Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Annika Nicole Brosig, M.Sc.

born in Ostfildern, Germany

Oral examination: February 12th, 2021

The RNA interactome of pediatric T-cell leukemia:

Identification of alterations in the RNA-bound proteome of T-ALL in the process of leukemogenesis and treatment with the HDAC inhibitor Vorinostat

Referees: Prof. Dr. Anne-Claude Gavin

Prof. Dr. Michael Brunner

This work has been carried out in the Molecular Medicine Partnership Unit (MMPU) at the European Molecular Biology Laboratory (EMBL) Heidelberg and the Medical Faculty of the University of Heidelberg from July 2016 to November 2020 supervised by Prof. Dr. med. Andreas Kulozik, PhD, and Prof. Dr. med. Matthias Hentze.

- Yes, it is significant – but is it relevant?

(Andreas Kulozik, group meeting)

ACKNOWLEDGEMENTS

I would like to thank everyone who has supported and helped me in accomplishing this PhD project. In particular, I am grateful to:

My supervisors **Andreas Kulozik** and **Matthias Hentze** for selecting me as a PhD student and giving me the opportunity to work in the inspiring environment of the MMPU, the EMBL and the University Hospital. Thank you for your advice and support throughout these four years and for guiding me through this phase of my career.

My Thesis Advisory Committee members **Anne-Claude Gavin**, **Michael Brunner**, **Wolfgang Huber** for great advice, valuable scientific input and lively discussions during my TAC meetings. Further, I would like to thank Anne-Claude and Michael for agreeing on reviewing my thesis, and Wolfgang and **Frauke Melchior** for being members on my Thesis Defence Committee.

The **Merck´sche Gesellschaft für Kunst und Wissenschaft**, **Christine Esdar** and **Stefan Becker** for supporting my research project with the PhD Scholarship in Translational Medicine.

Gaby Neu-Yilik for her support and her always honest and direct advice and feedback. Also, thank you for carefully reading my thesis: I will probably forever remember the "ganz heller weißer Schimmel" when trying to write something "knackig" in the future.

Beate Amthor, der guten Seele des Labors. Danke für deine Hilfe, deine Ruhe, dein Verständnis und dafür, dass es uns nie an Materialien mangelt, und dass Geräte immer gleich repariert werden, und und und... ☺

Yang Zhou, **Jonas Becker**, **Dasha Lavysh**, **Michael Backlund**, **Claudia Gruber** for helping out whenever my two hands were not enough for all my samples, for various funny lunch, coffee/tee and beer breaks, for unforgettable nights out and for sharing long miserable eRIC nights. Because of you the lab was more than just a pure working place. Here, my very special thanks go to Yang, we two shared more than just our lab room. Thank you for all your help, your dry humor, chatting with me about science and all the world and his wife… and that it also never felt uncomfortable to work next to you without talking for h.

All the **members of the Kulozik, Muckenthaler and Roth lab** and everyone else who helped me and with whom I shared fun times. Thank you, **Margit Happich**, **Gabi Tolle**, **Paulina Richter-Pechanska**, **Nicole Dickemann**, **Jasmin Keck**, **Chuli Fu**. Thank you **Büsra Erarslan-Uysal** for your company on courses and conferences and for making them so much more enjoyable ☺. And thank you **Katja Müdder** for always figuring something out to make it possible to use the flow cytometer or the Fusion although they were completely booked out, aaaand for all your sweets, muffins, cakes…

The **Proteomics Core Facility**: Thank you **Mandy Rettel** for processing all my samples as fast as possible, in particular directly before and after the EMBL shut-down and despite of the reduced capacity. The same for you **Frank Stein**, thank you for supporting my data analysis in spite of being home with your twins, your patients when helping me with graphs, plots and figures via email and on the phone. Also, thanks to you, **Jianguo Zhu**, for helping out with the disordered plots.

The **Flow Cytometry Core Facility**: Many thanks to **Malte Paulsen** and **Diana Ordonez** for teaching me everything about flow cytometry, MACS and FACS, for your help with setting up protocols and for spontaneously finding free slots for thymocyte analyses.

To **all members of the Hose Group** who teached me separating thymocytes on their auto-MACS systems and who supported me even when it was already way past their working hours.

All the EMBL and university staff (in particular **Esther Liebig**, **Christine Georgi-Turcotte**, **Jutta Mattern**, **Bärbel Fritz**, **Monika Lachner**, **Matija Grgurinovic**, **Vira Beck**) who helped me to survive bureaucracy whenever faced.

To my previous supervisors **Paul Wan**, **Helen Boyd** and **Joe Lewis** for sparking my interest in science, motivating me to pursue a PhD and supporting me in finding a PhD position.

To people of **my batch** and **friends in- and outside of EMBL or the MMPU**: **Mareike**, **Evelyn**, **Cora**, **Marina**, **Ricarda**, **Sanjana**, **Renato**, **Ashna**, **Areeb, Magdalena** and many I probably forgot – sorry for that! Thank you for detaching my brain from science during dinners out, Christmas markets or vacations and travel, and may other events. Thanks for your understanding when I got late due to delays in the lab or had to cancel last minute – once more: sorry! Thank you for all your motivating words. Thank you for thesis writing emergency packages Mareike and Ricarda

Und schließlich möchte ich mich bei meiner **Familie** bedanken, bei meinen **Eltern**, meinem **Bruder**, meinen **Großeltern**, meinen **Paten** und dem ganzen Rest der „duppligen" Verwandtschaft (wie meine Tante euch immer alle liebevoll zusammenfasst ☺). Ihr alle habt mich immer unterstützt wo ihr nur konntet, wart für mich da, habt an mich geglaubt und mich aufgefangen, wenn es mit irgendetwas doch mal nicht so lief. Vielen Dank euch allen! Ohne euch wäre ich jetzt nicht da wo ich bin!

Ganz besonderer Dank gilt meinem Mann **Sebastian**. Danke, dass du mich schon die ganze lange Achterbahnfahrt durch das Studium, die Auslandsaufenthalte und den PhD begleitest. Wir haben schon einiges gemeinsam gemeistert und du hast es bestimmt nicht einfach mit mir ;) Danke, dass du immer zu mir stehst, dich mit mir freust und lacht, mich unterstützt, mich auffängt und mich während diesem PhD immer wieder motiviert hast. Ich liebe dich!

Thank you all very much!

Vielen Dank euch allen!

ABSTRACT

Pediatric T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological cancer characterized by excessive proliferation of malignant T-lymphocyte progenitors. Current treatment regimens achieve cure rates of more than 80 % but include aggressive high-dose, multi-agent chemotherapies causing severe adverse effects. Further, T-ALL represents a particular challenge in pediatric oncology, as the prognosis in the event of a relapse is poor and only 10 – 20 % of patients survive. Therefore, the search for new targets that impact on T-ALL and the development of new treatment options with less toxicity and a reduced risk of relapse is of high clinical importance. To date, research on T-ALL has mainly provided insights on the transcriptional level of the disease. However, an increasing number of studies report posttranscriptional dysregulation; a level of regulation governed by RNA-binding proteins (RBPs). Hence, this work aimed at uncovering RBPs that may be involved in the origin and development of T-ALL and have so far remained unknown. For this purpose, the method "enhanced RNA interactome capture" (eRIC) was carefully developed further and for the first time an RNA interactome was obtained from freshly isolated human tissue, the thymus. This normal thymocyte RNA interactome served as a counterpart to RNA interactomes derived from cell lines representing T-ALL. The RNA interactomes of normal thymocytes and T-ALL cells differ profoundly in size, but present typical characteristics of T-cells such as low metabolic activity, a remarkably high number of cytoskeletal proteins, and proteins related to immune or T-cell development, activation and function. Interestingly, a number of previously not captured RBPs and proteins so far not known to bind RNA were identified (e.g. SASH3, CAPZB, MIA3). As a possible model for the dynamic alteration of an RNA interactome after chemotherapy, the T-ALL RNA interactome was also analyzed upon treatment with the histone deacetylase (HDAC) inhibitor SAHA. Several RBPs with altered RNA binding activity were found, including some eukaryotic initiation factors, in particular EIF4A1. An ultimate and direct comparative analysis of the RNA interactomes indicates a globally increased RNA-binding activity in T-ALL compared to normal thymocytes. In conclusion, this work comprises the first comprehensive investigation of the RNA-binding proteomes of T-ALL and normal T-cell progenitors, and thus, offers unprecedented insights into alterations in the RBP-landscape upon leukemogenesis.

ZUSAMMENFASSUNG

Bei der akuten T-Zell Lymphoblastischen Leukämie (T-ALL) handelt es sich um eine aggressive Art von Blutkrebs, der die klonale Ausbreitung maligner Vorläuferzellen von T-Lymphozyten zugrunde liegt. Derzeitige Behandlungsansätze erzielen bei 80 % der betroffenen Kinder und Jugendlichen zwar einen dauerhaften Erfolg, bestehen aber aus der Kombination verschiedener, aggressiver und hoch dosierter Chemotherapeutika und führen zu schweren Nebenwirkungen. Zudem stellt T-ALL in der pädiatrischen Onkologie eine besondere Herausforderung dar, da bei Auftreten eines Rezidivs die Prognose schlecht ist und nur 10 - 20 % der Patienten überleben. Aufgrund dessen ist die Erforschung neuer Zielstrukturen und die Entwicklung effektiverer Medikamente mit geringerer Toxizität und verringertem Rückfallrisiko von hoher klinischer Relevanz. Die bisherige Erforschung von T-ALL ermöglicht hauptsächlich einen Einblick in die transkriptionelle und epigenetische Ebene der Erkrankung. Allerdings berichten immer mehr Studien auch über Regulationsstörungen auf posttranskriptioneller Ebene; einer Regulationsebene die von RNA-bindenden Proteinen (RBPs) dominiert wird.

Ziel dieser Arbeit war es, bisher unentdeckte RBPs zu identifizieren, die an der Entstehung und Entwicklung von T-ALL beteiligt sind. Zu diesem Zweck wurde die Methode „enhanced RNA interactome capture“ (eRIC) sorgfältig weiterentwickelt und erstmals ein RNA Interaktom aus frisch isoliertem menschlichem Gewebe, dem Thymus, gewonnen. Dieses normale, aus gesundem Gewebe gewonnene Thymozyten RNA Interaktom diente anschließend als Pendant zu Interaktomen, die aus T-ALL Zelllinien gewonnen wurden. Die RNA-Interaktome von normalen Thymozyten und T-ALL-Zellen unterscheiden sich deutlich in ihrer Größe, präsentieren jedoch typische Eigenschaften von T-Zellen wie eine geringe metabolische Aktivität, eine auffällig hohe Anzahl an Proteinen des Zytoskeletts und Proteine, die mit Immun- bzw. T-Zellentwicklung, -aktivierung und -funktion im Zusammenhang stehen. Interessanterweise wurden auch einige Proteine identifiziert, die bisher noch nicht als RBPs erfasst wurden oder deren RNA-bindende Funktion bisher noch nicht bekannt war (z.B. SASH3, CAPZB, MIA3). Als ein mögliches Modell für die dynamische Veränderung eines RNA-Interaktoms nach einer Chemotherapie wurde zudem das T-ALL RNA-Interaktom nach der Behandlung der Zellen mit dem Histon-Deacetylasen (HDAC) Inhibitor SAHA analysiert. Dabei wurden mehrere RBPs mit veränderter RNA-Bindeaktivität gefunden; unter anderem einige eukaryotischen Initiationsfaktoren (eIFs), insbesondere EIF4A1. Ein direkter Vergleich der beiden RNA

Interaktome in einer gemeinsamen Analyse deutet darauf hin, dass die globale RNA-Bindeaktivität bei T-ALL gegenüber normalen Thymozyten erhöht ist. Schlussendlich beinhaltet diese Dissertation die erste umfassende Untersuchung von RNA-bindenden Proteomen von T-ALL und normalen Thymozyten und ermöglicht dadurch völlig neue Einblicke in die Veränderungen der RBP-Zusammensetzung nach der Leukemogenese.

ABBREVIATIONS

7AAD	7-aminoactinomycin D
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
APC	Allophycocyanin
APL	acute promyelocytic leukemia
BSA	Bovine serum albumin
BV605	BD Horizon Brilliant™ Violet dye 605
cDNA	Complementary DNA
CHX	Cyclohexamide
CLL	Chronic lymphoblastic leukemia
DAVID	Database for annotation, visualization and integrated discovery
DC	Detergent compatible
DC	Cluster of differentiation
DCs	Dendritic cells
DMSO	Dimethyl sulfoxide
DN	Double negative
DP	Double positive
dsRBD	Double-stranded RNA-binding domain
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eRIC	Enhanced RNA interactome capture
EtOH	Ethanol
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
FC	Fold change
FDR	False discovery rate
FITC	Fluorescein isothiocyanate
FMO	Fluorescence Minus One
FSC	Forward scatter

gDNA	Genomic DNA
GO	Gene ontology
HRP	Horseradisch peroxidase
HuR	Hu-antigen R (also: ELAV-like protein 1)
K_{Ac}	Acetyl lysine
KH	K Homology domain
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography–tandem MS
MACS	Magnetic Activated Cell Sorting
MHC	Major histocompatibility complex
MS	Mass spectrometry
NaCl	Sodium chloride
o/n	Over night
PAZ	Piwi/Argonaute/Zwille
PBS	Phosphate-buffered saline
PBS	Phosphate-buffered saline
PE	Phycoerythrin
PFA	Paraformaldehyde
PI	Propidium iodide
PS	Phospholipid phosphatidylserine
PTM	Posttranslational modification
PVDF	Polyvinylidene difluoride
qPCR	Quantitative polymerase-chain-reaction
Rae1	messenger RNA export factor
RBD	RNA-Binding domain
RBP	RNA-binding protein
RGG	Arginine/glycine domain
RIC	RNA interactome capture
RIPA	Radioimmunoprecipitation assay buffer
RNP	Ribonuceloprotein particle
RRM	RNA recognition motif
RT	Room temperature

SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acids in cell culture
SP	Single positive
SSC	Side scatter
T-ALL	T-cell acute lymphoblastic leukemia
T-ALL	T-cell acute lymphoblastic leukemia
TBST	Tris-buffered saline with Tween20
TC	Tissue culture
TCR	T-cell receptor
TEC	Thymic epithelial cells
TSA	Trichostatin A
UnR	Upstream-of-N-Ras (also: Cold shock domain-containing protein E1)
UTR	Untranslated region
V(D)J	Variable-Diversity-Joining
ZnF	Zink finger

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 T-cell ontogeny and leukemogenesis

The process of development and production of all blood and immune cells is referred to as hematopoiesis (**Figure 1**)[1]. Pluripotent hematopoietic stem cells are located in the bone marrow and produce common myeloid and lymphoid progenitors that give rise to the myeloid and lymphoid lineage[2]. The myeloid lineage comprises amongst others red blood cells, platelets and myeloblasts that further differentiate into granulocytes (neutrophil, eosinophil, basophil) and monocytes/macrophages. For the lymphoid lineage, lymphoid progenitors further develop to lymphoblasts and differentiate into B-cells, T-cells and natural killer (NK) cells.[3,4]

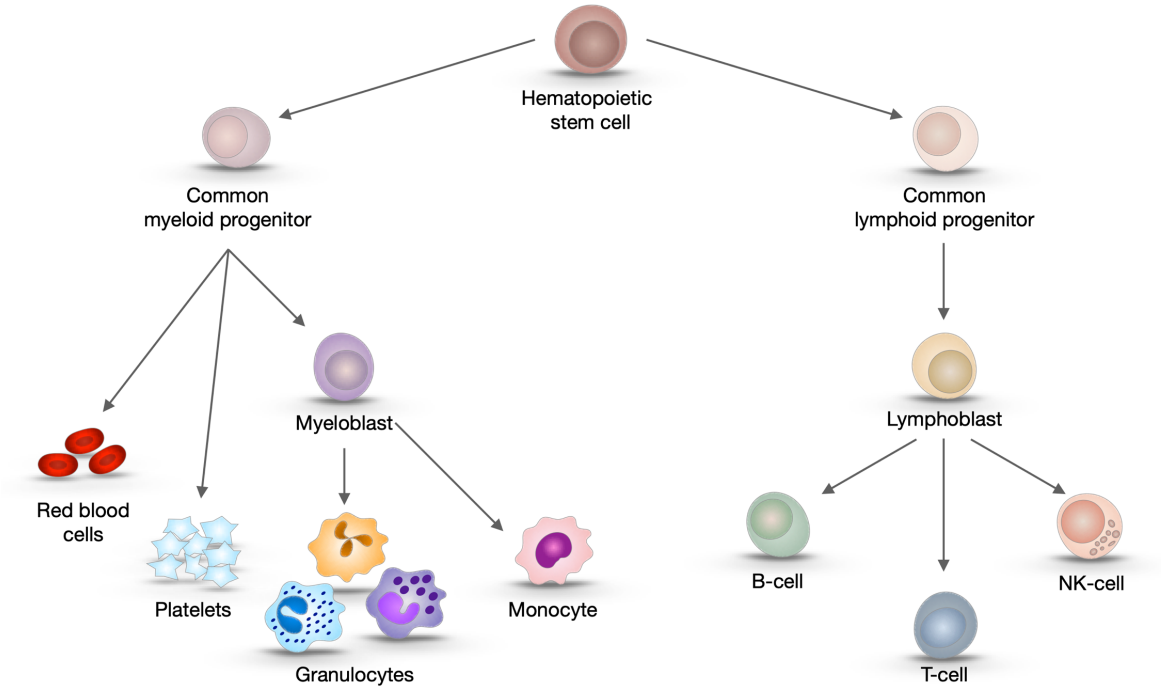


Figure 1: Schematic representation of hematopoiesis. All blood cells, including the immune cells, derive from hematopoietic stem cells in the bone marrow. These pluripotent cells produce two further types of stem cells, common myeloid and lymphoid progenitors, that give rise to the myeloid and lymphoid lineage. The myeloid branch is represented by red blood cells, platelets and myeloblasts, from which granulocytes (neutrophils, basophils, eosinophils) and monocytes/macrophages arise. The lymphoid branch is represented by lymphoblasts that generate B-cells, T-cells and natural killer (NK) cells. For simplicity, the schematic representation shows only major cell types. Individual stages of development are not depicted. Adapted from Roth-Walter and colleagues (2013)[3] and Kind and Goodell (2011)[5].

The development and differentiation of all these blood and immune cells, except of T-cell development, takes place in the bone marrow. T-cell development and maturation is a highly regulated process taking place in the thymus. Therefore, this type of cells is termed thymus-dependent (T) lymphocytes or in short T-cells.[4,6]

1.1.1 T-cell development takes place in the thymus

The thymus is a primary lymphoid organ located in the upper thorax above the heart. It consists of a multitude of lobules with an outer cortex and an inner medulla. These lobules are colonized by thymocytes; i.e. T-cell progenitors and a variety of intra-thymic cells that provide distinct areas of specialized microenvironments for T-cell differentiation (**Figure 2**).[2,4,7,8] In normal T-cell development, lymphoid progenitors from the bone marrow migrate via postcapillary venules into the cortico-medullary region in the thymus[7,9]. Through thymic epithelial cells (TECs), these lymphoid progenitors receive signals via Notch1 that irreversibly induce commitment to the T-cell lineage and maintain T-cell development.[4] The lymphoid progenitors migrate within the thymic lobules, guided by chemokines and adhesion molecules produced by intra-thymic cells[8,10,11]. The progenitors pass the distinct microenvironment areas and interact with the intra-thymic cells to accomplish differentiaton[7]. Initially, they migrate outwards to the sub-capsular region while they undergo extensive proliferation and T-cell receptor (TCR) gene rearrangements. In the cortex, they are subjected to positive selection. Surviving T-cell progenitors with functional TCR continue further to the medulla, while they undergo negative selection eliminating self-reactive T-cell progenitors. Remaining mature T-cells exit the thymus and migrate to the periphery.[8,11,12]

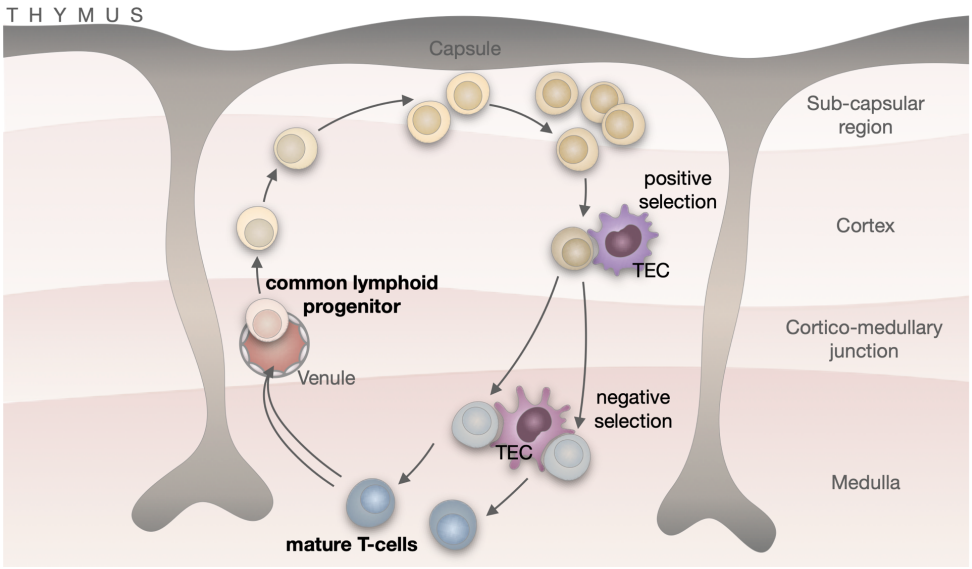


Figure 2: Different steps of T-cell development occur in distinct thymic microenvironments. Early lymphoid progenitors migrate via the blood stream from the bone marrow to the thymus[4]. They enter the thymus through postcapillary venules in the cortico-medullary region, commit to the T-cell lineage and migrate into the sub-capsular region[11]. While migrating, the T-cell progenitors undergo proliferation and TCR gene rearrangements. Subsequently, the functionality of the developing TCR is tested through interaction with TECs: first by positive selection within the cortex, and second by negative selection within the medulla. Mature T-cells, expressing a functional but harmless TCR, emigrate the thymus through the postcapillary venules in the cortico-medullary region and migrate towards the secondary lymphoid organs (i.e. lymph nodes, spleen, appendix and tonsils).[4,7,11] Figure adapted from Murphy and Weaver[4] and Savino *et al*[7].

1.1.2 Classification of T-cell development by clusters of differentiation

The T-cell development process is accompanied by characteristic cell surface proteins which allow to trace the differentiation using flow cytometry.[7,9] Based on the expression of these proteins, so called clusters of differentiation (CD), T-cell maturation has been classified into defined stages[4]. Three major developmental stages are commonly differentiated: *double negative* (DN), *double positive* (DP) and *single positive* (SP), depending on the presence of the co-receptors CD4 and CD8. In human, the DN stage can be further subdivided into the most immature human lymphoid progenitors DN1 (CD34+, CD38-, CD1a-), DN2 (CD34+, CD38+, CD1a-) and DN3 (CD34+, CD38+, CD1a+). Next, the human thymocytes mature via an immature single positive stage (ISP/DN4, CD1a+, CD4+), to early (CD3-), late (CD3+) DPs (CD1a+, CD4+, CD8+), and ultimately into mature SP helper (CD3+, CD4+) or cytotoxic (CD3+, CD8+) T-cells expressing a functional TCR (**Figure 3**).[4,13,14]

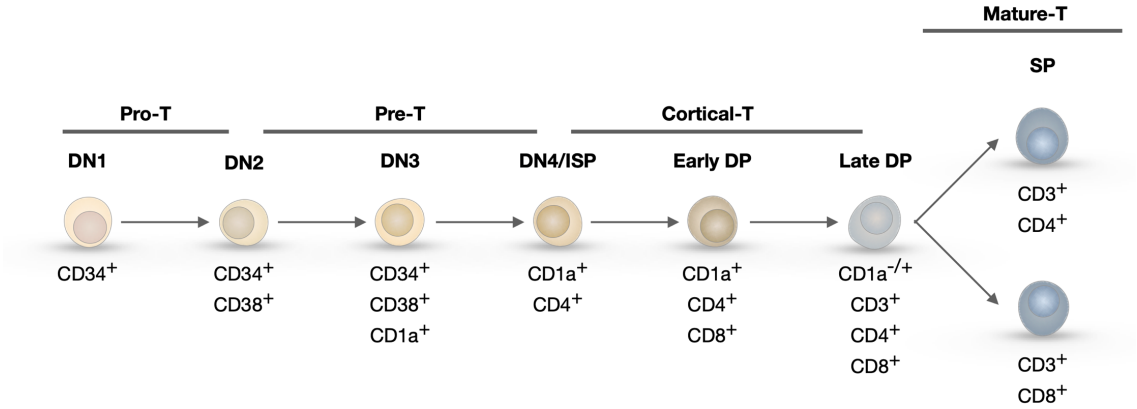


Figure 3: Overview of the clusters of differentiation commonly used to classify developing human T-cells into defined maturation stages. T-cell progenitors enter the thymus as early double negatives (DN). While developing to DN1 (CD34+, CD38-, CD1a-), DN2 (CD34+, CD38+, CD1a-) and DN3 (CD34+, CD38+, CD1a+), they migrate through the thymus, commit to the T-cell lineage and undergo TCR gene rearrangement. From an immature single positive stage (ISP/DN4, CD1a+, CD4+), they develop via early (CD3-) and late (CD3+) double positives (DP: CD1a+, CD4+, CD8+) into mature single positive (SP) helper (CD3+, CD4+) or cytotoxic (CD3+, CD8+) T-cells, expressing a functional TCR and capable to migrate out of the thymus. More classical definitions of T-cell maturation stages are based on the expression of CDs correlating to defined thymic compartments: pro-T, pre-T, cortical-T and medullary/mature-T. Adapted from published overviews of human T-cell development[2,15,16].

Conventionally, T-cell developmental stages are defined by CDs corresponding to defined intra-thymic differentiation stages. Yet uncommitted lymphocyte progenitors in the bone marrow and in the thymus express CD34, CD33, CD7 and cytoplasmic CD3 (cCD3). This developmental stage is referred to as *pro-T*. Upon T-cell commitment and TCR gene rearrangement, the T-cell progenitors can be identified by expressing CD7, CD2, CD5, cCD3 and both, CD4 and CD8 (*pre-T*). T-cells developing in the thymic cortex (*cortical-T*) are initially defined by CD1 expression and co-expression of CD4/CD8. Later stages of cortical T-cells already express surface CD3 and lose CD1 expression. T-cells that ultimately mature in the medulla and possess the capacity to exit the

thymus are defined by CD1 negativity and the expression of CD3 and either CD4 or CD8 (*medullary/mature-T*).[15,17–19]

1.1.3 Notch signaling and T-cell lineage commitment

Notch proteins are highly conserved transmembrane receptors[20]. In mammals, four Notch homologues have been identified (Notch 1 – 4), which interact with five Notch ligands that belong to the Jagged family (Jagged 1 and 2) and Delta-like family (DL1, DL3 and DL4)[21–26]. Notch signaling is involved in various aspects of development. In T-cell development, it is essential to initiate and maintain the T-cell transcription programs throughout the whole maturation process.[4,27–31] Interactions between T-cell progenitors expressing Notch1 and thymic stroma cells expressing Notch ligand DL4 induce a double proteolytic cleavage of the receptor, first by the metalloprotease ADAM10[a] and second by the γ secretase complex. The intracellular Notch1 domain is released from the membrane and activates the expression of Notch target genes in the nucleus. FBXW7[b] recognizes the PEST domain of activated NOTCH1 and triggers the termination of Notch signaling by proteasomal degradation of the receptor (**Figure 4**).[32]

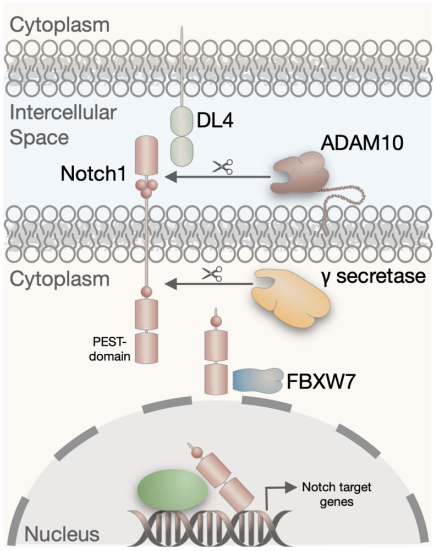


Figure 4: Overview of NOTCH1 signaling in T-cell development. Interaction of Notch1, expressed by T-cell progenitors, with DL4 on thymic stroma cells prompts a double proteolytic cleavage of the receptor by the metalloprotease ADAM10 and the γ secretase complex. The released Notch1 domain activates the expression of Notch target genes in the nucleus. Notch signaling is terminated by FBXW7 recognizing the PEST domain of activated Notch1 resulting in proteasomal degradation. Adapted from Van Vlierberghe and Ferrando (2012)[32].

Notch signaling in T-cell progenitors triggers the expression of two major transcription factors and initiates T-cell commitment and differentiation: GATA Binding Protein 3 (GATA3) and T-cell factor-1 (TCF1). Together, these transcription factors induce the expression of several T-cell lineage specific genes such as the recombinase activating gene 1 (Rag1), a major regulator of T-cell receptor gene rearrangements, or genes encoding the CD3 complex. Further, the expression of B-cell lymphoma/leukemia 11B (BCL11B) is induced, a transcription factor crucial for restricting alternative cell fates and activating the T-cell gene expression program, and thus, for committing to the T-cell lineage.[4,33]

[a] ADAM10: A disintegrin and metallopeptidase domain 10
[b] FBXW7: F-box and WD repeat domain containing 7

1.1.4 Assembly of the TCR and positive/negative selection

The assembly of a functional TCR complex is a hallmark in T-cell differentiation. The TCR is a heterodimeric receptor comprising of two covalently linked polypeptide chains, either an alpha (α) and beta (β) chain or a gamma (γ) and delta (δ) chain.[14,34] Each chain is comprised of a variable domain for antigen recognition and a constant domain[34]. As the T-cell progenitors mature, they can give rise to two distinct T-cell lineages: αβ T-cells and γδ T-cells. However, only 5 %[34] of the normal circulating T-cells express a γδ TCR, whereas the majority of T-cell progenitors commit to the αβ T-cell lineage and diverge further into the helper and the cytotoxic T-cell sub-lineages.[4,14] Together with the associated CD3 complex, providing the signaling components, and either the CD8 (cytotoxic T-cells) or the CD4 (helper T-cells) co-receptor, for the interaction with antigen presenting cells, the TCR complex is formed (**Figure 5**)[4,34].

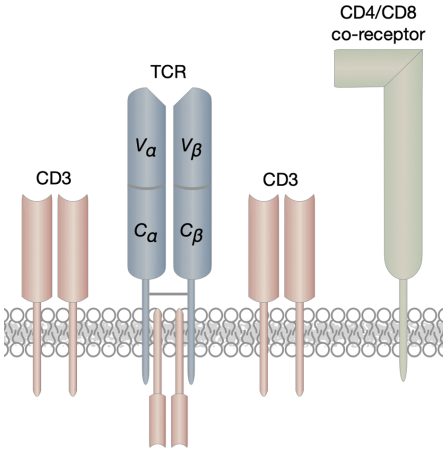


Figure 5: Schematic representation of the T-cell receptor (TCR) complex. The TCR is a heterodimeric receptor typically composed of two covalently linked polypeptide chains α and β; less frequent of an γ and an δ chain. To form the TCR complex, this TCR is complemented with the CD3 complex, providing signaling components, and the CD4 or CD8 co-receptor, to facilitate the interaction with antigen presenting cells. Adapted from Actor (2014)[34].

From V(D)J recombination to TCR expression

Development of a functional TCR involves an ordered rearrangement of the variable (V), diversity (D) and joining (J) gene segments of the TCR (referred to as *V(D)J recombination*) before the variable domains of the α, β, γ and δ chain genes are assembled. Through random combinations of the various V-, D-, J segments and the addition or deletion of extra nucleotides at the junctions, the generation of a high diversity of antigen receptors is generated.[4,14,15] Taking the αβ TCR as an example, for the combination of the α chain a set of 70 Vα segments, 61 Jα segments and 1 constant segment is available. For the β chain, 52 V_β segments, 13 J_β segments, 2 D_β segments and 2 constant segments are for disposition.[3]

At first the TCR γ and TCR δ loci undergo rearrangement and γδ T-cells branch off at the DN2 and DN3 stage. Little is known about the development of the γδ T-cell lineage.[4,6] Rearrangement of the Vβ, Dβ, and Jβ segments of the TCR β locus starts in DN2 cells and continues till the DN3 stage[4]. It is controlled by the enzymes RAG1 and RAG2 of the V(D)J recombinase complex that targets conserved recombination signal sequences flanking the individual gene segments[14,31]. In DN3 to DN4/ISP cells, the TCR β chain is expressed on the cell surface and paired with a surrogate pre-TCRα chain (pTα) to test the functionality of the TCR β chain (*β-selection*). Thymocytes that fail to go through successful β-chain rearrangement undergo apoptosis. For thymocytes with a functional pre-TCR complex, ligand-independent signaling triggers the expression of both CD4 and CD8. As such, the developing thymocytes process to the DP stage. At the same time, the signaling induces degradation of the RAG enzymes and results in an arrest of further β chain rearrangements. During the subsequent phase of intensive proliferation, RAG1 and RAG2 remain repressed. However, after the proliferative phase, RAG1 and RAG2 are expressed again and accumulate to initiate successive rearrangements of the Vα, Dα, and Jα segments.[4,14,35–37] This ensures that cells with successful β chain rearrangement give rise to a large number of DP cells with independently rearranged α chain genes. As such, a single functional TCR β chain is assembled with numerous TCR α chains.[4] Taking the random rearrangement of the α and β chains gene segments, and the inaccuracy of joining, as well as the combination of chains during $\alpha\beta$ T-cell development into account, these mechanisms enable a potential repertoire of 10^{18} antigen specificities for the TCR[3].

Positive and negative selection of cells expressing a functional TCR

Whereas T-cell development until the $\alpha\beta$ receptor production is independent of the interaction with antigens or the major histocompatibility complex (MHC), the developmental decisions in DP and SP thymocytes are made based on the binding affinity to self-peptide:self-MHC complexes[4]. Developing DP T-cells express both co-receptor molecules CD4 and CD8, and start to express low levels of the TCR $\alpha\beta$ complex on the cell surface. These cells possess a limited life span of about 3 – 4 days before they undergo programmed cell death. Their further survival depends on a rescue by *positive selection*.[4] Therefore, the thymocytes migrate through the thymic cortex and interact with cortical TECs. The cortical TECs present antigens on their MHC molecules to the developing T-cells. Thymocytes that express a non-functional receptor undergo death-by-neglect. In contrast, T-cell progenitors that recognize self-MHC receive survival-signals

and become positively selected for further maturation.[4,38] In addition, the interaction with MHC molecules during positive selection coordinates the expression of the co-receptors CD4 or CD8. Although T-cells recognize the antigens presented on MHC molecules by virtue of the αβ TCR, the co-receptors CD4 and CD8 are required to assist the TCR-mediated binding to MHC. Cytotoxic T-cells bind to MHC class I molecules[c] via the CD8 co-receptor and helper T-cells bind to MHC class II molecules[d] via the CD4 co-receptor.[4,38–40]

The importance of MHC molecules for the commitment to either the helper (CD4) or the cytotoxic (CD8) T-cell lineage was shown in transgenic mice. In mice transgenic for TCRs restricted to MHC class I, only T-cells expressing CD8 were developed. In turn, mice transgenic for TCRs restricted to MHC class II developed only T-cells expressing CD4.[4,40,41] Analogously, both mice and humans lacking MHC class II molecules show CD8 expressing T-cells but only few CD4 expressing T-cells, whereas individuals lacking MHC class I molecules lack CD8 expressing T-cells[4].

In addition to positive selection, developing T-cells in and after DP stage are subjected to *negative selection*. While, the thymocytes start to express high levels of the TCR and either the co-receptor CD4 or CD8, they transition to the DP stage and migrate deeper into the thymus. Hereby, they interact with medullary TECs, macrophages and dendritic cells in cortex and medulla. To ensure the absence of reactivity against self-antigens, those T-cell progenitors that are too strongly stimulated upon recognition of self-peptide:self-MHC complexes are eliminated. Both, T-cell progenitors and the interacting thymic cells can induce the elimination of the self-reactive cells and therefore prevent potential autoimmune responses.[4,16,38]

In sum, this tightly regulated multi-step process for T-cell development ensures the maturation of T-cells expressing a functional but not harmful TCR complex[11]. It has been estimated that only about 2 % of the thymocytes survive these selection processes and differentiate into mature T-cells leaving the thymus[4].

[c] MHC class I molecules: Expressed on most cells and present antigenic peptides generated from intracellular cytosolic proteins[4].

[d] MHC class II molecules: Expressed on antigen-presenting cells and present antigenic peptides derived from internalized extracellular pathogens[4].

1.1.5 Types and functions of mature T-cells

Different sub-types of mature T-cells exit the thymus. As stated before, the majority of thymocytes mature to T-cells of the αβ-lineage.[4] These αβ T-cells recognize target antigens presented by MHC molecules depending on their sub-type. Helper T-cells express the co-receptor CD4 and recognize internalized extracellular peptides derived from pathogens presented via MHC class II on the surface of antigen-presenting cells. Upon activation, these cells proliferate into a larger number of effector helper T-cells and secrete mediators to activate other immune cells (e.g. macrophage, cytotoxic T-cells) and stimulate B-cells to mature and produce antibodies (**Figure 6a**).[3,4] Cytotoxic T-cells express the co-receptor CD8 and react to intracellular antigens presented via MHC class I molecules; e.g. virus infected cells presenting abnormal proteins. These cells proliferate into effector cytotoxic T-cells and trigger apoptosis of the infected cell by secreting perforin and granzymes (**Figure 6b**).[3] Further subtypes of T-cells are memory and regulatory T-cells. Memory T-cells are generated after an infection to ensure a rapid response upon re-exposure to the antigen.[3,4] Regulatory T-cells prevent autoimmune diseases by moderating and terminating inflammatory processes induced by different stimuli or pathogens[42].

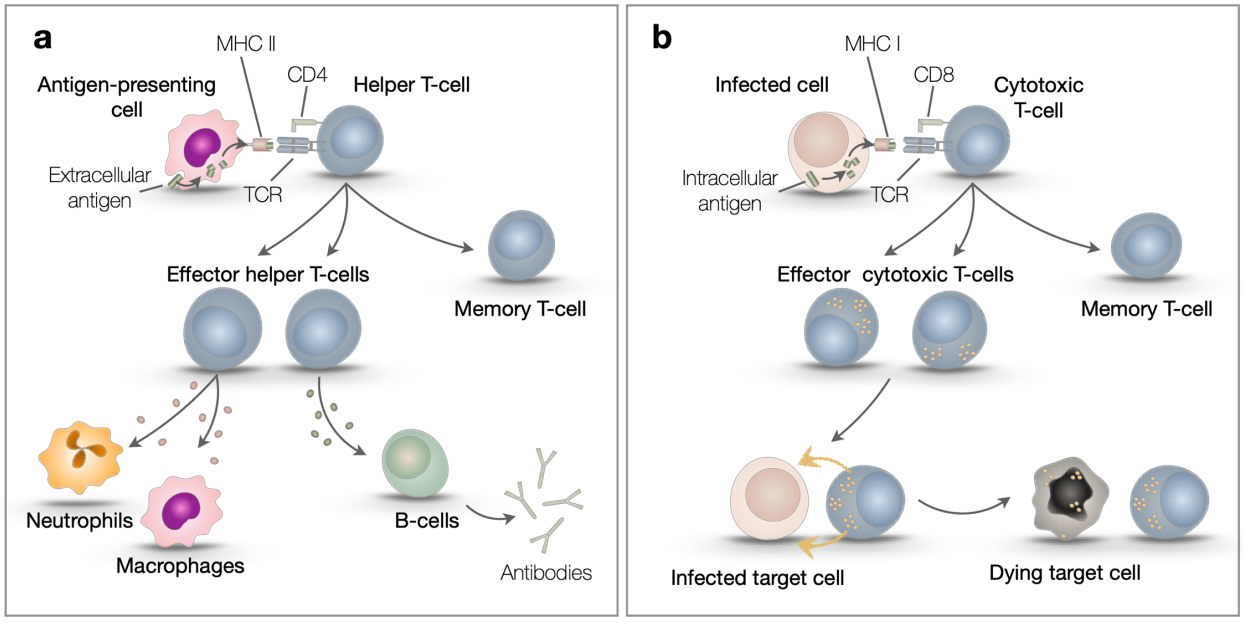


Figure 6: Activation and action of helper and cytotoxic T-cells. (a) Helper T-cells express a TCR complex comprising the co-receptor CD4. They recognize internalized extracellular peptides derived from pathogens and presented via MHC class II on the surface of antigen-presenting cells. Upon activation, they proliferate into memory T-cells and effector helper T-cells. As effector helper T-cells they secrete mediators to activate to activate other immune cells and stimulate B-cells to mature and produce antibodies. (b) Cytotoxic T-cells express an TCR complex comprising the co-receptor CD8. They react to intracellular antigens presented via MHC class I molecules, e.g. virus infected cells presenting abnormal proteins. Upon activation, these cells proliferate into memory T-cells and effector cytotoxic T-cells that secrete perforin and granzymes to trigger apoptosis of the infected cell. Schematic representation adapted from Roth-Walter *et al*[3].

As briefly mentioned in the previous section, a minor population of T-cells bear the γδ receptor. This sub-lineage of T-cells differs from the αβ T-cell lineage as these cells lack the expression of the CD4 and CD8 co-receptors. Unlike αβ T-cells, γδ T-cells seem not to be MCH restricted but recognize their target antigens directly. They are thought to be involved in tissue homeostasis, repair and the innate defense system, however, little is known about their antigens.[3–5] A further sub-lineage of T-cells are the NK T-cells which recognize self- and pathogen-derived lipids on CD1 molecules[3,4].

Given these important roles, proper functionality of T-cells is not only crucial for the immune defense. If developing thymocytes become self-reactive, inappropriately activate oncogenes or loose tumor suppressor gene activity and escape the strictly regulated selection processes, this can cause deleterious effects such as autoimmune diseases or cancer.[3,4,11,42,43]

1.1.6 Leukemogenesis and classification of leukemia

Leukemia broadly refers to different types of hematopoietic cancers. Similar to the term carcinogenesis, which describes the development of cancer, leukemogenesis refers to the development of leukemia. Leukemogenesis can occur at any stage during normal hematopoiesis. Hereby, hematopoietic (progenitor) cells are thought to undergo malignant transformation by acquiring multiple genetic and epigenetic changes which result in uncontrolled clonal expansion of the defective cells.[44–46] Leukemia is classified based on the origin of the malignant cells and the speed of leukemia progression. Chronic leukemias are thought to be derived from rather mature cells and progress slowly.[47,48] Early symptoms are typically mild e.g. patients may have vague symptoms years before diagnosis[49]. In contrast, acute leukemias develop quickly as the malignant cells originate from either hematopoietic stem cells without lineage differentiation or immature but already lineage-determined progenitor cells, which both proliferate fast. A further classification of leukemia depends on whether the malignant cells descend from the myeloid lineage (**Figure 1**, left branch) or lymphoid lineage (**Figure 1**, right branch). As an example, acute myeloid leukemia (AML) refers to acute leukemia that involves malignant cells originating from the myeloid lineage, whereas acute lymphoid leukemia (ALL) refers to that originating from the lymphoid lineage. [47,48] In addition, leukemias in which mature or already more lineage-determined progenitor cells are affected, can be further subdivided on the basis of the cell type, e.g. B-cell acute lymphoblastic leukemias (precursor (pre)-B-ALL) or T-cell acute lymphoblastic leukemia (T-ALL)[50].

1.2 T-cell acute lymphoblastic leukemia

ALL is the most common childhood malignancy worldwide and accounts for 26 % of all pediatric cancers[51,52]. About 15 % and 25 % of all pediatric and adult patients with ALL, are further diagnosed as T-ALL[53,54]. T-ALL is more common among males than females and has a median age of onset of 9 years[32,53,55,56]. The disease is characterized by malignant transformation, block in further differentiation and enhanced proliferation of lymphocyte progenitors, so called blasts, in the bone marrow, blood and extramedullary sites (**Figure 7**).

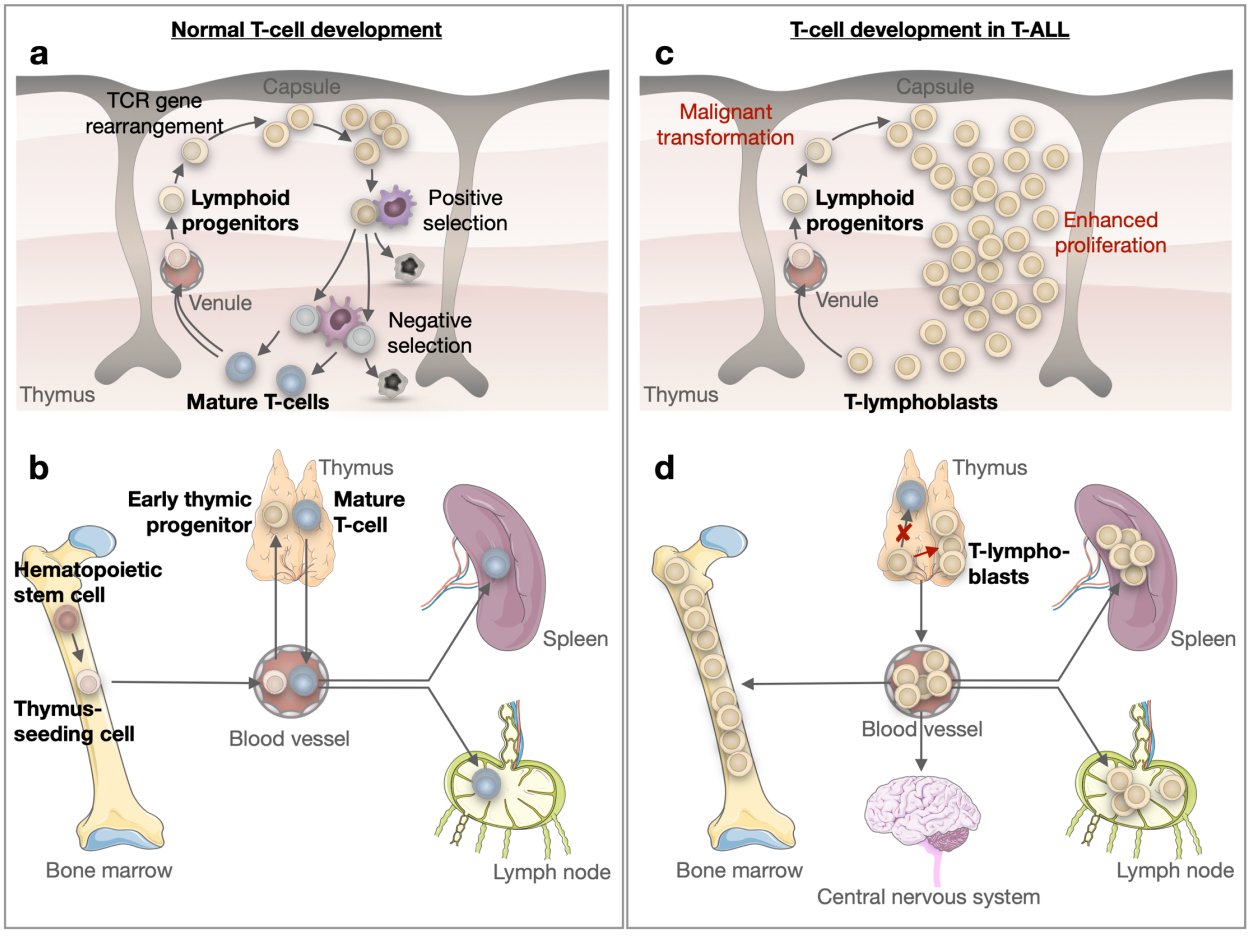


Figure 7: Normal T-cell development versus T-cell development in T-ALL. (a) In normal T-cell development is a tightly regulated process with several selection steps in distinct thymic microenvironments and yields defined numbers of fully matured and functional T-cells. (b) Lymphoid progenitors originating from hematopoietic stem cells in the bone marrow seed the thymus via the blood stream. Upon maturation, they migrate into the periphery and other lymphoid organs, e.g. spleen and lymph nodes. (c) T-cell differentiation in T-ALL is characterized by malignant transformation typically in early stages of the thymic development, resulting in a block in further differentiation and enhanced proliferation of immature T-lymphoblasts. (d) The malignant blasts exit the thymus and invade the blood, the extramedullary sites and the bone marrow, leading to disturbed hematopoiesis and an infiltration of the central nervous system. Figure modified from Karrman and Johansson (2017)[57].

However, the exact events and mechanisms underlying the transformation are not well understood.[43,52,58] The rapid proliferation and subsequent accumulation of the poorly differentiated blasts lead to disturbed hematopoiesis and critically low numbers of erythrocytes,

other white blood cells and platelets[59]. As a consequence, clinical characteristics and symptoms include high numbers of leukemic blasts in the peripheral blood or in the bone marrow (typical white blood cell counts are > 100,000/µl in T-ALL patients[60]), easy bruising and bleeding, pallor, fatigue and lethargy, fever as well as bone pain from infiltrating blasts[50,54,58,61]. About 50 – 60 % of patients show enlarged lymph nodes, liver and spleen[61]. In addition, enlargement of the thymus and meningeal infiltration of the central nervous system are observed[59,61].

The diagnosis ALL is established by a blast cell count of at least 25 % in the bone marrow or peripheral blood. Evaluation by flow cytometry, immunophenotyping, cytogenetic testing and assessment of the morphology are used to confirm the diagnosis and further classify the type of leukemia. Classification and risk stratification facilitate the assessment of prognosis and allow an adaption of the treatment regimen and intensity.[58]

1.2.1 Treatment of T-ALL

Despite of the unfavorable clinical characteristics described above, the prognosis for children with T-ALL has improved considerably with the introduction of intensified chemotherapy and nowadays modern treatment regimens achieve cure rate of 75 – 85 %[32,43,54,62]. However, the treatment consists of an aggressive high-dose, multi-agent chemotherapy and comes at the cost of severe short- and potentially long-term adverse effects[53,54]. In addition, the clinical outcome for patients with poor response, failure of induction and/or relapsed leukemia remains extremely poor[32,53,63]. Therefore, the search for new therapeutic targets that impact on T-ALL and the development of more effective drugs with less toxicity and adverse effects is of high clinical importance. This requires a comprehensive understanding of the molecular events leading to the disease and the biological characteristics of T-ALL.[32,54]

1.2.2 Classification of T-ALL by clusters of differentiation

As described before, T-ALL blasts are thought to be derived from malignant thymocytes that can arise at different stages during normal T-cell differentiation[54]. In consequence, the European Group for the Immunological classification of Leukemias (EGIL) classified T-ALL into four immunophenotypic subtypes corresponding to the normal thymic development (**Table 1**). As cytoplasmic (c) expression of CD3 occurs already early in T-cell development and precedes the membrane (m) expression, the classification is based on the presence of cCD3 or mCD3 and five further cell surface markers found in defined stages of thymic differentiation[64,65]. In addition to

cCD3, immature or pro-T-ALL express only CD7 whereas pre-T-ALL express additional CD2 and/or CD5 and/or CD8. Cortical T-ALL is, in addition to the presence of cCD3+, defined exclusively by CD1a positivity regardless of any other cell surface marker. Classification into medullary or mature T-ALL is based on the presence of mCD3 positivity and CD1a negativity. Two further subtypes are distinguished, depending on the membrane expression of an αβ TCR (group A) or a γδ TCR (group B).[15,65] However, T-ALL is a highly heterogeneous type of cancer and as many as 25 %[66] of all childhood T-ALLs do not conform to any classification[15,17,55]. Nevertheless, most T-ALLs show CD1a positivity and as such the majority of T-ALL incidences are classified as cortical T-ALL[67].

Table 1: Classification of T-ALL into four subtypes based on the definitions of the EGIL[65].

T-ALL subtype	Cell surface markers
Pro T-ALL (T-I)	cytoplasmic CD3+, CD7+
Pre T-ALL (T-II)	cytoplasmic CD3+, CD2+ and/or CD5+ and/or CD8+
Cortical T-ALL (T-III)	cytoplasmic CD3+, CD1a+
Mature T-ALL (T-IV)	membrane CD3+, CD1a-

In analogy to fresh leukemia samples for diagnosis, several studies characterized established T-ALL cell lines and correlated them either to different stages of T-cell development or assigned them into the EGIL classification subtypes [17,68–70]. Burger *et al.* tested 16 T-ALL cell lines that are frequently used for immunological and molecular studies[17]. All cell lines were shown positive for the T-cell lineage marker CD7, thus confirming their T-cell origin. Further, the high heterogeneity of T-ALL was underlined as none of the cell lines was found to be alike. Since the majority of T-ALL cases are diagnosed as cortical, it is not surprising that in total five T-ALL cell lines were proposed to correlate to the cortical phenotype. Similar to EGIL, these cell lines were considered as cortical T-ALL based on the expression of CD1a. In addition, Burger and co-workers correlated the cell lines further to the normal differentiation of human thymocytes based on CD3. For example, the cell lines P12 and MOLT3 were correlated to early (CD3-) and DND41 to late (CD3+) cortical T-ALL. Such studies aid the T-ALL research as they ease the search for appropriate T-ALL cell lines serving as disease models.[17]

In addition to the EGIL classification several other classifications have been proposed. These include further immunophenotypic classifications, classifications based on the status of the TCR or molecular-genetic classifications depending on genetic subtypes or epigenetic features.[15,32,52]

1.2.3 Genetic and epigenetic landscape of T-ALL

To date, research has mainly provided insights into T-ALL biology on a genetic and epigenetic level[53]. On the genetic level, T-ALL was found to be a highly heterogeneous type of cancer. Multiple oncogenes and tumor suppressors cooperate and give rise to malignant T-cells with aberrant differentiation, cell cycle control and proliferation.[15,17,53,54,71] Nevertheless, a number of recurrent features have been identified (**Figure 8**)[54]. In about 50 % of T-ALLs, an abnormal karyotype has been found[54,72]. This may be explained by thymocytes having a higher susceptibility for chromosomal translocations due to incorrect TCR recombination during T-cell development [54]. Such chromosomal translocations and intra-chromosomal rearrangements frequently place developmentally important transcription factors under the control of the strong promoter and enhancer elements of the TCRβ – TCRγ (chromosome 7) or TCRα – TCRδ (chromosome 14) genes, resulting in an aberrant overexpression of these transcription factors in T-cell progenitors (e.g. HOX11, LYL1 and TAL1[e])[15,32,54,72]. In addition, chromosomal translocations commonly generate fusions of genes, resulting in chimeric proteins with oncogenic activity. Examples are SIL – TAL1 or MLL[f] – partner-gene fusions found in up to 30 % and 8 % of T-ALLs respectively[15,54]. Frequently, cytogenetically cryptic abnormalities result in the loss of tumor suppressors, such as the most common deletions in the CDKN2A/B locus leading to a loss of p16/INK4A and p14/ARF[g] found in 70 %[72] of T-ALLs[15,54]. It is hypothesized that cryptic abnormalities arise during V(D)J recombination as other genetic loci may be accessible and as such are susceptible to illegitimate activity of the recombinase enzymes RAG1 and RAG2[54]. In addition, activating mutations have been found in T-ALL. For example, more than 50 %[32,54] of all T-ALLs show constitutive activation of NOTCH1 signaling by gain of function mutations of the NOTCH receptor or loss of function of the FBXW7 ubiquitin ligase[54,72]. In consequence, T-cell specific transcription factors are abnormally expressed and can function as oncogenes[59]. As Notch1 signaling regulates hematopoietic stem cell self-renewal and is essential for commitment to the T-cell lineage, it was proposed that mutated active Notch1 could be a driver for the transformation of T-cell progenitors[15,54].

[e] HOX11: homeobox-11 (also TLX1); LYL1: lymphoblastic leukemia-derived sequence 1; TAL1: T-cell acute lymphocytic leukemia 1 (also SCL);

[f] SIL: SCL-interrupting locus; MLL: mixed-lineage leukemia

[g] CDKN2A: cyclin dependent kinase 2a; INK4a: cyclin-dependent kinase inhibitor 4a; ARF: alternate reading frame tumor suppressor

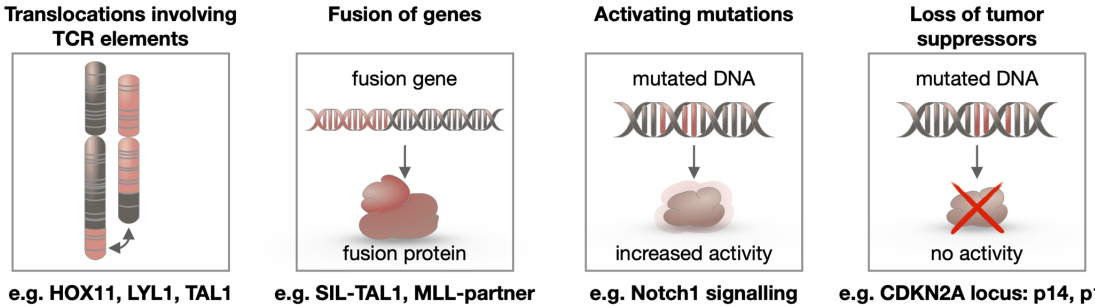


Figure 8: Overview of recurrent genetic features found in T-ALL. Approximately 50 % of all TALLs show translocations[54,72]. Frequently, these place transcription factors adjacent to strongly expressed TCR elements, resulting in transcription factor overexpression (e.g. HOX11, LYL1 and TAL1). Others yield fusions of transcription factor genes resulting in chimeric proteins with oncogenic activity (e.g. SIL – TAL1 or MLL – partner-gene). Further aberrations are activating mutations (e.g. constitutive active NOTCH signaling) and cryptic abnormalities that result in the loss of tumor suppressors (e.g. deletions in the CDKN2A/B locus leading to a loss of p16/INK4A and p14/ARF)[32,54].

Over the past decade studies have demonstrated that not only genetic but also epigenetic changes influence cancer onset, progression and treatment response[73]. These changes include alterations in DNA methylation and histone modification which perturb transcriptional regulation and consequently modify gene expression (**Figure 9**)[72–74]. Typically, cancer cells exhibit global hypomethylation and localized hypermethylation in CpG-rich promotor regions[63]. Global DNA hypomethylation triggers chromosomal instability and activation of tumor promoting genes, whereas local DNA hypermethylation has been linked to repression of tumor suppressor genes[53,73,75,76]. In T-ALL, different studies reported hypermethylation of promotor regions of tumor suppressors, such as CDKN1A, CDKN2B, SYK, C/EBPA, PAX5 and ABCG2[h], causing an abrogation of their expression (**Figure 9a**)[63]. DNA methylation, and especially repression of tumor suppressor genes by hypermethylation, has been the most widely studied epigenetic modification in cancers. However, in recent years aberrant modifications to proteins that package DNA, such as histone proteins, have gained increasing interest.[77–79] Modifications on histone proteins affect their interaction and thus the accessibility of the DNA. Histone modifications are created by *writer* proteins that covalently attach modifying groups, *reader* proteins recognize and bind to modifications and *eraser* proteins remove the modifying groups[77,80,81]. In a recent study, Huether and colleagues sequenced the genes of 633 epigenetic regulators in more than 1000 pediatric tumors of 21 cancer types[82]. In T-ALL the frequency of mutations in genes encoding proteins involved in epigenetic regulation was amongst the highest. In particular, mutations and alterations in histone writers, readers and erasers were found to be more frequent in T-ALL compared to other childhood cancers.[72,82]

[h] CDKN1A/2B: cyclin dependent kinase 1a/2b; SYK: spleen tyrosine kinase; C/EBPA: CCAAT/enhancer-binding protein α; PAX5: paired box protein 5; ABCG2: ATP-binding cassette sub-family G member 2

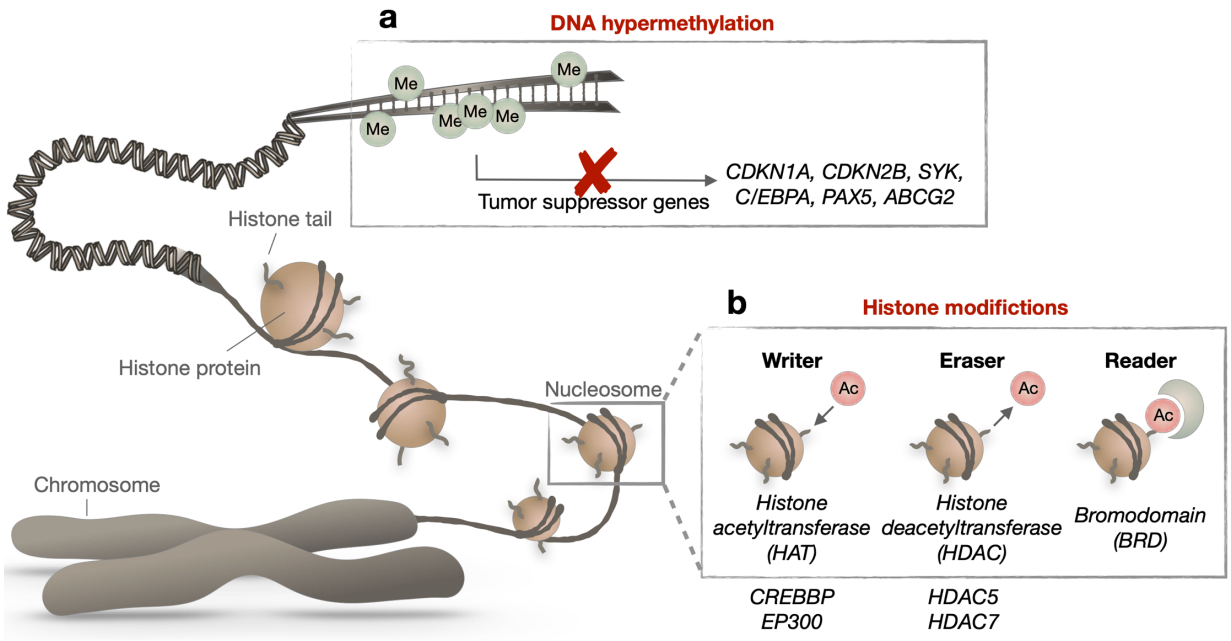


Figure 9: Overview of epigenetic alterations in T-ALL. Epigenetic alterations are changes in gene function occurring independent of changes in the DNA sequence[83]. Instead, these alterations occur through chemical modifications directly on DNA (e.g. DNA methylation (Me) at cytosine residues) or on proteins that package the DNA (e.g. acetylation of histone tails). (a) In general, DNA hypermethylation at promotor regions of tumor suppressor genes is associated with their repression. In T-ALL in particular, the tumor suppressors CDKN1A, CDKN2B, SYK, C/EBPA, PAX5 and ABCG2 were shown to be repressed[63]. (b) Histone modifications are created by *writer* proteins that covalently attach modifying groups, *reader* proteins recognize and bind modifications and *eraser* proteins remove the modifying groups[77,80,81]. In case of histone acetylation, acetyl groups (Ac) are covalently attached to lysins on histone proteins by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs). Bromodomain (BRD) containing proteins recognize the acetylation marks. In T-ALL, mutations in the HATs CREBBP and EP300 and in the HDACs HDAC5 and HDAC7 were identified[53,82,84–86]. Figure adapted from Arrowsmith (2012)[77] et a. and from Maleszewska and Kaminska (2013)[87].

