were mimicking. TSAd Phe^{290} migrated similarly to unphosphorylated TSAd, while TSAd Glu^{290} migrated similarly to phosphorylated TSAd. This behavior did not change upon phosphatase inhibition treatment, which confirmed the notion that Tyr^{290} is crucial for TSAd phosphorylation. When we tested the TSAd mutants' ability to interact with other proteins, we found that TSAd Phe^{290} binding capacity was essentially unchanged, while TSAd Glu^{290} was strongly inhibited. This was even more apparent, when compared with TSAd Δex7, which displayed a similar binding pattern to TSAd WT and TSAd Phe^{290}, but with much lower intensity.

Lastly, we measured CD3ζ phosphorylation upon TCR stimulation. It was shown previously, that TSAd-deficient cells display increased CD3ζ phosphorylation, which cannot be reversed to normal by TSAd Phe^{290} reconstitution. We could confirm this with our mutated cell lines. Both TSAd Phe^{290} and TSAd Δex7 displayed significantly higher amount of CD3ζ tyrosine phosphorylation after TCR stimulation compared to TSAd WT cells. These indicate that TSAd Tyr^{290} regulates TSAd function.

Our results pinpoint the importance of TSAd pTyr^{290} for the protein's function. However, pTyr^{290} does not necessarily form a binding site for other proteins. In paper III, we hypothesized that pTyr^{290} participates in the regulation of TSAd conformation.

General discussion

Disordered protein regions

An intrinsically disordered region is a part of the protein that lacks a stable three-dimensional structure as defined by currently available scientific methods (Dunker et al., 2001). At first, it was shown that most disordered regions of the proteins evolve faster than structured regions (Brown et al., 2002). However, a few years later it was found that linear motifs within the disordered regions, recognized by other domains (SH2, SH3 or Ser/Thr kinases), are surprisingly conserved (Ren et al., 2008). Additionally, the number of proteins containing unstructured regions of 50 amino acids in length or more is much higher in eukaryotes than in bacteria or archaea (Dunker et al., 2000, Basile et al., 2019). This suggests that the evolutionary purpose of developing an intrinsically disordered region is to provide the proteins with higher flexibility. The flexibility could allow complex organisms to produce multifunctional proteins, whose role would differ within a signaling pathway or between various cells.

What are the biological benefits of having a disordered structure? There are a substantial number of features which characterize disordered regions and define their unique function in the intracellular signaling.

Firstly, short linear motifs present in disordered regions are more accessible to binding domains as they are constantly exposed (Ren et al., 2008). If the motif would be hidden within the structured domain, the binding conformation would be much more restricted as in the classic model of lock-and-key protein-protein interactions. The unstructured regions can bind proteins with high specificity, but low affinity, which makes them ideal signaling hubs – both specific and reversible (Dunker et al., 2005). However, such a perspective has been recently challenged (Ivarsson and Jemth, 2019) and it was speculated, that disordered

structure actually provides a low specificity with a high affinity, due to its flexibility. Even if the binding affinity is low, the disordered structure can adapt its surface, increasing the recognition surface and therefore, increasing the interaction stability and its affinity (Sugase et al., 2007). Additionally, disordered linear structures allow the binding motifs to overlap (Ren et al., 2008). This greatly enhances the functionality of the protein within the signaling pathway and/or allows it to participate in distinct signaling pathways.

Secondly, studies show that phosphorylation events predominantly happen within disordered regions of the proteins (Iakoucheva et al., 2004), what is now commonly used in phosphorylation site prediction softwares (Iakoucheva et al., 2004, Li et al., 2008b). This indicates that a certain degree of sequence flexibility is necessary for the phosphorylation to occur. Even the kinase activation loops, which need a phosphorylated amino acid residue to become active, undergo significant conformational changes during the phosphorylation process. It is possible that during the conformation change, the loops become partially disordered, allowing the phosphorylation to happen. In fact, the EF loop of the Lck SH2 domain, where $pTyr^{192}$ is located, is also one of the most flexible parts of the domain (Zvelebil et al., 1995). The post-translational control over the unstructured region is much easier due to the aforementioned reasons. There are multiple examples of phosphorylation in the disordered regions which shape the protein conformation (Khan et al., 2017, Kulkarni et al., 2017, Bah et al., 2015). Such events can stabilize protein interactions (Khan et al., 2017), or on the contrary, destabilize them (Bah et al., 2015, Kulkarni et al., 2017). In addition, controlling the structure through post-translational modifications can establish the order of events within the signaling pathway. The post-translational modification can regulate which motifs within the disordered region are exposed at various stages of the

signaling cascade. Alternatively, it can relegate the protein to other signaling pathways.

Thirdly, disordered regions can serve as linkers between functional domains of the protein. It facilitates the function of the protein through the flexibility of its domains (Gonfloni et al., 1997). These regions can also allow proteins to perform auto-inhibition through interaction with its own protein binding domains (Boggon and Eck, 2004). Alternatively spliced protein variants are enriched in unstructured regions (Buljan et al., 2013), which is another way to increase the signaling complexity of the cell. TSAd itself is an example, as exon 7, which encodes much of its disordered C-terminal tail, is subjected to alternative splicing (Granum et al., 2006). Finally, disordered regions were shown to be a crucial element of the formation of phase-separated microdomains within the cell (Pak et al., 2016, Banjade et al., 2015). A phase-separated microdomain is a membraneless cellular compartment which concentrates certain molecules. These microdomains gained a lot of interest in the T cell studies, as it was shown that LAT, Grb2, Gads, Sos1, SLP76, NCK1 and N-WASp form such clusters, which are actively transported via actin association to the center of IS (Su et al., 2016, Ditlev et al., 2019).

Although, we do not have any direct results, showing with certainty the role of TSAd pTyr^{290} in shaping the conformation of the molecule, we have indirect evidence that this might be the case. The description of the functions of the intrinsically disordered regions presented here should also support our notion that the current knowledge of TSAd fits within the field of unstructured proteins.

Lck-TSAd interaction via TSAd pTyr^{290}

Our initial interest in the assessment of the interactome of TSAd pTyr^{290} came from our previous results, which showed that this peptide preferentially bound to Lck-SH2 domain harboring the Y192E mutation

(Granum et al., 2014). In comparison, TSAd pTyr^{305} preferentially bound the Lck-SH2 WT domain. Therefore, we aimed to simulate the conditions where Lck would be predominantly phosphorylated on Lck Tyr^{192} (pervanadate treated Jurkat cells) or dephosphorylated (PP2 treated Jurkat cells) and compared its binding to the TSAd phosphopeptides. Although our assumption may be valid, it bears unintentional consequences. The samples treated with pervanadate have an excessive amount of all kinds of phosphopeptides, also unphysiological, which can easily outcompete the lower affinity binding of TSAd phosphopeptides to target proteins. Additionally, the excessive phosphorylation can change the proteins' conformation. For example, most of the specific binding partners of TSAd pTyr^{305} are pronounced in PP2 treated samples, while their binding is considerably reduced in pervanadate treated cells as measured by mass spectrometry. Therefore, it is probably impossible to assess the specificity of TSAd phosphopeptides to the Lck SH2 domain with pTyr^{192} in that manner. A better approach may be to repeat the TSAd peptides pull-down in Lck Y192F and Y192E mutated cell lines as mimics of Lck SH2 domain with Tyr^{192} and pTyr^{192} respectively.

TSAd Y290E lack of interaction with other molecules does not necessarily have to exclude the possibility of Lck SH2 domain interaction with TSAd pTyr^{290}. If pTyr^{290} was indeed changing TSAd molecule conformation, it would do it only while interacting with other proteins. The new conformation could still expose pTyr^{290} as a binding site. However, it is possible that the conformational change obtained through amino acid substitution results in the molecule folding in an unphysiological manner. Consequently, the abnormal folding blocks the engagement with other proteins. The order of events can be a crucial factor for TSAd function. Previous work from our group indicate that TSAd is processively phosphorylated by Lck (Granum et al., 2008), which support the notion of TSAd actions being sequential.

Other parts of TSAd and Lck can play a more important role in the TSAd-Lck interaction. For example, Lck SH2 domain can bind other tyrosines on TSAd and the Lck SH3 domain can bind prolines on TSAd (Sundvold-Gjerstad et al., 2005, Andersen et al., 2019). The TSAd SH2 domain specificity has been mapped with a peptide array (Li et al., 2008a) and a dominating motif is pYXN. However, it is likely that the TSAd SH2 domain may have additional binding partners that do not have this motif. For instance, VEGFR2 pTyr^{951}, which is a known TSAd SH2 domain ligand, contains the pYVG motif (Matsumoto et al., 2005). It may thus be that also the TSAd SH2 domain participates in the TSAd-Lck interaction. In any of these cases, the measurement of the interaction between TSAd pTyr^{290} and the Lck SH2 domain on the protein level will inevitably be skewed by the binding of other sites.

Lck adaptor molecules

TSAd is not the only Lck adaptor protein, which phenotype is hard to grasp. Other Lck adaptors include Lnk (Li et al., 2000), ALX (Shapiro et al., 2008) and LIME (Brdickova et al., 2003, Hur et al., 2003). Both Lnk and ALX contain an SH2 domain, and all three molecules contain multiple phosphosites, whose functions were investigated. Similarly to TSAd, all three of them have viable KO mouse models with various abnormalities of the immune system (Perchonock et al., 2006, Velazquez et al., 2002, Gregoire et al., 2007). In contrast, mouse models of SHC, yet another Lck adaptor molecule, revealed that SHC KO die at embryonic stage and conditional SHC KO mice showed that SHC is essential for T cell development (Fukushima et al., 2006). However, the seemingly redundant nature of TSAd, Lnk, ALX and LIME, actually points toward the less obvious conclusion, that the mutant mice were not challenged properly. The redundant function of the protein is assigned by the scientist as an explanation for the lack of benefits from expressing the protein in tested conditions. However, the protein can have an irreplaceable role in specific conditions that happens in nature. Our

organisms are challenged daily by an enormous number of external (and internal) dangers. Therefore, testing them all out in mouse models is a tremendous amount of work. Until new screening methods are developed to systematically assess all aspects of the immune system, a discovery connecting the disease/treatment model with one of the proteins mentioned above will be random. Ideally, a molecular mechanism is identified first, through a broad spectrum of *in vitro* screening methods, and subsequently translated into physiological relevance.

The interactome of the Lck SH2 domain

The initial aim of paper II was the identification of proteins that preferentially interact with Lck SH2 domain upon phosphorylation of Tyr^{192}. Lck Tyr^{192} is located in the EF loop of the SH2 domain (Eck et al., 1994). It is known that both the EF and the BG loops, define the domain's specificity (Liu et al., 2019, Kaneko et al., 2010). In fact, the shift in the specificity of the Lck $pTyr^{192}$ may be due to altered conformation and intramolecular interaction between $pTyr^{192}$ and Arg^{184} within the SH2 domain pocket [Huszenicza et. al., submitted]. The SH2 domain is not the only protein binding domain in which specificity can be shifted by the phosphotyrosine. For instance, phosphotyrosines within the SH3 domain were also reported to affect its interactome (Merő et al., 2019).

To the best of our knowledge, there is no published report of an unbiased screen for Lck SH2 domain interaction partners in T cells. However, peptide array analysis, as well as low-throughput, such as single molecule studies, have revealed that the Lck SH2 domain has a number of confirmed or suspected binding partners in T cells (**Appendix Tab. 1**). It is therefore of interest to investigate the influence of $pTyr^{192}$ on the Lck SH2 domain interactome, as it has previously been poorly addressed. In the initial paper from 1996 (Couture et al., 1996b), the authors narrowed down their investigation to the comparison of anti-

pTyr blots of pull-down with isolated Lck SH2 domains (WT, Y192F and Y192E) performed in Jurkat cell lysates, not discussing which specific proteins were affected by the Lck Y192 mutations. In a previous study from our group, we also focused on the interactome of the isolated Lck SH2 domains (Granum et al., 2014), and showed that there are phosphoproteins and phosphopeptides that preferentially bind to the isolated Lck Y192E SH2 domain mutant. However, we did not examine the binding capacity of the Lck Y192F SH2 domain. In paper II, we addressed that negligence by assessing both phosphoproteins and specific proteins binding to all three Lck SH2 domains: WT, Y192F and Y192E. Our analysis revealed that the WT and the Lck Y192F domain have essentially the same binding capacity, while Lck Y192E has a decreased binding of the Lck SH2 domain to most proteins.

Most signaling molecules, including Lck itself, contain multiple sites for interaction with other signaling molecules. Thus, interactions observed with the isolated Lck SH2 domains may not directly predict how full-length intact Lck will interact with the same protein target. The most striking example is VAV1, whose interactions with the isolated Lck SH2 domains differed from what we observed with the intact, endogenous Lck. Binding of VAV1 to isolated Lck SH2 domains, but not to the intact Lck molecules, was clearly affected by the Tyr^{192} mutation. It shows that other VAV1 and Lck domains (for example SH3) are more critical for their interaction. Moreover, in the Lck mutant cell lines, the amount of VAV1 phosphorylation was affected by the absence of Lck, but not by the Tyr^{192} mutation. Although VAV1 is a known substrate for Lck (Gulbins et al., 1993), also ZAP70 and Fyn phosphorylate VAV1 (Michel et al., 1998). This may explain why there is no significant net effect of Lck $pTyr^{192}$ on Lck interaction with VAV1.

As a consequence of lower binding affinity to classical Lck targets and hyperphosphorylation of Lck Tyr^{505} (Courtney et al., 2017), Lck Y192E fails to activate the T cell signaling pathway. However, our data also show

that Lck Y192E is enzymatically active, as it can phosphorylate its targets and to some extent initiate NFAT translocation. Yet, this activation was not sufficient to initiate CD69 expression. Our data confirm that a certain level of NFAT nuclear translocation must be reached to initiate activation induced protein expression (Fiering et al., 1990). Even if the Lck Y192E mutant is not physiologically accurate representation of pTyr^{192}, it clearly shows the importance of Lck Tyr^{192} for Lck functions.

Phosphorylation of the Tyr^{192} in the Lck SH2 domain

The Lck Tyr^{192} site is conserved in the other kinases from Src family (Src Tyr^{216}, Fyn Tyr^{214} etc.). It is thus possible that the mechanism of the kinase regulation described in paper II is a common characteristic for the whole family. It was reported that phosphorylation of that site in both Src (Stover et al., 1996) and Fyn (Smida et al., 2007) increases their activity. In both papers, the authors explained increased kinase activity with the blocked binding of C-terminal end inhibitory phosphotyrosine. However, in the case of Fyn, phosphorylation of Tyr^{214} was accompanied by increased phosphorylation of the inhibitory tyrosine (Tyr^{531}), which implies that the regulatory mechanism may be similar to that of Lck Tyr^{192}. The basic scheme of regulation (active/inactive conformation ruled by two phosphosites) is the same for the whole Src family (Boggon and Eck, 2004). Therefore, it is highly likely that the phosphosite located in the SH2 domain can be a third regulator of the Src kinase family activity.

To our knowledge, there have been no previous reports on the function of the phosphorylation of Lck Ser^{194} and Ser^{213}. In paper II, we observed a direct correlation between pTyr^{192} (i.e. Glu^{192}) and increased phosphorylation of Ser^{194} and Ser^{213}. However, the function or mechanism of regulation of these sites is unknown. Presence of these two phosphosites correlates with hyperphosphorylation of other Lck tyrosine

sites, which are targets of CD45: Tyr^{505} (in Lck Y192E mutant) and Tyr^{394} (in Lck Y192E/Y505F mutant). Ser^{194} is located in the EF loop, while Ser^{213} is located at the beginning of the BG loop on the opposite side of the Lck SH2 domain groove in relation to Tyr^{192}. This shared location suggests a plausible explanation. Phosphorylation of these serines can be a side effect of pTyr^{192} (or Glu^{192}) – where the altered charge (or steric hindrance) of pTyr^{192} could shift the serine residues, increasing their accessibility to serine kinases. Determining the function of Lck Ser^{194} and Ser^{213} requires further exploration. It is possible that they can also alter Lck SH2 domain specificity similar to pTyr^{192}.

Another surprising observation in terms of Lck phosphorylation was that we did not find pTyr^{192} phosphopeptides in the MS analysis of the Lck IP from the Lck Y505F mutant lysates. This could be a methodological defect or a consequence of the introduced mutation. It is possible that pTyr^{505} binding to the Lck SH2 domain (inactive Lck conformation) stabilizes pTyr^{192}, analogous to pTyr^{505} itself being stabilized by Lck SH2 domain binding (Gervais et al., 1993). Thus, introduction of Y505F, which prevents the Lck closed conformation, makes pTyr^{192} exposed to the activity of phosphatases and therefore, the site can be easily dephosphorylated. On the other hand, this may imply that Lck Y505F and Lck Y192F/Y505F are equal, as both would lack the possibility of Tyr^{192} being phosphorylated. If both mutated cell lines had a constantly prototypic SH2 domain specificity, both would have similarly enhanced signaling. Yet, we can clearly see a superiority of Lck Y192F/Y505F over Lck Y505F activity. Together, this suggests that even if pTyr^{192} is absent in Lck Y505F samples, the phosphorylation must happen in the intact cells, consequently affecting the TCR signaling. This issue requires more attention in future studies of Lck Tyr^{192} phosphorylation regulation.

Lck Y192E binding partners

Although we performed an unbiased MS screening for changes in the interactome of Lck Tyr^{192} mutants, we were unable to identify any interaction partners specific for Lck Y192E (**Fig. 6 A and B**). There are several possibilities that could explain this. Firstly, as mentioned above, the interactions observed between intact proteins and between isolated protein domains or peptides can differ drastically. The signaling network of proteins is a highly dense mesh, which rely on multiple binding/interaction sites. Changing a single interaction site can have a redundant effect or a phenotype observed only under very specific conditions. Therefore, increasing the physiological relevance of the model may not necessarily provide better results than testing protein interactions in isolation.

Secondly, the interaction of an SH2 domain in a kinase can be transient to support phosphorylation in a processive manner (Lewis et al., 1997). Therefore, it could be impossible to capture all "Lck SH2 domain-target protein" interactions with the methodology used in this thesis. A feasible alternative would be to identify novel phosphorylation events caused by the change in SH2 domain specificity. We are currently developing a protocol for phosphoproteomics in samples enriched for pTyr peptides/proteins (Abe et al., 2017, Lind et al., 2012). However, this proved to be more technically challenging than anticipated and the method will require additional optimization.

Thirdly, the phenotype of Lck Y192E, which directly affects Lck activity in the cellular model, is a significant challenge for our experiments. Lck activity directly corresponds to the amount of generated phosphosites in the TCR signaling pathway. Therefore, Lck activity directly affects the amount of available SH2 domain binding targets. In the case of Lck Y192E, the amount of phosphopeptides is practically non-existent, while for Lck Y192E/Y505F the amount of phosphopeptides is a few times

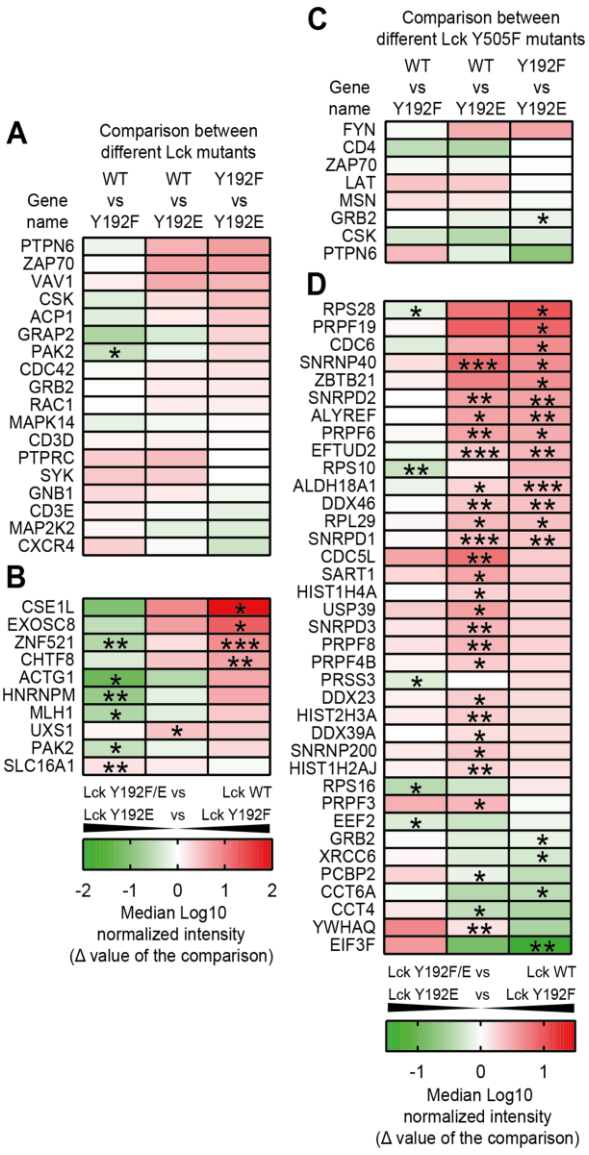


Figure 6. Results of Lck IP in Lck mutants analyzed with mass spectrometry. Jurkat cell lines stably expressing different versions of Lck (Lck WT, Lck Y192F, Lck Y192E) (A-B) or (Lck Y505F, Lck Y192F/Y505F, Lck Y192E/Y505F) (C-D) were stimulated for 2 min with anti-TCR antibody. After stimulation, cells were lysed and subjected to anti-Lck co-immunoprecipitation. Samples were analyzed with mass spectrometry. Experiments were repeated three times. Student's T test (two-sample, unpaired, two-tailed) was performed between each set of samples of Lck variants. Heat-maps present the median of the difference between each set of samples. The hit was considered significant if the p-value was <0.05 in 3 different imputations. A and C) Data was filtered using a list of 100 known Lck interaction partners taken from the STRING database. B and D) All statistically significant comparisons found in the results. Presented p-value is a median of 3 different imputations shown accordingly: * - p-value <0.05; ** - p-value <0.01; *** - p-value <0.001.

higher than that of Lck Y505F. This makes it very challenging to find novel targets in the Lck Y192E cells, while the results from Lck Y192E/Y505F cells will be skewed by Lck hyperactivity (**Fig. 6 C**

and D). Although the doubly mutated Lck cell lines proved to have intriguing phenotypes, they did not help us map the shift in the interactome of the Lck SH2 domain with the Tyr^{192} phosphorylation.

CRISPR/Cas9 as a tool for functional analysis of endogenously expressed proteins

The protocol for knock-in mutant generation, described in paper I is simple, relatively robust (easily scalable), and can deliver a mutated cell line within three weeks. In our experience, the most challenging part is the design of gRNA, ssODN repair templates and AS-PCR primers. All of these components are prone to unspecific interactions, which can disturb the experimental setup at various steps. Therefore, it is necessary to have a possibility for testing them before performing the actual experiment. We have proposed to establish AS-PCR prior to the mutation experiments, using plasmids encoding the expected WT or mutated protein products. We have proposed to test for mutated cells in bulk prior to limiting dilution by flow cytometry (for the presence of KO cells) or with AS-PCR (for the presence of knock-in mutation).

There are two alternative methods that we worked on, but were not sufficiently established prior to manuscript submission, namely the heteroduplex mobility assay and testing of gRNA targeting ability on plasmids.

