UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE

IN MEDICINA MOLECOLARE

CURRICULUM IN FISIOPATOLOGIA CELLULARE E MOLECOLARE

Dissecting the oncogenic mechanism of DUX4-IGH in acute lymphoblastic leukemia

DoS: Dr. Davide Gabellini

Second Supervisor: Prof. Jan Cools

Tesi di DOTTORATO di RICERCA di Daniele Campolungo

matr. 013902

Ciclo di dottorato XXXIV

SSD: BIO/11

Anno Accademico 2020/2021

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il/la sottoscritto/*I* Campolungo Daniele

Matricola / *registration number* 013902

nat_ a/ *born at* Recanati

il/*on* 10/06/1994

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

Dissecting the oncogenic mechanism of DUX4-IGH in acute lymhpoblastic leukemia

☒ AUTORIZZA la Consultazione della tesi / AUTHORIZES *the public release of the thesis*

☐ NON AUTORIZZA la Consultazione della tesi per …… mesi /DOES NOT AUTHORIZE *the public release of the thesis for …… months*

a partire dalla data di conseguimento del titolo e precisamente / *from the PhD thesis date, specifically*

Dal / *from* ……/……/…… Al / *to* ……/……/……

Poiché /*because*:

□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

□ ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

□ la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

Si rende noto che parti della tesi sono indisponibili in relazione all'utilizzo di dati tutelati da segreto industriale **(da lasciare solo se applicabile)** /*Please Note: some parts of the thesis are not available in relation to the norm of the use of information protected by trade secret* ***(To leave only if relevant)***

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data /Date 10/02/2022 Firma /Signature Daniele Campolungo

DECLARATION

This thesis has been composed by myself and has not been used in any previous application for a degree. Throughout the text I use both ‘I’ and ‘We’ interchangeably.

All the results presented here were obtained by myself, except for:

1) Bioinformatic analyses of RNA-sequencing data (Results, chapter 3.1.1, figure9 and chapter 3.1.4, figure 14) have been performed by Dr Anna Sofia Tascini, Center of Omics Sciences, San Raffaele Scientific Institute, Milan, Italy.

All sources of information are acknowledged by means of reference

ABSTRACT

In up to 10% of Acute Lymphoblastic Leukemia (ALL) patients, the disease is caused by translocations of the double homeobox 4 (DUX4) gene into the immunoglobulin heavy chain (IGH) locus, resulting in the overexpression in B-cell precursors of the DUX4-IGH fusion protein with an aberrant C-terminus. DUX4 is a transcription factor normally expressed in the germline and early embryo. Its aberrant reactivation in somatic cells mediates cellular toxicity by activating a pro-apoptotic transcriptional program. In contrast, DUX4-IGH displays transforming abilities and its expression induces leukemia. To date, the molecular mechanism through which DUX4-IGH induces and maintains leukemia remains poorly understood, impairing the possibility to develop targeted therapies.

By coupling Genome-wide transcriptomics (RNA-seq) and DNA-binding analyses (Cut&Tag) in B-ALL cells, we found that DUX4 and DUX4-IGH bind to and activate different gene sets, despite sharing the same DNA-binding domain. In contrast with the pro-apoptotic targets activated by DUX4, DUX4-IGH drives the expression of genes involved in cell adhesion and migration. Intriguingly, while DUX4 is similarly active in any cell type, DUX4-IGH displays a B-cell restricted activity.

By performing tandem affinity purification followed by quantitative mass spectrometry, we discovered that DUX4-IGH selectively interacts with the transcription factor GTF2I, which is highly expressed in B cells. Small molecule-mediated downregulation of GTF2I abrogates the ability of DUX4-IGH to activate its target genes and promote tumor spheroids formation. Strikingly, forced GTF2I expression is sufficient to allow non-B cells to become DUX4-IGH responsive. Our results support a novel mechanism of leukemogenesis by DUX4-IGH which requires selective interaction with GTF2I via its novel C-terminus generated upon rearrangement which is key to activate genes required for leukemogenesis, laying the basis for the development of new therapeutic approach to specifically target the activity of the fusion transcription factor.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