As described above, remodeling chromatin between the "open and active" or "closed and repressive" form plays a major role in epigenetic regulation processes[88]. The level of chromatin accessibility depends on the type and combination of modifications[77,81]. So far, sixteen different types of modification were found on histones, such as acetylation, methylation, phosphorylation and ubiquitination[80,81,88] Acetylation is one of the most well-characterized histone modifications[77,78]. Acetyl groups are covalently attached to lysins on histone proteins, and also on non-histone proteins, by writers called histone acetyltransferases (HATs) and removed by erasers, termed histone deacetylases (HDACs). Principal readers of acetylation marks are bromodomain (BRD) containing proteins. In general, acetylation neutralizes the positive charge associated with the ε-amino group of lysine and reduces the attraction to the negatively charged DNA; relaxing the DNA and promoting transcription.[77,80,88–93] Studies on epigenetics in T-ALL identified, amongst others, mutations in CREBBP and EP300[i], transcriptional co-activators with HAT activity, as well as in HDAC5 and HDAC7 (**Figure 9b**)[53,82,84–86]. Further, elevated expression

[i] CREBBP: CREB-binding protein (also CBP); EP300: E1A binding protein p300 (also p300)

levels or enhanced activity of particular HDACs and their association with oncogenic transcription factors was reported by several groups[82,94–97]. Similar to DNA hypermethylation, HDAC overexpression is thought to be linked to cancer development by repression of tumor suppressor genes[92,98,99]. However, the exact mechanism by which HDACs are involved in T-ALL remains poorly understood. Nevertheless, inhibitors of HDACs have been reported to reverse gene silencing and showed strong antitumor effects in the context of T-ALL[53,63,100].

1.2.4 HDAC inhibitors and their promise for the treatment of leukemia

In contrast to genetic abnormalities, epigenetic abnormalities are reversible[101]. Thus, epigenetic therapies have evolved as a new promising field within cancer therapeutics. Among the fastest moving areas of clinical research is the development of HDAC inhibitors (HDACi) as a potential strategy to reverse abnormal acetylation.[93,102,103] HDACs, and therefore their inhibitors, were named referring to histones, since the first identified substrates of these enzymes were histones[102]. However, over the past years more than 50 acetylated non-histone substrates were identified[104]. As such, HDACs are not only involved in the deacetylation of histones, leading to the repression of gene transcription, but also in the deacetylation of non-histone proteins, regulating important functions in cellular homeostasis; e.g. cell-cycle progression, differentiation and cell death[78]. As a consequence, it has become evident that HDAC is an inappropriate and misleading designation[102]. It would be better to regard them more broadly as "lysine deacetylases" (KDACs) and "lysine deacetylase inhibitors" (KDACi); as they are more and more frequently referred to in newer literature[102,104,105].

So far 18 human HDACs have been identified and categorized into four classes depending on their homology to yeast proteins[88,93,101,104]. Class I is comprised of HDAC1, 2, 3 and 8, which comprise a single deacetylase domain homologous to yeast RPD3[101]. This type of HDACs is predominantly localized in the nucleus[101,104]. Class II is subdivided into a and b: Class IIa contains HDAC4, 5, 7 and 9, which possess a long amino terminus and show homology to yeast HDA1. Class IIb contains HDAC6 and 10, both have two catalytic domains. Class II HDACS were shown to shuttle between nucleus and cytoplasm.[97,104] The catalytic domain of HDAC11 possesses features of both, class I and class II, and is therefore sometimes placed as a single HDAC in class IV[101,104]. HDACs in class III are termed sirtuins and show homology to yeast Sir2.[93,101,104] Of these 18 HDACs, 11 HDACs share common features and show zinc-dependency (class I, II and IV). They are referred to as "classical HDACs", whereas the class III HDACs are structurally

different to the other classes, zinc-independent but nicotinamide–adenine dinucleotide-dependent (Class III).[93,99,102,106] Most of the so far developed HDACi target the classical HDACs of class I, II and IV[103]. Hereby, the inhibitory effect of many HDACi is based on the zinc-dependency[77,103,107]. The first identified HDACi was the natural antibiotic trichostatin A (TSA). As TSA docks into the active site of classical HDACs, the hydroxyl group chelates the coordinating zinc-ion and thus inhibits the enzymatic activity[103]. Different types and classes of HDACi have been developed[88]. Several of these are based on a related pharmacophoric structure and exert their function based on a similar mechanism of action. This type of HDACi is exemplified by the hydroxamate Vorinostat[88,103].

Vorinostat, also termed suberoylanilide hydroxamic acid and therefore referred to as SAHA, is a small molecular weight compound (< 300[97]) which was further developed from DMSO and is nowadays one of the most advanced HDACi[97,103,108]. Due to its broad target spectrum including enzymes of both, class I and II, SAHA belongs to the so-called pan-HDACi[97]. Already at low concentrations (e.g. IC_{50} <86 nM[110], IC_{50} of 0.5 – 10 μM[111]) SAHA was reported to alter acetylation patterns, inhibit cell cycle progression and proliferation, and ultimately induce apoptosis in a wide variety of transformed cells, *in vitro* as well as in animal models[97,99,110]. These effects are thought to results from the inhibition of HDAC activity and subsequent accumulation of both acetylated histones and non-histone proteins[88,97]. Primarily the former was shown to selectively alter the expression of genes[88,104]. In JURKAT/CEM cells, a mature T-ALL cell line, SAHA-treatment for 16 h resulted in an alteration of 2,205 expressed genes (i.e. 22.1 %). Hereby a similar number of genes was up-regulated as repressed.[104,112] In addition, hypoacetylation-mediated alteration of the activity of non-histone proteins was observed. Amongst them numerous transcription factors, including p53 and STAT1, or cytoplasmic proteins such as α-tubulin and Hsp90.[j88,97,104]

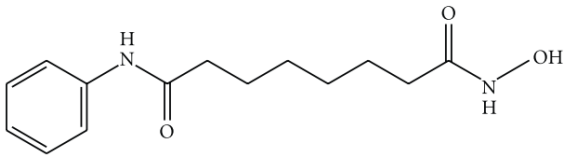


Figure 10: Chemistry and structure of the HDAC inhibitor Vorinostat. Structurally, Vorinostat is a linear small molecular weight (264.32 g/mol) hydroxamic acid compound with an empirical formula of $C_{14}H_{20}N_2O_3$; also referred to as suberoylanilide hydroxamic acid (SAHA).[109] Structural representation as shown by Masetti *et al.* (2011)[99].

[j] p53: Tumor protein p53; STAT1: Signal transducer and activator of transcription 1; HSP90: Heat shock protein 90

In 2006, the FDA approved SAHA as Zolinza® (Merck) for the treatment of cutaneous T cell lymphoma. Hereby SAHA was the first HDACi to obtain approval, which was a milestone within the field of epigenetic therapies.[73,77] The approval of SAHA and the remarkable efficacy shown in cutaneous T cell lymphoma suggest that it might become available for inclusion in T-ALL treatment regimens[53,73]. In T-ALL cell lines and patient blasts, SAHA-treatment induced growth inhibition and apoptosis[106]. These growth inhibitory effects may in part be attributed to a HDACi-mediated decrease in the expression of the major transcription factor oncogene TAL1 (see previous section)[53,113]. As pan-HDACi with its specific biological effects on cellular processes but low toxicity to normal cells, SAHA appears to be suited for combination with the currently used T-ALL chemotherapeutics[53,99]. Different studies evaluated combinations with SAHA: Leclerc *et al.*[114] reported that methotrexate in combination with SAHA synergistically increased apoptosis in the T-ALL cell lines, Gruhn *et al.*[95] showed synergistic effects for etoposide plus SAHA and Looseveld *et al.* found synergism when combining vincristine with SAHA[115]. Despite of the available data and these encouraging results, the efficacy of SAHA for the treatment of T-ALL has yet to be established and further clinical trials are required[99]. In addition, a potential clinical application is thought to be significantly improved by an increased understanding of the mode of action underlying the anti-tumor effects of SAHA in T-ALL[102].

1.3 RBPs – major players in health and disease

Studies on genetic and epigenetic alterations provided insights on the transcriptional level of gene expression. However, gene expression is also regulated on post-transcriptional levels. RNA binding proteins (RBPs) are the major players and global coordinators of these posttranscriptional steps.[116,117] Therefore, it is not surprising that aberrant RBP activity is involved in many diseases[118,119].

1.3.1 RBP biology

Immediately after transcription, a myriad of proteins associate with the RNA and mediate virtually every stage of its lifecycle[117,119,120]. These RBPs directly interact via RNA-binding domains (RBDs) with their target RNAs[121]. In doing so, the RBPs can either form stable ribonucleoprotein (RNP) particles or interact transiently and in dynamic exchange with the RNA to carry out their functions[121,122].

Given the various functions of RBPs, a large repertoire of diverse RBDs seems likely[123]. In contrast, only about > 40 canonical RBDs are known[124,125]. Among the most common and well characterized RBDs are the RNA recognition motif (RRM), which is found in more than 50 % of all RBPs, the K homology (KH) domain, the RGG (Arg-Gly-Gly) box, several zinc finger motifs (ZnF), the DEAD/DEAH box and the double stranded RNA-binding domain (dsRBD)[122,124,126]. Typically, RBDs contain around 60 – 100 amino acid residues which adopt an $\alpha\beta$ topology[125]. Some of these RBDs can additionally bind DNA or proteins[127]. The different binding modes of the RBDs range from nonspecific binding to sequence-specific and/or structure-specific[121]. However, most individual domains commonly bind only to short stretches of RNA with relatively low affinity[123]. To achieve high specificity and target the vast variety of RNAs, RBPs follow a modular design, i.e. they combine the limited number of RBDs in various structural arrangements to expand the binding repertoire[123,128]. Further, RBPs comprise many intrinsically disordered regions which increase flexibility and adaptability to facilitate binding to various structures[129]. Such combinatorial binding allows individual RBPs to interact with hundreds to thousands of different RNAs[130]. In addition to RBDs, RBPs commonly contain further domains that confer catalytic activity or enable protein-protein interactions[123,124]. This allows the formation of large RNP complexes or even whole macromolecular "machines" with specific functions e.g. the ribosome or splicosome[124,131]. Individual RBPs in these large RNPs serve as structural elements for RNP packaging and act as "adaptors" which create versatile macromolecular binding surfaces[122,123,130]. Hereby, higher-order interactions between RBPs on specific mRNAs are commonly involved in coordinating regulatory functions. Thus, the composition of RBPs in the RNP dictates the fate of its target RNA[132].

The binding of RBPs to the RNA is highly dynamic and constantly changing[121]. The processes that drive theses changes in RNP composition involve, amongst others, post-translational modifications (PTMs) to RBPs[119,131]. PTMs are thought to directly regulate and fine-tune RBP-activity[119,121,131]. A recent study revealed that the RNA-binding sites of RBPs constitute hot spots for defined PTMs; e.g. for lysine acetylation and tyrosine phosphorylation[128]. An example for PTM-dependent RBP activity represents SAM68[k], which requires tyrosine phosphorylation to mediate RNA binding[119,125]. Such post transcriptional regulation via RBP activity directly affects RNA stability and/or translation and thus facilitates a fast alteration of gene expression[121,133,134].

[k] SAM68: Src-associated protein in mitosis of 68 kDa

Moreover, it allows precise adjustment of RNA regulation and an appropriate response to internal and external stimuli[121]. Thus, the composition of the RNA-bound proteome (referred to as "RNA interactome") is always context-dependent. Although certain RBPs seems to be constitutively and ubiquitously active (referred to as "housekeeping" RBPs), many RBPs appear to be restricted in their expression and/or RNA-binding activity.[131]

1.3.2 The role of RBPs in normal post-transcriptional gene regulation

As stated at the beginning of this chapter, RBPs exert essential roles in post-transcriptional regulation of gene expression[116,117]. Upon transcription in the nucleus, newly synthesized RNAs undergo *(a) 5'-end capping* to form a 7-methylguanosine cap, *(b) pre-mRNA splicing* to remove introns and ligate remaining exons, *(c) 3'-end cleavage and polyadenylation*, *(d) mRNA export and localization* in the cytoplasm, *(e) translation*; and ultimately degradation (*(f) mRNA stability*)[122,135].

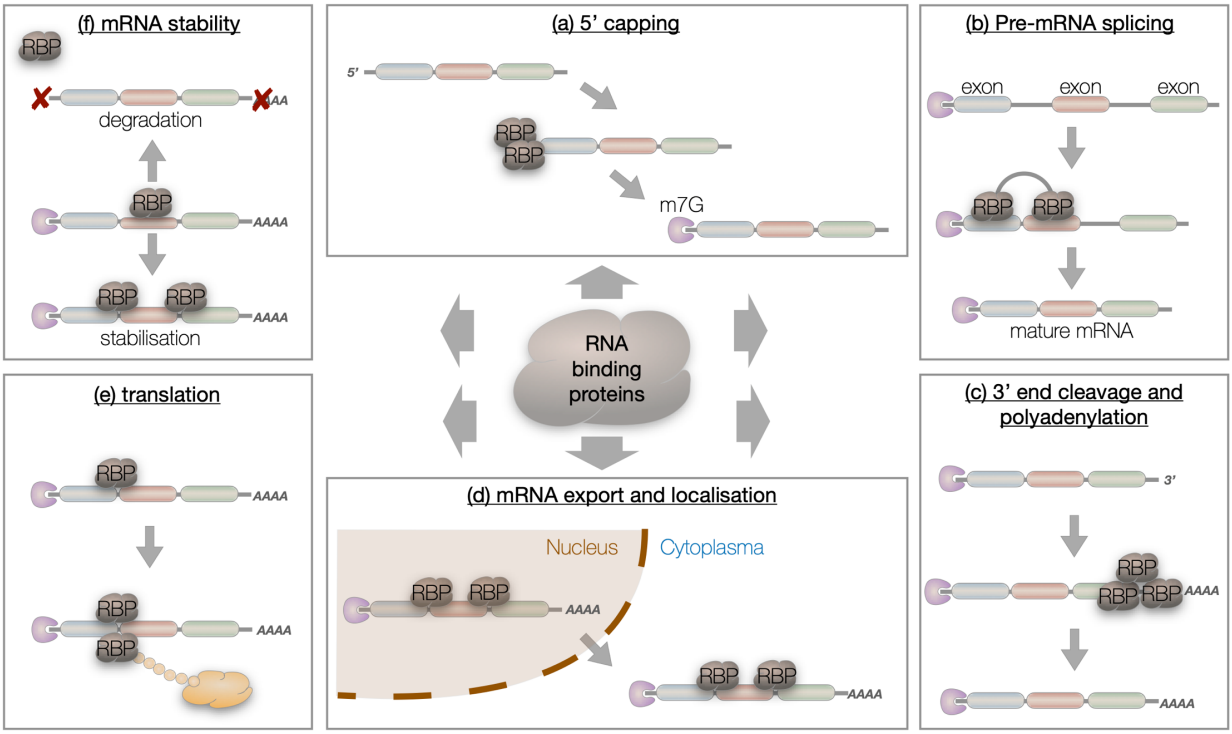


Figure 11: Schematic overview of the role of RBPs in regulating post-transcriptional gene expression. (a) 5'-end capping: RBPs in the cap binding complex add a 7-methylguanosine cap (m7G) to the 5'- end of the nascent mRNA. (b) Pre-mRNA splicing: a large number of RBPs forms a macromolecular RNP (spliceosome), removes non-coding introns and ligates the remaining exons. (c) 3'-end cleavage and polyadenylation: RBPs cleave the fully transcribed pre-mRNA at defined site within the 3'-end and add between 150 – 200 adenosine residues (poly(A)-tail, -AAAA). (d) mRNA export and localization: mRNA is through the nuclear pore complex into the cytoplasm is mediated by a multitude of RBPs that form and package the mRNA into an export competent mRNP complex. Upon export the mRNA is either directed for cytoplasmic compartmentalization or for (e) translation to the polysomes (RBP complex orchestrating translation). (f) mRNA stability: transcript stability depends on specific RBPs associating with the transcript or on alterations to the poly(A) tail and the cap. Figure adapted from Sutherland *et al.* (2015)[136].

For these processes, RBPs bind to sequence and/or structural motifs within the nascent mRNA and, together with noncoding RNAs, form the messenger RNP (mRNP).[121,122,131] Particularly the untranslated regions (UTRs) of mRNAs often serve for interactions with RBDs as they contain various RBP binding sites with cis-acting regulatory functions[117,122]. Some of the RBPs of the mRNP remain stably bound whereas others are subject to dynamic exchange. The composition of RBPs and their relative position on the mRNA mediate the mRNA processing events.[122]

In the mRNP, the nuclear cap binding complex, a heterodimer of the RBP CBP20 and CBP80[l], adds the cap-structure and remains bound to the caps of the pre- and newly synthesized mRNAs. It asserts the cap function during pre-mRNA processing and in directing the translational machinery.[136,137] Splicing is catalyzed by the spliceosome, a macromolecular RNP "machine" composed of five small nuclear RNPs[m], called U1-U6, and more than 100 different further proteins. Within the spliceosome, RBP-RNA as well as RNA-RNA interactions are formed and consistently rearranged in order to align reactive groups during catalysis.[136,138] Endolytic cleavage and the addition of the poly(A) tail involves a polyadenylation machinery comprising of six multimeric proteins: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors I and II (CFI and CFII), poly(A) polymerase (PAP), and poly(A)-binding protein II (PABII). These (RNA-binding) proteins come together and first mediate 3'-end cleavage of the nascent mRNA before they facilitate the coupled polyadenylation.[136,139] Next, the now mature mRNA is exported into the cytoplasm. This is mediated by a multitude of RBPs serving as export adapters, ATPase/RNA helicases and nuclear pore complex associated proteins that form and package the mRNA into an mRNP complex competent for export. During the export, the RBPs of the mRNP either remain within the nucleus or chaperone the mRNA into the cytoplasm. The ultimate cytoplasmic localization is determined by the RBPs that remain bound to the mRNA and the ones newly recruited in the cytoplasm.[122,135,136,140,141] Only transcripts directed to polyribosomes undergo translation; one of the most important events through which RBPs play a major role in post transcriptional gene regulation. Hereby, regulation is mainly achieved by modifying the poly(A) tail of the transcript, the composition of the associating RBPs in its mRNP and thus its entailing localization. Binding of the RBP PABP[n] to the poly(A) tail regulates

[l] CBP20 and CBP80: cap binding protein 20 and 80

[m] The splicosome consists of five small nuclear RNAs (snRNAs) with associated RBPs, forming the small nuclear ribonucleic particles (snRNPs) U1-U6.

[n] PABP: poly(A)-binding protein

its length. Poly(A) tail lengthening is associated with increased translation rates.[136] PABPs and the CBP20/80 complex at the 5´ cap are essential for translation initiation. The CBP20/80 complex interacts with translation initiation factor 4G (eIF4G[o]) to recruit the ribosome. Functional circularization of the mRNA and efficient translation by polyribosomes is promoted by further interactions between the 5´cap, eIF4E, eIF4G, the poly(A) tail and PABPs.[122] However, transcripts not directed to polysomes but to cytoplasmic compartmentalization in other RNPs are kept in a transcriptionally quiescent state and undergo processing, sorting, storage, or degradation. Some of these RNPs protect the transcript and remain transcriptionally silent awaiting signals to time protein expression (processing bodies), others stall translation (stress granule) or include exosomes for mRNA degradation (exosome).[122,136]

mRNA stability is a further aspect of post-transcriptional regulation controlled mainly through RBPs. Unwanted transcripts undergo degradation, whereas mRNA stability of others is maintained. RNA degradation is triggered by exonucleolytic shortening of the poly(A) tail and decapping.[136] Further and remarkable examples for regulation of mRNA stability are "moonlighting" enzymes. These metabolic enzymes serve as RBPs and regulate the expression of mRNAs involved in their pathway[142]. A prominent example is ACO1/IRP1[p] which balances cellular iron levels. In iron-deplete cells, IRP1 is at loss of its co-factor iron. It acts as a RBP and binds to the iron responsive elements (IRE, cis-acting element) in the 3'-UTR of the transcript encoding the transferrin receptor which is required for cellular iron up-take. This interaction promotes mRNA stabilization and transferrin receptor expression. In iron-repleted cells, IRP1 utilizes iron as co-factor and convert to the enzyme ACO1 which metabolizes citrate and isocitrate. The IRE of the transferrin receptor transcript remains unbound and the mRNA is subjected to degradation.[143–145] In such and further ways, RBPs orchestrate the fate of their target RNAs from transcription to decay and maintain cellular homeostasis.

[o] EIF4G and EIF4E: Eukaryotic translation initiation factor 4 G and 4 E

[p] ACO1: Aconitase 1; IRP1: iron regulatory protein 1

1.3.3 The role of RBPs in disease

Because of the important and widespread functions RBPs exert to control gene expression, aberrations in their function are associated with many pathological conditions[117,121,146,147]. These include neuropathies, muscular atrophies, metabolic disorders and cancer[117,125]. In addition, many RBPs can be linked to Mendelian diseases[117].

Particularly neurodegenerative disorders and cancer are frequently associated with malfunctioning RBPs. For the former, this is attributed to the high degree of alternative splicing in the brain, a process mainly controlled by RBPs.[119,125,148] Cancer in contrast, is a highly heterogeneous disease and RBP dysfunction can be involved in most, if not in all aspect of its development and progression[119,120]. Because individual RBPs can interact with many different target mRNAs, and affect various biological processes (see section 1.3.2), aberrations of single RBPs can involve in all hallmarks of cancer, and thus contribute to uncontrolled cell proliferation, resistance to apoptosis, evading immune surveillance, sustained angiogenesis and induction of invasion and metastasis.[119,120,129] The defects in RBP function can either constitute a primary cause of cancer (driver) or a consequence (passenger)[129,146]. Either way, there are various mechanisms leading to such defects, including genetic alteration or changes in RBP expression and/or activity[117,147]. Although genomic alterations are typical for cancer cells, they seem to play a minor role in RBP malfunctioning in cancer[129]. A recent large-scale analysis of mutations, copy numbers and gene expression patterns of 1348 RBP genes in 11 solid tumors showed that only ~15 % of RBPs contain mutations that impact on the protein sequence[129,149]. A further study reports that amongst the ~500 cancer driver mutations discovered so far, only several tens are associated with RBPs[150]. Nevertheless, other studies showed that RBPs are significantly abnormally expressed in cancer cells compared to their normal cell equivalents and that RBP expression appears to correlate with patient prognosis.[129,147,149,151] However, the alterations in RBP expression levels were not as strong as those of other cancer-related genes (e.g. transcription factors)[129,151]. But since individual RBPs regulate multiple mRNAs, even small alterations in their expression can lead to large-scale disruptions or affect global gene expressing[119,129]. Above all, however, it is thought that RBP malfunction in cancer is not only due to abnormal gene expression, but mainly due to deregulation of RBP activity[119,129]. As described in section 1.3.1, PTMs regulate RBP activity, i.e. binding properties, function, localization or turnover, and active sites within RBPs are hotspots for different types of PTMs[117,129]. Moreover, disease-associated mutations at the PTM sites of RBPs and even more frequently directly adjacent (PTM ± 2) were reported[117].

Despite of the indications and the evidence of RBPs to be involved in cancer, only a relatively small number of RBPs were characterized as cancer-traits so far. One example is SAM68, which is directly linked to oncogenesis by controlling alternative splicing of cyclin D1 and by mediating the generation of a stable isoform of the splicing factor SF2/ASF, both typical protooncogenes frequently deregulated in cancer[125,152]. Further, the RNA-binding activity of SAM68 is regulated by PTM and the increased acetylation levels found in most cancer cells compared to normal cells were shown to enhance the binding of SAM68 to its target RNAs[129,153]. In addition, SAM68 was found to be overexpressed in a variety of tumors including prostate, breast, renal and cervical cancer[125,152]. A further prominent example is the human antigen R (HuR). HuR is involved in mRNA stabilization and regulates the splicing of target mRNAs that encode proteins involved in sustained proliferation, impeded apoptosis, enhanced angiogenesis and promotion, as exemplified by the protooncogenes of different cyclins (D1, E1, A2 and B1), the antiapoptotic factors Bcl-2, Mcl-1 and SIRT1 or the proangiogenic factors VEGF and HIF-1α[q]. Notably, HuR is also involved in the repression of the proapoptotic factor c-Myc. The subcellular localization of HuR, and thus its RNA-binding, is regulated by PTMs. Whereas in normal cells HuR is localized in in the nucleus, it is often found in the cytoplasm in cancer cells. Therefore, high cytoplasmic levels correlate with a high degree of malignancy and can serve as a prognostic factor for poor clinical outcome. In addition, overexpression of HuR has been reported in many different types of cancer, including lymphomas, breast, brain, lung and colon cancer.[129,147,152] Both SAM68 and HuR, exemplify that abnormal post-translational modification to RBPs can deregulate their binding to mRNA targets independent of alteration in their gene expression[150,152]. Further, the large spectrum of mRNAs regulated by these proteins underscores the potential of RBPs to affect nearly all steps of tumorigenesis and suggest that targeting RBP activity might serve as a good strategy for therapeutic intervention.

A growing body of literature suggests that dysregulation of RBP function in involved in leukemia as well. Deletions in the RBPs ZFP36L1 and ZFP36L2 were reported to cause impaired thymic development and T-cell lymphoblastic leukemia[154]. Further, defects in the RBP RPL10 and RPL 22 were found to be involved in 7.9 % and 20 % of pediatric T-ALLs, respectively[155]. In addition, a recent study performed a CRISPR RBD screen and identified a network of upregulated and physically interacting RBPs in AML. In a comparison with RBPs of a counterscreen, RBP

[q] Bcl-2: B-cell lymphoma 2; Mcl-1: myeloid cell leukemia 1; SIRT1: Sirtuin-1; VEGF: vascular endothelial growth factor; HIF-1α: hypoxia-inducible factor 1 α

dependencies unique to AML as well as to T-ALL were found.[156] These findings suggest that a number of RBPs may be involved in T-ALL, which have so far remained unknown and undiscovered.

1.3.4 Analysis of RBPs

Since the realization that RBPs exert important roles in normal post transcriptional regulation of gene expression and in disease, approaches for their analysis have received a great deal of attention. Many methods to study RBPs, their binding domains and sites, their mode of action and their network have been developed.[121,131,157] Meanwhile, the number of known RBPs has increased to more than a thousand, amongst them several hundred that lack apparent links to RNA-binding and/or known RBDs (enigmaRBPs)[131,158]. Identification and characterization of new RBPs, including enigmaRBPs, require more than sequence and/or structural information of classical RBDs and depend on direct experimentation. Due to increasing interest and progress in the past years, the number of methods for studying RBPs and their interactions has considerably expanded. In general, there are two main types of methods: protein-centric approaches to study interactions of given RBPs and RNA-centric or global approaches for a proteome-wide discovery of RBPs.[158,159]

Protein-centric approaches to study interactions of given RBPs

Earliest approaches to determine binding specificity and targets of RBPs of interest were based on RNA immunoprecipitation (IP) followed by either sequencing (RIP-seq) or microarray analysis (RIP-chip) to detect the co-purified RNAs[121,160–163]. However, the RNA-RBP complexes are formed under non-physiological condition in cell lysates and the stringency of wash conditions of RIP approaches requires careful evaluation to minimize non-specific interactions. Such IPs allow only relatively mild conditions to preserve the RNA-RBP interactions and RNA is a sticky molecule. Thus, significant backgrounds can hamper data analysis or RNAs with low binding affinity to the RBPs may not be recovered.[159,164,165] These limitations have been overcome with the development of methods involving covalent crosslinking of RBPs to their target RNAs in living cells; referred to as crosslinking-immunopurification (CLIP) strategies[159,164]. Covalent crosslinking preserved RNA-RBP complexes formed *in vivo* and allows denaturing conditions and thus a more stringent purification of RBP-RNA complexes. Disruption of protein-protein interactions results in reduced off-target RNAs that only indirectly associate with the RBP of interest[165]. Conventional UV-crosslinking at a wavelength of 254 nm creates short-lived radicals at naturally photoreactive

nucleotide bases that can attack and form covalent bonds with amino acids in very close proximity, which is typically only the case for direct interactions[157,165,166]. However, conventional crosslinking efficiency is generally low (5 %)[164]. Yields can be increased by metabolic RNA labeling with UV-reactive nucleoside analogs, such as 4-thiouridine (4SU), and crosslinking at 365 nm (Photoactivatable ribonucleoside CLIP; PAR-CLIP)[167]. Several CLIP approaches have been developed, amongst them revised protocols that use high-throughput sequencing (HITS) or allow individual-nucleotide resolution CLIP (iCLIP)[158,164,165]. Very recent improvements to the efficiency of key biochemical reactions of iCLIP protocols facilitate mid-throughput library generation and led to the development of infrared CLIP (irCLIP) and enhanced CLIP (eCLIP). Since, the number of data sets generated by eCLIP is continuously growing.[129,158,165,168] For more detailed features of the individual CLIP protocols the reader is referred to recent reviews by Wheeler *at al.* (2018)[165] and Ramanathan *et al.* (2019)[158].Although further approaches to determine the RNA targets and specific binding sites of RBPs have been developed, both RIP and CLIP-based approaches, are the state-of-the art methods in the field[159,165].

Global approaches for a proteome-wide discovery of RBPs

Global, RNA-centric approaches to identify RBPs have long been used in the field of RNA research but have quickly advanced over the past years[117,131,169]. Initial approaches to identify the interactome of poly(A)-RNA were based oligo(dT) sepharose chromatography and have purified RNA-associated proteins from untreated cells[170], UV-irradiated polysomal fractions[171] and intact cells[172]. Years later, first global screens were based on the hybridization of fluorescently labelled mRNA to protein microarrays and facilitated the identification of ~200 RBPs from yeast[173]. A further approach screened for RBPs using immobilized RNA baits and stable isotope labeling by amino acids in cell culture (SILAC)-labeled mammalian cell extracts combined with protein identification by mass spectrometry (MS)[174]. Such screens are prone to pick up nonspecific interactions as the RNA-RBP complexes form *in vitro* under non-physiological condition and because direct RNA-proteins interactions cannot be distinguished from indirect protein-protein interactions[117,142]. To address these limitations, two laboratories developed a new approach to simultaneously recover and characterize the whole RNA-binding proteome *in vivo*: RNA interactome capture (RIC)[169,175]. This method combines UV-crosslinking, oligo(dT) capture and quantitative MS to isolate and identify RBPs bound to their mRNA targets directly in living cells (**Figure 12**)[159].

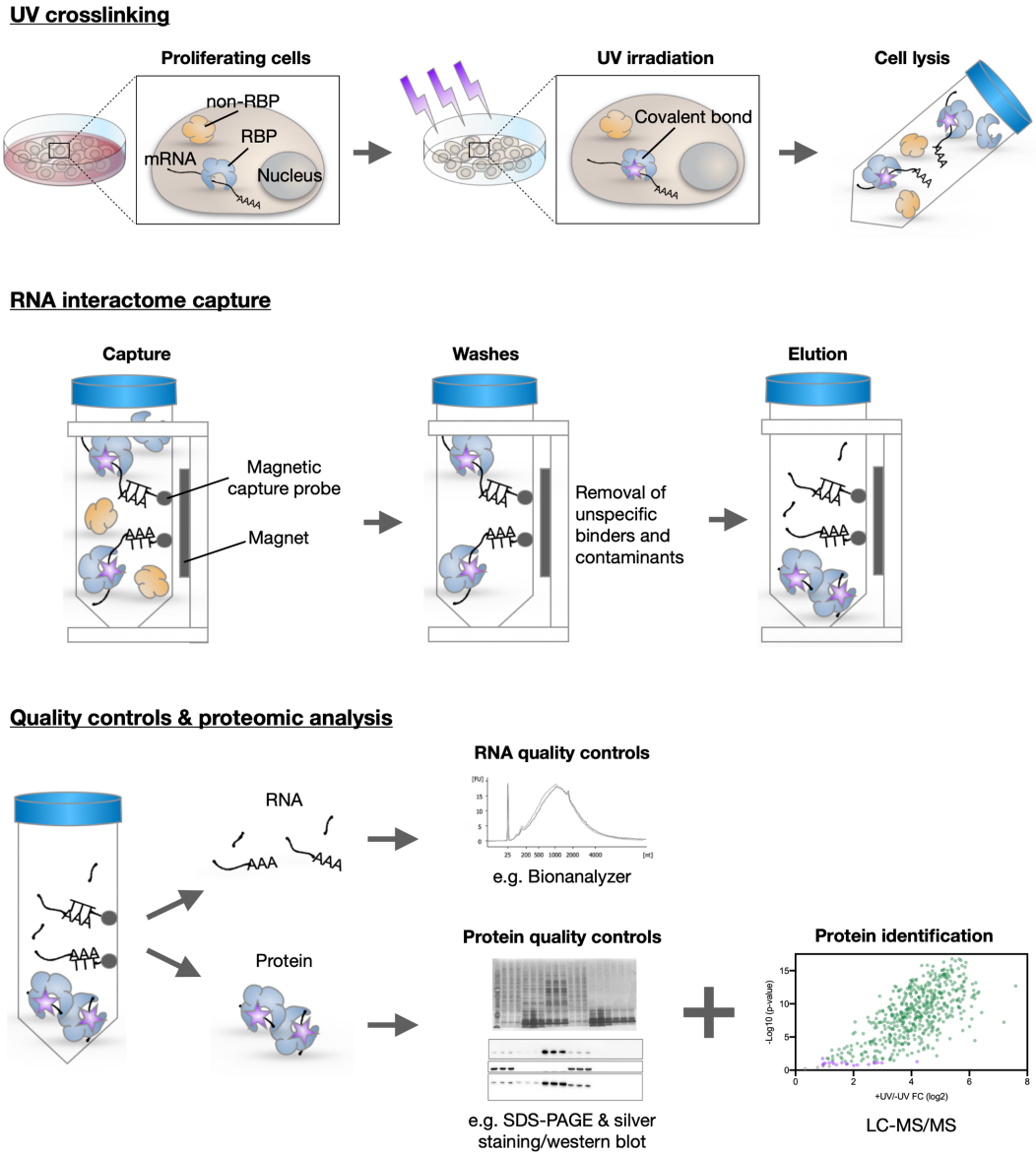


Figure 12: Overview of the RNA interactome capture methodology. In a first step, the *UV-crosslinking*, monolayers of proliferating cells are exposed to UV-light crosslinking RBPs to their target mRNAs. Upon cell lysis, the actual *RNA interactome is capture* in performed. UV-crosslinked RBPs are captured via the poly(A)-tail of their target RNA by adding magnetic probes hybridizing to the poly(A) stretches. The samples are subjected to a series of stringent washes to remove unspecific binders and contaminants. Next, the RBPs are eluted either by disrupting the hybridization of the pol(A)-tail to the oligo(dT)-based probe (heat, RIC) or by cleaving the target RNA (RNases, eRIC). The capture is monitored by *quality controls* on RNA (Bioanalyzer) and protein (SDS-PAGE followed by silver staining and western blot) level before the protein eluate is subjected to *proteomic analysis*. Using liquid chromatography-tandem mass spectrometry (LC-MS/MS), the captured proteins identified and the RNA interactome is determined. Figure modified from Castello *et al.* (2016)[176].

For RIC, monolayers of living cells are irradiated with UV light[169,175]. Both, conventional UV-crosslinking at 254 nm[175] or 4SU incorporation and subsequent photoactivatable-ribonucleoside-enhanced crosslinking (PAR-CL) at 365 nm[169], can be used[177]. As described in the previous paragraph, proteins and RNA in direct contact become covalently linked; i.e. RBP-RNA interactions are "frozen" under physiological conditions. Upon cell lysis, RBPs bound to polyadenylated RNAs are indirectly captured by hybridization to oligo(dT) magnetic beads and

purified by a series of stringent washes. Next, the RBPs are released from the RNA by RNase treatment, cleaved into peptides by tryptic digest and analyzed on a liquid chromatography-tandem mass spectrometry (LC-MS/MS) platform.[175] To identify the isolated RBPs either label-free, stable isotope labeling (SILAC) or isobaric labeling (Tandem Mass Tag; TMT) quantification methods are applicable[169,175,178]. Proteins significantly enriched in crosslinked samples over non-crosslinked controls are identified as RBPs and represent the RNA interactome[175]. Typically, the quality of RIC is monitored in parallel on the RNA and protein levels. For RNA analysis, RT-qPCR or analysis by Bioanalyzer can be used to confirm the enrichment of mRNAs. Proteins isolated by RNA interactome capture can be analyzed by gel electrophoresis, combined with silver staining and western blotting against well-known RBPs (positive control), or non-RBPs (negative control) to confirm selectivity and specificity.[175,177] Originally, this method has been developed with HeLa and HEK293 cells but meanwhile it was adapted to different model systems, including yeast[179,180] (*S.cerevisiae*), parasites[181,182] (*Plasmodium, Trypanosoma*), plants[183,184] (*A. thaliana*) and flies[185,186] (*D. melanogaster*)[131].

A few years after the development of RIC, Castello *et al.*[128] extended their approach by introducing a partial protease digest after elution, followed by a further round of oligo(dT) capture. This new approach, named RBDmap, allows to determine the regions within RBPs interacting with RNA.[120,128] These RIC-based studies revealed many new and unconventional RBPs without classical RBDs but rich in disordered regions and discovered unexpected RBPs with enzymatic activity[131,159]. However, RIC has shown limitations: contaminations with rRNA, DNA-binding and other proteins that could be falsely identified as mRNA-binders have been reported[178,187]. To overcome these limitations, our laboratory further developed the original RIC protocol and devised enhanced RNA interactome capture (eRIC)[178]. This new method employs locked nucleic acid (LNA) oligonucleotides instead of oligo(dT) magnetic beads. LNA nucleosides are nucleic acid analogues with a methylene bridge to "lock" the ribose in an ideal position for Watson-Crick binding. Thus, incorporation of LNA nucleosides into RNA oligonucleotides improves the hybridization properties with the complementary strand. As a result, LNA-based probes exhibit increased affinity and higher thermal stability compared to regular RNA oligonucleotides.[188,189]. Thus, eRIC combines several advantages and is superior to other approaches: Crosslinking induces covalent bonds only at "zero distance" (≤2 Å), does not promote protein-protein crosslinking and is applied to intact cells; i.e. "freezes" native RBP-RNA interactions. Due to the use of LNA-based probes, the hybridization stability persists even under

stringent purification conditions (high temperature, salt, detergents) which remove unspecific binders and empower RBPs isolation with greater specificity.[175,178] Therefore, eRIC facilitates comparative analyses of changes in the RNA interactome, for instance before and after treatment[178], and thus, opened the opportunity to directly investigate the role of RBPs in T-ALL in general and to test the effect of modulating PTMs on these RBPs in particular.

2 AIMS

Studies on genetic and epigenetic alterations provided insights on the transcriptional level of gene expression in T-ALL. However, gene expression is also regulated on post-transcriptional levels by processes such as mRNA splicing, transport, surveillance, decay and translation. RBPs are the major players and global coordinators of these posttranscriptional steps.[116,117] Thus, disturbed interactions between RBPs and their target RNAs emerge as a novel principle in human diseases including cancer[117,125,147].

The activity of RBPs is regulated and fine-tuned though PTMs and the binding sites of RBPs were reported to be hotspots for such PTMs, including lysine acetylation[119,121,128,131]. Moreover, disease-associated mutations at or close to PTM sites have been found and suggest that aberrantly regulated and post-transcriptionally modified RBPs are involved in tumorigenesis[117]. As HATs and HDACs have a broad target spectrum, including a number of non-histone proteins, RBPs are likely subject to the reversible acetylation by HATs and HDACs[190,191]. Analogous to histones and DNA, acetylation of RBPs may influence their binding affinity to RNA. Thus, HDAC inhibition may alter the acetylation status of RBPs and impact on RBP-RNA interactions and ultimately result in changes of the RNA metabolism. Furthermore, elevated HDAC activity and antileukemic activity of HDAC inhibitors in ALL and T-ALL have been reported[53].

Therefore, the overarching goal of this project was to define the RNA interactome of T-ALLs and normal T-cell progenitors. Further, it was hypothesized that RBP activity may be modulated by HDAC inhibition; a therapeutic principle originally conceptualized for epigenetic modulation of the transcriptional profile of tumor cells, which may also regulate the function of non-histone proteins in general and of RBPs in particular.

In order to achieve this goal, this work aimed at answering the following specific questions:

1) How does the mRNA interactome of T-ALLs differ from that of normal T-lymphocyte precursors?
2) How does the mRNA interactome change upon HDAC inhibition using the HDAC inhibitor SAHA?

3 MATERIALS & METHODS

All general reagents and solvents were obtained from Carl Roth (Karlsruhe, DE), Sigma -Aldrich (St. Louis (MO), US), Merck Millipore (Darmstadt, DE) or Thermo Fischer Scientific (Waltham (MA), US) unless otherwise stated.

3.1 General biochemical and molecular biology methods

Commercial reagents, kits and assays were utilized according to the manufacturer's recommendations unless otherwise indicated.

3.1.1 DC protein assay

Protein quantification was carried out using the DC (detergent compatible) protein assay (Bio-Rad Laboratories, Hercules (CA), US). This colorimetric assay is based on a two-step reaction: A first reaction of protein and copper in an alkaline copper tartrate solution (Reagent A) and second a reduction of folin in a dilute folin solution (Reagent B) by the copper-treated protein[192,193]. A concomitant color formation is the consequence of the protein (especially the amino acids tyrosine and tryptophan) losing oxygen atoms during the reaction, producing reduced species of characteristic blue color. The blue product can be detected using a spectrophotometer at a maximum absorbance of 750 nm and is dependent on concentration of protein with the sample.[193r]

Standard assay protocol in cuvettes

The total protein amount of cell lysates for RNA interactome capture experiments was determined using the standard assay protocol. For each measurement a standard curve of five dilutions of purified BSA (10 mg/ml; New England BioLabs, Ipswich (MA), US) between 0 – 1.2 µg/µl was generated using the respective buffer of the sample. Samples were diluted by a factor of four in water. Following the instruction manual 100 µl of sample/standard were mixed with 500 µl Reagent A' (ratio: 20 µl of Reagent S to each 1 ml of Reagent A) before the addition of 4 ml Reagent B. After 15 – 20 min incubation in the dark, the samples were measured in cuvettes at 650 nm using a BioPhotometer plus (Eppendorf, Hamburg, DE).

[r] Bio-Rad Laboratories: DC Protein Assay Instruction Manual (https://www.bio-rad.com/webroot/web/pdf/lsr/literature/LIT448.pdf); Requested: 2020.04.01

Microplate assay protocol

Quantification of protein prior to scintillation counting, SDS-PAGE and immunoblotting or autoradiographic analysis was performed using the microplate assay protocol. A standard curve of six dilutions between 0 – 1.5 µg/µl was prepared using purified BSA (10 mg/ml, New England BioLabs, Ipswich (MA), US) and the corresponding sample buffer. Samples were diluted 1:4 in water. For each standard/sample 5 µl were plated into a costar 96-well clear flat bottom polystyrene microplate (Corning, New York (NY), US) prior to the addition of 25 µl Reagent A' (ratio: 20 µl of Reagent S to each 1 ml of Reagent A) and 200 µl Reagent B. Bubbles were removed using a lighter and the plate was incubated for 15 – 20 min in the dark. Microplates were measured using a SpectraMax M2 microplate reader (Molecular Devices, San Jose (CA), US). Each standard/sample was measured in triplicates.

3.1.2 SDS-PAGE analysis

For separation of proteins by molecular weight, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Proteins were denatured and uniformly negatively charged using a loading buffer containing the ionic detergent SDS. After applying an electric field, the SDS-bound proteins thus migrate through the polyacrylamide matrix towards the positively charged electrode at a rate proportional to their charge density. Due to a sieving effect of the gel, proteins with higher molecular weight will migrate more slowly through the gel matrix compared to proteins of higher molecular weight, leading to protein separation.[s] Here, the SDS-PAGE analysis was performed using either the XCell SureLock Mini-Cell Electrophoresis System (Life Technologies, Carlsbad (CA), US) or the Novex Midi Gel Electrophoresis System (Thermo Fischer Scientific, Waltham (MA), US). For regular cell lysates aliquots of 5 – 15 µg total protein per lane and for (enhanced) RNA interactome samples 0.01 – 0.1 % of the input and ~10 % of the final eluate per lane were prepared. Samples were mixed with loading buffer (6x loading buffer (7 ml 4x Tris-CL/SDS pH 6.8, 3.6 ml glycerol, 1 g DTT, 1 mg bromophenol blue, aliquots stored at -20 °C); 4x Tris-Cl/SDS pH 6.8 (0.5 M Tris-Cl, 0.4 % SDS)) and heated for 10 min at 70 °C in a thermomixer (Eppendorf, Hamburg, DE). Between 1 – 5 µl of either the PageRuler Prestained Protein Ladder (10 to 180 kDa) or the PageRuler Plus Prestained Protein Ladder (10 to 250 kDa),

[s] Thermo Fisher Scientific web page: Overview of Electrophoresis (https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-electrophoresis.html); Requested: 2020.04.02

served as size standards. Proteins were resolved on 4 – 12 % Bis-Tris pre-cast protein gels corresponding to the respective electrophoresis system using NuPAGE MOPS SDS running buffer for 40 – 50 min at 200 V using a Consort EV265 power supply (Turnhour, BEL). Next, the separated proteins were visualized either by staining, by transfer onto a membrane for detection by immunoblotting or on a film for autoradiographic analysis.

3.1.3 Coomassie staining of protein polyacrylamide gels

For Coomassie blue staining of proteins in a gel, InstantBlue Ultrafast Protein Gel Stain (Expedeon, Heidelberg, DE) was used. In the acidic staining solution Coomassie brilliant blue acts as an anion and reacts with the NH_3^+ residues of proteins, resulting in color formation[194]. After SDS-PAGE, the gel was removed from the cassette and rinsed with ultra-pure water before incubation in InstantBlue solution on a DRS-12 platform shaker (NeoLab, Heidelberg, DE) for at least 30 min. Next, the gel was washed repeatedly in ultra-pure water to remove background and improve sensitivity prior to documentation using a Fusion-FX7 Spectra imaging platform (Vilber Lourmat, DE).

3.1.4 Silver staining of protein polyacrylamide gels

Silver staining was performed using the SilverQuest silver staining kit (Life Technologies, Carlsbad (CA), US). This staining is based on color development on protein bands due to a reduction of silver ions to metallic silver[t]. First, the gel was washed in ultrapure water before it was fixed in fixing solution on a DRS-12 platform shaker (NeoLab, Heidelberg, DE) for at least 20 min, but usually overnight. The following steps were carried out according to the instruction manual. Ultimately, the stained gel was incubated in ultrapure water on the platform shaker overnight prior to documentation using a Fusion-FX7 Spectra imaging platform (Vilber Lourmat, DE).

[t] Novex by Life Technologies: SilverQuest™ Silver Staining Kit – User guide (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/silverquest_man.pdf); Requested: 2020.04.02

3.1.5 Western Blot analysis

Detection of specific proteins was carried out by western blot analysis. Proteins were separated using SDS-PAGE (as described in section 3.1.2) and transferred onto a membrane. For electrophoretic transfer, the protein-containing gel is placed in direct contact with the membrane and an electric field is applied. The proteins migrate from the gel towards the positively charged electrode onto the membrane. Once on the membrane, proteins become immobilized by hydrophobic interactions and detectable by immunological procedures.[195u] Here, western blot was performed under semi-dry conditions. A polyvinylidene difluoride (PVDF) membrane (0.2 µm Hybond-P, GE Healthcare; Amersham, GB) was activated for at least 1 min in methanol.

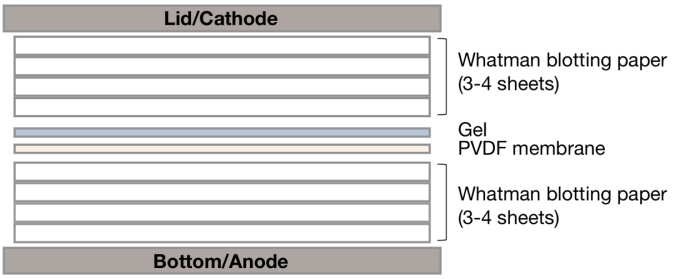


Figure 13: Schematic assembly of a western blot sandwich for semi-dry protein transfer. The sandwich includes 3 – 4 sheets of Whatman blotting paper soaked in anode buffer on the bottom, 3 – 4 sheets of Whatman blotting paper soaked in cathode buffer on the top and an activated PVDF membrane with the gel in close contact. The PVDF membrane is positioned between the gel and the anode to enable the negatively charged proteins to migrate from the gel onto the membrane.

Three to four sheets of Whatman blotting paper (3 mm Chr, GE Healthcare; Amersham, UK) were soaked in anode buffer (1.8 g Tris base, 50 ml Methanol, to 500 ml H_2O) and cathode buffer (1.5 g Tris base, 2.6 g 6-Aminohexanoic acid, 50 ml Methanol, 250 µl 20 % SDS, to 500 ml H2O), respectively, and assembled in a western blot sandwich as shown in **Figure 13**. Proteins were transferred using a Biometra Fastblot blotting chamber (Analytik Jena, Jena, DE) in combination with a Major Science MP-300V power supply (Saratoga (CA), US) at 1 mA per 1 cm^2 membrane for 1.5 – 2 h. To prevent unspecific binding, the membrane was blocked in blocking solution (5 % milk in TBST (20 mM Tris pH 7.7, 150 mM NaCl, 0.03 %Tween)) at RT for at least 1 h. After blocking, the proteins of interest were specifically detected by immunodetection using primary antibodies (Table 2) at indicated concentrations and incubation times under gentle agitation. Next, the membrane was rinsed three times with TBST for at least 1 min each, before incubation for 1 at RT with the corresponding peroxidase (HRP) coupled secondary antibody (Table 2) under gentle agitation. After three washes with TBST for at least 10 min, the membrane was drained

[u] Thermo Fisher Scientific web page: Overview of Western Blotting (https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-western-blotting.html); Requested: 2020.04.02

from superfluous TBST and imaged using the Western Lightning Plus-ECL Enhanced Luminol Reagent and Oxidizing Reagent (PerkinElmer, Waltham (MA), US) in a Fusion-FX7 Spectra imaging platform (Vilber Lourmat, DE).

Table 2: Primary and secondary antibodies for immunodetection.

Antibody	Source	Dilution; Incubation	Supplier; Number
Acetylate lysine	rabbit	1:1000 (5 % BSA in TBST); o/n	Cell Signaling Technology, Danvers (MA), US; 9441S
CSDE1/UnR	rabbit	1:1000 (5 % milk in TBST); o/n	Proteintech, Rosemont (IL), US; 13319-1-AP
Histone H3	rabbit	1:1000 (5 % milk in TBST); o/n	Cell Signaling Technology, Danvers (MA), US; 9715S
Histone H4	mouse	1:1000 (5 % milk in TBST); o/n	Abcam, Cambridge, UK; Ab31830
HnRNP K	rabbit	1:1000 (5 % milk in TBST); o/n	Proteintech, Rosemont (IL), US; 11426-1-AP
HuR	rabbit	1:10,000 (5 % milk in TBST); o/n	Proteintech, Rosemont (IL), US; 11910-1-AP
α-tubulin	mouse	1:5000 (5 % milk in TBST); 1 – 1.5 h	Sigma-Aldrich; St. Louis (MO), US; T5168
β-actin	mouse	1:10,000 (5 % milk in TBST); 1 – 1.5 h	Sigma-Aldrich; St. Louis (MO), US; A1978
anti-rabbit (HRP-conjugated)	goat	1:5000 (5 % milk in TBST); 1 – 1.5 h	Sigma-Aldrich; St. Louis (MO), US; A9044
anti-mouse (HRP-conjugated)	rabbit	1:10,000 (5 % milk in TBST); 1 – 1.5 h	Sigma-Aldrich; St. Louis (MO), US; A0545

3.1.6 Global protein synthesis

Global protein biosynthesis was analyzed by metabolic labeling experiments. Cells were cultured in methionine-free media supplemented with a radiolabeled source of the low-energy beta-emitting radioisotope ^{35}S methionine. Newly synthesized methionine-containing proteins are isolated prior to analysis by SDS-PAGE and exposure to a film (autoradiography) or quantification by scintillation counting. For autoradiographs, the Coomassie stained SDS-PAGE gel serves as control to confirm equal protein loading, whereas the ^{35}S incorporation is determined by scintillation counting is normalized by the total protein concentration.

Cells were seeded at a density of $0.5x10^6$ – $0.8x10^6$ cells/well in 2 ml complete RPMI media in six-well plates (Sarstedt, Leicester, UK). Post seeding, plates were incubated overnight in the incubator before treatment with 2 µl of either 100 % DMSO (0.1 % final conc., negative control) or 25 mg/ml cycloheximide (CHX) dissolved in 100 % DMSO (25 µl/ml final conc., positive control) or SAHA at 0.5 mM, 1.0 mM, 2.5 mM and 5 mM in 100 % DMSO (0.5 µM, 1.0 µM, 2.5 µM and 5 µM final conc.) for 3 and 7 h. Then, cells were washed in 5 ml tubes with 3 ml PBS by 5 min

of centrifugation at 400 g and aspiration of the supernatant before the complete RPMI media was replaced by depletion medium (Gibco RPMI without methionine, 5 ml dialyzed FBS, 500 µl P/S) containing DMSO, CHX and SAHA at concentrations as stated above. After 30 min incubation in the incubator to deplete cells of methionine, newly synthesized proteins were labeled for 30 min by the addition of 4 µl ^{35}S methionine (10 mCi/ml, Hartmann Analytic, Braunschweig, DE) per well (full treatment time of 4 h and 8 h, see **Figure 14**).

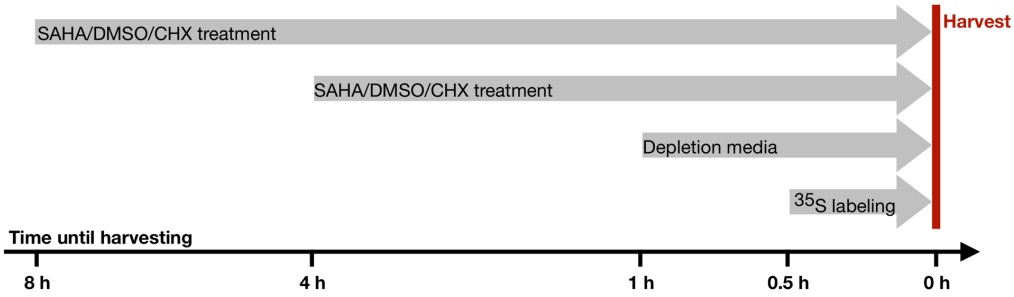


Figure 14: Experimental overview of the global protein synthesis analysis. Cells were treated with SAHA and controls (DMSO, CHX) for 8 h and 4 h before harvesting. 1 h before harvesting, cells were washed and complete RPMI was replaced by depletion media containing SAHA and controls (DMSO, CHX). Methionine was depleted from internal stores for 0.5 h before ^{35}S methionine was added. After 0.5 h ^{35}S methionine incorporation into newly synthesized proteins, cells were harvested and lysed.

Before harvesting, cells were transferred to 2 ml tubes and washed twice with 1.5 ml PBS by 5 min of centrifugation at 400 g and aspiration of the supernatant. Next, cells were lysed in 100 µl ice-cold RIPA buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 % NP-40, 0.1 % SDS, 0.5 % Deoxycholate), by vortexing and incubation on ice for 20 min. After 5-10 min centrifugation at 4 °C and 16,000 g the supernatant was recovered and the lysate was stored at -20 °C. Protein concentration of the lysate was determined by DC protein assay.

Autoradiography

For autoradiographic analysis, lysate equaling 10 µg total protein per sample were separated by SDS-Page as described in section 3.1.2. To visualize proteins, the gels were stained using Coomassie blue (see section 3.1.3), before they were photographed and dried on Whatman blotting paper (3 mm Chr, GE Healthcare; Amersham, UK) in a Biometra mididry D62 gel tryer (Analytik Jena, Jena, DE) for 2 h at 80 °C. Radioactive signal was detected by placing an autoradiography film (Amersham Hyperfilm MP High Performance, GE Healthcare, Upsala, SE) on the dried gel for several hours. Films were developed using a Kodak M35 X-OMAT Processor (Kodak, Rochester (NY), US). Coomassie stained gels served as loading control.

Quantification by scintillation counting

To quantify the newly synthesized proteins by scintillation counting, 10 µl lysate were spotted on 10 mm glass microfiber filters (Whatman, Little Chalfont, UK) and allowed to dry. Then, proteins were precipitated by incubation in ice-cold 15 % TCA for at least 30 min on ice, before washing once in ice-cold PBS and twice in ice-cold 100 % EtOH. The filters were dried by placing on Whatman blotting paper (3 mm Chr; GE Healthcare; Amersham, UK) before submerging in scintillation buffer (Rotiszint eco plus, Carl Roth, Karlsruhe, DE). Radioactivity of methionine-labeled proteins was quantified by scintillation counting for 10 min per sample in a 1414 liquid scintillation counter (Wallac WinSpectral, PerkinElmer, Waltham (MA), US). Measured scintillation counts were normalized by the total protein concentration and visualized by GraphPad Prism 8. Data were plotted as means ± SD (Standard Deviation) and analyzed for a statistical difference using ANOVA (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ and ****$p \leq 0.0001$. $p \leq 0.05$ was considered as significant).

3.1.7 Bioanalyzer

RNA and RNA fragments were analyzed using the RNA 6000 Pico Kit (Agilent Technologies, Waldbronn, DE). Nucleic acids are separated based on their size when migrating through a set of microchannels within an RNA chip in an electric field[v]. Crude RNA from crude lysates was extracted using either an RNA Clean & Concentrator-5 or 25 Kit (Zymo Research, Irvine (CA), US) prior to DNase treatment using the TURBO DNA-free Kit (Life Technologies, Carlsbad (CA), US) following the manufacturer's instructions. If necessary, the RNA was afterwards concentrated by applying the RNA Clean & Concentrator-5 kit. Concentration of RNA samples was measured using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham (MA), US) and aliquots were diluted to 5 – 10 ng/µl in RNase free water (Aqua ad iniectabila Mini-Plasco connect; Braun, Melsungen, GE) for analysis. RNA analysis was performed as described in the manufacturer's Quick Start Guide using 1:150 diluted Nano Chip RNA Ladder on an Agilent 2100 Bioanalyzer system (Agilent Technologies, Waldbronn, DE).

[v] Agilent RNA 6000 Pico Kit Quick Start Guide: (https://www.agilent.com/cs/library/usermanuals/Public/G2938-90049_RNA6000Pico_QSG.pdf); Requested: 2020.04.03

3.1.8 RT-qPCR analysis

Quantitative analysis of RNA was performed by real-time polymerase-chain-reaction (qPCR). For qPCR analysis, RNA is first transcribed into complementary DNA (cDNA) using the enzyme reverse transcriptase (cDNA synthesis). The resulting cDNA serves as a template for the subsequent qPCR reaction. During qPCR, synthesis of the corresponding DNA strand by a polymerase leads to the formation of a new DNA double-strand. Incorporation of a double-stranded DNA-binding dye (e.g. SYBRGreen) into this DNA double strand results in fluorescence signal and allows to quantify the amount of amplified product. For real-time PCR, the fluorescence intensity and the corresponding template quantity is measured while the amplification is still progressing. Here, RNA of 5 – 7 µl of non-crosslinked RIC eluates and eRIC heat-eluates was retrotranscribed to cDNA. First, volumes of the RNA samples were adapted to 11 µl using Aqua (Aqua ad iniectabila Mini-Plasco connect; Braun, Melsungen, GE) before they were mixed with 1 µl dNTPs (10 mM, dNTP-Set 1, Carl Roth, Karlsruhe, DE) and 1 µl random hexamers (300 ng/ µl, Invitrogen by Thermo Fisher Scientific, Waltham (MA), US) followed by an incubation at 65 °C for 5 min and cooling down on ice for at least 1 min. Next, 4 µl 5x First-Strand buffer, 1 µl DTT (0.1 M), 1 µl RiboLock RNase-Inhibitor (Waltham (MA), US) and 1 µl SuperScript III Reverse Transcriptase (Invitrogen by Thermo Fisher Scientific, Waltham (MA), US) were added as a master mix and the reaction was performed for 5 min at 25 °C, 60 min at 50 °C and 15 min at 70 °C. For control reactions and genomic DNA (gDNA) analysis, reverse transcriptase was omitted and replaced by Aqua (Aqua ad iniectabila Mini-Plasco connect; Braun, Melsungen, GE). cDNA was diluted 1:10 to prepare a working solution. Purified total RNA was retrotranscribed and served as standard for further analysis. qPCR was performed as published before, using PrimaQuant SYPRGreen Master Mix with ROX (Steinbrenner Laborsysteme, Wiesenbach, DE) and a StepOnePlus Real-Time PCR System (Life Technologies, Carlsbad (CA), US) together with previously published primer pairs[178,196].

All primers from 5′ to 3′ (forward: f, reverse: r):

- 28 S rRNA: f: TTACCCTACTGATGATGTGTTG TTG
  r: CCTGCGGTTCCTCTCGTA
- RPS6: f: TGAAGTGGACGATGAACGCA
  r: CCATTCTTCACCCAGAGCGT
- 18 S rRNA: f: GAAACTGCGAATGGCTCATTAAA
  r: CACAGTTATCCAAGTGGGAGAGG

3.2 Cell culture

All cell culture work was carried out under sterile conditions within a laminar air flow hood (Safe 2020 Class II Biological Safety Cabinet, Thermo Fisher Scientific, Waltham (MA), US). The working surface and any working materials used inside the hood were wiped with 70 % ethanol before use. Only cell culture grade sterile reagents, plastic and glass ware was used and opened/closed exclusively within the hood. To prevent any mycoplasma contamination, cell lines were handled in a mycoplasma-free culture room apart from primary cells and the working space was regularly cleaned with PromoKine mycoplasma ExS spray (PromoCell, Heidelberg, DE). General plasticware and tissue culture consumables was obtained from Sarstedt (Leicester, UK) unless otherwise stated.

3.2.1 Cell lines and general culture conditions

All T-ALL cells used for this study were established cell lines originating from peripheral blood of T-ALL patients and were obtained from the German Collection of Microorganisms and Cell Cultures (Leibniz Institute DSMZ, Braunschweig, DE).

Established human T-ALL cell lines:

- DND41 (ACC 525): 13-year-old, male, cortical T-ALL
- JURKAT (ACC 282): 14-year-old, male in relapse, mature T-ALL
- MOLT3 (ACC 82): 19-year-old, male in relapse, cortical T-ALL
- P12/ICHIKAWA (ACC 34): 7-year-old, male, cortical T-ALL

All cell lines were generally cultured as a suspension in TC flasks with hydrophobic growth surface in Gibco RPMI 1640 (1X, [+] L-Glutamine; Thermo Fisher Scientific, Waltham (MA), US) medium supplemented with heat-inactivated 10 % Fetal Bovine Serum (FBS; Gibco, Thermo Fisher Scientific, Waltham (MA), US) and 1 % Penicillin-Streptomycin solution (Sigma -Aldrich St. Louis (MO), US); in the following referred to as complete media. Cells were maintained at 37 °C within a humidified atmosphere of 5 °C CO_2 in a BINDER CB series CO_2 incubator (Binder, Tuttlingen, DE).