Heteroduplex mobility assay (**Fig. 7**) – The assay requires a PCR product spanning the site of the cut, amplified from the bulk cells after transfection. Running such a DNA sample on a 10% polyacrylamide gel drastically increases the electrophoresis resolution (in contrast to an agarose gel). It allows for observation of shorter and longer PCR products, as well as, heteroduplexes formed by differentially mutated sequences, which typically migrate with a significant delay. As a result, the bands' pattern of the cells successfully cut by Cas9 differs from WT

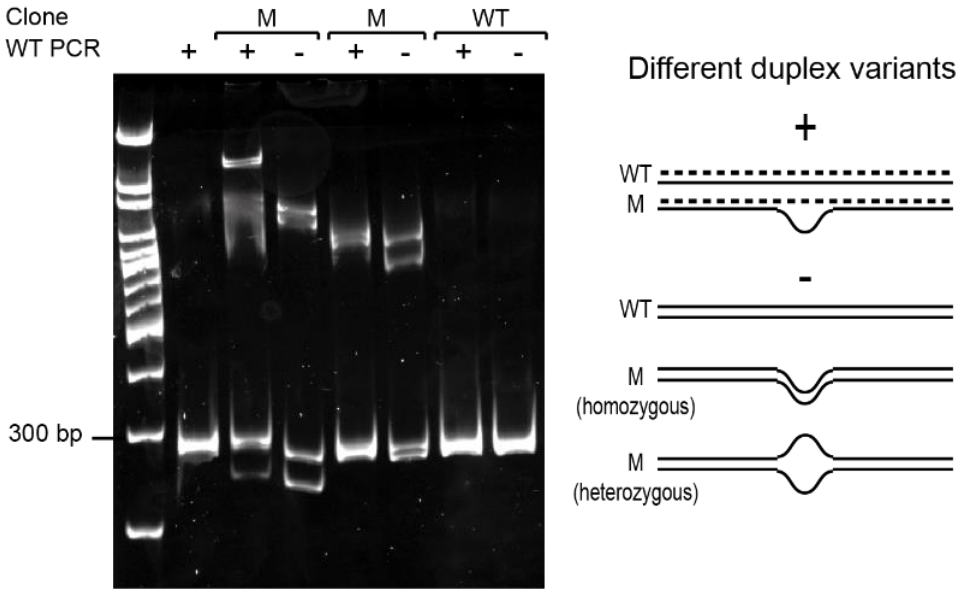


Figure 7. Heteroduplex mobility assay. PCR products spanning the knock-in site were mixed with wild-type (WT) PCR product (+) or left alone (-). These samples were resolved by polyacrylamide gel electrophoresis and visualized with ethidium bromide staining. Mutant (M) sequences form different duplex variants in comparison to WT. These duplex variants can be observed in the gel as a specific pattern of DNA bands.

cells. The assay works well when applied on the single clones, but experiments with bulk cells are inconsistent.

Testing gRNA targeting ability on plasmids – The assay requires simultaneous transfection of cells with plasmids encoding gRNA-Cas9 and the target protein. If the gRNA was designed properly, the expression of the target protein should gradually decrease within a couple of days, as a consequence of the target protein plasmid being cut by Cas9. The protein expression can be monitored by a fluorescent tag or with a proper antibody by flow cytometry or western blotting.

We cannot exclude the possibility that TSAd-targeting gRNAs failed due to unknown technical reasons. Therefore, we have to convincingly show that TSAd-targeting gRNAs, which failed to generate the mutated cell lines, are actually able to target the TSAd genomic sequence. Bulk measurements of TSAd KO generation by flow cytometry was not

optimal due to lack of monoclonal antibodies with high specificity, thus we need to establish other methods to test the gRNA targeting abilities.

Our and others' results (van Overbeek et al., 2016, Shen et al., 2018) point out towards sequence related outcomes of CRISPR/Cas9 genome editing. There are two elements that define the outcome: the gRNA sequence and the local sequence context (van Overbeek et al., 2016). Both of these elements influence the DNA repair mechanisms, which decide the actual outcome. It shows that both the CRISPR/Cas9 system, as well as the DNA repair mechanisms are not only controlled by rules which assure the repeatable outcome, but they can also be predicted to certain degree (Shen et al., 2018).

The CRISPR/Cas9 system is not the only technique that could have been used to explore the function of phosphosites described in this thesis. Transient transfection of cDNA-containing plasmids, which was the major technique used by our lab until now, has been broadly used in molecular biology. While it is easy to perform, it is not perfect, as the protein expression is not physiological nor uniform between cells. Moreover, the corresponding endogenously expressed protein can affect the results if it is not suppressed by siRNA knock-down. On the other hand, developing a stably transfected cell line is time-consuming, laborious and inefficient. Additionally, the random nature of the DNA acquisition can cause an unspecific phenotype.

Before CRISPR/Cas9 system came to dominate the genome editing field, there were other engineered nucleases in development i.e.: meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector-based nucleases (TALENs) (Urnov, 2018). The CRISPR/Cas9 system became the primary choice for genome editing, because of its feasibility, as demonstrated in paper I.

As mentioned in Methodological considerations, Jurkat cells are not a perfect cellular model. Although, the cells should mirror the proximal

TCR signaling quite precisely, their distal TCR signaling is not necessarily authentic. Therefore, the phenotype of Lck mutants which affects the proximal TCR signaling should be genuine, while the phenotype of TSAd mutants which might affect both proximal and distal TCR signaling may be distorted. Nonetheless, most methods are tested at first on the cell lines due to their robustness and simple maintenance. The next step is to apply our methods in primary T cells. Although primary T cells are challenging to maintain due to limited viability, it should be possible to establish mutated clones before they reach exhaustion.

Future perspectives

In future studies, it would be interesting to start combining single amino acid mutations in different proteins in single cell lines. We have already shown in paper II, that such an approach, combining two amino acid mutations in one protein, can produce an unexpected outcome. Therefore, it is possible that a combination of certain mutations in two proteins can rescue the phenotype of a single mutation. It is also possible to observe a synergetic effect. Such experiments would not be as easy as they are nowadays, if it was not for CRISPR/Cas9 genome editing.

Applying the knowledge from paper II into clinical practice can be more than challenging. There is still a lot of unknowns about Lck activity regulation in T cells. Our work is revealing more knowledge gaps than it is closing. For example: Which phosphatases dephosphorylate Lck upon TCR activation? Which phosphatases regulate Lck pTyr^{192} dephosphorylation? How stable/dynamic is the switch between Tyr^{192}/pTyr^{192}? Would hyperactive Lck Y192E/Y505F lead to autoimmune responses?

Although, the transition from basic science to clinical applications is a long way, one can imagine a CAR T cell designed with intracellular motifs which promote phosphorylation or dephosphorylation of Lck Tyr^{192}. For

example, it may be possible to recruit specific phosphatases or adaptor proteins to the CAR which can regulate Lck Tyr^{192} phosphorylation status. Scientists have just begun to think about modulation of intracellular signaling with specifically designed CAR T cells (Karlsson et al., 2015). It could thus be possible to achieve CAR T cells with lower/higher threshold of activation based on the Lck Tyr^{192} phosphorylation status. Alternatively, knowing the site which regulates Lck activity, it may be possible to design small molecular drugs which could block or activate it (Merk et al., 2018). Nonetheless, Lck Tyr^{192} regulation of Lck activity requires much deeper understanding before being applied for clinical use.

The phenotypes of TSAd Y290F and Y290E mutants have not been explored sufficiently as of yet. There are a number of known TSAd-related signaling abnormalities that should be tested with our mutants. We have not explored mutants' interaction with Itk (Berge et al., 2010, Andersen et al., 2019), nor their influence on Itk phosphorylation (Berge et al., 2010). The interaction of TSAd mutants and Lck needs further studies (Marti et al., 2006, Granum et al., 2008), as well as TSAd mutants' influence on Lck phosphorylation (Granum et al., 2014). We have started testing how TSAd mutants can affect conjugate formation (Granum et al., 2014, Abrahamsen et al., 2018), however, the experiments need to be repeated for a proper evaluation.

In addition, there are more substantial questions which should be answered. We need to prove that TSAd pTyr^{290} affects protein conformation. For this purpose, we could design a FRET probe that would monitor TSAd conformational changes (Ohashi et al., 2007, LeBlanc et al., 2018) or perform structural studies of TSAd mutated proteins (Martin et al., 2016). Both methods can be challenging, especially due to the unstructured nature of TSAd. Moreover, although the knowledge about TSAd is steadily growing, there is still no consensus on what the actual role of TSAd is in T cells. The mouse TSAd KO strain

is viable and has a very mild phenotype. Therefore, there is a need for wide screenings in the context of TSAd's role in T cells. It is even possible, that TSAd has more important functions downstream of other receptors than the TCR. Our mutated cell lines should be of great use for any kind of *in vitro* screening setup. In particular, due to the importance of Tyr^{290} for TSAd function.

References

ABE, Y., NAGANO, M., TADA, A., ADACHI, J. & TOMONAGA, T. 2017. Deep Phosphotyrosine Proteomics by Optimization of Phosphotyrosine Enrichment and MS/MS Parameters. *J Proteome Res,* 16**,** 1077-1086.

ABRAHAM, R. T. & WEISS, A. 2004. Jurkat T cells and development of the T-cell receptor signalling paradigm. *Nat Rev Immunol,* 4**,** 301-8.

ABRAHAMSEN, G., SUNDVOLD-GJERSTAD, V., HABTAMU, M., BOGEN, B. & SPURKLAND, A. 2018. Polarity of CD4+ T cells towards the antigen presenting cell is regulated by the Lck adapter TSAd. *Sci Rep,* 8**,** 13319.

ANDERSEN, T. C. B., KRISTIANSEN, P. E., HUSZENICZA, Z., JOHANSSON, M. U., GOPALAKRISHNAN, R. P., KJELSTRUP, H., BOYKEN, S., SUNDVOLD-GJERSTAD, V., GRANUM, S., SORLIE, M., BACKE, P. H., FULTON, D. B., KARLSSON, B. G., ANDREOTTI, A. H. & SPURKLAND, A. 2019. The SH3 domains of the protein kinases ITK and LCK compete for adjacent sites on T cell-specific adapter protein. *J Biol Chem***,** jbc. RA119. 008318.

ANDERSON, M. S., VENANZI, E. S., KLEIN, L., CHEN, Z., BERZINS, S. P., TURLEY, S. J., VON BOEHMER, H., BRONSON, R., DIERICH, A., BENOIST, C. & MATHIS, D. 2002. Projection of an immunological self shadow within the thymus by the aire protein. *Science,* 298**,** 1395-401.

ANDREOTTI, A. H., SCHWARTZBERG, P. L., JOSEPH, R. E. & BERG, L. J. 2010. T-cell signaling regulated by the Tec family kinase, Itk. *Cold Spring Harb Perspect Biol,* 2**,** a002287.

ANTHIS, N. J., HALING, J. R., OXLEY, C. L., MEMO, M., WEGENER, K. L., LIM, C. J., GINSBERG, M. H. & CAMPBELL, I. D. 2009. β integrin tyrosine phosphorylation is a conserved mechanism for regulating talin-induced integrin activation. *Journal of Biological Chemistry,* 284**,** 36700-36710.

ARPAIA, E., SHAHAR, M., DADI, H., COHEN, A. & ROIFMAN, C. M. 1994. Defective T cell receptor signaling and CD8+ thymic selection in humans lacking zap-70 kinase. *Cell,* 76**,** 947-58.

BAGHIROVA, S., HUGHES, B. G., HENDZEL, M. J. & SCHULZ, R. 2015. Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells. *MethodsX,* 2**,** 440-5.

BAH, A., VERNON, R. M., SIDDIQUI, Z., KRZEMINSKI, M., MUHANDIRAM, R., ZHAO, C., SONENBERG, N., KAY, L. E. & FORMAN-KAY, J. D. 2015. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature,* 519**,** 106-9.

BALLEK, O., VALECKA, J., MANNING, J. & FILIPP, D. 2015. The pool of preactivated Lck in the initiation of T-cell signaling: a critical re-evaluation of the Lck standby model. *Immunol Cell Biol,* 93**,** 384-95.

BANJADE, S., WU, Q., MITTAL, A., PEEPLES, W. B., PAPPU, R. V. & ROSEN, M. K. 2015. Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc Natl Acad Sci U S A,* 112**,** E6426-35.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BARREIRA, M., RODRIGUEZ-FDEZ, S. & BUSTELO, X. R. 2018. New insights into the Vav1 activation cycle in lymphocytes. *Cell Signal,* 45**,** 132-144.

BASILE, W., SALVATORE, M., BASSOT, C. & ELOFSSON, A. 2019. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS computational biology,* 15**,** e1007186.

BERGE, T., GRONNINGSAETER, I. H., LORVIK, K. B., ABRAHAMSEN, G., GRANUM, S., SUNDVOLD-GJERSTAD, V., CORTHAY, A., BOGEN, B. & SPURKLAND, A. 2012. SH2D2A modulates T cell mediated protection to a B cell derived tumor in transgenic mice. *PLoS One,* 7**,** e48239.

BERGE, T., SUNDVOLD-GJERSTAD, V., GRANUM, S., ANDERSEN, T. C., HOLTHE, G. B., CLAESSON-WELSH, L., ANDREOTTI, A. H., INNGJERDINGEN, M. & SPURKLAND, A. 2010. T cell specific adapter protein (TSAd) interacts with Tec kinase ITK to promote CXCL12 induced migration of human and murine T cells. *PLoS One,* 5**,** e9761.

BOGGON, T. J. & ECK, M. J. 2004. Structure and regulation of Src family kinases. *Oncogene,* 23**,** 7918-27.

BONIFACINO, J. S., GERSHLICK, D. C. & DELL'ANGELICA, E. C. 2016. Immunoprecipitation. *Curr Protoc Cell Biol,* 71.

BRDICKOVA, N., BRDICKA, T., ANGELISOVA, P., HORVATH, O., SPICKA, J., HILGERT, I., PACES, J., SIMEONI, L., KLICHE, S., MERTEN, C., SCHRAVEN, B. & HOREJSI, V. 2003. LIME: a new membrane Raft-associated adaptor protein involved in CD4 and CD8 coreceptor signaling. *J Exp Med,* 198**,** 1453-62.

BROUNS, S. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J., SNIJDERS, A. P., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science,* 321**,** 960-4.

BROWN, C. J., TAKAYAMA, S., CAMPEN, A. M., VISE, P., MARSHALL, T. W., OLDFIELD, C. J., WILLIAMS, C. J. & DUNKER, A. K. 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol,* 55**,** 104-10.

BULJAN, M., CHALANCON, G., DUNKER, A. K., BATEMAN, A., BALAJI, S., FUXREITER, M. & BABU, M. M. 2013. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr Opin Struct Biol,* 23**,** 443-50.

BURKHARDT, A. L., STEALEY, B., ROWLEY, R. B., MAHAJAN, S., PRENDERGAST, M., FARGNOLI, J. & BOLEN, J. B. 1994. Temporal regulation of non-transmembrane protein tyrosine kinase enzyme activity following T cell antigen receptor engagement. *J Biol Chem,* 269**,** 23642-7.

CEROTTINI, J. C., NORDIN, A. A. & BRUNNER, K. T. 1970. Specific in vitro cytotoxicity of thymus-derived lymphocytes sensitized to alloantigens. *Nature,* 228**,** 1308-9.

CHAKRABORTY, A. K. & WEISS, A. 2014. Insights into the initiation of TCR signaling. *Nature immunology,* 15**,** 798.

CHATILA, T., SILVERMAN, L., MILLER, R. & GEHA, R. 1989. Mechanisms of T cell activation by the calcium ionophore ionomycin. *J Immunol,* 143**,** 1283-9.

CHEN, L. & FLIES, D. B. 2013. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol,* 13**,** 227-42.

CHEN, Z., TROTMAN, L. C., SHAFFER, D., LIN, H. K., DOTAN, Z. A., NIKI, M., KOUTCHER, J. A., SCHER, H. I., LUDWIG, T., GERALD, W., CORDON-CARDO, C. & PANDOLFI, P. P. 2005. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature,* 436**,** 725-30.

CHIANG, G. G. & SEFTON, B. M. 2001. Specific dephosphorylation of the Lck tyrosine protein kinase at Tyr-394 by the SHP-1 protein-tyrosine phosphatase. *J Biol Chem,* 276**,** 23173-8.

CHOI, Y. B., KIM, C. K. & YUN, Y. 1999. Lad, an adapter protein interacting with the SH2 domain of p56lck, is required for T cell activation. *The Journal of Immunology,* 163**,** 5242-5249.

CHUAI, G. H., WANG, Q. L. & LIU, Q. 2017. In Silico Meets In Vivo: Towards Computational CRISPR-Based sgRNA Design. *Trends Biotechnol,* 35**,** 12-21.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

COURTNEY, A. H., AMACHER, J. F., KADLECEK, T. A., MOLLENAUER, M. N., AU-YEUNG, B. B., KURIYAN, J. & WEISS, A. 2017. A Phosphosite within the SH2 Domain of Lck Regulates Its Activation by CD45. *Mol Cell,* 67**,** 498-511 e6.

COUTURE, C., DECKERT, M., WILLIAMS, S., RUSSO, F. O., ALTMAN, A. & MUSTELIN, T. 1996a. Identification of the site in the Syk protein tyrosine kinase that binds the SH2 domain of Lck. *J Biol Chem,* 271**,** 24294-9.

COUTURE, C., SONGYANG, Z., JASCUR, T., WILLIAMS, S., TAILOR, P., CANTLEY, L. C. & MUSTELIN, T. 1996b. Regulation of the Lck SH2 domain by tyrosine phosphorylation. *J Biol Chem,* 271**,** 24880-4.

CROTTY, S. 2015. A brief history of T cell help to B cells. *Nat Rev Immunol,* 15**,** 185-9.

DAI, K. Z., HARBO, H. F., CELIUS, E. G., OTURAI, A., SORENSEN, P. S., RYDER, L. P., DATTA, P., SVEJGAARD, A., HILLERT, J., FREDRIKSON, S., SANDBERG-WOLLHEIM, M., LAAKSONEN, M., MYHR, K. M., NYLAND, H., VARTDAL, F. & SPURKLAND, A. 2001. The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis. *Genes Immun,* 2**,** 263-8.

DAI, K. Z., JOHANSEN, F. E., KOLLTVEIT, K. M., AASHEIM, H. C., DEMBIC, Z., VARTDAL, F. & SPURKLAND, A. 2004. Transcriptional activation of the SH2D2A gene is dependent on a cyclic adenosine 5'-monophosphate-responsive element in the proximal SH2D2A promoter. *J Immunol,* 172**,** 6144-51.

DALGLEISH, A. G., BEVERLEY, P. C., CLAPHAM, P. R., CRAWFORD, D. H., GREAVES, M. F. & WEISS, R. A. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature,* 312**,** 763-7.

DATLINGER, P., RENDEIRO, A. F., SCHMIDL, C., KRAUSGRUBER, T., TRAXLER, P., KLUGHAMMER, J., SCHUSTER, L. C., KUCHLER, A., ALPAR, D. & BOCK, C. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods,* 14**,** 297-301.

DI BARTOLO, V., MEGE, D., GERMAIN, V., PELOSI, M., DUFOUR, E., MICHEL, F., MAGISTRELLI, G., ISACCHI, A. & ACUTO, O. 1999. Tyrosine 319, a newly identified phosphorylation site of ZAP-70, plays a critical role in T cell antigen receptor signaling. *J Biol Chem,* 274**,** 6285-94.

DITLEV, J. A., VEGA, A. R., KOSTER, D. V., SU, X., TANI, T., LAKODUK, A. M., VALE, R. D., MAYOR, S., JAQAMAN, K. & ROSEN, M. K. 2019. A composition-dependent molecular clutch between T cell signaling condensates and actin. *Elife,* 8.

DRAPPA, J., KAMEN, L. A., CHAN, E., GEORGIEV, M., ASHANY, D., MARTI, F. & KING, P. D. 2003. Impaired T cell death and lupus-like autoimmunity in T cell-specific adapter protein-deficient mice. *J Exp Med,* 198**,** 809-21.

DUNKER, A. K., CORTESE, M. S., ROMERO, P., IAKOUCHEVA, L. M. & UVERSKY, V. N. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J,* 272**,** 5129-48.

DUNKER, A. K., LAWSON, J. D., BROWN, C. J., WILLIAMS, R. M., ROMERO, P., OH, J. S., OLDFIELD, C. J., CAMPEN, A. M., RATLIFF, C. M., HIPPS, K. W., AUSIO, J., NISSEN, M. S., REEVES, R., KANG, C., KISSINGER, C. R., BAILEY, R. W., GRISWOLD, M. D., CHIU, W., GARNER, E. C. & OBRADOVIC, Z. 2001. Intrinsically disordered protein. *J Mol Graph Model,* 19**,** 26-59.

DUNKER, A. K., OBRADOVIC, Z., ROMERO, P., GARNER, E. C. & BROWN, C. J. 2000. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform,* 11**,** 161-71.

DUPLAY, P., THOME, M., HERVE, F. & ACUTO, O. 1994. p56lck interacts via its src homology 2 domain with the ZAP-70 kinase. *J Exp Med,* 179**,** 1163-72.

ECHEVERRI, C. J., BEACHY, P. A., BAUM, B., BOUTROS, M., BUCHHOLZ, F., CHANDA, S. K., DOWNWARD, J., ELLENBERG, J., FRASER, A. G., HACOHEN, N., HAHN, W. C., JACKSON, A. L., KIGER, A., LINSLEY, P. S., LUM, L., MA, Y., MATHEY-PREVOT, B., ROOT, D. E., SABATINI, D. M., TAIPALE, J., PERRIMON, N. & BERNARDS, R. 2006. Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat Methods,* 3**,** 777-9.

ECK, M. J., ATWELL, S. K., SHOELSON, S. E. & HARRISON, S. C. 1994. Structure of the regulatory domains of the Src-family tyrosine kinase Lck. *Nature,* 368**,** 764-9.

ERPEL, T., SUPERTI-FURGA, G. & COURTNEIDGE, S. A. 1995. Mutational analysis of the Src SH3 domain: the same residues of the ligand binding surface are important for intra- and intermolecular interactions. *EMBO J,* 14**,** 963-75.

FAGRAEUS, A. 1948. Antibody Production in relation to the Development of Plasma Cells. In vivo and in vitro Experiments. *Acta Medica Scandinavica,* 130.

FARGNOLI, J., BURKHARDT, A. L., LAVERTY, M., KUT, S. A., VAN OERS, N. S., WEISS, A. & BOLEN, J. B. 1995. Syk mutation in Jurkat E6-derived clones results in lack of p72syk expression. *J Biol Chem,* 270**,** 26533-7.

FIERING, S., NORTHROP, J. P., NOLAN, G. P., MATTILA, P. S., CRABTREE, G. R. & HERZENBERG, L. A. 1990. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev,* 4**,** 1823-34.

FLAJNIK, M. F. & KASAHARA, M. 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet,* 11**,** 47-59.

FREEBURN, R. W., WRIGHT, K. L., BURGESS, S. J., ASTOUL, E., CANTRELL, D. A. & WARD, S. G. 2002. Evidence that SHIP-1 contributes to phosphatidylinositol 3,4,5-trisphosphate metabolism in T lymphocytes and can regulate novel phosphoinositide 3-kinase effectors. *J Immunol,* 169**,** 5441-50.

FU, Y., FODEN, J. A., KHAYTER, C., MAEDER, M. L., REYON, D., JOUNG, J. K. & SANDER, J. D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol,* 31**,** 822-6.

FUHRMANN, F., LISCHKE, T., GROSS, F., SCHEEL, T., BAUER, L., KALIM, K. W., RADBRUCH, A., HERZEL, H., HUTLOFF, A. & BAUMGRASS, R. 2016. Adequate immune response ensured by binary IL-2 and graded CD25 expression in a murine transfer model. *Elife,* 5**,** e20616.

FUKUSHIMA, A., HATANAKA, Y., CHANG, J. W., TAKAMATSU, M., SINGH, N. & IWASHIMA, M. 2006. Lck couples Shc to TCR signaling. *Cell Signal,* 18**,** 1182-9.

GERVAIS, F., CHOW, L., LEE, J., BRANTON, P. & VEILLETTE, A. 1993. The SH2 domain is required for stable phosphorylation of p56lck at tyrosine 505, the negative regulatory site. *Molecular and cellular biology,* 13**,** 7112-7121.