ALL: acute lymphoblastic leukemia

AML: acute myeloid leukemia

ATAC-seq: assay for transposase-accessible chromatin sequencing

AYA: adolescents and young adults

CAR-T: chimeric antigen receptor T-cell

CBP: cAMP-response element binding protein

CIC: containing protein capicua

CLL: chronic lymphocytic leukemia

CML: chronic myelogenous leukemia

CNS: central nervous system

CTA: cancer Testis Antigens

CNV: copy number variation

CR: complete remission

ChIP-seq: chromatin immunoprecipitation sequencing

CUT&Tag: cleavage under targets & tagmentation

CUTAC: cleavage under targeted accessible chromatin

DUX: double homeobox

DUX4: Double homeobox 4

DUX4-r: rearranged DUX4

EFS: event-free survival

ERVs: endogenous retrovirus elements

ETV: E-Twenty six Variant

FISH: fluorescence in-situ hybridization

FSHD: facioscapulohumeral muscular dystrophy

GEP: gene expression profile

GWAS: genome-wide association studies

HMG: high mobility group box

HSATII: human satellite II

iAMP21: intrachromosomal amplification of chromosome 21

IGH: immunoglobulin heavy chain

MaLR: mammalian apparent LTR-retrotransposons

MRD: minimal residual disease

NGS: next-generation sequencing

OS: overall survival

p300: E1A binding protein P300

RNA-seq: Ribonucleic acid sequencing

swALL: enriched lineage switching

TAD: transcription activation domain

TAP-MS: tandem affinity purification and mass spectrometry

WTS: whole transcriptome sequencing

ZGA: zygotic genome activation

LIST OF FIGURES AND TABLES

1. INTRODUCTION

1.1 ACUTE LYMPHOBLASTIC LEUKEMIA

1.1.1 Key Leukemia Statistics

Leukemia includes a group of blood cancers, usually originating in the bone marrow and resulting in the production of high numbers of abnormally developed blood cells called blasts or leukemia cells. Nearly 500.000 leukemia cases are diagnosed each year worldwide, causing more than 300.000 annual deaths (https://gco.iarc.fr/today/online-analysis-table). The vast majority of leukemia cases, in particular acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myelogenous leukemia (CML), occurs mainly in adults, whereas acute lymphoblastic leukemia (ALL) is prevalent in children and young adults (Stefania *et al*, 2015). Accounting for almost 1 out of 3 cancers, ALL represents the most common pediatric malignancy and the most frequent cause of death from cancer at young age (Bunin *et al*, 1996; Linet *et al*, 1999). The incidence of ALL follows a bimodal distribution, with the first peak occurring in childhood and a second peak occurring around the age of 50(Paul *et al*, 2016). While dose intensification strategies have led to a significant improvement in outcomes for pediatric patients, prognosis for the elderly remains very poor. Despite a high rate of response to chemotherapy, only 30–40% of adult patients with ALL will achieve long-term remission (Jabbour *et al*, 2015). Consequently, most deaths from ALL (about 4 out of 5) occur in adults, with cure rates below 40% despite pediatric-inspired chemotherapy regimens (Stock *et al*, 2019). While the prognosis for pediatric ALL cases has improved dramatically during the last decades resulting in today's cure rates of more than 90%, chemotherapy is associated with life-long health sequelae, relapse can be experience by up to 20% of cases and survival following relapse is poor (Hunger & Mullighan, 2015; Inaba & Mullighan, 2020).

1.1.2 Acute Lymphoblastic Leukemia

Like cancer in general, ALL arises as a direct result of mutations leading to uncontrolled self-renewal and proliferation, differentiation block, and decreased apoptosis of the leukemic cells (Hanahan & Weinberg, 2011) (Figure 1). In the case of leukemia, driving genetic lesions are acquired early during lymphoid development, resulting in block of differentiation at immature cellular stages. B-cell acute lymphoblastic leukemia (B-ALL) represents the most common ALL form and comprises more than 30 genetic subtypes characterized by founding chromosomal alterations including chromosomal aneuploidy (gains and losses of whole chromosomes), or chromosomal rearrangements that result in deregulation of genes by their juxtaposition to strong enhancers as well as the generation of chimeric fusion genes. These genetic alterations often converge in cellular pathways that block differentiation and promote the proliferation of immature B cells, thus defining the biology of ALL subtypes (Iacobucci & Mullighan, 2017; Iacobucci *et al*, 2021)

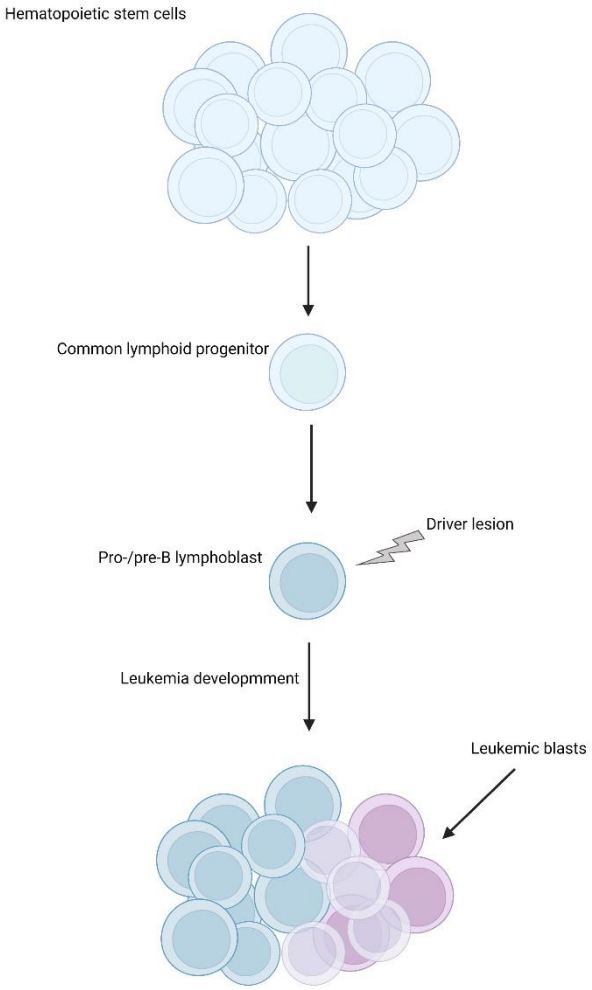


Figure 1. B-ALL development. *In ALL, the driver genetic lesion is acquired in early B progenitor cells, which cause block of differentiation and uncontrolled expansion of immature blasts in the bone marrow. Image created with Biorender.*