3.2.2 Sub culturing of T-ALL cells

Cells were passaged at least three times a week, depending on their density. Cell density and concentration were assessed by microscopic inspection using an CKX41 inverted microscope (Olympus Corporation, Tokyo, JP) and a TC20 automated cell counter (Bio-Rad Laboratories, Hercules (CA), US) to obtain live cell counts. For general propagation, cell cultures of high density and/or with exhausted medium were diluted with fresh and pre-warmed (~37 °C) complete media. Only for cell cultures with very high density or acidic pH (yellowish coloring of phenol red) media was changed completely. Cells were gently centrifuged for 5 min at 400 g and the depleted media was decanted before resuspension in fresh and pre-warmed (~37 °C) complete media at a lower seeding density. All cells were maintained between ~$0.3x10^6$ – $1.5x10^6$ cells/ml with DND41 cells generally growing better in higher cell densities.

3.2.3 Collection and preparation of thymi

Thymocytes were obtained by disaggregating human thymus tissue obtained from children undergoing corrective cardiac surgery (without extra cardiac diagnosis)[w]. All thymus samples were transported in ice-cold complete medium and prepared immediately after collection. Throughout disaggregation and during later down-stream processing, thymus samples and resulting thymocyte suspensions were constantly kept in ice-cold medium/ PBS or on ice unless otherwise stated. For disaggregation, collected thymi were sliced into thin pieces and placed between two pieces of nylon mesh (100 μm; Technical Fabrics, Denbigh, UK) in a petri dish with complete media. By very gently pressing the plunger of a 5 ml syringe on the thymus slices, the tissue was disaggregated and thymocytes were dissociated and extracted through the mesh. Next, the extracted thymocytes were collected and passed through a 40 μm cell strainer cap (Falcon, Corning (NY), US) into 50 ml centrifuge tubes, centrifuged for 5 min at 400 g and carefully resuspended in ice-cold PBS. Throughout the course of this project it was found that during resuspension, care needs to be taken to avoid resuspension of the lower and red part of the pellet. Resuspended cells were transferred again through a 40 μm cell strainer cap into fresh 50 ml centrifuge tubes while avoiding to transfer turbid and highly viscous slime of cell suspension. The remaining viscous slime and the red parts of the pellet were discarded. These steps helped to increase the purity of viable, cortical T-cell progenitors for later stages of the

[w] Thymus was obtained only after written consent of the donor's parents was given and with approval of the ethics committees of the EMBL and of the University Hospital Heidelberg.

processing. Then, cell number and viability of the isolated thymocytes was determined by trypan blue staining and counting of the cells using a TC20 automated cell counter (Bio-Rad Laboratories, Hercules (CA), US). Thymocyte suspensions with sufficient high cell numbers and viability were further processed for either flow cytometry, FACS, MACS or UV-crosslinking.

3.2.4 Cryopreservation

To minimize genetic changes and avoid aging, cells were cryopreserved for long-time storage and sub cultured for up to ~25 passages before another ampoule of cells was resuscitated. For cryopreservation, cells were pelleted by gentle centrifugation for 5 min at 400 g and the medium was removed by careful aspiration of the supernatant. Next, the cell pellet was gently re-suspended in freezing media consisting of FBS with 10 % DMSO (CryoMACS, Miltenyi Biotec, Bergisch Gladbach, DE) to a concentration of ~$1.0x10^7$ cell/ml and aliquoted (1 ml per vial) into 1.8 ml Nunc Cryogenic storage vials (Roskilde, DK). Cells were slowly frozen by placing the vials in a Mr. Frosty freezing container in a -80 °C freezer (MDF-U7 4V Ultra-Low Temperature Freezer; SANYO Electric, Etten-Leur, NL) over night before transferring to a liquid nitrogen storage tank (MVE XC47/11; PrincetonCryo, Pipersville (PA), US). To resuscitate cells, the cryovials containing frozen cells were removed from the liquid nitrogen storage and immersed into a 37 °C water bath (SUB AQUA Pro; Grant Instruments, Cambridge, UK). As soon as the cell suspension was almost completely thawed, it was transferred into 50 ml centrifuge tubes and slowly diluted with complete culture media at ~4 °C to prevent heat-shock effects. The remaining DMSO was removed by centrifugation at 400 g for 5 min and decanting the supernatant. Cells were gently resuspended in 5-10 ml fresh, pre-warmed (~37 °C) complete media and transferred into a T25 TC flask. To optimize recovery, cells were directly placed in the incubator over night before they were counted and appropriately diluted to ~$0.8x10^6$ cell/ml the next morning. Cells were allowed to recover for about two weeks before experiments were carried out.

3.2.5 SAHA-treatment of T-ALL cell lines

SAHA (Vorinostat) was obtained from Selleckchem (Houston (TX), US) and Tocris Bioscience (Bristol, UK) and stored as stock solution of 10 mM in 100 % DMSO in aliquots at -80 °C. All centrifugations were performed for 5 min at 400 g unless otherwise stated.

Identification of appropriate treatment conditions

For the identification of acetylation-dependent alterations in the RNA-binding proteome, it was important to induce protein hyperacetylation without inducing major secondary effects. To select such treatment conditions, different SAHA concentrations and incubation times were tested. Cells were suspended at a density of ~0.5x10^6 cells/ml in fresh complete RPMI medium and equal volumes of 4 – 5 ml each were transferred to 60 mm cell culture dishes. After ~4 – 5 h post seeding SAHA-treatment was started. Cells were treated for 2, 4, 8, 16 and 20 h with either 0.1 % (final conc., control) or SAHA at final concentrations of 0.5 µM, 1.0 µM, 2.5 µM and 5 µM in DMSO (stocks: 0.5 mM, 1.0 mM, 2.5 mM and 5 mM in 100 % DMSO), respectively. After the treatment, cells were counted and the cell suspension of each condition was divided into three aliquots, one equaling 1.0x10^6 cells for cell cycle analysis (see section 3.3.4), one equaling 0.5x10^6 cells for apoptosis analysis (see section 3.3.5) and one of 1.5x10^6 cells was used to investigate protein acetylation. For the latter, cells were pelleted and lysed in 100 µl ice-cold RIPA buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 % NP-40, 0.1 % SDS, 0.5 % Deoxycholate), by incubating for 20 min on ice and occasional vortexing. After 10 min centrifugation at 4 °C and 16,000 g the supernatant was recovered and the lysate was stored at -20 °C. Protein concentration of the lysate was determined by DC protein assay (see section 3.1.1) before the protein acetylation was analyzed by SDS-PAGE and immunoblotting (see section 3.1.2 & 3.1.5).

SAHA-treatment for RNA interactome capture

Before RNA interactome capture, cells were treated with 5 µM SAHA for 6 h. A suspension of 0.5 – 0.8x10^6 proliferating cells/ml was prepared to in fresh complete RPMI medium and seeded at equal volumes into T175 tissue culture flasks. The next day, cells were treated with final concentrations of 5 µM SAHA in 0.1 % DMSO with 0.1 % DMSO (non-treated control; vehicle only), respectively. After 6 h incubation, cells were counted and harvested for UV-crosslinking followed by RNA interactome capture (see section 3.4).

3.3 Flow cytometry and cell sorting

Using flow cytometry, cells were analyzed for the expression of cell surface markers, internal molecules and DNA content. After staining with fluorochrome-conjugated antibodies or marker dyes, cells are individually illuminated by a laser beam while flowing in a fluid stream. Upon illumination, the signals emitted by the analyzed cells are detected, converted to digital signals and correlated to characteristics. Fluorescence Activated Cell Sorting (FACS), allows to sort cells possessing such characteristics into separate test tubes for further processing or analysis. As modern flow cytometers and sorters are equipped with several lasers, they possess the ability to detect different colors of fluorescence, and therefore different characteristics, simultaneously. Although FACS enables powerful multiparameter sorting, the capacity to sort large numbers of cells is limited when comparing to Magnetic Activated Cell Sorting (MACS).[197] This is attributed to the frequency of analysis and sorting. For MACS, cells are stained with specific antibodies coupled to magnetic microparticles and separated in a magnetic field. Such magnetic sorting provides only one parameter for separation but is significantly more rapid than flow sorting, as several cells are analyzed and separated simultaneously.[198] In this study, general flow cytometry analysis was performed using a LSR Fortessa flow cytometer (BD Biosciences, San Jose (CA), US). Apoptosis and cell cycle analysis was performed on an Accuri C6 flow cytometer (BD Biosciences, San Jose (CA), US). Isolation of thymocyte sub-populations was performed using a FACSAria Fusion cell sorter (BD Biosciences San Jose (CA), US) or an autoMACS Pro Separator (Miltenyi Biotec, Bergisch Gladbach, DE). Fluorochrome conjugated antibodies and general flow cytometry reagents are listed in Table 3.

Table 3: Information on antibodies and reagents used for flow cytometry and FACS.

Reagent	Fluorochrome	Dilution	Clone	Supplier; Catalogue number
Human CD7	APC	1:100	M-T701	Becton Dickinson (Franklin Lakes (NJ), US), 561604
Human CD1a	PE	1:200	HI149	Becton Dickinson (Franklin Lakes (NJ), US),561754
Human CD3	APC-Cy7	1:100	SK7; Leu-4	Becton Dickinson (Franklin Lakes (NJ), US), 561800
Human CD4	FITC	1:100	RPA-T4	Becton Dickinson (Franklin Lakes (NJ), US), 561005
Human CD8	BV605	1:200	SK1	Becton Dickinson (Franklin Lakes (NJ), US), 564115
Aqua Fixable Dead Cell Stain				Thermo Fischer Scientific (Waltham (MA), US), L34957
7-AAD Viability Staining Solution				Becton Dickinson (Franklin Lakes (NJ), US), 55925
Human BD Fc block				Becton Dickinson (Franklin Lakes (NJ), US), 564219

All antibody mixtures were prepared by diluting the respective antibodies (see Table 3) and volumes in FACS staining buffer (PBS containing 2 mM EDTA and 2 % BSA) unless otherwise stated. After centrifugation for 10 min at 13000 rpm and 4 °C, the supernatant was transferred

into a fresh tube, stored on ice and protected from light until staining. After staining and directly before acquisition, all cell suspensions were strained through a 100 µl nylon mesh to eliminate cell clumps. Compensation beads were prepared by mixing 200 µl FACS staining buffer with 1 µl of antibody and 1 drop of compensation beads (BD Biosciences, San Jose (CA), US). For each experiment, the background fluorescence was measured using both unstained cells and Fluorescence minus one (FMO) controls comprising of a cell mixture stained with all fluorochromes expect of one. As such the gating parameters between the positive and negative cell population were set. Dissociated but unsorted thymocytes (the original thymocyte suspension, not subjected to the FACS or MACS; referred to as "total thymus") were used as control when analyzing thymocyte sub-populations. Staining panels and gating strategies for thymocyte sub-populations are well-known and were already established in the laboratory due to previous work[199,200]. Data analysis was performed using FlowJo software (FlowJo LLC, Ashland (OR), US).

3.3.1 Immunophenotyping of T-ALL cell lines

The analysis of heterogeneous cell populations in order to identify the presence and proportion of certain sub-populations is referred to as immunophenotyping. The cell populations of interest are identified by antibodies against markers expressed by the cells. A specific set of such markers is known as immunophenotype. To confirm the cortical immunophenotype of P12, DND41 and MOLT3 cells proposed by Burger *et al.,* these T-ALL cell lines were analyzed for the expression of CD7, CD1a, CD3, CD4 and CD8 using analytical flow cytometry[17]. Total thymus served as a control to set gates and was stained analogously as described in section 3.3.2. All centrifugation steps were performed at 400 g for 5 min at 4 °C.

Staining of T-ALL cell lines

For each T-ALL cell line, a fully stained sample and a FMO for CD1a of $2x10^6$ cells were prepared. First, cells were washed in 1 ml FACS staining buffer and resuspended in 200 µl FACS staining buffer. After addition of 5 µl 7AAD viability staining solution, samples were incubated for 10 min at RT. Then, cells were washed again with 1 ml FACS staining buffer. Next, cell pellets were resuspended in 200 µl of the respective antibody mix, followed by 30 min incubation at 4 °C in the dark. Cells were washed once more in 1 ml FACS staining buffer and resuspended in 300 µl PBS followed by flow cytometric analysis.

Acquisition and data analysis

All gates, except of CD1a, were selected based on the well-known populations of total thymus control. Initial gating was set for lymphocytes. Then, doublets were thoroughly removed by first assessing side scatter (SSC) area vs. width and second forward scatter (FFC) height vs. area, cells with enlarged width or area were not considered. Dead cells were excluded by 7AAD staining. Gates for CD1a were selected based on the FMO controls. Further gating was performed following the importance of the cell surface markers for classification similar to Burger *et al.*[17] and by selecting subpopulations with highest frequency: Cell populations positive for the T-cell lineage marker CD7 were selected first, followed by detection of CD1a positives corresponding to the cortical phenotype. Classification into early/late cortical was achieved by analyzing CD3. By detection of the antigens CD4 and CD8 the maturation stage was further defined.

3.3.2 Analytical flow cytometry and FACS isolation of thymocyte sub-populations

For both, analytical flow cytometry and isolation of thymocytes by FACS, thymi were disaggregated and prepared as described in section 3.2.3. 3 x 10^6 thymocytes were analyzed for viability and the expression of the cell surface marker CD7 and CD1a unless otherwise stated. All centrifugation steps were performed at 300 g for 10 min at 4 °C.

Staining of thymocytes without fixation

Thymocytes were washed in 1 ml FACS staining buffer and resuspended in 200 µl FACS staining buffer. After addition of 10 µl 7AAD viability staining solution, samples were incubated for 10 min at RT, before cells were washed again with 1 ml FACS staining buffer. Thymocytes were once more resuspended in 200 µl FACS staining buffer and 8 µl human Fc block was added. After 10 min incubation at RT, thymocytes were washed with 1 ml FACS staining buffer to clean unbound Fc block. After resuspension in 200 µl antibody mix, thymocytes were incubated 20 – 30 min at 4 °C in the dark. After a final wash in 1 ml FACS staining buffer, thymocytes were resuspended in 300 µl PBS and analyzed.

Staining of thymocytes for subsequent fixation

Thymocytes were washed in 1 ml FACS staining buffer and resuspended in 200 µl FACS staining buffer. After addition of 5 µl human Fc block, thymocytes were incubated for 10 min at RT to remove excess Fc block. The pelleted cells were resuspended in 200 µl antibody mix and incubated for 20 – 30 min at 4 °C in the dark. Next, the thymocytes were washed in 1 ml PBS and resuspended in 150 µl Aqua dead cell stain solution (1:500 diluted in PBS). After 15 min incubation at RT, the thymocytes were once more washed in 1 ml PBS. After resuspension in 150 µl PBS, the thymocytes were fixed by adding 150 µl of a 4 % paraformaldehyde (PFA) solution. Cell suspensions were stored at 4 °C in the dark until acquisition for at most 24 h.

Acquisition and data analysis

After staining, samples were analyzed and the subpopulations of interest were identified. Gates were defined using FMO controls. First, bulk thymus was gated for lymphocytes. From these, only single cells were further selected by assessing SSC-area vs. SSC-width and FSC-height vs. FCS-area, cells with increased area or width were excluded. Dead cells were excluded either by 7AAD or Aqua dead cell staining. Next, only thymocytes positive for the T-cell lineage marker CD7 were selected to avoid contamination with stromal cells such as epithelial cells from the cortex and medulla or dendritic cells. Ultimately, cortical T-cell progenitors were identified based on CD1a positivity.

3.3.3 MACS separation of CD1a positive thymocytes

Before labeling, thymocytes were dissociated and prepared as described in section 3.2.3. All centrifugation steps were performed at 300 g for 10 min at 4 °C. Magnetic labeling was performed with the reagents listed in Table 4 using MACS buffer (PBS containing 2 mM EDTA and 0.5 % BSA) unless otherwise stated.

Table 4: Antibodies (with clone), magnetic beads and Fc block used for MACS separation.

Reagents	Clone	Catalogue number*
Human CD11c-Biotin	MJ4-27G12	130-114-104
Human CD303-Biotin	AC144	130-113-753
Human CD1a-PE	HI149	130-097-868
FcR Blocking Reagent		130-059-901
Anti-PE MicroBeads UltraPure		130-105-639
Anti-Biotin MicroBeads		130-090-485

*All reagents were obtained from Miltenyi Biotec (Bergisch Gladbach, DE)

Sequential MACS separation of CD1a positive T-cell progenitors

Initial isolation of CD1a positive T-cell progenitors by MACS was performed in a sequential MACS separation. In a first negative selection, CD1a positive dendritic cells (DCs) were depleted by magnetic labeling of the DC-markers CD11c and CD303. In a subsequent positive selection, the cortical T-cell progenitor target population was isolated via magnetically labelled CD1a. Labeling was performed as recommended by the manufacturer's protocol or after consultation of the technical support. For the negative selection, $10x10^6$ thymocytes were pelleted and resuspended in 500 µl MACS buffer. Then, 5 µl of FcR blocking reagent were added. After 10 min incubation at RT, 10 µl CD303-Biotin and 20 µl CD11c-Biotin antibody were added followed by 20 min incubation at 4 °C. After addition of another 5 µl FcR blocking reagent, the cell suspension was gently mixed and washed in 1 ml MACS buffer. The cell pellet was resuspended in 400 µl MACS buffer and 100 µl of anti-Biotin MicroBeads were added and incubated for 15 min at 4 °C. After washing in 10 ml MACS buffer, thymocytes were resuspended in up to 2 ml MACS buffer. Magnetic separation was performed as negative selection in sensitive mode. The eluate with the DCs, initially retained by the magnetic field, were collected as "Dendritic cells" fraction. The flow through (containing the target thymocyte sub-population) was collected for subsequent CD1a-cell surface staining and magnetic labeling. The thymocytes in the flow through were pelleted and resuspended in 500 µl MACS buffer. Next, 5 µl FcR blocking reagent were added and incubated for 10 min at RT before 20 µl CD1a antibody were added followed by20 min incubation at 4 °C in the dark. Once more 5 µl FcR blocking reagent were added and the sample was mixed gently. Thymocytes were washed with 1 ml MACS buffer and resuspended in 80 µl MACS before adding 20 µl anti-PE MicroBeads. After 15 min incubation at 4 °C in the dark, the thymocytes were washed in 15 ml MACS buffer and resuspended in up to 2 ml MACS buffer. Magnetic separation was performed in positive selection and sensitive mode. Again, flow through as well as the target cells, which were first retained by the magnetic field and then eluted ("CD1a positive" fraction) were collected. To assess the purity of the isolated cells, aliquots of both "flow-through" fractions as well as from the "Dendritic cells" and "CD1a positive" fractions, together with unsorted total thymus (original thymocyte suspension prior separation) were analyzed subsequently by flow cytometry according to section 3.3.2. For this purpose, the "flow-through", "Dendritic cells" and "CD1a positive" fractions were further stained with Aqua dead stain, CD7 and CD1a antibodies before the cells were fixed for subsequent analysis (fresh labeling of CD1a

for the "CD1a positive" fraction was not required). Aliquots of the unsorted total thymus were stained as described in "Staining of thymocytes for subsequent fixation".

Optimization and reduction of reagent consumption

Further isolations of CD1a positive cells by MACS were performed only as single separations. For optimizing the reduction of reagent consumption, thymocytes were pelleted and resuspended in 200 – 980 µl MACS buffer before adding 5 µl of FcR blocking reagent and incubating for 10 min at RT. Between 1 – 20 µl CD1a-PE antibody were added and incubated for 20 – 30 min at 4 °C in the dark before washing the thymocytes in 1 ml MACS buffer. After resuspension in 80 – 200 µl MACS buffer 1 – 5 µl anti-PE MicroBeads were added and incubated for 15 min protected from light. Cells were washed in 15 ml MACS buffer and resuspended in up to 2 ml MACS buffer. For separation, normal positive selection and positive selection in sensitive mode were compared. In addition, positive selection in standard mode with a double column was tested. For the optimized protocol FcR blocking reagent was only added once before antibody addition. Further, only 5 µl CD1a-PE antibody per $10x10^6$ thymocytes in 200 µl MACS buffer and 20 µl beads per $10x10^6$ thymocytes in 80 µl were used. MACS separation was performed in positive selection in sensitive mode. Purity of flow through and "CD1a positive" fraction was analyzed by flow cytometry as described in previous section. The number of thymocytes in each fraction was determined to investigate the efficiency of the separations.

Large scale separation of CD1a positive T-cell progenitors

First, the thymocyte suspension was prepared to a final concentration of fife times $1.5x10^8$ cells per 3 ml in MACS buffer. Then, 75 µl of FcR blocking reagent were added and the thymocyte suspension was incubated for 10 min at RT. Next, 75 µl of CD1a antibody were added followed by further incubation for 20 min at 4 °C protected from light. The cells were washed in 15 ml MACS buffer and resuspended in 1.2 ml fresh MACS buffer before adding 300 µl Anti-PE MicroBeads and incubating 15 min at 4 °C with gentle rotation (RM-Multi 1; StarLab, Brussels, BE). Cells were once more washed in 15 ml MACS buffer before resuspension in up to 2 ml MACS buffer for separation. Magnetic separation was performed as positive selection in sensitive mode. The effluents were collected and pooled as "flow-through" fraction. The eluates containing the cells initially retained by the magnetic field were collected and pooled as "CD1a positive" fraction. To evaluate the purity of the isolated target cells, aliquots of both the "flow-through" and "CD1a positive" fractions, together with the unsorted total thymus (original thymocyte

suspension prior separation) were analyzed subsequently by flow cytometry as described in section 3.3.2. The "flow-through" and "CD1a positive" fractions were further stained with Aqua dead stain and for CD7 prior to fixation and subsequent analysis (fresh labeling of CD1a was not required). Aliquots of the unsorted total thymus were stained as described in "Staining of thymocytes for subsequent fixation".

3.3.4 Cell cycle analysis

Propidium iodide (PI) binds to dsDNA and enables cell cycle analysis by measuring the DNA content of cells and allowing to distinguish different phases of the cell cycle. The fluorescence intensity of the PI stained cells correlates with the DNA content. Cells in G_0 or G_1 phase of the cell cycle contain a diploid number of chromosomes, and as such diploid DNA content (2n). Therefore, the detected fluorescence signal is lower compared to cells in G2 and just prior to mitosis (M), which contain twice the DNA amount (4n). During S-phase, DNA is synthesized and doubled, thus cells possess a DNA content between 2n and 4n. Plotting DNA content against cell number results in the classical DNA profile of a proliferating cell culture **(Figure 15)**.[x] As PI cannot cross intact plasma membranes, permeabilization of the cells, by fixation and/or treatment with detergents such as Triton X-100, is required. In addition, PI stains RNA, thus the staining procedure includes RNase treatment. All centrifugation steps prior to fixation were performed at 400 g for 5 min at 4 °C. After fixation, centrifugation speed was increased to 1000 g to minimize loss. A number of $1x10^6$ cells was collected, washed in 5 ml and thoroughly resuspended in 500 µl ice-cold PBS. While gently vortexing, cells were fixed by dropwise addition of 1.2 ml of -20 °C ethanol (100 %) and immediately placing the cells at -20 °C. The next day, the cells were pelleted, resuspended in 5 ml PBS and allowed to hydrate for 15 min at RT. The cells were pelleted once

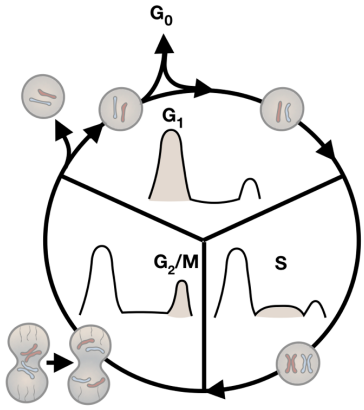


Figure 15: Cell cycle stages and corresponding DNA profile. In G_1 (gap) phase, cells prepare for replication and contain a diploid number of chromosomes (2n). In S (synthesis) phase, cells replicate and double the DNA. They possess a broad distribution of DNA content between G_1 to G_2. In G_2 (gap) phase, replication of DNA is complete, cells contain twice the amount of DNA (4n) compared to G_1. In M (mitosis) phase, cells divide leading to a DNA content of 2n.

[x] Institute of Cell and Molecular Science, Flow Cytometry Core Facility: Cell cycle analysis (http://www.icms.qmul.ac.uk/flowcytometry/uses/cellcycleanalysis/cellcycle/index.html); Requested: 2020.04.27

more and resuspended in 300 µl of PI mix (10 ml PBS, 10 µl Triton-X100, 200 µl of PI (1 mg/ml; Sigma -Aldrich St. Louis (MO), US), 200 µl of DNase- and protease-free RNase A (10 mg/ml; Thermo Fisher Scientific, Waltham (MA), US) before incubating for 30 min at RT protected from light. For each sample, at least 15.000-gated events were acquired and the cell-cycle distributions were analyzed in linear mode. Cells were identified plotting SSC-area vs. FSC-area while gating out obvious debris. Clumps and doublets were excluded by pulse shape analysis (PI-height vs. PI-area). Selected single cells were visualized as histogram plot of cell count against PI-area and the proportion of cells in G_1, S, or G_2/M phase was calculated applying the Dean Jett Fox algorithm. The number of cells in each cell cycle stage was visualized as a proportion of all cells in cycle (including sub-G_1) using GraphPad Prism 8. Data were plotted as means ± SD (Standard Deviation) and analyzed for a statistical difference using ANOVA (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ and ****$p \leq 0.0001$. $p \leq 0.05$ was considered as significant).

3.3.5 Apoptosis analysis

As described in the previous section, PI binds to dsDNA but can only enter the cell if the plasma membrane has lost its integrity, hence only in late apoptosis or necrosis. As such, it is only indicative of cell death but not specific for apoptosis. In early stages of apoptosis, the membrane phospholipid phosphatidylserine (PS) is translocated from the inner to the outer leaflet of the plasma membrane, thereby exposing PS to the external cellular environment. The Ca^{2+} dependent phospholipid-binding protein Annexin V has a high affinity for PS and binds to cells with exposed PS. Thus, fluorochromes conjugated to Annexin V can be used in flow cytometric analysis to already detect cells that are in early stages of apoptosis. The combination with PI as vital dye, allows to identify early apoptotic cells (PI neg., Annexin V pos.), late apoptosis (PI pos., Annexin V pos.), while cells with intact membranes appear as PI and Annexin V negative.[y] In this study, analysis of apoptosis was performed using the FITC Annexin V Apoptosis detection Kit I (Becton Dickinson Franklin Lakes (NJ), US) following the manufacturer's instructions. All centrifugation steps were performed at 300 g for 10 min at 4 °C. A number of $1x10^6$ cells was collected, washed in 1 ml PBS before and resuspended in 500 µl Annexin V binding buffer. Then, 100 µl of the cell suspension were transferred to a fresh tube, and 5 µl of Annexin V-FITC and

[y] BD Biosciences: Detection of Apoptosis Using the BD Annexin V FITC Assay on the BD FACSVerse™ System (https://www.bdbiosciences.com/documents/BD_FACSVerse_Apoptosis_Detection_AppNote.pdf); Requested: 2020.04.27

5 µl of PI (diluted to 10 µg/ml) were added. The cells were gently vortexed and incubated for 15 min at RT in the dark. Ultimately, 300 µl of Annexin V binding buffer were added and the samples were analyzed within 1 h. For each sample, at least 15,000-gated events were acquired Cells were identified plotting SSC-area vs. FSC-area, gating out obvious debris. Doublets were thoroughly removed by first assessing SSC-area vs. -width and FFC-height vs. -area, excluding cells with enlarged width or area. Then, selected single cells were visualized as histogram plot of cell count against FITC-area and by plotting PI-areas vs. FITC-area. For the former, the percentage of the total apoptotic fraction of cells was determined, for the later the proportion of cells in early and late apoptosis was calculated separately. The percentages were expressed as proportion to all single cells and visualized using GraphPad Prism 8. Data were plotted as means ± SD and analyzed for a statistical difference using ANOVA (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ and ****$p \leq 0.0001$. $p \leq 0.05$ was considered as significant).

3.4 RNA interactome capture methods

For the identification of RBPs, RNA interactome capture techniques developed in our laboratory were employed. Briefly, living cells are irradiated with UV light at 254 nm to cross-link naturally photo-reactive nucleotide bases, especially pyrimidines, to specific amino acids such as Phe, Trp, Tyr, Cys and Lys. RBPs which are directly bound to RNA *in vivo* are covalently coupled to their target RNA. After cell lysis, covalently linked RBPs are captured via the polyadenylated RNA using either oligo(dT) or locked nucleic acid (LNA)-coupled magnetic beads and subjected to stringent washes. By RNase treatment, the RBPs are released from the RNA and analyzed by quantitative mass spectrometry (MS). Additionally, the quality of the RNA interactome is experimentally monitored on the RNA and protein level. For RNA analysis, a Bioanalyzer is used to confirm enrichment of mRNA over rRNA. Protein isolation is monitored by gel electrophoresis, followed by silver staining and western blotting to confirm selectivity and specificity.[117,177,178] For all RNA interactome capture related experiments, DNA/RNA LoBind tubes (Eppendorf, Hamburg, DE) were used to reduce the loss of material due to oligonucleotide binding to the tube surface. All buffers were autoclaved before the addition of LiDS and DTT to inactivate e.g. proteases or nucleases etc.

General RNA interactome buffers:

- Lysis buffer:
  20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 1 mM EDTA, 5 mM DTT, 0.5% LiDS
- Buffer 1:
  20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 1 mM EDTA, 5 mM DTT, 0.1% LiDS
- Buffer 2:
  20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 1 mM EDTA, 5 mM DTT, 0.02% NP40
- Buffer 3:
  20 mM Tris-HCl (pH 7.5), 200 mM LiCl, 1 mM EDTA, 5 mM DTT, 0.02% NP40
- Elution buffer:
  20 mM Tris-HCl (pH 7.5), 1 mM EDTA
- 10x RNase buffer:
  100 mM Tris-HCl (pH 7.5), 1.5 M NaCl

3.4.1 Oligonucleotide coupling to carboxylated magnetic beads

For eRIC capture experiments, LNA-modified capture probes are used. The probe comprised of a primary amine at the 5′ end, a C6 linker region and 20 thymidine (T) nucleosides with every second being an LNA thymidine (T+) (/5AmMC6/+TT+TT+TT+T T+TT+TT+TT+TT+TT+TT)[178,201]. The probe was coupled to carboxylated M-PVA C11 magnetic beads (PerkinElmer, Waltham (MA), US) via the 5′ primary amine using EDC as described before[178,196]. For the coupling procedure, the LNA oligonucleotides (Qiagen, Hilden, DE) were resuspended in nuclease-free water to a final concentration of 100 µM and either immediately coupled or stored at -20 °C. Next, the resuspended LNAs were diluted 1:5 in an EDC solution (20 mg/ml; N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride in 50 mM MES buffer pH 6). The bead slurry was washed 3x in 5 volumes of 50 mM MES buffer pH 6 and prepared to 200 µl aliquots of original bead slurry (50 mg/ml) per tube. The beads were pelleted by magnetic force, the washing buffer was removed and resuspended in 1.2 ml of LNA-EDC-MES solution. Coupling was performed for 5 h at 50 °C at 800 rpm. To rescue condensed LNA-EDC-MES solution from the LiDS, tubes were shortly centrifuged from time to time and the beads gently resuspended by vortexing. After the coupling step, beads were washed twice in 1.5 ml PBS, resuspended in 1.2 ml of 200 mM ethanolamine pH 8.5 and incubated for 1 h at 37 °C and 800 rpm to inactivate residual carboxyl residues. Ultimately, the beads were washed 3x in 1.5 ml NaCl and stored in 200 µl of 0.1 % PBS-Tween 20 at 4 °C.

3.4.2 Recycling coupled LNA-beads after RNA interactome capture

Coupled LNA-beads were washed after each round of RNA interactome capture and re-used up to six times as described before[178,196]. As the poly(A) stretches hybridized to the LNA probe are resistant to RNase treatment, recycling of the beads requires heat elution. Further, residual RNases need to be removed of the beads by a series of washes. Therefore, beads were resuspended in 400 µl nuclease-free water and incubated for 10 min at 90 °C and 800 rpm followed by centrifugation. The beads were magnetized and the supernatant was aspirated before the bead slurry cooled down. The beads were washed 3x in in 1 ml of nuclease-free water, 3x in 1 ml of lysis buffer and stored in 200 µl of 0.1 % PBS-Tween 20 at 4 °C until use.

3.4.3 UV cross linking Cell lysate preparation

Originally, monolayers of cells are exposed to UV irradiation, directly lysed in the required volume of lysis buffer and homogenized by means of a syringe with narrow needle. Lymphocytes, however, grow in suspension with a tendency to form clumps. Hence, the protocol was adapted to suspension cells[178]. As homogenization through a narrow needle was inefficient for thymocytes and yielded cloudy viscous lysates, the use of a Precellys 24 tissue homogenizer (Bertin Technologies, Montigny-le-Bretoneux, FR) was incorporated into the protocol as outlined in Backlund *et al.*[196]. UV-crosslinking and cell lysis for final experiments of this study were performed as described below. All centrifugation steps were performed at 400 g for 5 min at 4 °C. After harvesting, cells and samples were constantly maintained at 4 °C or on ice.

T-ALL cell lines

Per condition, ~1x10^8 cells for both, UV-crosslinked samples (+UV) and non-crosslinked control (-UV) were employed. Cells were collected, pelleted and adjusted to 40 ml per +UV sample or -UV control in ice-cold PBS in 50 ml centrifuge tubes. The +UV sample was split into two non-coated petri dishes (145x20 mm^2 (639102), Greiner Bio-One, Frickenhausen, DE) placed on a pre-cooled metal plate and irradiated with UV light at 254 nm (150 mJ/cm^2) in a Stratalinker UV Crosslinker 2400 (Stratagene, La Jolla (CA), US). Cells were collected from the dishes and transferred to 50 ml centrifuge tubes. For -UV controls, UV-crosslinking was omitted. Both, +UV sample and -UV control, were pelleted and resuspended in max. 3 ml lysis buffer (supplemented with one tablet cOmplete Protease Inhibitor Cocktail (Roche, Basel, CH) per 50 ml buffer but lacking LiDS to avoid foam during homogenization). Cell lysates were transferred to pre-cooled

7 ml Precellys tubes (Bertin Technologies, Montigny-le-Bretoneux, FR) containing 500 µl of 0.5 mm zirconia/silica beads (Carl Roth, Karlsruhe, DE). Lysates were homogenized for 2x 30 sec at 5000 rpm with 20 sec pause and 4 °C. 10 % LiDS was added to a final concentration of 0.5 % and the bottom of the tubes was punctured with a heated needle (24-gauge; BD Microlance, Drogheda, IE). The Precellys tubes were transferred into fresh tubes (15 ml tube, 76 x 20 mm; Sarstedt, Nümbrecht, DE) and the homogenized lysates were collected by centrifugation. Remaining lysate was collected from the beads by washing it down twice with 3 ml complete lysis buffer ([+] protease inhibitor, 0.5 % LiDS). Aliquots of 25 – 50 µl were taken for protein measurement using the DC protein assay. All samples were snap-frozen in liquid nitrogen and stored at -80 °C until further processing.

Thymocytes

Thymocytes were dissociated as described in section 3.2.3 and either directly UV-crosslinked or after MACS separation (see section 3.3.3). UV-crosslinking and homogenization of sufficient numbers of thymocytes was carried out in successive batches. For each batch, the thymocytes were pelleted and prepared to ~100 – 200x10^6 cells in 40 ml per +UV sample or -UV control respectively. Analogously to the T-ALL cell lines, the +UV sample was split between two non-coated petri dishes (145 x 20 mm (639102), Greiner Bio-One, Frickenhausen, DE) placed on a pre-cooled metal plate and irradiated with UV light at 254 nm (150 mJ/cm^2) in a Stratalinker UV Crosslinker 2400 (Stratagene, La Jolla (CA), US). Next, the +UV sample was further splitted. Thymocytes in suspension were collected from the dishes and transferred to 50 ml centrifuge tubes. 1 ml lysis buffer (supplemented with one tablet cOmplete Protease Inhibitor Cocktail (Roche, Basel, CH) per 50 ml buffer, reduced LiDS concentration of 0.02 %) was added to each dish. Using a cell scraper (39 cm Sarstedt, Newton (NC), US), attached thymocytes were scraped from the dish into the lysis buffer and transferred to pre-cooled 7 ml Precellys tubes (Bertin Technologies, Montigny-le-Bretoneux, FR) containing 500 µl of 0.5 mm zirconia/silica beads (Carl Roth, Karlsruhe, DE). For -UV controls, UV-crosslinking was omitted. Both +UV sample and -UV control thymocytes were pelleted and resuspended in lysis buffer ([+] protease inhibitor, 0.02 % LiDS; 1 ml for +UV and 3 ml for -UV). Cell lysates were transferred to pre-cooled 7 ml Precellys tubes containing 500 µl zirconia beads (combining the +UV lysates obtained from attached thymocytes and the thymocytes from suspension). Homogenization, further processing and storage of the thymocyte lysates was performed analogously to the T-ALL cell lines.

3.4.4 RNA interactome capture (RIC)

The original RNA interactome capture method using oligo(dT) beads performed as described before by Castello *et al.*[175] with some modifications. Over the course of this study, system-specific optimizations were introduced to further develop the protocol as described in the following. Lysates were allowed to thaw at 37 °C in a water bath (SUB AQUA Pro; Grant Instruments, Cambridge, UK) until almost completely defrosted, before incubation at 60 °C for 10 min. Immediately after incubation, the samples were cooled down on ice for about 1 – 2 min and clarified by centrifugation for 10 min at maximal speed and 4 °C. As DTT is temperature sensitive, extra DTT was added to each sample to a final concentration of 5 mM. All samples were adjusted to the same protein concentration and sample volume. Aliquots of 100 µl were taken as "input" for full proteome analysis and quality controls. Oligo(dT)$_{25}$ magnetic beads (New England BioLabs, Ipswich (MA), US) were equilibrated and washed 3x in three volumes of lysis buffer and prepared to the original bead slurry using fresh lysis buffer. The samples were incubated with 300 µl of the equilibrated oligo(dT)$_{25}$ magnetic beads for 1 h at 4 °C with gentle rotation to capture RNA-bound proteins. After incubation, the beads were pelleted by magnetic force and the supernatant was transferred to fresh tubes for a second round of capture. The beads were subjected to a series of washes: once with lysis buffer, twice with buffer 1, buffer 2 and buffer 3 with each being performed for 5 min at 4 °C with gentle rotation using 8 ml ice-cold buffer. After the last wash, beads were resuspended in 900 µl buffer 3 and transferred to 1.5 ml tubes for the elution. Finally, RNA-protein complexes were eluted in 150 µl elution buffer for 5 min at 55 °C and 800 rpm in a thermomixer. Two rounds of capture were performed and the eluates of both were combined. An aliquot of 30 µl was taken for RNA quality controls. Next, 1.7 µl of 10 % SDS was added to the remaining eluates before they were concentrated in a SpeedVac (Concentrator 5301; Eppendorf, Hamburg, DE) at 30 °C to 100 µl. RBPs were released by RNase treatment, adding 11 µl of 10x RNase buffer, 0.5 µl of 1 M DTT, 1 µl of 1 % NP40 and 0.5 µl of RNase A (~200 U, R5503, Sigma-Aldrich, St. Louis (MO), US) and RNase T1 (~200 U, R1003, Sigma-Aldrich, St. Louis (MO), US) respectively to each eluate and incubation for 60 min at 37 °C. Aliquots of 25 µl (~25 %) of each eluate were taken for protein quality controls and the remaining eluates were subjected to TMT labeling and mass spectrometry. All samples and aliquots were frozen in liquid nitrogen and stored at -80 °C until further processing. For comparing RIC to eRIC, lysates and RIC eluates were adjusted to correspond to eRIC lysates and eluates.

3.4.5 Enhanced RNA interactome capture (eRIC)

The enhanced RNA interactome capture, using LNA-coated magnetic beads instead of oligo(dT) beads, was essentially performed as described in our recent publications but with minor adaptations[178,196]. As before, lysates were thawed at 37 °C in a water bath (SUB AQUA Pro; Grant Instruments, Cambridge, UK) until almost completely defrosted and directly incubated at 60 °C for 10 min to increase protein denaturation. Immediately after incubation, the samples were cooled on ice for about 1 – 2 min, while extra DTT was added to each sample (5 mM final). Lysates were clarified by centrifugation for 10 min at maximal speed and 4 °C and all samples were adapted to the same protein concentration and sample volume. Aliquots of 100 – 200 µl were taken as "input" for full proteome analysis and quality controls. The required amount of LNA-coupled magnetic beads was pooled together and equilibrated 3x in three volumes of lysis buffer before adjusting to the original volume of bead slurry with fresh lysis buffer. To capture RNA-protein complexes, 300 µl of the equilibrated LNA-beads were added to each sample and incubated for 1 h at 37 °C with gentle rotation in an incubator. After the capture, beads were pelleted by magnetic force, the supernatant was transferred to a fresh tube and stored either at 4 °C (when used on the same day) or snap-frozen in liquid nitrogen and stored at -80 °C o/n for a second round of capture. As for RIC, the beads were subjected to a series of washes using 8 ml buffer pre-heated to 37 °C: once with lysis buffer, twice with buffer 1, buffer 2 and buffer 3 with each being performed for 5 min with gentle rotation. After the second wash with buffer 3, beads were resuspended in 900 µl of fresh buffer 3 and transferred to 1.5 ml tubes for the elution. At first, a "pre-elution" was performed to eliminate unspecific binders. The beads were incubated in 170 µl of nuclease-free water for 5 min at 40 °C and 800 rpm in a thermomixer. Next, the bead suspension was divided: 20 µl of beads were transferred to a fresh tube to perform "heat elution" for RNA quality controls and the remaining 150 µl of beads to perform "RNase elution" for protein analysis. For heat elution, the beads were magnetized to remove the supernatant, then resuspended in 20 µl elution buffer and incubated for 10 min at 90 °C and 800 rpm. Immediately thereafter, the tubes were quickly centrifuged to remove condensation form the lid and the beads were pelleted to recover the supernatant ("heat eluate"). For RNase elution, the beads were first collected on the magnet to recover the supernatant ("pre-"), before resuspension in 150 µl 1x RNase buffer. 1x RNase buffer was prepared immediately before use from a 10x RNase stock and supplemented to 5 mM DTT, 0.01 % NP40 and RNase A (~200 U, R5503, Sigma -Aldrich, St. Louis (MO), US) and RNase T1 (~200 U, R1003, Sigma -Aldrich, St. Louis (MO), US) using 1.7 µl of each

RNase per 2.5 ml 1x RNase buffer. RNA was digested for 1 h at 37 °C and 800 rpm before beads were magnetized and the supernatant ("RNase eluate") was recovered to a fresh tube, which was placed again on a magnet to completely remove the beads from the RNase eluate. All eluates were maintained at 4 °C (when processed on the same day) or snap-frozen in liquid nitrogen and stored at-80 °C. The beads were washed and recycled as described in section 3.4.2 before a second round of capture was performed. Afterwards, the eluates of both capture rounds were combined and further processed. Pre-eluates were processed exactly as RIC eluates (see section 3.4.4). RNase eluates were supplemented with 1.5 µl of 10 % SDS before they were concentrated in a SpeedVac (Concentrator 5301; Eppendorf, Hamburg, DE) at 30 °C to 100 µl. Aliquots of 20 µl (~20 %) of each eluate were taken for protein quality controls and the remaining eluates were subjected to TMT labeling and mass spectrometry. All samples and aliquots were frozen in liquid nitrogen and stored at -80 °C until further processing. For comparing RIC to eRIC, lysates and eRIC eluates were adjusted to correspond to RIC lysates and eluates.

3.4.6 Sample preparation for quality controls and mass spectrometry

Upon RIC/eRIC, controls and samples were processed to monitor the capture quality and identify the RNA interactome by proteomic analysis. To assess the quality on the protein level, aliquots of the input and the eluate (RIC) or RNase-eluate (eRIC) were prepared for SDS-PAGE followed by silver staining and western blotting as described in section 3.1.2. Pre-eluates were prepared analogously. Capture quality on the RNA level was evaluated by RT-qPCR analysis and/or bioanalyzer. For the analysis by bioanalyzer, 50 – 100 µl input were processed analogously to crude cell lysate following the steps described in section 3.1.7. For RIC eluates and eRIC heat-eluates 1 µl of the combined eluate (from capture round 1 and 2) was diluted with Aqua (Aqua ad iniectabila Mini-Plasco connect; Braun, Melsungen, GE) to a final concentration of 5 or 10 µg/µl RNA before the analysis as described in detail in section 3.1.7. For RT-qPCR analysis, RIC eluates and eRIC heat-eluates were processed as described in section 3.1.8. To analyze the full proteome, an aliquot equaling 25 µg of total protein was taken from the input and diluted to a total volume of 95 µl using Aqua (Aqua ad iniectabila Mini-Plasco connect; Braun, Melsungen, GE). Upon addition of 1 µl Novagen ultra-pure Benzonase (25 U/µl; EMD Millipore Corporation, Burlington (MA), US), digestion of DNA and RNA was allowed for 40-60 min at 37 °C before the samples were snap-frozen in liquid nitrogen and stored at -80 °C.

Further processing was performed as described previously[178,196]. In brief, disulphide bridges in captured proteins were reduced using 10 mM DTT in 50 mM HEPES pH 8.5 for 30 min at 56 °C before they were alkylated using 20 mM 2-chloroacetamide in 50 mM HEPES pH 8.5 at RT in the dark for 30 min. Subsequently, samples were processed according to the SP3 protocol of Hughes and colleagues[202,203] and subjected to tryptic digest (sequencing grade; Promega, Madison (WI), US) using an enzyme:protein ratio of 1:50 at 37 °C o/n. Peptides were labelled using either TMT6plex[204] or TMT10plex[205] or Isobaric Label Reagent (Thermo Fischer Scientific (Waltham (MA), US) following the manufacturer's instructions. Further clean-up of the samples was achieved using an OASIS® HLB µElution Plate (Waters, Milfords (MA), US). Next, an offline high pH reverse phase fractionation was performed using an Agilent 1200 Infinity high-performance liquid chromatography system, equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm; Phenomenex, Torrance (CA), US), which resulted in 12 fractions for mass spectrometry data aquisition[184].

3.4.7 Mass spectrometry data acquisition

Data acquisition was performed as previously described by Perez-Perri *et al.* (2018)[178] and Backlund *et al.* (2020)[196].

Liquid chromatography

For liquid chromatography (LC), an UltiMate 3000 RSLC nano LC system (Dionex, Sunnyvale (CA), US) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5 µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å; Waters, Milfords (MA), US) was used. Trapping was performed with a constant flow of trapping solution (0.05 % TCA in water) at 30 µl/min onto the trapping column for 6 min, before peptide elution. The latter was carried out via the analytical column, running solvent A (0.1 % formic acid in water) with a constant flow of 0.3 µl/min and increasing percentage of solvent B (0.1 % formic acid in acetonitrile). The outlet of the analytical column was coupled directly to an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fischer Scientific, Waltham (MA), US) using the Nanospray Flex™ ion source in positive ion mode. The length of the LC gradient was chosen depending on the sample type: for RNA interactome samples of cell lines 90 min and of thymocytes 60 min, for full proteomes (lysate) of cell lines 120 min and for thymocytes 90 min.

MS method for TMT4-6

The peptides were introduced into the Fusion Lumos using a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective, Woburn (MA), US) with an applied spray voltage of 2.4 kV. The capillary temperature was set at 275 °C. The full mass scan was acquired with a mass range of 375 – 1500 *m/z* in profile mode in the orbitrap with a resolution of 60,000. The filling time was set to a maximum of 50 ms. Data dependent acquisition (DDA) was performed with the resolution of the orbitrap set to 15,000, with a fill time of 54 ms and a limitation of 1×10^5 ions. A normalized collision energy of 38 was applied. MS/MS data were acquired in profile mode.

MS method for TMT8-10

The MS method was carried out analogously to the one used for TMT4-6 with minor changes: using a resolution of 120,000 in the orbitrap for the full mass scan, a DDA resolution of the orbitrap setting of 30,000 and a fill time of 64 ms.

3.4.8 Mass spectrometry data analysis

Data analysis was conducted similarly to recent studies[178,196]. The acquired data were processed using IsobarQuant[206] and Mascot (v2.2.0) by searching against the Uniprot *Homo sapiens* proteome database (UP000005640), containing common contaminants and reversed sequences. The following search parameters were used: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (N-term), Oxidation (M) and TMT10 (N-term) (variable modifications). For both, the full scan and for MS/MS spectra, a mass error tolerance of 10 ppm and 0.02 Da was applied. A tryptic cleavage pattern with a maximum of two missed cleavages was selected. In addition, the minimum peptide length was set to seven amino acids. A further search requirement for protein identification was a number of at least two unique peptides. The false discovery rate (FDR) for both, peptide and protein level, was set to 0.01. In a next step, the raw output files (protein.txt) of IsobarQuant were further analyzed using R (ISBN 3-900051-07-0). To ensure a high data quality of the final eRIC experiments, only proteins quantified with at least two unique peptides in at least two replicates were further considered for the analysis. First, the limma package[207] was employed to clean the raw "signal_sum" columns for batch effects using the "removeBatchEffect" function, before a further normalization was carried out using a variance stabilization normalization method (vsn package[208]). For RNA interactome samples, separate normalization coefficients for crosslinked (+UV) and non-crosslinked (-UV) conditions

were estimated in order to maintain differences in the absolute protein amount. Missing values were imputed with the k-nearest neighborhood (knn) method using the Msnbase package[209]. Next, the data were tested for differential expression using the limma package[207]. Thus, whenever testing the RNA interactome enrichment of +UV over -UV conditions, the *p*-value and FDR output of the limma package were used. In such a comparison, a protein was classified as "hit" (also referred to as "high confidence RBP") with a fold-change of at least two (FC ≥ 2) and an FDR smaller than 5 % (FDR < 0.05). Proteins with an FC ≥ 1.5 and an FDR < 0.2 were annotated as "candidates" (also referred to as "candidate RBPs"). When comparing +UV against other +UV conditions (+SAHA versus -SAHA) conditions, the *t*-value output of the limma package was applied to the "fdrtool" function of the fdrtool package[210] to derive *p*-values and FDRs (*q*-value output). For these comparisons, "hits" (also referred to as "SAHA-responsive") were defined as those proteins with an FDR < 0.05 and an *p*-value < 0.05. Proteins with an FDR ≥ 0.05 and an *p*-value < 0.05 were considered "candidates". For comparisons between T-ALL versus thymocytes, all corresponding datasets (-SAHA only) were reanalyzed in an additional analysis, in which they were treated as replicates and +UV and -UV conditions were distinguished. For the differential expression analysis using the limma package, only limma-generated FDR-values were used for hit classification. Both, the ggplot2 R package[211] and GraphPad Prism 8 were used to generate the graphical representations.

3.4.9 Hit classification and annotations, GO and STRING analyses

After hit identification, high confidence RBPs and SAHA-responsive proteins were compared to each other, classified, annotated and further analyzed. Comparisons of identified RBPs and RNA interactomes of different cell types or conditions were conducted using Venn diagrams: either Venny 2.1 (Oliveros, J.C. (2007–2015), http://bioinfogp.cnb.csic.es/tools/venny/index.html), InteractiVenn[212] or BioVenn[213] (for quantitative Venn diagrams). The analysis of the relative disorder rank of hit proteins, as described by Sridharan *et al.*[214], was performed by Jianguo Zhu (EMBL; Savitski & Steinmetz laboratory) using the D²P²data-base[215]. Enrichment analyses of protein domains were based on Pfam and performed with DAVID 6.7 (Database for Annotation, Visualization and Integrated Discovery)[216,217], using *Homo sapiens* (all genes in database) as reference list. Up to top ten of the significantly enriched proteins domains, i.e. with lowest adjusted *p*-values (Bonferroni multiple testing corrected, adj. *p*-value < 0.05), were graphically represented using GraphPad Prism 8. The enrichment analyses of GO terms for experiments

related to technical adaptions and optimizations were performed with DAVID 6.7 as well. Similarly, *Homo sapiens* was used as reference list, however, top ten terms were based on counts in the respective datasets. For pilot experiments again up to top ten of the significantly enriched proteins domains (lowest adjusted p-values Bonferroni multiple testing corrected, adj. p-value < 0.05), were represented. For final experiments, performed in at least duplicates, GO term enrichment analyses were carried out using PANTHER[218]: Proteins identified in the respective full proteome served as background. The following parameters were selected: type of analysis: PANTHER Overrepresentation Test (Released 20200407), annotation data set: GO molecular function complete or GO biological process complete, test type: Fisher's exact test corrected for multiple testing with the Bonferroni method. Up to the top ten significantly enriched GO terms (i.e. lowest Bonferroni multiple testing corrected adj. *p*-values and < 0.05) were selected. Identical, highly similar and repetitive terms were omitted. Graphical representations were generated using GraphPad Prism 8. Protein-protein interaction networks and further functional enrichment analyses were performed using STRING[219,220]. The following parameters were used: meaning of network edges: confidence, minimum required interaction score: high confidence (0.700). Classification of hit proteins into enzymes and metabolic enzymes was based on RBPbase (v0.1.8 alpha; A Comprehensive Database of Eukaryotic RNA-binding Proteins (RBPs) with their RBP Annotations, hosted and maintained thy Hentze laboratory at EMBL); https://rbpbase.shiny.embl.de/). Annotation with known RBDs was carried out by comparing hit proteins to previous mRNA RIC/eRIC approaches[128,169,177–179,187], analogously to Backlund *et al.*[196]. Graphics were created with GraphPad Prism 8. To categorize proteins into functional classes PANTHER was used. For this type of analysis, "functional classification viewed in graphics charts" and "pie chart" were selected. Classification of proteins in into "RNA-binding", "nucleic acid binding" and "not related" was based on the functional annotation tool powered by DAVID. The generated functional annotation table was manually searched for processes, functions and terms related to RNA-binding and nucleic acid binding. Proteins not associated with any of these two categories were considered as not related. For graphical visualization GraphPad Prism 8 was used. Similarly, tables with proteins related to blood/immune/T-cells were generated. Lists of proteins were annotated with the functional annotation tool of both DAVID 6.7 and 6.8 using the following parameters: disease: OMIM diseases, functional categories: COG Ontology, SP-PIR Keywords and UP SEQ Feature, gene ontology: GO Term BP/CC/MF with the extension ALL and DIRECT, pathways: BBID, Biocharta and KEGG Pathway. Subsequently, the functional

annotation tables were searched with the following keywords: defense, immune, leukocyte, lymphocyte, T-cell, hemopoiesis and leukemia. Tables with proteins matching pathways, processes, functions and terms related to these keywords were created manually. Only selected pathways, processes, functions and terms were included into the table.

4 RESULTS & DISCUSSION

4.1 Identifying and obtaining cells recapitulating cortical T-ALL and cells corresponding to their normal T-cell progenitors

Although T-ALL is a highly heterogeneous type of cancer, and as many as 25 % of all childhood T-ALLs do not conform to any immunophenotypic classification, most T-ALLs show CD1a positivity (cortical immunophenotype)[55]. Therefore, this study focuses on the cortical subtype of T-ALL. As such cortical T-ALL cell lines and corresponding healthy thymocytes were required to allow the identification of alterations in the RNA-binding proteome upon leukemogenesis.

4.1.1 T-ALL cell lines representing cortical T-ALL

Cortical T-ALL is defined by the expression of CD1a regardless of the activity of any other T-cell markers[65]. Based on a study of Burger *et al.*[17], who classified several frequently used T-ALL cell lines as cortical immunophenotype, the three T-ALL cell lines P12, DND41 and MOLT3 were selected for this investigation. However, literature was found to be inconsistent and partially contradictory regarding their expression of relevant cell surface markers. Therefore, it was necessary to confirm the immunophenotype and especially the expression of CD1a. Using analytical flow cytometry, these three T-ALL cell lines were analyzed for the same cell surface marker panel Burger *et al.*[17] employed for their classification of T-ALL cells, namely CD7, CD1a, CD3, CD4 and CD8 (**Figure 16**).

To reliably set the gating, total thymocytes served as controls. Human thymocytes can be obtained by disaggregating human thymus tissue. The thymus tissue was obtained from otherwise healthy patients undergoing heart surgery during which the thymus was removed in order to access the heart. From previous work, the main cell populations obtained from thymic tissue were well-known[199]. As expected, all cells were highly positive for the T-cell lineage marker CD7. Importantly, the previously described cortical immunophenotype was confirmed. Albeit at different frequencies, the majority of cells of each of the cell lines expressed the marker CD1a, accounting for a cortical stage of T-cell development. Further, all cell lines, including the cell line DND41 reported previously as CD3 positive[17], were found to be mostly CD3 negative and were considered as early cortical. In addition, DND41 was double negative for CD4 and CD8 and P12 showed a single positive phenotype expressing CD4 only. CD4/CD8 double positivity was found

for MOLT3, although this cell line was reported to be CD4 positive only[17]. Based on these results, all three cell lines were considered suitable representative cell types for the cortical phenotype of T-ALL.

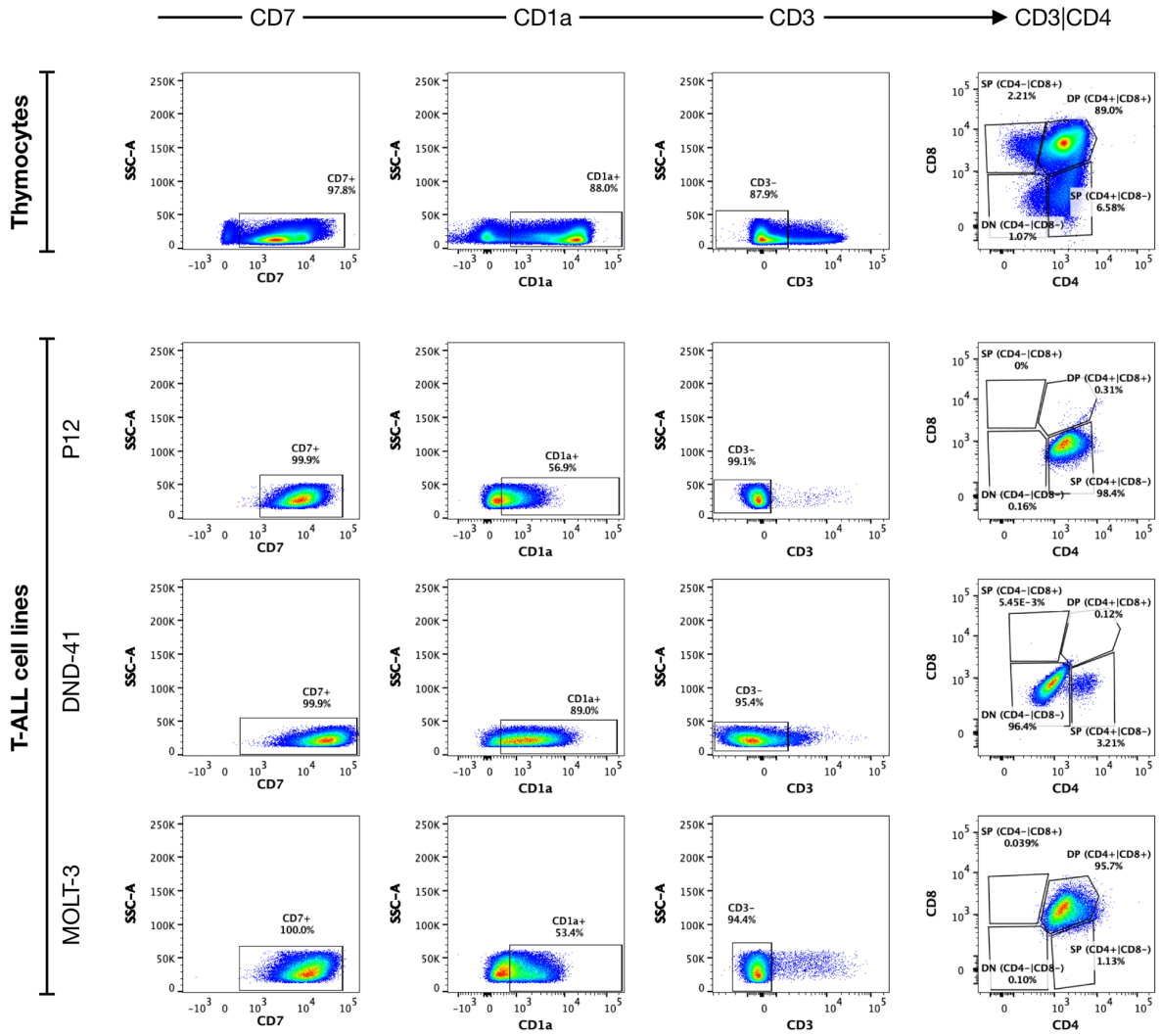


Figure 16: Confirmation of the immunophenotype of selected T-ALL cell lines. Cell lines and freshly disaggregated thymocytes were stained analogously. All gates, except CD1a, were selected based on the well-known populations of total thymus. Gates for CD1a were selected based on FMO controls. Gating was performed following the importance of the cell surface markers for classification similar to Burger *et al.*[17] and by selecting subpopulations with highest frequency: Cell populations positive for the T-cell lineage marker CD7 were selected first, followed by detection of CD1a positives corresponding to the cortical phenotype. Classification into early/late cortical was achieved by analyzing CD3. By detection of the antigens CD4 and CD8 the maturation stage was further defined. Depicted frequencies refer to the respective parent population.

4.1.2 Thymocyte sub-populations resembling normal cortical T-cell progenitors

Close links between T-cell ontogeny and leukemogenesis have been reported[15,54]. To further enhance the understanding of T-ALL leukemogenesis, this project aimed at identifying differences in the RNA interactome of T-ALL cells and normal thymocytes. Thus, it was important to study the T-cell progenitors that are most similar to the T-ALL cell lines.

Assessment of thymocyte sub-population abundance

To obtain cells as similar as possible to the T-ALL cell lines, total thymus was analyzed for the abundance of thymocyte subpopulations with the same subset of cell surface markers as confirmed for the selected cortical T-ALL cell lines. Whereas the total percentage of the thymocyte sub-populations with a set of cell surface markers corresponding to MOLT3 was relatively high (89.9 %), the total frequencies of P12- and DND41-similar thymocytes were surprisingly low (6.58 % and 1.07 %). Taken together with the relatively high number of cells required for RNA interactome capture, sorting for these closely related subpopulations was found to be unfeasible. In contrast, isolation of a more broadly defined subpopulation based solely on the expression of CD7 (cell surface marker defining T-cells) and CD1a (cell surface marker defining cortical T-ALL) was considered a feasible alternative. Therefore, the viable CD7 and CD1a positive subpopulation was isolated from total thymus by FACS and re-analyzed to determine its abundance (**Figure 17**). The results reveal a purity of ~70 % of viable CD7 and CD1a expressing T-cell progenitors. However, the costs to obtain sufficient amounts of this CD1a positive subpopulation by FACS were still estimated to be too high and the required sorting times too long. However, Magnetic Activated Cell Sorting (MACS) was considered as a feasible alternative to FACS.