GHOSH, S. & BALTIMORE, D. 1990. Activation in vitro of NF-κB" by phosphorylation of its inhibitor IκB". *Nature,* 344**,** 678.

GILLIS, S. & WATSON, J. 1980. Biochemical and biological characterization of lymphocyte regulatory molecules. V. Identification of an interleukin 2-producing human leukemia T cell line. *J Exp Med,* 152**,** 1709-19.

GIOIA, L., SIDDIQUE, A., HEAD, S. R., SALOMON, D. R. & SU, A. I. 2018. A genome-wide survey of mutations in the Jurkat cell line. *BMC Genomics,* 19**,** 334.

GONFLONI, S., WILLIAMS, J. C., HATTULA, K., WEIJLAND, A., WIERENGA, R. K. & SUPERTI-FURGA, G. 1997. The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *EMBO J,* 16**,** 7261-71.

GRANUM, S., ANDERSEN, T. C., SORLIE, M., JORGENSEN, M., KOLL, L., BERGE, T., LEA, T., FLECKENSTEIN, B., SPURKLAND, A. & SUNDVOLD-GJERSTAD, V. 2008. Modulation of Lck function through multisite docking to T cell-specific adapter protein. *J Biol Chem,* 283**,** 21909-19.

GRANUM, S., SUNDVOLD-GJERSTAD, V., DAI, K. Z., KOLLTVEIT, K. M., HILDEBRAND, K., HUITFELDT, H. S., LEA, T. & SPURKLAND, A. 2006. Structure function analysis of SH2D2A isoforms expressed in T cells reveals a crucial role for the proline rich region encoded by SH2D2A exon 7. *BMC Immunol,* 7**,** 15.

GRANUM, S., SUNDVOLD-GJERSTAD, V., GOPALAKRISHNAN, R. P., BERGE, T., KOLL, L., ABRAHAMSEN, G., SORLIE, M. & SPURKLAND, A. 2014. The kinase Itk and the adaptor TSAd change the specificity of the kinase Lck in T cells by promoting the phosphorylation of Tyr192. *Sci Signal,* 7**,** ra118.

GREGOIRE, C., SIMOVA, S., WANG, Y., SANSONI, A., RICHELME, S., SCHMIDT-GIESE, A., SIMEONI, L., ANGELISOVA, P., REINHOLD, D., SCHRAVEN, B., HOREJSI, V., MALISSEN, B. & MALISSEN, M. 2007. Deletion of the LIME adaptor protein minimally affects T and B cell development and function. *Eur J Immunol,* 37**,** 3259-69.

GULBINS, E., COGGESHALL, K. M., BAIER, G., KATZAV, S., BURN, P. & ALTMAN, A. 1993. Tyrosine kinase-stimulated guanine nucleotide exchange activity of Vav in T cell activation. *Science,* 260**,** 822-5.

HASEGAWA, K., MARTIN, F., HUANG, G., TUMAS, D., DIEHL, L. & CHAN, A. C. 2004. PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells. *Science,* 303**,** 685-9.

HEM, C. D., SUNDVOLD-GJERSTAD, V., GRANUM, S., KOLL, L., ABRAHAMSEN, G., BUDAY, L. & SPURKLAND, A. 2015. T cell specific adaptor protein (TSAd) promotes interaction of Nck with Lck and SLP-76 in T cells. *Cell Commun Signal,* 13**,** 31.

HS LU, C., LIU, K., TAN, L. P. & YAO, S. Q. 2012. Current chemical biology tools for studying protein phosphorylation and dephosphorylation. *Chemistry–A European Journal,* 18**,** 28-39.

HUR, E. M., SON, M., LEE, O. H., CHOI, Y. B., PARK, C., LEE, H. & YUN, Y. 2003. LIME, a novel transmembrane adaptor protein, associates with p56lck and mediates T cell activation. *J Exp Med,* 198**,** 1463-73.

IAKOUCHEVA, L. M., RADIVOJAC, P., BROWN, C. J., O'CONNOR, T. R., SIKES, J. G., OBRADOVIC, Z. & DUNKER, A. K. 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res,* 32**,** 1037-49.

IVARSSON, Y. & JEMTH, P. 2019. Affinity and specificity of motif-based protein-protein interactions. *Curr Opin Struct Biol,* 54**,** 26-33.

KAIZUKA, Y., DOUGLASS, A. D., VARMA, R., DUSTIN, M. L. & VALE, R. D. 2007. Mechanisms for segregating T cell receptor and adhesion molecules during immunological synapse formation in Jurkat T cells. *Proc Natl Acad Sci U S A,* 104**,** 20296-301.

KANEKO, T., HUANG, H., ZHAO, B., LI, L., LIU, H., VOSS, C. K., WU, C., SCHILLER, M. R. & LI, S. S. 2010. Loops govern SH2 domain specificity by controlling access to binding pockets. *Sci Signal,* 3**,** ra34.

KAPLUN, L., FRIDMAN, A. L., CHEN, W., LEVIN, N. K., AHSAN, S., PETRUCELLI, N., BARRICK, J. L., GOLD, R., LAND, S., SIMON, M. S., MORRIS, R. T., MUNKARAH, A. R. & TAINSKY, M. A. 2012. Variants in the Signaling Protein TSAd are Associated with Susceptibility to Ovarian Cancer in BRCA1/2 Negative High Risk Families. *Biomark Insights,* 7**,** 151-7.

KAPOOR-KAUSHIK, N., HINDE, E., COMPEER, E. B., YAMAMOTO, Y., KRAUS, F., YANG, Z., LOU, J., PAGEON, S. V., TABARIN, T., GAUS, K. & ROSSY, J. 2016. Distinct Mechanisms Regulate Lck Spatial Organization in Activated T Cells. *Front Immunol,* 7**,** 83.

KARLSSON, H., SVENSSON, E., GIGG, C., JARVIUS, M., OLSSON-STROMBERG, U., SAVOLDO, B., DOTTI, G. & LOSKOG, A. 2015. Evaluation of Intracellular Signaling Downstream Chimeric Antigen Receptors. *PLoS One,* 10**,** e0144787.

KHAN, S. H., MCLAUGHLIN, W. A. & KUMAR, R. 2017. Site-specific phosphorylation regulates the structure and function of an intrinsically disordered domain of the glucocorticoid receptor. *Sci Rep,* 7**,** 15440.

KLEIN, L., HINTERBERGER, M., WIRNSBERGER, G. & KYEWSKI, B. 2009. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol,* 9**,** 833-44.

KOLLTVEIT, K. M., GRANUM, S., AASHEIM, H. C., FORSBRING, M., SUNDVOLD-GJERSTAD, V., DAI, K. Z., MOLBERG, O., SCHJETNE, K. W., BOGEN, B., SHAPIRO, V. S., JOHANSEN, F. E., SCHENCK, K. & SPURKLAND, A. 2008. Expression of SH2D2A in T-cells is regulated both at the transcriptional and translational level. *Mol Immunol,* 45**,** 2380-90.

KOLLTVEIT, K. M., SCHREURS, O., OSTREM, J., SOLAND, T. M., KHUU, C., BERGE, T., MESSELT, E., HAYASHI, K., GRANUM, S., SPURKLAND, A. & SCHENCK, K. 2010. Expression of the T-cell-specific adapter protein in oral epithelium. *Eur J Oral Sci,* 118**,** 159-67.

KOMOR, A. C., KIM, Y. B., PACKER, M. S., ZURIS, J. A. & LIU, D. R. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature,* 533**,** 420-4.

KULKARNI, P., JOLLY, M. K., JIA, D., MOONEY, S. M., BHARGAVA, A., KAGOHARA, L. T., CHEN, Y., HAO, P., HE, Y., VELTRI, R. W., GRISHAEV, A., WENINGER, K., LEVINE, H. & ORBAN, J. 2017. Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity. *Proc Natl Acad Sci U S A,* 114**,** E2644-E2653.

LANZAVECCHIA, A. 1985. Antigen-specific interaction between T and B cells. *Nature,* 314**,** 537-9.

LEBLANC, S. J., KULKARNI, P. & WENINGER, K. R. 2018. Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules,* 8**,** 140.

LEE-FRUMAN, K. K., COLLINS, T. L. & BURAKOFF, S. J. 1996. Role of the Lck Src homology 2 and 3 domains in protein tyrosine phosphorylation. *J Biol Chem,* 271**,** 25003-10.

LEWIS, L. A., CHUNG, C. D., CHEN, J. Y., PARNES, J. R., MORAN, M., PATEL, V. P. & MICELI, M. C. 1997. The Lck SH2 phosphotyrosine binding site is critical for efficient TCR-induced processive tyrosine phosphorylation of the zeta-chain and IL-2 production. *Journal of Immunology,* 159**,** 2292-2300.

LI, L., WU, C., HUANG, H., ZHANG, K., GAN, J. & LI, S. S. 2008a. Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res,* 36**,** 3263-73.

LI, T., LI, F. & ZHANG, X. 2008b. Prediction of kinase-specific phosphorylation sites with sequence features by a log-odds ratio approach. *Proteins,* 70**,** 404-14.

LI, X., PADHAN, N., SJOSTROM, E. O., ROCHE, F. P., TESTINI, C., HONKURA, N., SAINZ-JASPEADO, M., GORDON, E., BENTLEY, K., PHILIPPIDES, A., TOLMACHEV, V., DEJANA, E., STAN, R. V., VESTWEBER, D., BALLMER-HOFER, K., BETSHOLTZ, C., PIETRAS, K., JANSSON, L. & CLAESSON-WELSH, L. 2016. VEGFR2 pY949 signalling regulates adherens junction integrity and metastatic spread. *Nat Commun,* 7**,** 11017.

LI, Y., HE, X., SCHEMBRI-KING, J., JAKES, S. & HAYASHI, J. 2000. Cloning and characterization of human Lnk, an adaptor protein with pleckstrin homology and Src homology 2 domains that can inhibit T cell activation. *J Immunol,* 164**,** 5199-206.

LIM, W. A. & PAWSON, T. 2010. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell,* 142**,** 661-7.

LIN, Y., CRADICK, T. J., BROWN, M. T., DESHMUKH, H., RANJAN, P., SARODE, N., WILE, B. M., VERTINO, P. M., STEWART, F. J. & BAO, G. 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic acids research,* 42**,** 7473-7485.

LIND, S. B., ARTEMENKO, K. A. & PETTERSSON, U. 2012. A strategy for identification of protein tyrosine phosphorylation. *Methods,* 56**,** 275-83.

LINDSTEN, T., LEE, K. P., HARRIS, E. S., PETRYNIAK, B., CRAIGHEAD, N., REYNOLDS, P. J., LOMBARD, D. B., FREEMAN, G. J., NADLER, L. M., GRAY, G. S. & ET AL. 1993. Characterization of CTLA-4 structure and expression on human T cells. *J Immunol,* 151**,** 3489-99.

LIU, H., HUANG, H., VOSS, C., KANEKO, T., QIN, W. T., SIDHU, S. & LI, S. S. 2019. Surface Loops in a Single SH2 Domain Are Capable of Encoding the Spectrum of Specificity of the SH2 Family. *Mol Cell Proteomics,* 18**,** 372-382.

LIU, K. I., RAMLI, M. N., WOO, C. W., WANG, Y., ZHAO, T., ZHANG, X., YIM, G. R., CHONG, B. Y., GOWHER, A., CHUA, M. Z., JUNG, J., LEE, J. H. & TAN, M. H. 2016. A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat Chem Biol,* 12**,** 980-987.

LORENTZEN, A. R., SMESTAD, C., LIE, B. A., OTURAI, A. B., AKESSON, E., SAARELA, J., MYHR, K. M., VARTDAL, F., CELIUS, E. G., SORENSEN, P. S., HILLERT, J., SPURKLAND, A. & HARBO, H. F. 2008. The SH2D2A gene and susceptibility to multiple sclerosis. *J Neuroimmunol,* 197**,** 152-8.

MA, Z. & FINKEL, T. H. 2010. T cell receptor triggering by force. *Trends Immunol,* 31**,** 1-6.

MACIEJEWSKI, P. M., PETERSON, F. C., ANDERSON, P. J. & BROOKS, C. L. 1995. Mutation of serine 90 to glutamic acid mimics phosphorylation of bovine prolactin. *J Biol Chem,* 270**,** 27661-5.

MANGER, B., WEISS, A., WEYAND, C., GORONZY, J. & STOBO, J. D. 1985. T cell activation: differences in the signals required for IL 2 production by nonactivated and activated T cells. *J Immunol,* 135**,** 3669-73.

MARTI, F., GARCIA, G. G., LAPINSKI, P. E., MACGREGOR, J. N. & KING, P. D. 2006. Essential role of the T cell-specific adapter protein in the activation of LCK in peripheral T cells. *J Exp Med,* 203**,** 281-7.

MARTIN, E. W., HOLEHOUSE, A. S., GRACE, C. R., HUGHES, A., PAPPU, R. V. & MITTAG, T. 2016. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J Am Chem Soc,* 138**,** 15323-15335.

MATSUMOTO, T., BOHMAN, S., DIXELIUS, J., BERGE, T., DIMBERG, A., MAGNUSSON, P., WANG, L., WIKNER, C., QI, J. H., WERNSTEDT, C., WU, J., BRUHEIM, S., MUGISHIMA, H., MUKHOPADHYAY, D., SPURKLAND, A. & CLAESSON-WELSH, L. 2005. VEGF receptor-2 Y951 signaling and a role for the adapter molecule TSAd in tumor angiogenesis. *EMBO J,* 24**,** 2342-53.

MATZINGER, P. 1994. Tolerance, danger, and the extended family. *Annu Rev Immunol,* 12**,** 991-1045.

MAYYA, V. & DUSTIN, M. L. 2016. What scales the T cell response? *Trends in immunology,* 37**,** 513-522.

MCNEILL, L., SALMOND, R. J., COOPER, J. C., CARRET, C. K., CASSADY-CAIN, R. L., ROCHE-MOLINA, M., TANDON, P., HOLMES, N. & ALEXANDER, D. R. 2007. The differential regulation of Lck kinase phosphorylation sites by CD45 is critical for T cell receptor signaling responses. *Immunity,* 27**,** 425-37.

MECHTA, F., LALLEMAND, D., PFARR, C. M. & YANIV, M. 1997. Transformation by ras modifies AP1 composition and activity. *Oncogene,* 14**,** 837-47.

MELLACHERUVU, D., WRIGHT, Z., COUZENS, A. L., LAMBERT, J. P., ST-DENIS, N. A., LI, T., MITEVA, Y. V., HAURI, S., SARDIU, M. E., LOW, T. Y., HALIM, V. A., BAGSHAW, R. D., HUBNER, N. C., AL-HAKIM, A., BOUCHARD, A., FAUBERT, D., FERMIN, D., DUNHAM, W. H., GOUDREAULT, M., LIN, Z. Y., BADILLO, B. G., PAWSON, T., DUROCHER, D., COULOMBE, B., AEBERSOLD, R., SUPERTI-FURGA, G., COLINGE, J., HECK, A. J., CHOI, H., GSTAIGER, M., MOHAMMED, S., CRISTEA, I. M., BENNETT, K. L., WASHBURN, M. P., RAUGHT, B., EWING, R. M., GINGRAS, A. C. & NESVIZHSKII, A. I. 2013. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods,* 10**,** 730-6.

MERK, D., FRIEDRICH, L., GRISONI, F. & SCHNEIDER, G. 2018. De Novo Design of Bioactive Small Molecules by Artificial Intelligence. *Mol Inform,* 37**,** 1700153.

MERŐ, B., RADNAI, L., GÓGL, G., TŐKE, O., LEVELES, I., KOPRIVANACZ, K., SZEDER, B., DÜLK, M., KUDLIK, G. & VAS, V. 2019. Structural insights into the tyrosine phosphorylation–mediated inhibition of SH3 domain–ligand interactions. *Journal of Biological Chemistry,* 294**,** 4608-4620.

MICHEL, F., GRIMAUD, L., TUOSTO, L. & ACUTO, O. 1998. Fyn and ZAP-70 are required for Vav phosphorylation in T cells stimulated by antigen-presenting cells. *J Biol Chem,* 273**,** 31932-8.

MILLER, A. T., WILCOX, H. M., LAI, Z. & BERG, L. J. 2004. Signaling through Itk promotes T helper 2 differentiation via negative regulation of T-bet. *Immunity,* 21**,** 67-80.

MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & ALMENDROS, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology,* 155**,** 733-40.

MOLINA, T. J., KISHIHARA, K., SIDEROVSKI, D. P., VAN EWIJK, W., NARENDRAN, A., TIMMS, E., WAKEHAM, A., PAIGE, C. J., HARTMANN, K. U., VEILLETTE, A. & ET AL. 1992. Profound block in thymocyte development in mice lacking p56lck. *Nature,* 357**,** 161-4.

MOORE, J. K. & HABER, J. E. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol Cell Biol,* 16**,** 2164-73.

MOUSSA, P., ABRAHAMSEN, G., FODIL, N., GOPALAKRISHNAN, R. P., MANCINI, M., DISSEN, E., SAETHER, P. C., WILTSHIRE, S. A., BOIVIN, G. A., CAIGNARD, G., SPURKLAND, A. & VIDAL, S. M. 2015. Reduced MCMV Deltam157 viral clearance in the absence of TSAd. *Sci Rep,* 5**,** 9219.

NEGISHI, I., MOTOYAMA, N., NAKAYAMA, K., NAKAYAMA, K., SENJU, S., HATAKEYAMA, S., ZHANG, Q., CHAN, A. C. & LOH, D. Y. 1995. Essential role for ZAP-70 in both positive and negative selection of thymocytes. *Nature,* 376**,** 435-8.

NEJAD, S., BRYCESON, Y., DISSEN, E., SUNDVOLD-GJERSTAD, V., NAPER, C., ROLSTAD, B., SPURKLAND, A. & VAAGE, J. T. 2004. cDNA cloning of a rat orthologue of SH2D2A encoding T-cell-specific adaptor protein (TSAd): expression in T and NK cells. *Immunogenetics,* 56**,** 338-42.

NG, D. H., WATTS, J. D., AEBERSOLD, R. & JOHNSON, P. 1996. Demonstration of a direct interaction between p56lck and the cytoplasmic domain of CD45 in vitro. *J Biol Chem,* 271**,** 1295-300.

NIKA, K., SOLDANI, C., SALEK, M., PASTER, W., GRAY, A., ETZENSPERGER, R., FUGGER, L., POLZELLA, P., CERUNDOLO, V., DUSHEK, O., HOFER, T., VIOLA, A. & ACUTO, O. 2010. Constitutively active Lck kinase in T cells drives antigen receptor signal transduction. *Immunity,* 32**,** 766-77.

NIKA, K., TAUTZ, L., ARIMURA, Y., VANG, T., WILLIAMS, S. & MUSTELIN, T. 2007. A weak Lck tail bite is necessary for Lck function in T cell antigen receptor signaling. *J Biol Chem,* 282**,** 36000-9.

NORTHROP, J. P., ULLMAN, K. S. & CRABTREE, G. R. 1993. Characterization of the nuclear and cytoplasmic components of the lymphoid-specific nuclear factor of activated T cells (NF-AT) complex. *J Biol Chem,* 268**,** 2917-23.

NOTTURNO, F., PACE, M., DE ANGELIS, M. V., CAPORALE, C. M., GIOVANNINI, A. & UNCINI, A. 2008. Susceptibility to chronic inflammatory demyelinating polyradiculoneuropathy is associated to polymorphic GA repeat in the SH2D2A gene. *J Neuroimmunol,* 197**,** 124-7.

OHASHI, T., GALIACY, S. D., BRISCOE, G. & ERICKSON, H. P. 2007. An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins. *Protein Sci,* 16**,** 1429-38.

PAGEON, S. V., TABARIN, T., YAMAMOTO, Y., MA, Y., NICOVICH, P. R., BRIDGEMAN, J. S., COHNEN, A., BENZING, C., GAO, Y., CROWTHER, M. D., TUNGATT, K., DOLTON, G., SEWELL, A. K., PRICE, D. A., ACUTO, O., PARTON, R. G., GOODING, J. J., ROSSY, J., ROSSJOHN, J. & GAUS, K. 2016. Functional role of T-cell receptor nanoclusters in signal initiation and antigen discrimination. *Proc Natl Acad Sci U S A,* 113**,** E5454-63.

PAK, C. W., KOSNO, M., HOLEHOUSE, A. S., PADRICK, S. B., MITTAL, A., ALI, R., YUNUS, A. A., LIU, D. R., PAPPU, R. V. & ROSEN, M. K. 2016. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell,* 63**,** 72-85.

PALACIOS, E. H. & WEISS, A. 2004. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. *Oncogene,* 23**,** 7990-8000.

PARK, D., CHOI, Y. B., HAN, M. K., KIM, U. H., SHIN, J. & YUN, Y. 2001. Adaptor protein Lad relays PDGF signal to Grb2 in lung cells: a tissue-specific PDGF signal transduction. *Biochem Biophys Res Commun,* 284**,** 275-81.

PARK, D., PARK, I., LEE, D., CHOI, Y. B., LEE, H. & YUN, Y. 2007. The adaptor protein Lad associates with the G protein beta subunit and mediates chemokine-dependent T-cell migration. *Blood,* 109**,** 5122-8.

PARK, E., CHOI, Y., AHN, E., PARK, I. & YUN, Y. 2009. The adaptor protein LAD/TSAd mediates laminin-dependent T cell migration via association with the 67 kDa laminin binding protein. *Exp Mol Med,* 41**,** 728-36.

PEARLMAN, S. M., SERBER, Z. & FERRELL, J. E., JR. 2011. A mechanism for the evolution of phosphorylation sites. *Cell,* 147**,** 934-46.

PELOSI, M., DI BARTOLO, V., MOUNIER, V., MEGE, D., PASCUSSI, J. M., DUFOUR, E., BLONDEL, A. & ACUTO, O. 1999. Tyrosine 319 in the interdomain B of ZAP-70 is a binding site for the Src homology 2 domain of Lck. *J Biol Chem,* 274**,** 14229-37.

PERCHONOCK, C. E., FERNANDO, M. C., QUINN, W. J., 3RD, NGUYEN, C. T., SUN, J., SHAPIRO, M. J. & SHAPIRO, V. S. 2006. Negative regulation of interleukin-2 and p38 mitogen-activated protein kinase during T-cell activation by the adaptor ALX. *Mol Cell Biol,* 26**,** 6005-15.

PEREZ-VILLAR, J. J., WHITNEY, G. S., SITNICK, M. T., DUNN, R. J., VENKATESAN, S., O'DAY, K., SCHIEVEN, G. L., LIN, T. A. & KANNER, S. B. 2002. Phosphorylation of the linker for activation of T-cells by Itk promotes recruitment of Vav. *Biochemistry,* 41**,** 10732-40.

PETERS, D., TSUCHIDA, M., MANTHEI, E. R., ALAM, T., CHO, C. S., KNECHTLE, S. J. & HAMAWY, M. M. 2000. Potentiation of CD3-induced expression of the linker for activation of T cells (LAT) by the calcineurin inhibitors cyclosporin A and FK506. *Blood,* 95**,** 2733-41.

PHILIPSEN, L., REDDYCHERLA, A. V., HARTIG, R., GUMZ, J., KASTLE, M., KRITIKOS, A., POLTORAK, M. P., PROKAZOV, Y., TURBIN, E., WEBER, A., ZUSCHRATTER, W., SCHRAVEN, B., SIMEONI, L. & MULLER, A. J. 2017. De novo phosphorylation and conformational opening of the tyrosine kinase Lck act in concert to initiate T cell receptor signaling. *Sci Signal,* 10**,** eaaf4736.