1.1.3 Genetic Risk Variants

There are several genetic factors associated with an increased risk of developing ALL. Down syndrome, familial cancer syndromes and specific DNA polymorphisms represent inherited susceptibilities, at least in a subset of ALL cases (Buitenkamp *et al*, 2014; Moriyama *et al*, 2015; Noetzli *et al*, 2015). In addition, genome-wide association studies (GWAS) have identified polymorphic variants in several genes associated with ALL. Risk variants are frequently at/near hematopoietic transcription factor or tumor suppressor genes, including *ARID5B, BAK1, CDKN2A/CDKN2B, BMI1-PIP4K2A, CEBPE, ELK3, ERG, GATA3, IGF2BP1, IKZF1, IKZF3, USP7*, and *LHPP* (Gocho & Yang, 2019; Mullighan, 2012; Perez-Andreu *et al*, 2013). Moreover, germline genomic analysis has identified additional susceptibility variants in sporadic ALL *(NBN, ETV6, FLT3, SH2B3*, and *CREBBP*), Down syndrome- associated B-ALL (*IKZF1, NBN, RTEL1*) and rare germline mutations in *PAX5* and *ETV6*, which are linked to familial ALL (de Smith *et al*, 2019; Qian *et al*, 2019b). Only few environmental risk factors are generally associated with ALL such radiation and certain chemicals(Hunger & Mullighan, 2015), but these associations explain only a very small minority of cases and the relative risk associated with each variant is typically low. On the other hand, accumulation of such mutations, particularly with increasing age, may result in an increased ALL risk. Noteworthy, despite these examples of genetic predisposition, in most cases ALL appears as a *de novo* malignancy in previously healthy individuals, as most patients have no recognized inherited factors.

1.1.4 Classification of ALL

Through the immunophenotype, that is the cell-surface and/or cytoplasmic expression of lineage markers, ALL is broadly classified into B-cell or T-cell lymphoblastic leukemia, reminiscent of normal lymphoid maturation stage. Accounting for nearly 80% of diagnosed cases, B-ALL represents the most common form of acute lymphoblastic leukemia in both young and adult population (Inaba *et al*, 2013; Roberts & Mullighan, 2020). B-ALL comprises over twenty distinct subtypes which are defined by recurrent, disease-initiating genetic abnormalities, such as chromosomal gains or losses as well as chromosomal rearrangements that deregulate oncogenes or encode

chimeric fusion oncoproteins (Gu *et al*, 2019; Pui *et al*, 2019; Schwab & Harrison, 2018; Li *et al*, 2021; Kimura & Mullighan, 2020). These subtype-defining genetic alterations vary according to age and ethnicity, are both somatic and/or germline and are associated with distinct gene expression patterns which converge on specific cellular pathways (Kimura & Mullighan, 2020; Li *et al*, 2021). Secondary mutations are characteristic of B-ALL and their nature and prevalence vary according to the disease subtype (Iacobucci & Mullighan, 2017; Qian *et al*, 2019a). These alterations may be acquired or enriched during disease progression and often perturb lymphoid development by affecting lymphoid transcription factors (IKZF1, PAX5, EBF1), tumor suppressors (CDKN2A/CDKN2B, RB1), and regulators of cell-cycle, apoptosis or transcription (ETV6, ERG), thereby influencing leukemogenesis and treatment response(Iacobucci & Mullighan, 2017; Mullighan *et al*, 2007a). Nonetheless, the biology of B-ALL is determined mainly by the first genetic hit, which is usually represented by chromosomal translocations and intrachromosomal rearrangements giving rise to chimeric proteins, often displaying different functions with respect to the wild type counterpart. For this reason, a deeper understanding of the molecular mechanism lying at the basis of leukemia development is needed for the development of efficient and specific therapies targeting key driver lesions in B-ALL.

1.1.5 Clinical Characteristics of B-ALL

Most B-ALL clinical manifestations reflect the accumulation of malignant, poorly differentiated lymphoid cells within the bone marrow, peripheral blood and extramedullary sites. Early signs of the disease can be non-specific and include fatigue, loss of appetite, bone pain and lymph nodes swelling, which are usually accompanied by signs of bone marrow failure such as anemia, thrombocytopenia and leukopenia (Berger *et al*, 2015). At later stages of the disease, involvement of extramedullary sites can cause lymphadenopathy, splenomegaly or hepatomegaly in 20% of patients (Jabbour *et al*, 2005). Central nervous system (CNS) involvement at time of diagnosis occurs in 5–8% of patients and presents most commonly as cranial nerve deficit or meningismus (Gu *et al*, 2016; Jabbour *et al*, 2015). In ALL, the hijacking of the hematopoietic system by leukemia blasts compromises the production of other blood

cell types essential for oxygen transport (erythrocytes) or coagulation (platelets), which leads to major insufficiencies in the patient's immunological and physiological system, ultimately resulting in death. Noteworthy, acute lymphoblastic leukemia has a much more rapid progression (weeks) compared to chronic leukemias (months).

1.1.6 Diagnostic, Prognostic and Therapeutic Implications of B-ALL

Diagnosis B-ALL diagnosis is established by the presence of 20% or more lymphoblasts in the bone marrow or peripheral blood (Bielorai *et al*, 2013), followed by the identification of genetic factors by conventional karyotyping, fluorescence in-situ hybridization (FISH) and targeted-molecular analyses, which have been historically used to classify ALL and to risk-stratify patients with the disease. However, the alterations thus identified do not establish the genetic basis of the disease for all cases, failing to reveal the nature of the genetic alterations driving leukemogenesis is a significant number of patients. Importantly, in recent years the rapid development and implementation of next-generation sequencing (NGS) techniques revolutionized the understanding of the B-ALL genomic landscape. In particular, integrative genome-wide sequencing studies allowed the identification and characterization of cryptic genetic alterations, structural DNA variations and gene expression signatures which define new B-ALL (Mullighan *et al*, 2008; Andersson *et al*, 2015; Tzoneva *et al*, 2018) subtypes.

Prognosis Different B-ALL subtypes converge on distinct gene expression patterns with prognostic and therapeutic significance. As an example, current risk stratification and treatment protocols incorporate age, sex, white blood cell count, established cytogenetic alterations, and response to initial therapy as measured by levels of minimal residual disease (MRD). This last aspect, in particular, represents a central component in risk stratification and sequencing-based approaches have also been used successfully to analyze antigen receptor rearrangements and quantitate MRD more sensitively than flow cytometric or conventional polymerase chain reaction–based approaches(Faham *et al*, 2012).