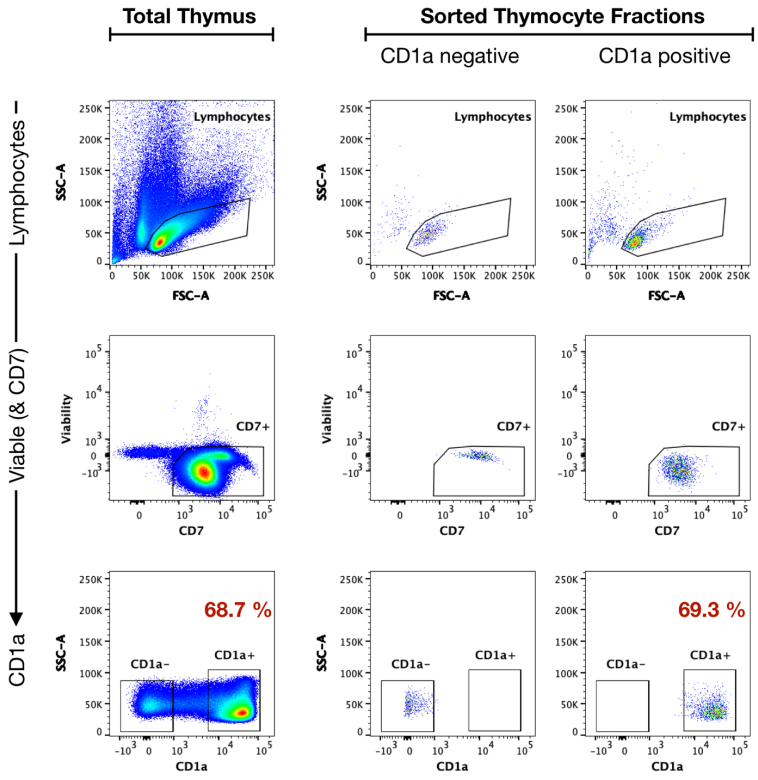


Figure 17: FACS sorting for CD1a positive T-cell progenitors from human thymus. Cell surface markers of freshly disaggregated thymocytes were stained for CD7 (APC) and CD1a (PE). Cell viability was detected via 7AAD staining. Total thymocytes were analyzed by gating for lymphocytes, selecting single cells (not shown) and sorting for viable CD7 positives. Form these, CD1a positives and negatives were sorted. Both, CD1a positive and negative fraction, were collected and re-analyzed. The total frequencies of CD1a positive subpopulations are depicted in red.

Protocol set-up for the isolation of cortical T-cell progenitors by MACS

Compared to FACS, MACS offers superior capacity. After staining with specific antibodies coupled to magnetic microparticles, cells are separated simultaneously in a magnetic field. As a result, MACS enables faster processing.[221] In contrast to the powerful multi-parameter sorting of FACS, however, magnetic sorting provides only one parameter for each separation. As the cell surface marker CD1a is also expressed on dendritic cells (DCs), a sequential MACS separation protocol was established[198,222]. In an initial negative selection DCs were depleted using magnetic labels for the DC-markers CD11c and CD303, and in a subsequent positive selection the cortical T-cell population was isolated using a magnetically labeled anti-CD1a antibody. Purity and viability of the obtained fractions were assessed by flow cytometry (**Figure 18**).

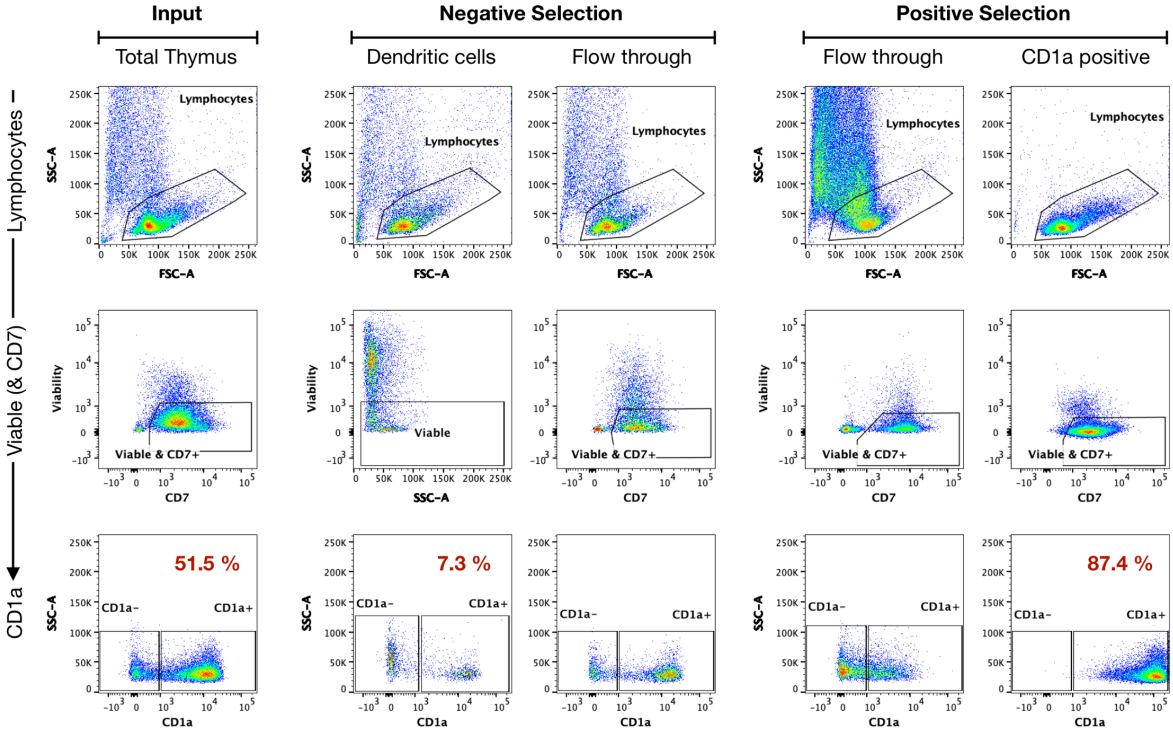


Figure 18: Flow cytometry analysis of CD1a positive T-cell progenitor isolation by sequential MACS separation. Cell surface markers of dendritic cells in freshly disaggregated total thymus were stained using CD303 (biotin) and CD11c (biotin) antibodies and depleted by negative selection. Subsequently, cortical T-cell progenitors were stained by CD1a (PE) and isolated by positive selection using microbeads ultrapure (anti-PE). Total thymus (left column), DC fraction and corresponding flow through fraction containing target cells from negative selection (middle column) as well as flow through and the CD1a positive target fraction from positive selection (right column) were further stained with CD7 (APC) and AQUA viability dye, before fixing with paraformaldehyde and analysis by flow cytometry. CD1a PE-labels were directly detected by flow cytometry. All gates were selected based on the well-known population of total thymus (control). Gates for CD1a determined before based FMO controls. Samples were analyzed by gating for lymphocytes (top panel) and selection of single cells (not shown). Subsequently, CD1a positives and negatives (bottom panel) were sorted from viable cells (DC fraction) or from viable CD7 positives (middle panel). Total frequencies of the CD1a positive subpopulations are depicted in red.

A proportion of 51.5 % of the total thymocytes accounted for the target population (i.e. viable and CD1a positive T-cell progenitors). Sequential MACS separation yielded high purity and enriched these target cells to 87.4 %, demonstrating the successful isolation of cortical thymocytes. Analysis of the DC fraction from negative selection revealed that amongst the enriched DCs 7.3 % show the same CD1a positive immuno-phenotype as the target cells. Due to this negligible amount of CD1a positive DCs, and with regard to the additional time and cost consuming processing steps required for sequential MACS separation, the negative selection step was deleted from the protocol. Assessment of thymocyte isolation by single MACS separation yielded similarly high purity and enrichment of viable CD1a positive T-cell progenitors (data not shown).

In a next step different positive selection programs for MACS separation were tested. Positive selection in "sensitive mode" with a single-column or in "standard mode" using a double-column are recommended programs for isolations with recovery as highest priority. In small-scale test experiments both programs were compared and yielded similar amounts of isolated cells with high purity and similar enrichment of the target cells (data not shown). Therefore, further MACS separations were performed by positive selection in "sensitive mode" with a single-column.

Optimization and reduction of reagent consumption

Although MACS separation proved superior to FACS, the estimated costs of reagent consumption for large scale isolation of thymocytes remained high. Strong signal of the CD1a label in the flow cytometry analysis, indicated that the labeling procedure is amenable to reagent optimization without sacrificing yield and quality of the thymocyte isolation. Therefore, different amounts and ratios of staining antibodies and magnetic microbeads were tested aiming at a reduction where possible.

Decreasing the amount of microbeads did not affect the purity of the target fraction. The number of isolated target cells, however, was substantially less. In contrast, a reduction of CD1a antibody to one quarter of the volume recommended in the user's manual was possible without compromising yield and purity. Further, a series of centrifugation steps followed by repeated decanting of the easily soluble part of the formed pellet while discarding the viscous part of the disaggregated thymus was found key to drastically reduce the amount of cell debris observed by flow cytometry (data not shown). Ultimately both protocols, the original staining procedure with reagent volumes similar to the recommendations of the user's manual as well as the optimized protocol using only one quarter of CD1a antibody and adapted reagent ratios, were compared.

Experiments were carried out in parallel using thymocytes of the same individual (**Figure 19**). The low amount of cell debris (~10 %) in the total thymus input reflects the additional centrifugation steps mentioned above. A high percentage of 83.6 % of viable T-cell progenitors positive for the cell surface marker CD1a was observed in the input. For both staining procedures, original and optimized, similar purities of less than 1 % CD1a negative cells and 86.8 % and 87.5 %, respectively, of CD1a positive cells were obtained. Both protocols yielded virtually identical cell numbers equaling ~50 % of the input cells. Thus, these findings validate the optimization and reduction of reagent consumption in the MACS staining protocol.

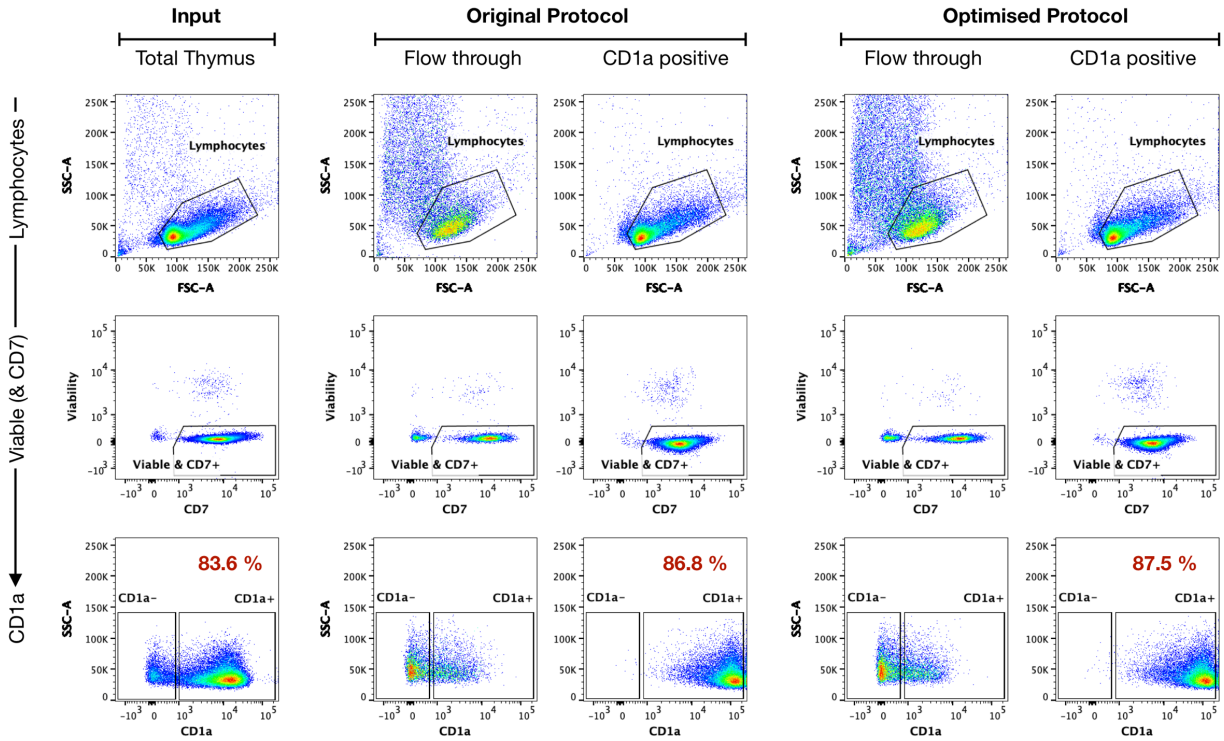


Figure 19: Monitoring the optimization of reagent consumption of thymocyte isolation by MACS using flow cytometry. Freshly disaggregated thymocytes were stained by CD1a (PE) and isolated by positive selection using microbeads ultrapure (anti-PE). Total thymus (left column), fractions of thymocyte isolation using the original staining protocol (middle column) as well as fractions obtained with the optimized staining protocol (right column) were further stained with CD7 (APC) and AQUA viability dye, before fixing with paraformaldehyde and analysis by flow cytometry. CD1a PE-labels were directly detected by flow cytometry. All gates were selected based on the well-known population of total thymus (control). Gates for CD1a determined before using based in Fluorescence Minus One (FMO) controls. All samples were analyzed by gating for lymphocytes (top panel) and selection of single cells (not shown). CD1a positives and negatives (bottom panel) were sorted from viable CD7 positives (middle panel). Total frequencies are depicted in red.

4.1.3 UV-crosslinking of thymocytes

In general, a protein yield of ~20 mg is required to obtain optimal results by RNA interactome capture (personal communication, Hentze laboratory). Since thymocytes are smaller than most other cell types, it was expected that higher cell numbers than usual have to be crosslinked. Therefore, UV-crosslinking with primary thymocytes was assessed using total thymocytes before performing large scale MACS isolation and UV-crosslinking. First UV-crosslinking trials revealed that thymocytes strongly adhere to the cell culture dishes after being exposed to UV light and cannot be washed down or drained. Different types of dishes (hydrophobic coating TC dish, uncoated petri dish, glass dish) were tested. Although best results were achieved using uncoated petri dishes, substantial amounts of cells remained attached. Using a cell scraper for dissociation resulted in damaged cells and in low protein amounts. Collection of the thymocytes directly in lysis buffer and using a syringe for cell lysis did not completely sheer precipitated DNA and resulted in viscous cell lysates. Ultimately, the UV-crosslinking protocol was subdivided into first collecting the cells still in suspension, followed by centrifugation and resuspension in a small volume, and second scraping attached cells directly into a low volume of lysis buffer. After combining both fractions, cells were lysed using the Precellys homogenizer. Using this approach, clear non-viscous lysates were obtained. Although high numbers of total thymocytes were used, these tests yielded very little protein. Therefore, even higher cell numbers (~2×10^8 – 3×10^8) of cortical thymocytes were required for RNA interactome capture than originally anticipated.

4.1.4 Large scale testing of cortical thymocyte isolation by MACS followed by subsequent UV-crosslinking

After reduction of the reagent consumption and optimization of the UV-crosslinking protocol, a first large-scale test of the cortical T-cell progenitor isolation by MACS and subsequent UV-crosslinking was performed. To achieve gentle thymus disaggregation, and obtain sufficient numbers of thymocytes within a reasonable time, two operators were required. In three independent experiments ~6×10^8 – 7.5×10^8 thymocytes (input) each were subjected to MACS separation. MACS was performed in four to five batches each to avoid exceeding the column capacity. Nevertheless, for two out of three tests an unexpectedly low total yield of ~2×10^8 – 3×10^8 isolated thymocytes was obtained. Since for subsequent UV-crosslinking, the cell suspension had to be split into UV-crosslinked and non-irradiated control, input material per condition was further decreased. To monitor the MACS separation and verify the purity of the cells used for

UV-crosslinking, aliquots of the respective total thymus and the CD1a positive fraction were analyzed by flow cytometry (**Figure 20**). High percentages of ~85 %, 70 % and 81 % target cells (CD1a positives) were observed in the three total thymus specimens (input). By MACS separation the purity was further increased to ~90 % target cells (output). Consequently, MACS separated fractions were considered to be sufficiently pure and representing the normal cortical T-cell equivalent to the cortical T-ALL cell lines. Thus, all three CD1a positive fractions were pooled and the amount of total protein of was determined. For UV-crosslinked and non-crosslinked samples amounts between 6 – 8 mg of total protein were obtained. However, RNA interactome capture requires a total of ~20 mg of protein per condition. Thus, the amount of protein obtained from ~1.8 x 10^9 – 2.25 x 10^9 total thymocytes input was still too low to obtain sufficient CD1a positive thymocytes for one RNA interactome capture. Two of the total thymocyte fractions analyzed showed a high percentage of viable, CD7 and CD1a positive thymocytes already before

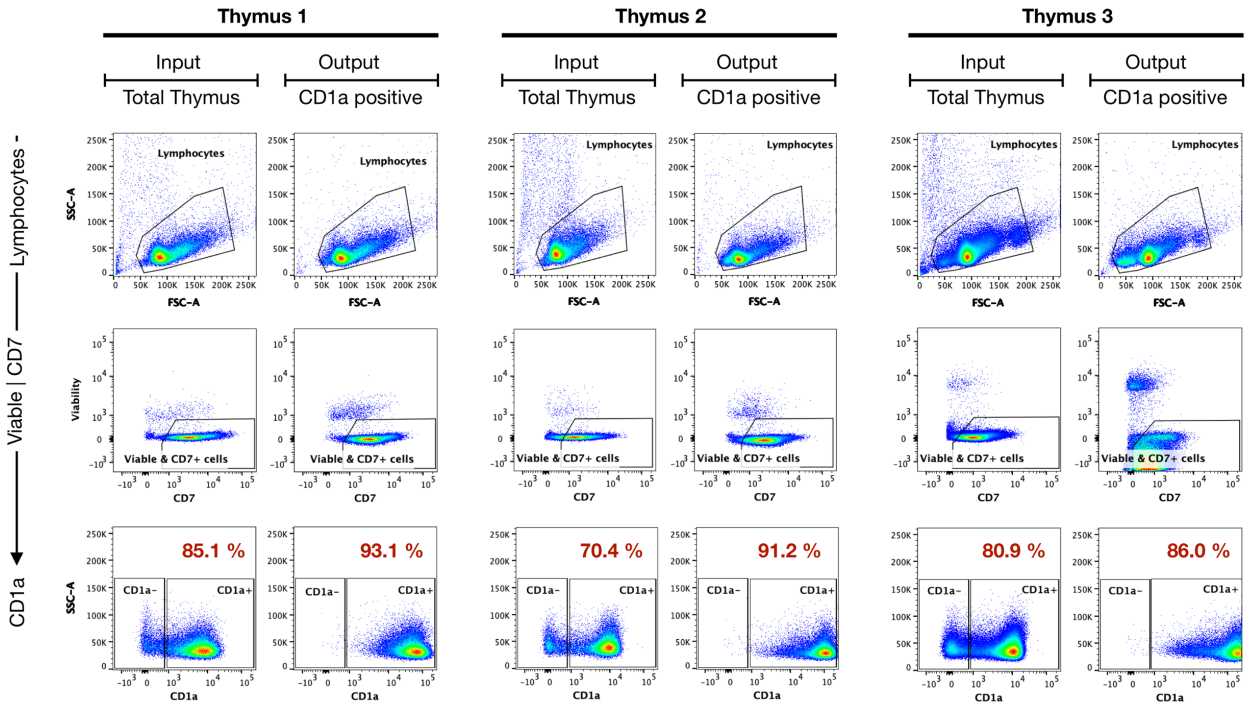


Figure 20: Quality assessment of large-scale thymocyte isolations by MACS using flow cytometry. Freshly disaggregated thymocytes were stained by CD1a (PE) and isolated by positive selection using microbeads ultrapure (anti-PE). Total thymus control, flow through as well as CD1a positive fractions were further stained with CD7 (APC) and AQUA viability dye, before fixing with paraformaldehyde and analysis by flow cytometry. CD1a PE-labels were directly detected by flow cytometry. All gates were selected based on the well-known population of total thymus (input). Gates for CD1a determined before using based in Fluorescence Minus One (FMO) controls. All samples were analyzed by gating for lymphocytes and selection of single cells (not shown). CD1a positives and negatives were sorted from viable CD7 positives. Total frequencies are depicted in red.

MACS separation. MACS isolation further increased the percentage of this subpopulation by 5 – 8 % only. Hence, the expenditure on MACS separation was not proportionate to the increase of purity – especially considering that the prolonged handling time likely induces stress and corresponding secondary effects on the thymocytes which might compromise on the "cell quality" and ultimately lead to the identification of false positives/negatives in the RNA-binding proteome.

In order to decide if using the total thymocyte fraction directly after disaggregation is an alternative, the differences between the total thymocyte lysate and the CDa1 positive thymocyte lysate were evaluated by MS (**Figure 21**). Comparison of the median signal sum of both fractions before and after normalization shows, that the signal sum is similarly distributed and that the effect of normalization is very modest (**Figure 21a**). Proteins identified in each fraction were next compared to each other to evaluate how different the protein compositions before and after enrichment of CD1a positives are (**Figure 21b**). Overall, a high correlation of protein abundance between the CD1a positive enriched and the total thymus sample was observed. However, a small number of proteins appear differentially enriched or only present in the total thymus fraction. To identify these differentially abundant proteins, the data were analyzed in more detail (**Figure 21c**). In line with previous correlations, the majority of proteins show a smaller than two-fold difference in abundance with and without MACS separation. 79 proteins were identified as "hits", i.e. as proteins with significantly differential abundance. Next, these hits were analyzed by STRING network and Gene Ontology (GO) domain analyses searching for RNA-binders and representative features (**Figure 21d+e**). Results revealed that the number of proteins associated with RNA biology-related functions and processes is low. The good overall correlation and the low number of RNA-binding related proteins with significantly differential abundance demonstrates that an increase of the CD1a positive subpopulation of ~5 – 8 % by MACS separation is not a worthwhile expenditure. Based on these findings, the MACS separation was omitted and further experiments were performed using total thymocytes only.

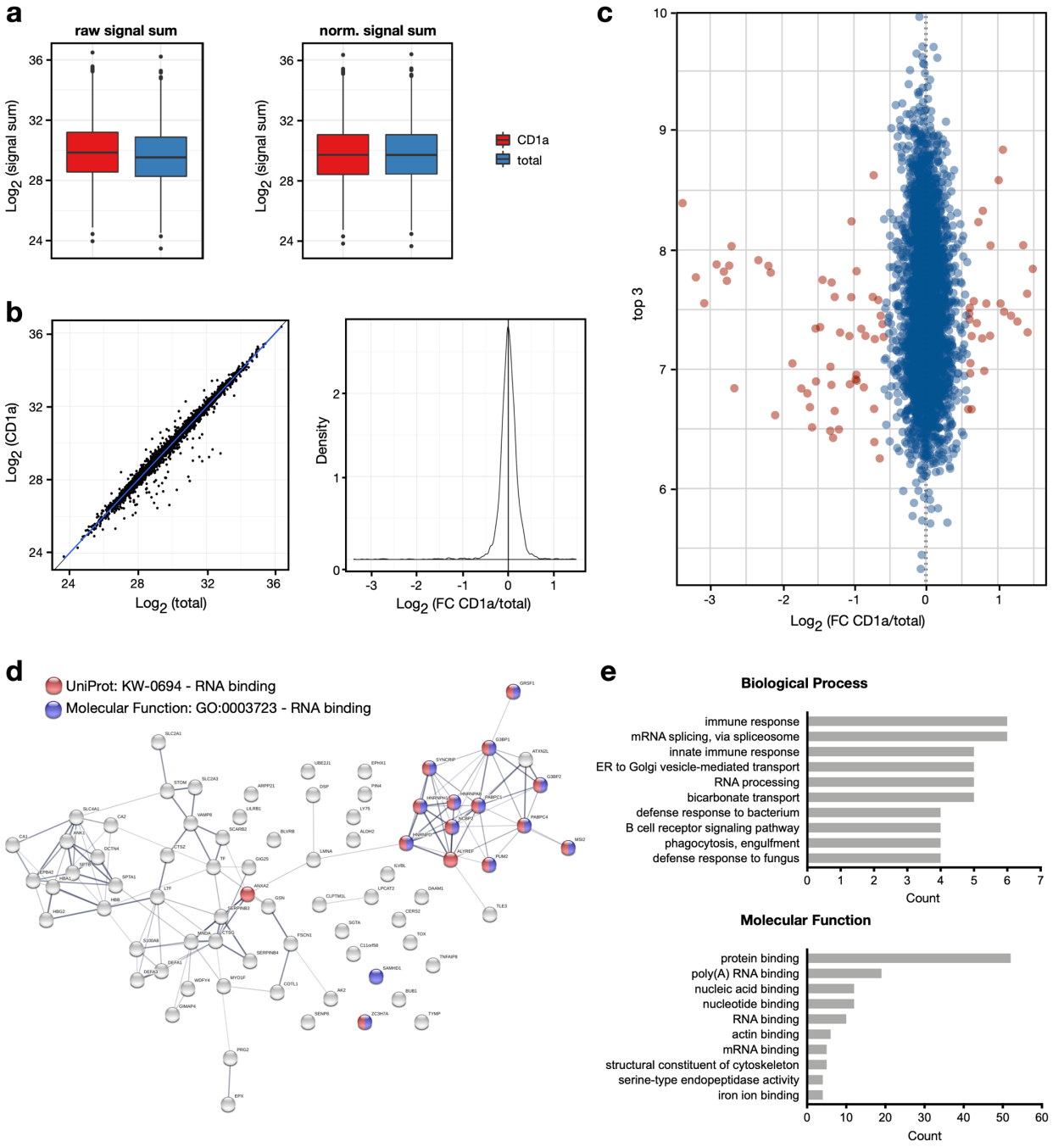


Figure 21: Analysis of the thymocyte proteome before and after enrichment of CD1a positives by MACS. Thymus was disaggregated and cleaned by subsequent centrifugation steps and discarding the viscous part of the cell pellets (total thymus). After staining, CD1a positive cells were isolated by MACS separation (CD1a). Differences in the proteome of both fractions were analyzed by MS. (a) Boxplot of the signal sum of proteins identified in each fraction. (b) Scatter plot showing the signal sum distribution and density blot, both visualizing the correlation of identified proteins. (c) Scatter plot depicting the fold change correlation of proteins identified from the total thymocyte fraction and the isolated CD1a positive fraction. Proteins with a fold-change (FC) of at least 50 % and a false discovery rate (FDR) smaller than 5 % were considered as "hits" (depicted in red), proteins not considered as hits are shown in blue. (d) STRING network analysis of proteins with highly different abundance (hits) between both fractions. Red dots represent proteins associated with RNA binding based on UniProt and blue dots represent proteins associated annotated with the GO function RNA binding. (e) GO domain analysis of the differentially abundant hit proteins using DAVID Functional Annotation Tools[216,217].

4.2 System-specific adaption of the RNA interactome protocol

The RNA interactome methodology was developed using HeLa cells and had so far only been applied to further adherent cells, such as HuH7, HEK293 and RAW264.7. Monolayers of adherent cells can be directly exposed to UV light for cross-linking[175,177]. In contrast, lymphocytes grow in suspension, with a tendency to form clumps and are in general of smaller size. For instance, HeLa cells attached to a surface show the shape of a half-sphere with an average diameter of 20 µm and were reported to have a volume of up to 2600 µm³. Leukocytes have a spherical shape, diameters of around 10 – 13 µm (JURKAT cells), and volumes of around 125 µm³ [223–225]. In addition, lymphocytes are not only smaller, they also have an unfavorable nuclear-cytoplasmic ratio (4:1 – 3:1), and are thus likely to have a reduced protein content as well as a relatively high DNA content compared to larger cells such as HeLa[226]. The RNA interactome capture protocol had so far not been experimentally established for such substantially smaller cells growing in suspension. Therefore, it was necessary to assess the technical steps of the RBP capture protocol using representative cells and to undertake system specific adaptions where needed.

4.2.1 Initial RNA interactome capture assessment using T-cells

To test the feasibility of the RNA interactome capture method with T-cells, initial trial experiments were carried out using the readily available, and in the laboratory well-known, T-ALL cell line JURKAT. A total number of around ~0.5 - 1 x 10^8 cells per condition (UV-crosslinked (+UV) and non-crosslinked (-UV) control) was estimated as appropriate, based on differences in volume and protein content of HeLa and T-ALL cells. For UV-crosslinking of RBPs to RNA, the cells were pelleted and resuspend in a volume of PBS resulting in a monolayer. For cell lysis and homogenization, a Precellys homogenizer step was included into protocol since the normally performed syringe lysis did not efficiently homogenize thymocytes lysates in previous test experiments (see section 4.1.3). RBPs covalently bound to polyadenylated RNAs were captured on oligo(dT) magnetic beads and subjected to stringent washes to reduce contaminations. The eluted samples were treated with RNases or proteinase K to purify the RBPs or RNAs, respectively, for further analysis. Protein and RNA quality were monitored by silver staining, immunoblotting and the use of a Bioanalyzer chip (**Figure 22**).

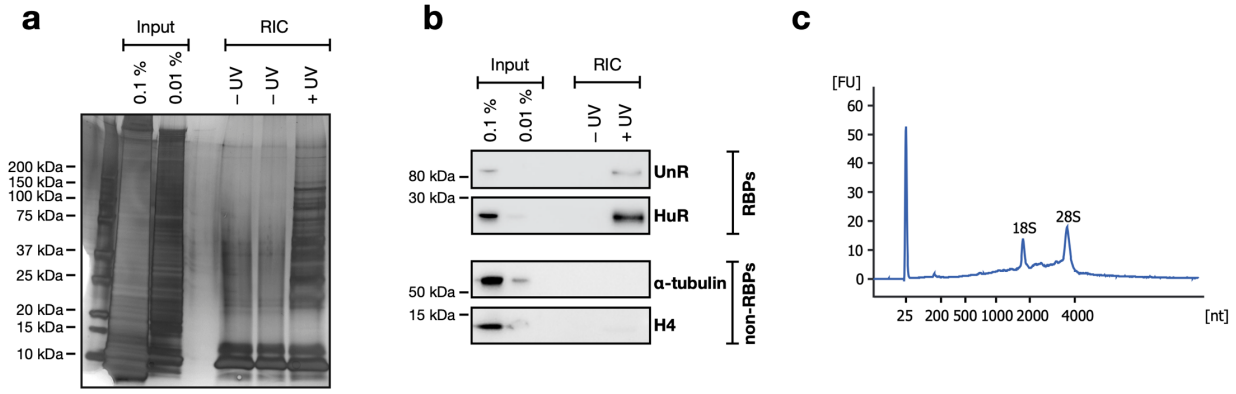


Figure 22: Monitoring the adapted mRNA interactome capture protocol on protein and RNA level. RBPs are covalently bound to polyadenylated RNAs in living cells by UV-crosslinking (+UV). Non-irradiated samples serve as negative controls (-UV). Following cell lysis, the RBPs are captured on oligo(dT) magnetic beads out of the whole cell lysate (input). After stringent washes, proteins are eluted (RIC). (a-b) Samples were digested with RNases, and released proteins were analyzed by SDS-PAGE combined with silver staining (a) and by western blotting against UnR, α-tubulin, HuR, and histone H4 (b). (c) Proteins were digested with proteinase K and released RNAs were analyzed by bioanalyzer.

As can be seen in **Figure 22a**, silver staining reveals a pattern of isolated proteins (RIC) that differs notably from the pattern of the whole cell proteome (input). Further, complex protein patterns are obtained from the UV-crosslinked samples (+UV), whereas control reactions from non-irradiated cells (-UV) are remarkably clean. These results indicate a successful enrichment of RBPs. The western blot analysis depicted in **Figure 22b**, reveals that UnR and HuR, two well-known RBPs, are both enriched in UV-crosslinked eluates and undetectable in the non-irradiated samples. The abundant cellular proteins α-tubulin and histone H4 are depleted from the eluates of both, UV-crosslinked and non-irradiated samples. These results confirm selectivity against protein abundance of the RNA capture protocol and demonstrate specificity for RNA binding. By using a Bioanalyzer Chip, an enrichment of nucleic acids corresponding to the same size range as mRNA and a partial depletion of rRNA (which is the most abundant RNA species in cells (~80 – 90 % of total RNA) was confirmed (**Figure 22c**)[227]. However, compared to previous RNA interactome captures performed in the laboratory, the observed mRNA enrichment is considerably lower. Thus, although aiming on RBPs bound to polyadenylated mRNAs, the results indicate that both, rRNA together with mRNA, are present in the eluted samples.

Following successful quality control analyses, the JURKAT samples were subjected to a proteomic test analysis. The released proteins were cleaved into peptides using trypsin and identified on a liquid chromatography-tandem MS (LC-MS/MS) platform. In total, 801 proteins were identified in the UV-crosslinked sample. Of these, 446 proteins were significantly enriched over the non-irradiated control ("hits"; false discovery rate (FDR) of less than 0.05 and fold change (FC) of at least 2) and identified as JURKAT RBPs. This JURKAT RNA interactome was

benchmarked with a Gene Ontology (GO) domain analysis using the DAVID (Database for Annotation, Visualization and Integrated Discovery) Functional Annotation Tools to search for representative features (**Figure 23**)[216,217]. As expected, in the GO domain *Molecular Function* functions involving interaction with RNA are highly over-represented. Consistently, in the GO domain *Biological Process* "RNA processing" is the most over-represented category followed by further RNA biology-related functions and processes such as "mRNA metabolic process", "translation", and "RNA splicing". In contrast, rRNA and DNA-related functions and processes (e.g. "ribonucleotide binding", "structure-specific DNA binding") are still present but less enriched. Noteworthy, these results are in line with the assumption, that in addition to mRNA, rRNA was present in the eluted samples. However, overall proteins implicated in RNA-binding far exceed proteins related to rRNA and DNA-binding.

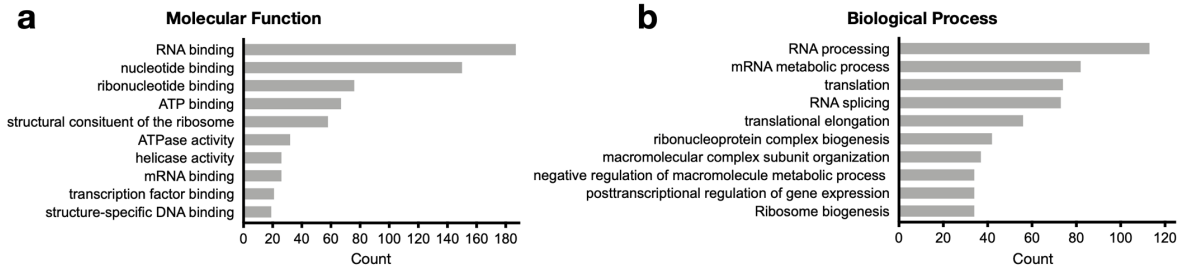


Figure 23: GO domain analysis of the JURKAT RNA interactome using DAVID Functional Annotation Tools[216,217]. (a) GO domain analysis of Molecular Function, and (b) Biological Process. RNA biology-related terms, functions and processes are highly over-represented. DNA-binding and rRNA-binding or related functions and processes are present but under-represented.

In addition, this JURKAT RNA interactome was compared to previously published RNA interactomes of HEK293, HuH7 and HeLa cells (**Figure 24**)[169,177,179]. As anticipated, the 446 RBPs identified in the JURKAT RNA interactome show a considerable overlap with RNA interactomes of all three other cell lines. These overlapping RBPs are likely to represent "housekeeping" RBPs. Further, a number of RBPs is unique to each cell line and 76 RBPs were identified only in JURKAT cells and may be JURKAT specific. Overall, the large number of identified RBPs that match pub-

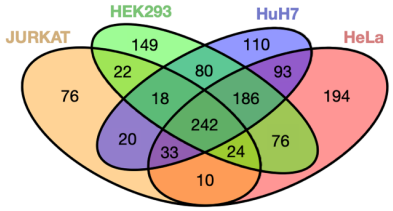


Figure 24: Comparison of the JURKAT RNA interactome with the published RNA interactomes. Venn diagram showing the number of common and different proteins between the RNA interactome of JURKAT cells and the previously published RNA interactomes of HEK293, HuH7 and HeLa[169,177,179]. RBPs overlapping between all cell lines are thought to represent the "housekeeping" RNA interactome. 76 RBPs of a total of 446 RBPS identified are specific to the JURKAT RNA interactome.

lished data ultimately validate the specificity and selection of RBPs by RNA interactome capture and the undertaken alterations in the protocol. Taken together, the results of these proof-of-principle experiments demonstrated that RNA interactome capture studies facilitate the identification of RBPs involved in T-ALL despite of the fact that the employed T-ALL cells are small, growing in suspension and relatively high in their DNA content compared to other cell types. Nevertheless, these results also indicate the necessity for an optimization of the RNA interactome capture protocol to increase the mRNA enrichment, reduce rRNA and genomic DNA contaminations, and thus increase the specificity for RBPs.

4.2.2 RNA interactome capture employing established cortical T-ALL cell lines

Burger *et al.*[17] characterized and assigned several established T-ALL cell lines to the cortical subtype. Using analytical flow cytometry, this cortical immuno-phenotype was confirmed. Based on these findings, the cell lines P12, DND41, and MOLT3 were selected as representative for cortical T-ALL (see section 4.1.1).

RNA interactome capture using cortical T-ALL cell lines

To undertake further system-specific adaptions directly with the cells that will be ultimately used for the experiments performed here, the RNA interactome capture protocol was tested next with the three selected cortical T-ALL cell lines. As before, the quality of these RNA interactome captures was monitored on both, protein and RNA level (**Figure 25**). Silver staining of all three cell lines reveals patterns of eluted RBPs (eluate) that differ notably from the patterns of input samples and are absent in the non-irradiated controls (**Figure 25a**). For immunoblotting, the two well-known RBPs UnR and HuR, served as controls for successful RNA interactome isolation (**Figure 25b**). Both proteins are present in UV-crosslinked eluates and not detected in the non-irradiated samples. Conversely, the abundant cellular proteins α-tubulin and histone H4 are depleted from the eluates of both, UV-crosslinked and non-irradiated samples. These findings demonstrate selectivity against protein abundance and confirm specificity for RNA binding. The RNA quality analysis using a Bioanalyzer Chip (**Figure 25c**) shows enrichment of nucleic acids corresponding to the same size range as mRNA, albeit relatively low, and a partial depletion of rRNA. However, an unexpectedly high additional peak at 2000 nucleotides imbricates the rRNA peaks, indicating genomic DNA contamination. These findings are likely to reflect the expected high relative DNA content of the immature early cortical T-ALL cell lines and indicate that mRNA together with increased DNA and rRNA contaminations are present in these eluates.

Overall, these results clearly demonstrate that the RNA interactome capture protocol in principal works with the selected cortical T-ALL cell lines. However, low mRNA enrichment combined with rRNA and DNA contaminations remained an issue. Indeed, the degree of DNA contamination appears to be increased compared to the RNA interactome captures performed with JURKAT cells. This may reflect the fact that early cortical T-ALL cell lines are more immature compared to mature T-ALL cell line JURKAT. Thus, it is likely that these cells contain an even higher relative DNA content leading to the low mRNA enrichment while at the same time causing increased unspecific binding. Further, the amounts of input material may have been too low aggravating the issue of low mRNA enrichment and thus favoring increased unspecific binding.

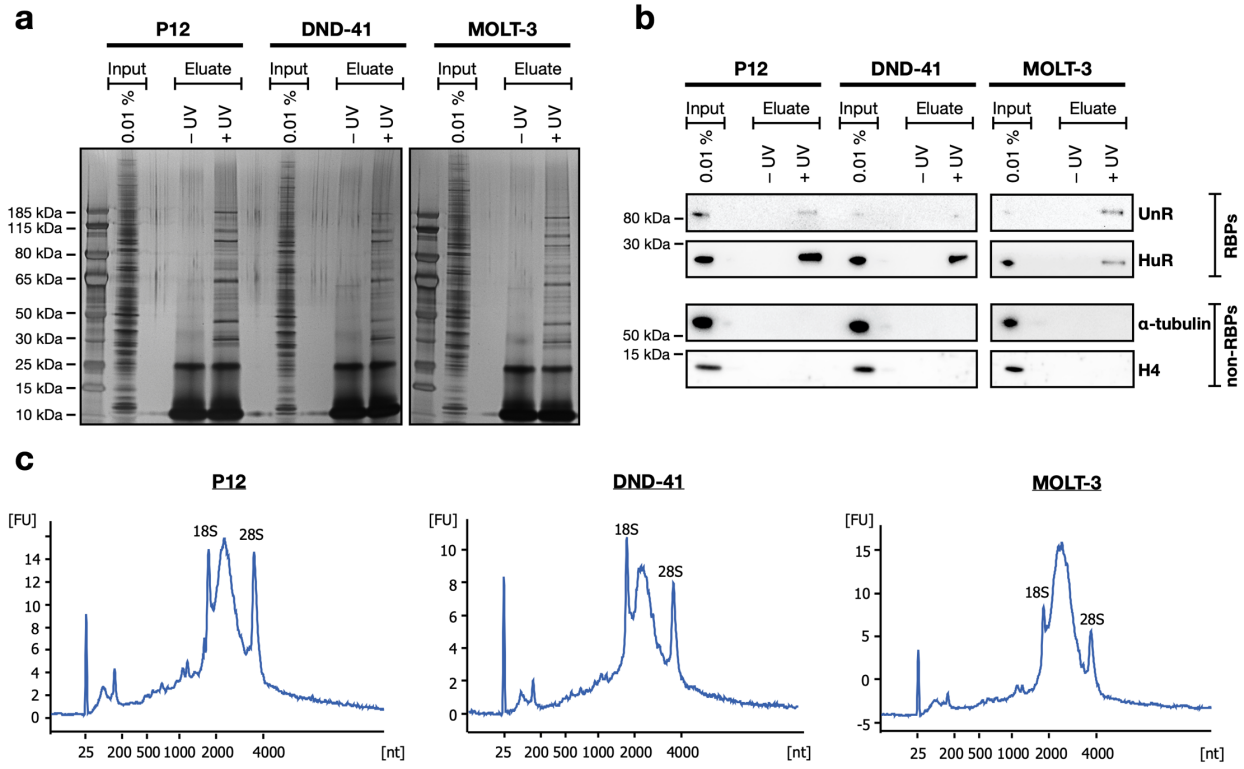


Figure 25: Monitoring the adapted RNA interactome capture protocol with selected cell lines representing T-ALL on protein and RNA level. RBPs are covalently bound to polyadenylated RNAs in living cells by UV-crosslinking (+UV). Non-irradiated samples serve as negative controls (-UV). Following cell lysis, the RBPs are captured on oligo(dT) magnetic beads out of the whole cell lysate (input). After stringent washes, proteins are eluted (eluate). (a-b) Samples were digested with RNases, and released proteins were analyzed by SDS-PAGE combined with silver staining (a) and by western blotting against UnR, α-tubulin, HuR, and histone H4 (b). (c) Proteins of aliquots taken after the elution were digested with proteinase K and the released RNAs were analyzed by bioanalyzer.

Taken together, these experiments showed that the use of cortical T-ALL cell lines requires increased amounts of input and underline the need for further system-specific adaptions, aiming to further increase the mRNA enrichment while reducing rRNA and DNA contaminations.

Addressing the need for a potential internal spike-in control

In order to identify the complete repertoire of RBPs involved in T-ALL, this project investigated the RBP composition in healthy T-cell precursors and in different T-ALL cell lines, with and without SAHA-treatment. Ideally, this identification of RBPs would be performed in one single MS run to allow best comparability between the analyzed samples. However, these studies were likely to require several proteomic analyses by LC-MS/MS. To address the need for a potential spike-in control for comparing RBPs across different MS runs in advance, the RNA interactome capture samples of P12, DND41 and MOLT3 were analyzed for potential suitable internal standard proteins common to all three of them. In case of insufficient similarity, large quantities of RNA interactome captures of different T-ALL cells should have been prepared, pooled and added to into each MS run as spike-in controls. Hence, a set of proteins serving as internal standards was highly favored.

From a total of 4886 proteins detected in the full proteomes, 922 proteins were also present in the RNA interactome capture samples, whereas 40 proteins were exclusively found in samples after oligo(dT) pull-down (**Figure 26a**). For the full proteome the signal sum distributions of all three T-ALL cell lines were equally distributed. In contrast, the signal sum distribution of the RNA interactome capture samples was globally higher for UV-crosslinked samples compared to non-irradiated controls, accounting for enrichment in irradiated samples. However, the extent of enrichment appeared to differ between the cell lines. Whereas for MOLT3 a relatively large average difference of the signal sum distributions from the +UV sample compared to the -UV control implies strong enrichment, the in comparison relatively small average difference in the signal sum distribution of DND41 suggests less enrichment (**Figure 26b**).

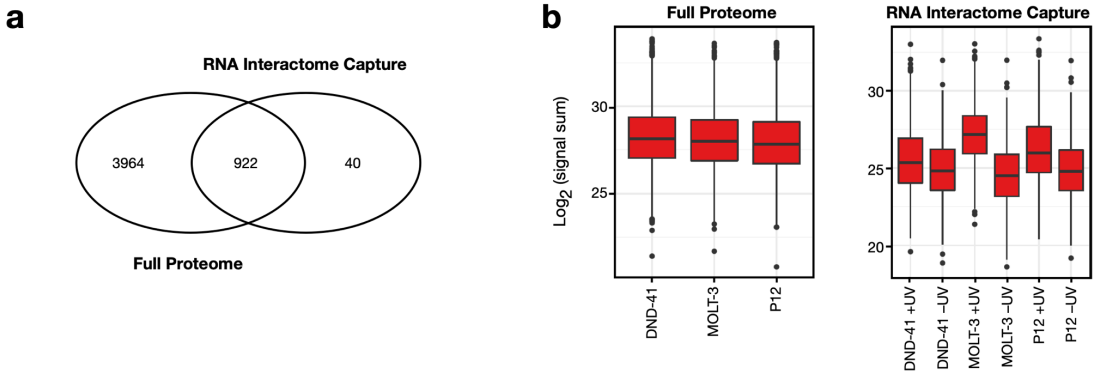


Figure 26: Overview of the proteomic analysis of the cortical T-ALL cell lines P12, DND41 and MOLT3. Following oligo(dT) capture and RNase treatment (RNA interactome), irradiated elution samples containing isolated proteins (+UV), non-irradiated controls (-UV) and whole cell lysates (full proteome) were subjected to proteomic analyses by MS. (a) Protein identification overview comparing the total number of proteins identified in the full proteome (left) to

the total number of proteins identified in the RNA interactome capture samples (right). (b) Boxplot of the signal sum obtained of proteins identified in the full proteome (left) and RNA interactome (right) of the cortical T-ALL cell lines.

To investigate this apparently differential enrichment and to obtain insights into the extent of similarity of the identified proteins across the three cortical T-ALL cell lines, the MS data were further analyzed (**Figure 27**). In a first step, proteins identified in the full proteomes were compared to each other to evaluate how similar the protein compositions in the respective cell lines were before enrichment by RNA interactome capture. Overall, good correlations were observed, however, several proteins showed already highly differential abundance in the full proteome (**Figure 27a**). Second, the correlation of the protein enrichment with the abundance in the full proteome was assessed. Whereas, DND41 shows enrichment of both, abundant and less abundant proteins, MOLT3 data show a stronger enrichment of highly abundant proteins, and in P12 more less abundant proteins were stronger enriched (**Figure 27b**). Further, the protein enrichment between the cell lines was compared, revealing over all a good correlation. Nevertheless, substantial differences can be seen for some proteins when compared to MOLT3 (**Figure 27c**). Ultimately, the difference between protein enrichments were compared to the difference in full proteome abundance of the respective cell lines (**Figure 27d**). These comparisons allowed to identify proteins which did not show differential abundance already in the full proteomes and were constantly enriched in RNA interactome samples of all cell lines, i.e. proteins which are suitable to serve as internal standards. In total, a promising set of 197 of such possible standard proteins was identified, which are depicted in **Figure 27** as red dots.

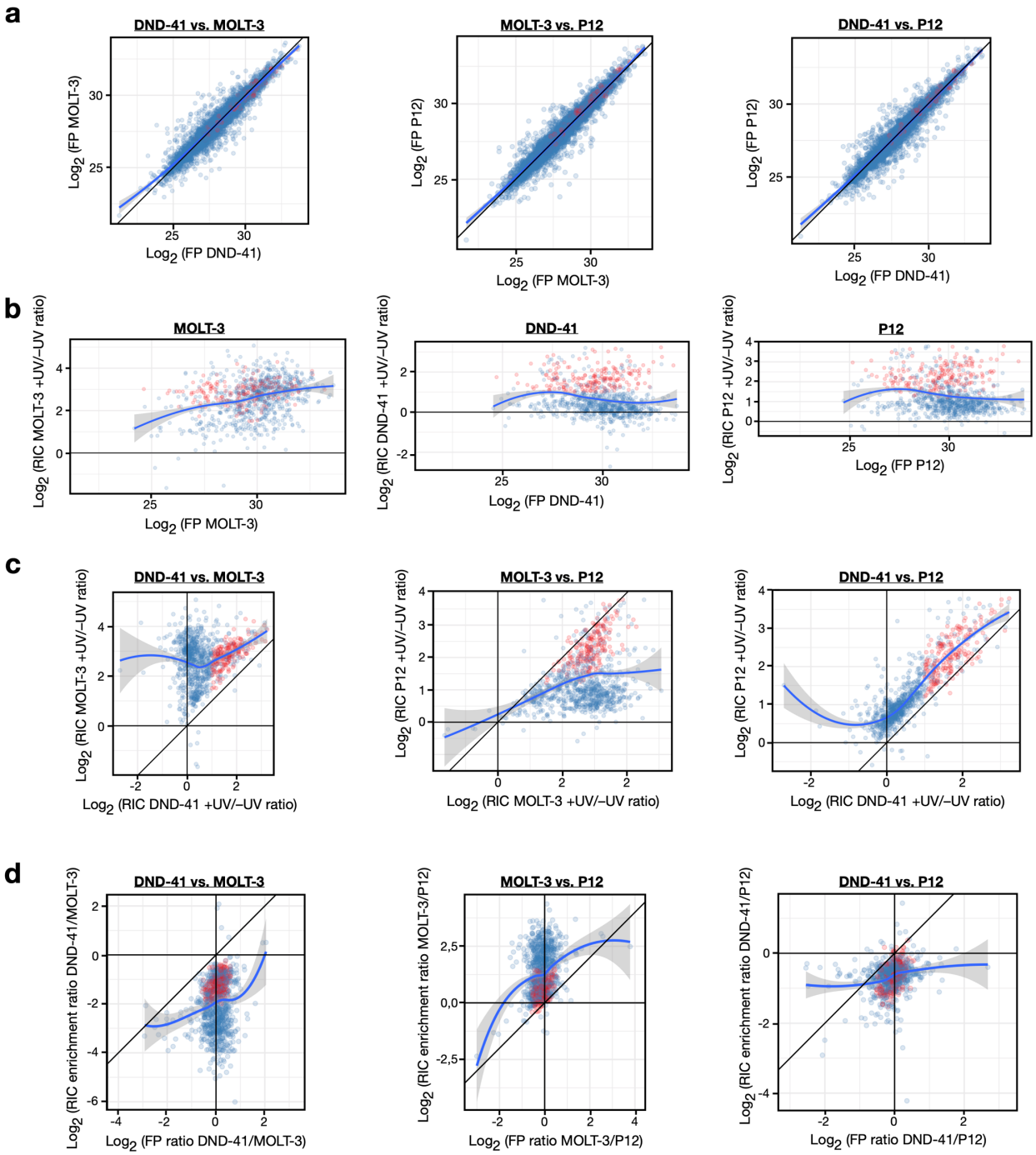


Figure 27: Mass spectrometry data analysis of the cortical T-ALL cell lines P12, DND41 and MOLT3. After protein identification, data from full proteome (FP) and RNA interactome (RIC) samples were analyzed for similarity in the protein composition between the different cell lines. Each dot represents one protein. Proteins constantly and similarly enriched in all three cell lines represent potential standard proteins and are depicted by red dots. Blue line represents the local regression line fit. (a) Scatter plots depicting the correlation of the full proteomes of the three cell lines. (b) Scatter plots of the protein enrichment of UV-crosslinked samples over controls compared to the full proteome of the respective cell line. (c) Scatter plots comparing the protein enrichment of UV-crosslinked samples over controls amongst the three cell lines (d). Scatter plots comparing the protein enrichment ratio of different cell lines to the full proteome ratio of the respective cell lines.

The obtained set of proteins was benchmarked with a GO domain analysis using DAVID[216,217] annotations to search for representative features (**Figure 28**). Despite the observed signs for rRNA and DNA contamination, functions involving interaction with RNA are over-represented in the GO domain *Molecular Function*. Consistently, "RNA processing" is the most over-represented category in the GO domain *Biological Process*, followed by further RNA biology-related functions and processes such as "mRNA metabolic process" and "RNA splicing". As expected, DNA- and rRNA-related terms are also present (e.g. "DNA binding", "transcription", "structural constituent of the ribosome"). These results are in line with the bioanalyzer results and the previous assumption, that in addition to mRNA, DNA and rRNA were present in the eluates.

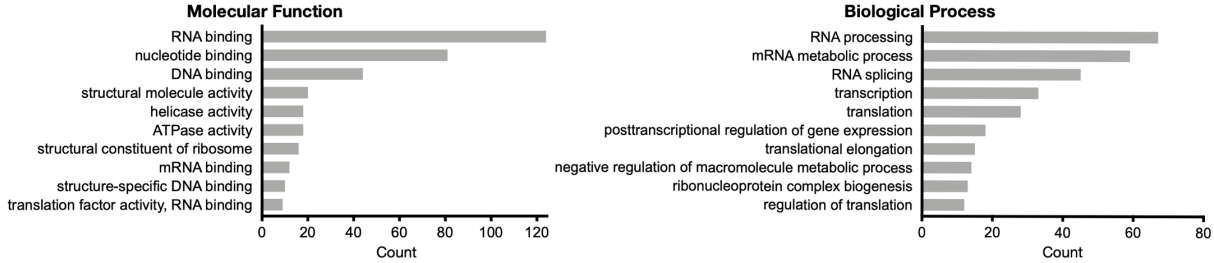


Figure 28: GO domain analysis of a set of 197 potential internal standard proteins using the DAVID Functional Annotation Tool[216,217]. GO domain analysis of Biological Process (left) and Molecular Function (right). RNA-biology related terms, functions and processes are highly over-represented.

However, overall terms linked to RNA-binding far exceed terms related to DNA-binding, thus, validating the check for internal standard proteins from this series of test RNA interactome captures. Although these data and the obtained set of possible internal standard proteins were generated from single experiments, the results demonstrated sufficient similarity between the cortical T-ALL cell lines so that a generation of spike-in controls was considered redundant.

4.2.3 From oligo(dT)-based RNA interactome capture (RIC) to LNA-based enhanced RNA interactome capture (eRIC)

Similar issues of yielding relatively low mRNA amounts, while capturing relatively high DNA and rRNA contaminations, when applying the RNA interactome capture to T-cells were observed simultaneously by Joel Perez-Perri, a colleague in the Hentze laboratory (personal communication). Therefore, we further optimized the classical RNA interactome capture method using oligo(dT) magnetic beads, developing the enhanced RNA interactome capture (eRIC) technique using locked nucleic acid (LNA) oligonucleotides[178]. LNA nucleosides are nucleic acid analogues comprising a methylene bridge, in which the ribose is "locked" in the ideal position for Watson-Crick binding. When LNA nucleosides are incorporated into RNA oligonucleotides,

hybridization properties with the complementary strand are improved. As a result, LNA-based oligonucleotides exhibit increased affinity and higher thermal stability compared to regular RNA oligonucleotides.[188,189,z] As such, the LNA technology offers the possibility to perform the RNA interactome capture at higher temperatures and allows the incorporation of a medium temperature, low salt (pure water) pre-elution step into the protocol, before ultimately eluting RNA by heat elution at 90 °C or RNase elution. These measures lead to reduced DNA and rRNA contaminations whilst the increased affinity improved the yield. Therefore, the performance of this newly developed eRIC protocol with the cortical T-ALL cell lines was tested and compared to the original oligo(dT) RNA interactome capture. For this purpose, the cell line P12 was used as a representative T-ALL cell line, as it offers the best growth behavior and handling. In order to obtain best comparability, the input of all samples was adapted to an amount of 20 mg protein per condition for each comparison. In addition, the classical RIC protocol was compared to the eRIC protocol using syringe lysis and homogenization for both approaches. However, cell lysis and homogenization by syringe proved difficult for primary thymocyte samples and the use of the Precellys cell homogenizer in later stages of this project was required. Therefore, a further comparison involved the originally published eRIC protocol (using a syringe) versus eRIC upon introduction of Precellys for cell lysis and homogenization and the associated changes in the protocol. For these comparisons, two biological replicates were used for each condition. For each replicate, proliferating P12 cells were either irradiated or not and next equally split for each protocol: 1) *RIC syringe*, 2) *eRIC syringe* and 3) *eRIC Precellys*.

For all RNA interactome capture protocols, silver staining shows profoundly different patterns of isolated proteins (eluate) compared to the whole cell proteomes (input). Whereas complex protein patterns and enrichment of several bands is observed in the UV-crosslinked samples, control samples of non-irradiated cells are (despite of the RNase bands) clean. Similar but weaker patterns are obtained from eRIC pre-eluates (**Figure 29a**). These results indicate a successful capture but also some loss of RBPs in the pre-elution step. A comparison of the band patterns of RIC versus eRIC eluates reveals notable differences, indicating that specific proteins are captured by both methods with differential efficiency, which likely reflects unspecific binders associated

[z] Qiagen webpage: What is LNA and Why is it Such a Powerful Research Tool? (https://www.qiagen.com/de/service-and-support/learning-hub/technologies-and-research-topics/lna/benefits-of-lna/); Requested: 2020.09.16

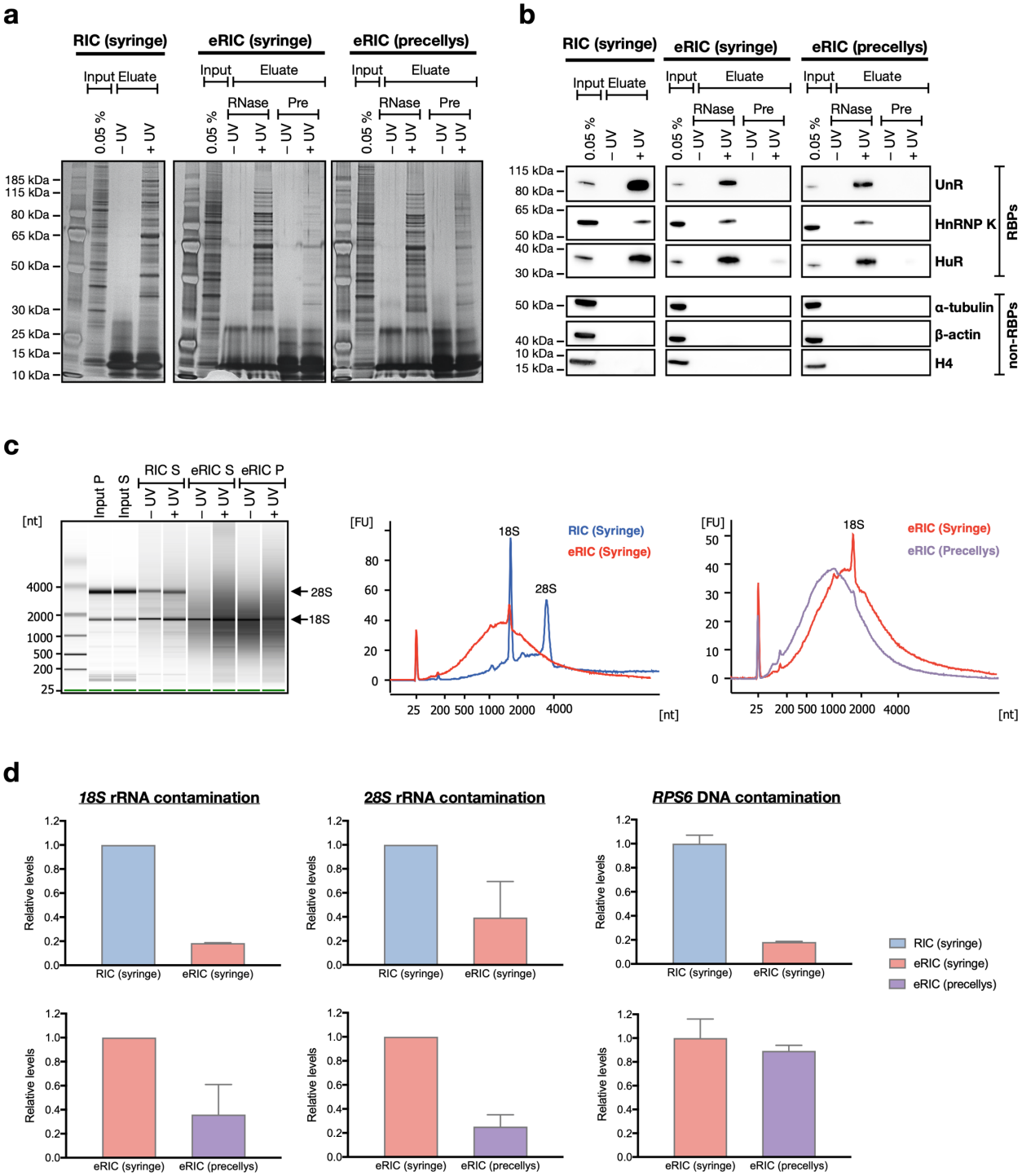


Figure 29: Monitoring the comparison of the original oligo(dT)- based RNA interactome capture (RIC) and the newly developed LNA-based enhanced RNA interactome capture (eRIC) on protein and RNA level. For each condition, RBPs in at least 100 x 10^6 P12 cells are in vivo covalently bound to polyadenylated RNAs by UV-crosslinking (+UV). Non-irradiated samples serve as negative controls (-UV). Following cell lysis, either by syringe or by Precellys, the RBPs are captured out of the whole cell lysate (input) either by traditional RNA interactome capture via oligo(dT) magnetic beads or by eRIC using magnetic beads linked to LNA-based oligonucleotides. After stringent washes, proteins in the oligo(dT) protocol are directly eluted (eluate). Proteins captured by eRIC are eluted after a 40 °C step in water (pre-elution) either by a 90 °C step (heat elution) to preserve the RNA or by RNases (RNase-eluate). (a-b) Released proteins from RNase eluted samples of eRIC and elution samples of oligo(dT) pull-downs after RNase digest were analyzed by SDS-PAGE combined with silver staining (a) and by western blotting against UnR, HnRNP K, HuR, β-actin, α-tubulin and histone H4 (b). (c-d) Nucleic acids isolated by RIC or eRIC were analyzed by bioanalyzer (c) and by RT-qPCR (d). Data show mean ±SD derived from RT-qPCR experiments performed in triplicates.

with rRNA and DNA contamination and proteins captured by eRIC only due to its improved enrichment compared RIC. In contrast, the band patterns of eRIC eluates obtained after syringe versus Precellys lysis and homogenization appear alike, suggesting that the capture efficiency after introduction of the Precellys lysis remains similar. Western blot analysis shows strong enrichment of the well-known RBPs UnR, HnRNP K and HuR in all UV-crosslinked eluates and their absence in non-irradiated control eluates. Conversely, the abundant cellular proteins β-actin, α-tubulin and histone H4 are depleted from the eluates of all conditions. In line with the silver staining results, the RBP HuR is present in the pre-eluate of the UV-crosslinked sample of eRIC confirming the loss of a negligible amount of target proteins along with the unspecific binders (**Figure 29b**). Taken together, these protein quality control experiments demonstrated high selectivity against protein abundance and specificity for RNA binding with all three RNA interactome capture protocols. Compared to previous RNA interactome captures the UnR signal increased in the RIC eluate as well, which is likely to result from both, increased protein input and improved RBP enrichment.