QIAN, X., LI, G., VASS, W. C., PAPAGEORGE, A., WALKER, R. C., ASNAGHI, L., STEINBACH, P. J., TOSATO, G., HUNTER, K. & LOWY, D. R. 2009. The Tensin-3 protein, including its SH2 domain, is phosphorylated by Src and contributes to tumorigenesis and metastasis. *Cancer Cell,* 16**,** 246-58.

RAJAGOPAL, K., SOMMERS, C. L., DECKER, D. C., MITCHELL, E. O., KORTHAUER, U., SPERLING, A. I., KOZAK, C. A., LOVE, P. E. & BLUESTONE, J. A. 1999. RIBP, a novel Rlk/Txk-and itk-binding adaptor protein that regulates T cell activation. *Journal of Experimental Medicine,* 190**,** 1657-1668.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013b. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.

READ, A. P. 2009. Stewart Scherer: A short guide to the human genome. Springer.

REN, S., UVERSKY, V. N., CHEN, Z., DUNKER, A. K. & OBRADOVIC, Z. 2008. Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. *BMC Genomics,* 9 Suppl 2**,** S26.

RIBEIRO, L. F., RIBEIRO, L. F. C., BARRETO, M. Q. & WARD, R. J. 2018. Protein Engineering Strategies to Expand CRISPR-Cas9 Applications. *Int J Genomics,* 2018**,** 1652567.

RICHARD, K. C., BERTOLESI, G. E., DUNFIELD, L. D., MCMASTER, C. R. & NACHTIGAL, M. W. 2006. TSAd interacts with Smad2 and Smad3. *Biochem Biophys Res Commun,* 347**,** 266-72.

SACHSE, S. M., LIEVENS, S., RIBEIRO, L. F., DASCENCO, D., MASSCHAELE, D., HORRE, K., MISBAER, A., VANDERROOST, N., DE SMET, A. S., SALTA, E., ERFURTH, M. L., KISE, Y., NEBEL, S., VAN DELM, W., PLAISANCE, S., TAVERNIER, J., DE STROOPER, B., DE WIT, J. & SCHMUCKER, D. 2019. Nuclear import of the DSCAM-cytoplasmic domain drives signaling capable of inhibiting synapse formation. *EMBO J,* 38.

SAKAGUCHI, S., YAMAGUCHI, T., NOMURA, T. & ONO, M. 2008. Regulatory T cells and immune tolerance. *Cell,* 133**,** 775-87.

SELA, M., BOGIN, Y., BEACH, D., OELLERICH, T., LEHNE, J., SMITH-GARVIN, J. E., OKUMURA, M., STAROSVETSKY, E., KOSOFF, R., LIBMAN, E., KORETZKY, G., KAMBAYASHI, T., URLAUB, H., WIENANDS, J., CHERNOFF, J. & YABLONSKI, D. 2011. Sequential phosphorylation of SLP-76 at tyrosine 173 is required for activation of T and mast cells. *EMBO J,* 30**,** 3160-72.

SHAH, N. H., LOBEL, M., WEISS, A. & KURIYAN, J. 2018. Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. *Elife,* 7**,** e35190.

SHAN, X., CZAR, M. J., BUNNELL, S. C., LIU, P., LIU, Y., SCHWARTZBERG, P. L. & WANGE, R. L. 2000. Deficiency of PTEN in Jurkat T cells causes constitutive localization of Itk to the plasma membrane and hyperresponsiveness to CD3 stimulation. *Mol Cell Biol,* 20**,** 6945-57.

SHAPIRO, M. J., NGUYEN, C. T., AGHAJANIAN, H., ZHANG, W. & SHAPIRO, V. S. 2008. Negative regulation of TCR signaling by linker for activation of X cells via phosphotyrosine-dependent and -independent mechanisms. *J Immunol,* 181**,** 7055-61.

SHARP, P. A. 1999. RNAi and double-strand RNA. *Genes Dev,* 13**,** 139-41.

SHAW, J. P., UTZ, P. J., DURAND, D. B., TOOLE, J. J., EMMEL, E. A. & CRABTREE, G. R. 1988. Identification of a putative regulator of early T cell activation genes. *Science,* 241**,** 202-5.

SHEN, M. W., ARBAB, M., HSU, J. Y., WORSTELL, D., CULBERTSON, S. J., KRABBE, O., CASSA, C. A., LIU, D. R., GIFFORD, D. K. & SHERWOOD, R. I. 2018. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature,* 563**,** 646-651.

SHEN, X. & TOMARAS, G. D. 2011. Alterations of the B-cell response by HIV-1 replication. *Curr HIV/AIDS Rep,* 8**,** 23-30.

SHIPKOVA, M. & WIELAND, E. 2012. Surface markers of lymphocyte activation and markers of cell proliferation. *Clin Chim Acta,* 413**,** 1338-49.

SJÖLIN-GOODFELLOW, H., FRUSHICHEVA, M. P., JI, Q., CHENG, D. A., KADLECEK, T. A., CANTOR, A. J., KURIYAN, J., CHAKRABORTY, A. K., SALOMON, A. R. & WEISS, A. 2015. The catalytic activity of the kinase ZAP-70 mediates basal signaling and negative feedback of the T cell receptor pathway. *Sci. Signal.,* 8**,** ra49-ra49.

SLAYMAKER, I. M., GAO, L., ZETSCHE, B., SCOTT, D. A., YAN, W. X. & ZHANG, F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science,* 351**,** 84-8.

SMERDEL, A., DAI, K. Z., LORENTZEN, A. R., FLATO, B., MASLINSKI, S., THORSBY, E., FORRE, O. & SPURKLAND, A. 2004. Genetic association between juvenile rheumatoid arthritis and polymorphism in the SH2D2A gene. *Genes Immun,* 5**,** 310-2.

SMIDA, M., POSEVITZ-FEJFAR, A., HOREJSI, V., SCHRAVEN, B. & LINDQUIST, J. A. 2007. A novel negative regulatory function of the phosphoprotein associated with glycosphingolipid-enriched microdomains: blocking Ras activation. *Blood,* 110**,** 596-615.

SOMMERS, C. L., SAMELSON, L. E. & LOVE, P. E. 2004. LAT: a T lymphocyte adapter protein that couples the antigen receptor to downstream signaling pathways. *Bioessays,* 26**,** 61-7.

SPURKLAND, A., BRINCHMANN, J. E., MARKUSSEN, G., PEDEUTOUR, F., MUNTHE, E., LEA, T., VARTDAL, F. & AASHEIM, H. C. 1998. Molecular cloning of a T cell-specific adapter protein (TSAd) containing an Src homology (SH) 2 domain and putative SH3 and phosphotyrosine binding sites. *J Biol Chem,* 273**,** 4539-46.

ŠTEFANOVÁ, I., HEMMER, B., VERGELLI, M., MARTIN, R., BIDDISON, W. E. & GERMAIN, R. N. 2003. TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. *Nature immunology,* 4**,** 248.

STEINMAN, R. M. 1991. The dendritic cell system and its role in immunogenicity. *Annu Rev Immunol,* 9**,** 271-96.

STIRNWEISS, A., HARTIG, R., GIESELER, S., LINDQUIST, J. A., REICHARDT, P., PHILIPSEN, L., SIMEONI, L., POLTORAK, M., MERTEN, C., ZUSCHRATTER, W., PROKAZOV, Y., PASTER, W.,

STOCKINGER, H., HARDER, T., GUNZER, M. & SCHRAVEN, B. 2013. T cell activation results in conformational changes in the Src family kinase Lck to induce its activation. *Sci Signal,* 6**,** ra13.

STOVER, D. R., FURET, P. & LYDON, N. B. 1996. Modulation of the SH2 binding specificity and kinase activity of Src by tyrosine phosphorylation within its SH2 domain. *J Biol Chem,* 271**,** 12481-7.

STRAUS, D. B., CHAN, A. C., PATAI, B. & WEISS, A. 1996. SH2 domain function is essential for the role of the Lck tyrosine kinase in T cell receptor signal transduction. *J Biol Chem,* 271**,** 9976-81.

STRAUS, D. B. & WEISS, A. 1992. Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor. *Cell,* 70**,** 585-593.

SU, X., DITLEV, J. A., HUI, E., XING, W., BANJADE, S., OKRUT, J., KING, D. S., TAUNTON, J., ROSEN, M. K. & VALE, R. D. 2016. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science,* 352**,** 595-9.

SUGASE, K., DYSON, H. J. & WRIGHT, P. E. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature,* 447**,** 1021-5.

SUN, W., KESAVAN, K., SCHAEFER, B. C., GARRINGTON, T. P., WARE, M., JOHNSON, N. L., GELFAND, E. W. & JOHNSON, G. L. 2001. MEKK2 associates with the adapter protein Lad/RIBP and regulates the MEK5-BMK1/ERK5 pathway. *J Biol Chem,* 276**,** 5093-100.

SUN, W., WEI, X., KESAVAN, K., GARRINGTON, T. P., FAN, R., MEI, J., ANDERSON, S. M., GELFAND, E. W. & JOHNSON, G. L. 2003. MEK kinase 2 and the adaptor protein Lad regulate extracellular signal-regulated kinase 5 activation by epidermal growth factor via Src. *Mol Cell Biol,* 23**,** 2298-308.

SUN, Z., LI, X., MASSENA, S., KUTSCHERA, S., PADHAN, N., GUALANDI, L., SUNDVOLD-GJERSTAD, V., GUSTAFSSON, K., CHOY, W. W., ZANG, G., QUACH, M., JANSSON, L., PHILLIPSON, M., ABID, M. R., SPURKLAND, A. & CLAESSON-WELSH, L. 2012. VEGFR2 induces c-Src signaling and vascular permeability in vivo via the adaptor protein TSAd. *J Exp Med,* 209**,** 1363-77.

SUNDVOLD-GJERSTAD, V., GRANUM, S., MUSTELIN, T., ANDERSEN, T. C., BERGE, T., SHAPIRO, M. J., SHAPIRO, V. S., SPURKLAND, A. & LEA, T. 2005. The C terminus of T cell-specific adapter protein (TSAd) is necessary for TSAd-mediated inhibition of Lck activity. *European journal of immunology,* 35**,** 1612-1620.

SUNDVOLD, V., TORGERSEN, K. M., POST, N. H., MARTI, F., KING, P. D., ROTTINGEN, J. A., SPURKLAND, A. & LEA, T. 2000. T cell-specific adapter protein inhibits T cell activation by modulating Lck activity. *J Immunol,* 165**,** 2927-31.

TAN, E. P., LI, Y., DEL CASTILLO VELASCO-HERRERA, M., YUSA, K. & BRADLEY, A. 2015. Off-target assessment of CRISPR-C as9 guiding RNA s in human i PS and mouse ES cells. *Genesis,* 53**,** 225-236.

TAYLOR, K. E., WONG, Q., LEVINE, D. M., MCHUGH, C., LAURIE, C., DOHENY, K., LAM, M. Y., BAER, A. N., CHALLACOMBE, S., LANFRANCHI, H., SCHIODT, M., SRINIVASAN, M., UMEHARA, H., VIVINO, F. B., ZHAO, Y., SHIBOSKI, S. C., DANIELS, T. E., GREENSPAN, J. S., SHIBOSKI, C. H. & CRISWELL, L. A. 2017. Genome-Wide Association Analysis Reveals Genetic Heterogeneity of Sjogren's Syndrome According to Ancestry. *Arthritis Rheumatol,* 69**,** 1294-1305.

TOMLINSON, M. G., KANE, L. P., SU, J., KADLECEK, T. A., MOLLENAUER, M. N. & WEISS, A. 2004. Expression and function of Tec, Itk, and Btk in lymphocytes: evidence for a unique role for Tec. *Mol Cell Biol,* 24**,** 2455-66.

URNOV, F. D. 2018. Genome Editing BC (before CRISPR): lasting lessons from the “old testament”. *The CRISPR Journal,* 1**,** 34-46.

VAN OVERBEEK, M., CAPURSO, D., CARTER, M. M., THOMPSON, M. S., FRIAS, E., RUSS, C., REECE-HOYES, J. S., NYE, C., GRADIA, S. & VIDAL, B. 2016. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Molecular cell,* 63**,** 633-646.

VEILLETTE, A., BOOKMAN, M. A., HORAK, E. M. & BOLEN, J. B. 1988. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56lck. *Cell,* 55**,** 301-8.

VELAZQUEZ, L., CHENG, A. M., FLEMING, H. E., FURLONGER, C., VESELY, S., BERNSTEIN, A., PAIGE, C. J. & PAWSON, T. 2002. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. *J Exp Med,* 195**,** 1599-611.

VORONOVA, A. F., BUSS, J. E., PATSCHINSKY, T., HUNTER, T. & SEFTON, B. M. 1984. Characterization of the protein apparently responsible for the elevated tyrosine protein kinase activity in LSTRA cells. *Mol Cell Biol,* 4**,** 2705-13.

WANG, H., KADLECEK, T. A., AU-YEUNG, B. B., GOODFELLOW, H. E., HSU, L. Y., FREEDMAN, T. S. & WEISS, A. 2010. ZAP-70: an essential kinase in T-cell signaling. *Cold Spring Harb Perspect Biol,* 2**,** a002279.

WANGE, R. L., MALEK, S. N., DESIDERIO, S. & SAMELSON, L. E. 1993. Tandem SH2 domains of ZAP-70 bind to T cell antigen receptor zeta and CD3 epsilon from activated Jurkat T cells. *J Biol Chem,* 268**,** 19797-801.

WEBER, J. R., ØRSTAVIK, S., TORGERSEN, K. M., DANBOLT, N. C., BERG, S. F., RYAN, J. C., TASKÉN, K., IMBODEN, J. B. & VAAGE, J. T. 1998. Molecular cloning of the cDNA encoding pp36, a tyrosine-phosphorylated adaptor protein selectively expressed by T cells and natural killer cells. *Journal of Experimental Medicine,* 187**,** 1157-1161.

WEDEL, J., STACK, M. P., SETO, T., SHEEHAN, M. M., FLYNN, E. A., STILLMAN, I. E., KONG, S. W., LIU, K. & BRISCOE, D. M. 2019. T Cell-Specific Adaptor Protein Regulates Mitochondrial Function and CD4(+) T Regulatory Cell Activity In Vivo following Transplantation. *J Immunol***,** ji1801604.

WEISS, L., WHITMARSH, A. J., YANG, D. D., RINCON, M., DAVIS, R. J. & FLAVELL, R. A. 2000. Regulation of c-Jun NH(2)-terminal kinase (Jnk) gene expression during T cell activation. *J Exp Med,* 191**,** 139-46.

WILSON, I. A. & GARCIA, K. C. 1997. T-cell receptor structure and TCR complexes. *Curr Opin Struct Biol,* 7**,** 839-48.

WIRTH, T. C., XUE, H. H., RAI, D., SABEL, J. T., BAIR, T., HARTY, J. T. & BADOVINAC, V. P. 2010. Repetitive antigen stimulation induces stepwise transcriptome diversification but preserves a core signature of memory CD8(+) T cell differentiation. *Immunity,* 33**,** 128-40.

WU, J., KATREKAR, A., HONIGBERG, L. A., SMITH, A. M., CONN, M. T., TANG, J., JEFFERY, D., MORTARA, K., SAMPANG, J., WILLIAMS, S. R., BUGGY, J. & CLARK, J. M. 2006. Identification of substrates of human protein-tyrosine phosphatase PTPN22. *J Biol Chem,* 281**,** 11002-10.

WU, L. W., MAYO, L. D., DUNBAR, J. D., KESSLER, K. M., OZES, O. N., WARREN, R. S. & DONNER, D. B. 2000. VRAP is an adaptor protein that binds KDR, a receptor for vascular endothelial cell growth factor. *J Biol Chem,* 275**,** 6059-62.

XU, C., GAGNON, E., CALL, M. E., SCHNELL, J. R., SCHWIETERS, C. D., CARMAN, C. V., CHOU, J. J. & WUCHERPFENNIG, K. W. 2008. Regulation of T cell receptor activation by dynamic membrane binding of the CD3ε cytoplasmic tyrosine-based motif. *Cell,* 135**,** 702-713.

YAMAGUCHI, H. & HENDRICKSON, W. A. 1996. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature,* 384**,** 484-9.

YAMASAKI, S., TAKAMATSU, M. & IWASHIMA, M. 1996. The kinase, SH3, and SH2 domains of Lck play critical roles in T-cell activation after ZAP-70 membrane localization. *Mol Cell Biol,* 16**,** 7151-60.

YAO, Z., YOON, S., KALIE, E., RAVIV, Z. & SEGER, R. 2010. Calcium regulation of EGF-induced ERK5 activation: role of Lad1-MEKK2 interaction. *PLoS One,* 5**,** e12627.

YASUDA, K., KOSUGI, A., HAYASHI, F., SAITOH, S., NAGAFUKU, M., MORI, Y., OGATA, M. & HAMAOKA, T. 2000. Serine 6 of Lck tyrosine kinase: a critical site for Lck myristoylation, membrane localization, and function in T lymphocytes. *J Immunol,* 165**,** 3226-31.

YEUNG, M. L., HOUZET, L., YEDAVALLI, V. S. & JEANG, K. T. 2009. A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J Biol Chem,* 284**,** 19463-73.

YOUNG, J., ZASTROW-HAYES, G., DESCHAMPS, S., SVITASHEV, S., ZAREMBA, M., ACHARYA, A., PAULRAJ, S., PETERSON-BURCH, B., SCHWARTZ, C., DJUKANOVIC, V., LENDERTS, B., FEIGENBUTZ, L., WANG, L., ALARCON, C., SIKSNYS, V., MAY, G., CHILCOAT, N. D. & KUMAR, S. 2019. CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-target Activity and Its Relevance in Crop Improvement. *Sci Rep,* 9**,** 6729.

YURCHAK, L. K. & SEFTON, B. M. 1995. Palmitoylation of Either Cys-3 or Cys-5 Is Required for the Biological-Activity of the Lck Tyrosine Protein-Kinase. *Molecular and Cellular Biology,* 15**,** 6914-6922.

ZHANG, W., SLOAN-LANCASTER, J., KITCHEN, J., TRIBLE, R. P. & SAMELSON, L. E. 1998. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell,* 92**,** 83-92.

ZHOU, S. Y., SHOELSON, S. E., CHAUDHURI, M., GISH, G., PAWSON, T., HASER, W. G., KING, F., ROBERTS, T., RATNOFSKY, S., LECHLEIDER, R. J., NEEL, B. G., BIRGE, R. B., FAJARDO, J. E., CHOU, M. M., HANAFUSA, H., SCHAFFHAUSEN, B. & CANTLEY, L. C. 1993. Sh2 Domains Recognize Specific Phosphopeptide Sequences. *Cell,* 72**,** 767-778.

ZHU, J. & PAUL, W. E. 2010. Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors. *Immunol Rev,* 238**,** 247-62.

ZINKERNAGEL, R. & DOHERTY, P. 1974. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system.

ZINKERNAGEL, R. M. & DOHERTY, P. C. 1975. H-2 compatability requirement for T-cell-mediated lysis of target cells infected with lymphocytic choriomeningitis virus. Different cytotoxic T-cell specificities are associated with structures coded for in H-2K or H-2D. *J Exp Med,* 141**,** 1427-36.

ZVELEBIL, M. J., PANAYOTOU, G., LINACRE, J. & WATERFIELD, M. D. 1995. Prediction and analysis of SH2 domain-phosphopeptide interactions. *Protein Eng,* 8**,** 527-33.

Appendix

Table 1. Overview of the known and suspected Lck SH2 domain's binding partners. ? - The phosphotyrosine motif has not been identified yet. * - The interaction is speculated not to be based on phosphotyrosine recognition.

Gene	pY site	Motif	Ref.
ACP1	pY131	QLIIEDPyYGNDSDF	(Granum et al., 2014)
	pY132	LIIEDPYyGNDSDFE	(Granum et al., 2014)
AXL	pY752	GVENSEIyDYLRQGN	(Granum et al., 2014)
	pY814	QEPDEILyVNMDEGG	(Granum et al., 2014)
BCAR1	pY12	NVLAKALyDNVAESP	(Granum et al., 2014)
	pY128	SKAQQGLyQVPGPSP	(Granum et al., 2014)
	pY165	PSPATDLyQVPPGPG	(Granum et al., 2014)
	pY222	PTRVGQGyVYEAAQP	(Granum et al., 2014)
	pY234	AQPEQDEyDIPRHLL	(Granum et al., 2014)
	pY249	APGPQDIyDVPPVRG	(Granum et al., 2014)
	pY267	SQYGQEVyDTPPMAV	(Granum et al., 2014)
	pY287	RDPLLEVyDVPPSVE	(Granum et al., 2014)
	pY306	PSNHHAVyDVPPSVS	(Granum et al., 2014)
	pY327	PLLREETyDVPPAFA	(Granum et al., 2014)
	pY372	PPPAPDLyDVPPGLR	(Granum et al., 2014)
	pY387	RPGPGTLyDVPRERV	(Granum et al., 2014)
	pY410	GVVDSGVyAVPPPAE	(Granum et al., 2014)
	pY664	EGGWMEDyDYVHLQG	(Granum et al., 2014)
	pY666	GWMEDYDyVHLQGKE	(Granum et al., 2014)
CBL	pY368	IKVTQEQyELYCEMG	(Granum et al., 2014)
	pY674	SSSANAIySLAARPL	(Granum et al., 2014)
	pY731	QQIDSCTyEAMYNIQ	(Granum et al., 2014)
	pY774	SENEDDGyDVPKPPV	(Granum et al., 2014)
CD247	?	?	(Lewis et al., 1997)
	pY72	QQGQNQLyNELNLGR	(Granum et al., 2014)
CD28	pY209	TRKHYQPyAPPRDFA	(Granum et al., 2014)
CD3E	pY188	PPVPNPDyEPIRKGQ	(Granum et al., 2014)
CD5	pY453	ASHVDNEySQPPRNS	(Granum et al., 2014)
	pY487	DNSSDSDyDLHGAQR	(Granum et al., 2014)
CD79A	pY182	GLDAGDEyEDENLYE	(Granum et al., 2014)
	pY188	EYEDENLyEGLNLDD	(Granum et al., 2014)
CD79B	pY196	GMEEDHTyEGLDIDQ	(Granum et al., 2014)
CSF3R	pY727	LPTLVQTyVLQGDPR	(Granum et al., 2014)
	pY752	GTSDQVLyGQLLGSP	(Granum et al., 2014)
	pY787	LTPSPKSyENLWFQA	(Granum et al., 2014)
CSK	pY184	VAAQDEFyRSGWALN	(Granum et al., 2014)
	pY263	VEEKGGLyIVTEYMA	(Granum et al., 2014)
CTNND2	pY1124	NTGISTLyRNSYGAP	(Granum et al., 2014)
	pY1154	QEPSRKDyETYQPFQ	(Granum et al., 2014)
	pY1197	GLKSTGNyVDFYSAA	(Granum et al., 2014)
DAPP1	pY139	KVEEPSIyESVRVHT	(Granum et al., 2014)
DLG1	pY760	YEVDGRDyHFVTSRE	(Granum et al., 2014)
DOK1	pY296	LDSPPALyAEPLDSL	(Granum et al., 2014)
	pY341	KPLYWDLyEHAQQQL	(Granum et al., 2014)
	pY362	DPKEDPIyDEPEGLA	(Granum et al., 2014)
	pY398	ARVKEEGyELPYNPA	(Granum et al., 2014)
	pY409	YNPATDDyAVPPPRS	(Granum et al., 2014)
DOK2	pY299	PRGQEGEyAVPFDAV	(Granum et al., 2014)
	pY345	PPRPDHIyDEPEGVA	(Granum et al., 2014)
DOK3	pY208	PMEENSIySSWQEVG	(Granum et al., 2014)
	pY398	PPGNEHLyENLCVLE	(Granum et al., 2014)
	pY432	SPTTSPIyHNGQDLS	(Granum et al., 2014)