Treatment An improved knowledge concerning the genetic basis of ALL and the identification of dysregulated pathways associated with therapeutic targets has had a huge impact on treatment outcomes, mostly through the development of precision

medicine. The best example is provided by the constitutively active kinases in newly diagnosed BCR-ABL1 and BCR-ABL1-like B-ALL, which are now druggable by a variety of single or combinatorial TKIs (Slayton *et al*, 2018; Foà *et al*, 2011, 2020), as well as in combination with chemotherapy (Fielding *et al*, 2014). Nonetheless, chemotherapeutic treatment still represents the main treatment for ALL patients (Terwilliger & Abdul-Hay, 2017), and comprises three main phases: remission induction, consolidation (or intensification), and maintenance, with the goal to achieve complete remission and to restore normal hematopoiesis. Such treatment regimen, however, is associated with life-long health sequelae and more than 50% of adult and up to 20% of pediatric patients experience relapse (Jabbour *et al*, 2015).

A refined molecular classification could greatly improve the treatment strategy. The wide range of genetic alterations defining leukemia subtypes display distinct gene expression patterns, which converge on specific cellular pathways. The identification of such pathways is crucial for the management of the disease, allowing the discovery of new therapeutic vulnerabilities. In recent years, multiagent therapy regimens which include target inhibitors (e.g., imatinib), immunomodulators, monoclonal antibodies, and chimeric antigen receptor T-cell (CAR-T) therapy, are transforming the clinical practice from chemotherapy drugs to personalized medicine in the field of risk-directed disease management (Inaba *et al*, 2017; Li *et al*, 2021).

1.1.7 Primary and Relapsed B-ALL

Cure rates for pediatric ALL now exceeds 90%. However, outcomes for adolescents and young adults (AYA) and adults remain poor, since only up to 60% and 40% of them, respectively, will achieve long-term remission (Stock *et al*, 2014; Roberts, 2018; Jabbour *et al*, 2015; Terwilliger & Abdul-Hay, 2017). This could be partially explained by the fact that, despite treatment of adult ALL is largely modeled after the multiagent chemotherapy regimen utilized in pediatric ALL, adults with the disease tend to have higher risk features at the time of diagnosis, more comorbidities, and increasing age that often requires dose reductions (Sive *et al*, 2012; Terwilliger & Abdul-Hay, 2017; Jemal *et al*, 2004). Furthermore, despite cure rate of primary ALL has greatly improved with risk-adjusted therapy, relapsed ALL is still a leading cause of cancer-related death for

all ages mainly due to therapy resistance, usually induced by mutations acquired during disease progression, which are also influenced by the bone marrow microenvironment (Roberts, 2018). In particular, recent genomic analyses of paired primary and relapsed ALL show that, while predominant clones at diagnosis are often eradicated, acquired alterations drive drug-resistance in minor leukemic clones, which will eventually re-populate the leukemic niche (Schroeder *et al*, 2019). Hence, genomic and gene expression profiling of primary and relapsed B-ALL could reveal potential new therapeutic approaches to target specific vulnerabilities of many subtypes, several of which have been validated in preclinical models and are being formally evaluated for possible clinical trials (Kimura & Mullighan, 2020).

1.1.8 Genetic Basis of B-ALL

Prior the advent of NGS, techniques such as FISH and targeted molecular analyses were used to identify recurring chromosomal abnormalities including aneuploidy, chromosomal rearrangements and/or known gene fusions. However, in a large proportion of cases, historically denoted as B-other, recurrent B-ALL-associated genomic alterations remain unidentified. Genomic and transcriptomic profiling of B-ALL patients, in particular whole transcriptome sequencing (WTS), allowed the identification of several novel leukemia-driving genetic alterations emerging from the B-other subgroup (Gu *et al*, 2019; Li *et al*, 2018a; Liu *et al*, 2016). These include cryptic rearrangements not identifiable by conventional approaches, subtypes that "phenocopy" already established subtypes and share similar gene expression profile but display different driver alterations, as well as subtypes defined by a single point mutation. Below, I will briefly report the current molecular classification of B-ALL subtypes.

Chromosomal Aneuploidies

High hyperdiploidy (nonrandom gain of at least five chromosomes) is present in ~25% of childhood ALL patients, but accounts for <5% of adolescents and young adults (16–39 yr old; AYA) and adults and is associated with favorable outcome. Alterations involving the Ras pathway (*KRAS, NRAS, FTL3, PTPN11*) and epigenetic

modifiers (*CREBBP, WHSC1*) are frequent genetic events in high hyperdiploid (Paulsson *et al*, 2015).

Hypodiploid ALL comprises two subtypes with distinct transcriptional profiles and genetic alterations according to the severity of aneuploidy. Low-hypodiploid (31-39 chromosomes) subtype is highly rare in children (<1%) but increases with age, accounting for 5% of AYAs and over 10% of adults (Iacobucci *et al*, 2021; Roberts, 2018). At the genetic level it is characterized by common *IKZF2* deletions and *TP53* sequence mutations, and it is associated with a very poor outcome(Moorman *et al*, 2007). Near-haploid ALL (24-30 chromosomes) is characterized by Ras-activating mutations and *IKZF3* alterations, and accounts for ~2% of childhood ALL and <1% of AYAs and adults (Iacobucci *et al*, 2021). This subtype shows intermediate prognosis.