Bioanalyzer Chip evaluation of the eluates revealed substantially different nucleic acid profiles for RIC and eRIC. For eRIC, the enrichment of nucleic acids corresponding to the same size range as mRNA is profoundly increased compared to RIC. Although the pattern of RIC shows partial depletion for both, 18S and 28S rRNA, compared to the input, the rRNA peaks in the profile of eRIC are substantially further decreased. Strikingly, the 28S rRNA appears to be completely depleted from the eRIC eluates and as such the bioanalyzer band pattern of eRIC is prevailed by an even smear between ~500 and ~4000 nucleotides; a size range associated with polyadenylated mRNA. In contrast, RIC eluates are dominated by the rRNA bands. Comparing the profiles of *eRIC syringe* versus *eRIC Precellys*, the differences show even further depletion of 18S rRNA and a minor shift in the length distribution of the nucleic acids. These results indicate improved mRNA capture efficiency and rRNA depletion with eRIC compared to RIC, the latter especially when using the Precellys for lysis and homogenization (**Figure 29c**). To further assess the reduction of rRNA and DNA contaminations by eRIC, eluates were analyzed by qPCR. For both, 18S and 28S rRNA, a strong reduction was observed when comparing *RIC syringe* to *eRIC syringe*. Comparing *eRIC syringe* versus *eRIC Precellys*, rRNA contamination was further decreased. Co-purified DNA was monitored by analysis of RPS6 encoding DNA without reverse transcription. Whereas *eRIC syringe* compared to *RIC syringe* yields a drastic reduction of co-purified DNA, the introduction of Precellys lysis (i.e. *eRIC syringe* vs. *eRIC Precellys*) does not influence the degree

of DNA contamination. Hence, the qPCR findings for rRNA contamination reflect the bioanalyzer results, i.e. they confirm strong rRNA depletion by eRIC, especially when using a Precellys homogenizer, and demonstrate that by using eRIC, DNA contamination can be reduced (**Figure 29d**).

Next, the RNA interactome capture protocols were compared by proteomic analysis. Proteins significantly enriched in the +UV sample over the -UV control (FDR < 0.05, FC > 2) were considered with high confidence as RBPs ("hits"). Using *RIC syringe*, 427 RBPs were identified, for *eRIC syringe* 426 RBPs, and for *eRIC Precellys* 454 RBPs were identified (**Figure 30a**). An overlap and detailed comparison of '*RIC syringe* versus *eRIC syringe*' and '*eRIC syringe* versus *eRIC Precellys*' revealed 67 and 34 significantly differential enriched proteins, respectively (**Figure 30b**). As expected, with eRIC higher +UV/-UV ratios were obtained. In addition, *eRIC syringe* and *eRIC Precellys* show both a high degree of correlation (**Figure 30a & b**). Focusing on high confidence RBPs only, a number of 51 (26 and 25 RBPs for *´RIC syringe* vs. *eRIC syringe*`) and 58 (15 and 43 for *´eRIC syringe* vs. *eRIC Precellys*`) proteins are considered as RNA-binders exclusively by one RNA interactome capture protocol. However, the majority of RBPs overlap between all three RNA interactome capture protocols (**Figure 30c**). Noteworthy, STRING network analysis uncovers that RBPs significantly more enriched or unique to *RIC syringe* are associated with the ribosome complex (**Figure 30d**). These results are in line with the decreased rRNA contaminations found in the preceding RNA quality controls. Overall, mass spectrometry analysis reveals that all three protocols identify a similar number of RBPs. Although a number of specific RBPs are unique to each protocol or differentially enriched, the enrichment and identification of the majority of RBPs remain at least similarly efficient.

In conclusion, eRIC demonstrated superior mRNA enrichment and depletion of genomic DNA and rRNA contaminations compared to RIC. Further, the results obtained here with P12 cells are largely in agreement with the published eRIC studies using JURKAT cells[178]. The use of the Precellys homogenizer, instead of a syringe, did not compromise RBP enrichment or data quality. Thus, further RNA interactome capture studies were performed using the newly developed LNA-based eRIC protocol with Precellys lysis and homogenization.

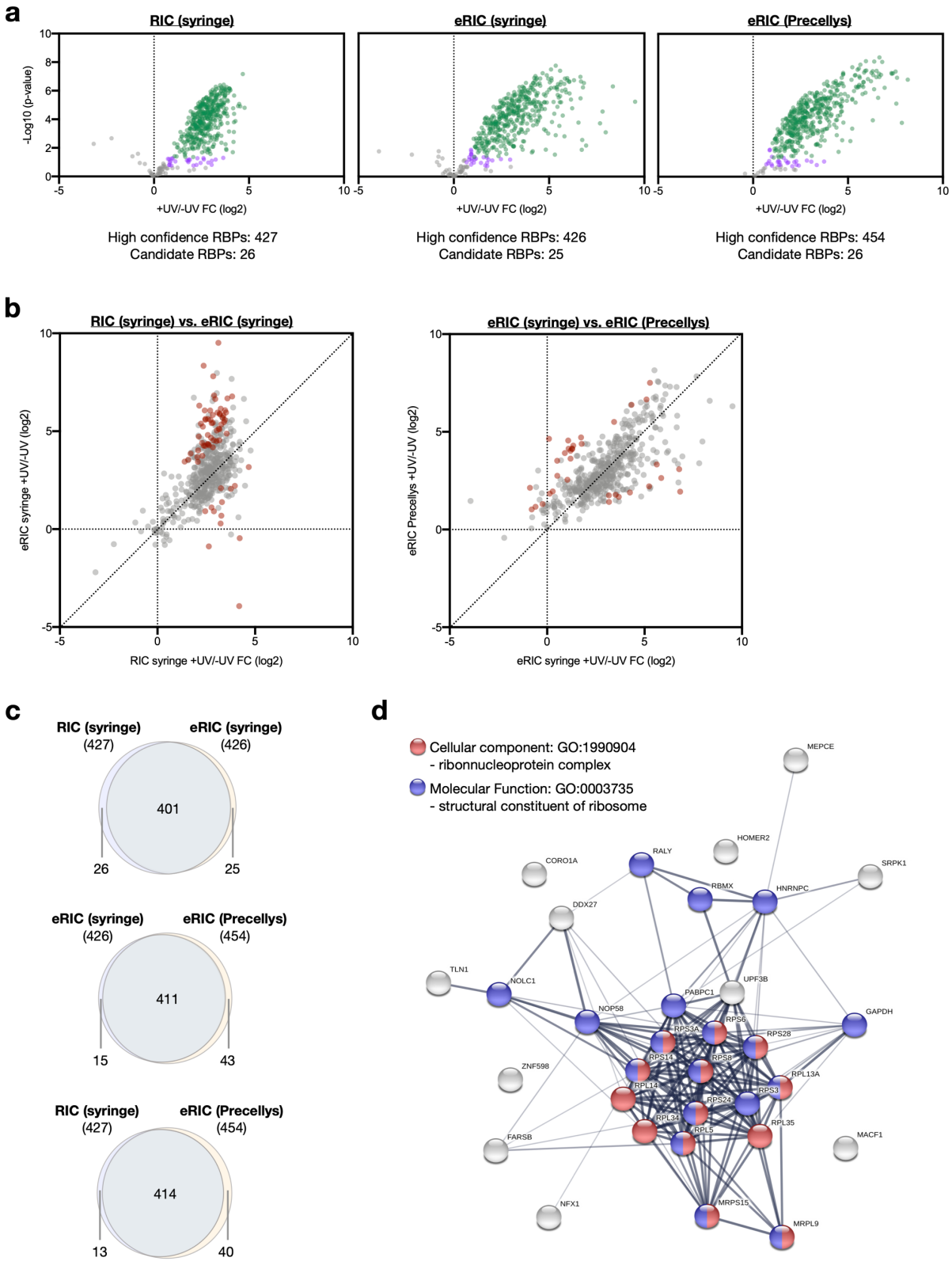


Figure 30: Mass spectrometry data analysis of the comparisons 'RIC syringe vs eRIC syringe' and 'eRIC syringe vs eRIC Precellys'. (a) Volcano plots depicting the protein enrichment of UV-crosslinked (+UV) over non-crosslinked (−UV) eluates as log2-FC ratio (x-axis) versus the p values (-log10; y-axis). Proteins significantly enriched in the +UV sample compared to the −UV control with a FC > 2 and an FDR < 0.05 classified as "hits", and were considered as high confidence RBPs (depicted in green). Proteins with an +UV/−UV enrichment of FC > 1.5 and FDR < 0.2 were considered as "candidates" (purple). "Unspecific" non-RBPs are depicted in grey. (b) Scatter plot comparing the protein enrichment of UV-crosslinked samples over controls (Log2-FC) of RIC syringe to eRIC syringe as well as of eRIC syringe to eRIC Precellys. Proteins showing differential enrichment with an FDR < 0.05 and FC > 2 between the two

compared methods were considered as "hits" (displayed in red). (c) Venn diagram comparing the number of RBPs identified by each RNA interactome capture protocol amongst each other. (d) STRING network analysis of RBPs significantly more enriched by RIC syringe or unique to RIC syringe. Red dots represent proteins associated with the ribonucleoprotein complex and blue dots represent proteins associated with structural constituents of the ribosome.

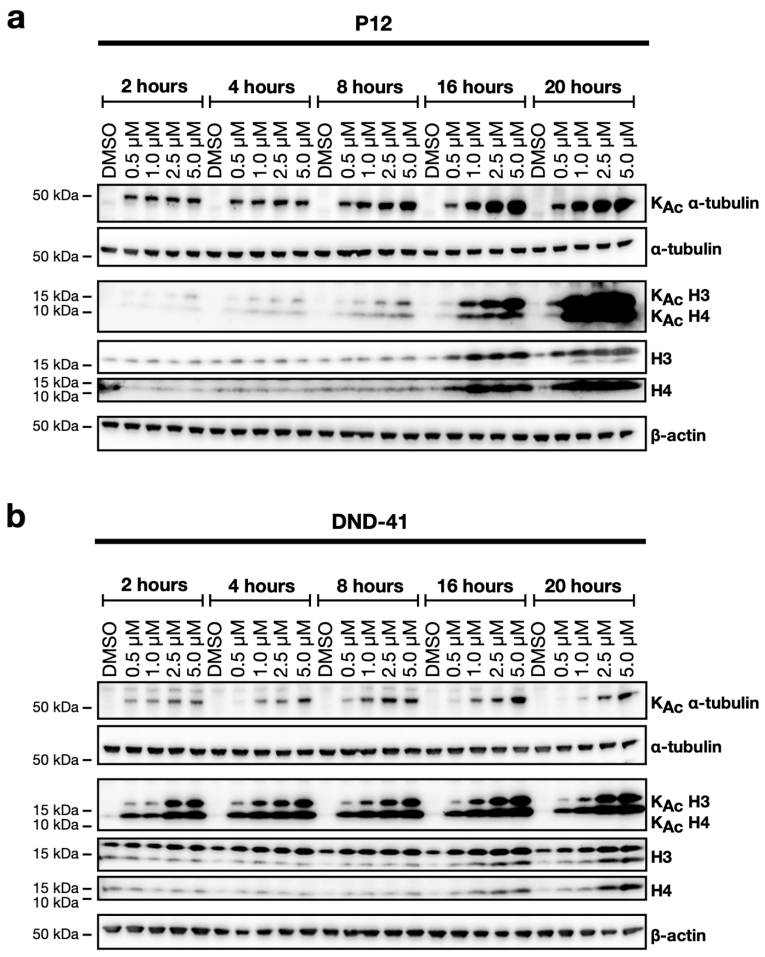
4.3 Identification of appropriate conditions for SAHA-treatment

In addition to identifying the RNA-bound proteomes of T-ALL cells and of their respective normal origin cell equivalents, this project aimed at identifying changes in the RBP composition of T-ALL cells upon HDAC inhibition. Several studies have already demonstrated an anti-leukemic activity of HDAC inhibition in T-ALL, either alone or in combination with other drugs[106,191]. The target spectrum of HDACs is not limited to histones only, but includes non-histone proteins as well[78,104,190,191]. In addition, the binding sites of RBPs are hotspots for posttranslational modifications (PTMs), such as lysine acetylation, and RBP activity is thought to be fine-tuned by PTMs[119,121,128,131]. Thus, RBPs are likely subjects of deacetylation by HDACs and it was hypothesized that interaction with the RNA may be regulated by such post-translational modification. Therefore, the three T-ALL cell lines used in this study were treated with the pan-HDAC inhibitor SAHA (Vorinistat, Zolinza) prior to RNA interactome capture. SAHA is an approved drug for the treatment of cutaneous T-cell lymphoma and may become available for inclusion in T-ALL treatment regimens[53,73,228]. Indeed, SAHA and other HDAC inhibitors are currently under clinical investigation for leukemia treatment[53,92,97,191]. High sensitivity of primary ALL and T-ALL cells towards HDAC inhibition has been reported[97,191,229]. SAHA in particular was shown to affect various biological processes including induction of cell cycle arrest, apoptosis and histone hyperacetylation[97,99,191,229]. These effects are thought to be caused by an increased acetylation of transcription factors such as p53 and increased acetylation of cytoplasmic proteins such as α-tubulin and HSP90[99,109]. Thus, first steps included the identification of an appropriate SAHA concentration and incubation time for the treatment of the T-ALL cell lines, aiming at protein hyperacetylation without yet inducing major secondary effects.

4.3.1 Induction of protein acetylation

To identify suitable treatment parameters, the effect of SAHA on protein acetylation was characterized by immunoblotting. For these experiments, each of the three cortical T-ALL cell lines was treated with either DMSO (negative control) or SAHA at 0.5, 1.0, 2.5 and 5.0 µM for 2, 4, 8, 16 and 20 h. Using antibodies against acetyl-lysine (K_{Ac}) total α-tubulin, histone 3 (H3), and histone 4 (H4), the levels of acetylated tubulin and histone proteins were analyzed. As expected,

SAHA induced tubulin and histone acetylation in a time and concentration dependent manner across all three cell lines. Overall, protein acetylation is rapidly induced and starts to increase considerably on average around 4 – 8 h at 2.5 – 5.0 μM SAHA. After 16 – 20 h treatment and more than 1.0 μM SAHA, protein acetylation appeared to show a dramatic increase, especially for P12 and MOLT3. However, beginning at around 8 h incubation, the expression levels of total histones start to rise as well (**Figure 31**). These findings are consistent with published reports. Tiffon and colleagues observed a similarly rapid induction of protein acetylation when treating cutaneous T-cell lymphoma cells (HUT78 cells) with SAHA. However, in these cells acetylation levels of histones reached their maximum already within 3 h[230]. Flatmark and colleagues, in contrast, observed a similar concentration-dependent increase in acetylation after 12 and 24 h of SAHA-treatment in colon cancer cells (HCT116 cells)[231]. Thus, the extent of protein acetylation induced by SAHA appears to strongly depend on the cell type.



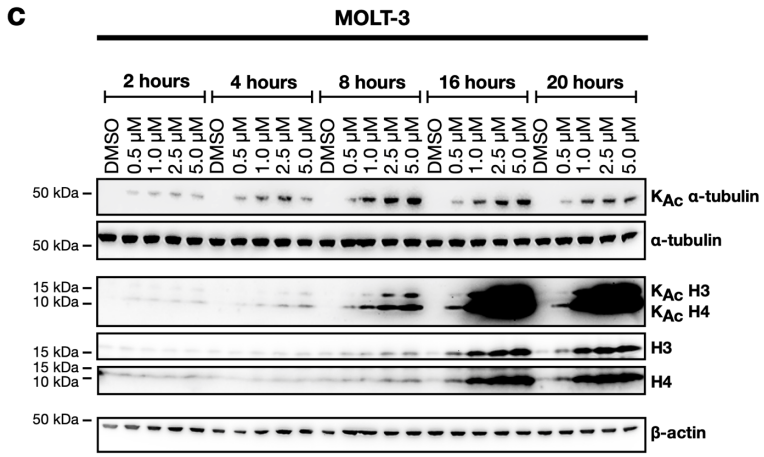


Figure 31: HDAC inhibitor SAHA-induced protein acetylation. Cortical T-ALL cell lines were treated for the stated incubation times with either 0.1 % DMSO or the indicated increasing concentrations of SAHA. After protein extract preparation, the levels of acetylated tubulin and histone were analyzed by western blot immunostaining using an acetyl-lysine (KAc) antibody and comparing to the signal obtained by total α-tubulin, histone 3 (H3) as well as histone 4 (H4) antibodies. Anti β-actin was detected as a loading control. Results shown are representative of at least three experiments.

Overall, the results from all three T-ALL cell lines used here, suggest that a SAHA concentration of 2.5 – 5.0 µM for a treatment period of 4 – 8 h is most suitable for RNA interactome capture studies; allowing to archive maximum hyperacetylation without yet increasing protein expression.

4.3.2 Effect on the induction of apoptosis

Simultaneously to the characterization of the effects of SAHA on protein acetylation, the induction of apoptosis was monitored by Annexin-V-fluorescein isothiocyanate (FITC)/propidium iodide (PI) dual staining flow cytometry. All three T-ALL cell lines were exposed to the same increasing SAHA concentrations and incubation times as before (**Figure 32**).

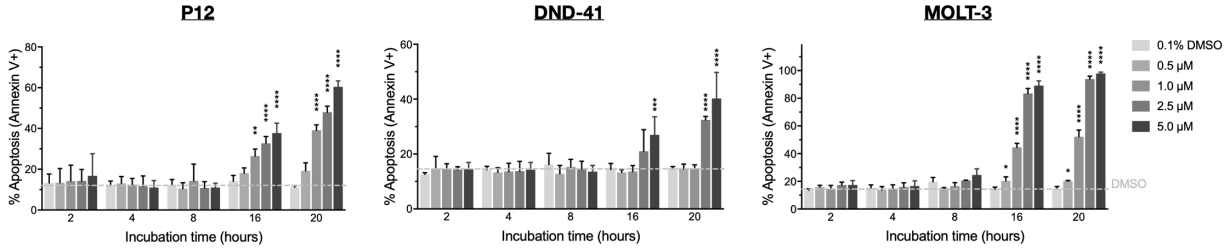


Figure 32: The effect of HDAC inhibitor SAHA on induction of apoptosis in cortical T-ALL cell lines. Cortical T-ALL cell lines were treated with either 0.1 % DMSO or the indicated increasing concentrations of SAHA for the stated incubation times. Induction of apoptosis was analyzed using the Annexin-V-fluorescein isothiocyanate (FITC) / propidium iodide (PI) dual staining assay by flow cytometric analysis. The rate of apoptotic cells was quantified by plotting the percentages of annexin V positive cells. Data show mean ±SD derived from three independent experiments. Statistically significant differences compared to cells treated with DMSO control are shown (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$).

For all of them, the population of apoptotic cells started to progressively increase after longer treatment (> 8 h) correlating with the SAHA dose and treatment length. In DND41 cells, the effect was relatively delayed and only a high dose of SAHA-treatment of 2.5 – 5.0 μM for 16 – 20 h induced apoptosis with a maximum of ~40 % apoptotic cell. In contrast, the same dose of SAHA induced apoptosis in almost a 100 % of MOLT3 cells. Compared to the rapid induction of protein acetylation, the induction of apoptosis requires an exposure to SAHA for an extended time period. This observation is in line with literature suggesting that apoptosis induced by HDAC inhibition is mediated by an "indirect" regulation of gene expression: not through "direct" and "fast" hyperacetylation of histones and promotors but rather through the activity of hyperacetylated transcription factors mediating the expression programs of pro-apoptotic genes[88]. In addition, these results show differential sensitivity of the T-ALL cell lines towards SAHA and account for a SAHA-treatment period of a maximum of 8 h for RNA interactome capture studies.

4.3.3 Cell cycle parameters

To reduce the likelihood that differences in the RNA interactomes are caused by secondary effects, the impact of SAHA on the cell cycle was evaluated as well. P12, DND41 and MOLT3 cells were treated with the same increasing SAHA doses and incubation times as before. After ethanol fixation and PI-staining the proportion of cells in each cell cycle phase was analyzed by flow cytometry (**Figure 33**). Overall, SAHA had no or only modest effects on cell cycle parameters for up to 8 h treatment. For longer incubation times, SAHA-treatment noticeably affected the cell cycle progression in a dose-dependent manner. For all cell lines, the proportion of cells in the G0/G1 phase decreased gradually. Whereas treatment appears to delay the G2/M phase in P12 and DND41 cells, leading to an accumulation of cells in S phase, MOLT3 cells directly progress into G2/M phase.Thus, these findings suggest that SAHA-treatment induces cell cycle arrest in S and G2/M phase in these T-ALL cell lines and contrast literature stating that malignant cell lines generally arrest in G1 upon SAHA-treatment[109]. Matching previous results, MOLT3 cells show the highest and DND41 cells the lowest sensitivity to SAHA-treatment. Despite temporary statistically significant increases in the proportion of cells in G0/G1 for MOLT3, the effect on cell cycle progression for up to 8 h of treatment were very modest and mostly not statistically significant. In line with the effects on apoptosis, these results supported the decision to use SAHA-treatment for less than 8 h for the RNA interactome capture studies.

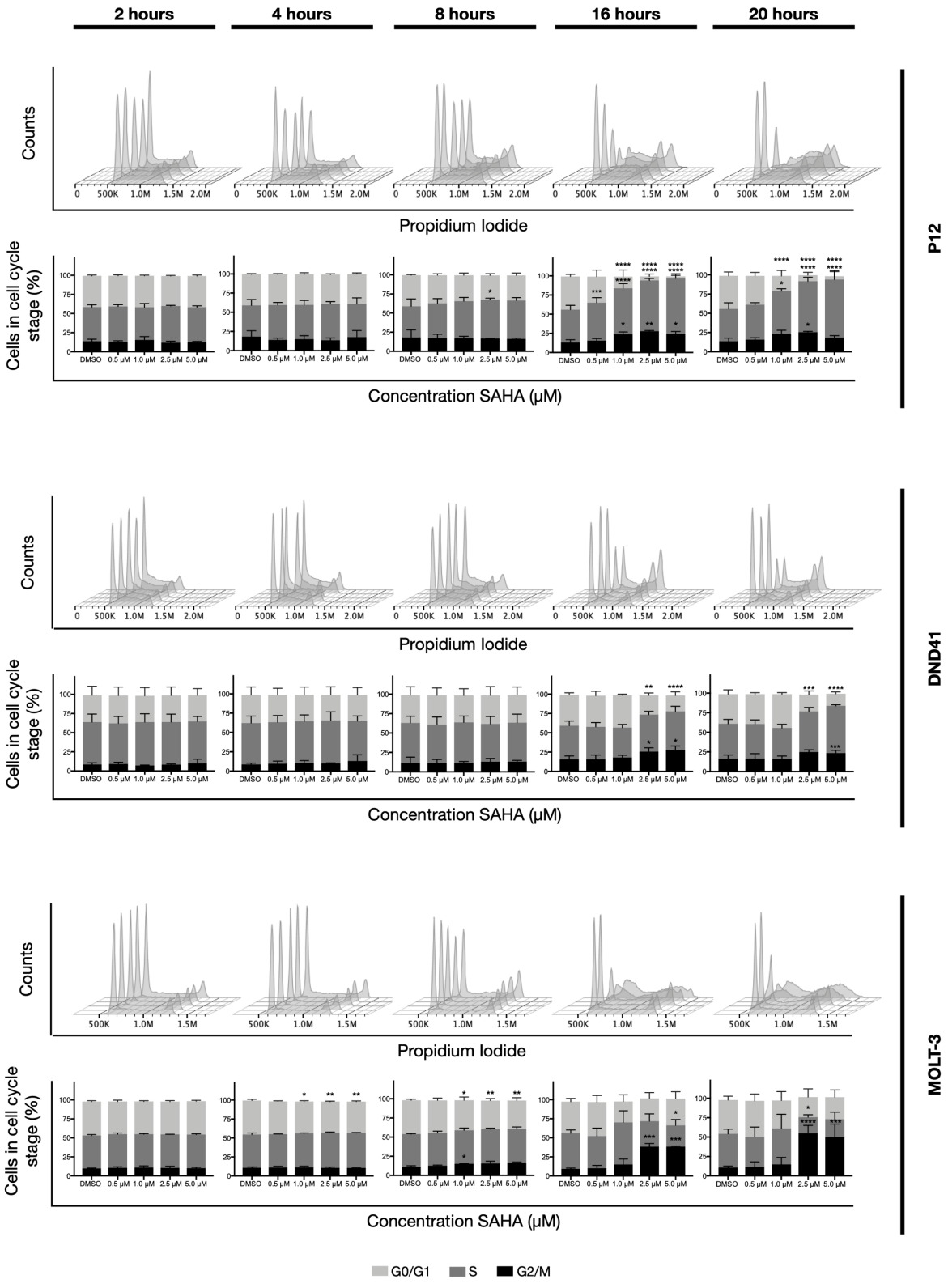


Figure 33: The effect of HDAC inhibitor SAHA on cell cycle progression in cortical T-ALL cell lines. Cortical T-ALL cell lines were treated with either 0.1 % DMSO or the indicated increasing concentrations of SAHA and incubated for 2, 4, 8, 16 and 20 h. Cellular DNA content was determined by flow cytometric analysis upon PI-staining. Images display representative cell cycle profiles of at least three independent experiments. The axes indicate cell counts versus the PI signal. The graphs underneath the cell cycle profiles visualize the proportion of cells in different phases of the cell cycle. Data show mean ±SD derived from three independent experiments. Statistically significant differences compared to cells treated with DMSO control are shown (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$).

4.3.4 Influence on global protein biosynthesis

To ensure that the SAHA-treatment parameters identified so far do not compromise global protein biosynthesis, newly synthesized proteins were metabolically labeled and analyzed. T-ALL cell lines were treated either with DMSO (negative control) or cycloheximide (positive control) or SAHA at 0.5, 1.0, 2.5 and 5.0 µM for 4 and 8 h. Cells were metabolically labeled by ^{35}S methionine and its incorporation was analyzed by autoradiography of labeled proteins and scintillation counting (**Figure 34**). Coomassie blue staining of the same gels as the autoradiographs confirmed equal overall amounts of proteins. For all three cell lines and across all SAHA concentrations and incubation times, alterations in protein translation are very modest and not statistically significant (~10 %) compared to the respective DMSO controls. Only upon cycloheximide treatment both, autoradiographs and scintillation counts show dramatically reduced levels of incorporated ^{35}S-labeled methionine and thus indicate severely diminished protein biosynthesis. Hence, any SAHA dose of up to 5.0 µM and treatment length of up to 8 h is suitable for RNA interactome capture studies of the T-ALL cells used.

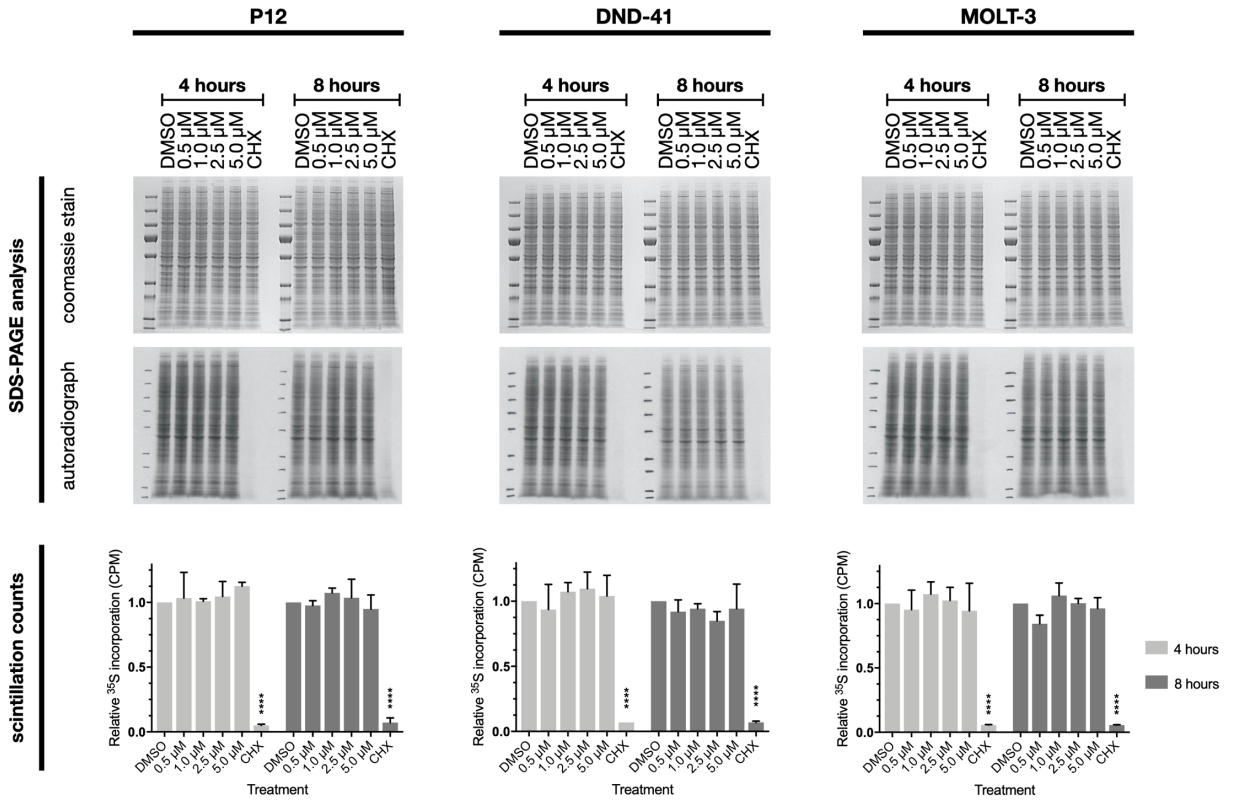


Figure 34: Effect of the HDAC inhibitor SAHA on global protein biosynthesis. T-ALL cell lines were treated with either 0.1 % DMSO (pos. control) or 25 µg/ml cycloheximide (CHX, neg. control) or the indicated increasing concentrations of SAHA and incubated for 4 and 8 h. Cells were metabolically labeled by ^{35}S methionine incorporation. ^{35}S incorporation was analyzed by autoradiography of proteins separated by SDSP-PAGE. Corresponding coomassie stains were used as loading controls. Autoradiograms and coomassie stains of the same respective gels are representative of at least three independent experiments. In addition, ^{35}S incorporation was quantified by scintillation counting and normalized to the total protein amount. Data represent mean ±SD derived from three independent experiments. Statistically significant differences compared to cells treated with DMSO are shown (****$p \leq 0.0001$).

Taken all data together, a treatment length of 4 – 8 h with 2.5 – 5.0 µM SAHA allows the induction of noticeable protein hyperacetylation without evoking secondary effects such as apoptosis, cell cycle perpetuation or influencing global de novo protein biosynthesis. Considering that a higher SAHA concentration readily induces strong protein acetylation, and thus more easily detectable alterations, and since overall incubation times longer than 8 h started to induce secondary effects, for the ultimate RNA interactome capture studies a treatment with 5 µM SAHA for 6 h was determined, aiming at maximum protein acetylation while limiting secondary effects.

4.4 Experimental design and validation of the selected set-up by pilot comparative RNA interactome capture

Following the technical adaptions and the identification of optimal SAHA-treatment conditions, the experimental design for the full comparative eRIC screen was determined. Based on preceding RNA interactome capture studies, it was known that non-crosslinked controls usually contain only few proteins. Moreover, the background was typically constant. Thus, it was decided to combine the two -UV controls (untreated and SAHA-treated) of each T-ALL cell line similar to the RNA interactome capture study published by Liepelt *et al.*[232]. Reducing the background to one common one per cell line simplifies the later statistical analysis and eases the comparability between the treatment-conditions. In addition, such an experimental set-up reduces the number of conditions for the proteomic analyses of untreated and SAHA-treated T-ALL cell lines to nine (**Figure 35, samples 1 – 9**). Using the TMT 11plex label system for proteomic analysis, up to eleven samples can be analyzed within one run. Thus, two more samples could be measured within this analysis. It was reasoned that the RNA interactomes of both, T-ALL cell lines (untreated versus SAHA-treated) and thymocytes, could be analyzed directly together within the same mass spec run. Such a set-up offers optimal comparability between all analyzed conditions and facilitates a direct comparison of ´healthy versus T-ALL` as well as ´T-ALL before and after SAHA-treatment`. To this end, the thymocytes of several individuals needed to be combined to one +UV and -UV control. Previous experiments in our laboratory had already revealed a high degree of similarity between thymocyte samples of different individuals, and thus, combining thymocytes of different individuals to one sample is not expected to compromise data quality[199,200]. A schematic representation of the complete planned experimental design is shown in **Figure 35**.

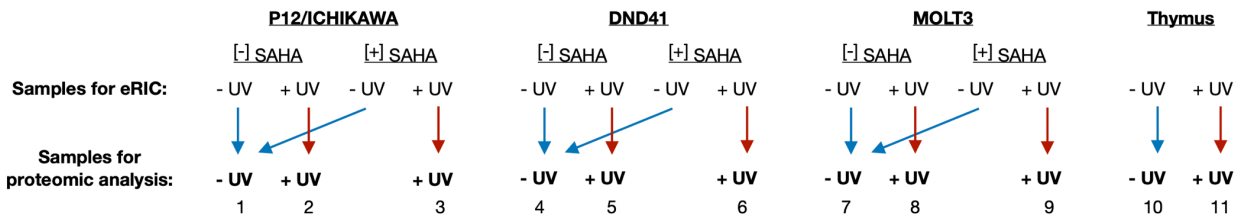


Figure 35: Experimental design for the direct comparative eRIC of normal thymocytes and T-ALL cell lines with and without SAHA-treatment. The identification of the RNA interactomes of T-ALL cell lines and thymocytes is performed in one combined proteomic analysis. Background controls (-UV) of both conditions ([-] SAHA and [+] SAHA) for each cell line are equally combined (blue arrows). UV-crosslinked samples are analyzed individually (red arrows). "Thymus" refers to a combination of thymocytes of three different healthy individuals.

In order to test this experimental set-up and validate the selected treatment conditions, a pilot RNA interactome capture was carried out. Thymocytes from three different individuals were collected and combined (data not shown). T-ALL cells lines were either left untreated (DMSO; vehicle) or were treated with 5 µM SAHA for 6 h. After crosslinking and cell lysis, eRIC was performed. As before, the quality of these RNA interactome captures was monitored by SDS-PAGE followed by silver-staining and immunoblotting of aliquots of the whole cell lysate (input) and the RNase-eluted samples (eluate). Silver staining results are shown in **Figure 36a**. All cell types show a different pattern of eluted proteins (lanes 3, 4, 13, 14, 22, 23, 30) compared to the pattern of the input (lanes 1, 2, 6, 7, 11, 12, 15, 16, 20, 21, 24, 25, 29, 31). In addition, +UV eluates (lanes 3, 4, 13, 14, 22, 23, 30) display complex protein patterns. The -UV controls (lanes 8, 9, 17, 18, 26, 27, 32), in contrast, are strikingly clean, except for the added RNases (lane 34) and residual cellular proteins. These results indicate selective enrichment of RBPs against protein abundance. Interestingly, the eluate obtained from thymocytes (lane 30) shows a similar but different and weaker pattern compared to the eluates from T-ALL cells (lanes 3, 4, 13, 14, 22, 23), although the pattern and amount of input (lanes 1, 2, 6, 7, 11, 12, 15, 16, 20, 21, 24, 25 versus lanes 29, 31) appear mostly similar, suggesting differences in the protein enrichment. Further, SAHA-treatment did not noticeably influence the patterns of input or eluates (compare uneven and even lanes). The specificity and efficiency of the pilot eRIC were further assessed by immunoblot (**Figure 36b**). Immunoblotting against the well-known RNA-binders UnR, HnRNP K and HuR revealed their presence in the +UV eluates of all three T-ALL cell lines (lanes 3, 4, 11, 12, 19, 20). Whereas HnRNP K and HuR are present in thymocytes as well, although to a lesser extent (longer exposure required), UnR is undetectable in both the input (lane 25) and the eluate (lane 26). In addition, all three RBPs are depleted from the -UV controls (lanes 7, 8, 15, 16, 23, 24, 28). Despite their cellular abundance, α-tubulin, β-actin, histone H3 and H4 are not present in the eluates (lanes 3, 4, 11, 12, 19, 20, 26). Using an anti-acetyl lysine antibody (K_{Ac}), bands attributed to

acetylated tubulin and histone H3 and H4 can be visualized. This reveals stronger staining for SAHA-treated samples, reflecting increased protein acetylation (compare uneven and even lanes). Therefore, the western blot analysis ratifies the specificity of eRIC for RNA binding. In line with the silver staining, these results indicate differential protein expression in normal thymocytes versus T-ALL cell lines (e.g. UnR) as well as differential protein enrichment and thus RNA-binding activity (e.g. HnRNP K and HuR). Further, successful induction of hyperacetylation by SAHA-treatment compared to control samples is validated. In summary, eRIC selectively and specifically enriched RNA-binders in both cellular systems, although the protein enrichment and thus potentially the RNA-binding activity differed. In addition, the treatment with 5 μM SAHA for 6 h successfully induced hyperacetylation without compromising the quality of the eRIC selection.

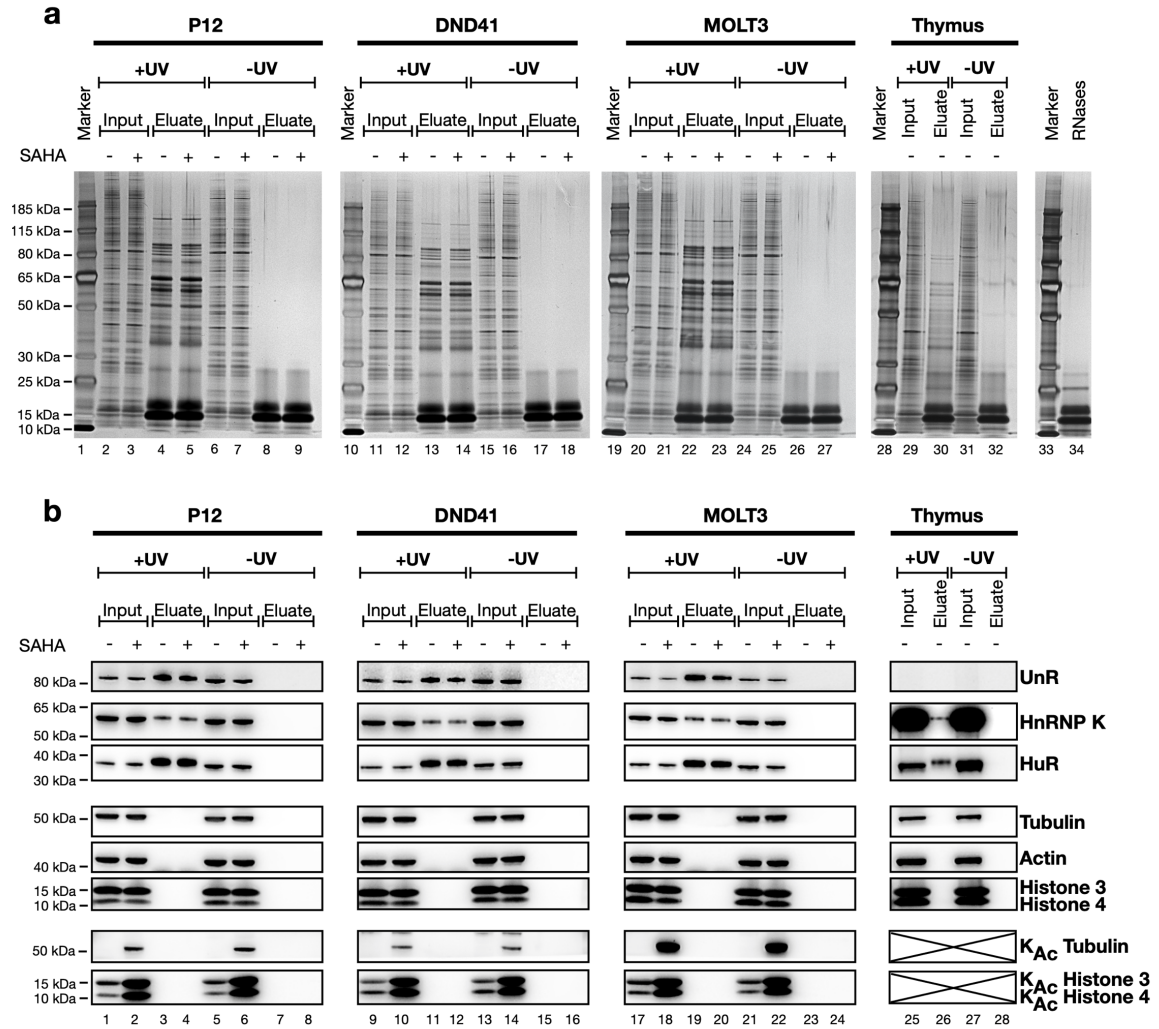


Figure 36: Monitoring the pilot comparative eRIC on protein level. Thymocytes from three different healthy individuals were collected and combined. T-ALL cells were either not treated (0,1 % DMSO vehicle-only) or 5 μM SAHA for 6 h. By *in-vivo* UV-crosslinking (+UV), RBPs were covalently bound to their target mRNAs. Non-irradiated samples served as negative controls (-UV). Following cell lysis, RBPs were isolated from the whole cell lysate (input) by eRIC and eluted by RNases (eluate). (a-b) Eluted proteins were analyzed by SDS-PAGE and subsequent (a) silver staining and (b) western blotting against well-known RNA-binders (UnR, HnRNP K, HuR), cellular abundant non-RBPs (α-tubulin, β-actin, histone H3 and H4) and anti-acetyl lysine (K_{Ac}).

Next, the pilot samples were subjected to proteomic analysis. An overview of this analysis is shown in **Figure 37**. For all T-ALL cell line samples, the signal sum distribution is globally higher for +UV conditions compared to the -UV controls indicating protein enrichment. For the thymus sample the signal sum distribution of both, +UV and -UV condition, is low. Moreover, the median signal sum of the +UV condition is below the median signal sum of -UV background condition, arguing against significant protein enrichment in this sample. Although the preceding quality controls already implied differences in protein enrichment between both cellular systems, a clear enrichment in the +UV condition over the -UV control was observed. These findings, however, clearly confirm lower protein enrichment of eRIC with thymocytes. Even though the signal sum distributions of the T-ALL cell lines indicate successful protein enrichment, the total number of total proteins identified from all eleven conditions in this analysis amounts to only 297. Of these, the numbers of proteins that classify as RBPs (+UV/-UV log2-FC > 2 and FDR < 0.05) ranges between 77 for the thymus sample and 200 for the T-ALL cell line MOLT3. Compared to previous eRIC experiments with T-ALL cell lines, at least 400 RBPs were anticipated and these numbers are unexpectedly low.

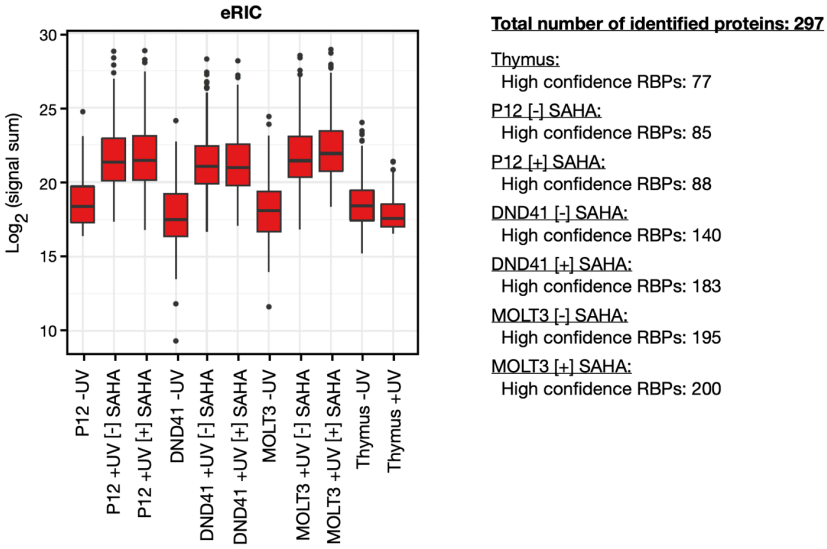


Figure 37: Proteomic analysis of the comparative pilot RNA interactome capture. Irradiated (+UV) and non-irradiated (-UV) eRIC RNase eluates of untreated thymocytes and the T-ALL cell lines P12, DND41 and MOLT3 either untreated (DMSO; vehicle-only) or treated with 5 µM SAHA for 6 h, were subjected to a proteomic pilot analysis by LC-MS/MS. Boxplot (left side) shows the signal sum distribution (y-axis, log2 scale) of proteins identified in the respective conditions (x-axis). Overview of the number of identified proteins in the analysis and the number of proteins considered as RBPs (i.e. proteins enriched in the +UV sample compared to the −UV control with a log2-FC > 2 and an FDR < 0.05) in the indicated conditions (right side).

Interestingly, similarly observations of low protein numbers for eRIC performed with primary cells were observed by colleagues in our group, working with primary osteoblast cells and T-cells from blood respectively (unpublished data). These observations suggest that in general fewer proteins are enriched when applying eRIC to primary cells compared to established cell lines. Primary cells typically show lower proliferation rates compared to their established cell line equivalents, and thus might be transcriptionally less active. As such they might simply harbor proteomes with lower RNA-binding activity.

Since eleven conditions and the use of the TMT 11plex label system require the analysis of more labels compared to previous proteomic analyses with less conditions, longer gradients (here: 12 fractions, 120 min each) were applied to resolve these labels. Thus, the low number of identified proteins may in part be attributed to increased sample dilution: less abundant proteins may be too diluted to reach the detection threshold. In consequence, proteins otherwise identified by deeper analyses may be missed. Indeed, the intensities of proteins identified in this pilot analysis were lower compared to previous eRIC experiments with less conditions. In addition, sample fractionation and QC files of the LC-MS/MS analysis revealed that thymocyte proteins/peptides were highly underrepresented in this analysis (data not shown). These findings indicate, that the high number of conditions analyzed concurrently within the same LC-MS/MS run may have limited the detection of enriched RBPs.

In order to confirm the lower enrichment obtained with thymocytes, a further test eRIC was performed. Thymocyte lysates collected during the optimization of the thymocyte isolation and UV-crosslinking procedure (section 4.1.4) served as input. As before, capture quality was controlled by silver staining and immunoblotting using aliquots of the inputs and eluates. Silver staining reveals the typical complex patterns for +UV samples, which are notably different from those of the inputs, and clean -UV conditions (apart from the added RNases), demonstrating protein selectivity of eRIC applied to thymocytes (**Figure 38a**). Immunoblot analysis discloses the presence of the RBPs UnR, HnRNP K and HuR in +UV conditions and their absence in -UV conditions (**Figure 38b**). However, UnR is barely detectable in the eluate of the second thymus lysate. The non-RBPs α-tubulin, β-actin, histone H3 and H4 are only present in the input and not in the eluates. Thus, the western blot results validate the specificity obtained with eRIC of thymocytes. Although UnR is highly expressed and enriched by eRIC using T-ALL cell lines, its expression or RNA-binding activity in thymocytes appears to be reduced or dependent on the donor individual. In contrast to previous proteomic analysis (TMT 11plex), this analysis (TMT

4plex) yielded substantially higher signal sum distributions for +UV conditions compared to the -UV conditions, suggesting an increased protein detection when measuring fewer samples in one MS run (**Figure 38c**). Nevertheless, the number of proteins identified (169) and classified as RBPs (114) is still lower than expected. Benchmarking of the identified RBPs by GO domain analysis, however, shows that RNA-biology related processes and RNA-binding functions are highly enriched (**Figure 38d**). Thus, these findings confirm that in thymocytes eRIC indeed enriches substantially fewer RBPs than eRIC with T-ALL cell lines, but also that the obtained enrichment is still specific for RNA-binders and thus reliable.

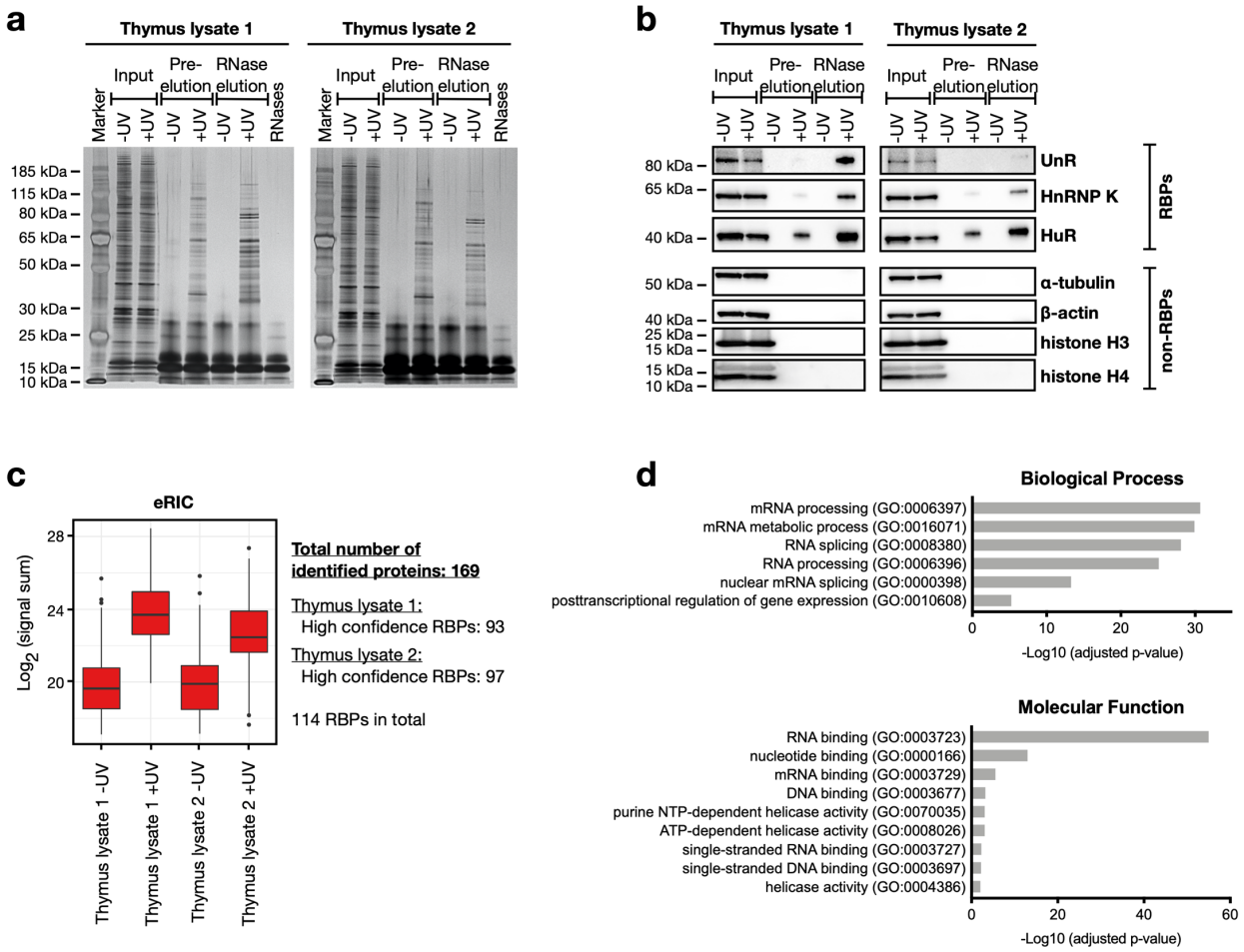


Figure 38: Overview of the pilot eRIC and proteomic analysis using thymocytes only. Two lysates (UV-crosslinked (+UV) and non-irradiated (-UV) condition each), originally obtained during the optimization of the thymocyte isolation and UV-crosslinking procedure, were used as input. After stringent washes, proteins captured by eRIC were subjected to a 40 °C pre-elution step in water to removing residual contaminants and unspecific binders and then to the final elution by RNases (RNase elution). Eluted proteins were analyzed by SDS-PAGE and subsequent (a) silver staining and (b) immunoblotting against known RBPs (UnR, HnRNP K, HuR) and abundant non-RBPs (α-tubulin, β-actin, histone H3 and H4). (c) Boxplot of the signal sum distribution (y-axis, log2 scale) of proteins identified in the respective thymocyte conditions. Right: summary of the number of in total identified proteins and the respective numbers of proteins considered as RBPs (+UV over −UV conditions with a log2-FC > 2, FDR < 0.05). (d) GO domain analysis of proteins classified as RBP using DAVID[216,217]. GO-term analysis of the domains *Biological Process* and *Molecular Function*.

Overall, these pilot RNA interactome captures showed, that planned experimental design for a full comparative eRIC screen would compromise the analysis depth for all samples. Analyzing all T-ALL cell line samples and both thymocyte conditions within one LC-MS/MS run using the TMT 11plex label system requires longer gradients and dilutes the samples. Less abundant proteins are not detected resulting in a reduced number of identified proteins. For samples of lower complexity, such as the thymocyte eluates, the relatively low number of identified proteins is further reduced. To ensure best resolution and data quality, it was decided to perform the eRIC analyses for each cellular system (T-ALL cell lines and thymocytes) separately. Although this does not allow a direct comparison of absolute numbers, this experimental set-up still allows to investigate differences between T-ALL and normal thymocytes.

4.5 RNA interactome capture unravels the RNA-bound proteome of pediatric T-ALL and identifies alterations upon SAHA-treatment and compared to normal T-cell progenitors

To ultimately address the major goal of this project, i.e. to uncover the complete RNA-binding proteome of T-ALL, including its changes in response to HDAC inhibitor treatment and upon leukemogenesis, a series of eRIC analyses was performed. In a first set of experiments, eRIC was applied to T-ALL cell lines, preceded either by normal culture conditions or by treatment with the HDAC inhibitor SAHA. These analyses addressed the characterization of the collective of RBPs in T-ALL and the identification of changes in its composition upon HDAC inhibition. In a second set of experiments, thymocytes of different donors were collected and subjected to eRIC in order to explore the RBP composition of normal T-lymphocyte progenitors. Next, the RNA interactomes of both, normal T-lymphocyte progenitors and T-ALL cells, were compared to each other to identify alterations in the RNA interactome upon leukemogenesis.

4.5.1 Exploring the RNA-bound proteome of T-ALL and its response to HDAC inhibition using SAHA

To identify and explore the T-ALL RNA interactome and its response to HDAC inhibition by SAHA, proliferating P12, DND41 and MOLT3 cells were either grown under normal culture conditions ([-] SAHA; DMSO vehicle-only) or treated for 6 h with 5 μM SAHA ([+] SAHA) prior the eRIC analyses. For subsequent proteomic analyses, the two -UV controls ([-] SAHA and [+] SAHA) of each of the three T-ALL cell lines were combined as depicted in **Figure 39**.

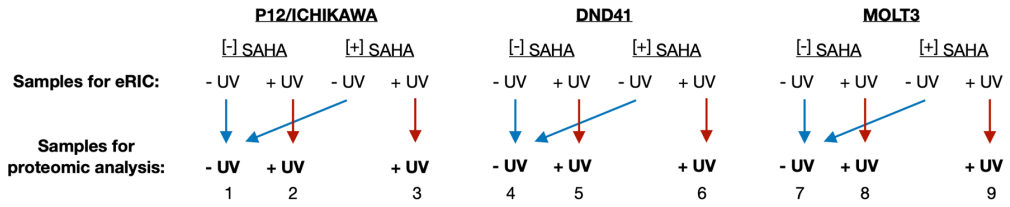


Figure 39: Experimental set-up for the comparative eRIC from untreated and SAHA-treated T-ALL cells. The three cortical T-ALL cell lines P12 (also ICHIKAWA), DND41 and MOLT3 were either grown under native conditions ([-] SAHA; DMSO vehicle-only) or treated with 5 µM SAHA for 6 h ([+] SAHA). Background controls (-UV) of both conditions ([-] SAHA and [+] SAHA) for each cell line are equally combined (blue arrows). UV-crosslinked samples are analyzed individually (red arrows). Three biological replicates were used for each cell line and each condition.

These eRIC analyses, were performed in three independent biological replicates of each cell line and each condition. Full proteomes of each cell line and each treatment condition were analyzed in parallel.

eRIC of untreated versus SAHA-treated T-ALL cell lines and corresponding quality control analyses

eRIC performance and sample quality were monitored using aliquots from the whole cell lysates (input) and from eluates of UV-crosslinked and non-crosslinked conditions. Representative results are shown in **Figure 40.** SDS-PAGE followed by silver staining (**Figure 40a**) reveals complex protein patterns for the +UV eluates of all three cell lines (lanes 4 & 5, 12 & 13, 20 & 21) that profoundly differ compared to the inputs (lanes 1, 2, 6, 7, 10, 11, 14, 15, 18, 19, 22, 23). In contrast, -UV eluates are clean (lanes 8 & 9, 16 & 17, 24 & 25), showing only the added RNases (lane 26) and traces of residual proteins. These findings indicate successful RBP enrichment and reflect the high selectivity combined with the low backgrounds obtained by the eRIC methodology. Treatment with the HDAC inhibitor SAHA ([+] SAHA) did not visibly alter the protein patterns of inputs or eluates (compare uneven to even lanes), reflecting the moderate treatment to limit secondary affects. In addition, eRIC performance and specificity was evaluated by western blot analysis (**Figure 40b**). Results show that the known RNA-binders UnR, HnRNP K and HuR (positive controls) are present in the inputs (lanes 1, 2, 5, 6, 9, 10, 13, 14, 17, 18, 21, 22) and +UV eluates of all three T-ALL cell lines (lanes 2 & 3, 11 & 12, 20 & 21), but not in the eluates of -UV conditions (lanes 7 & 8, 15 & 16, 19 & 20). Abundant non-RBPs such as α-tubulin, β-actin, histone H3 and histone H4 (negative controls) are exclusively detectable in the inputs (lanes 1, 2, 5, 6, 9, 10, 13, 14, 17, 18, 21, 22) and not in any of the eluates (lanes 3, 4, 7, 8, 11, 12, 15, 16, 19, 20, 23, 24). Further, inputs of SAHA-treated ([+] SAHA) samples, show increased acetylation (K_{Ac}) of tubulin and histone H3 and H4 compared to non-treated ([-] SAHA) samples (compare uneven

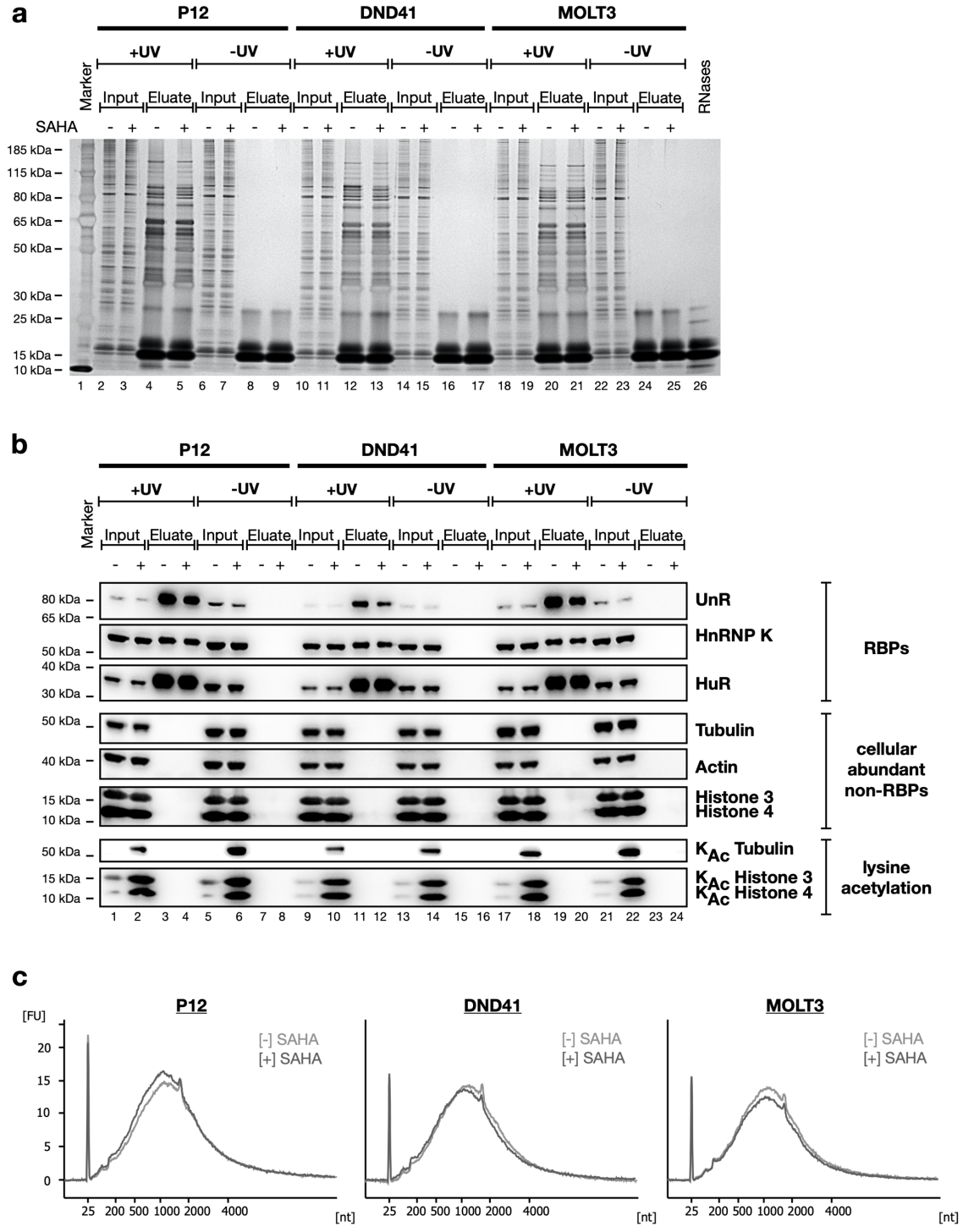


Figure 40: Quality controls on protein and RNA level confirm specific RBP enrichment of eRIC applied to T-ALL cell lines. RBPs were UV-crosslinked (+UV) to polyadenylated RNAs. Non-crosslinked conditions served as background controls (-UV). To isolate the RBPs from the whole cell lysate (input), eRIC was applied: RBPs were captured via magnetic beads linked to LNA-based oligonucleotides, subjected to a series of stringent washes and eluted (eluate) either by 90 °C (RNA quality controls) or RNases (protein quality controls, proteomic analysis). Quality of the eRIC was monitored on protein and RNA level. Proteins from inputs and eluates were separated by SDS-PAGE and detected by subsequent (a) silver staining and (b) immunoblotting against known RBPs (UnR, HnRNP K, HuR), abundant non-RBPs (α-tubulin, β-actin, histone H3 and H4) and anti-acetyl lysine (K_{Ac}). (c) Profiles of isolated nucleic acids were assessed by bioanalyzer. (c) Profiles of enriched nucleic acids were evaluated by bioanalyzer. eRIC analyses and associated quality controls were performed in triplicates, representative results are depicted.

and even lanes). These results demonstrate high specificity of eRIC for RBPs and hyperacetylation in SAHA-treated samples. Analysis of the eluted RNA by bioanalyzer (**Figure 40c**) yields similar profiles for all cell lines and conditions: a single peak spanning a nucleotide range of ~200 – 3000 nucleotides; typically associated with polyadenylated mRNA. Signal attributed to 18S and 28S rRNA, usually observed in inputs or RIC eluates (discussed in section 4.2.3; Figure 29c), is either not detected or very modest. Thus, bioanalyzer data validate mRNA enrichment and depletion of rRNA contaminations.