Gene	pY site	Motif	Ref.
ESR1	pY537	CKNVVPLyDLLLEML	(Granum et al., 2014)
ESR2	pY488	CKNVVPVyDLLLEML	(Granum et al., 2014)
EZR	pY478	PPPPPPVyEPVSYHV	(Granum et al., 2014)
	pY483	PVYEPVSyHVQESLQ	(Granum et al., 2014)
FAS	pY232	SDVDLSKyITTIAGV	(Granum et al., 2014)
FYN	pY185	SETTKGAySLSIRDW	(Granum et al., 2014)
	pY213	RKLDNGGyYITTRAQ	(Granum et al., 2014)
	pY531	FTATEPQyQPGENL	(Granum et al., 2014)
GAB2	pY266	TEFRDSTyDLPRSLA	(Granum et al., 2014)
	pY324	PATPLSAyQIPRTFT	(Granum et al., 2014)
	pY409	RASSCETyEYPQRGG	(Granum et al., 2014)
	pY614	KSTGSVDyLALDFQP	(Granum et al., 2014)
	pY643	TSDEKVDyVQVDKEK	(Granum et al., 2014)
IFNAR2	pY337	PRTSGGGyTMHGLTV	(Granum et al., 2014)
	pY411	DPFPEEDySSTEGSG	(Granum et al., 2014)
IL2RB	pY387	VYFTYDPySEEDPDE	(Granum et al., 2014)
	pY418	LSGEDDAyCTFPSRD	(Granum et al., 2014)
	pY536	LPLNTDAyLSLQELQ	(Granum et al., 2014)
IL3RB	pY593	SFDFNGPyLGPPHSR	(Granum et al., 2014)
	pY711	DPGVASGyVSSADLV	(Granum et al., 2014)
	pY766	VKSGFEGyVELPPIE	(Granum et al., 2014)
	pY882	KALKQQDyLSLPPWE	(Granum et al., 2014)
ITK	pY512	RFVLDDQyTSSTGTK	(Granum et al., 2014)
JAK3	pY785	NSLISSDyELLSDPT	(Granum et al., 2014)
	pY980	LLPLDKDyYVVREPG	(Granum et al., 2014)
KHDRBS1	pY440	GAYREHPyGRY	(Granum et al., 2014)
KIR2DL2	pY109	GSVTHSPyQLSAPSD	(Granum et al., 2014)
	pY126	DIVITGLyEKPSLSA	(Granum et al., 2014)
	pY302	QDPQEVTyTQLNHCV	(Granum et al., 2014)
KIT	pY568	EEINGNNyVYIDPTQ	(Granum et al., 2014)
	pY578	IDPTQLPyDHKWEFP	(Granum et al., 2014)
	pY936	SESTNHIySNLANCS	(Granum et al., 2014)
LAT	pY110	GANSVASyENEGASG	(Granum et al., 2014)
	pY156	ADEDEDDyHNPGYLV	(Granum et al., 2014)
	pY200	SMESIDDyVNVPESG	(Granum et al., 2014)
	pY220	SLDGSREyVNVSQEL	(Granum et al., 2014)
	pY225	EEEGAPDyENLQELN	(Granum et al., 2014)
LAX1	pY93	RQRAKNIyDILPWRQ	(Granum et al., 2014)
LCK	pY192	NLDNGGFyISPRITF	(Granum et al., 2014)
	pY505	FFTATEGQyQPQP	(Granum et al., 2014)
	SH3 domain	pY independent*	(Lee-Fruman et al., 1996)
LCP2	pY113	SSFEEDDyESPNDDQ	(Granum et al., 2014)
	pY128	DGEDDGDyESPNEEE	(Granum et al., 2014)
	pY145	PVEDDADyEPPPSND	(Granum et al., 2014)
LIME1	?	?	(Hur et al., 2003)
	pY235	ALAGDLAyQTLPLRA	(Granum et al., 2014)
	pY254	SGPLENVyESIRELG	(Granum et al., 2014, Brdickova et al., 2003)
LYN	pY193	SLDNGGYyISPRITF	(Granum et al., 2014)
	pY264	FGEVWMGyYNNSTKV	(Granum et al., 2014)
	pY305	HDKLVRLyAVVTREE	(Granum et al., 2014)
	pY315	VTREEPIyIITEYMA	(Granum et al., 2014)
	pY500	QSVLDDFyTATEGQY	(Granum et al., 2014)
	pY507	YTATEGQyQQQP	(Granum et al., 2014)
MUC1	pY1212	PMSEYPTyHTHGRYV	(Granum et al., 2014)
	pY1218	TYHTHGRyVPPSSTD	(Granum et al., 2014)
	pY1229	SSTDRSPyEKVSAGN	(Granum et al., 2014)
NEDD9	pY12	NLMARALyDNVPECA	(Granum et al., 2014)
	pY92	TFGQQKLyQVPNPQA	(Granum et al., 2014)
	pY106	AAPRDTIyQVPPSYQ	(Granum et al., 2014)
	pY112	IYQVPPSyQNQGIYQ	(Granum et al., 2014)
	pY118	SYQNQGIyQVPTGHG	(Granum et al., 2014)
	pY164	PVRTGHGyVYEYPSR	(Granum et al., 2014)
	pY166	RTGHGYVyEYPSRYQ	(Granum et al., 2014)
	pY168	GHGYVYEyPSRYQKD	(Granum et al., 2014)
	pY177	SRYQKDVyDIPPSHT	(Granum et al., 2014)

Gene	pY site	Motif	Ref.
NEDD9	pY189	SHTTQGVyDIPPSSA	(Granum et al., 2014)
	pY317	VGSQNDAyDVPRGVQ	(Granum et al., 2014)
	pY345	PQERDGVyDVPLHNP	(Granum et al., 2014)
	pY629	ERSWMDDyDYVHLQG	(Granum et al., 2014)
	pY631	SWMDDYDyVHLQGKE	(Granum et al., 2014)
NFKBIA	pY305	FTEDELPyDDCVFGG	(Granum et al., 2014)
PAG1	pY163	GLGMEGPyEVLKDSS	(Granum et al., 2014)
	pY181	NMVEDCLyETVKEIK	(Granum et al., 2014)
	pY341	LTVPESTyTSIQGDP	(Granum et al., 2014)
	pY359	PSSCNDLyATVKDFE	(Granum et al., 2014)
	pY387	SEEPEPDyEAIQTLN	(Granum et al., 2014)
	pY417	LVPKENDyESISDLQ	(Granum et al., 2014)
PECAM1	pY690	PLNSDVQyTEVQVSS	(Granum et al., 2014)
PIK3CA	pY508	SREAGFSySHAGLSN	(Granum et al., 2014)
PIK3R1	pY463	FQEKSREyDRLYEEY	(Granum et al., 2014)
	pY467	SREYDRLyEEYTRTS	(Granum et al., 2014)
	pY470	YDRLYEEyTRTSQEI	(Granum et al., 2014)
	pY607	NENTEDQySLVEDDE	(Granum et al., 2014)
	pY688	FAEPYNLySSLKELV	(Granum et al., 2014)
PLCG1	pY472	KLAEGSAyEEVPTSM	(Granum et al., 2014)
	pY771	IGTAEPDyGALYEGR	(Granum et al., 2014)
	pY775	EPDYGALyEGRNPGF	(Granum et al., 2014)
	pY783	EGRNPGFyVEANPMP	(Granum et al., 2014)
	pY1253	EGSFESRyQQPFEDF	(Granum et al., 2014)
PLCG2	pY1217	LNNQLFLyDTHQNLR	(Granum et al., 2014)
PLD2	pY179	RLLTMSFyRNYHAMT	(Granum et al., 2014)
	pY470	GRWDDLHyRLTDLGD	(Granum et al., 2014)
PRKCA	pY657	SDFEGFSyVNPQFVH	(Granum et al., 2014)
PRKCD	pY64	STFDAHIyEGRVIQI	(Granum et al., 2014)
	pY155	IKQAKIHyIKNHEFI	(Granum et al., 2014)
	pY630	KVKSPRDySNFDQEF	(Granum et al., 2014)
PTK2	pY148	KPTLNFFyQQVKSDY	(Granum et al., 2014)
	pY397	SVSETDDyAEIIDEE	(Granum et al., 2014)
	pY407	IIDEEDTyTMPSTRD	(Granum et al., 2014)
	pY570	GDFGLSRyMEDSTYY	(Granum et al., 2014)
	pY861	PIGNQHIyQPVGKPD	(Granum et al., 2014)
	pY925	DRSNDKVyENVTGLV	(Granum et al., 2014)
PTK2B	pY402	CSIESDIyAEIPDET	(Granum et al., 2014)
	pY440	EGFFGEVyEGVYTNH	(Granum et al., 2014)
	pY881	DRTDDLVyLNVMELV	(Granum et al., 2014)
PTPN6	pY61	IQNSGDFyDLYGGEK	(Granum et al., 2014)
	pY64	SGDFYDLyGGEKFAT	(Granum et al., 2014)
	pY377	MQRAYGPySVINCGE	(Granum et al., 2014)
	pY536	QKGQESEyGNITYPP	(Granum et al., 2014)
	pY564	SKHKEDVyENLHTKN	(Granum et al., 2014, Štefanová et al., 2003)
PTPN11	pY63	IQNTGDYyDLYGGEK	(Granum et al., 2014)
	pY66	TGDYYDLyGGEKFAT	(Granum et al., 2014)
	pY304	PNEPVSDyINANIIM	(Granum et al., 2014)
	pY584	REDSARVyENVGLMQ	(Granum et al., 2014)
PTPRC	?	pY independent*	(Courtney et al., 2017, Ng et al., 1996)
	pY705	NGDAGSNyINASYID	(Granum et al., 2014)
	pY978	RNSNVIPyDYNRVPL	(Granum et al., 2014)
PXN	pY31	FLSEETPySYPTGNH	(Granum et al., 2014)
	pY40	YPTGNHTyQEIAVPP	(Granum et al., 2014)
	pY181	PGALSPLyGVPETNS	(Granum et al., 2014)
	pY375	QPYCEKDyHNLFSPR	(Granum et al., 2014)
RAF1	pY340	RGQRDSSyYWEIEAS	(Granum et al., 2014)
	pY341	GQRDSSYyWEIEASE	(Granum et al., 2014)
SH2B3	?	?	(Li et al., 2000)
SH2D2A	pY280	PKPSNPIyNEPDEPI	(Granum et al., 2014, Marti et al., 2006, Granum et al., 2008)
	pY290	PDEPIAFyAMGRGSP	(Granum et al., 2014, Marti et al., 2006, Granum et al., 2008)
	pY305	GEAPSNIyVEVEDEG	(Granum et al., 2014, Marti et al., 2006, Granum et al., 2008)

Gene	pY site	Motif	Ref.
SH3BP2	pY174	YPTDNEDyEHDDEDD	(Granum et al., 2014)
	pY183	HDDEDDSyLEPDSPE	(Granum et al., 2014)
	pY448	GDDSDEDyEKVPLPN	(Granum et al., 2014)
SHC1	pY349	EEPPDHQyYNDFPGK	(Granum et al., 2014)
	pY427	ELFDDPSyVNVQNLD	(Granum et al., 2014)
SIT1	pY90	SVEEVPLyGNLHYLQ	(Granum et al., 2014)
	pY95	PLYGNLHyLQTGRLS	(Granum et al., 2014)
	pY169	SGPEPELyASVCAQT	(Granum et al., 2014)
	pY188	ASFPDQAyANSQPAA	(Granum et al., 2014)
SKAP1	pY219	LSSLTIPyEEDEEEE	(Granum et al., 2014)
	pY232	EEEKEETyDDIDGFD	(Granum et al., 2014)
	pY271	EKEEEDIyEVLPDEE	(Granum et al., 2014)
	pY295	TRRKGVDyASYYQGL	(Granum et al., 2014)
SYK	pY203	ARDNNGSyALCLLHE	(Granum et al., 2014)
	pY244	LWQLVEHySYKADGL	(Granum et al., 2014)
	pY323	STVSFNPyEPELAPW	(Granum et al., 2014)
	pY348	LPMDTEVyESPYADP	(Granum et al., 2014)
	pY518	ADENyYKAQTHG	(Couture et al., 1996a)
	pY519	ADENYyKAQTHG	(Couture et al., 1996a)
TEK	pY1048	GMTCAELyEKLPQGY	(Granum et al., 2014)
	pY1108	TYVNTTLyEKFTYAG	(Granum et al., 2014)
VAV1	pY826	GWWRGEIyGRVGWFP	(Granum et al., 2014)
ZAP70	?	?	(Yamasaki et al., 1996, Straus et al., 1996, Duplay et al., 1994)
	pY164	AHERMPWyHSSLTRE	(Granum et al., 2014)
	pY315	MPMDTSVyESPYSDP	(Granum et al., 2014)
	pY319	TSVYESPySDPEELK	(Pelosi et al., 1999)
	pY474	LLNVNRHyAKISDFG	(Granum et al., 2014)
	pY492	ALGADDSyYTARSAG	(Granum et al., 2014)

I

A simple and efficient workflow for generation of knock-in mutations in Jurkat T cells using CRISPR/Cas9

Paweł Borowicz[1]*, Hanna Chan[1]*, Daniel Medina[1], Simon Gumpelmair[1], Hanna Kjelstrup[1] and Anne Spurkland[1§]

**Contributed equally*

[1]Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, Norway.

Abstract

CRISPR/Cas9 is a powerful gene editing tool allowing for specific gene manipulation at targeted sites in the genome. Here we used CRISPR/Cas9 mediated gene editing to better characterise signalling pathways during T cell activation. Knock-in mutations were introduced in Jurkat T cells by homologous directed repair using single stranded oligodeoxynucleotides. Specifically, we aimed to create targeted mutations at two loci within *LCK*, a constitutively expressed gene, and at three loci within *SH2D2A*, whose expression is induced upon T cell activation. Here we present a simple workflow that can be applied by any laboratory equipped for cell culture work, utilising basic flow cytometry, western blotting and PCR techniques. Our data reveal that gene editing may be locus-dependent and can vary between target sites, also within a gene. In our two targeted genes, on average 2% of the clones harboured homozygous mutations as assessed by allele-specific PCR and subsequent sequencing. We highlight the importance of decreasing the clonal heterogeneity and developing robust screening methods to accurately select for correct knock-in mutations. Our workflow may be employed in other immune cell lines, and acts as a useful approach for decoding functional mechanisms of proteins of interest.

Introduction

Immunity to infection depends on adequate intracellular signalling in many different immune cell types. Elucidation of these signalling pathways and the function of single signalling proteins, have often been performed in cell lines and rodents using a variety of protein specific inhibitors, knock-down assays and knockout techniques to target the protein of interest. Although these techniques have proven effective, many signalling molecules contain multiple functional domains. Thus, knocking out an entire gene may hide domain-specific activity of the encoded protein. To explore the function of specific sub-domains, cDNA encoding recombinant proteins with domain specific mutations can be forcibly expressed in both primary cells and in cell lines. However, such exogenous expression of cDNA may not be regulated by cellular processes and the resulting recombinant protein might compete with the endogenous protein.

Methods for genome editing of specific loci in intact cells in order to alter endogenous protein activity through specific single point mutations, have been hampered by low efficiency and technical difficulties [1]. However, with the recent discovery of clustered regularly interspaced palindromic repeats (CRISPR)/Cas (CRISPR-associated protein) system, a prokaryotic adaptive immune system [2], researchers now have a powerful gene editing tool to precisely introduce gene edits in various living cells and organisms [3].

In general, the molecular machinery of CRISPR/Cas consists of two major components: a DNA binding domain directing the machinery to a defined locus and an effector domain that facilitates DNA cleavage resulting in a double-strand break (DSB). While several CRISPR/Cas systems exist, the most widely used system adapted for genome editing is CRISPR/Cas9, derived from *Streptococcus pyogenes* [4]. Cas9 is an endonuclease, evolved to be guided to the target site by a single guide RNA (sgRNA) that binds to a complementary target DNA strand. Specific cleavage of a double stranded DNA by Cas9 requires binding of Cas9 to the sgRNA-target DNA complex as well as the presence of a protospacer adjacent motif (PAM) in the target DNA sequence [5, 6]. In this case, the Cas9 recognises a 5'-NGG PAM sequence immediately downstream of the sgRNA binding site [7] (Supplementary Fig. 1).

In a eukaryotic cell, there are several DNA mechanisms in place to repair the DSB. The most common and rapid repair occurs through non-homologous end joining (NHEJ), where micro-homologous overhangs are used to join the DSB [8]. Although small insertions or deletions may occur in NHEJ, it is relatively accurate. In contrast, microhomology mediated end joining, another DSB repair mechanism, aligns broken strands with homologous sequences of 5-25 bp in length and is highly error prone [9]. Microhomology mediated end joining often results in insertions, deletions (indels) and inversions; some more frequently occurring than others. A majority of indels result in frame shifts of the sequence and consequently occurrence of premature stop codons, resulting in knockout mutations. The CRISPR/Cas9 system works with high efficiency, and as such has become a popular tool to create knockout mutations [6, 10, 11].

Alternatively, DNA breaks can be repaired by homology-directed repair (HDR). By combining a CRISPR/Cas9 cleavage with DNA templates, defined point mutations can be introduced [4]. The DNA template for HDR can be provided as a single stranded deoxynucleotide (ssODN), commonly referred to as a repair template, and contains homologous arms complementary to either the target or non-target DNA strand. Recently, it has been shown that knock-in mutations work with higher efficiency when ssODNs are complementary to the non-target DNA strand, as this is more available to binding post-cleavage [12]. The repair template ssODNs contain the desired knock-in mutation and an additional synonymous knock-in mutation

[§]To whom correspondence may be addressed: Department of Molecular Medicine, Institute of Basic Medical Sciences, Tel.: +47 22851125, E-mail: anne.spurkland@medisin.uio.no

Abbreviations list: HDR – homologous directed repair, CRISPR – clustered regularly interspaced palindromic repeats, Cas9 – CRISPR-associated protein, DSB – double-strand break, sgRNA – single guide RNA, PAM – protospacer adjacent motif, NHEJ – non-homologous end joining, ssODN – single stranded deoxynucleotide, LCK – lymphocyte-specific protein tyrosine kinase, TSAd – T cell specific adaptor protein, SNP – single nucleotide polymorphism, AS-PCR – Allele specific PCR, WT – Wild-type, GFP – Green fluorescent protein.

for the PAM site to prevent repeated binding of sgRNA-Cas9 and re-cutting of the target strand. Together with the host DNA repair machinery, the ssODN mediates HDR which introduces the desired mutations into the gene of interest (Supplementary Fig. 1C-D).

To better characterise signalling pathways controlling T cell activation, we have employed the CRISPR/Cas9 system to introduce targeted point mutations by HDR in Jurkat T cells using ssODNs. Specifically, we were interested in how phosphorylation associated domains of lymphocyte-specific protein tyrosine kinase (LCK) and T cell specific adaptor protein (TSAd) function in T cell activation [13, 14]. We thus sought to introduce knock-in mutations in the *LCK* and *SH2D2A* genes that encode for LCK and TSAd respectively, in order to either abolish phosphotyrosine binding (TSAd-SH2) or to mimic phosphorylation dependent alteration of SH2 domain specificity (LCK-SH2) [13]. Additionally, we mutated two other phosphotyrosines: LCK Y505 and TSAd Y290, which both are known ligands for the LCK SH2 domain [13]. In parallel efforts, we performed experiments to generate Jurkat T cells homozygous for the non-synonymous single nucleotide polymorphism (SNP), rs926103, in *SH2D2A*. This SNP is known to affect the interaction of TSAd with LCK [15], hence we aimed to determine how this SNP affects endogenously expressed LCK and TSAd in intact cells.

Altogether we attempted to edit five loci in these two genes in multiple knock-in experiments. We successfully introduced knock-in mutations within exon 6 and exon 13 of the *LCK* gene and in exon 7 of *SH2D2A*. However, knock-in mutation experiments in the early exons of *SH2D2A* were unsuccessful, implying that the gene editing system may be influenced by its target site.

We successfully generated knock-in mutated Jurkat cells in repeated independent experiments for three of the five targeted loci. There were striking differences in mutation efficiency within the same gene. Based on our results, we here provide an overview of a simple workflow used for generating and identifying knock-in mutations in cell lines, and we discuss potential challenges faced when editing constitutively expressed genes compared to genes that are upregulated following T cell activation.

Results

Heterogeneity of cell lines may affect interpretation of genome editing results

To study LCK and TSAd mediated phosphotyrosine dependent signalling in Jurkat T cells, we aimed to create point mutations, Y192F, Y192E and Y505F in *LCK* (Supplementary Fig. 2A), and N52S, R120K, Y290F and Y290E in *SH2D2A* (Supplementary Fig. 2B). Here we report our results from a total of 31

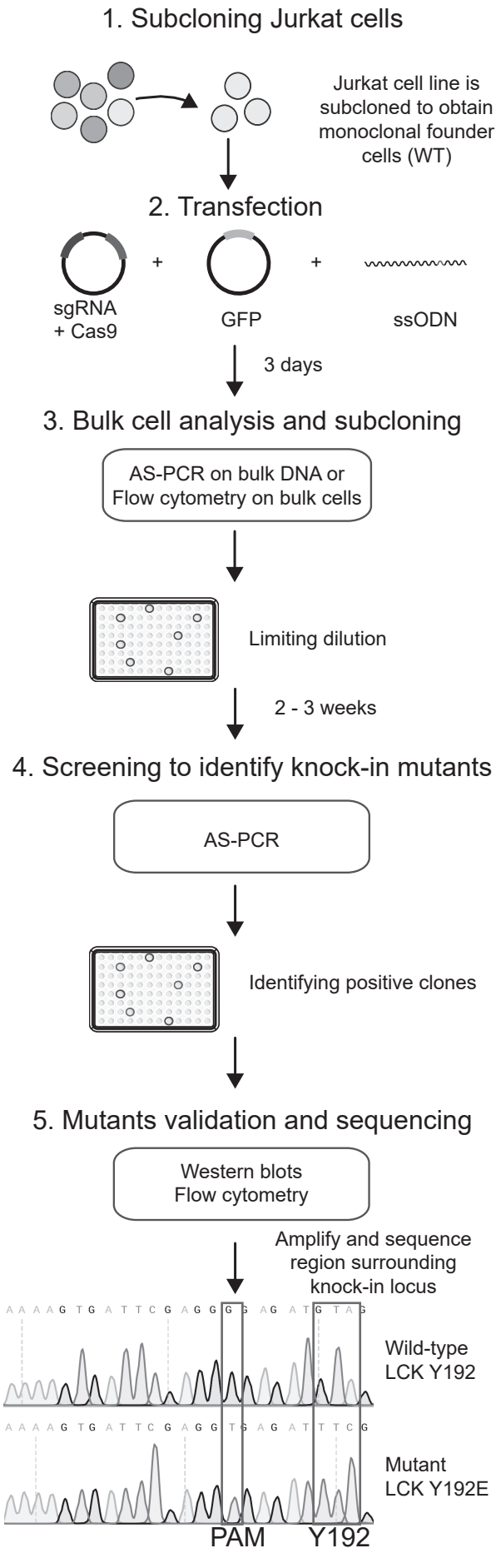


Figure 1. Workflow for creating knock-in mutations in Jurkat T cells using CRISPR/Cas9. Schematic depiction of mutation strategy: Jurkat T cells are subcloned and assessed for expression of CD3 and the protein of interest prior to CRISPR/Cas9 gene editing. Single clones are co-transfected with sgRNA/Cas9 encoding plasmid, GFP encoding plasmid and the ssODN repair template. Three days after transfection, bulk DNA is tested by AS-PCR to detect evidence of knock-in mutations or flow cytometry to detect evidence of knockout of the gene of interest. A limiting dilution is carried out and after 2-3 weeks single clones are screened by AS-PCR for successful knock-in mutations. Positive clones are further validated using Western blotting and flow cytometry before being confirmed by sequencing. Heterozygous samples are subjected to a second round of limiting dilution to exclude possible polyclonality.