Intrachromosomal amplification of chromosome 21 (iAMP21) is more common in older children and accounts for 1% of childhood ALL(Harrison, 2015). Two germline genomic alterations, the Robertsonian translocation rob (15;21) and ring chromosome 21 are associated with elevated risk of iAMP2. This subtype shows poor outcome and high rate of relapse when treated as standard risk. However, intensive therapy can greatly improve the outcome (Moorman *et al*, 2013; Heerema *et al*, 2013).

Recurrent Chromosomal Translocations and Gene Fusions

KMT2A-rearranged on chromosome 11q23 to over 80 different partner genes define a subtype of leukemia with both lymphoid and myeloid features and associated with poor prognosis (El Chaer *et al*, 2020; Kimura & Mullighan, 2020). This subtype shows a distinct gene expression signature with overexpression of *HOX* cluster genes and HOX cofactors (Armstrong *et al*, 2002; Yeoh *et al*, 2002). Alteration of PI3K and Ras pathways is common in KMT2A-rearranged ALL (Gu *et al*, 2019; Andersson *et al*, 2015; Valentine *et al*, 2014). Moreover, KMT2A rearrangement is associated with altered chromatin patterning including H3K79 methylation, which has stimulated development of novel therapeutic approaches including inhibition of chromatin remodeling complexes (Winters & Bernt, 2017; Chen *et al*, 2015; Klossowski *et al*, 2020).

ETV6-RUNX1 translocation is the most common alteration in childhood ALL (25%) and is characterized by excellent prognosis. The ETV6-RUNX1 fusion is considered a leukemia-initiating alteration arising in utero, as demonstrated by the identification in umbilical cord blood (Sundaresh & Williams, 2017). However, the prolonged latency from birth to clinically manifest leukemia indicates that ETV6-RUNX1 requires cooperating genetic events to induce leukemia, consistent with the heterogeneity in the sub-clonal composition of ETV6-RUNX1 ALL (Mullighan *et al*, 2008; Sundaresh & Williams, 2017).

ETV6-RUNX1-like is defined by having a gene expression profile and immunophenotype (CD27 positive, CD44 low to negative) similar to ETV6-RUNX1 ALL but lacking the ETV6-RUNX1 fusion (Zaliova *et al*, 2017; Lilljebjörn *et al*, 2016). This alteration is most commonly found in children (3%) and is associated with intermediate to favorable prognosis.

E2A-rearranged encoding E2A-PBX1 by the t(1;19)(q23;p13) translocation, is present in ∼5% of children and less commonly found in AYAs and adults. Previously considered a high-risk subtype, it is now associated with a favorable outcome on contemporary ALL therapies (Barber *et al*, 2007; Burmeister *et al*, 2010). On the other hand, *E2A* translocation with the *HLF* gene defines a rare subtype of ALL (<1% in all ages) with a distinct transcriptional profile that is typically associated with an overall survival of <2 yr from diagnosis (Inaba *et al*, 1996). Interestingly, primary leukemic cells harboring E2A-HLF, but not E2A-PBX1, show sensitivity to the BCL2 inhibitor venetoclax (ABT-199), identifying a new therapeutic option for the more fatal subtype (Fischer *et al*, 2015).

BCR-ABL1 ALL is uncommon in children (2%–5% of patients), but accounts for at least 25% of adults (Roberts et al. 2014a, 2017a). Although historically considered a high-risk subtype, the incorporation of tyrosine kinase inhibitors (TKIs) into the standard treatment regimen significantly improved clinical outcomes for BCR-ABL1-positive ALL (Slayton *et al*, 2018; Foà *et al*, 2020, 2011). Secondary cooperative mutations are *IKZF1*, *PAX5* and *CDKN2A*/*B* deletions (Foà *et al*, 2011; Iacobucci *et al*, 2009; Mullighan *et al*, 2008), which have been associated with unfavorable outcome irrespective of TKI exposure (Slayton *et al*, 2018; Martinelli *et al*, 2009).

BCR-ABL1-like patients have similar transcriptional profiles to BCR-ABL+ ALL but lack the BCR-ABL1 fusion gene (Mullighan *et al*, 2009; Den Boer *et al*, 2009). It is a highly heterogenous subtype characterized by several rearrangements, copy number alterations, and sequence mutations that activate tyrosine kinase or cytokine receptor signaling. The frequency of such genetic events varies with age and this subtype can fall into four main groups based on the alterations. Similar to BCR-ABL1 ALL, the incidence, prevalence and outcome of Ph-like ALL increases with age (Reshmi *et al*, 2017; Roberts *et al*, 2014, 2017).

MEF2D-rearranged subtype displays the N-terminal portion of *MEF2D* fused to several partner genes, retaining its DNA binding domain (Liu *et al*, 2016; Kimura & Mullighan, 2020) and resulting in transcriptional activation of MEF2D targets (Gu *et al*, 2016). MEF2D-rearranged ALL is characterized by an aberrant immunophenotype (low or absent expression of CD10, high expression of CD38 and cytoplasmic μ chain), mature B-ALL-like morphology, and distinct expression profiles (Kentaro Ohki *et al*, 2019; Liu *et al*, 2016). Of importance, dysregulated MEF2D targets includes overexpression of HDAC9, which confers therapeutic sensitivity to HDAC inhibitors such as Panobinostat.

ZNF384-rearranged is fused with multiple, different partners. ZNF384-rearranged ALL displays a unique transcriptional signature and is often diagnosed as B-ALL with aberrant expression of the myeloid markers CD13 and/or CD33. This subtype accounts for ∼5% of children and up to 7% in AYA patients, and is associated with older age (Liu *et al*, 2016).