Overall, these findings document that eRIC selectively and specifically enriched RBPs against a background of abundant cellular proteins in all three cell lines. SAHA-treatment successfully induced protein hyperacetylation without negatively affecting the capture quality or eRIC performance.

Proteomic identification of the T-ALL RNA interactome

Following successful quality control analyses, eRIC eluates were labeled by TMT 9plex and subjected to LC-MS/MS in three independent replicate experiments (see section 3.4.7 for details). Full proteomes were analyzed in parallel. Acquired data were cleaned for batch effects, normalized and only proteins quantified with at least two unique peptides were considered. **Figure 41** shows an overview of the proteomic analysis. The signal sum distributions of the full proteomes of all three cell lines and replicates are equally distributed. For eRIC eluates in contrast, the signal sum distributions of UV-crosslinked conditions are expected to be higher than the corresponding non-crosslinked controls. Indeed, all +UV conditions show similar and clearly higher signal sum distributions compared to the -UV controls, for which similar distributions can be observed as well (**Figure 41a**). These results demonstrate enrichment in +UV eluates over -UV controls with similar enrichment efficiency. In line with these results, principal component analysis of eRIC eluates clusters all +UV conditions and all -UV conditions of the three cell lines and replicates respectively in close proximity (**Figure 41b**). Ultimately, a total of 5830, 5831 and 6024 proteins were identified from the full proteomes and 465, 637 and 511 proteins from the eRIC eluates (**Figure 41c**). These numbers of total proteins match the anticipated range based on previous eRIC experiments with T-ALL cell lines (data not shown).

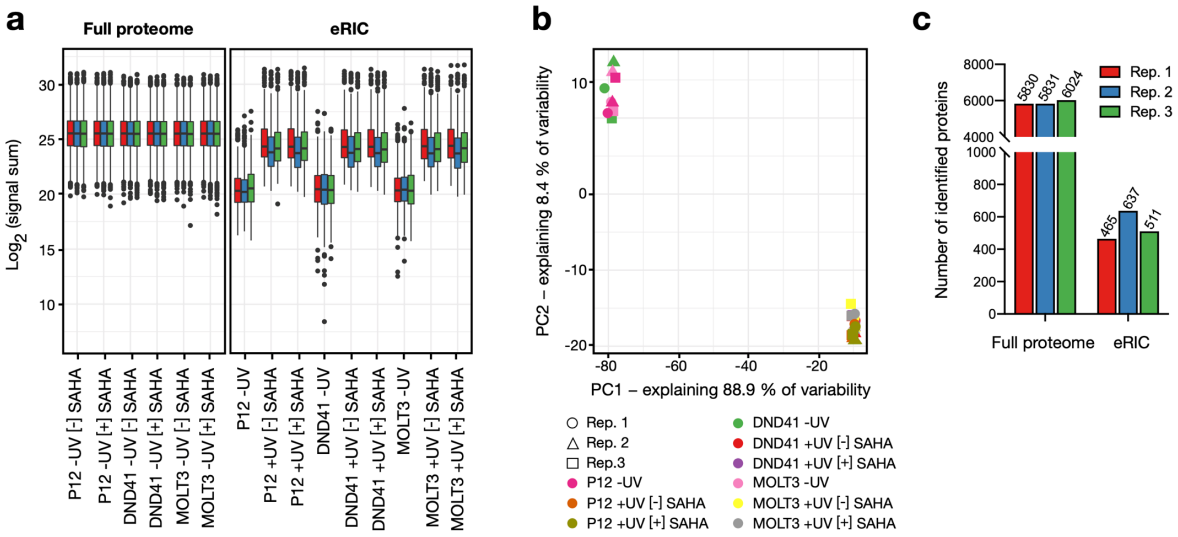


Figure 41: Proteomic analyses for the identification of the RNA-binding proteome of T-ALL and its response to SAHA-treatment. Whole cell lysates (full proteome) as well as irradiated (+UV) and non-irradiated (-UV) eRIC eluates of the T-ALL cell lines P12, DND41 and MOLT3, either untreated (DMSO; vehicle-only) or treated with 5 µM SAHA for 6 h, were subjected to proteomic analysis by LC-MS/MS. Three biological replicates for each cell line and each condition were analyzed. Data were cleaned of batch effects, normalized and only proteins quantified with at least two unique peptides were considered. (a) Boxplot shows the signal sum distribution (y-axis, log2 scale) of proteins identified in the respective conditions (x-axis). Different replicates are depicted in red (replicate 1), blue (replicate 2) and green (replicate 3). (b) Principal component analysis of the analyzed eRIC conditions. Each shape represents one replicate and each color corresponds to one cell line and condition. (c) Overview of the number of proteins identified in the full proteome and in the RNA interactome capture (eRIC) in each replicate.

Overall, these results demonstrate (i) successful and similarly efficient protein enrichment for +UV over -UV conditions with all three cell lines in all three replicates, (ii) good capture performance and robustness for eRIC with P12, DND41 and MOLT, and (iii) high data quality suitable for to reliably identify the T-ALL RNA interactome. To also further ensure good data quality, only proteins identified in at least two replicates were considered for the following downstream analyses.

Comprehensive analysis of the RNA-bound proteome of T-ALL

Typically for RNA interactome capture studies[175,177,178,196], proteins significantly enriched in the UV-crosslinked condition over the non-crosslinked background control (fold-change (FC) > 2 and FDR < 0.05) are considered as "high confidence RBPs". Applying these criteria, a number of 478, 465 and 477 RBPs were identified for the T-ALL cell lines P12, DND41 and MOLT3, respectively (**Figure 42a**, Supplementary table 1). Of these, 459 RBPs were common to all three cell lines, 15 RBPs were identified in at least two and 12 RBPs were found exclusively in one T-ALL cell line only (**Figure 42b**). This adds up to a total of 486 unique RBPs from all three T-ALL cell lines, which were defined here as the collective set of RBPs expressed in T-ALL: the "T-ALL RNA interactome" (Supplementary table 2).

Earlier studies identified numerous unexpected proteins as RBPs, amongst them several (metabolic) enzymes[128,142,233]. In line with these observations, ~15 % (71) of the RBPs within the T-ALL RNA interactome constitute enzymes and ~2 % (10) represent metabolic enzymes (**Figure 42c, left**). Similar percentages were previously observed within the JURKAT RNA interactome (mature T-ALL cells) [178], whereas, for example, the RNA interactome of hepatocellular carcinoma cells (HuH7)[196] harbors profoundly higher percentages of enzymes and especially of metabolic enzymes. This difference is likely due to the fact that hepatocytes are amongst the most metabolically active cell types[234]. Such (metabolic) enzymes were shown to have second functions as RBPs and in some cases can regulate the expression of their own or their target mRNAs[142,235,236] (see section 1.3.2). In addition to enzymes, previous studies identified many unorthodox RBPs that lack classical RBDs[173,175,233,237]. Instead of a stable tertiary structure, these RBPs are enriched in disordered regions[173,233,237–239]. The T-ALL interactome reflects these findings as well: about half (254) of the RBPs do not contain canonical RBDs (**Figure 42c, right**) and an analysis for intrinsically disordered regions shows a significantly higher disorder rank for the T-ALL RBPs here identified relative to a random set of proteins of similar size (**Figure 42d**). Disordered regions can contribute to RBP binding and increase the flexibility and adaptability of RBPs, thereby facilitating modular binding and binding to various structures[129,233]. The other half of RBPs within the T-ALL RNA interactome contains known RBDs (**Figure 42c, right**). The most prevalent ones are RNA recognition motifs (RRM) and KH domains (**Figure 42e**). These results match published reports of RNA interactomes of other cellular systems. For instance, both RBDs were also identified as the most significant ones in the RNA interactomes of HuH7[196] and RAW264.7[232] cells.

Next, the T-ALL RNA interactome was searched for representative features by enrichment analyses. As expected, GO enrichment analysis within the domain *Molecular Function* reveals "RNA binding" to be the most significantly enriched function (**Figure 42f**). Within the GO domain *Biological Process,* terms and processes related to (m)RNA (metabolic) processing and RNA splicing are the most significantly overrepresented ones (**Figure 42g**). Consistently, STRING network analysis shows a major network of numerous RBPs associated with RNA splicing (**Figure 42h**). Two further predominant networks are associated with the ribosome and RNA transport. Notably, although not surprising for blood/immune cells, a number of RBPs within the T-ALL RNA interactome is involved in "leukocyte activation" and "regulation of hematopoiesis" (significantly enriched within the GO domain *Biological Process;* FDR of 0.0051 and 0.0288, respect-

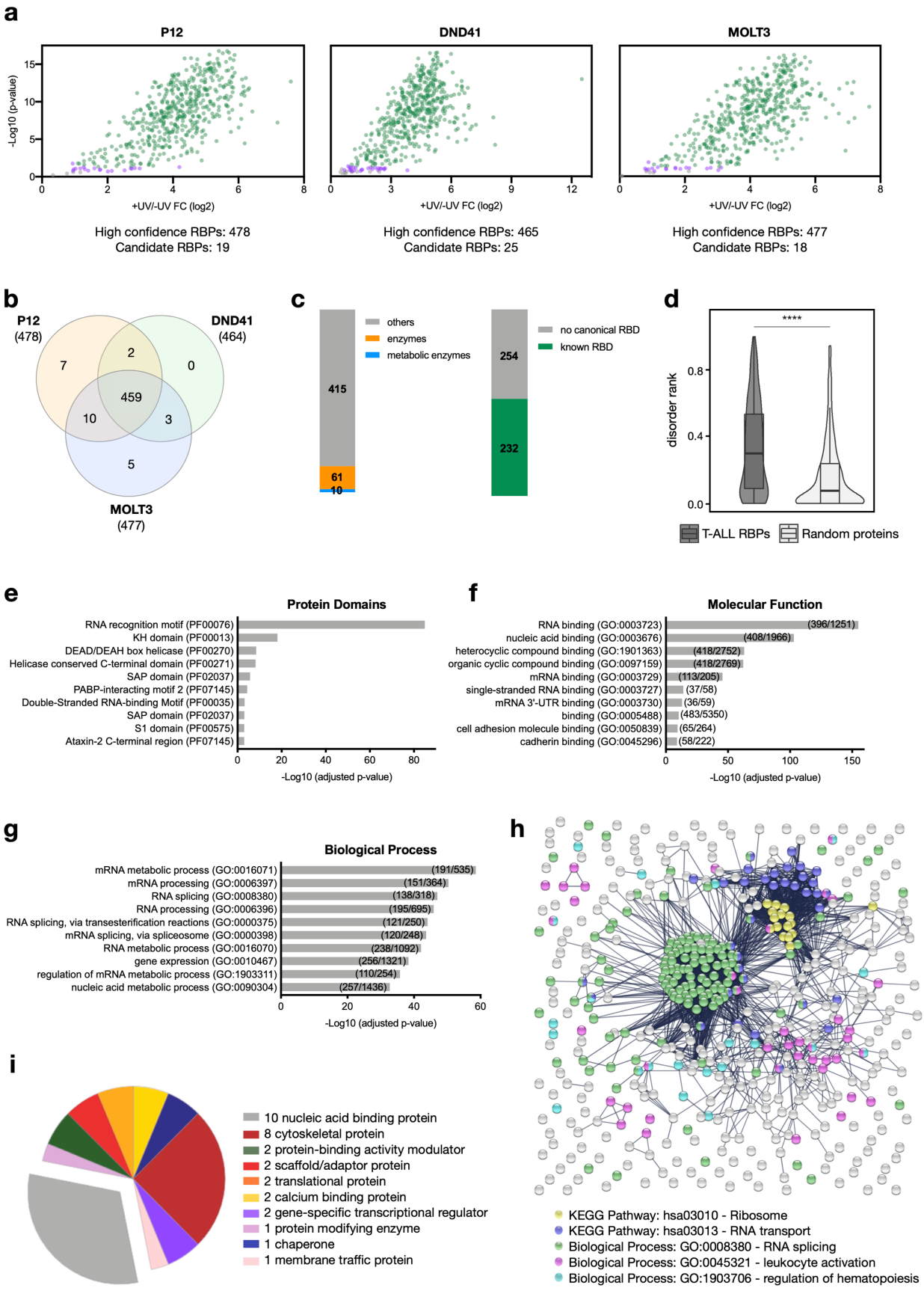


Figure 42: Identification and characterization of the T-ALL RNA interactome. (a) Volcano plots showing the log2-FC ratio of UV-crosslinked (+UV) over non-crosslinked (-UV) (x-axis) versus the *p* value (-log10; y-axis) for each T-ALL cell line. Proteins significantly enriched in the +UV sample compared to the −UV control with a FC > 2 and an FDR < 0.05 are considered as "high confidence RBPs" (shown in green). Proteins with an +UV/-UV enrichment of FC > 1.5 and FDR < 0.2 are considered as "candidates" (depicted in purple). "Unspecific" non-RBPs are depicted in

grey. (b) Venn diagram visualizing the overlap of the RBPs identified in the indicated T-ALL cell lines. Numbers in brackets represent the number of RBPs identified in the respective T-ALL cell line. (c) Number of RBPs in the T-ALL RNA interactome that are known (metabolic) enzymes and that contain known RBDs. (d) Violin plot displaying the disorder rank of the identified 486 T-ALL RBPs relative to an equal number of random proteins (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$). (e) Ten most significantly enriched (adj. p value < 0.05) protein domains within the T-ALL RNA interactome based on Pfam. (f & g) Gene ontology (GO) domain analysis showing the ten most significantly enriched (f) molecular functions and (g) biological processes (with lowest adj. p values; Bonferroni multiple testing corrected). Within brackets: numbers of RBPs identified by eRIC versus the number of proteins identified in the respective full proteome. (h) STRING protein network analysis of the T-ALL RNA interactome. RBPs connected in sub-networks and their associated pathways and functions are displayed as colored dots. (i) Panther protein class annotation of T-ALL RBPs with blood/immune cell relation. Several proteins could not be classified.

tively). Neither of these two processes or any other related ones were found to be signifycantly enriched when subjecting the HeLa RNA interactome (compiled from RBPbase[aa]) to STRING analysis (data not shown). For further examination, the T-ALL RNA interactome was annotated using DAVID[216,217] and manually searched for RBPs associated with terms related to blood/immune cells (**Table 5**).

Table 5: T-ALL RBPs associated with pathways, functions, processes and terms related to blood/immune/T-cells.

Gene symbol	Gene name	Selected pathways, functions, processes and terms related to blood/immune/T-cells
ACIN1*	apoptotic chromatin condensation inducer 1	immune system process, immune system development, regulation of myeloid leukocyte differentiation, hemopoiesis, positive regulation of cell differentiation, hemopoietic or lymphoid organ development, homeostasis of number of cells
ADAR	adenosine deaminase, RNA specific	hematopoietic progenitor cell differentiation, innate immune response, defense response to virus, definitive hemopoiesis, hematopoietic stem cell homeostasis
ANXA1	annexin A1	defense response, inflammatory response, regulation of leukocyte migration, positive regulation of T-helper 1 cell differentiation, negative regulation of T-helper 2 cell differentiation, alpha-beta T cell differentiation
ANXA11	annexin A11	immune system process, immune response
APOBEC3G	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	immune system process, regulation of immune system process, defense response, immune response, regulation of immune effector process
ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	immune system process, immune response
CALR	calreticulin	immune system process, MHC protein complex assembly
CORO1A*	coronin, actin binding protein, 1A	leukocyte homeostasis, regulation of immune system process, regulation of lymphocyte activation, defense response, immune response, regulation of T cell proliferation, T cell homeostasis, regulation of T cell activation
DBNL	drebrin-like	immune system process, immune response
DDX3X	DEAD-box helicase 3, X-linked	response to virus, innate immune response
EZR	ezrin	leukocyte adhesion, Leukocyte transendothelial migration, negative regulation of T cell receptor signaling pathway
FYB	FYN binding protein	immune system process, immune response, T cell receptor signaling pathway

[aa] RBPbase: v0.1.8 alpha; A Comprehensive Database of Eukaryotic RNA-binding Proteins (RBPs) with their RBP Annotations; hosted and maintained by EMBL: Hentze Group; https://rbpbase.shiny.embl.de/). Date of access: 14.07.2020.

GRB2	growth factor receptor-bound protein 2	T cell receptor signaling pathway, IL-2 Receptor Beta Chain in T cell Activation, Chronic myeloid leukemia, Acute myeloid leukemia, leukocyte migration
HCLS1	hematopoietic cell-specific Lyn substrate 1	immune system process, immune system development, hemopoiesis, hemopoietic or lymphoid organ development, negative regulation of leukocyte apoptotic process
HIST1H4A	histone cluster 1, H4a	regulation of immune system process, regulation of myeloid cell differentiation
HSPD1*	heat shock 60kDa protein 1	somatic diversification of immune receptors, activation of immune response, regulation of T cell mediated immunity, regulation of T cell activation, leukocyte proliferation
ILF2	interleukin enhancer binding factor 2	immune system process, immune response
ILF3	interleukin enhancer binding factor 3	defense response to virus, Antiviral defense
LCP1	lymphocyte cytosolic protein 1	immune system process, immune response, lymphocyte activation during immune response, T cell activation
MIA3	melanoma inhibitory activity family, member 3 (also: TANGO1)	regulation of immune system process, regulation of leukocyte migration, negative regulation of lymphocyte migration
MSN	moesin	immune system process, leukocyte adhesion, leukocyte migration
MYH9*	myosin, heavy chain 9, non-muscle	immune system process, immune system development, hemopoiesis, establishment of T cell polarity, leukocyte differentiation, T cell activation, hemopoietic or lymphoid organ development
NKAP	NFKB activating protein	T cell differentiation in thymus, positive regulation of alpha-beta T cell differentiation, hematopoietic stem cell proliferation
NPM1	nucleophosmin 1	Leukemia, acute myeloid, Leukemia, acute promyelocytic, NPM/RARA type
NUMA1	nuclear mitotic apparatus protein 1	Leukemia, acute promyelocytic, NUMA/RARA type
PAK2	p21 protein (Cdc42/Rac)-activated kinase 2	T cell receptor signaling pathway, T cell activation
PCBP2	poly(rC) binding protein 2(PCBP2)	innate immune response, defense response to virus
PIK3R1*	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	immune system process, immune system development, hemopoiesis, T cell receptor signaling pathway, Chronic myeloid leukemia, Acute myeloid leukemia, T cell receptor signaling pathway, lymphocyte differentiation, immune system development, leukocyte differentiation, T cell activation
PRKDC*	DNA-dependent protein kinase catalytic subunit	immune response, hemopoiesis, lymphoid progenitor cell differentiation, T cell lineage commitment, T cell differentiation, T cell differentiation in the thymus, T cell receptor V(D)J recombination, hemopoietic or lymphoid organ development
PURB	purine-rich element binding protein B	regulation of immune system process, regulation of cell differentiation, regulation of myeloid cell differentiation
RANBP1	Ran-specific GTPase-activating protein	Human T-cell leukemia virus 1 infection (HTLV-I infection)
RBM15	RNA binding motif protein 15	regulation of immune system process, regulation of cell differentiation, regulation of myeloid cell differentiation, acute megakaryoblastic leukemia
RC3H1*	ring finger and CCCH-type zinc finger domains 1	regulation of immune system process, regulation of immune response, regulation of T cell activation, regulation of T cell differentiation, regulation of T-helper cell differentiation, regulation of CD4-positive alpha beta T cell differentiation, regulation of T cell activation
RPA1	replication protein A1, 70kDa	immune system process, immune system development, hemopoietic or lymphoid organ development, hemopoiesis

RPL22*	ribosomal protein L22	immune system process, immune system development, hemopoiesis, lymphocyte differentiation, T cell differentiation, T cell activation, alpha-beta T cell differentiation, hemopoietic or lymphoid organ development
RPS14	ribosomal protein S14	immune system process, immune system development, hemopoiesis, hemopoietic or lymphoid organ development
RYDEN	Shiftless antiviral inhibitor of ribosomal frameshifting protein	defense response to virus
SASH3*	SAM and SH3 domain containing 3	regulation of immune system process, regulation of lymphocyte activation, regulation of T cell mediated immunity, regulation of T cell differentiation, regulation of CD4-positive, alpha beta T cell differentiation, regulation of alpha-beta T cell activation, regulation of lymphocyte proliferation
SBDS	Shwachman-Bodian-Diamond syndrome	immune system process, immune system development, hemopoietic or lymphoid organ development
TIAL1	TIA1 cytotoxic granule-associated RNA binding protein-like 1	defense response
TUBB	tubulin, beta pseudogene 2	immune effector process, immune system process, defense response, immune response, lymphocyte mediated immunity, leukocyte mediated immunity
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5	immune system process, immune system development, hemopoiesis, hemopoietic or lymphoid organ development, hemopoietic stem cell differentiation, hemopoiesis, hemopoietic or lymphoid organ development
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6;	immune system process, immune system development, somatic diversification of immune receptors, V(D)J recombination
YTHDF2	YTH domain family, member 2	immune system process, immune response, humoral immune response
YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	immune response, immune effector process, cell activation during immune response, leukocyte mediated immunity, leukocyte degranulation
ZC3HAV1	zinc finger CCCH-type containing, antiviral 1	innate immune response, defense response to virus

* RBPs directly associated with T-cell pathways, functions and/or processes.

In total 46 T-ALL RBPs involved in blood/immune cell specific functions, pathways and processes were identified, nine of them are specifically linked to T-cells. Examples are PRKDC (DNA-dependent protein kinase catalytic subunit), which is involved in T-cell lineage commitment and T-cell differentiation in the thymus, or SASH3 (SAM and SH3 domain-containing protein 3), which is specifically involved in the differentiation and activation of $\alpha\beta$ T-cells. To categorize these blood/immune cell linked RBPs and to further investigate their function, they were analyzed using PANTHER[218] protein class annotation. This analysis showed that for these proteins, functions linked to various processes are reported. As anticipated, the majority of RBPs are classified as nucleic acid binding proteins. Interestingly, cytoskeletal proteins constitute the second most prevalent protein class (**Figure 42i**). These findings likely reflect T-cells undergoing extensive cytoskeleton-dependent processes along with their activation[240,241], such as integrin-

mediated cell adhesion, formation of immunological-synapses, cytoskeletal polarization, receptor sequestration and signaling as well as changes in size and shape in general[240]. Already while in contact with B-cells, T-cells were shown to change their spherical to an elongated and/or flattened shape[241]. Similar observations were found by Liepelt and colleagues[232] when analyzing the macrophage RNA interactome. Macrophages require extensive cytoskeleton remodeling for their chemotactic and phagocytic functions[232,242]. Thus, some RBPs identified here are likely to link immune and cytoskeletal functions. For instance coronin-1A (CORO1A), which is a constituent of the cytoskeleton and reported to contribute to TCR$\alpha\beta$-induced cell signaling[243]. An investigation of the RNA-binding functions and roles of such RBPs specifically linked to features of T-/T-ALL cell differentiation will likely yield interesting results.

To also identify new RBPs specific to this T-ALL RNA interactome, it was compared to previously published RNA interactomes and data sets. Comparisons with the two classical RIC RNA interactomes of the human cell lines HeLa[177] and HEK293[169] (**Figure 43a**), and with the eRIC RNA interactome of the mature T-ALL cell line JURKAT (**Figure 43b**), show a considerable overlap of ~75 % (364 RBPs) and ~63 % (308 RBPs), respectively. An overlap with all previously identified human RBPs (human RNA interactome: 1,803 RBPs in total; Backlund *et al.*[196]) reveals 413 shared RBPs (~85 %), while 73 RBPs are unique to the T-ALL RNA interactome identified here (**Figure 43c**). Functional enrichment analysis using STRING accounts for an over-representation of regulatory proteins amongst these unique RBPs: GO processes such as "regulation of organelle organization", "regulation of cytoskeleton organization" and "regulation of cell cycle" are enriched. This is in line with previous findings and substantiates that several T-ALL RBPs may be involved in the process of cellular re-organization required for T-cell activation. In addition, several of the RBPs unique to the T-ALL RNA interactome were found to be annotated with the GO term "RNA-binding" (**Figure 43d**). Therefore, all 73 RBPs were annotated using DAVID[216,217] and manually categorized into 20, 8 and 45 RBPs which are already known to be "RNA-binding and/or related", "nucleic acid binding and/or related", and so far "not related" at all, respectively (**Figure 43e,** Supplementary table 3). These 45 RBPs so far not known to relate to RNA or nucleic acid binding bear a broad range of protein functions. However, as observed in preceding analyses, a major share of these RBPs constitutes cytoskeletal proteins (**Figure 43f**). To investigate whether some of these new RBPs specific to the T-ALL RNA interactome are associated with blood/immune cell related functions as well, the RBPs of both analyses were compared to each other. Indeed, a total of five proteins constitute new RBPs linked to immune/blood cell processes

(**Figure 43g**): PIK3R1, MIA3, NUMA1, SASH3, RANBP1. (i) *PIK3R1* encodes an isoform of the phosphatidylinositol 3-kinase (PI3K) p85 regulatory subunit[244]. PI3K plays a major role in regulating the cell cycle, proliferation and apoptosis and is upregulated in many tumor types[245]. Published reports show *PIK3R1*/*PTEN*[bb] mutations in ALL[246], gene abnormalities in the PI3K/AKT pathway[cc] in T-ALL[247], and that its inhibition decreases cell proliferation in CLL[248], AML[249,250] and ALL[251]. (ii) The melanoma inhibitory activity protein 3 (MIA3, also TANGO1) constitutes a transmembrane protein localized at the endoplasmic reticulum involved in the transport of bulky cargo[252,253]. According to literature, it is broadly expressed; except in cells of hematopoietic origin[252,254]. Interestingly, MIA3 is identified in all three T-ALL cell lines. Thus, MIA3 expression might be ectopic in T-ALL and is potentially relevant for leukemogenesis. (iii) *NUMA1* encodes the nuclear mitotic apparatus protein (NuMA) a microtubule-binding protein involved in mitotic cell division. It focuses microtubules to the poles of the mitotic spindle and bundles spindle microtubules to centrosomes.[255,256] NuMA was identified, together with its partner gene retinoic acid receptor alpha, to be involved in the development of acute promyelocytic leukemia (APL)[257]. (iv) SAM and SH3 domain-containing protein 3 (SASH3; also SH3-domain containing protein expressed in lymphocytes (SLY/SLY1)) is thought to function as an adapter protein and is expressed exclusively in lymphocytes[258]. In T-cells, it is required as anti-apoptotic protein during thymic development and for signaling through the TCR[259]. Further, SASH3 is a prerequisite for full activation of adaptive immunity[258]. (v) The Ran-specific GTPase-activating protein 1 (RANBP1) plays a role in RAN-dependent nucleocytoplasmic transport, nuclear assembly, spindle assembly and mitotic cell cycle regulation[260,261]. It has been proposed to be involved in the development of adult T-cell leukemia (ATL), which is a T-cell neoplasm developed by up to 5 % of individuals after an HTLV-1 (Human T-cell leukemia virus type 1) infection. Interaction with the viral regulatory protein Tax is thought to disrupt the RanBP1-dependent regulation of centriole cohesion[262,263].

[bb] PTEN: lipid phosphatase; functional antagonist of PI3K[288]

[cc] AKT: serine/threonine kinase; involved in the signal transduction of PI3K[245]

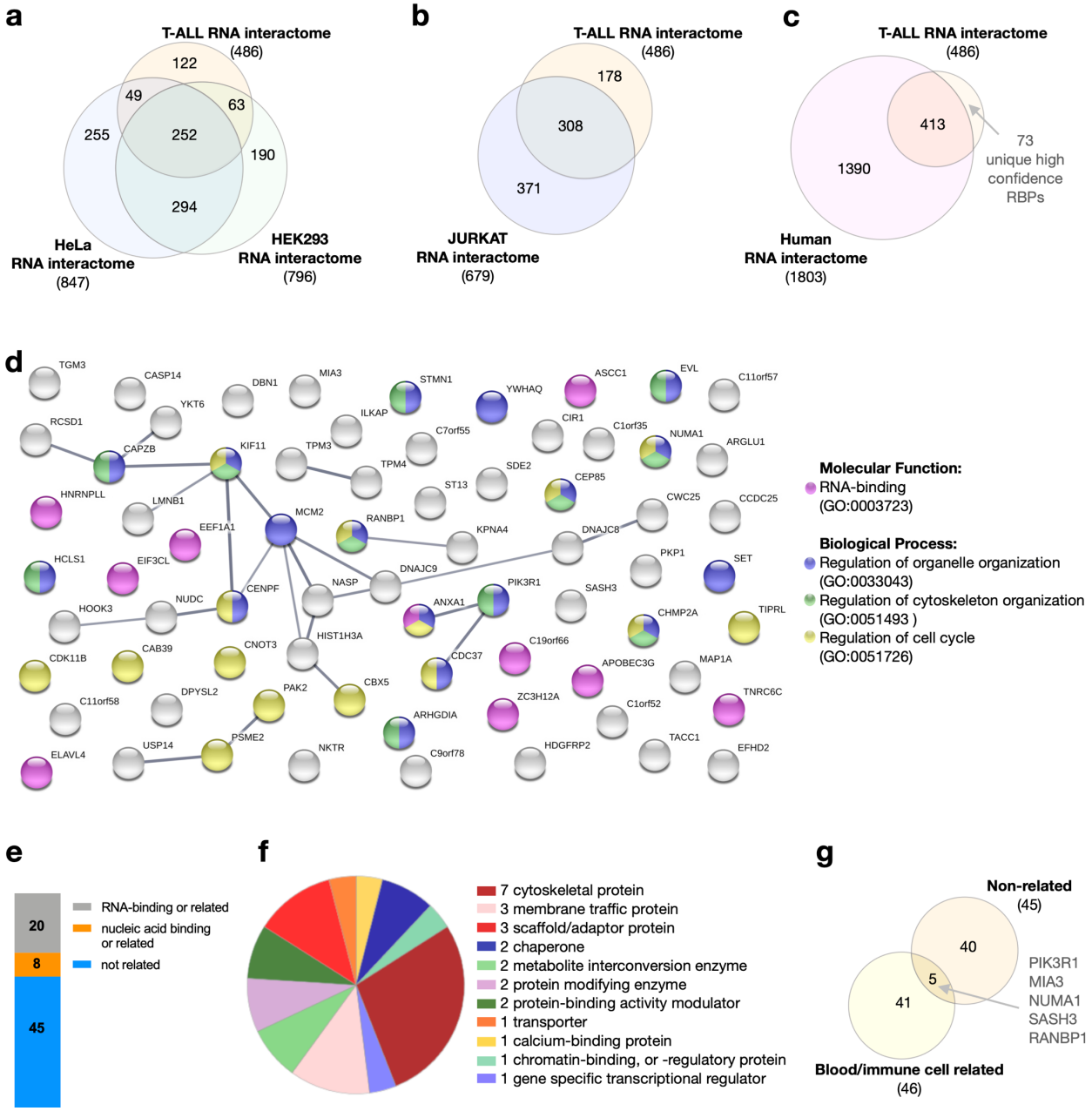


Figure 43: The T-ALL RNA interactome in comparison with other previous RNA interactome captures and RBPs specifically identified in T-ALL. Venn diagrams showing the overlap of the here identified T-ALL RBPs with the RBPs identified (a) by RIC from HeLa[177] and HEK293[169], (b) by eRIC from JURKAT cells[178] and (c) with all previously identified human RBPs (based on the dataset used by Backlund *et al.*[196]). (d) STRING network analysis of RBPs found exclusively in this T-ALL RNA interactome capture. RBPs connected in sub-networks and their associated GO domain processes and functions are displayed as colored dots. (e) Categorization of the RBPs unique to this T-ALL RNA interactome capture into proteins known to be involved in RNA-binding and related, nucleic acid binding and related and into proteins not related at all. (f) Panther protein class annotation of proteins here identified as RBPs without relation to RNA- or nucleic acid binding. Several proteins could not be classified. (g) Venn diagram comparing the RBPs without prior relation to RNA- or nucleic acid binding to the RBPs linked to blood/immune cell related functions (**Table 5**).

In summary, this T-ALL RNA interactome strongly reflects features of its cellular origin, the T-cell. Especially naive and memory T-cells show low metabolic activity and undergo extensive cytoskeleton rearrangements to become activated[240,241,264,265]. Accordingly, the T-ALL RNA interactome harbors profoundly lower numbers of (metabolic) enzymes compared to the RNA interactome of metabolically active cells[196] and several analyses reveal a notably high number of cytoskeletal proteins. Further, immune/blood cell associated GO terms are overrepresented and several identified RBPs are involved in pathways and processes linked to T-cell development, activation and function, and thus, underline the biological validity of the data set generated here. Five of the new RBPs were found amongst the proteins linked to immune/T-cells and provide interesting starting points for follow-up studies investigating whether their unknown RNA-binding functions contribute to T-ALL initiation and maintenance. These studies will need to involve an extensive validation of the RNA-binding activity which may be achieved by polynucleotide kinases (PNK) assay or CLIP experiments. Further characterization may then include RBDmap to identify the sites for RNA-binding and RNAi or CRISP/Cas9 approaches to analyze to role of the RNA-binding function by knockdown studies. Finally, the translation of these findings into primary T-ALL cells will need to be assessed. In addition, the T-ALL RNA interactome here identified shows features typical for RBPs and previously identified RNA interactomes such as a high relative disorder rank, classical RBDs, enrichment of RNA-binding related GO terms and RNA splicing as a major protein network, thus, verifying a specific and reliable determination of RBPs.

Identification of SAHA-responsive RBPs

Next, the effect of SAHA-induced HDAC inhibition on the T-ALL RNA interactome was analyzed. A number of 468, 466 and 472 high confidence RBPs (FC > 2 and FDR < 0.05) were identified in SAHA-treated samples of P12, DND41 and MOLT3, respectively (**Figure 44a**, Supplementary table 4). A comparison of these RBPs to those identified in untreated samples (see previous section; **Figure 42a**) reveals a large overlap for all three T-ALL cell lines. In total, 16 unique high confidence RBPs from all samples show a response to SAHA (**Figure 44b**). Here, it should be noted that RBPs specific to one condition (untreated or SAHA-treated) were identified in the other corresponding condition as well but considered as "high confidence" RBPs only in one of them due to differential enrichment (i.e. +UV/-UV ratio less than two-fold or FDR ≥ 0.05).

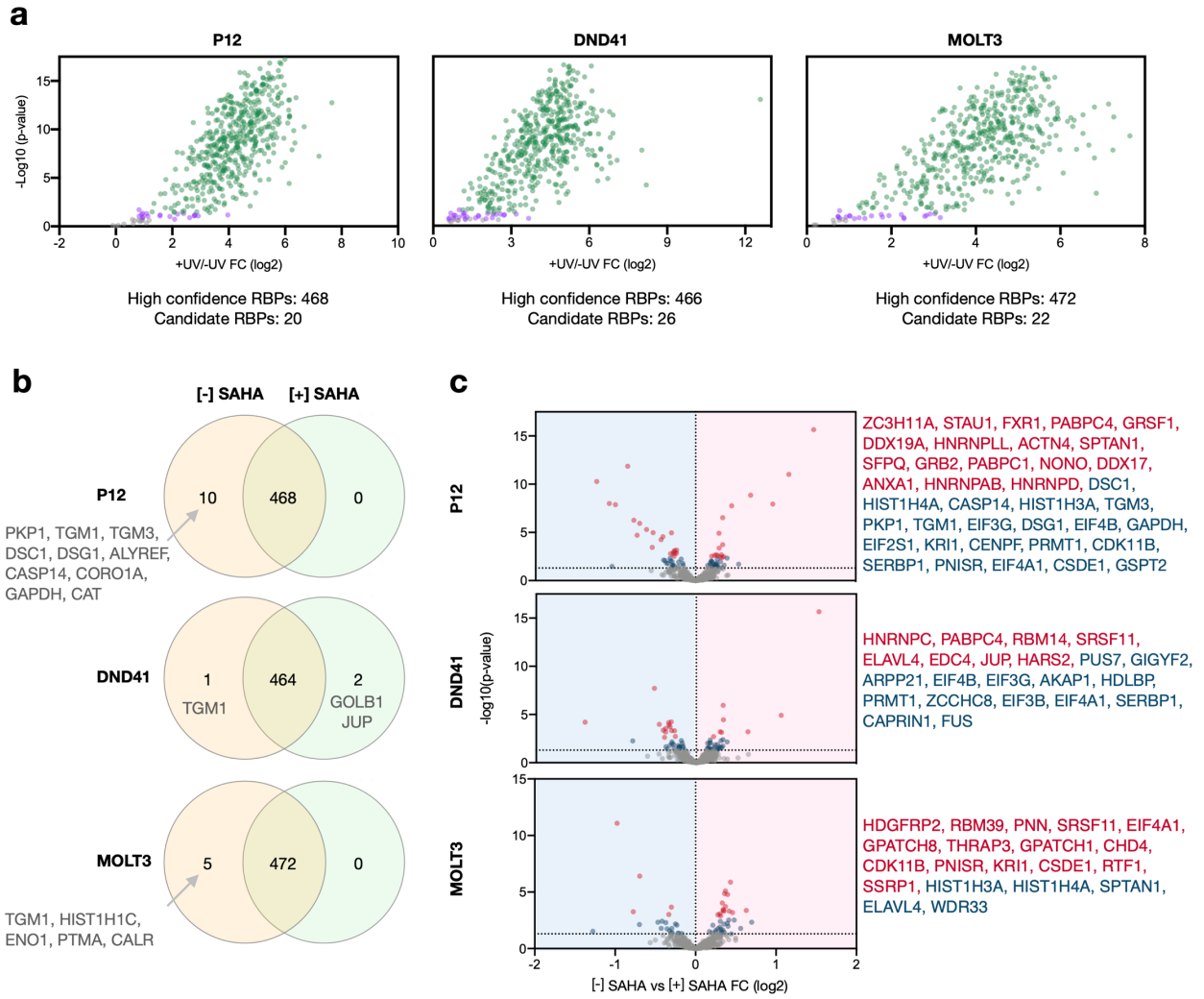


Figure 44: Identification of changes in the RBP composition and activity of T-ALL cells upon HDAC inhibition. T-ALL cell lines (P12, DND41 and MOLT3) were either grown untreated ([-] SAHA; DMSO vehicle-only) or SAHA-treated ([+] SAHA; 5 µM SAHA for 6 h) and analyzed by comparative eRIC. (a) Volcano plots showing the log2-FC ratios of UV-crosslinked (+UV) over non-crosslinked (-UV) SAHA-treated samples (x-axis) versus the respective *p* values (-log10; y-axis) for each T-ALL cell line. Proteins significantly enriched in the +UV sample compared to the −UV control with a FC > 2 and an FDR < 0.05 are considered as "high confidence RBPs" (shown in green). Proteins with an +UV/-UV enrichment of FC > 1.5 and FDR < 0.2 are considered as "candidates" (depicted in purple). "Unspecific" non-RBPs are depicted in grey. (b) Venn diagrams overlapping high confidence RBPs identified from untreated and SAHA-treated samples of each T-ALL cell line. (c) Volcano plots visualizing SAHA-induced effects of all proteins identified from UV-crosslinked conditions for each T-ALL cell line. Differential abundance analysis between SAHA-treated versus untreated samples was tested: x-axes depict the log2-FC ratio of SAHA-treated versus untreated samples of UV-crosslinked condition, y-axes depict the respective *p* values of the comparisons. Proteins showing SAHA-induced fold changes with an FDR < 0.05 are depicted in red; "SAHA-responsive" proteins. Proteins with *p* values < 0.05 but an FDR > 0.05 are depicted in blue. Horizontal dashed line represents the *p* value = 0.05 threshold. Right panel shows gene symbols of SAHA-responsive proteins (FDR < 0.05): upregulated proteins are shown in red, downregulated proteins in blue.

To investigate the response of the T-ALL RNA interactome in more detail, all proteins identified in UV-crosslinked eRIC eluates of untreated and SAHA-treated samples were tested for differential abundance. This analysis yielded 38, 24 and 20 hits (FDR < 0.05), here referred to as "SAHA-responsive RBPs", in P12, DND41 and MOLT3, respectively. In each cell line several of these RBPs show an enhanced RNA-binding activity (red), while others show a decreased (blue)

RNA-binding activity. For MOLT3, however, considerably more of the SAHA-responsive RBPs show an increase in RNA-binding activity (**Figure 45c**). Integrating the hits of all three cell lines, a total of 64 unique SAHA-responsive RBPs are identified. 14 SAHA-responsive RBPs are common to at least two cell lines, amongst them three eukaryotic initiation factors (eIFs). One of these eIFs, EIF4A1, is the only RBP responding to SAHA in all three T-ALL cell lines (**Figure 45a**). A GO enrichment analysis reveals that RNA-binding and translation-related functions and processes are significantly overrepresented amongst the SAHA-responsive RBPs. In addition, processes associated with skin development are enriched as well (**Figure 45b**), which may hypothetically be related to the activation of skin development in wound healing by the immune system in general and by cutaneous T-cells in particular[266,267]. Moreover, the activation and deactivation of genes for cellular proliferation and differentiation along with this process involves epigenetic and posttranslational modifications[267]. String network analysis reflects these findings. The three major protein networks are associated with mRNA splicing, translation initiation complex formation and keratinization. Furthermore, several proteins are annotated as known RNA-binders and, in line with their identification here as SAHA-responsive RBP, most of the proteins are already linked to acetylation, thus validating the biological relevance of these findings (**Figure 45c**). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is amongst these proteins linked to acetylation. As an RBP, GAPDH is involved in the regulation of mRNA stability and translation[268] and, as an enzyme, it is essential for glycolysis[269]. In the latter function, GAPDH is known to be subject to reversible acetylation by the lysine acetyltransferase PCAF[dd] and the deacetylase HDAC5; upon acetylation at lysine 254 its activity in response to glucose in known to increase and to promote cell proliferation and cancerogenesis[269]. A closer assessment of quantitative differences in RNA-binding reveals that SAHA-responsive RBPs overall tend to show opposing RNA-binding activity in P12 (and DND41) cells compared to MOLT3 cells: i.e. increased RNA-binding activity in P12 but decreased in MOLT3 and vice versa. Illustrative examples are SPTAN1, SERBP1 and most eIFs, such as EIF4A1 and EIF3G (**Figure 45d**). Further, the majority of the SAHA-responsive RBPs are identified in the full proteome as well but the abundance of most of them is not or only marginally altered upon SAHA-treatment. Interestingly, all SAHA-responsive RBPs annotated by STRING with keratinization are identified in the interactomes only but in none of the respective full proteomes. Hence, these proteins may

[dd] PCAF: P300/CBP-associated factor or K(lysine) acetyltransferase 2B (KAT2B)

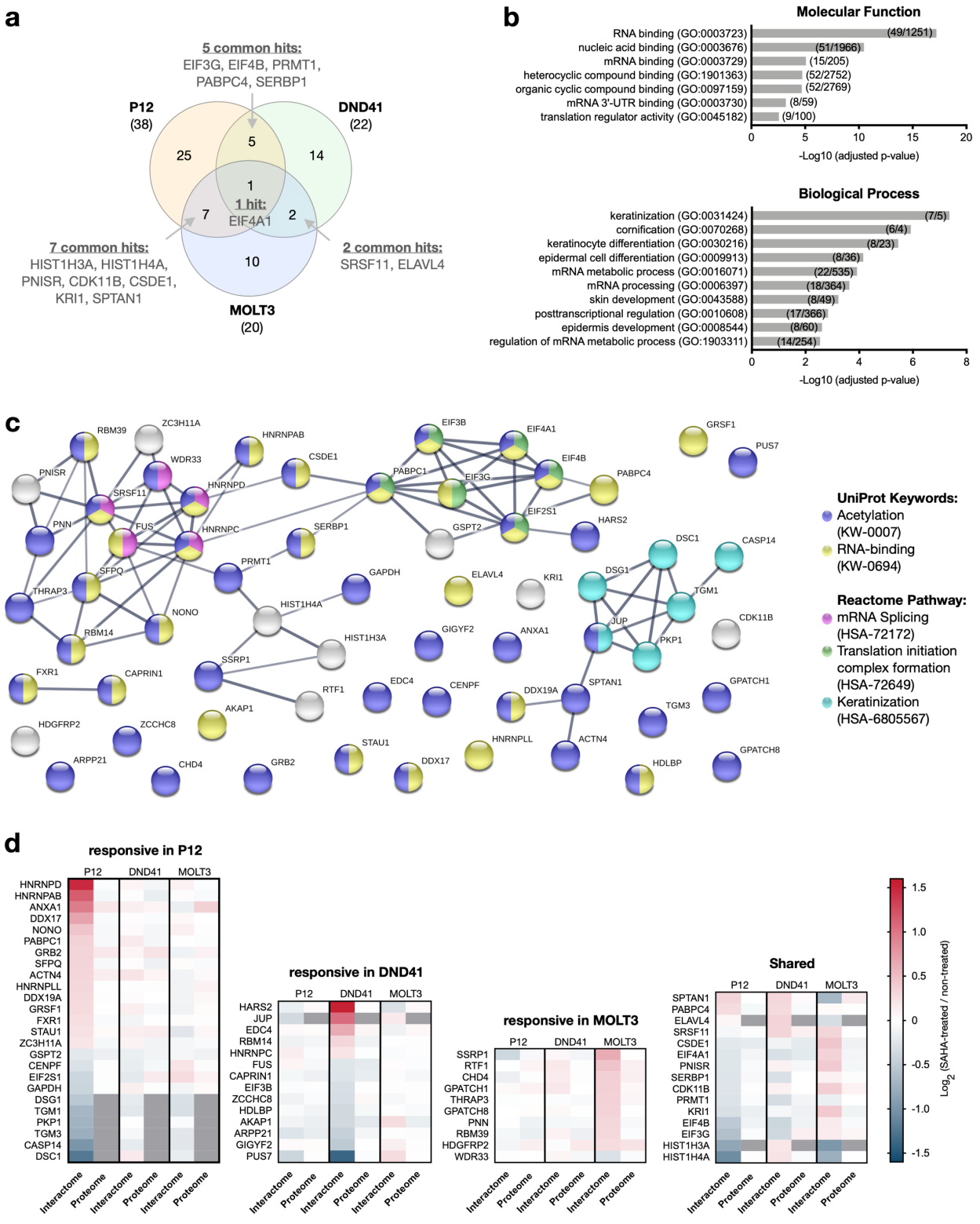


Figure 45: Characterization of SAHA-responsive proteins and their RNA-binding activity. (a) Venn diagram showing the overlap of SAHA-responsive proteins identified from each of the three T-ALL cell lines. (b) GO-domain analysis showing the leading significantly (adj. p value < 0.05) enriched *biological processes* and *molecular functions* (lowest adj. p values; Bonferroni multiple testing corrected). Within brackets: numbers of RBPs identified by eRIC versus the number of proteins identified in the respective full proteome. (c) STRING protein network analysis of SAHA-responsive RBPs. RBPs connected in sub-networks and their associated keywords and pathways are displayed as colored dots. (d) Heat maps depicting the alterations in RNA-binding activity upon SAHA-treatment of responsive (FDR < 0.05) RBPs (interactome) and the respective alterations in the abundance (proteome). The color-scale on the right shows the log2-FC ratios of SAHA-treated versus untreated UV-crosslinked samples. Not identified proteins are depicted in grey.

be of low abundance and have been missed by the analysis of the full proteome due to its high complexity compared to the smaller RNA interactome. Hypothetically, the abundance and RNA-binding activity of these proteins could relate to the activation of wound healing upon injury.

Overall, several proteins in the eRIC eluates clearly displayed differential RNA-binding activity upon SAHA-treatment. Most of these SAHA-responders are known RNA-binders and proteins already linked to acetylation, for example eIFs. In addition, some of these proteins have also been linked to T-ALL before. For instance, the RNA helicase EIF4A1 was shown to be involved in promoting T-ALL development and maintenance *in vivo*[270] and the serine/arginine-rich splicing factor SRSF11 is member of a protein family involved in exon skipping, which was found to happen more frequently in T-ALL cells compared to normal T-cells[271]. Hypothetically, the therapeutic effect of SAHA may thus be related to altered splicing patterns, to modulated translation efficiency or to the metabolic functions exerted by some of these proteins. This interpretation is in line with published reports showing that the functions of eIFs are controlled by reversible posttranslational modifications, for instance, binding of the eukaryotic initiation factor eIF2 is regulated by phosphorylation[272] and the subcellular localization of eIF5A is controlled by acetylation[273]. Since EIF4A1 is identified as a SAHA-responsive RBP in all three cell lines and was previously found to be acetylated in HeLa cells by RBDmap[128], a potential regulatory function between acetylation and RNA-binding activity for this and other eIFs is hypothesized. Shedding light on the mechanism underlying this particular correlation between HDAC inhibition and RNA-binding activity may expand our knowledge on RBP-regulation and may add to the mechanistic understanding of the antileukemic effect of SAHA.

4.5.2 Identification of the RNA-bound proteome of normal primary thymocytes

T-ALL cells are thought to be derived from malignant thymocytes that can arise at different stages during normal intra-thymic T-cell development[46,54,55]. However, the majority of T-ALLs originate from a cortical stage of T-cell ontogeny and therefore, this thesis project focused on cortical T-ALL[55]. To study normal T-cell progenitors that are most similar to cortical T-ALL, a protocol to efficiently isolate and UV-crosslink sufficient numbers of cortical thymocytes was established in section 4.1. Ultimately, for eRIC analyses, thymocytes of three different, healthy individuals were collected and either UV-crosslinked or left unexposed as background controls. Lysates were equally split and subjected to two independent replicates of eRIC analysis (**Figure 46**). Whole cell lysates of thymocytes of each individual were analyzed in parallel.

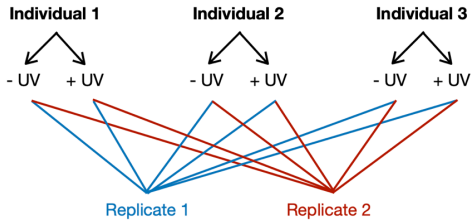


Figure 46: Schematic overview of the experimental set-up for analyzing normal, cortical thymocytes by eRIC. Thymocytes of three healthy individuals were isolated and irradiated with UV light (+UV) to crosslink RBPs to their target RNA. Equal amounts of thymocytes were left non-irradiated to serve as background controls (-UV). Lysates of each individual and condition were equally split and two independent replicates of eRIC analyses were performed.

Obtaining sufficient normal cortical T-cell progenitors (thymocytes)

In a first step, sufficient numbers of thymocytes with a high proportion of CD7 (T-cell lineage marker) and CD1a (cortical maturation stage marker) positives were collected. Thymi of different individuals were disaggregated, cleaned by repeated centrifugation and discarding the viscous part of the cell pellets, before the UV-crosslinking the thymocytes in four to five batches each (see section 4.1.4). Aliquots of the cell suspensions used for UV-crosslinking were stained and analyzed by FACS to determine, if the purity suffices for eRIC. Thymocyte lysates of the three thymi depicted in **Figure 47** were selected for the final RNA interactome captures. All three thymi yielded > 80 % viable, CD7 and CD1a positive target cells considered as cortical T-cell progenitors. Cell lysate of each individual and condition (+UV and -UV each) were pooled, yielding sufficient material for two replicates of eRIC.

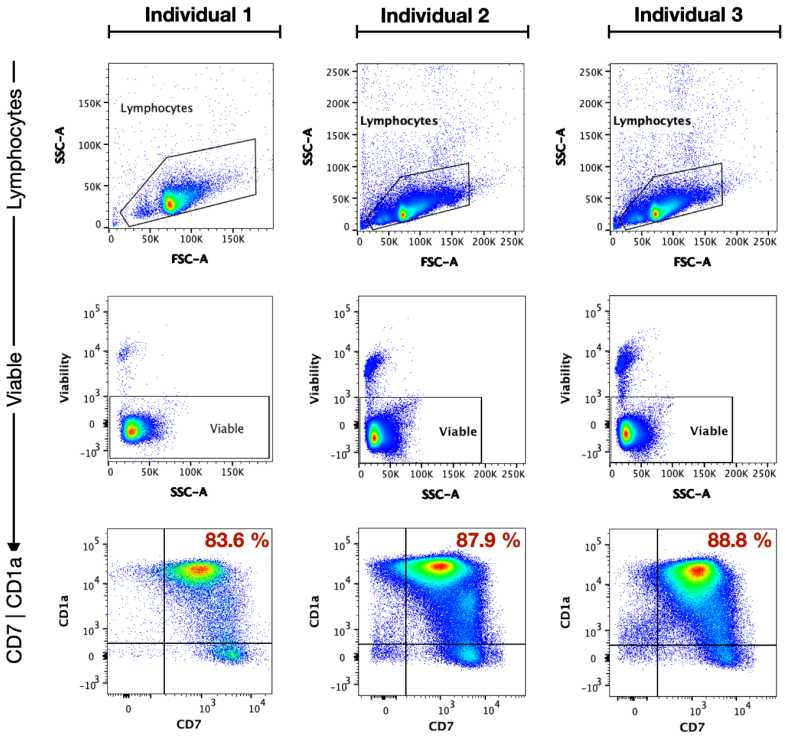


Figure 47: Evaluation of the percentage of CD1a positive target cells in total thymi using flow cytometry. Freshly disaggregated thymocytes from three individuals were stained with CD7 (APC), AQUA viability dye and CD1a (PE) before fixation with paraformaldehyde and analysis by flow cytometry. Gates were selected based on the well-known

populations of total thymus and by using FMO controls. All samples were analyzed by gating for lymphocytes and selection of single cells (not shown). CD7 and CD1a positives were sorted from viable subpopulations. Total frequencies of the viable CD7 and CD1a target population are depicted in red.

eRIC of normal thymocytes and corresponding quality control analyses

As before, eRIC performance and quality was evaluated on protein and RNA level using aliquots from the crude lysates (input) and from eluates of UV-crosslinked and non-crosslinked conditions. Representative results are depicted in **Figure 48**. SDS-PAGE and silver staining (**Figure 48a**) show distinctive patterns for all +UV conditions (lanes 5, 10, 15) profoundly different from the complex band patterns of the inputs (lanes 2, 3, 7, 8, 12, 13). All -UV conditions (lanes 4, 9, 14) are clean, exempting the RNase bands (lane 17) and smears of residual proteins. These results suggest selective protein enrichment by eRIC. Although the +UV eluates show similar band patterns, the intensity amongst the individual samples differs. Despite the equal protein amounts used for eRIC and the same percentages of eluate applied to SDS-PAGE, the +UV eluate pattern observed from individual 1 (lane 5) shows the highest intensity, while the band pattern of the other two individuals (lanes 10 and 15) are notably weaker, with individual 3 (lane 15) being the weakest. These differences indicate differential enrichment efficiency between the individuals. Further, the specificity of eRIC was evaluated by western blot (**Figure 48b**). +UV eluates are expected to include known RBPs (UnR, HnRNP K, HuR) while cellular abundant non-RBPs (α-tubulin, β-actin, histones H3 and H4) should be depleted and exclusively present in the inputs. As expected, HuR and HnRNP K are present in the +UV eluates of all three individuals (lanes 4, 8, 12); although HnRNP K to a lesser extent. UnR, however, is only present in samples of individual 1 (lane 4) and not detectable in inputs (lanes 5 & 6, 9 & 10) and +UV eluates (lanes 8 and 12) of the individuals 2 and 3. As expected, the non-RBPs α-tubulin, β-actin, histones H3 and H4 are only detectable in the inputs. These findings demonstrate specific enrichment of RBPs but also confirm suspected differential enrichment between the individuals. Regarding UnR, the results are in line with the data of previous eRIC experiments using thymocytes (chapter 4.4): UnR appears to be detectable only in some individuals and may be subject to differential expression or depend on certain stimuli. Next, the profiles of the eluted nucleic acids were assessed by bioanalyzer (**Figure 48c**). Samples of all three individuals show a single peak encompassing ~200 – 3000 nucleotides, a size range attributed polyadenylated mRNA. Signal corresponding to 18S and 28S rRNA, usually observed in inputs or RIC eluates (discussed in section 4.2.3; Figure 29c), is not detected. Thus, bioanalyzer data validate an enrichment of mRNA and a successful depletion of rRNA contaminations.

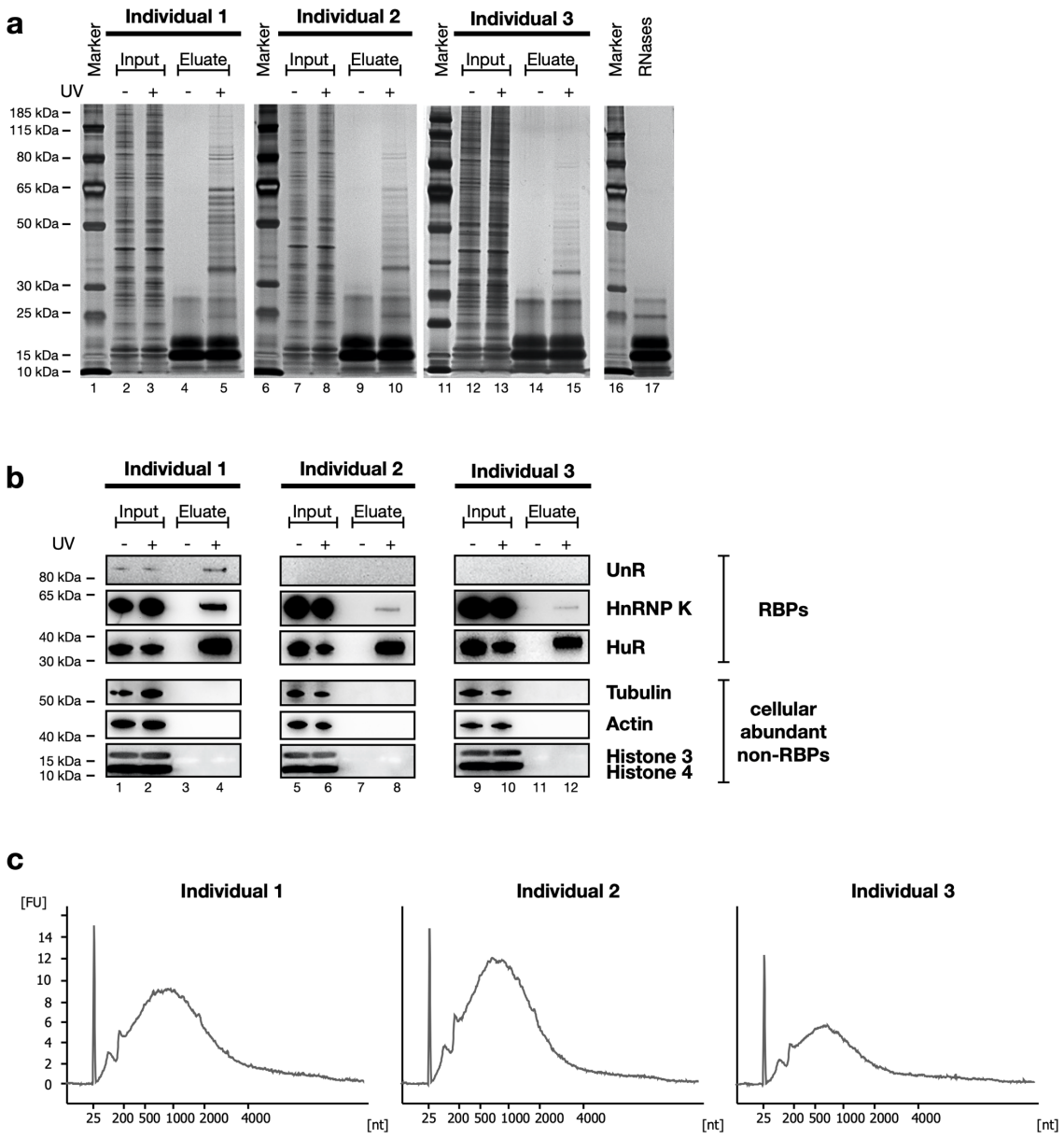


Figure 48: Quality controls monitoring the RBP isolation of eRIC applied to thymocytes on protein and RNA level. Whole cell lysates (input) of crosslinked (+UV) and non-crosslinked (-UV) thymocytes of three healthy individuals were subjected to eRIC. RBPs crosslinked to their target RNAs are captured on magnetic beads linked to LNA-based oligonucleotides and isolated by stringent washes and elution (eluate) at either at 90 °C (RNA quality controls) or by RNases (protein quality controls, proteomic analysis). To evaluate the quality of the RBP isolation, proteins from inputs and eluates were analyzed by SDS-PAGE combined with (a) silver staining and (b) western blotting against known RBPs (UnR, HnRNP K, HuR) as positive control and abundant non-RBPs (α-tubulin, β-actin, histone H3 and H4) as negative controls. (c) Profiles of enriched nucleic acids were evaluated by bioanalyzer. Two replicates of RNA interactome capture and associated quality controls were performed, representative results are depicted.

In sum, these quality controls validate that eRIC selectively and specifically enriched RBPs against a background of abundant cellular proteins. Although differential enrichment efficiency between the individuals was observed, the data of all three individuals showed enrichment of proteins in +UV conditions over the -UV controls.

Proteomic identification of the normal thymocyte RNA interactome

Next, the proteins isolated by eRIC were labeled by 6plex TMT and analyzed by LC-MS/MS in two independent replicate experiments (see section 3.4.7 for details). Corresponding full proteomes were analyzed in parallel. The mass spectrometry data were cleaned for batch effects, normalized and only proteins quantified with at least two unique peptides were further considered. A boxplot visualizing the acquired signal sums is depicted in **Figure 49** and reveals equal distributions for full proteomes. For eRIC eluates, the signal sum medians of +UV conditions are in generally higher than those of the corresponding -UV controls, reflecting protein enrichment. The only exception are the eluates of individual 3 in replicate 2. Here, in contrast, the median signal sum of the +UV eluate is lower than the median signal sum of the respective -UV eluate. In fact, the median signal sum of the +UV eluate of individual 3 from replicate 2 is even lower than the one from all other -UV eluates. After a careful analysis of the proteins detected in these eRIC eluates, a mix-up of conditions or labels was excluded (not shown). Taken together with the relatively low enrichment efficiency observed with samples of individual 3 in preceding quality controls (see **Figure 48**), these results raised doubts concerning the protein enrichment and the reliability of data from individual 3.