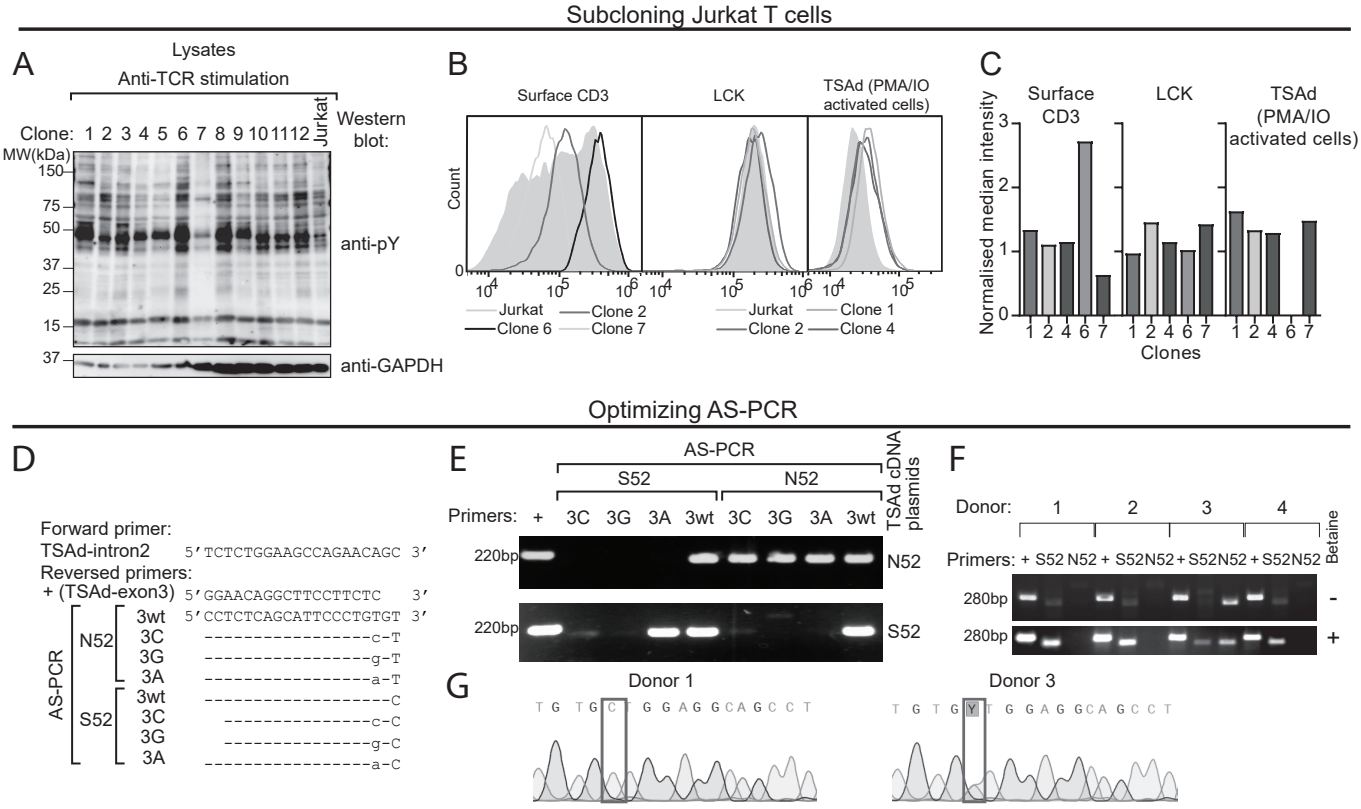


Figure 2. Initial necessary steps prior to genome editing: sub-cloning of chosen cell line and AS-PCR optimisation. (A) Analysis of clonal variation of Jurkat T cell response to TCR stimulation. Western blot shows tyrosine phosphorylation in Jurkat T cell subclones after 2 min stimulation with anti-CD3 antibodies (B-C) Analysis of surface CD3, LCK and TSAd expression levels in Jurkat T subclones as assessed with flow cytometry. (B) Selected histograms of data presented as graphs in (C). (D) Sequences of AS-PCR primers used to screen for knock-in mutations of N52S in SH2D2A. The additional mismatches introduced in position -3 from the 3' end of the primer are indicated. (E) AS-PCR was performed on cDNA encoding TSAd-N52 (top gel) or TSAd-S52 (bottom gel). Primers correspond to those presented in (D). (F) AS-PCR performed on isolated DNA from blood donor samples using the N/S52 3A primers (D). (G) Sequencing analysis of the N/S52 region of DNA from donor 1 and donor 3 in (F). The nucleotide determining the N/S52 codon is indicated with a red box.

gene-editing experiments of *LCK* and *SH2D2A* in Jurkat cells performed over a period of three years. Our current overall mutation and cloning workflow, based on results presented in the following section, is depicted in Fig. 1.

Similar to other cancer cell lines, Jurkat T cell lines are inherently heterogenous [22]. During the course of our studies, we noticed that our Jurkat T cells had heterogeneous expression of surface CD3, which may be a clonotypic property. Our protein of interest, LCK and TSAd, directly or indirectly interact with CD3 to initiate or modulate downstream signalling events in the T cell. Therefore, it was essential that Jurkat T cell clones with uniform expression of surface CD3, LCK and TSAd were selected for our gene-editing experiments. Thus, prior to performing genome editing experiments involving isolation of single clones by limiting dilution, we subcloned our Jurkat T cell line creating monoclonal founder cells, hereafter referred to as WT. The resulting clones were analysed for response to anti-CD3 stimulation, expression of surface CD3 and expression of our two proteins of interest, LCK and TSAd (Fig. 2). The results showed considerable variability in the Jurkat T cell clones' response to anti-CD3 stimulation when assessed with anti-phosphotyrosine immunoblotting (Fig. 2A). Some clones (including clone 6) responded more strongly than the mother cell line (Jurkat), while other clones (including clone 7) responded relatively weaker. This variability was better correlated to the amount of CD3 expressed on the cell surface than to the amount of LCK expressed (Fig. 2C).

Establishment of method for reliable screening of knock-in mutants

To identify successful targeted point mutations in the Jurkat T cell clones as well as in bulk cells prior to limiting dilution, we used allele-specific PCR (AS-PCR) (Fig. 1), a technique also used to detect SNPs [23, 24]. For all loci, we used a common primer in one orientation and two alternate detection primers in the other orientation. The detection primers were complimentary at the 3'-end either to the WT sequence or to the knock-in mutation, including the PAM site mutation if relevant (Supplementary figure 3).

Despite lacking exonuclease activity, Taq polymerase tolerates a certain degree of 3' mismatch in the primer [25]. To further destabilise the primer-template complex and ensure specific amplification of either WT or the mutated variant, we in addition introduced mismatches 2-5 bases upstream of the 3'-end in the detection primers [26]. The effect of one such additional mismatch in the +3 position from the 3' end was systematically tested for the rs926103 SNP (with alleles here referred to as N52 (asparagine) and S52 (serine)) (Fig. 2D). As seen from Fig. 2E, when the detection primers lacking a destabilizing +3 nucleotide were tested against cDNA-containing plasmids encoding one or the other of the two variants, none of them reliably distinguished the two variants from each other. However, for both detection primers, inclusion of an A base in position +3 was sufficient to discriminate one allele from the alternative variant. The two alternate bases C or G, had a similar effect only for the primer detecting the allele encoding N52 (Fig. 2E).

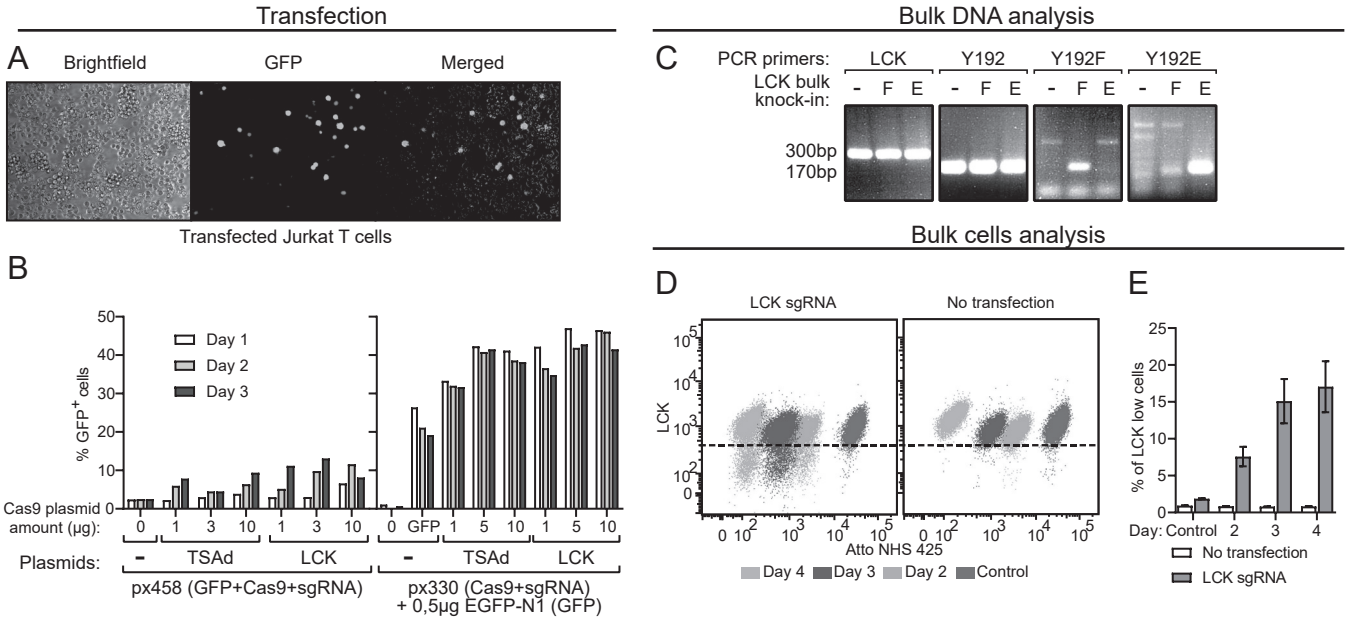


Figure 3. Transfection and CRISPR/Cas9 setup validation on bulk cells. (A) Cells were co-transfected with sgRNA-Cas9 and GFP encoding plasmids. The GFP signal serves as indication of successful transfection. Images were taken one day after transfection. (B) Cells were transfected with the indicated plasmids. The GFP signal was subsequently assessed up to 3 days after transfection. (C) 3 days after transfection, DNA isolated from bulk cells was analysed by AS-PCR to detect the presence of knock-in mutation. "-" symbolises untransfected cells. (D) Cells were transfected with sgRNA-Cas9 plasmid targeting LCK or left untreated. Samples were collected three days in a row, barcoded with varying CTV concentrations and analysed by flow cytometry. (E) Graph presents the median and range of the percentage of cells with low LCK expression obtained from experiments (n=2) as in (D).

As genomic DNA is more complex than a cDNA containing plasmid, we subsequently used the TSAd-N/S52 detection primers containing the +3A nucleotide to genotype the DNA of four anonymous blood donors. Donors 1, 2 and 4 were found to carry the S52 allele, while donor 3 was found to carry the N52 allele. Addition of betaine to the PCR mixture [27], clearly improved the specificity of the PCR, and revealed that donor 3 also carried the S52 allele (Fig. 2F). These genotyping results were confirmed by sequencing (Fig. 2G). Taken together, establishment of AS-PCR for testing of knock-in mutations should be done prior to gene editing experiments. Systematic testing of alternate mismatches in the 3'-end of detection primers for AS-PCR should be performed, to ensure a specific screening assay for knock-in mutations. This is particularly important when the desired knock-in mutation only differs by one or two nucleotides from the WT sequence.

Transfection and bulk analysis of sgRNA targeting efficiency

Co-expression with cDNA encoding green fluorescent protein (GFP) may allow monitoring of transfection efficiency with plasmids encoding both sgRNA and Cas9 (Fig. 3A). Initial experiments using the alternative PX458 plasmid encoding both sgRNA, Cas9 and GFP in cis, showed very low GFP expression efficiency (less than 10% of positive cells) (Fig. 3B). Attempts to isolate GFP positive cells by cell sorting, did not yield in any mutated cells (data not shown). In all experiments reported here, we thus co-transfected cells with px330 plasmids encoding the sgRNA and Cas9 together with a plasmid encoding GFP (pEGFP-N1) (Fig 3B).

Limiting dilution to obtain clonal cells is time consuming, thus if possible, a few days post transfection, the bulk of cells should be monitored for CRISPR-Cas9 targeting efficiency prior to further experimental work. With AS-PCR established prior to genome editing (Fig. 2E-G), the bulk of transfected cells can be tested for the presence of knock-in mutations already at day 3 post-transfection, as shown in Fig. 3C. However, we noticed that when the detection primers were complementary to the ssODN repair template, false positive AS-PCR may result, due to excess ssODN present in the media or intracellularly after the transfection (data not shown). In addition, provided the targeted gene product can be detected by flow cytometry, initial experiments with sgRNA-Cas9 transfection in the absence of ssODN repair template may allow assessment of Cas9 targeting efficiency through loss of the target protein in the bulk of transfected cells, as a consequence of generated knockouts. An example of this in the bulk cells expressing Cas9 and sgRNA targeting Lck Y192 is shown in Fig. 3D and E.

Controlling the heterogeneity of generated mutants

To obtain single clones with homozygous knock-in mutations, the cells were subjected to limiting dilution three days after the transfection. Clones were subsequently screened for successful knock-in mutations 2-3 weeks later, using AS-PCR (Fig. 4A). Positive clones were further tested with detection primers for the WT allele (Fig. 4B), to reveal possible heterozygosity. Further characterisation of the clones included analysis of expression of the targeted protein by Western blotting (Fig. 4C and D) and flow cytometry (Fig. 4E and F). Western blotting may reveal changes in protein expression or length. Thus, performing a high percentage PAGE (12% or 4-20% gradient gels depending on the protein of interest) could be a measure of precaution (Supplementary Fig. 4). It is especially useful for target sites at the C-terminus as they are unlikely to generate a knockout (e.g. Lck Y505), but the change in protein length can be observed. Presence of more than one clone in a sample, may be detected by flow cytometry, provided that one of the clones is hemizygous or knockout for the protein of interest (Fig. 4E). Since both of our target genes affect T cell activation and TCR-signalling, we also monitored surface expression of the CD3 again. As shown in Fig. 4F, in the LCK knockout cell lines we observed a trend towards lower surface expression of CD3 compared to the starting Jurkat T cells clone and the LCK expressing daughter clones (Fig. 4F).

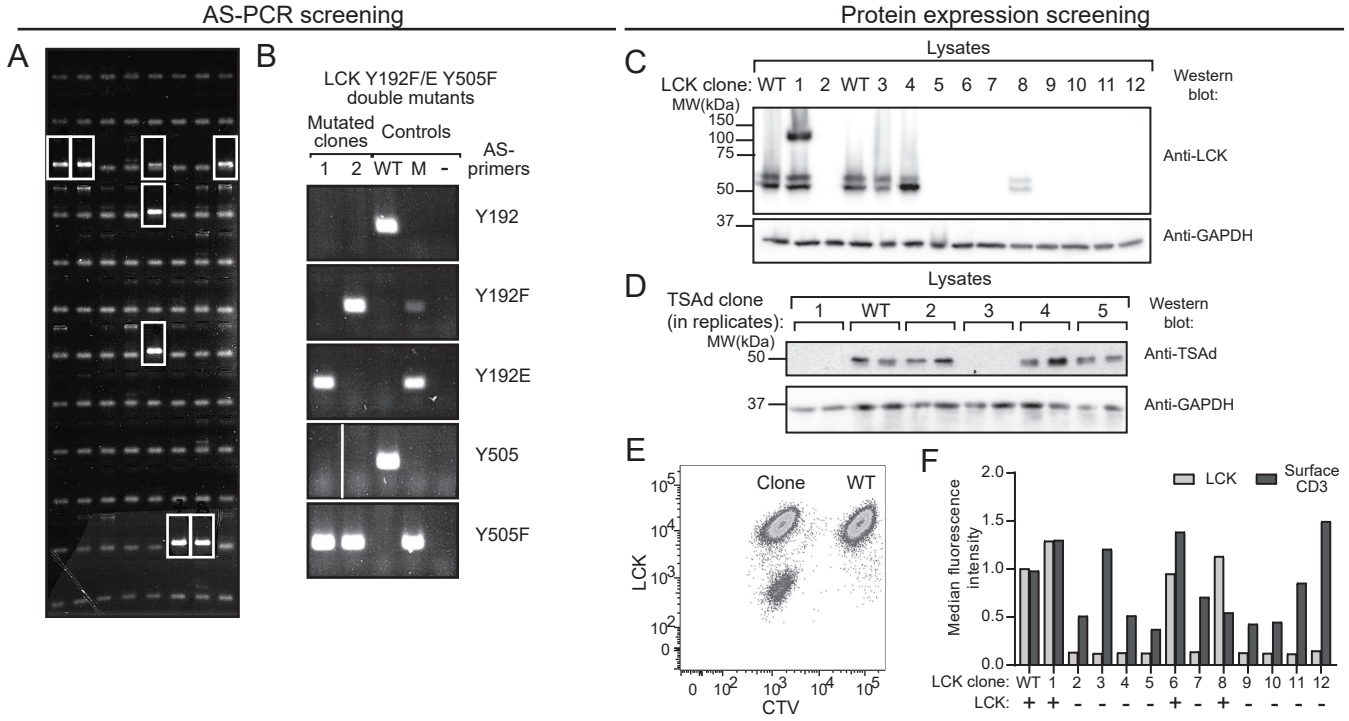


Figure 4. Screening knock-in clones with AS-PCR, western blotting and flow cytometry. (A) 2-3 weeks after limiting dilution, single clones were screened by AS-PCR to identify clones harbouring the knock-in mutation. Initial screening was performed with a 96-well 2% agarose gel, here testing the presence of Lck Y192F mutation. Samples with positive AS-PCR result are shown with white boxes (B) Positive hits from analysis performed as in A were verified with 16-well 2% agarose gel for better resolution. (C and D) Western blot show the expression of indicated proteins in various LCK mutated (C) and TSAd mutated clones. (E) Example of the analysis of clonal homogeneity of one LCK mutated clone as assessed by flow cytometry. (F) Monitoring with flow cytometry the surface CD3 expression and intracellular LCK expression of various mutated LCK clones.

Sequencing of knock-in mutants reveal a wide range of mutations and deletions

The screening methods described above provide most of the information needed to identify knock-in mutants. However these screening methods may not reveal knockout mutants resulting from indels at the C-terminus, nor will they reveal possible additional silent mutations or mutations resulting in amino acid substitutions outside of the targeted codon. Thus, all clones must also be verified by sequencing.

During the course of our experiments we sequenced amplicons of approximately 250 bp covering the targeted locus in a total of 54 clones for LCK Y192, 10 clones for LCK Y505, 15 clones for TSAd Y290 and 2 clones for TSAd R120. For LCK Y192 and TSAd Y290, clones with suspected gene disruption were also included for sequencing, explaining the higher number of sequenced clones in comparison with the other loci. Fig. 5A shows the sequencing trace around *LCK* codon 192 of a WT clone as well as five mutants. We used EMBOSS Needle pairwise sequence alignment tool to deconvolute the sequences and establish the two sequences present in the clones. Examples of possible sequences obtained in LCK Y192 or LCK Y505 mutated clones are shown in Fig. 5B and C respectively. Some sequence outcomes were more likely to occur than others. We observed multiple instances of hemizygous knock-in mutants (where one allele is a knockout and the other allele harbours the correct knock-in mutation). Clones determined to be WT/knock-in heterozygous could either truly be heterozygous or a mixture of a WT clone with a knock-in homozygous mutant clone. As a rule of thumb, we subjected all heterozygous clones, as determined by AS-PCR (Fig. 4B), to a second round of limiting dilution. Even in the presence of a correctly mutated locus of interest, the remaining sequence sometimes contained additional mutations (indels or substitutions, i.e. sequence 1 and 2 in Fig 5B and sequence 1 in Fig 5C) in the vicinity of the targeted locus.

Differing knock-in efficiency in the LCK and the SH2D2A genes

A total of three different knock-in mutations in *LCK* at two different loci (the codons encoding Y192 and Y505 respectively) were performed in Jurkat T cells. The mutation frequency as estimated by AS-PCR after limiting dilution was between 5-13 % of the total number of clones screened (Fig. 6A). Subsequent genotyping by sequencing revealed that nearly 80% of the AS-PCR positive clones were non-homozygous mutants (heterozygous or hemizygous) or knock-out giving a false-positive signal (Fig 6B). Among the total number of clones isolated by limiting dilution for each independent experiment, on average 2% harboured the Y192E and Y192F and 1% harboured the Y505F homozygous knock-in mutation (Fig. 6C).

CRISPR/Cas9 mediated cleavage of DNA is independent of whether the strand is sense or non-sense with respect to mRNA transcript for the respective proteins. After cleavage of DNA, the target strand remains bound to the DNA binding domain of Cas9, whereas the non-target DNA strand becomes exposed and available for pairing with the ssODN (Richardson et al., 2016) (Supplementary Fig. 1A-C). During the course of these studies, and as suggested by Richardson et al, we thus created asymmetric ssODN that were 103-127-nt long, with the shorter homology arm of the ssODN pairing with the non-target DNA strand 36-nt away from the cut site. We were however unable to detect an improved mutation frequency when using asymmetric ssODN for mutation of the same locus (Fig. 6A and C).