NUTM1-rearranged is a rare distinct subtype observed exclusively in children. In this subtype, *NUTM1* has been shown to be fused with multiple 5' partners, but most commonly with BRD9 (Femke M. Hormann *et al*, 2019; Boer *et al*, 2021). Given this involvement, bromodomain or HDAC inhibitors may result effective for specific targeting of these patients (Schwartz *et al*, 2011).

DUX4-IGH involves translocations of *DUX4* to the immunoglobulin heavy chain (*IGH*) locus as early leukemia-initiating event resulting in expression of a DUX4 isoform carrying the DNA-binding domain of DUX4 but lacking the normal C-terminal domain, which is commonly fused to random amino acids encoded by the *IGH* locus,

hence giving rise to chimeric proteins called DUX4-IGH (Zhang *et al*, 2016a; Yasuda *et al*, 2016; Lilljebjörn *et al*, 2016; Liu *et al*, 2016). This subtype comprises up to 10% of B-ALL patients, with a slight peak of incidence in AYAs, and is characterized by a very distinctive gene expression signature and immunophenotype (CD2± and CD371+) (Slamova *et al*, 2014; Schinnerl *et al*, 2019). DUX4-IGH was shown to bind to an intragenic region of the ETS transcription factor ERG, which results in its transcriptional deregulation and expression of multiple aberrant coding and non-coding *ERG* isoforms (Zhang *et al*, 2016a). Moreover, deletions of *ERG* have been reported in up to 70% of DUX4-IGH cases. Nonetheless, *ERG* deletions are commonly polyclonal and not present in all DUX4-IGH B-ALL cases, suggesting that it represents only a secondary event in the pathogenesis of the DUX4-IGH ALL subtype. For this reason, the molecular mechanism through which this oncogenic fusion drives and maintains leukemia is currently unclear, severely impairing the possibility to develop targeted therapies.

Since the role of DUX4-IGH in the pathogenesis of ALL represents the focus of my PhD project, a more detailed description of this subtype and the current knowledge concerning this rearrangement is provided below.

1.2 DOUBLE HOMEOBOX 4 (DUX4)

1.2.1 DUX4 Structure

Double homeobox 4 (*DUX4*) encodes for a sequence-specific transcription factor belonging to the family of *double homeobox* (*DUX*) genes, which is exclusive to placental mammals and includes the paralogs *DUXA*, *DUXB*, *DUXC* as well as the rodent *Dux* and *Duxbl* (Leidenroth & Hewitt, 2010; Clapp *et al*, 2007). The *DUX4* protein coding sequence is contained within each unit of the *D4Z4* macrosatellite repeats, which are present in up to 100 copies on the subtelomeric regions of human chromosomes 4q35 and 10q26 (Hewitt *et al*, 1994; Gabriëls *et al*, 1999; van der Maarel *et al*, 2007). At the structural level, DUX4 possesses an amino-terminal DNA binding domain, composed by two, closely spaced homeoboxes mediating DUX4 contact with the consensus DNA sequence 5'-TAATCTAATCA-3' (Lee *et al*, 2018). Interestingly, despite the two homeoboxes share a high similarity in their sequence, as well as in their structure (Li *et al*, 2018b; Lee *et al*, 2018), they bind in a head-to-head fashion to different core sequences 5'-TAAT-3' and 5'-TGAT-3' within the DUX4 consensus (Lee *et al*, 2018). At the carboxy terminal portion, DUX4 possess a transcription activation domain (TAD) through which it interacts with the

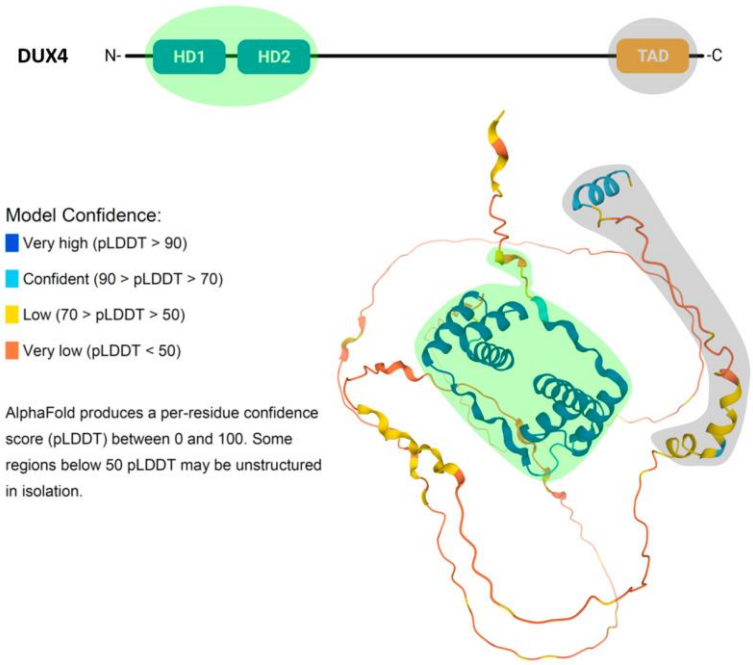


Figure 2. Double Homeobox 4 (DUX4). *Schematic representation DUX4 structural features: the two N-terminal homeoboxes (HD1 and HD2) are highlighted in green, while the C-terminal transcription activation domain TAD in grey. 3D structure prediction of DUX4 via AlphaFold denotes a well-defined tertiary structure of the DNA-binding domain and a less well-structured C-terminal TAD. The rest of the protein is predicted to be disordered. Modified from Mocciaro & Runfola et al. 2021.*

transcriptional coactivators CBP and p300, thus mediating DUX4 transcriptional activity (Geng *et al*, 2012; Bosnakovski *et al*, 2008a) (Figure 2).