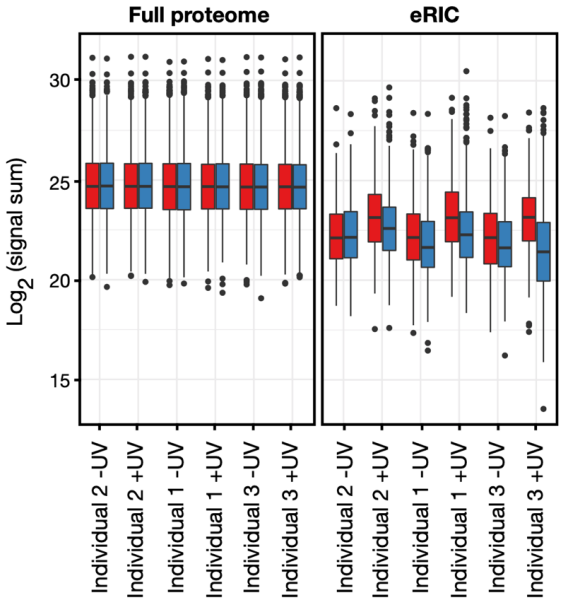


Figure 49: Mass spectrometry data acquisition from thymocyte full proteomes and eRIC eluates. Whole cell lysates (full proteome) and irradiated (+UV) and non-irradiated (-UV) eRIC eluates of thymocytes from three individuals were labeled by 6plex TMT and analyzed by LC-MS/MS in two independent mass spectrometry replicates. Data were cleaned of batch effects, normalized and only proteins quantified with at least two unique peptides were considered. Boxplot shows the signal sum distribution (y-axis, log2 scale) of proteins identified in the respective conditions (x-axis). Replicate 1 is depicted in red, replicate 2 is depicted in blue.

Therefore, all samples obtained from individual 3 were excluded from further analysis. Mass spectrometry data acquired from individuals 1 and 2 were normalized anew, and again only proteins quantified with at least two unique peptides were further considered. **Figure 50** shows an overview of this renewed proteomic analysis comprising only of the individuals 1 and 2. The signal sums of the full proteomes are equally distributed. As expected from previous analyses, the median signal sums of the +UV conditions are globally higher than the ones of the corresponding -UV conditions, confirming protein enrichment (**Figure 50a**). These finding are reflected by principal component analysis of eRIC eluates, which clusters all +UV eluates and all -UV eluates respectively in close proximity (**Figure 50b**). In total, a number of 4287 (replicate 1) and 4429 (replicate 2) proteins was identified from full proteomes. From eRIC eluates, 296 (replicate 1) and 755 (replicate 2) proteins were identified (**Figure 50c**). Based on the pilot RNA interactome captures and the lower enrichment obtained with thymocytes compared to the T-ALL cell lines, 296 identified proteins match the anticipated range. A number of 755 identified proteins, however, far exceeded the expectation. Neither the preceding quality controls, the MS quality control files (data not shown) nor the signal sum distributions are indicative of such a profound difference in the number of proteins identified from both replicates.

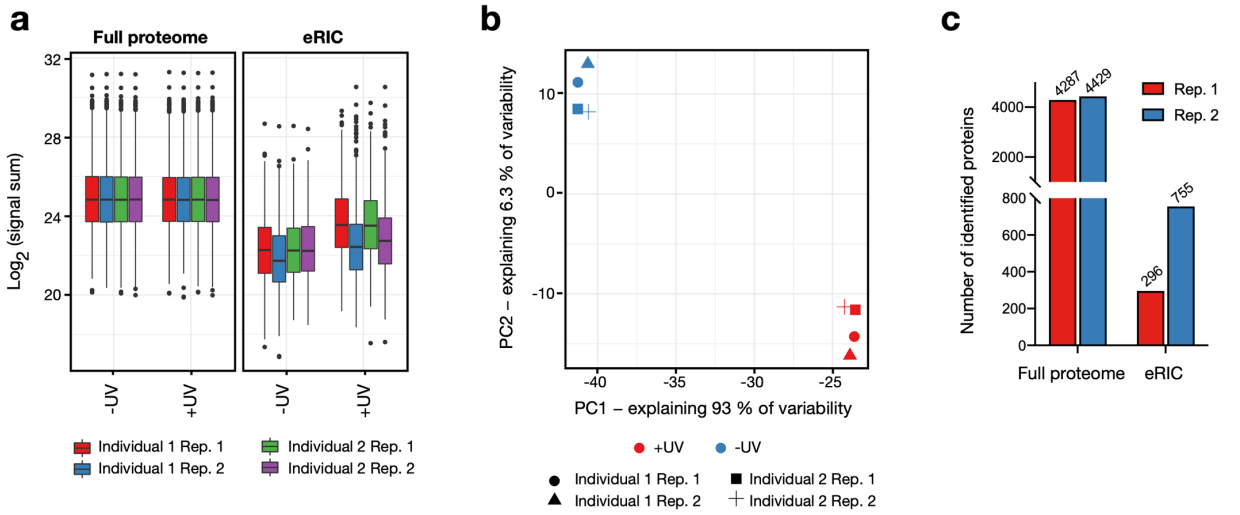


Figure 50: Proteomic analysis for the identification of the thymocyte RNA interactome. Mass spectrometry data acquired from two independent LC-MS/MS replicates of whole cell lysates (full proteome) and irradiated (+UV) as well as non-irradiated (-UV) eRIC eluates of thymocytes from two individuals were analyzed. Data were cleaned of batch effects, normalized and only proteins quantified with at least two unique peptides were considered. (a) Boxplot shows the signal sum distribution (y-axis, log2 scale) of proteins identified in the respective conditions (x-axis). (b) Principal component analysis of the analyzed eRIC conditions. Each shape represents one individual and replicate. Red color corresponds to +UV condition, blue color corresponds to -UV conditions. (c) Overview of the number of proteins identified in the full proteome and in the RNA interactome capture (eRIC) in each replicate.

Comprehensive analysis of the RNA-bound proteome in thymocytes

Given that the amendments in previous section reduced the number of individuals for further consideration to two only and that the number of proteins identified from both replicates are very different, the analysis was adapted accordingly. Originally, the data were obtained from samples of both individuals analyzed together in two mass spectrometry runs. For the analysis, however, each individual in each run was considered as one replicate of an "average" thymus. Thus, the number of replicates within the analysis amounted to four (**Figure 51a)**. Although compromising the depth of the RNA interactome, only proteins identified from all four replicates were further considered to ensure high data reliability. Applying the usual criteria for "hit" identification (FC > 2 for UV-crosslinked over non-crosslinked condition and an FDR < 0.05), a number of 120 high confidence RBPs was identified, representing the "thymocyte RNA interactome" identified here (**Figure 51b**, Supplementary table 5). Analogously to preceding chapter 4.5.1, the thymocyte RNA interactome was analyzed for typical features. This revealed that ~9 % (11) of the identified thymocyte RBPs constitute enzymes (**Figure 51c)** and that ~23 % (28) of them lack canonical RBDs (**Figure 51d)**. Nevertheless, an analysis for intrinsically disordered regions, a typical feature for non-canonical RBPs in particular[131,274], shows a significantly higher rank of disorder for the thymocyte RBPs identified here relative to a random set of proteins of similar size (**Figure 51e**). Amongst the RBPs that harbor classical RBDs, RRMs are highly overrepresented, followed by an enrichment of further typical RBDs such as KH and DEAD box helicase domains (**Figure 51f**). A GO-domain analysis reveals that RNA splicing and RNA processing related *Biological Processes* and RNA-binding linked *Molecular Functions* lead the lists of significantly enriched terms (**Figure 51g**). Consistently, STRING network analysis yields a major network of RBPs involved in RNA splicing and a smaller network linked to translation initiation factor activity. Illustrative examples are several heterogeneous nuclear ribonucleoproteins (hnRNPs), such as hnRNP K and HnRNP D, and eukaryotic initiation factors (eIFs) such as eIF4A1 and eIF4B (**Figure 51h**). Thus, this thymocyte RNA interactome exhibits typical features of classical RBPs and previous RNA interactomes, accounting for a reliable identification of RBPs. For the T-ALL RNA interactome, STRING analysis revealed an enrichment of a number of RBPs annotated with GO-terms related to blood/immune cells. Similarly, the thymocyte RNA interactome shows an enrichment of RBPs linked to "defense response and "positive regulation of innate immune response" (significantly

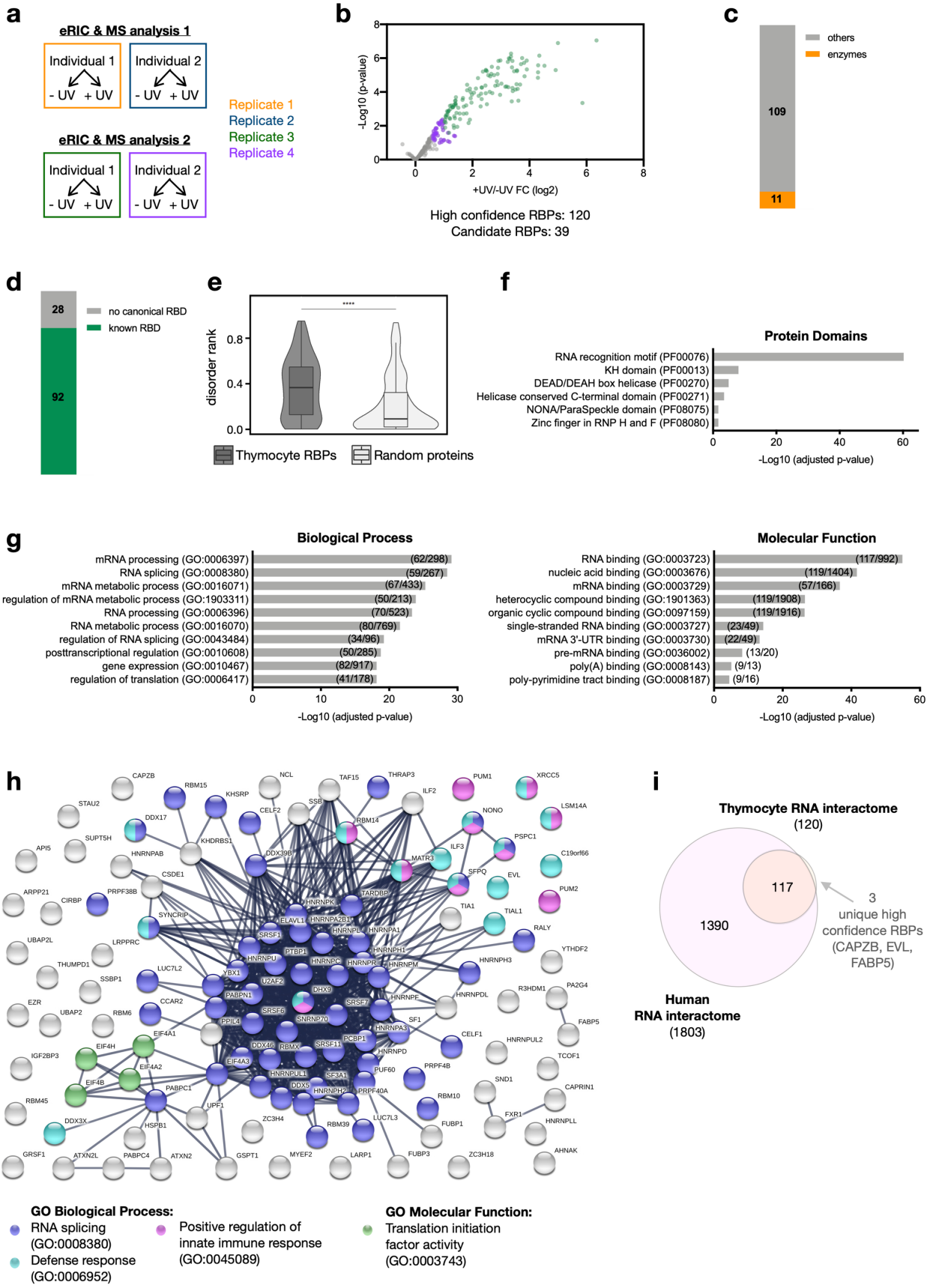


Figure 51: Identification and characterization of the thymocyte RNA interactome. (a) Schematic overview of the experimental set-up and conditions ultimately considered to identify the thymocyte RNA-binding proteome. (b) Volcano plots depicting the log2-FC ratio of UV-crosslinked (+UV) over non-crosslinked (−UV) eluates (x-axis) and the *p* values (-log10; y-axis) of the proteins identified by RNA interactome capture. Proteins significantly enriched in the +UV sample compared to the -UV control with a FC > 2 and an FDR < 0.05 classified as "hits", and were considered as

high confidence RBPs (depicted in green). Proteins with an +UV/−UV enrichment of FC > 1.5 and FDR < 0.2 were considered as "candidates" (purple). "Unspecific" non-RBPs are depicted in grey. (c) Number of thymocyte RBPs that are known enzymes. (d) Number of thymocyte RBPs that contain known RBDs. (e) Violin plot displaying the disorder rank of the identified 120 thymocyte RBPs relative to an equal number of random proteins (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$). (f) Graphical representation of protein domains significantly enriched (adj. p value < 0.05) within the thymocyte RNA interactome based on Pfam. (g) Gene ontology domain analysis of *Biological Process* (left) and *Molecular Function* (right). Ten most significantly enriched processes and functions, i.e. with lowest adj. p values (Bonferroni multiple testing corrected) are graphically represented. In brackets: numbers of RBPs identified by eRIC versus the number of proteins identified in the respective full proteome. (h) STRING protein network analysis of the thymocyte RNA interactome. RBPs connected in sub-networks and their associated GO domain processes and functions are displayed as colored dots. (i) Venn diagram showing the overlap of the thymocyte RBPs with previously known human RBPs.

enriched within the GO domain *Biological Process;* FDR of 0.0417 and 6.01×10^{-5}, respectively (**Figure 51h**). An annotation of the thymocyte RNA interactome by DAVID[216,217] and screening for RBPs associated with pathways, processes and functions related to blood/immune cells yielded 9 RBPs (XRCC5, YTHDF2, ILF2, ILF3, RBM15, DDX3X, TIAL1, EZR, RYDEN) already previously identified within the T-ALL RNA interactome. Hence, the thymocyte RNA interactome reflects features typical for its cellular origin.

For an identification of RBPs specific to the thymocyte RNA interactome, it was compared to previously identified human RBPs (human RNA interactome: 1,803 RBPs in total; Backlund *et al.*[196]). Almost all thymocyte RBPs (97.5 %) have been identified as human RBPs before and only three RBPs (EVL, CAPZB, FABP5) are unique to the thymocyte RNA interactome (**Figure 51i**). (i) The Ena-VASP-like protein (EVL) is one of three members of the Ena/VASP (vasodilator-stimulated phosphoprotein) family which regulate the recruitment of monomeric actin to the barbed end of the actin filament and block its capping. In normal T-cells, EVL acts as cytoskeletal effector protein and is required for diapedesis and the trafficking of activated T-cells.[275] In T-ALL, the highly expressed EVL gene was found to be involved in gene fusions resulting in the overexpression of the partner gene[247]. (ii) CAPZB constitutes the β subunit of the F-actin-capping protein CAPZ, which is a regulator of cytoskeletal organization known to cap the growing end of actin filaments[276]. Literature on CAPZB reports a role in regulating cell differentiation, migration and morphology[276,277]. Previous studies showed differential expression of CAPZB in leukemia: Dehgan-Nayeri and colleagues found decreased CAPZB expression in ALL samples compared to normal samples and Chiaretti and collegues found particularly high levels of CAPZB expression in ALL at relapse. Hence, CAPZB expression was suggested as a potential biomarker for the dexamethasone-effect on ALL.[278,279] Interestingly, in the T-ALL RNA interactome described before CAPZB was identified as unique RBP as well. (iii) The third protein identified here as a unique RBP in the thymocyte RNA interactome is the fatty acid-binding

protein 5 (FABP5). FABP5 is a lipid-binding protein involved in the uptake and intracellular transport of long-chain fatty acids. In general, FABPs are implicated in regulating gene expression in the context of cell growth and differentiation.[280–282] The validation of the RNA-binding function of these proteins and the determination of their role in T-ALL will be an interesting goal of follow-up studies and may further our knowledge of the mechanisms involved in the malignant transformation of thymocytes.

In sum, these findings describe the discovery of the first human thymocyte RNA interactome. Compared to the corresponding T-ALL RNA interactome and RNA interactomes of other cell types[169,175,177,196], the number of identified RBPs was small. This may be in part attributed to the depth of the enrichment of the analysis but it is also likely to reflect the nature of the cellular source. A previous study already predicted a similar number of ~150 RBPs to be expressed in normal human thymus[283]. In addition, previous human RNA interactome data are all obtained from cell lines with cancerous origin, the thymocyte RNA interactome identified here, however, is the first one obtained from normal human primary tissue. Cancerous cells are known to have undergone metabolic reprogramming during oncogenesis[284], to show gene dysregulation[285] and further abnormalities resulting from genomic instability[286] such as gene fusions and resulting oncogenes. Thus, normal thymocytes are highly likely to exhibit lower metabolic activity, reduced gene expression and thus lower and/or different RNA-binding activity compared to leukemic T-cells. As a consequence, such "quiescent" or low RNA-binding activity proteins might simply escape a capture and their detection as RBPs. Intriguingly, some of the features of the thymocyte RNA interactome resemble those of the nuclear RNA interactome identified by eRIC of HuH7 cells upon nuclear-cytoplasmic fractionation[196]: relatively small, few new RBPs and non-canonical RBDs. These similarities are in line with the small size and the high nuclear-cytoplasmic ratio of approximately 3:1 – 4:1 of T-cells and their progenitors[226]. In addition, and in analogy to the T-ALL RNA interactome, the thymocyte RNA interactome comprises several RBPs associated with pathways, function and processes linked to blood/immune cells and is enriched for related processes. As such, the thymocyte RNA interactome likewise reflects the nature of its cellular source and thus further underlines the biological validity of the identified thymocyte RBPs.

4.5.3 Comparative analysis of RBP-composition identified from cortical T-ALL and normal T-cell progenitors

Several studies report close links between normal T-cell development and leukemogenesis[15,46,54]. To further the understanding of T-ALL leukemogenesis, the identification of differences in the RBP compositions between T-ALL cells and normal thymocytes was part of this project. Therefore, the data obtained from all T-ALL cell lines and from the two technically validated thymi were reanalyzed. This analysis facilitates an approximation of "average T-ALL" versus "average thymocytes". An overlap of all high confidence RBPs from both cell types shows that almost all of the thymocyte RBPs are also identified in T-ALL. However, three RBPs are unique to the thymocyte RNA interactome (**Figure 52a**). Two of them, AHNAK (neuroblast differentiation-associated protein AHNAK) and HSPB1 (heat shock protein beta-1) are known RNA-binders. FABP5 (fatty acid-binding protein 5) was identified as a new RBP in the previous section. This analysis only allows the comparison of proteins showing significant RNA-binding activity exclusively in either one of the two cell types. Therefore, the analysis was extended to take all proteins identified in both RNA interactomes into account, to compare for differential RNA-binding activity (UV-crosslinked over non-crosslinked ratio), and to consider protein abundance in the full proteome (**Figure 52b**). Overall, this analysis reveals an enrichment correlation (Person's correlation test: t-test, correlation coefficient of 0.71, *p*-value of 2.2×10^{-16}) shifted towards T-ALL. As already observed previously (chapter 4.4), protein enrichment in general is decreased in thymocytes compared to T-ALL cell lines. In line with this finding, none of the proteins that show pronounced abundance in the T-ALL full proteome (red shape) shows more RNA-binding activity in thymocytes (red shapes on blue background). Nevertheless, several proteins are similarly or noticeably more abundant in the thymocyte full proteome (white to purple shapes) and show recognizably more apparent RNA-binding in T-ALL (white to purple shapes on red background, proteins with noticeably pronounced abundance in the thymocyte full proteome are labelled in black). Amongst these RBPs are heterogeneous nuclear ribonucleoproteins (hnRNPs) such as HnRNP C, HnRNP K and especially HnRNP A1 and HnRNP A2B1. It is noteworthy that several hnRNPs were detected with notably high signal intensities in the thymocyte RNA interactome (data not shown). hnRNPs comprise a family of well-known RNA-binders with diverse roles in RNA metabolism including RNA splicing, stabilization and translational regulation[287]. Amongst the ones only identified as RBPs in T-ALL are members of the histone H1 and H4 family (HIST1HD, HIST1H1E, HIST1H4A), and proteins

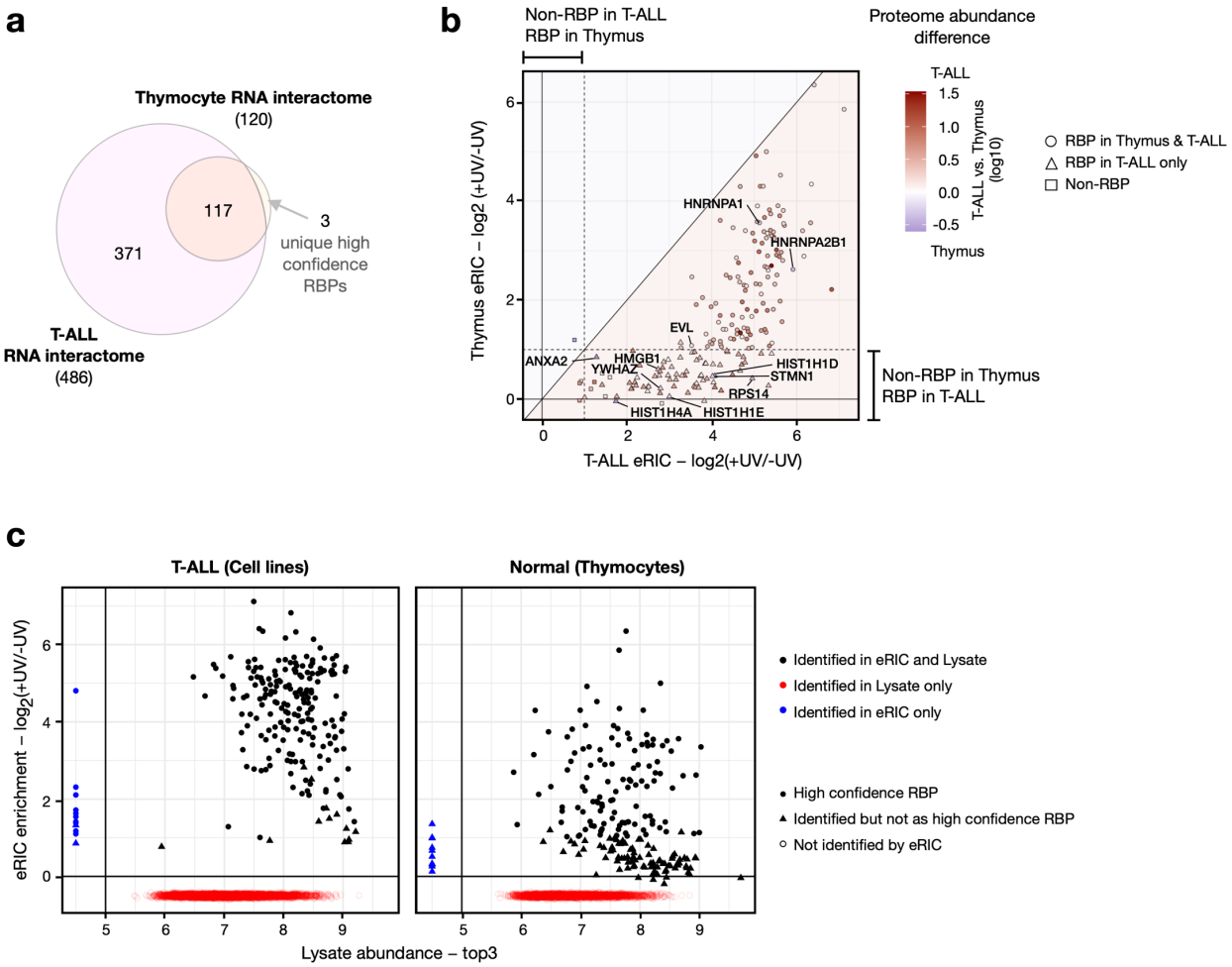


Figure 52: Comparative analysis of the RNA interactomes identified in normal thymocytes and T-ALL cell lines. (a) Venn diagram showing the overlap of the RBPs identified in the T-ALL RNA interactome and the thymocyte RNA interactome. Numbers in brackets represent the number of RBPs identified in each RNA interactome. (b) Correlation of the RNA binding activity (reflected by the UV-crosslinked over non-crosslinked fold change ratio (log2)) between T-ALL (x-axis) and thymocyte (y-axis) RBPs. For simplicity, only proteins identified in both RNA interactomes are shown. Shapes indicate the identification of proteins a non-RBPs (square), as RBP in T-ALL only (triangle) or as RBP in both, thymus and T-ALL (circle). The color-scale on the right reflects the abundance difference in the full proteomes between T-ALL and thymocyte samples (log10). Red or purple shapes indicate proteins with higher abundance in T-ALL or thymus, respectively. Shapes on red or purple background show higher RNA-binding activity in the T-ALL RNA interactome or in the thymocyte RNA interactome, respectively. Dashed lines represent the 2-fold enrichment threshold of +UV/-UV ratios for an identification as high confidence RBP in the respective RNA interactome. Proteins below both dashed lines are identified in both RNA interactomes but do not show significant RNA-binding activity in either. Proteins above both dashed lines are high confidence RBP in both RNA interactomes. Proteins identified in both RNA interactomes that show higher RNA-binding activity in T-ALL although similar or more abundant in thymocytes are labelled in black. (c) Correlation between eRIC enrichment and lysate abundance. eRIC enrichment (y-axis) is reflected by the UV-crosslinked over non-crosslinked fold change ratio (log2) of the identified proteins. Lysate abundance (x-axis) is reflected by the top3 values (based on MS1 signal intensity) of the respective proteins in the full proteome. Shapes indicate the type of protein identification by eRIC: not identified (circle) or identified but not as high confidence RBPs (triangle) or identified as high confidence RBPs (dot). Color of the shapes indicate the condition in which the proteins are identified: lysate (full proteome) and/or eRIC. For simplicity, only proteins identified in both cellular sources by eRIC and lysate respectively are shown.

linked to blood/immune cells before (**Table 5**), like XRCC5 (X-ray repair cross-complementing protein 5) and YWHAZ (14-3-3 protein zeta/delta) and CORO1A (Coronin-1A). Prompted by the observed shift in the enrichment correlation, the differential eRIC protein enrichment with T-ALL cells and thymocytes was assessed in more detail and relative to the respective full proteome (**Figure 52b**). Whereas eRIC with T-ALL cell lines shows predominantly strong enrichment of already highly abundant proteins, eRIC enrichment in thymocytes yields proteins ranging from low to high abundance. Further, the analysis reveals that the enrichment with eRIC using thymocytes is in general lower than the enrichment obtained with T-ALL cell lines. In line with these findings, more proteins identified in thymocytes are not considered as high confidence RBPs (triangular shapes). As such, this analysis confirms system-specific differences in the eRIC enrichment using T-ALL cell lines and thymocytes as already indicated by earlier analyses (chapter 4.4), which has to be taken into consideration for direct comparisons of RBPs from different cellular systems.

In sum, this comparative analysis showed that the considerably larger T-ALL RNA interactome comprises almost all RBPs identified in thymocytes and confirms earlier assumptions that the RBP enrichment by eRIC with thymocytes differs from the one obtained from T-ALL cell lines. Although the RNA-binding activity appears to be generally lower in thymocytes than in T-ALL, this analysis further revealed that some of the proteins with higher abundance in the thymocyte full proteome are highly enriched in T-ALL RNA interactome, and thus, indicated potential differences in their RNA-binding activity between normal and disease state. Validating the apparent differential RNA-binding activity of these RBPs in T-ALL versus thymocytes by other experimental means may further our understanding of oncogenic pathways underlying T-ALL.

As described in the previous section, it is highly likely that primary cells, compared to their corresponding cancerous cells, harbor RNA-binding proteomes with lower activity. eRIC approaches with other primary cells (osteoblast cells, T-cells from peripheral blood) in our group revealed similar low-complexity RNA interactome in non-malignant cells. The possibility to align the RBP enrichment of normal primary tissues and their cancerous pendants might facilitate more direct comparisons of T-ALL and thymocyte RBPs in future projects. The use of whole tissue slices rather than isolated primary cells may improve the methodology to allow more efficient crosslinking and a deeper analysis of the thymocyte RNA interactome and thus may facilitate such an alignment.

5 SUMMARY & CONCLUSIONS

The overarching goal of this project was to uncover RNA-binding proteins (RBPs) that may be involved in the development of T-cell acute lymphoblastic leukemia (T-ALL) and may be exploited for developing novel treatment strategies. Since T-ALL originates from immature thymic T-cell precursors this project included a comparison of the RNA interactomes of these cell types. When compared to lines of leukemic and other malignant cells, normal human thymic T-cell precursors are characterized by their small size, high nucleo-cytoplasmic ratio and low protein content[226]. Further, these primary cells cannot be expanded in culture and thus generating high confidence RNA-interactomes of these cells represented a substantial technical challenge. Therefore, the first phase of this work focused on developing the technology to enable the reliable and reproducible generation of RNA interactomes of primary human thymic T-cell precursors and cultured T-ALL cells. This challenge was met by adapting and optimizing the novel locked nucleic acid (LNA)-based "enhanced RNA interactome capture" (eRIC) technology for the use with these cells. Compared to the original RNA interactome capture method using oligo(dT) magnetic beads, eRIC allowed to prepare RNA interactomes with increased sensitivity and specificity. Finally, the first high-confidence RNA interactomes from uncultured human primary cells, thymocytes, could be generated and compared to those derived from T-ALL cells; the direct malignant counterparts to human thymocytes.

It has been an unexpected finding that the RNA interactome obtained from primary thymocytes is profoundly smaller than that the RNA interactomes of corresponding T-ALL cells, and indeed, smaller than all other reported RNA interactomes obtained from cultured malignant cells. This observation may likely be explained by finding a globally reduced abundance of RNA binding proteins in these primary normal cells, thus resulting in an overall lower RNA-binding activity in thymocytes. Hence, a number of thymocyte RBPs may not be sufficiently abundant or active for efficient crosslinking and detection under the same experimental conditions applied to rapidly growing and metabolically active cell lines. The technology may thus require further modifications to the UV-crosslinking step facilitating eRIC of larger quantities of source material to efficiently capture RBP subsets with low abundance or low RNA-affinity. The use of whole tissue slices rather than isolated and suspended cells of the organ may represent such an improvement of the methodology to further facilitate crosslinking and RBP capture.

In addressing the first aim of this project, an ultimate number of 120 and 486 high confidence RBP were identified from primary thymocytes and T-ALL cell lines, respectively. These RBPs provide a snapshot of the RBP compositions of normal human T-cell progenitors and T-ALL, thus unveiling important previously unknown biological differences between these leukemic cells and the normal parent cell type. The RNA interactomes of both, normal precursors and T-ALL cells, present typical characteristics of RBPs and RNA interactome, such as an increased proportion of disordered proteins, the presence of classical RNA-binding domains (RBD) and an enrichment of RNA-binding related gene ontology (GO) terms and protein networks. Further, both interactomes reflect the T-cell identity by harboring many proteins related to immune or T-cell development, activation and function, a remarkably high number of cytoskeletal proteins serving functions in TCR-induced cell signaling and T-cell activation, and a small number of RBPs with metabolic activity. When compared to the normal thymic precursors, the RNA interactomes of the T-ALL cells were characterized by differential binding of several RBPs. This class of RBPs is exemplified by CORO1A, a constituent of the cytoskeleton involved in T-cell signaling and proliferation[243], which was reproducibly more abundant in the thymocyte total proteomes but only showed RNA binding in T-ALL. Finally, the RNA interactomes revealed three previously unknown RBPs with potentially particularly relevant biological functions. (i) SASH3 is a signaling adaptor protein known to be expressed exclusively in lymphocytes[258] and required for thymic development[259]. However, it has been unknown that the SASH3 function may depend on its role as an RBP. (ii) MIA3 is a transmembrane protein here identified as a high confidence RBP in all three T-ALL cell lines, although this protein has previously been reported not to be expressed in cells of hematopoietic origin[252,254]. At this point, it must remain speculative which biological function this RBP may serve in the process of leukemogenesis, although its highly reproducible and T-ALL specific identification in the RNA interactome analysis strongly suggests an important role. (iii) CAPZB is the β subunit of the F-actin-capping protein CAPZ and was found in the RNA interactomes of both, normal T-cell precursors and T-ALL cells. Previous studies reported lower CAPZB expression in ALL samples compared to normal bone marrow samples and particularly high expression levels in B-ALL at relapse[278,279]. It will be an interesting goal of further studies to characterize the functional implications of these differences and possibly any perspectives towards developing novel treatment strategies, which may be based on these differences.

The second aim of this project was based on the previous observation that the RBDs of many RBPs contain amino acid residues that are expected to be subject of post-translational modification in general and acetylation in particular. Thus, the effect of modulating acetylation on the RNA interactome of T-ALLs was analyzed by treatment with suberoylanilide hydroxamic acid (SAHA; Vorinostat), a broad-spectrum inhibitor of histone, or more generally lysine deacetylases (HDAC/KDAC). Following intensive optimization of drug doses and treatment kinetics to minimize secondary effects of the influence of SAHA on de-novo protein synthesis and cell viability, the T-ALL RNA interactomes in response to HDAC inhibitor treatment were assessed. This comparative analysis revealed the differential binding of several SAHA-responsive RBPs. These included proteins already annotated with the term acetylation and known to exhibit activities depending on acetylation, such as GAPDH[269], thus validating the experimental approach. In addition, the comparative assessment revealed previously unknown SAHA-responsive proteins with potentially relevant links to T-ALL, as exemplified by EIF4A1 and SRSF11. (i) EIF4A1 is a RNA helicase required for translation initiation and known to play a role in promoting T-ALL development and maintenance *in vivo*[270]. (ii) SRSF11 is a member of the serine/arginine-rich splicing factor family, which is involved in exon skipping, a phenomena recently found to be more frequent in T-ALL cells compared to normal T-cells[271]. Thus, the therapeutic effect of SAHA in T-ALL may potentially be related to a modulation of the splicing pattern or the translation efficiency.

Taken together, this work represents the first comprehensive analysis of the RNA-binding proteomes of T-ALL and primary human thymocytes, and offers unprecedented insights into alterations in the RBP-landscape upon leukemogenesis. My results extend the understanding of human RNA interactomes by unveiling unexpected features RNA-protein interactions in primary normal human cells and provide new testable hypotheses regarding the functional role of T-ALL specific RNA-binding proteins and their relevance for the development of new treatment options.

BIBLIOGRAPHY

1. Jagannathan-Bogdan, M. & Zon, L. I. Hematopoiesis. *Development* **140**, 2463–2467 (2013).
2. Staal, F. J. T., Wiekmeijer, A. S., Brugman, M. H. & Pike-Overzet, K. The functional relationship between hematopoietic stem cells and developing T lymphocytes. *Ann. N. Y. Acad. Sci.* **1370**, 36–44 (2016).
3. Roth-Walter, F., Jensen-Jarolim, E. & Stockinger, H. Common Concepts of Immune Defense. in *Principles and Comparative Aspects of Adaptive Immunity* 243–266 (2013).
4. Murphy, K. & Weaver, C. *Janeway's immunobiology*. (Garland Science, Taylor & Francis Group, LLC, 2017).
5. King, K. Y. & Goodell, M. A. Inflammatory modulation of HSCs : viewing the HSC as a foundation for the immune response. **11**, 685–692 (2011).
6. Rothenberg, E. V. & Taghon, T. Molecular Genetics of T cell Development. *Annu. Rev. Immunol.* **23**, 601–649 (2005).
7. Savino, W., Mendes-Da-Cruz, D. A., Lepletier, A. & Dardenne, M. Hormonal control of T-cell development in health and disease. *Nat. Rev. Endocrinol.* **12**, 77–89 (2016).
8. Petrie, H. T. & Zúñiga-Pflücker, J. C. Zoned Out: Functional Mapping of Stromal Signaling Microenvironments in the Thymus. *Annu. Rev. Immunol.* **25**, 649–679 (2007).
9. Lind, E. F., Prockop, S. E., Porritt, H. E. & Petrie, H. T. Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development. *J. Exp. Med.* **194**, 127–134 (2001).
10. Scimone, M. L., Aifantis, I., Apostolou, I., Von Boehmer, H. & Von Andrian, U. H. A multistep adhesion cascade for lymphoid progenitor cell homing to the thymus. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7006–7011 (2006).
11. Dzhagalov, I. & Phee, H. How to find your way through the thymus: a practical guide for aspiring T cells. *Cell. Mol. Life Sci. Vol.* **69**, 663–682 (2012).
12. Bousso, P. & Robey, E. A. Dynamic Behavior of T Cells and Thymocytes in Lymphoid Organs as Revealed by Two-Photon Microscopy. *Immunity* **21**, 349–355 (2004).
13. Staal, F. J. T., Weerkamp, F., Langerak, A. W., Hendriks, R. W. & Clevers, H. Transcriptional control of t lymphocyte differentiation. *Stem Cells* **19**, 165–179 (2001).
14. Dik, W. A. *et al*. New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201**, 1715–1723 (2005).
15. Graux, C., Cools, J., Michaux, L., Vandenberghe, P. & Hagemeijer, A. Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia* **20**, 1496–1510 (2006).
16. Pike-Overzet, K., van der Burg, M., Wagemaker, G., van Dongen, J. J. M. & Staal, F. J. T. New insights and unresolved issues regarding insertional mutagenesis in X-linked SCID gene therapy. *Mol. Ther.* **15**, 1910–1916 (2007).
17. Burger, R., Hansen-Hagge, T. E., Drexler, H. G. & Gramatzki, M. Heterogeneity of T-acute lymphoblastic leukemia (T-ALL) cell lines: suggestion for classification by immunophenotype and T-cell receptor studies. *Leuk. Res.* **23**, 19–27 (1999).
18. Haynes, B. F., Denning, S. M., Le, P. T. & Singer, K. H. Human intrathymic T cell differentiation. *Semin.*

Immunol. **2**, 67–77 (1990).

19. Spits, H., Lanier, L. L. & Phillips, J. H. Development of human T and natural killer cells. *Blood* **85**, 2654–70 (1995).
20. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: Cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
21. Bettenhausen, B., Hrabe de Angelis, M., Simon, D., Guenet, J. L. & Gossler, A. Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. *Development* **121**, 2407–2418 (1995).
22. Dunwoodie, S. L., Henrique, D., Harrison, S. M. & Beddington, R. S. P. Mouse Dll3: A novel divergent Delta gene which may complement the function of other Delta homologues during early pattern formation in the mouse embryo. *Development* **124**, 3065–3076 (1997).
23. Lindsell, C. E., Shawber, C. J., Boulter, J. & Weinmaster, G. Jagged: A mammalian ligand that activates notch1. *Cell* **80**, 909–917 (1995).
24. Shawber, C., Boulter, J., Lindsell, C. E. & Weinmaster, G. Jagged2: A Serrate-like gene expressed during rat embryogenesis. *Dev. Biol.* **180**, 370–376 (1996).
25. Shutter, J. R. *et al.* D114, a novel Notch ligand expressed in arterial endothelium. *Genes Dev.* **14**, 1313–1318 (2000).
26. Aster, J. C., Pear, W. S. & Blacklow, S. C. Notch Signaling in Leukemia. *Annu. Rev. Pathol. Mech. Dis.* **3**, 587–613 (2008).
27. Seo, W. & Taniuchi, I. Transcriptional regulation of early T-cell development in the thymus. *Eur. J. Immunol.* **46**, 531–538 (2016).
28. Osborne, B. & Miele, L. Notch and the Immune System. *Immunity* **11**, 653–663 (1999).
29. Pui, J. C. *et al.* Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11**, 299–308 (1999).
30. Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547–558 (1999).
31. Rothenberg, E. V, Moore, J. E. & Yui, M. A. Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* **8**, 9–21 (2008).
32. Van Vlierberghe, P. & Ferrando, A. The molecular basis of T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **122**, 3398–3406 (2012).
33. Rothenberg, E. V. Transcriptional drivers of the T-cell lineage program. *Current Opinion in Immunology* **24**, 132–138 (2012).
34. Actor, J. K. T Lymphocytes: Ringleaders of Adaptive Immune Function. in *Introductory Immunology* (ed. Actor, J. K. B. T.-I. I.) 42–58 (Academic Press, 2014).
35. Von Boehmer, H. *et al.* Pleiotropic changes controlled by the pre-T-cell receptor. *Curr. Opin. Immunol.* **11**, 135–142 (1999).
36. McMahan, C. J. & Fink, P. J. RAG reexpression and DNA recombination at T cell receptor loci in peripheral CD4+ T cells. *Immunity* **9**, 637–647 (1998).
37. Yannoutsos, N. *et al.* The role of recombination activating gene (RAG) reinduction in thymocyte development

in vivo. *J. Exp. Med.* **194**, 471–480 (2001).

38. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: What thymocytes see (and don't see). *Nature Reviews Immunology* **14**, 377–391 (2014).
39. Kisielow, P., Teh, H. S., Blüthmann, H. & von Boehmer, H. Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. *Nature* **335**, 730–733 (1988).
40. von Boehmer, H. *et al.* The Expression of CD4 and CD8 Accessory Molecules on Mature T Cells is not Random but Correlates with the Specificity of the αβ Receptor for Antigen. *Immunol. Rev.* **109**, 143–152 (1989).
41. Singer, A., Adoro, S. & Park, J.-H. Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. *Nat. Rev. Immunol.* **8**, 788–801 (2008).
42. Vignali, D. A. A., Collison, L. W. & Workman, C. J. How regulatory T cells work. *Nature Reviews Immunology* **8**, 523–532 (2008).
43. Durinck, K. *et al.* Novel biological insights in T-cell acute lymphoblastic leukemia. *Experimental Hematology* **43**, 625–639 (2015).
44. Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T. Acute Myeloid Leukemia - Overview. in *Atlas of Hematopathology* (eds. Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T. B. T.-A. of H. (Second E.) 293–302 (Academic Press, 2018).
45. Irons, R. D. & Stillman, W. S. The process of leukemogenesis. *Environ. Health Perspect.* **104 Suppl**, 1239–1246 (1996).
46. Uckun, F. M. *et al.* Clinical features and treatment outcome of childhood T-lineage acute lymphoblastic leukemia according to the apparent maturational stage of T- lineage leukemic blasts: A Children's Cancer Group study. *J. Clin. Oncol.* **15**, 2214–2221 (1997).
47. Krause, D. S. & Van Etten, R. A. Right on target: eradicating leukemic stem cells. *Trends in Molecular Medicine* **13**, 470–481 (2007).
48. Misaghian, N. *et al.* Targeting the leukemic stem cell: The Holy Grail of leukemia therapy. *Leukemia* **23**, 25–42 (2009).
49. Byrd, J. C., Stilgenbauer, S. & Flinn, I. W. Chronic Lymphocytic Leukemia. *Hematology* **2004**, 163–183 (2004).
50. Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: Rationale and important changes. *Blood* **114**, 937–951 (2009).
51. Ward, E., DeSantis, C., Robbins, A., Kohler, B. & Jemal, A. Childhood and adolescent cancer statistics, 2014. *CA. Cancer J. Clin.* **64**, 83–103 (2014).
52. Nordlund, J. & Syvänen, A. C. Epigenetics in pediatric acute lymphoblastic leukemia. *Seminars in Cancer Biology* **51**, 129–138 (2018).
53. Peirs, S. *et al.* Epigenetics in T-cell acute lymphoblastic leukemia. *Immunol. Rev.* **263**, 50–67 (2015).
54. Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat. Rev. Immunol.* **8**, 380–390 (2008).
55. Gorczyca, W. *Atlas of differential diagnosis in neoplastic hematopathology*. (CRC Press, 2014).
56. Dores, G. M., Devesa, S. S., Curtis, R. E., Linet, M. S. & Morton, L. M. Acute leukemia incidence and patient survival among children and adults in the United States, 2001-2007. *Blood* **119**, 34–43 (2012).
57. Karrman, K. & Johansson, B. Pediatric T-cell acute lymphoblastic leukemia. *Genes Chromosomes and Cancer* **56**,

89–116 (2017).

58. Terwilliger, T. & Abdul-Hay, M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J.* **7**, e577 (2017).

59. Belver, L. & Ferrando, A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nature Reviews Cancer* **16**, 494–507 (2016).

60. Crist, W. M. *et al.* Clinical features and outcome in childhood T-cell leukemia-lymphoma according to stage of thymocyte differentiation: a Pediatric Oncology Group Study. *Blood* **72**, 1891–1897 (1988).

61. Madhusoodhan, P. P., Carroll, W. L. & Bhatla, T. Progress and Prospects in Pediatric Leukemia. *Curr. Probl. Pediatr. Adolesc. Health Care* **46**, 229–241 (2016).

62. Schrappe, M. *et al.* Late MRD response determines relapse risk overall and in subsets of childhood T-cell ALL: Results of the AIEOP-BFM-ALL 2000 study. *Blood* **118**, 2077–2084 (2011).

63. Van Der Meulen, J., Van Roy, N., Van Vlierberghe, P. & Speleman, F. The epigenetic landscape of T-cell acute lymphoblastic leukemia. *Int. J. Biochem. Cell Biol.* **53**, 547–557 (2014).

64. Campana, D., Thompson, J. S., Amlot, P., Brown, S. & Janossy, G. The cytoplasmic expression of CD3 antigens in normal and malignant cells of the T lymphoid lineage. *J. Immunol.* **138**, 648–55 (1987).

65. Bene, M. C. *et al.* Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia* **9**, 1783–6 (1995).

66. Liang, D.-C. & Pui, C.-H. Childhood acute lymphoblastic leukemia. in *Postgraduate Haematology* (eds. Hoffbrand, V. A., Catovsky, D. & Tuddenham, E. G. D.) 542–560 (2005).

67. Hoelzer, D. & Gökbuget, N. Adult acute lymphoblastic leukemia. in *Postgraduate Haematology* (eds. Hoffbrand, V. A., Catovsky, D. & Tuddenham, E. G. D.) 525–541 (Blackwell Publishing, 2005).

68. Minowada, J. *et al.* Marker profiles of human leukemia and lymphoma cell lines. *J. Cancer Res. Clin. Oncol.* **101**, 91–100 (1981).

69. Matsuo, Y. *et al.* Expression and distribution of cytoplasmic CD3 antigen and TcR i-chain with or without expression of surface CD3 TcR α/β or TcR γ/δ complexes. in *Leukocyte Typing* (eds. Knapp, W. et al.) (Oxford University Press, 1989).

70. Drexler, H. G., Gaedicke, G. & Minowada, J. Isoenzyme studies in human leukemia-lymphoma cell lines - V. Induction of differentiation by T-cell derived differentiation-inducing activity. *Leuk. Res.* **11**, 85–96 (1987).

71. Harris, N. L. *et al.* A revised European-American classification of lymphoid neoplasms: A proposal from the International Lymphoma Study Group. *Blood* **84**, 1361–1392 (1994).

72. Bhojwani, D., Yang, J. J. & Pui, C. H. Biology of childhood acute lymphoblastic leukemia. *Pediatr. Clin. North Am.* **62**, 47–60 (2015).

73. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome-biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).

74. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).

75. Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science (80-.).* **300**, 455 (2003).

76. Ehrlich, M. DNA methylation in cancer: Too much, but also too little. *Oncogene* **21**, 5400–5413 (2002).

77. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: A new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).

78. Ropero, S. & Esteller, M. The role of histone deacetylases (HDACs) in human cancer. *Mol. Oncol.* **1**, 19–25 (2007).

79. Esteller, M. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene* **21**, 5427–5440 (2002).

80. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12–27 (2012).

81. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

82. Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* **5**, 3630 (2014).

83. Meyers, R. A. *Epigenetic Regulation and Epigenomics*. (Wiley-Blackwell, 2012).

84. Mullighan, C. G. *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* **471**, 235–241 (2011).

85. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).

86. Simon, C. *et al.* A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev.* **26**, 651–656 (2012).

87. Maleszewska, M. & Kaminska, B. Is glioblastoma an epigenetic malignancy? *Cancers* **5**, 1120–1139 (2013).

88. Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* **5**, 769–784 (2006).

89. Roth, S. Y., Denu, J. M. & Allis, C. D. Histone Acetyltransferases. *Annu. Rev. Biochem.* **70**, 81–120 (2001).

90. Thiagalingam, S. *et al.* Histone deacetylases: Unique players in shaping the epigenetic histone code. in *Annals of the New York Academy of Sciences* **983**, 84–100 (New York Academy of Sciences, 2003).

91. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 536–550 (2014).

92. Glozak, M. A. & Seto, E. Histone deacetylases and cancer. *Oncogene* **26**, 5420–5432 (2007).

93. Barneda-Zahonero, B. & Parra, M. Histone deacetylases and cancer. *Mol. Oncol.* **6**, 579–589 (2012).

94. Moreno, D. A. *et al.* Differential expression of HDAC3, HDAC7 and HDAC9 is associated with prognosis and survival in childhood acute lymphoblastic leukaemia: Research paper. *Br. J. Haematol.* **150**, 665–673 (2010).

95. Gruhn, B. *et al.* The expression of histone deacetylase 4 is associated with prednisone poor-response in childhood acute lymphoblastic leukemia. *Leuk. Res.* **37**, 1200–1207 (2013).

96. Sonnemann, J., Gruhn, B., Wittig, S., Becker, S. & Beck, J. F. Increased activity of histone deacetylases in childhood acute lymphoblastic leukaemia and acute myeloid leukaemia: support for histone deacetylase inhibitors as antileukaemic agents. *Br. J. Haematol.* **158**, 664–666 (2012).

97. Richon, V. M. Cancer biology: Mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor. *Br. J. Cancer* **95**, 2–6 (2006).

98. Oki, Y. & Issa, J.-P. Review: Recent Clinical Trials in Epigenetic Therapy. *Rev. Recent Clin. Trials* **1**, 169–182 (2008).

99. Masetti, R., Serravalle, S., Biagi, C. & Pession, A. The Role of HDACs Inhibitors in Childhood and Adolescence Acute Leukemias. *J. Biomed. Biotechnol.* **2011**, 1–9 (2011).

100. Plass, C. *et al.* Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nature Reviews Genetics* **14**, 765–780 (2013).

101. Ceccacci, E. & Minucci, S. Inhibition of histone deacetylases in cancer therapy: Lessons from leukaemia. *Br. J. Cancer* **114**, 605–611 (2016).

102. Carey, N. & La Thangue, N. B. Histone deacetylase inhibitors: gathering pace. *Curr. Opin. Pharmacol.* **6**, 369–375 (2006).

103. Stimson, L., Wood, V., Khan, O., Fotheringham, S. & La Thangue, N. B. HDAC inhibitor-based therapies and haematological malignancy. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **20**, 1293–1302 (2009).

104. Dokmanovic, M., Clarke, C. & Marks, P. A. Histone Deacetylase Inhibitors: Overview and Perspectives. *Mol. Cancer Res.* **5**, 981–989 (2007).

105. Patel, J., Pathak, R. R. & Mujtaba, S. The Biology of Lysine Acetylation Integrates Transcriptional Programming and Metabolism. in *Epigenetics and Pathology: Exploring Connections Between Genetic Mechanisms and Disease Expression* (ed. Ayyanathan, K.) 120 (CRC Press Taylor & Francis Group, 2013).

106. Tsapis, M. *et al.* HDAC inhibitors induce apoptosis in glucocorticoid-resistant acute lymphatic leukemia cells despite a switch from the extrinsic to the intrinsic death pathway. *Int. J. Biochem. Cell Biol.* **39**, 1500–1509 (2007).

107. Bradner, J. E. *et al.* Chemical phylogenetics of histone deacetylases. *Nat. Chem. Biol.* **6**, 238–243 (2010).

108. Marks, P. A. & Breslow, R. Dimethyl sulfoxide to vorinostat: Development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* **25**, 84–90 (2007).

109. Bubna, A. Vorinostat-An overview. *Indian J. Dermatol.* **60**, 419 (2015).

110. Richon, V. M., Garcia-Vargas, J. & Hardwick, J. S. Development of vorinostat: Current applications and future perspectives for cancer therapy. *Cancer Letters* **280**, 201–210 (2009).

111. Kelly, W. K. *et al.* Phase I study of an oral histone deacetylase inhibitor, suberoylanilide hydroxamic acid, in patients with advanced cancer. *J. Clin. Oncol.* **23**, 3923–3931 (2005).

112. Peart, M. J. *et al.* Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3697–3702 (2005).

113. Cardoso, B. A. *et al.* TAL1/SCL is downregulated upon histone deacetylase inhibition in T-cell acute lymphoblastic leukemia cells. *Leukemia* **25**, 1578–1586 (2011).

114. Leclerc, G. J., Mou, C., Leclerc, G. M., Mian, A. M. & Barredo, J. C. Histone deacetylase inhibitors induce FPGS mRNA expression and intracellular accumulation of long-chain methotrexate polyglutamates in childhood acute lymphoblastic leukemia: Implications for combination therapy. *Leukemia* **24**, 552–562 (2010).

115. Loosveld, M. *et al.* Therapeutic Targeting of c-Myc in T-Cell Acute Lymphoblastic Leukemia (T-ALL). *Oncotarget* **5**, 3168–3172 (2014).

116. Kim, M. Y., Hur, J. & Jeong, S. Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep.* **42**, 125–130 (2009).

117. Castello, A., Fischer, B., Hentze, M. W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends Genet.* **29**, 318–327 (2013).

118. Strein, C., Alleaume, A. M., Rothbauer, U., Hentze, M. W. & Castello, A. A versatile assay for RNA-binding proteins in living cells. *Rna* **20**, 721–731 (2014).

119. Hong, S. RNA Binding Protein as an Emerging Therapeutic Target for Cancer Prevention and Treatment. *J.*

Cancer Prev. **22**, 203–210 (2018).

120. Moore, S., Järvelin, A. I., Davis, I., Bond, G. L. & Castello, A. Expanding horizons: new roles for non-canonical RNA-binding proteins in cancer. *Curr. Opin. Genet. Dev.* **48**, 112–120 (2018).
121. Sternburg, E. L. & Karginov, F. V. Global Approaches in Studying RNA-Binding Protein Interaction Networks. *Trends Biochem. Sci.* 1–11 (2020).
122. Moore, M. J. From birth to death: The complex lives of eukaryotic mRNAs. *Science* **309**, 1514–1518 (2005).
123. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology* **8**, 479–490 (2007).
124. Kishore, S., Luber, S. & Zavolan, M. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief. Funct. Genomics* **9**, 391–404 (2010).
125. Lukong, K. E., Chang, K. wei, Khandjian, E. W. & Richard, S. RNA-binding proteins in human genetic disease. *Trends Genet.* **24**, 416–425 (2008).
126. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **582**, 1977–1986 (2008).
127. Cléry, A. & Allain, F. H.-T. From Structure to Function of RNA Binding Domains. in *Madame Curie Bioscience Database [Internet]* (Landes Bioscience, 2013).
128. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **63**, 696–710 (2016).
129. Pereira, B., Billaud, M. & Almeida, R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in Cancer* **3**, 506–528 (2017).
130. Müller-Mcnicoll, M. & Neugebauer, K. M. How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics* **14**, 275–287 (2013).
131. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **19**, 327–341 (2018).
132. Lee, S. R. & Lykke-Andersen, J. Emerging roles for ribonucleoprotein modification and remodeling in controlling RNA fate. *Trends in Cell Biology* **23**, 504–510 (2013).
133. Elkon, R., Zlotorynski, E., Zeller, K. I. & Agami, R. Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* **11**, 259 (2010).
134. Hao, S. & Baltimore, D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.* **10**, 281–288 (2009).
135. Corbett, A. H. Post-transcriptional regulation of gene expression and human disease. *Curr. Opin. Cell Biol.* **52**, 96–104 (2018).
136. Sutherland, J. M., Siddall, N. A., Hime, G. R. & McLaughlin, E. A. RNA binding proteins in spermatogenesis: An in depth focus on the Musashi family. in *Asian Journal of Andrology* **17**, 529–536 (Medknow Publications, 2015).
137. Schoenberg, D. R. & Maquat, L. E. Re-capping the message. *Trends in Biochemical Sciences* **34**, 435–442 (2009).
138. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, 1–2 (2011).
139. Proudfoot, N. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Current Opinion in Cell Biology* **16**, 272–278 (2004).

140. Iglesias, N. & Stutz, F. Regulation of mRNP dynamics along the export pathway. *FEBS Letters* **582**, 1987–1996 (2008).

141. Idler, R. K. & Yan, W. Control of messenger RNA fate by review RNA-binding proteins: An emphasis on mammalian spermatogenesis. *Journal of Andrology* **33**, 309–337 (2012).

142. Castello, A., Hentze, M. W. & Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol. Metab.* **26**, 746–757 (2015).

143. Hentze, M. W. *et al.* Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science (80-.).* **238**, 1570–1573 (1987).

144. Casey, J. L. *et al.* Iron-responsive elements: Regulatory RNA sequences that control mRNA levels and translation. *Science (80-.).* **240**, 924–928 (1988).

145. Muckenthaler, M. U., Galy, B. & Hentze, M. W. Systemic Iron Homeostasis and the Iron-Responsive Element/Iron-Regulatory Protein (IRE/IRP) Regulatory Network. *Annu. Rev. Nutr.* **28**, 197–213 (2008).

146. Brinegar, A. E. & Cooper, T. A. Roles for RNA-binding proteins in development and disease. *Brain Research* **1647**, 1–8 (2016).

147. Kechavarzi, B. & Janga, S. C. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol.* **15**, 1–16 (2014).

148. Gabut, M., Chaudhry, S. & Blencowe, B. J. SnapShot: The Splicing Regulatory Machinery. *Cell* **133**, (2008).

149. Sebestyén, E. *et al.* Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **26**, 732–744 (2016).

150. Wang, Z. L. *et al.* Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. *Cell Rep.* **22**, 286–298 (2018).

151. Wang, J., Liu, Q. & Shyr, Y. Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC Genomics* **16**, S5 (2015).

152. Wurth, L. Versatility of RNA-Binding Proteins in Cancer. *Comp. Funct. Genomics* **2012**, 178525 (2012).

153. Babic, I., Jakymiw, A. & Fujita, D. J. The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity. *Oncogene* **23**, 3781–3789 (2004).

154. Hodson, D. J. *et al.* Deletion of the RNA-binding proteins ZFP36L1 and ZFP36L2 leads to perturbed thymic development and T lymphoblastic leukemia. *Nat. Immunol.* **11**, 717–724 (2010).

155. Fancello, L., Kampen, K. R., Hofman, I. J. F., Verbeeck, J. & De Keersmaecker, K. The ribosomal protein gene RPL5 is a haploinsufficient tumor suppressor in multiple cancer types. *Oncotarget* **8**, 14462–14478 (2017).

156. Wang, E. *et al.* Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell* **35**, 369-384.e7 (2019).

157. Walker, J. M. *Post-Transcriptional Gene Regulation*. (2016).

158. Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA–protein interactions. *Nat. Methods* **16**, 225–234 (2019).

159. Marchese, D., de Groot, N. S., Lorenzo Gotor, N., Livi, C. M. & Tartaglia, G. G. Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* **7**, 793–810 (2016).

160. Steitz, J. A. Immunoprecipitation of ribonucleoproteins using autoantibodies. *Methods Enzymol.* **180**, 468–481 (1989).

161. Tenenbaum, S. A., Lager, P. J., Carson, C. C. & Keene, J. D. Ribonomics: Identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* **26**, 191–198 (2002).

162. Gilbert, C. & Svejstrup, J. Q. RNA Immunoprecipitation for Determining RNA-Protein Associations In Vivo. *Curr. Protoc. Mol. Biol.* **75**, 27.4.1-27.4.11 (2006).

163. Trifillis, P., Day, N. & Kiledjian, M. Finding the right RNA: Identification of cellular mRNA substrates for RNA-binding proteins. *RNA* **5**, 1071–1082 (1999).

164. Darnell, R. B. HITS-CLIP: Panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip. Rev. RNA* **1**, 266–286 (2010).

165. Wheeler, E. C., Van Nostrand, E. L. & Yeo, G. W. Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdiscip. Rev. RNA* **9**, (2018).

166. Meisenheimer, K. M. & Koch, T. H. Photocross-linking of nucleic acids to associated proteins. *Critical Reviews in Biochemistry and Molecular Biology* **32**, 101–140 (1997).

167. Hafner, M. *et al.* Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

168. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).

169. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* **46**, 674–690 (2012).

170. Lindberg, U. & Sundquist, B. Isolation of messenger ribonucleoproteins from mammalian cells. *J. Mol. Biol.* **86**, 451–468 (1974).

171. Greenberg, J. R. Ultraviolet light-induced crosslinking of mRNA to proteins. *Nucleic Acids Res.* **6**, 715–732 (1979).

172. Wagenmakers, A. J. M., Reinders, R. J. & van Venrooij, W. J. Cross-linking of mRNA to Proteins by Irradiation of Intact Cells with Ultraviolet Light. *Eur. J. Biochem.* **112**, 323–330 (1980).

173. Scherrer, T., Mittal, N., Janga, S. C. & Gerber, A. P. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One* **5**, (2010).

174. Butter, F., Scheibe, M., Mörl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10626–10631 (2009).

175. Castello, A. *et al.* System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491–500 (2013).

176. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Proteins by RNA Interactome Capture. in 131–139 (2016).

177. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).

178. Perez-Perri, J. I. *et al.* Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun.* **9**, (2018).

179. Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **6**, (2015).

180. Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* **20**, 127–133

(2013).

181. Bunnik, E. M. *et al.* The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. *Genome Biol.* **17**, (2016).

182. Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D. & Erben, E. D. Gene expression regulatory networks in Trypanosoma brucei: Insights into the role of the mRNA-binding proteome. *Mol. Microbiol.* **100**, 457–471 (2016).

183. Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S. & Gehring, C. The RNA-binding protein repertoire of Arabidopsis thaliana. *Sci. Rep.* **6**, (2016).

184. Reichel, M. *et al.* In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. *Plant Cell* **28**, 2435–2452 (2016).

185. Sysoev, V. O. *et al.* Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nat. Commun.* **7**, 1–11 (2016).

186. Wessels, H. H. *et al.* The mRNA-bound proteome of the early fly embryo. *Genome Res.* **26**, 1000–1009 (2016).

187. Conrad, T. *et al.* Serial interactome capture of the human cell nucleus. *Nat. Commun.* **7**, 1–11 (2016).

188. Vester, B. & Wengel, J. LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary RNA and DNA †. *Biochemistry* **43**, 13233–13241 (2004).

189. Kauppinen, S., Vester, B. & Wengel, J. Locked nucleic acid (LNA): High affinity targeting of RNA for diagnostics and therapeutics. *Drug Discovery Today: Technologies* **2**, 287–290 (2005).