In parallel efforts, we performed experiments to introduce four different knock-in mutations in three different loci (N52, R120

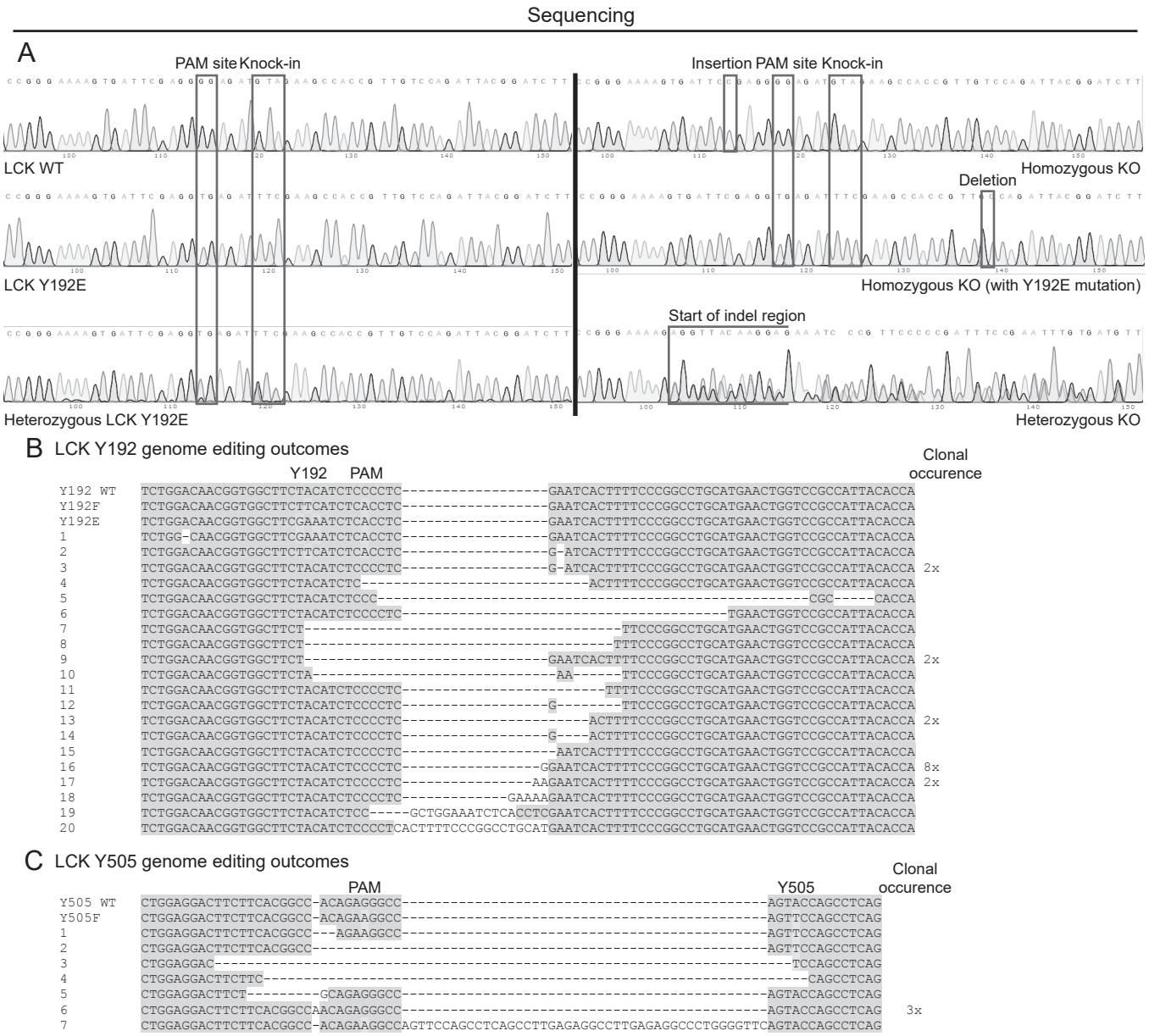


Figure 5. Screening knock-in clones with sequencing. Jurkat T cells were subjected to knock-in mutations for Y192E/F and Y505F in LCK. Amplicons of the region around the targeted loci from selected clones were subjected to Sanger sequencing. (A) Representative sequencing results showing various outcomes of the mutation. (B-C) Examples of various genome editing outcomes revealed by sequencing: (B) Sequences from cells targeted by LCK Y192 sgRNA (aberrant mutation outcomes are labelled 1-20) and (C) targeted by LCK Y505 sgRNA (aberrant mutation outcomes are labelled 1-7). Correctly introduced mutations are highlighted in yellow. Grey highlight represents unchanged WT sequence.

and Y290) in the *SH2D2A* gene. The knock-in mutation frequency after limiting dilution in Y290, as assessed by AS-PCR, was on the average 4% and 6% for the Y290E and Y290F mutations respectively (Fig. 6D-F). Sanger sequencing of amplicons of the mutated regions revealed that on average 2% of the total number of clones screened were homozygous for the knock-in mutation. The homozygous knock-in frequencies for Y290E or Y290F were comparable to the frequencies of the knock-in mutations in the *LCK* gene (Fig. 6B). In contrast, less than 1% of total clones were positive for R120K as determined by AS-PCR (Fig. 6D). Of the two clones that were positive for the knock-in allele by AS-PCR, genotyping by Sanger sequencing did not reveal any homozygous knock-in mutations (Fig. 6E and F). For mutation of N52 to S52, in two independent experiments we were unable to detect any successful knock-in mutations in bulk transfected Jurkat T cells prior to limiting dilution, neither did we observe any AS-PCR positive clones after limiting dilution (data not shown).

Epigenetic gene profiles correspond to knock-in efficiency

Our data show that the CRISPR/Cas9 knock-in mutation efficiency varied considerably both between genes and also between different loci within the same gene. Whereas LCK and TSAd are both expressed in T cells, LCK is constitutively expressed while TSAd expression is upregulated following T cell activation [17]. We therefore speculated that the expression pattern of the genes could have implications for gene editing. CRISPR/Cas9 target cleavage may be inhibited by nucleosomes [28] or other epigenetic mechanisms [29].

We hypothesised that *SH2D2A* is inaccessible to Cas9 in resting cells as a consequence of epigenetic modifications or local chromatin structure. The epigenetic profile of primary T cells from peripheral blood supports this notion (Fig. 6G), as the four exons at the 5' end of *SH2D2A*, encoding S52 and R120, are associated with multiple histone markers [30]. In contrast, the

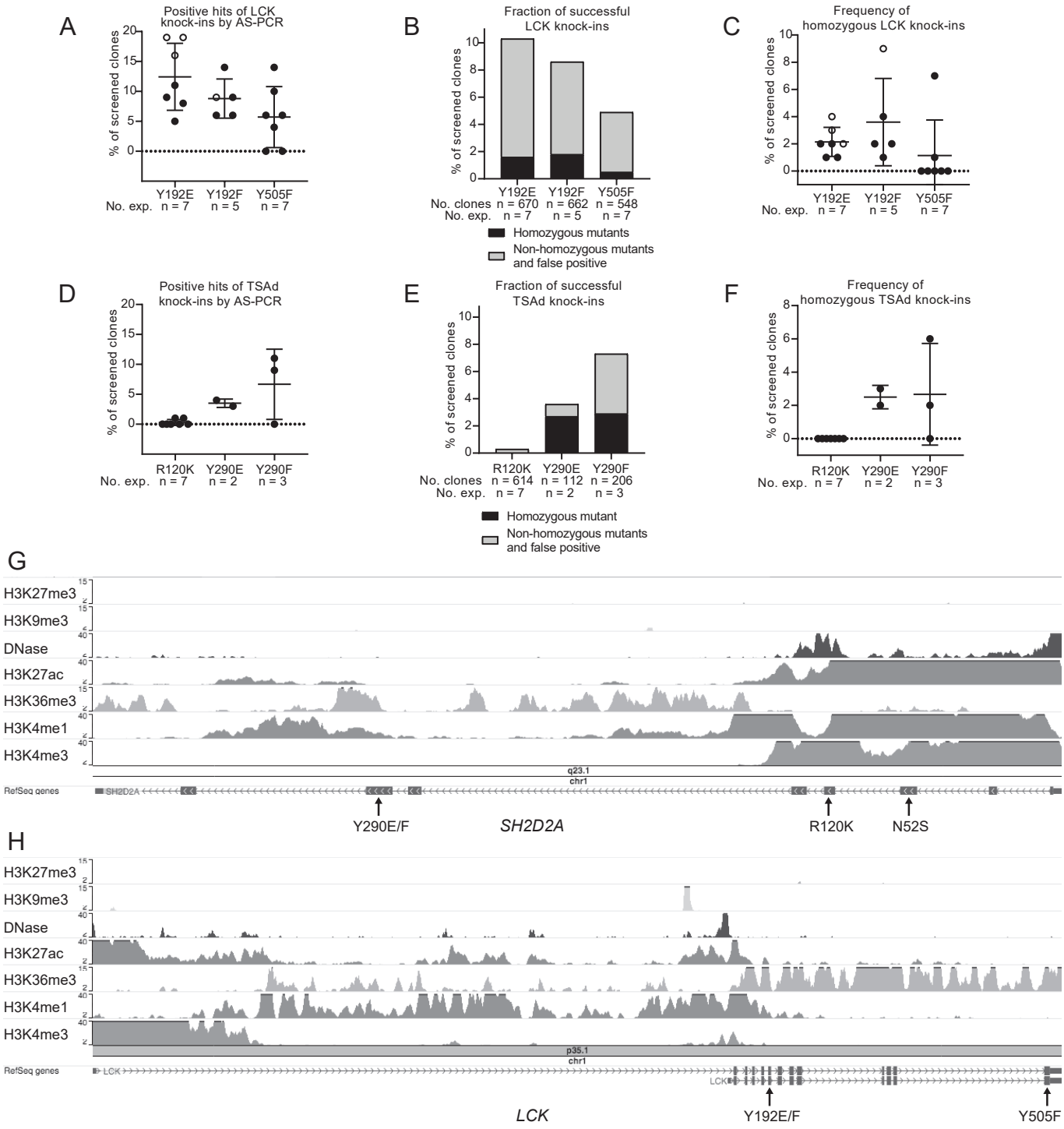


Figure 6. Efficiency of LCK and TSAd knock-ins. Jurkat T cells were subjected to knock-in mutations for Y192E/F and Y505F in LCK (A-C) or for R120K and Y290E/F in SH2D2A (D-F). Experiments where symmetric or asymmetric ssODN were used are indicated in open and filled symbols respectively. (A and D) Graphs show percentage of AS-PCR positive clones for the indicated knock-in mutation experiments. Each point represents one independent experiment. n is displayed below the graph for each targeted mutation. The bars indicate mean values ± standard deviation. (B and E) Graph shows true homozygous mutants as revealed by sequencing for the indicated mutations. Total n of clones and n of the experiments for each knock-in mutation is indicated below the graph. (C and F) Graphs show percentage of true homozygous mutants as confirmed by sequencing for the indicated mutations. Each point represents one independent experiment. n is displayed below the graph. The bars indicate mean values ± standard deviation. (G-H) Epigenetic modifications of (G) LCK and (H) SH2D2A in primary T cells from peripheral blood. The exons targeted for knock-in mutations in this study are indicated by the black arrows. The datasets are available at NIH Roadmap Epigenomics Mapping Consortium [30].

3' exons of *SH2D2A*, including exon 7 which contains the Y290 codon (Fig. 6G) has a different epigenetic profile, resembling that of *LCK* exon 6 and exon 13 encoding Y192 and Y505 respectively (Fig. 6H).

In an attempt to increase the accessibility of Cas9 to the TSAd locus, we activated Jurkat T cells by anti-CD3 or PMA/ionomycin (treatments which induce expression of TSAd in Jurkat cells) before or after transfection for gene editing of TSAd R120. Only one TSAd R120K heterozygous clone was obtained following activation of the Jurkat cells with anti-CD3 antibody prior to CRISPR/Cas9 transfection (Fig. 6E).

Discussion

In this study, we report our experience with gene editing of Jurkat T cells to elucidate domain specific functions of signalling molecules in T cell activation. Further, we report locus dependent knock-in mutation efficiencies, also within the same gene,

which may be due to local variation in chromatin structure. Our workflow for generation of knock-in mutants using the CRISPR/Cas9 system can easily be applied by laboratories equipped to use basic molecular biology techniques.

To successfully introduce a desired knock-in mutation using CRISPR/Cas9, there are two events which must occur concurrently in a precise manner. Firstly, the Cas9 machinery must be targeted to the correct location and secondly the Cas9 endonuclease must facilitate DNA cleavage. This involves accurate directing of Cas9 by the sgRNA to the target site and efficient binding to the PAM site respectively [7].

Unsuccessful CRISPR/Cas9 experiments may be caused by PAM site inaccessibility [31]. Several groups have addressed ways to improve Cas9 target efficiency, for instance by using algorithms to predict efficient sgRNA target sites, according to sequence features [32]. In the case of increasing knock-in efficiency, selecting efficient PAM sites that are <10-nt away from the site of the mutation, was proposed [33].

The limited availability of PAM sites near the mutation site, represent an important challenge for successful generation of knock-in mutations. The frequency of "GG" in the human genome has been reported to be 5,21% [34]. This means that "GG" should occur on the average every 42 bases. Accounting also for "CC" (representing the PAM site on the complementary strand) will give on average one PAM site within a 10 base distance from the mutation site of interest.

Since the first reports of gene disruption in Jurkat T cells using CRISPR/Cas9 [35, 36], there has been an ever increasing number of reports where CRISPR/Cas9 technology has been used to disrupt selected genes in Jurkat T cells (80 publications in PubMed, September 2019). However, to our knowledge, the efficiency of gene editing by HDR in Jurkat T cells using CRISPR/Cas9 technology has not previously been reported. The relatively low homozygous knock-in frequency compared to knockout frequencies observed in our study of Jurkat T cells, is consistent with previously reported data from human primary CD4+ T cells and CD34+ HSPCs [37]. Although the frequencies were low, we repeatedly introduced knock-in mutations across independent experiments with the same experimental setup (Fig. 6 A-C). We have successfully used the resulting mutants to study LCK function in T cell signalling (Borowicz et al., submitted).

Generation of knock-in mutants is time-consuming, and one single experiment takes a minimum of three weeks to complete. It is therefore advantageous to check the performance of the sgRNA already prior to the limiting dilution step, while the cells are still in bulk. However, this strategy also has its limitations. If detection primers are complementary to the ssODN repair template, initial bulk DNA analysis may give false positive results, as the ssODN may still be present in the medium or the transfected cells. This should however not be a problem during the clonal screening, as the ssODN would be sufficiently diluted or degraded by then. Hence, if possible, the detection primers should be non-complementary to the ssODN.

If the targeted site is located in the C-terminus of the protein (as in the case of LCK Y505), the potential mutations may not generate protein knockout, but rather truncated versions of the protein. Additionally, it is important to be aware of the existence of alternative spliced variants of the targeted protein. If the targeted mutation is located in an exon that can be omitted by alternative splicing, the mutation will not deplete the cell of all splicing variants of the targeted gene (as exemplified by TSAd Y290 in exon 7 of *SH2D2A*) [38].

CRISPR/Cas9 mediated genome editing involves a certain risk of generating off-target effects in the treated cells due to unspecific guidance of Cas9 [39]. However, the heterogeneity of the resulting mutated clones may also be a consequence of the inherent heterogeneity of the starting cell lines, as our analysis of the Jurkat T cell sub-clones clearly demonstrated (Fig. 2A-C). To control pre-existing heterogeneity, we thus strongly recommend to generate one or several founder cell lines. It is particularly important to ensure sufficient expression of key molecules crucial in pathways contributing to phenotypes in the knock-in cell line. In our case, establishing founder Jurkat T cell lines was necessary to ensure that the phenotypes of the later knock-in mutations were not biased by the variation in the expression of surface CD3, LCK or TSAd between the mutated clones.

An additional source of unspecific effects of CRISPR/Cas9 mediated gene editing is the stochastic nature of DNA repair mechanisms, which can create substantial local gene rearrangements [40] as we also observed (Fig 5C and D). A striking observation in our experiments was repeated similar repair patterns observed in independent mutation experiments. This fits with the notion that particular outcomes of NHEJ DNA repair to some extent can be anticipated. The software InDelphi [41], predicts the likelihood of given NHEJ results based on microhomology analysis. We have compared our results with InDelphi predictions in Supplementary Figure 5. Consistent with the InDelphi predictions, we observed that the sequences predicted to occur with the highest frequencies, were observed in our sequencing results.

Our workflow proposes methods for mutated clones' validation, which should ensure desired homozygous knock-in mutations and omit clones with unspecific mutations. In some of our knock-in mutants, we observed protein products of various length, which were not present in the WT samples. Even if the expressed knock-in mutated protein seems to be of correct length by Western blotting, DSB increases the chance of introducing random mutations into the gene of interest. Therefore, it should be carefully monitored by sequencing. We observed within the 250 bp sequenced, additional mutations occurring at some distance from the targeted codon. Thus, it is possible that we may have overlooked some additional mutations outside of the sequenced region.

A striking observation from our work, was the variable success rate for CRISPR/Cas9 targeting at different loci in the same gene, which might be explained by variable chromatin accessibility. This could be due to an error in the generation of the sgRNA or Cas9 from our px330 plasmid. However, for the R120 site in *SH2D2A*, we tested three different PAM sites and sgRNA target sites to improve efficiency, while still remaining in proximity to the desired site of mutation. Still, we were unable to detect any successful knock-in mutations. We thus hoped that activation of the Jurkat T cells prior to the mutation experiments, would change local chromatin structure and hence increase the accessibility of Cas9 to the target mutation site. However, despite different activation methods, we were unable to successfully increase the frequency of knock-in mutations at the R120 site and only obtained one heterozygous clone for R120K.

Although activation of Jurkat T cells through TCR prior to gene editing did not improve the mutation efficiency of TSAd R120, it does not exclude the possibility that local chromatin structure is a determining factor. Besides Cas9 accessibility, local chromatin domains might determine which DNA repair pathway is

taken to repair the DSB, i.e. through NHEJ or HDR [29]. Furthermore, it is known that NHEJ occurs more frequently and at a faster rate compared to HDR [8], which may result in HDR machinery being outcompeted by NHEJ. If this is the case, we would expect to see a high frequency of indels at the R120 site, which to a certain degree can be picked up by our screening methods. As we did not observe any abnormalities, it suggests that R120 was not successfully targeted by the sgRNA and Cas9. Several groups have addressed ways to promote HDR pathways, such as calibrating the cells to be in the G2 phase of the cell cycle, where HDR is suggested to be more active than NHEJ repair mechanisms [42]. Alternatively, inhibiting NHEJ pathways [43] might be a possible strategy to enhance HDR mediated knock-in efficiency.

In conclusion, we have successfully used CRISPR/Cas9 technology to target single domains in the *LCK* and *SH2D2A* genes using a setup which is simple and easy to establish for labs with basic skills in molecular biology. Our work reveals variability in mutation frequency, also within the same gene. Inability to target a given locus may be associated with the chromatin structure in the vicinity of the locus affecting sgRNA abilities to guide Cas9 to the desired genomic location.

Materials and Methods

Reagents. The plasmids used were pX330-U6-Chimeric_BB-CBh-hSpCas9 [4], pSpCas9(BB)-2A-GFP (PX458) [3] (Addgene), pEGFP-N1 (Clontech), pEF-HA-TSAd (encoding S52) [16]. sgRNA encoding oligonucleotides targeting the loci of interest were cloned into the pX330 or PX458 vectors. pEF-HA-TSAd-N52 was obtained using Quick Change Mutagenesis (Thermo Fischer Scientific). All new constructs were verified by sequencing. Oligonucleotides for generating sgRNA encoding plasmids, ultramer ssODN for HDR repair, as well as ssODN for PCR and sequencing referred to in the text were custom synthesized by Integrated DNA technologies (Coralville, Iowa). Primary antibodies used were: anti-CD3ε (clone OKT-3, American Type Culture Collection), anti-LCK (unconjugated or conjugated to AF647, clone 3A5, Santa Cruz Biotechnology), anti-SH2D2A (DyLight 488, clone 3C7, Origene), anti-phosphotyrosine (clone 4G10, Upstate Biotechnology), anti-GADPH (clone 6C5, Chemicon) and anti-TSAd antibodies raised against synthetic peptides of TSAd [17, 18] or the recombinant C terminus of TSAd (a generous gift from Virginia Shapiro). Secondary antibodies used were: AF647 goat anti-mouse IgG (H+L) (ThermoFisher), HRP conjugated goat anti-mouse (H+L) and HRP conjugated goat anti-rabbit antibodies (Jackson ImmunoResearch).

Cell cultures. Jurkat TAg cells, expressing large T antigen [19] (Jurkat T cells) were cultured in complete RPMI 1640 media (supplemented with 10% fetal calf serum, 1mM sodium pyruvate, 10mM HEPES, 1mM MEM nonessential amino acids, penicillin (100 U/ml), streptomycin (100µg/ml) (all from GIBCOBRL®, Thermo Fischer Scientific) and 50uM β-mercaptoethanol (Sigma)). To reduce the impact of clonal heterogeneity in the gene editing experiments, Jurkat T cells were subcloned by limiting dilution. Clones were assessed for expression of LCK, TSAd and surface CD3 by Western blotting and flow cytometry. Clones which displayed similar amounts as the median values of bulk Jurkat T cells were selected for the use in subsequent experiments.

Cell stimulation. For activation of Jurkat cells by OKT3, a T75 flask was coated with 10ml of 5µg/ml OKT3 for 1 hour. 1 x 107 cells were then activated at 37ºC, 5% CO2 for 16hrs. For activation by PMA/ionomycin, 1 x 107 Jurkat cells were incubated with 50 ng/ml PMA and 500 ng/ml ionomycin at 37ºC, 5% CO2 for 16hrs. After activation, cells were washed with PBS and resuspended in complete RPMI prior to transfection experiments.

Design of sgRNAs and mutation donor ssODNs. PAM sites where identified both on the sense and the non-sense strand in the immediate vicinity of the loci to mutate. Potential sgRNA (20-23 bp) located 5' to these PAM-sites were identified and evaluated using CRISPOR [20]. From one to three sgRNA sequences for each locus were chosen for experiments. 103-127 nt long ssODNs complementary to the non-target strand were designed based on the genomic sequence centred on the locus to mutate. Some of the ssODNs were symmetric with the Cas9 cut site located in the middle of the ssODN [4], while later in the course of experiments asymmetric ssODNs, with the Cas9 cut site approximately 36 nucleotides (nt) away from the cut site on the 5'-end of the non-target strand, were generated [12]. All ssODNs contained the desired codon and mutations of the PAM site when necessary.

Cell transfection. $1x10^7$ Jurkat T cells were transfected with 0,5µg of pEGFP-N1 plasmid, 5 µg of sg-RNA-pX330 plasmid as well as 1 µM of donor ssODN (IDT) (containing the desired mutation at the targeted locus as well as the adjacent PAM site) at 240V for 25 ms using a BTX electroporator (Genetronix). Three days after transfection, a limiting dilution in 96-well plates with 1 cell per well was performed. After two-three weeks of expansion, clones were screened for the defined knock-in mutation by allele-specific PCR. Successful knock-ins were assessed for CD3 expression as well as for the protein of interest using flow cytometry and Western blotting.

Genomic DNA extraction. To obtain genomic DNA for genotyping, Jurkat T cells were lysed in Tris-EDTA buffer (pH 8.0), 0,1% SDS (BioRad) and incubated with 0,2µg/µl proteinase K (Sigma) at 55ºC for 3 hours and subsequently at 80ºC for 10 min to deactivate proteinase K. Cell lysates were diluted 20 times with purified water before used for genotyping. DNA from nucleated blood cells of healthy anonymous donors from the Oslo University hospital blood bank was isolated using rapid salting out method [21].

Genotyping by allele specific-PCR and sequencing. Allele specific PCR (AS-PCR) assays were used to identify knock-in mutations. Wild-type (WT) and knock-in specific primers differed in the 3' end of the primers by two or more nucleotides, representing the WT or mutated codon sequence or the PAM site mutation. In most cases, to increase the specificity of the primer, a mismatch was introduced 3-4 nucleotides away from the 3' end of the primer. To increase the specificity of the PCR reaction, in some experiments betaine (final concentration 1.3M) was added. The PCR protocol with Taq Polymerase (PROMEGA) used for all AS-PCRs was: 95ºC for 2 min, 40 cycles for 95ºC for 15 s, 58ºC for 15 s, 72ºC for 15 s and final extension at 72ºC for 7 min. Resulting PCR products were analysed by electrophoresis at 90V for 30min on a 2% agarose gel. Samples for Sanger sequencing, to confirm knock-in mutations, were prepared by using primers flanking the knock-in region and the same PCR protocol as described above. Amplicons were purified with Wizard SV Gel and PCR Clean-Up System (PROMEGA) and custom sequenced (GATC Biotech).

Sequence analysis of heterozygous mutants. The sequencing results were translated by the DNA Baser Assembler (Heracle

BioSoft) to the standard nucleotide ambiguity codes and corrected manually, if necessary. The resulting sequences were aligned to the WT sequence of the targeted gene using pairwise sequence alignment with EMBOSS Needle. Heterozygous sequences were deconvoluted by manipulating the parameters of the algorithm and by manual rearrangements.

Flow cytometry. Cells were barcoded with three different concentrations of CellTrace Violet (ThermoFisher) according to the manufacturer's protocols. Different clones were either kept unlabelled, or stained with the varying concentrations of CTV and combined respectively. Surface stain (CD3): The combined samples were incubated for 20 min on ice with anti-CD3ε (OKT3) and fixed with 2% paraformaldehyde. Subsequently, fixed cells were incubated with AF647 goat anti-mouse IgG (H+L) antibodies. Intracellular stain (TSAd, LCK): The combined samples were fixed with 2% paraformaldehyde at room temperature, permeabilised with 0,1% saponin for 15 min and subsequently stained for 1h at room temperature with antibodies: anti-LCK (AF647), anti-SH2D2A (DyLight 488). Barcoded, fluorescently labelled samples were analysed with FACSCanto II flow cytometer (BD biosciences) and FlowJo software (Tree Star, Ashland).