1.2.2 Physiological Roles of DUX4

1.2.2.1 DUX4 in Embryo Development

Human DUX4 is transiently expressed in cleavage stage embryos, showing a peak at the 4-cell stage just before the onset of the zygotic genome activation (ZGA) (Hendrickson *et al*, 2017; De Iaco *et al*, 2017; Whiddon *et al*, 2017), which represents the first time during which the zygote transcribes its own genetic material (Vastenhouw *et al*, 2019). Subsequently, DUX4 gets strongly downregulated at both RNA and protein level at the 8-cell stage (Hendrickson *et al*, 2017; De Iaco *et al*, 2017) and remains epigenetically repressed in most somatic tissues. Dux, the DUX4 functional homolog in mouse, shows a similar expression pattern (Hendrickson *et al*, 2017; De Iaco *et al*, 2017). During this very short time-frame, DUX4 binds and activate transcription of cleavage-restricted genes, including *ZSCAN4*, *RFPLs*, *PRAME* and *TRIM* families(Whiddon *et al*, 2017; Hendrickson *et al*, 2017). DUX4 also drives expression of endogenous retroviruses, such as *HERVL*, which are known to be selectively transcribed at the cleavage stage. Similar results were obtained for Dux (Whiddon *et al*, 2017), raising the possibility that DUX proteins may have co-evolved with their targets, in different organisms, to regulate embryonic gene transcription. Given their germline-restricted nature, several DUX4 targets may induce adaptive immunity responses when aberrantly expressed in non-permissive environments. Such developmental-specific genes are often re-expressed in various malignancies (ex *PRAMEF* genes) (Malaguti *et al*, 2019), as part of Cancer Testis Antigens (CTA).

1.2.2.2 DUX4 in Somatic Tissues

While it is silent in most somatic tissues, DUX4 is normally expressed at relatively high levels in testis and thymus (Snider *et al*, 2010; Das & Chadwick, 2016). Although the exact function of DUX4 in these contexts is not known, a common aspect to testis and thymus is the high apoptosis rate characterizing developmental processes undergoing in these two tissues. Spermatogenesis is a complex process requiring the homeostasis of different cell types. Accordingly, testis development and sperm differentiation are characterized by high apoptosis rates: 75% of germ cells undergo

apoptosis at various stages (Shaha *et al*, 2010; Rodriguez *et al*, 1997). In thymus, one of the checkpoints required for the continuation of T cell development is the successful production of the T cell receptor β-chain (Kawazu *et al*, 2007b; Klein *et al*, 2019). Consequently, cells with a non-functional rearrangement are eliminated by apoptosis. Notably, the murine DUX family member Duxbl is selectively expressed in a short window of time during which the β-chain selection occurs and is involved in the induction of apoptosis in cells which failed the β-chain rearrangement (Kawazu *et al*, 2007a, 2007b). Testis and thymus are also associated with a defined immune status. Testis is an immunologically privileged site and thymus is involved in the production and maturation of immune cells, which could make it tolerant to several antigens. Whether DUX4 has a functional role in any of the above testis/thymus features is currently unknown.

1.2.3 Pathological Roles of DUX4

While the physiological roles of DUX4 are still under investigation, the consequences of its aberrant expression have been largely studied.

1.2.3.1 DUX4 in FSHD

Repeated sequences, such as the *D4Z4* macrosatellite repeats, are targeted by a number of repressive pathways including copy-number mediated silencing, DNA methylation and repressive histone modifications, which are responsible for maintaining *DUX4* epigenetically silenced in somatic cells (Himeda & Jones, 2019). However, because of their intrinsic instability, the deregulation of such genomic regions often disrupts normal cell physiology. In this respect, genetic defects associated with either deletion of a subset of *D4Z4* repeats or damaging variants in epigenetic regulators of *D4Z4* (*DNMT3B*, *LRIF1* and *SMCHD1*) leads to inefficient *DUX4* repression in skeletal muscle cells, whose expression is highly toxic causing muscle wasting and leading to facioscapulohumeral muscular dystrophy (FSHD). FSHD is one of the most prevalent neuromuscular diseases, characterized by progressive skeletal muscle weakness and wasting, and displaying high intra and interfamilial variability in age of onset, symptoms, presentation and progression (Deenen *et al*, 2014; Padberg, 2009; Sacconi *et al*, 2015; Wagner, 2019). Interestingly, while in early embryos, in testis and in thymus the DUX4 protein can be provided by either the 4q35 or 10q26 *D4Z4* regions, *DUX4*

expression in FSHD muscle cells is restricted to disease-permissive 4q35 alleles providing a polyadenylation signal, which stabilizes *DUX4* mRNA in FSHD myonuclei (F. *et al*, 2010; Lemmers *et al*, 2007). DUX4 cytotoxic effect in FSHD patients may be attributed to the cumulative effects of the numerous pathways it regulates. Indeed, when aberrantly re-expressed in terminally differentiated muscle cells, its transcriptional effects recapitulates its ability to activate cleavage stage-restricted genes (such as *ZSCAN4* family and endogenous retroviruses) and the aberrant activation of these transcriptional programs may be not tolerated by terminally differentiated muscle cells leading to cell death (Banerji & Zammit, 2021; Lim *et al*, 2020; Schätzl *et al*, 2021). Several pathways whose activation is associated with DUX4 expression are thought to contribute to the toxic effect of the transcription factor in FSHD. For example, DUX4 expression is associated with inflammation, hypoxia signaling and innate immune response pathways (Shadle *et al*, 2017; Lek *et al*, 2020). DUX4 has also been shown to affect myogenesis and sensitivity to oxidative stress at multiple levels (Knopp *et al*, 2016; Bosnakovski *et al*, 2008b; Dmitriev *et al*, 2016; Moyle *et al*, 2016; Bosnakovski *et al*, 2017; Geng *et al*, 2012). More recently, DUX4 roles in post-transcriptional regulation and protein homeostasis in FSHD have been reported (Feng *et al*, 2015; Homma *et al*, 2015; Jagannathan *et al*, 2019).