190. Singh, B. N. *et al.* Nonhistone protein acetylation as cancer therapy targets. *Expert Rev. Anticancer Ther.* **10**, 935–954 (2010).

191. Zhang, C. *et al.* Histone acetylation: Novel target for the treatment of acute lymphoblastic leukemia. *Clin. Epigenetics* **7**, 1–10 (2015).

192. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. Protein Measurementwith the Folin Phenol Reagent. *J. Biol. Chem.* **193**, 265–275 (1951).

193. Peterson, G. L. Review of the folin phenol protein quantitation method of lowry, rosebrough, farr and randall. *Anal. Biochem.* **100**, 201–220 (1979).

194. Sasse, J. & Gallagher, S. R. Staining proteins in gels. in *Current protocols in molecular biology* **Chapter 10**, Unit 10.6 (2009).

195. Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354 (1979).

196. Backlund, M. *et al.* Plasticity of nuclear and cytoplasmic stressresponses of RNA-binding proteins. *Nucleic Acids Res.* **48**, 4725–4740 (2020).

197. Givan, A. L. Flow Cytometry: An Introduction. in *Flow Cytometry Protocols* (eds. Hawley, T. S. & Hawley, R. G.) 1–29 (Humana Press, 2011).

198. Radbruch, A., Mechtold, B., Thiel, A., Milteny, S. & Pflüger, E. Flow Cytometry Part B. in *Methods in Cell Biology* (eds. Darzynkiewicz, Z., Robinson, J. P. & Crissman, H. A.) (Academic Press, 1994).

199. Erarslan-Uysal, B. Cell of origin in pediatric T-cell leukemia: An epigenetic approach to dissect T-ALL pathogenesis in the context of T-cell development. (Ruperto Carola University Heidelberg, 2020).

200. Erarslan-Uysal, B. *et al.* Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors. *EMBO Mol. Med.* 1–14 (2020).

201. Jacobsen, N. *et al.* Direct isolation of poly(A)+ RNA from 4 M guanidine thiocyanate-lysed cell extracts using locked nucleic acid-oligo(T) capture. *Nucleic Acids Res.* **32**, e64–e64 (2004).

202. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).

203. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).

204. Dayon, L. *et al.* Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal. Chem.* **80**, 2921–2931 (2008).

205. Werner, T. *et al.* Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal. Chem.* **86**, 3594–3601 (2014).

206. Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **10**, 1567–1593 (2015).

207. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

208. Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. in *Bioinformatics* **18**, (Oxford University Press, 2002).

209. Gatto, L. & Lilley, K. S. Msnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **28**, 288–289 (2012).

210. Strimmer, K. fdrtool: A versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).

211. Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc. Ser. A (Statistics Soc.* **174**, 245–246 (2011).

212. Heberle, H., Meirelles, V. G., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, (2015).

213. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - A web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, (2008).

214. Sridharan, S. *et al.* Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat. Commun.* **10**, (2019).

215. Oates, M. E. *et al.* D2P2: Database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508 (2013).

216. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

217. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

218. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

219. Snel, B., Lehmann, G., Bork, P. & Huynen, M. A. String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442–3444 (2000).

220. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

221. Radbruch, A. *et al.* Flow Cytometry, Part B, Volume 42. in *Methods in Cell Biology* (eds. Darzynkiewicz, Z., Robinson, J. P. & Crissman, H. A.) (Academic Press, 1994).

222. Ito, T. *et al.* A CD1a+/CD11c+ subset of human blood dendritic cells is a direct precursor of Langerhans cells. *J. Immunol.* **163**, 1409–19 (1999).

223. Zhao, L. *et al.* Intracellular water specific MR of microbead-adherent cells: Hela cell intracellular water diffusion. *Magn. Reson. Med.* **59**, 79–84 (2008).

224. Schmid-Schönbein, G. W., Shih, Y. Y. & Chien, S. Morphometry of human leukocytes. *Blood* **56**, 866–75 (1980).

225. Rosenbluth, M. J., Lam, W. A. & Fletcher, D. A. Force Microscopy of Nonadherent Cells: A Comparison of Leukemia Cell Deformability. *Biophys. J.* **90**, 2994–3003 (2006).

226. Turgeon, M. L. *Clinical Hematology: Theory and Procedures*. (Lippincott Williams & Wolkins, 2005).

227. Eun, H.-M. *Enzymes and Nucleic Acids: General Principles*. *Enzymology Primer for Recombinant DNA Technology* (Academic Press, 1996).

228. Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R. & Pazdur, R. FDA Approval Summary: Vorinostat for Treatment of Advanced Primary Cutaneous T-Cell Lymphoma. *Oncologist* **12**, 1247–1252 (2007).

229. Frismantas, V. *et al.* Ex vivo drug response profiling detects recurrent sensitivity patterns in drug-resistant acute lymphoblastic leukemia. *Blood* **129**, e26–e37 (2017).

230. Tiffon, C. E. *et al.* The histone deacetylase inhibitors vorinostat and romidepsin downmodulate IL-10 expression in cutaneous T-cell lymphoma cells. *Br. J. Pharmacol.* **162**, 1590–1602 (2011).

231. Flatmark, K. *et al.* Radiosensitization of colorectal carcinoma cell lines by histone deacetylase inhibition. *Radiat. Oncol.* **1**, 1–10 (2006).

232. Liepelt, A. *et al.* Identification of RNA-binding Proteins in Macrophages by Interactome Capture. *Mol. Cell. Proteomics* **15**, 2699–2714 (2016).

233. Järvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. *Cell Commun. Signal.* **14**, (2016).

234. Boon, R. *et al.* Amino acid levels determine metabolism and CYP450 function of hepatocytes and hepatoma cell lines. *Nat. Commun.* **11**, (2020).

235. Hentze, M. W. Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains? *Trends Biochem. Sci.* **19**, 101–103 (1994).

236. Hentze, M. W. & Preiss, T. The REM phase of gene regulation. *Trends Biochem. Sci.* **35**, 423–426 (2010).

237. Tsvetanova, N. G., Klass, D. M., Salzman, J. & Brown, P. O. Proteome-wide search reveals unexpected RNA-binding proteins in saccharomyces cerevisiae. *PLoS One* **5**, 1–12 (2010).

238. Xie, H. *et al.* Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* **6**, 1882–1898 (2007).

239. Lobley, A., Swindells, M. B., Orengo, C. A. & Jones, D. T. Inferring function using patterns of native disorder in proteins. *PLoS Comput. Biol.* **3**, 1567–1579 (2007).

240. Billadeau, D. D., Nolz, J. C. & Gomez, T. S. Regulation of T-cell activation by the cytoskeleton. *Nat. Rev. Immunol.* **7**, 131–143 (2007).

241. Lin, W. *et al.* Morphological change of CD4+ T cell during contact with DC modulates T-cell activation by accumulation of F-actin in the immunology synapse. *BMC Immunol.* **16**, (2015).

242. Rougerie, P., Miskolci, V. & Cox, D. Generation of membrane structures during phagocytosis and chemotaxis of macrophages: Role and regulation of the actin cytoskeleton. *Immunol. Rev.* **256**, 222–239 (2013).

243. Mugnier, B. *et al.* Coronin-1A links cytoskeleton dynamics to TCRαβ-induced cell signaling. *PLoS One* **3**, 3467 (2008).

244. Herrero-Gonzalez, S. & Di Cristofano, A. New Routes to Old Places: PIK3R1 and PIK3R2 Join PIK3CA and PTEN as Endometrial Cancer Genes. *Cancer Discov.* **1**, 106–107 (2011).

245. Shi, X. *et al.* Research progress on the PI3K/AKT signaling pathway in gynecological cancer (Review). *Mol. Med. Rep.* (2019).

246. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1211–1218 (2017).

247. Chen, B. *et al.* Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci.* **115**, 373–378 (2018).

248. Hoellenriegel, J. *et al.* The phosphoinositide 3′-kinase delta inhibitor, CAL-101, inhibits B-cell receptor signaling and chemokine networks in chronic lymphocytic leukemia. *Blood* **118**, 3603–3612 (2011).

249. Billottet, C. *et al.* A selective inhibitor of the p110δ isoform of PI 3-kinase inhibits AML cell proliferation and survival and increases the cytotoxic effects of VP16. *Oncogene* **25**, 6648–6659 (2006).

250. Nguyen, L. X. T., Sesay, A. & Mitchell, B. S. Effect of CAL-101, a PI3Kδ inhibitor, on ribosomal rna synthesis and cell proliferation in acute myeloid leukemia cells. *Blood Cancer J.* **4**, e228–e228 (2014).

251. Sanchez, V., Nichols, C., Kim, H., Gang, E. & Kim, Y.-M. Targeting PI3K Signaling in Acute Lymphoblastic Leukemia. *Int. J. Mol. Sci.* **20**, 412 (2019).

252. Saito, K. *et al.* TANGO1 Facilitates Cargo Loading at Endoplasmic Reticulum Exit Sites. *Cell* **136**, 891–902 (2009).

253. Wilson, D. G. *et al.* Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. *J. Cell Biol.* **193**, 935–951 (2011).

254. Bosserhoff, A. ., Moser, M. & Buettner, R. Characterization and expression pattern of the novel MIA homolog TANGO. *Gene Expr. Patterns* **4**, 473–479 (2004).

255. Wong, R. W., Blobel, G. & Coutavas, E. Rae1 interaction with NuMA is required for bipolar spindle formation. *Proc. Natl. Acad. Sci.* **103**, 19783–19787 (2006).

256. Endo, A., Moyori, A., Kobayashi, A. & Wong, R. W. Nuclear mitotic apparatus protein, NuMA, modulates p53-mediated transcription in cancer cells. *Cell Death Dis.* **4**, e713–e713 (2013).

257. Sukhai, M. A. *et al.* Myeloid leukemia with promyelocytic features in transgenic mice expressing hCG-NuMA-RARα. *Oncogene* **23**, 665–678 (2004).

258. Beer, S., Simins, A. B., Schuster, A. & Holzmann, B. Molecular cloning and characterization of a novel SH3 protein (SLY) preferentially expressed in lymphoid cells. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1520**, 89–93 (2001).

259. Reis, B., Pfeffer, K. & Beer-Hammer, S. The orphan adapter protein SLY1 as a novel anti-apoptotic protein required for thymocyte development. *BMC Immunol.* **10**, 38 (2009).

260. Koyama, M. & Matsuura, Y. An allosteric mechanism to displace nuclear export cargo from CRM1 and RanGTP by RanBP1. *EMBO J.* **29**, 2002–2013 (2010).

261. Arnaoutov, A. & Dasso, M. The Ran GTPase Regulates Kinetochore Function. *Dev. Cell* **5**, 99–111 (2003).

262. Peloponese, J.-M., Haller, K., Miyazato, A. & Jeang, K.-T. Abnormal centrosome amplification in cells through the targeting of Ran-binding protein-1 by the human T cell leukemia virus type-1 Tax oncoprotein. *Proc. Natl. Acad. Sci.* **102**, 18974–18979 (2005).

263. Giam, C.-Z. & Semmes, O. HTLV-1 Infection and Adult T-Cell Leukemia/Lymphoma—A Tale of Two Proteins: Tax and HBZ. *Viruses* **8**, 161 (2016).

264. Yusuf, I. & Fruman, D. A. Regulation of quiescence in lymphocytes. *Trends in Immunology* **24**, 380–386 (2003).

265. van der Windt, G. J. W. & Pearce, E. L. Metabolic switching and fuel choice during T-cell differentiation and memory development. *Immunol. Rev.* **249**, 27–42 (2012).

266. Ho, A. W. & Kupper, T. S. T cells and the skin: from protective immunity to inflammatory skin disorders. *Nat. Rev. Immunol.* **19**, 490–502 (2019).

267. Nascimento-Filho, C. H. V. *et al.* Skin wound healing triggers epigenetic modifications of histone H4. *J. Transl. Med.* **18**, 138 (2020).

268. White, M. R. & Garcin, E. D. The sweet side of RNA regulation: glyceraldehyde-3-phosphate dehydrogenase as a noncanonical RNA-binding protein. *Wiley Interdiscip. Rev. RNA* **7**, 53–70 (2016).

269. Li, T. *et al.* Glyceraldehyde-3-phosphate Dehydrogenase Is Activated by Lysine 254 Acetylation in Response to Glucose Signal. *J. Biol. Chem.* **289**, 3775–3785 (2014).

270. Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).

271. Zhou, Y. *et al.* Posttranslational Regulation of the Exon Skipping Machinery Controls Aberrant Splicing in Leukemia. *Cancer Discov.* **10**, 1388–1409 (2020).

272. Schmidt, C., Beilsten-Edmands, V. & Robinson, C. V. Insights into Eukaryotic Translation Initiation from Mass Spectrometry of Macromolecular Protein Assemblies. *J. Mol. Biol.* **428**, 344–356 (2016).

273. Ishfaq, M. *et al.* Acetylation regulates subcellular localization of eukaryotic translation initiation factor 5A (eIF5A). *FEBS Lett.* **586**, 3236–3241 (2012).

274. Takahashi, M. *et al.* Polyglutamine tract binding protein-1 is an intrinsically unstructured protein. *Biochim. Biophys. Acta - Proteins Proteomics* **1794**, 936–943 (2009).

275. Estin, M. L. *et al.* Ena/VASP proteins regulate activated T-cell trafficking by promoting diapedesis during transendothelial migration. *Proc. Natl. Acad. Sci.* **114**, E2901–E2910 (2017).

276. Bai, S. *et al.* Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol.* **9**, 54 (2011).

277. Mukherjee, K. *et al.* Actin capping protein CAPZB regulates cell morphology, differentiation, and neural crest migration in craniofacial morphogenesis. *Hum. Mol. Genet.* **25**, 1255–1270 (2016).

278. Dehghan-Nayeri, N. *et al.* Differential expression pattern of protein markers for predicting chemosensitivity of dexamethasone-based chemotherapy of B cell acute lymphoblastic leukemia. *Cancer Chemother. Pharmacol.* **80**, 177–185 (2017).

279. Chiaretti, S. Gene Expression Profiles of B-lineage Adult Acute Lymphocytic Leukemia Reveal Genetic Patterns that Identify Lineage Derivation and Distinct Mechanisms of Transformation. *Clin. Cancer Res.* **11**, 7209–7219 (2005).

280. Smathers, R. L. & Petersen, D. R. The human fatty acid-binding protein family: Evolutionary divergences and

functions. *Hum. Genomics* **5**, 170 (2011).

281. Zimmerman, A. W. & Veerkamp, J. H. New insights into the structure and function of fatty acid-binding proteins. *Cell. Mol. Life Sci.* **59**, 1096–1116 (2002).
282. Senga, S., Kobayashi, N., Kawaguchi, K., Ando, A. & Fujii, H. Fatty acid-binding protein 5 (FABP5) promotes lipolysis of lipid droplets, de novo fatty acid (FA) synthesis and activation of nuclear factor-kappa B (NF-κB) signaling in cancer cells. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1863**, 1057–1067 (2018).
283. Galante, P. A. F. *et al.* A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue: Identification of potential players in tumor formation. *RNA Biol.* **6**, 426 (2009).
284. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
285. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**, 629–643 (2017).
286. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
287. Bampton, A., Gittings, L. M., Fratta, P., Lashley, T. & Gatt, A. The role of hnRNPs in frontotemporal dementia and amyotrophic lateral sclerosis. *Acta Neuropathol.* **140**, 599–623 (2020).
288. Vazquez, F. & Sellers, W. R. The PTEN tumor suppressor protein: an antagonist of phosphoinositide 3-kinase signaling. *Biochim. Biophys. Acta - Rev. Cancer* **1470**, 21–35 (2000).

APPENDIX

Supplementary table 1: High confidence RBPs identified in untreated T-ALL cell lines (+UV/-UV ratio of at least two-fold, FDR < 0.05).

P12
ACIN1, ACTN4, ADAR, AGO2, AKAP1, AKAP8, ALYREF, ANKHD1, ANXA1, ANXA11, ANXA2, API5, APOBEC3G, AQR, ARGLU1, ARHGDIA, ARHGDIB, ARL6IP4, ARPP21, ASCC1, ATP5A1, ATXN2, ATXN2L, BAG4, BCLAF1, BRD2, BUD13, BZW1, C11ORF57, C11ORF58|SMAP, C1ORF52, C7ORF55|LUC7L2, C9ORF78, CAB39, CACTIN, CAPRIN1, CAPZB, CASP14, CAT, CBX5, CCAR1, CCAR2, CCDC25, CCDC9, CDC37, CDK11B|CDC2L1, CELF1, CELF2, CENPF, CEP85, CHD1, CHD4, CHERP, CHMP2A, CIR1, CIRBP, CNOT1, CNOT3, CNOT4, CORO1A, CPSF6, CPSF7, CSDE1, CSTF2T, CSTF3, CWC25, CWF19L2, DAZAP1, DBN1, DBNL, DBR1, DDX1, DDX17, DDX19A, DDX21, DDX23, DDX39B, DDX3X, DDX42, DDX46, DDX5, DDX6, DEK, DHX15, DHX30, DHX8, DHX9, DIAPH1, DKC1, DNAJC8, DNAJC9, DPYSL2, DSC1, DSG1, DUT, EDC4, EEF1A1|EEF1A1P5, EEF2, EFHD2, EFTUD2, EIF1AY|EIF1AX, EIF2S1, EIF3A, EIF3B, EIF3CL|EIF3C, EIF3E, EIF3G, EIF3H, EIF3J, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4ENIF1, EIF4G1, EIF4G2, EIF4H, EIF5B, ELAVL1, ELAVL4, EPRS, ESRP1, ETF1, EVL, EZR, FAM120A, FAM133B, FAM50A, FASTKD2, FKBP4, FLNA, FLNB, FMR1, FTSJ3, FUBP1, FUBP3, FUS, FXR1, FXR2, FYB, G3BP1, G3BP2, GAPDH, GARS, GDI2, GGCT, GIGYF2, GOLGB1, GPATCH1, GPATCH8, GRB2, GRSF1, GSPT1, GSPT2, HARS2, HCLS1, HDGFRP2, HDLBP, HELZ, HIST1H1B, HIST1H1D, HIST1H1E, HMGB1, HNRNPA0, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HOOK3, HSP90AA1, HSP90AB1, HSP90B1, HSPA5, HSPA8, HSPA9, HSPD1, HTATSF1, IGF2BP3, ILF2, ILF3, ILKAP, ISY1, KHDRBS1, KHSRP, KIF11, KPNA4, KPNB1, KRI1, KTN1, LARP1, LARP4, LARP4B, LCP1, LIN28B, LMNB1, LRPPRC, LRRFIP1, LSM14A, LSM14B, LSM3, LSM6, LSM8, LUC7L, LUC7L3, MACF1, MAP1A, MAP4, MAPRE1, MATR3, MCM2, MEX3C, MEX3D, MIA3, MMTAG2, MOV10, MRPL1, MRPS27, MRPS28, MRPS31, MSI2, MSN, MTPAP, MYEF2, MYH9, NASP, NCBP2, NCL, NCOA5, NKAP, NKTR, NOLC1, NONO, NOP56, NOP58, NPM1, NSUN2, NUDC, NUMA1, OLA1, P4HB, PA2G4, PABPC1, PABPC4, PABPN1, PAIP1, PAK2, PCBP1, PCBP2, PCBP3, PDAP1, PDCD11, PDCD6IP, PDIA4, PEBP1, PEG10, PFN1, PIK3R1, PKM, PKP1, PLEC, PNISR, PNN, PNPT1, POLR2A, POLR2B, PPIE, PPIG, PPIL4, PRKDC, PRMT1, PRPF38A, PRPF38B, PRPF40A, PRPF4B, PRPF8, PRRC2A, PRRC2C, PSME2, PSPC1, PTBP1, PTBP2, PTCD3, PUF60, PUM1, PUM2, PURA, PURB, PUS1, PUS7, PYM1, R3HDM1, R3HDM2, RACK1, RALY, RANBP1, RBBP6, RBM10, RBM12, RBM12B, RBM14, RBM15, RBM15B, RBM17, RBM22, RBM26, RBM3, RBM33, RBM39, RBM4, RBM45, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMX, RBMX2, RC3H1, RCC2, RCSD1, RNMT, RP9, RPA1, RPL10A, RPL13, RPL18, RPL22, RPL28, RPL34, RPL6, RPS11, RPS14, RPS2, RPS28, RPS3, RPS7, RPS8, RPSA, RRBP1, RSRC1, RSRC2, RTF1, RYDEN, SAFB, SAFB2, SASH3, SBDS, SCAF4, SDE2, SERBP1, SET, SF1, SF3A1, SF3B1, SF3B4, SFPQ, SFSWAP, SLBP, SLC4A1AP, SLIRP, SLTM, SND1, SNRNP200, SNRNP70, SON, SPEN, SPTAN1, SPTBN1, SRRM1, SRRM2, SRRT, SRSF1, SRSF10, SRSF11, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSB, SSBP1, SSRP1, ST13|ST13P5, STAU1, STAU2, STIP1, STMN1, STRAP, STRBP, SUB1, SUGP1, SUMO2, SUPT16H, SUPT5H, SUPT6H, SYNCRIP, SYNE2, TACC1, TAF15, TARDBP, TBRG4, TCOF1, TGM1, TGM3, THRAP3, THUMPD1, TIA1, TIAL1, TIPRL, TLN1, TNPO1, TNRC6B, TNRC6C, TOP1, TOP2A, TPM3, TPM4, TPR, TRA2A, TRA2B, TRNAU1AP, TUBB, TUFM, U2AF2, U2SURP, UBAP2, UBAP2L, UPF1, USP10, USP14, USP42, VCP, VIM, WDR33, XRCC5, XRCC6, XRN2, YBX1, YBX3, YKT6, YLPM1, YTHDF1, YTHDF2, YTHDF3, YWHAB, YWHAE, YWHAQ, YWHAZ, ZC3H10, ZC3H11A, ZC3H12A, ZC3H18, ZC3H4, ZC3H7A, ZC3H7B, ZC3HAV1, ZCCHC17, ZCCHC8, ZFR, ZNF326, ZRANB2
DND41
ACIN1, ACTN4, ADAR, AGO2, AKAP1, AKAP8, ALYREF, ANKHD1, ANXA1, API5, APOBEC3G, AQR, ARGLU1, ARHGDIA, ARHGDIB, ARL6IP4, ARPP21, ASCC1, ATP5A1, ATXN2, ATXN2L, BAG4, BCLAF1, BRD2, BUD13, BZW1, C11ORF57, C11ORF58|SMAP, C1ORF52, C7ORF55|LUC7L2, C9ORF78, CAB39, CACTIN, CAPRIN1, CAPZB, CBX5, CCAR1, CCAR2, CCDC25, CCDC9, CDC37, CDK11B|CDC2L1, CELF1, CELF2, CENPF, CEP85, CHD1, CHD4, CHERP, CHMP2A, CIR1, CIRBP, CNOT1, CNOT3, CNOT4, CPSF6, CPSF7, CSDE1, CSTF2T, CSTF3, CWC25, CWF19L2, DAZAP1, DBN1, DBNL, DBR1, DDX1, DDX17, DDX19A, DDX21, DDX23, DDX39B, DDX3X, DDX42, DDX46, DDX5, DDX6, DEK, DHX15, DHX30, DHX8, DHX9, DKC1, DNAJC8, DNAJC9, DPYSL2, DSC1, DSG1, DUT, EDC4, EEF1A1|EEF1A1P5, EEF2, EFHD2, EFTUD2, EIF1AY|EIF1AX, EIF2S1, EIF3A, EIF3B, EIF3CL|EIF3C, EIF3E, EIF3G, EIF3H, EIF3J, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4ENIF1, EIF4G1, EIF4G2, EIF4H, EIF5B, ELAVL1, ELAVL4, EPRS, ESRP1, ETF1, EVL, EZR, FAM120A, FAM133B, FAM50A, FASTKD2, FKBP4, FLNA, FLNB, FMR1, FTSJ3, FUBP1, FUBP3, FUS, FXR1, FXR2, FYB, G3BP1, G3BP2, GARS, GDI2, GGCT, GIGYF2, GPATCH1, GPATCH8, GRB2, GRSF1, GSPT1, GSPT2, HARS2, HCLS1, HDGFRP2, HDLBP, HELZ, HIST1H1B,

HIST1H1C, HIST1H1D, HIST1H1E, HIST1H3A|HIST3H3|HIST2H3A, HIST1H4A, HMGB1, HNRNPA0, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HOOK3, HSP90AA1, HSP90AB1, HSP90B1, HSPA5, HSPA8, HSPA9, HSPD1, HTATSF1, IGF2BP3, ILF2, ILF3, ILKAP, ISY1, KHDRBS1, KHSRP, KPNA4, KPNB1, KRI1, KTN1, LARP1, LARP4, LARP4B, LCP1, LIN28B, LMNB1, LRPPRC, LRRFIP1, LSM14A, LSM14B, LSM3, LSM6, LSM8, LUC7L, LUC7L3, MACF1, MAP1A, MAP4, MAPRE1, MATR3, MEX3C, MEX3D, MIA3, MMTAG2, MOV10, MRPL1, MRPS27, MRPS28, MRPS31, MSI2, MSN, MTPAP, MYEF2, MYH9, NASP, NCL, NCOA5, NKAP, NKTR, NOLC1, NONO, NOP56, NOP58, NPM1, NSUN2, NUDC, NUMA1, OLA1, P4HB, PA2G4, PABPC1, PABPC4, PABPN1, PAIP1, PAK2, PCBP1, PCBP2, PCBP3, PDAP1, PDCD11, PDCD6IP, PDIA4, PEBP1, PEG10, PFN1, PIK3R1, PKP1, PLEC, PNISR, PNN, PNPT1, POLR2A, POLR2B, PPIE, PPIG, PPIL4, PRKDC, PRMT1, PRPF38A, PRPF38B, PRPF40A, PRPF4B, PRPF8, PRRC2A, PRRC2C, PSME2, PSPC1, PTBP1, PTBP2, PTCD3, PUF60, PUM1, PUM2, PURA, PURB, PUS1, PUS7, PYM1, R3HDM1, R3HDM2, RACK1, RALY, RANBP1, RBBP6, RBM10, RBM12, RBM12B, RBM14, RBM15, RBM15B, RBM17, RBM22, RBM26, RBM3, RBM33, RBM39, RBM4, RBM45, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMX, RBMX2, RC3H1, RCC2, RCSD1, RNMT, RP9, RPA1, RPL13, RPL18, RPL22, RPL28, RPL34, RPL6, RPS11, RPS14, RPS2, RPS28, RPS3, RPS7, RPS8, RPSA, RSRC1, RSRC2, RTF1, RYDEN, SAFB, SAFB2, SASH3, SBDS, SCAF4, SDE2, SERBP1, SET, SF1, SF3A1, SF3B1, SF3B4, SFPQ, SFSWAP, SLBP, SLC4A1AP, SLIRP, SLTM, SND1, SNRNP200, SNRNP70, SON, SPEN, SPTAN1, SPTBN1, SRRM1, SRRM2, SRRT, SRSF1, SRSF10, SRSF11, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSB, SSBP1, SSRP1, ST13|ST13P5, STAU1, STAU2, STIP1, STMN1, STRAP, STRBP, SUB1, SUGP1, SUMO2, SUPT16H, SUPT5H, SUPT6H, SYNCRIP, SYNE2, TACC1, TAF15, TARDBP, TBRG4, TCOF1, TGM1, THRAP3, THUMPD1, TIA1, TIAL1, TIPRL, TLN1, TNPO1, TNRC6B, TNRC6C, TOP1, TOP2A, TPM4, TPR, TRA2A, TRA2B, TRNAU1AP, TUBB, TUFM, U2AF2, U2SURP, UBAP2, UBAP2L, UPF1, USP10, USP14, USP42, VIM, WDR33, XRCC5, XRCC6, XRN2, YBX1, YBX3, YKT6, YLPM1, YTHDF1, YTHDF2, YTHDF3, YWHAB, YWHAE, YWHAQ, YWHAZ, ZC3H10, ZC3H11A, ZC3H12A, ZC3H18, ZC3H4, ZC3H7A, ZC3H7B, ZC3HAV1, ZCCHC17, ZCCHC8, ZFR, ZNF326, ZRANB2

MOLT3

ACIN1, ACTN4, ADAR, AGO2, AKAP1, AKAP8, ALDOA, ALYREF, ANKHD1, ANXA1, ANXA11, API5, APOBEC3G, AQR, ARGLU1, ARHGDIA, ARHGDIB, ARL6IP4, ARPP21, ASCC1, ATP5A1, ATXN2, ATXN2L, BAG4, BCLAF1, BRD2, BUD13, BZW1, C11ORF57, C11ORF58|SMAP, C1ORF52, C7ORF55|LUC7L2, C9ORF78, CAB39, CACTIN, CALR, CAPRIN1, CAPZB, CBX5, CCAR1, CCAR2, CCDC25, CCDC9, CDC37, CDK11B|CDC2L1, CELF1, CELF2, CENPF, CEP85, CHD1, CHD4, CHERP, CHMP2A, CIR1, CIRBP, CNOT1, CNOT3, CNOT4, CPSF6, CPSF7, CSDE1, CSTF2T, CSTF3, CWC25, CWF19L2, DAZAP1, DBN1, DBNL, DBR1, DDX1, DDX17, DDX19A, DDX21, DDX23, DDX39B, DDX3X, DDX3Y, DDX42, DDX46, DDX5, DDX6, DEK, DHX15, DHX30, DHX8, DHX9, DIAPH1, DKC1, DNAJC8, DNAJC9, DPYSL2, DSC1, DSG1, DUT, EEF1A1|EEF1A1P5, EEF2, EFHD2, EFTUD2, EIF1AY|EIF1AX, EIF2S1, EIF3A, EIF3B, EIF3CL|EIF3C, EIF3E, EIF3G, EIF3H, EIF3J, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4ENIF1, EIF4G1, EIF4G2, EIF4H, EIF5B, ELAVL1, ELAVL4, ENO1, EPRS, ESRP1, ETF1, EVL, EZR, FAM120A, FAM133B, FAM50A, FASTKD2, FKBP4, FLNA, FLNB, FMR1, FTSJ3, FUBP1, FUBP3, FUS, FXR1, FXR2, FYB, G3BP1, G3BP2, GARS, GDI2, GGCT, GIGYF2, GOLGB1, GPATCH1, GPATCH8, GRB2, GRSF1, GSPT1, GSPT2, HARS2, HCLS1, HDGFRP2, HDLBP, HELZ, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H3A|HIST3H3|HIST2H3A, HIST1H4A, HMGB1, HNRNPA0, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HOOK3, HSP90AA1, HSP90AB1, HSP90B1, HSPA5, HSPA8, HSPA9, HSPD1, HTATSF1, IGF2BP3, ILF2, ILF3, ILKAP, ISY1, KHDRBS1, KHSRP, KIF11, KPNA4, KPNB1, KRI1, KTN1, LARP1, LARP4, LARP4B, LCP1, LIN28B, LMNB1, LRPPRC, LRRFIP1, LSM14A, LSM14B, LSM3, LSM6, LSM8, LUC7L, LUC7L3, MACF1, MAP1A, MAP4, MAPRE1, MATR3, MCM2, MEX3C, MEX3D, MIA3, MMTAG2, MOV10, MRPL1, MRPS27, MRPS28, MRPS31, MSI2, MSN, MTPAP, MYEF2, MYH9, NASP, NCBP2, NCL, NCOA5, NKAP, NKTR, NOLC1, NONO, NOP56, NOP58, NPM1, NSUN2, NUDC, NUMA1, OLA1, P4HB, PA2G4, PABPC1, PABPC4, PABPN1, PAIP1, PAK2, PCBP1, PCBP2, PCBP3, PDAP1, PDCD11, PDCD6IP, PDIA4, PEBP1, PEG10, PFN1, PIK3R1, PKM, PLEC, PNISR, PNN, PNPT1, POLR2A, POLR2B, PPIE, PPIG, PPIL4, PRKDC, PRMT1, PRPF38A, PRPF38B, PRPF40A, PRPF4B, PRPF8, PRRC2A, PRRC2C, PSME2, PSPC1, PTBP1, PTBP2, PTCD3, PTMA, PUF60, PUM1, PUM2, PURA, PURB, PUS1, PUS7, PYM1, R3HDM1, R3HDM2, RACK1, RALY, RANBP1, RBBP6, RBM10, RBM12, RBM12B, RBM14, RBM15, RBM15B, RBM17, RBM22, RBM26, RBM3, RBM33, RBM39, RBM4, RBM45, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMX, RBMX2, RC3H1, RCC2, RCSD1, RNMT, RP9, RPA1, RPL13, RPL18, RPL22, RPL28, RPL34, RPL6, RPS11, RPS14, RPS2, RPS28, RPS3, RPS7, RPS8, RPSA, RRBP1, RSRC1, RSRC2, RTF1, RYDEN, SAFB, SAFB2, SASH3, SBDS, SCAF4, SDE2, SERBP1, SET, SF1, SF3A1, SF3B1, SF3B4, SFPQ, SFSWAP, SLBP, SLC4A1AP, SLIRP, SLTM, SND1, SNRNP200, SNRNP70, SON, SPEN, SPTAN1, SPTBN1, SRRM1, SRRM2, SRRT, SRSF1, SRSF10, SRSF11, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSB, SSBP1, SSRP1, ST13|ST13P5, STAU1, STAU2, STIP1, STMN1, STRAP, STRBP, SUB1, SUGP1, SUMO2, SUPT16H, SUPT5H, SUPT6H, SYNCRIP, SYNE2, TACC1, TAF15, TARDBP, TBRG4, TCOF1,

TGM1, TGM3, THRAP3, THUMPD1, TIA1, TIAL1, TIPRL, TLN1, TNPO1, TNRC6B, TNRC6C, TOP1, TOP2A, TPM3, TPM4, TPR, TRA2A, TRA2B, TRNAU1AP, TUBB, TUFM, U2AF2, U2SURP, UBAP2, UBAP2L, UPF1, USP10, USP14, USP42, VIM, WDR33, XRCC5, XRCC6, XRN2, YBX1, YBX3, YKT6, YLPM1, YTHDF1, YTHDF2, YTHDF3, YWHAB, YWHAE, YWHAQ, YWHAZ, ZC3H10, ZC3H11A, ZC3H12A, ZC3H18, ZC3H4, ZC3H7A, ZC3H7B, ZC3HAV1, ZCCHC17, ZCCHC8, ZFR, ZNF326, ZRANB2

Supplementary table 2: 486 unique high confidence RBPs identified from P12, DND41 and MOLT3 referred to as the "T-ALL RNA interactome".

T-ALL RNA interactome
ACIN1, ACTN4, ADAR, AGO2, AKAP1, AKAP8, ALDOA, ALYREF, ANKHD1, ANXA1, ANXA11, ANXA2, API5, APOBEC3G, AQR, ARGLU1, ARHGDIA, ARHGDIB, ARL6IP4, ARPP21, ASCC1, ATP5A1, ATXN2, ATXN2L, BAG4, BCLAF1, BRD2, BUD13, BZW1, C11ORF57, C11ORF58|SMAP, C1ORF52, C7ORF55|LUC7L2, C9ORF78, CAB39, CACTIN, CALR, CAPRIN1, CAPZB, CASP14, CAT, CBX5, CCAR1, CCAR2, CCDC25, CCDC9, CDC37, CDK11B|CDC2L1, CELF1, CELF2, CENPF, CEP85, CHD1, CHD4, CHERP, CHMP2A, CIR1, CIRBP, CNOT1, CNOT3, CNOT4, CORO1A, CPSF6, CPSF7, CSDE1, CSTF2T, CSTF3, CWC25, CWF19L2, DAZAP1, DBN1, DBNL, DBR1, DDX1, DDX17, DDX19A, DDX21, DDX23, DDX39B, DDX3X, DDX3Y, DDX42, DDX46, DDX5, DDX6, DEK, DHX15, DHX30, DHX8, DHX9, DIAPH1, DKC1, DNAJC8, DNAJC9, DPYSL2, DSC1, DSG1, DUT, EDC4, EEF1A1|EEF1A1P5, EEF2, EFHD2, EFTUD2, EIF1AY|EIF1AX, EIF2S1, EIF3A, EIF3B, EIF3CL|EIF3C, EIF3E, EIF3G, EIF3H, EIF3J, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4ENIF1, EIF4G1, EIF4G2, EIF4H, EIF5B, ELAVL1, ELAVL4, ENO1, EPRS, ESRP1, ETF1, EVL, EZR, FAM120A, FAM133B, FAM50A, FASTKD2, FKBP4, FLNA, FLNB, FMR1, FTSJ3, FUBP1, FUBP3, FUS, FXR1, FXR2, FYB, G3BP1, G3BP2, GAPDH, GARS, GDI2, GGCT, GIGYF2, GOLGB1, GPATCH1, GPATCH8, GRB2, GRSF1, GSPT1, GSPT2, HARS2, HCLS1, HDGFRP2, HDLBP, HELZ, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H3A|HIST3H3|HIST2H3A, HIST1H4A, HMGB1, HNRNPA0, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HOOK3, HSP90AA1, HSP90AB1, HSP90B1, HSPA5, HSPA8, HSPA9, HSPD1, HTATSF1, IGF2BP3, ILF2, ILF3, ILKAP, ISY1, KHDRBS1, KHSRP, KIF11, KPNA4, KPNB1, KRI1, KTN1, LARP1, LARP4, LARP4B, LCP1, LIN28B, LMNB1, LRPPRC, LRRFIP1, LSM14A, LSM14B, LSM3, LSM6, LSM8, LUC7L, LUC7L3, MACF1, MAP1A, MAP4, MAPRE1, MATR3, MCM2, MEX3C, MEX3D, MIA3, MMTAG2, MOV10, MRPL1, MRPS27, MRPS28, MRPS31, MSI2, MSN, MTPAP, MYEF2, MYH9, NASP, NCBP2, NCL, NCOA5, NKAP, NKTR, NOLC1, NONO, NOP56, NOP58, NPM1, NSUN2, NUDC, NUMA1, OLA1, P4HB, PA2G4, PABPC1, PABPC4, PABPN1, PAIP1, PAK2, PCBP1, PCBP2, PCBP3, PDAP1, PDCD11, PDCD6IP, PDIA4, PEBP1, PEG10, PFN1, PIK3R1, PKM, PKP1, PLEC, PNISR, PNN, PNPT1, POLR2A, POLR2B, PPIE, PPIG, PPIL4, PRKDC, PRMT1, PRPF38A, PRPF38B, PRPF40A, PRPF4B, PRPF8, PRRC2A, PRRC2C, PSME2, PSPC1, PTBP1, PTBP2, PTCD3, PTMA, PUF60, PUM1, PUM2, PURA, PURB, PUS1, PUS7, PYM1, R3HDM1, R3HDM2, RACK1, RALY, RANBP1, RBBP6, RBM10, RBM12, RBM12B, RBM14, RBM15, RBM15B, RBM17, RBM22, RBM26, RBM3, RBM33, RBM39, RBM4, RBM45, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMX, RBMX2, RC3H1, RCC2, RCSD1, RNMT, RP9, RPA1, RPL10A, RPL13, RPL18, RPL22, RPL28, RPL34, RPL6, RPS11, RPS14, RPS2, RPS28, RPS3, RPS7, RPS8, RPSA, RRBP1, RSRC1, RSRC2, RTF1, RYDEN, SAFB, SAFB2, SASH3, SBDS, SCAF4, SDE2, SERBP1, SET, SF1, SF3A1, SF3B1, SF3B4, SFPQ, SFSWAP, SLBP, SLC4A1AP, SLIRP, SLTM, SND1, SNRNP200, SNRNP70, SON, SPEN, SPTAN1, SPTBN1, SRRM1, SRRM2, SRRT, SRSF1, SRSF10, SRSF11, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSB, SSBP1, SSRP1, ST13|ST13P5, STAU1, STAU2, STIP1, STMN1, STRAP, STRBP, SUB1, SUGP1, SUMO2, SUPT16H, SUPT5H, SUPT6H, SYNCRIP, SYNE2, TACC1, TAF15, TARDBP, TBRG4, TCOF1, TGM1, TGM3, THRAP3, THUMPD1, TIA1, TIAL1, TIPRL, TLN1, TNPO1, TNRC6B, TNRC6C, TOP1, TOP2A, TPM3, TPM4, TPR, TRA2A, TRA2B, TRNAU1AP, TUBB, TUFM, U2AF2, U2SURP, UBAP2, UBAP2L, UPF1, USP10, USP14, USP42, VCP, VIM, WDR33, XRCC5, XRCC6, XRN2, YBX1, YBX3, YKT6, YLPM1, YTHDF1, YTHDF2, YTHDF3, YWHAB, YWHAE, YWHAQ, YWHAZ, ZC3H10, ZC3H11A, ZC3H12A, ZC3H18, ZC3H4, ZC3H7A, ZC3H7B, ZC3HAV1, ZCCHC17, ZCCHC8, ZFR, ZNF326, ZRANB2

Supplementary table 3: Categorization of the RBPs unique to this T-ALL RNA interactome capture into proteins known to be involved in RNA-binding or related, nucleic acid binding or related and into proteins not related.

RNA-binding or related	Nucleic acid binding or related	not related
DDX3Y, EIF3CL, ELAVL4, ASCC1, EEF1A1, ANXA1, EIF1AY, TNRC6C, ZC3H12A, APOBEC3G, HNRNPLL, MMTAG2, RYDEN I C19of66, CNOT, DNAJC8, CIR1, YWHAQ, USP14, C9ORF78, PKP1	CBX5, HCLS1, HIST1H3A, KIF11, MCM2, PAK2, CDK11B, SDE2	DNAJC9, EFHD2, EVL, RCSD1, ARHGDIA, SASH3, SET, TIPRL, YKT6, ARGLU1, CAB39, CAPZB, CASP14, CDC37, CENPF, CHMP2A, C1ORF52, C11ORF57, C11ORF58, C7ORF55, CCDC25, CWC25, DPYSL2, DBN1, HDGFRP2, HOOK3, ILKAP, KPNA4, LMNB1, MIA3, NKTR, NASP, NUDC, NUMA1, PIK3R1, PSME2, RANBP1, ST13, STMN1, TACC1, TPM3, TPM4, CEP85, MAP1A, TGM3

Supplementary table 4: High confidence RBPs identified in SAHA-treated (5 μM, 6 h) T-ALL cell lines (+UV/-UV ratio of at least two-fold, FDR < 0.05).

P12
ACIN1, ACTN4, ADAR, AGO2, AKAP1, AKAP8, ANKHD1, ANXA1, ANXA11, ANXA2, API5, APOBEC3G, AQR, ARGLU1, ARHGDIA, ARHGDIB, ARL6IP4, ARPP21, ASCC1, ATP5A1, ATXN2, ATXN2L, BAG4, BCLAF1, BRD2, BUD13, BZW1, C11ORF57, C11ORF58|SMAP, C1ORF52, C7ORF55|LUC7L2, C9ORF78, CAB39, CACTIN, CAPRIN1, CAPZB, CBX5, CCAR1, CCAR2, CCDC25, CCDC9, CDC37, CDK11B|CDC2L1, CELF1, CELF2, CENPF, CEP85, CHD1, CHD4, CHERP, CHMP2A, CIR1, CIRBP, CNOT1, CNOT3, CNOT4, CPSF6, CPSF7, CSDE1, CSTF2T, CSTF3, CWC25, CWF19L2, DAZAP1, DBN1, DBNL, DBR1, DDX1, DDX17, DDX19A, DDX21, DDX23, DDX39B, DDX3X, DDX42, DDX46, DDX5, DDX6, DEK, DHX15, DHX30, DHX8, DHX9, DIAPH1, DKC1, DNAJC8, DNAJC9, DPYSL2, DUT, EDC4, EEF1A1|EEF1A1P5, EEF2, EFHD2, EFTUD2, EIF1AY|EIF1AX, EIF2S1, EIF3A, EIF3B, EIF3CL|EIF3C, EIF3E, EIF3G, EIF3H, EIF3J, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4ENIF1, EIF4G1, EIF4G2, EIF4H, EIF5B, ELAVL1, ELAVL4, EPRS, ESRP1, ETF1, EVL, EZR, FAM120A, FAM133B, FAM50A, FASTKD2, FKBP4, FLNA, FLNB, FMR1, FTSJ3, FUBP1, FUBP3, FUS, FXR1, FXR2, FYB, G3BP1, G3BP2, GARS, GDI2, GGCT, GIGYF2, GOLGB1, GPATCH1, GPATCH8, GRB2, GRSF1, GSPT1, GSPT2, HARS2, HCLS1, HDGFRP2, HDLBP, HELZ, HIST1H1B, HIST1H1D, HIST1H1E, HMGB1, HNRNPA0, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HOOK3, HSP90AA1, HSP90AB1, HSP90B1, HSPA5, HSPA8, HSPA9, HSPD1, HTATSF1, IGF2BP3, ILF2, ILF3, ILKAP, ISY1, KHDRBS1, KHSRP, KIF11, KPNA4, KPNB1, KRI1, KTN1, LARP1, LARP4, LARP4B, LCP1, LIN28B, LMNB1, LRPPRC, LRRFIP1, LSM14A, LSM14B, LSM3, LSM6, LSM8, LUC7L, LUC7L3, MACF1, MAP1A, MAP4, MAPRE1, MATR3, MCM2, MEX3C, MEX3D, MIA3, MMTAG2, MOV10, MRPL1, MRPS27, MRPS28, MRPS31, MSI2, MSN, MTPAP, MYEF2, MYH9, NASP, NCBP2, NCL, NCOA5, NKAP, NKTR, NOLC1, NONO, NOP56, NOP58, NPM1, NSUN2, NUDC, NUMA1, OLA1, P4HB, PA2G4, PABPC1, PABPC4, PABPN1, PAIP1, PAK2, PCBP1, PCBP2, PCBP3, PDAP1, PDCD11, PDCD6IP, PDIA4, PEBP1, PEG10, PFN1, PIK3R1, PKM, PLEC, PNISR, PNN, PNPT1, POLR2A, POLR2B, PPIE, PPIG, PPIL4, PRKDC, PRMT1, PRPF38A, PRPF38B, PRPF40A, PRPF4B, PRPF8, PRRC2A, PRRC2C, PSME2, PSPC1, PTBP1, PTBP2, PTCD3, PUF60, PUM1, PUM2, PURA, PURB, PUS1, PUS7, PYM1, R3HDM1, R3HDM2, RACK1, RALY, RANBP1, RBBP6, RBM10, RBM12, RBM12B, RBM14, RBM15, RBM15B, RBM17, RBM22, RBM26, RBM3, RBM33, RBM39, RBM4, RBM45, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMX, RBMX2, RC3H1, RCC2, RCSD1, RNMT, RP9, RPA1, RPL10A, RPL13, RPL18, RPL22, RPL28, RPL34, RPL6, RPS11, RPS14, RPS2, RPS28, RPS3, RPS7, RPS8, RPSA, RRBP1, RSRC1, RSRC2, RTF1, RYDEN, SAFB, SAFB2, SASH3, SBDS, SCAF4, SDE2, SERBP1, SET, SF1, SF3A1, SF3B1, SF3B4, SFPQ, SFSWAP, SLBP, SLC4A1AP, SLIRP, SLTM, SND1, SNRNP200, SNRNP70, SON, SPEN, SPTAN1, SPTBN1, SRRM1, SRRM2, SRRT, SRSF1, SRSF10, SRSF11, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSB, SSBP1, SSRP1, ST13|ST13P5, STAU1, STAU2, STIP1, STMN1, STRAP, STRBP, SUB1, SUGP1, SUMO2, SUPT16H, SUPT5H, SUPT6H, SYNCRIP, SYNE2, TACC1, TAF15, TARDBP, TBRG4, TCOF1, THRAP3, THUMPD1, TIA1, TIAL1, TIPRL, TLN1, TNPO1, TNRC6B, TNRC6C, TOP1, TOP2A, TPM3, TPM4, TPR, TRA2A, TRA2B, TRNAU1AP, TUBB, TUFM, U2AF2, U2SURP, UBAP2, UBAP2L, UPF1, USP10, USP14, USP42, VCP, VIM, WDR33, XRCC5, XRCC6, XRN2, YBX1, YBX3, YKT6, YLPM1, YTHDF1, YTHDF2, YTHDF3, YWHAB, YWHAE, YWHAQ, YWHAZ, ZC3H10, ZC3H11A, ZC3H12A, ZC3H18, ZC3H4, ZC3H7A, ZC3H7B, ZC3HAV1, ZCCHC17, ZCCHC8, ZFR, ZNF326, ZRANB2

DND41
TPR, HNRNPA2B1, HNRNPR, SF3A1, SFPQ, TIAL1, HNRNPF, SYNCRIP, HNRNPC, WDR33, NONO, ELAVL1, KHSRP, GRSF1, HNRNPK, CELF1, CSDE1, DDX1, HNRNPM, STAU2, HNRNPA3, MRPS28, FUBP3, CAPRIN1, HNRNPH1, FAM120A, TIA1, RBM17, CPSF6, CAB39, CENPF, STAU1, CELF2, SND1, FMR1, ZC3H11A, EIF4A1, RBM39, PSPC1, HNRNPL, FXR1, GRB2, SRSF11, CHD4, ZC3HAV1, RBM14, FUS, HNRNPH3, SF1, PUF60, SUGP1, GPATCH1, DDX23, ISY1, FLNA, HNRNPU, DDX19A, LRPPRC, PCBP2, FLNB, G3BP2, RBM12, PABPC4, FAM133B, STMN1, RALY, HNRNPDL, PRPF40A, EIF2S1, HNRNPH2, RBBP6, CACTIN, PUS1, TARDBP, DDX17, PUM1, ARPP21, PTBP2, KHDRBS1, RBM15, THRAP3, PDCD6IP, API5, C9ORF78, PRMT1, PTBP1, ZCCHC8, PRPF8, ZC3H7A, SPEN, PPIE, PURA, NSUN2, UBAP2L, TAF15, LARP4, RBM10, ATXN2, CDK11B|CDC2L1, GSPT2, TRA2A, FXR2, HNRNPLL|HNRPLL, EIF3G, PRKDC, HNRNPA1, PIK3R1, ARGLU1, HNRNPUL1, GPATCH8, PURB, TNRC6B, SRSF1, EIF4H, CNOT1, PUM2, PA2G4, UBAP2, EIF4G2, EIF4B, PCBP1, EZR, CIR1, CCAR2, PABPC1, ILF3, SRSF3, DKC1, ACIN1, NOP56, CSTF2T, RACK1, DDX6, SRSF9, HDLBP, U2AF2, GIGYF2, ATXN2L, NCL, DNAJC9, KRI1, MATR3, SRSF7, PDAP1, SERBP1, USP10, LARP1, DHX30, SFSWAP, EIF4G1, EIF4ENIF1, LUC7L, R3HDM2, BRD2, DDX42, SSBP1, UPF1, TOP2A, RBM8A, YTHDF2, MIA3, EIF4A3, PRPF4B, SON, CSTF3, PPIG, EIF1AY|EIF1AX, EIF3A, RBM26, YLPM1, SAFB2, RBM15B, ARL6IP4, LARP4B, ZFR, AKAP1, RPS3, MTPAP, AQR, RBM6, ELAVL4, LMNB1, PNN, FUBP1, BCLAF1, NKAP, NOP58, DEK, RBMX, ZCCHC17, HNRNPD, SRSF6, ETF1, SRSF2, HOOK3, THUMPD1, MMTAG2, TLN1, DDX46, FASTKD2, CHMP2A, RBM5, TRA2B, CBX5, CCDC9, DBNL, SSB, MRPS31, PRPF38A, RBMX2, HDGFRP2, CIRBP, FKBP4, MOV10, CEP85, MRPS27, SRRM1, RBM12B, SSRP1, SRRT, RBMS1, TBRG4, SF3B4, RP9, CHERP, SUPT5H, RPA1, CNOT4, SLBP, RBM3, SET, PEBP1, ZRANB2, RC3H1, DBR1, PNPT1, TCOF1, STRBP, G3BP1, RNMT, PTCD3, MYEF2, KPNA4, ARHGDIB, RBM22, CHD1, GSPT1, DDX5, APOBEC3G, ACTN4, SNRNP70, SF3B1, GARS, SPTAN1, STRAP, LSM14A, POLR2A, YTHDF1, DUT, SAFB, YBX1, ZC3H12A, MEX3D, HCLS1, SRSF10, SRSF4, ADAR, IGF2BP3, EIF3H, USP42, HARS2, XRCC5, TRNAU1AP, RBM45, FAM50A, AGO2, SUB1, C1ORF52, MSN, CWF19L2, U2SURP, LUC7L3, RTF1, EFTUD2, CAPZB, RPS11, P4HB, MEX3C, SUPT6H, TNRC6C, SRRM2, TPM4, NKTR, LSM6, ILKAP, C7ORF55|LUC7L2, EIF5B, DDX21, NUDC, POLR2B, AKAP8, HSPA5, YTHDF3, ST13|ST13P5, GDI2, ESRP1, HNRNPUL2, HSP90B1, R3HDM1, RPS14, CWC25, NUMA1, RSRC1, PAIP1, C11ORF58|SMAP, PRPF38B, SLIRP, MAP4, DDX3X, SPTBN1, SNRNP200, HNRNPAB, DDX39B, MACF1, PNISR, ZC3H4, LSM3, ASCC1, DHX8, ZC3H7B, YWHAE, OLA1, RCSD1, SLTM, HSPA8, HELZ, RPL28, XRN2, LIN28B, RPSA, EDC4, PABPN1, EIF3CL|EIF3C, MRPL1, HIST1H3A|HIST3H3|HIST2H3A, ZNF326, RPS28, SBDS, DHX9, FTSJ3, HSPA9, SUMO2, NOLC1, TNPO1, CCDC25, SYNE2, EVL, EIF4A2, TUBB, PRRC2C, RBM33, YWHAQ, RPS2, YKT6, PRRC2A, MAP1A, TIPRL, MSI2, RBM4B, LRRFIP1, LCP1, RPL22, DHX15, EEF2, SRSF5, YBX3, PAK2, EFHD2, PYM1, CNOT3, NPM1, ILF2, TOP1, PSME2, PPIL4, PFN1, C11ORF57, DAZAP1, VIM, SCAF4, RBM4, DPYSL2, RYDEN, EIF3B, LSM14B, ANKHD1, HIST1H1B, HNRNPA0, EIF3J, NASP, YWHAZ, NCOA5, CPSF7, MYH9, PCBP3, USP14, BUD13, RPL13, HIST1H1E, PEG10, RSRC2, SUPT16H, ZC3H10, DBN1, ATP5A1, HMGB1, HSP90AB1, RPS8, TACC1, RCC2, PDCD11, BAG4, RPL18, CCAR1, EPRS, HSPD1, DNAJC8, STIP1, RPL6, TUFM, KTN1, EIF3E, ANXA1, GGCT, SLC4A1AP, EEF1A1|EEF1A1P5, TPM3, XRCC6, PDIA4, SASH3, HIST1H4A, MAPRE1, ARHGDIA, HIST1H1C, HSP90AA1, RPL34, HIST1H1D, ZC3H18, LSM8, RPS7, HTATSF1, JUP, KPNB1, RANBP1, ALYREF, FYB, DSC1, CDC37, PKP1, BZW1, DSG1, SDE2, YWHAB, GOLGB1, PLEC, PUS7
MOLT3
ACIN1, ACTN4, ADAR, AGO2, AKAP1, AKAP8, ALDOA, ALYREF, ANKHD1, ANXA1, ANXA11, API5, APOBEC3G, AQR, ARGLU1, ARHGDIA, ARHGDIB, ARL6IP4, ARPP21, ASCC1, ATP5A1, ATXN2, ATXN2L, BAG4, BCLAF1, BRD2, BUD13, BZW1, C11ORF57, C11ORF58|SMAP, C1ORF52, C7ORF55|LUC7L2, C9ORF78, CAB39, CACTIN, CAPRIN1, CAPZB, CBX5, CCAR1, CCAR2, CCDC25, CCDC9, CDC37, CDK11B|CDC2L1, CELF1, CELF2, CENPF, CEP85, CHD1, CHD4, CHERP, CHMP2A, CIR1, CIRBP, CNOT1, CNOT3, CNOT4, CPSF6, CPSF7, CSDE1, CSTF2T, CSTF3, CWC25, CWF19L2, DAZAP1, DBN1, DBNL, DBR1, DDX1, DDX17, DDX19A, DDX21, DDX23, DDX39B, DDX3X, DDX3Y, DDX42, DDX46, DDX5, DDX6, DEK, DHX15, DHX30, DHX8, DHX9, DIAPH1, DKC1, DNAJC8, DNAJC9, DPYSL2, DSC1, DSG1, DUT, EEF1A1|EEF1A1P5, EEF2, EFHD2, EFTUD2, EIF1AY|EIF1AX, EIF2S1, EIF3A, EIF3B, EIF3CL|EIF3C, EIF3E, EIF3G, EIF3H, EIF3J, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4ENIF1, EIF4G1, EIF4G2, EIF4H, EIF5B, ELAVL1, ELAVL4, EPRS, ESRP1, ETF1, EVL, EZR, FAM120A, FAM133B, FAM50A, FASTKD2, FKBP4, FLNA, FLNB, FMR1, FTSJ3, FUBP1, FUBP3, FUS, FXR1, FXR2, FYB, G3BP1, G3BP2, GARS, GDI2, GGCT, GIGYF2, GOLGB1, GPATCH1, GPATCH8, GRB2, GRSF1, GSPT1, GSPT2, HARS2, HCLS1, HDGFRP2, HDLBP, HELZ, HIST1H1B, HIST1H1D, HIST1H1E, HIST1H3A|HIST3H3|HIST2H3A, HIST1H4A, HMGB1, HNRNPA0, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HOOK3, HSP90AA1, HSP90AB1, HSP90B1, HSPA5, HSPA8, HSPA9, HSPD1, HTATSF1, IGF2BP3, ILF2, ILF3, ILKAP, ISY1, KHDRBS1, KHSRP, KIF11, KPNA4, KPNB1, KRI1, KTN1, LARP1, LARP4, LARP4B, LCP1, LIN28B, LMNB1, LRPPRC, LRRFIP1, LSM14A, LSM14B, LSM3, LSM6, LSM8, LUC7L, LUC7L3, MACF1, MAP1A, MAP4, MAPRE1, MATR3, MCM2, MEX3C, MEX3D, MIA3, MMTAG2, MOV10, MRPL1, MRPS27, MRPS28, MRPS31, MSI2, MSN, MTPAP, MYEF2, MYH9, NASP, NCBP2, NCL, NCOA5,

NKAP, NKTR, NOLC1, NONO, NOP56, NOP58, NPM1, NSUN2, NUDC, NUMA1, OLA1, P4HB, PA2G4, PABPC1, PABPC4, PABPN1, PAIP1, PAK2, PCBP1, PCBP2, PCBP3, PDAP1, PDCD11, PDCD6IP, PDIA4, PEBP1, PEG10, PFN1, PIK3R1, PKM, PLEC, PNISR, PNN, PNPT1, POLR2A, POLR2B, PPIE, PPIG, PPIL4, PRKDC, PRMT1, PRPF38A, PRPF38B, PRPF40A, PRPF4B, PRPF8, PRRC2A, PRRC2C, PSME2, PSPC1, PTBP1, PTBP2, PTCD3, PUF60, PUM1, PUM2, PURA, PURB, PUS1, PUS7, PYM1, R3HDM1, R3HDM2, RACK1, RALY, RANBP1, RBBP6, RBM10, RBM12, RBM12B, RBM14, RBM15, RBM15B, RBM17, RBM22, RBM26, RBM3, RBM33, RBM39, RBM4, RBM45, RBM4B, RBM5, RBM6, RBM8A, RBMS1, RBMX, RBMX2, RC3H1, RCC2, RCSD1, RNMT, RP9, RPA1, RPL13, RPL18, RPL22, RPL28, RPL34, RPL6, RPS11, RPS14, RPS2, RPS28, RPS3, RPS7, RPS8, RPSA, RRBP1, RSRC1, RSRC2, RTF1, RYDEN, SAFB, SAFB2, SASH3, SBDS, SCAF4, SDE2, SERBP1, SET, SF1, SF3A1, SF3B1, SF3B4, SFPQ, SFSWAP, SLBP, SLC4A1AP, SLIRP, SLTM, SND1, SNRNP200, SNRNP70, SON, SPEN, SPTAN1, SPTBN1, SRRM1, SRRM2, SRRT, SRSF1, SRSF10, SRSF11, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SSB, SSBP1, SSRP1, ST13|ST13P5, STAU1, STAU2, STIP1, STMN1, STRAP, STRBP, SUB1, SUGP1, SUMO2, SUPT16H, SUPT5H, SUPT6H, SYNCRIP, SYNE2, TACC1, TAF15, TARDBP, TBRG4, TCOF1, TGM3, THRAP3, THUMPD1, TIA1, TIAL1, TIPRL, TLN1, TNPO1, TNRC6B, TNRC6C, TOP1, TOP2A, TPM3, TPM4, TPR, TRA2A, TRA2B, TRNAU1AP, TUBB, TUFM, U2AF2, U2SURP, UBAP2, UBAP2L, UPF1, USP10, USP14, USP42, VIM, WDR33, XRCC5, XRCC6, XRN2, YBX1, YBX3, YKT6, YLPM1, YTHDF1, YTHDF2, YTHDF3, YWHAB, YWHAE, YWHAQ, YWHAZ, ZC3H10, ZC3H11A, ZC3H12A, ZC3H18, ZC3H4, ZC3H7A, ZC3H7B, ZC3HAV1, ZCCHC17, ZCCHC8, ZFR, ZNF326, ZRANB2

Supplementary table 5: High confidence RBPs identified in thymocytes of two individuals (+UV/-UV ratio of at least two-fold, FDR < 0.05).

Thymocyte RNA interactome
AHNAK, API5, ARPP21, ATXN2, ATXN2L, C7ORF55|LUC7L2, CAPRIN1, CAPZB, CCAR2, CELF1, CELF2, CIRBP, CSDE1, DDX17, DDX39B, DDX3X, DDX46, DDX5, DHX9, EIF4A1, EIF4A2, EIF4A3, EIF4B, EIF4H, ELAVL1, EVL, EZR, FABP5, FUBP1, FUBP3, FXR1, GRSF1, GSPT1, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPLL|HNRPLL, HNRNPM, HNRNPR, HNRNPU, HNRNPUL1, HNRNPUL2, HSPB1, IGF2BP3, ILF2, ILF3, KHDRBS1, KHSRP, LARP1, LRPPRC, LSM14A, LUC7L3, MATR3, MYEF2, NCL, NONO, PA2G4, PABPC1, PABPC4, PABPN1, PCBP1, PPIL4, PRPF38B, PRPF40A, PRPF4B, PSPC1, PTBP1, PUF60, PUM1, PUM2, R3HDM1, RALY, RBM10, RBM14, RBM15, RBM39, RBM45, RBM6, RBMX, RYDEN, SF1, SF3A1, SFPQ, SND1, SNRNP70, SRSF1, SRSF11, SRSF6, SRSF7, SSB, SSBP1, STAU2, SUPT5H, SYNCRIP, TAF15, TARDBP, TCOF1, THRAP3, THUMPD1, TIA1, TIAL1, U2AF2, UBAP2, UBAP2L, UPF1, XRCC5, YBX1, YTHDF2, ZC3H18, ZC3H4