Lysis and Western blotting. Stimulated or treated cells were pelleted and lysed with 0.1% LDS/ 1% Triton solution containing 0,5% Triton X-100 (VWR International S.A.S), 50mM HEPES (GIBCOBRL®, ThermoFisher), 0,05% LDS (Merck), 0.05 M LiCl, 0.5 mM PMSF, 2.5 mM EDTA (pH 8.0), 1mM sodium vanadate and 1x SIGMAFAST protease inhibitor cocktail (all from Sigma). Cells were lysed for 45 min on ice. Lysates were sonicated briefly to break the DNA. The appropriate volume of loading buffer (containing 0.35 M Tris HCl, 10% SDS, 6% β-mercaptoethanol (all Sigma), 30% glycerol (VWR international S.A.S), 0.175 mM bromophenol blue (Fluka Ag), pH 6.8) was added to the samples. All samples were denatured by boiling at 95°C for 10 min. Denatured samples were run with SDS-PAGE and transferred onto PVDF membranes (Bio-Rad laboratories) with Trans-Blot Turbo Transfer System (Bio-Rad laboratories). The membranes were incubated with primary antibodies diluted in tris-buffered saline (pH 7.4), 0.1% Tween (Sigma-Aldrich) and 3% skimmed milk or 3% bovine serum albumin (Bio-Rad laboratories) (when combined with anti-pTyr antibodies). After overnight incubation at 4°C, the membranes were incubated with appropriate HRP-conjugated secondary antibody for 1h at room temperature. SuperSignal West Pico Stable Peroxide Solution (Pierce) was used to visualize bands using ChemiDocTM Imaging System (Bio-Rad Laboratories).

Author Contributions: A.S., P.B. and H.C. conceptualised the study. A.S., P.B., H.C., D.M. and S.G. designed experiments and analysed results. P.B. performed Western Blotting and flow cytometry experiments. P.B., H.C. D.M. and S.G. generated mutated and founder cell lines. P.B. deconvoluted mutated sequences. H.K. performed molecular cloning of all necessary constructs. A.S., P.B. and H.C. wrote the manuscript with the input from all co-authors. A.S. oversaw the project.

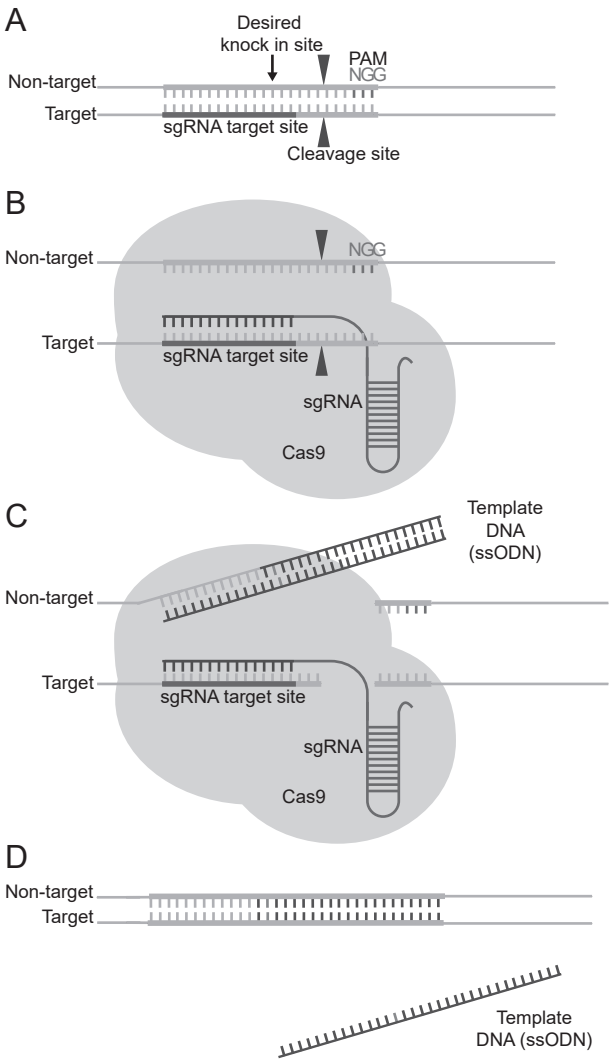
Competing interests statement: The authors declare no competing interests.

Funding: This work was supported by the University of Oslo, the Norwegian Cancer Society (grant 107561 - PR-2007-0193), Novo Nordisk Fonden, Unifor and Anders Jahres fond til vitenskapens fremme.

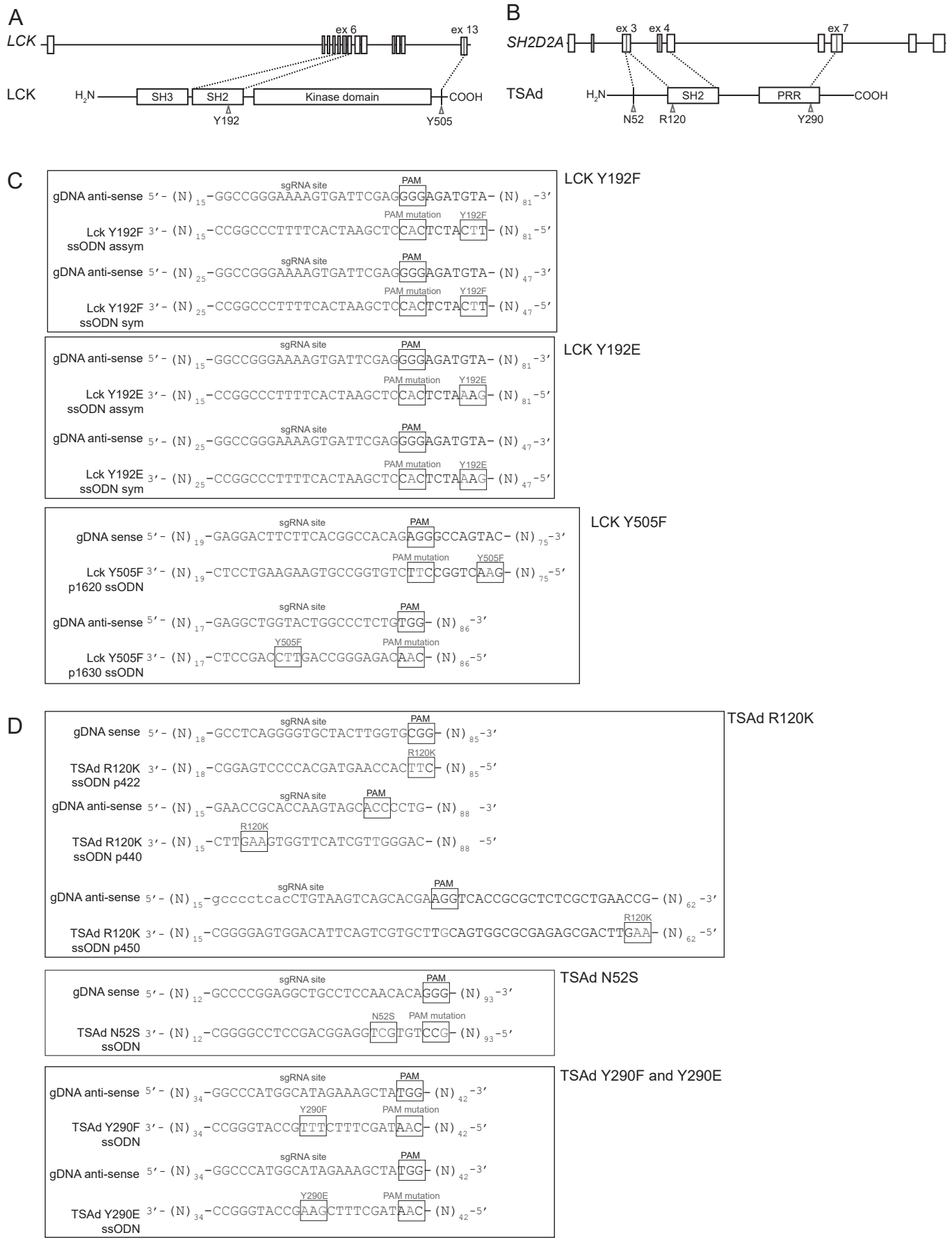
References

1. Urnov, F.D., Genome Editing BC (before CRISPR): lasting lessons from the "old testament". The CRISPR Journal, 2018. 1(1): p. 34-46.
2. Barrangou, R., et al., CRISPR provides acquired resistance against viruses in prokaryotes. Science, 2007. 315(5819): p. 1709-12.
3. Ran, F.A., et al., Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 2013. 8(11): p. 2281-2308.
4. Cong, L., et al., Multiplex genome engineering using CRISPR/Cas systems. Science, 2013. 339(6121): p. 819-23.
5. Hsu, P.D., E.S. Lander, and F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. 2014. p. 1262-1278.
6. Sander, J.D. and J.K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol, 2014. 32(4): p. 347-55.
7. Jinek, M., et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (New York), 2012.
8. Mao, Z., et al., Comparison of nonhomologous end joining and homologous recombination in human cells. DNA repair, 2008. 7: p. 1765-71.
9. McVey, M. and S.E. Lee, MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet, 2008. 24(11): p. 529-38.
10. Shalem, O., et al., Genome-scale CRISPR-Cas9 knockout screening in human cells. Science, 2014. 343: p. 84-87.
11. Sternberg, S.H. and J.A. Doudna, Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol Cell, 2015. 58(4): p. 568-74.
12. Richardson, C.D., et al., Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol, 2016. 34(3): p. 339-44.
13. Granum, S., et al., The kinase Itk and the adaptor TSAd change the specificity of the kinase Lck in T cells by promoting the phosphorylation of Tyr192. Science Signaling, 2014. 7: p. ra118-ra118.
14. Granum, S., et al., Modulation of Lck function through multisite docking to T cell-specific adapter protein. Journal of Biological Chemistry, 2008. 283: p. 21909-21919.
15. Kaplun, L., et al., Variants in the signaling protein TSAd are associated with susceptibility to ovarian cancer in BRCA1/2 negative high risk families. Biomarker Insights, 2012. 7: p. 151-157.
16. Sundvold, V., et al., T cell-specific adapter protein inhibits T cell activation by modulating Lck activity. J.Immunol., 2000. 165(6): p. 2927-2931.
17. Spurkland, A., et al., Molecular cloning of a T cell-specific adapter protein (TSAd) containing an Src homology (SH) 2 domain and putative SH3 and phosphotyrosine binding sites. Journal of Biological Chemistry, 1998. 273: p. 4539-4546.
18. Sundvold-Gjerstad, V., et al., The C terminus of T cell-specific adapter protein (TSAd) is necessary for TSAd-mediated inhibition of Lck activity. Eur J Immunol, 2005. 35(5): p. 1612-20.
19. Clipstone, N.A. and G.R. Crabtree, Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. Nature, 1992. 357(6380): p. 695-7.
20. Haeussler, M., et al., Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biology, 2016. 17: p. 148.
21. Koshy, L., et al., Evaluating genomic DNA extraction methods from human whole blood using endpoint and real-time PCR assays. Molecular biology reports, 2017. 44(1): p. 97-108.
22. Snow, K. and W. Judd, Heterogeneity of a human T-lymphoblastoid cell line. Exp Cell Res, 1987. 171(2): p. 389-403.
23. Ugozzoli, L. and R.B. Wallace, Allele-specific polymerase chain reaction. Methods, 1991. 2: p. 42-48.
24. Gaudet, M., et al., Allele-specific PCR in SNP genotyping. Methods Mol Biol, 2009. 578: p. 415-24.
25. Kwok, S., et al., Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res, 1990. 18(4): p. 999-1005.
26. Huang, M.-M., N. Arnheim, and M.F. Goodman, Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Research, 1992. 20: p. 4567-4573.
27. Lajin, B., A. Alachkar, and A. Alhaj Sakur, Betaine significantly improves multiplex tetra-primer ARMS-PCR methods. Mol Biotechnol, 2013. 54(3): p. 977-82.
28. Yarrington, R.M., et al., Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proceedings of the National Academy of Sciences, 2018. 115: p. 9351 LP - 9358.
29. Janssen, J.M., et al., The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. Molecular Therapy - Nucleic Acids, 2019. 16: p. 141-154.
30. Bernstein, B.E., et al., The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol, 2010. 28(10): p. 1045-8.
31. Uusi-Mäkelä, M.I.E., et al., Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). 2018.
32. Doench, J.G., et al., Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology, 2016. 34(2): p. 184.
33. Paquet, D., et al., Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature, 2016. 533.
34. Scherer, S., short guide to the human genome. 2008: Cold Spring Harbor

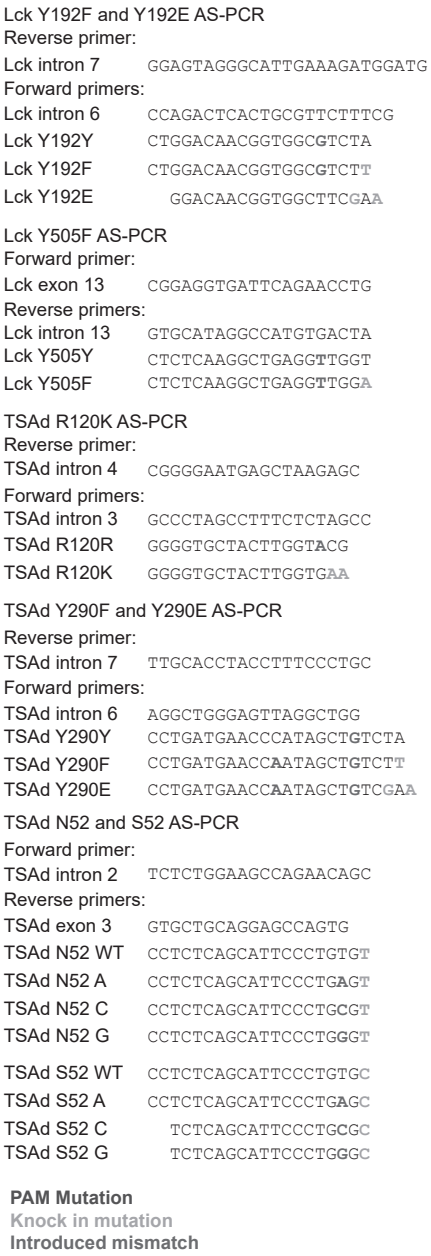
Laboratory Press.
35. Navarro, J.-M., et al., Site-and allele-specific polycomb dysregulation in T-cell leukaemia. Nature communications, 2015. 6: p. 6094.
36. Neggers, J.E., et al., Identifying drug-target selectivity of small-molecule CRM1/XPO1 inhibitors by CRISPR/Cas9 genome editing. Chemistry & biology, 2015. 22(1): p. 107-116.
37. Mandal, Pankaj K., et al., Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9. Cell Stem Cell, 2014. 15: p. 643-652.
38. Granum, S., et al., Structure function analysis of SH2D2A isoforms expressed in T cells reveals a crucial role for the proline rich region encoded by SH2D2A exon 7. BMC Immunology, 2006. 7.
39. Fu, Y., et al., High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology, 2013. 31: p. 822-826.
40. Kosicki, M., K. Tomberg, and A. Bradley, Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nature biotechnology, 2018. 36(8): p. 765.
41. Shen, M.W., et al., Predictable and precise template-free CRISPR editing of pathogenic variants. Nature, 2018. 563(7733): p. 646-651.
42. Gu, K., et al., Methodologies for Improving HDR Efficiency. Frontiers in Genetics, 2019. 9: p. 1-9.
43. Maruyama, T., et al., Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nature Biotechnology, 2015. 33: p. 538-542.



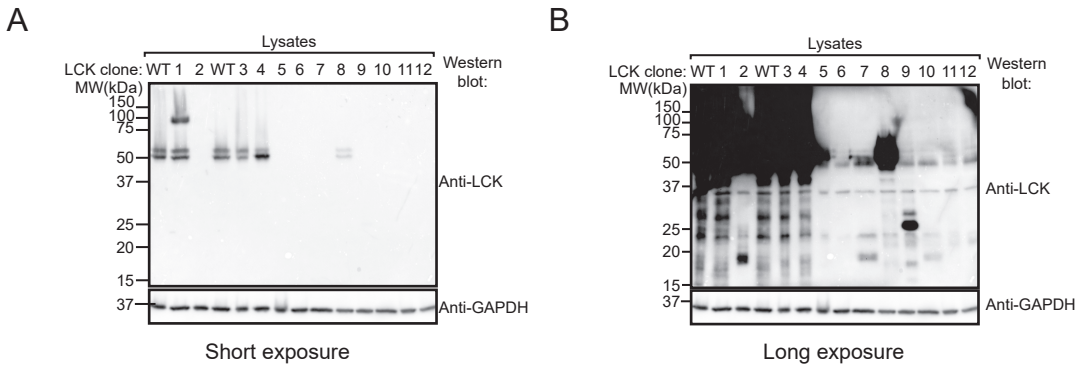
Supplementary Figure 1. Schematic depiction of CRISPR/Cas9 system mediating targeted point mutations. (A) When designing the CRISPR/Cas9, an optimal target site for the sgRNA to direct the Cas9 and a PAM site (pink) positioned in the non-target strand, directly downstream of the sgRNA target site, need to be identified. (B) When expressed in the cell, Cas9 (grey) complexes with the sgRNA, which contains the complementary target sequence (blue). The Cas9-sgRNA complex is directed to the target DNA strand and Cas9 interacts with the PAM site. (C) Cas9 mediates double strand DNA cleavage 3-4 bases upstream of PAM. Exonucleases degrade the exposed overhangs of the cleaved non-target strand. Homologous arms of the ssODN repair template bind to the non-target strand and repair based on the ssODN occurs. (D) The ssODN repair template containing the desired knock-in mutation is used to repair the DSB through homology directed repair, resulting in both non-target and target strand containing the desired mutations.



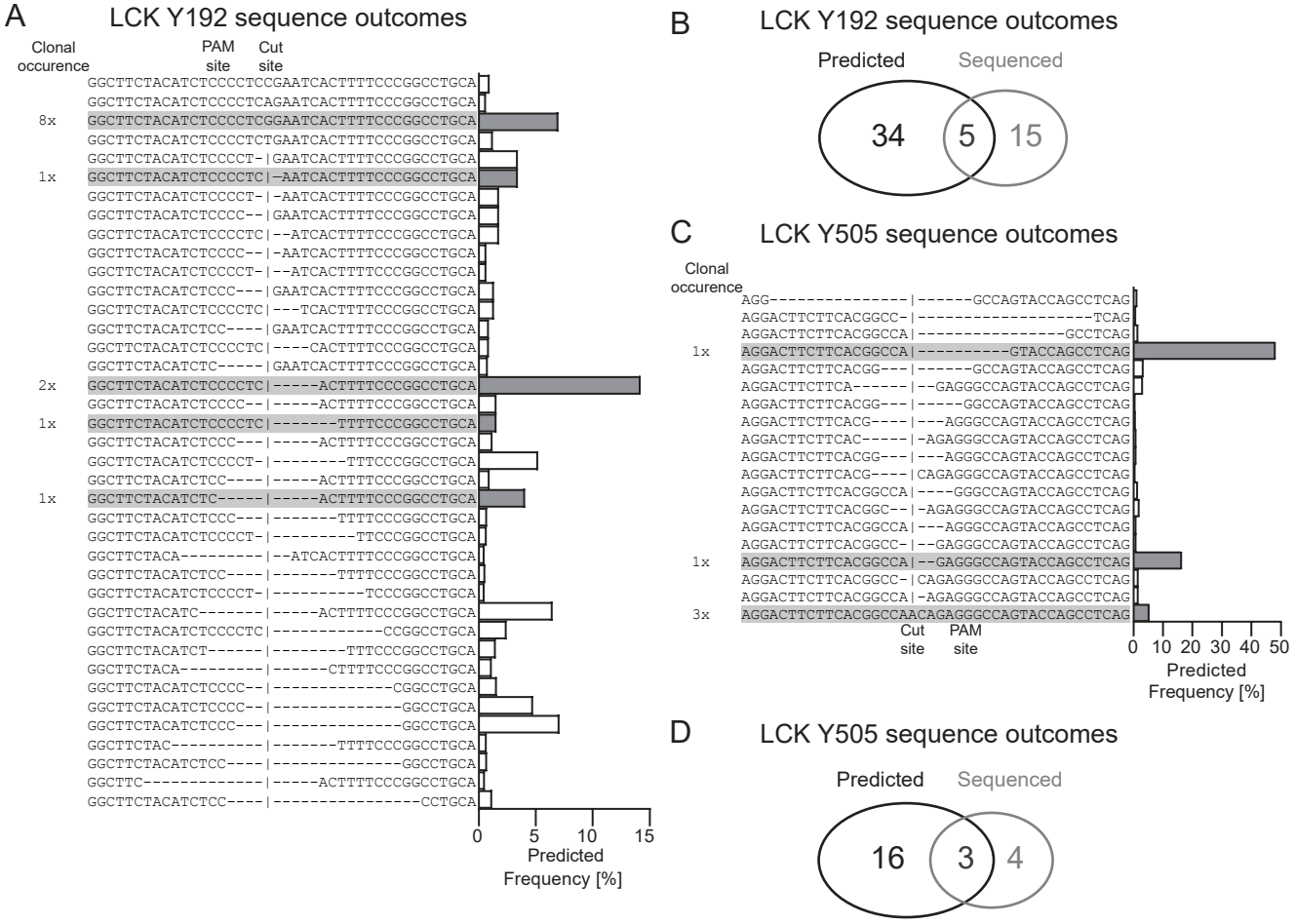
Supplementary Figure 2. Design of knock-in strategies for LCK and SH2D2A. (A) Knock-in mutations in exon 6 and exon 13 of LCK results in an amino acid substitution at positions Y192 and Y505 respectively in the LCK protein. (B) Knock-in mutations in exon 3, exon 4 and exon 7 of SH2D2A results in an amino acid substitution at positions N52, R120 and Y290 respectively in the TSAd protein. (C-D) Non-target strand containing the PAM site (gDNA) and the homologous repair donor template (ssODN) are indicated for Y192F, Y192E and Y505F for LCK knock-in mutations (C) and N52S, R120K, Y290E and Y290F for SH2D2A knock-in mutations (D). Whether the non-target strand is sense or non-sense with respective to the gene transcript is annotated.



Supplementary Figure 3. The list of AS-PCR primers used in this study. AS-PCR primers used to detect WT and knock-in mutations are listed for the respective knock-in mutations. All are listed in the 5' to 3' orientation. The exon/intron primers are the common forward/reverse primers which are matched with the detection forward/reverse primer for WT or the knock-in mutation. PAM mutations are highlighted in purple, knock-in mutations in green and introduced mismatches in red. When amplicons were prepared for sequencing, the intron/exon primers flanking the knock-in site were used.



Supplementary Figure 4. CRISPR/Cas9 genome editing creates protein products of abnormal length. (A-B) Full size images of the anti-Lck blot shown in Fig.4C. (A) Short and (B) long exposure.



Supplementary Figure 5. CRISPR/Cas9 genome editing outcomes predicted by InDelphi. (A and C) Frequency of the sequence outcomes predicted by the InDelphi software according to the parameters: lowest frequency 0.5%, indels ranging from +1bp to -60bp, microhomology deletions, microhomology-less deletions. Grey highlight indicates sequence outcomes discovered in our experiments, which corresponded to the sequences predicted by InDelphi. (B and D) Venn diagrams show the summary of the number of sequence outcomes that overlap between the InDelphi prediction and our results. (A and B) LCK Y192 target site. (C and D) LCK Y505 target site.