1.2.3.2 DUX4 in solid tumors

In recent years, *DUX4* aberrant re-expression has been reported in several forms of solid cancer (Preussner *et al*, 2018). As previously anticipated, developmental-specific genes such does activated by DUX4 are often re-expressed in various malignancies (CTAs) (Malaguti *et al*, 2019). Despite the immunogenicity of most DUX4 targets, activation of cleavage-stage transcriptional programs by DUX4 in cancer cells was shown to suppress of MHC class I-dependent antigen presentation, thereby promoting immune evasion (Chew *et al*, 2019). Moreover, such *DUX4*-positive cancers were shown to be characterized by a reduced immune infiltration, which is in contrast with the frequent presence of inflammation and lymphocytic infiltration observed in FSHD, in which the inflammatory response has been directly associated with DUX4 expression (Geng *et al*, 2012; Wang & Tawil, 2016; Arahata *et al*, 1995). On the other hand, DUX4 appears to act as tumor suppressor in colon cancer and synovial sarcoma by multiple mechanisms (Bury *et al*, 2019; DeSalvo *et al*, 2021). In synovial sarcoma, for example,

DUX4 expression is associated with upregulation of the growth control-genes *CDKN1A*, *EGR1* as well as stimulation and cell death, suggesting a role as a tumor suppressor inducing cell cycle arrest (DeSalvo *et al*, 2021). Furthermore, in colon cancer DUX4 has been recently suggested as a direct inhibitor of CDK1 activity, a critical cell cycle regulator, preventing its binding to its targets (Bury *et al*, 2019).

1.2.3.3 Herpesviridae Family Infection

Very recently, *DUX4* has been shown to be aberrantly re-expressed following infection by all viruses of the Herpesviridae family and to be a key transcriptional regulator in herpes virus infection. In particular, its ability to directly regulate the expression of *TRIM43*, a germline-restricted gene with an important role in viral genome replication, could potentially serve as biomarker of active herpes viral replication and pathogenesis (Full *et al*, 2019; C. *et al*, 2021).

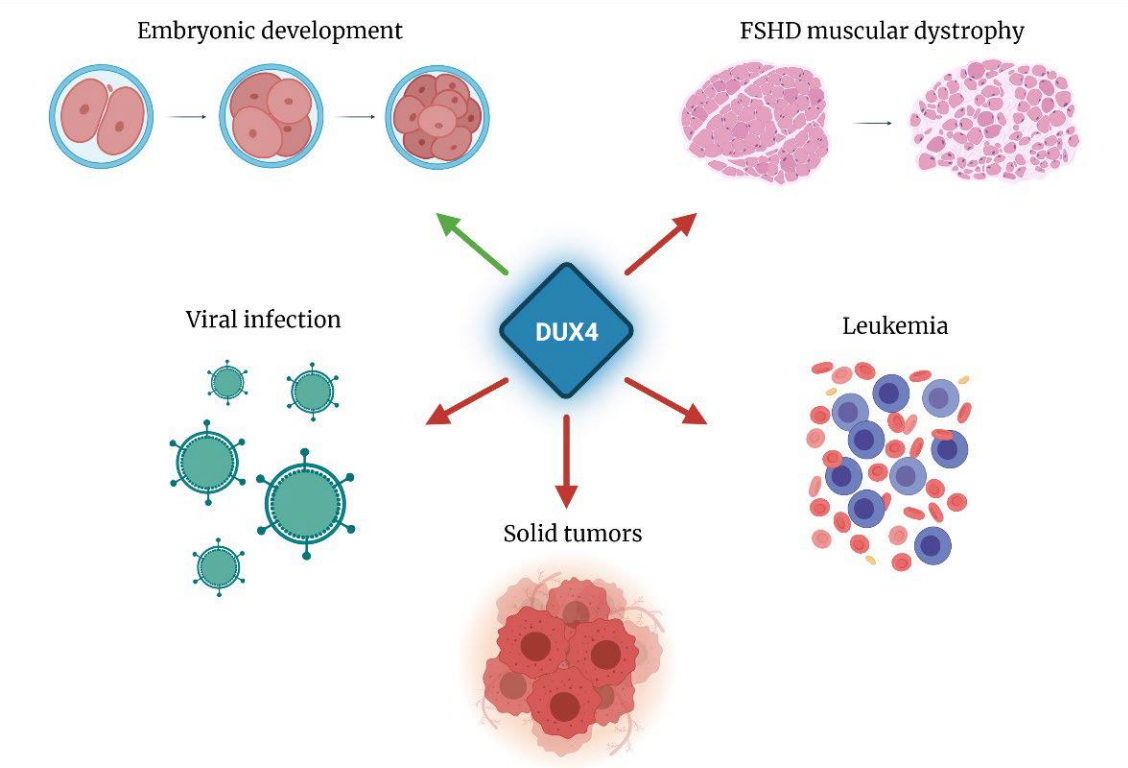


Figure 3. Roles of DUX4 in physiology and pathology. *DUX4 is physiologically expressed during early embryogenesis (green arrow) and subsequently silenced in most somatic tissues. DUX4 aberrant expression is associated with several pathological conditions (red arrows). DUX4 is re-expressed as direct consequence of Herpesviridae infection. DUX4 pathological gain of expression is associated with FSHD muscular dystrophy. Several forms of neoplasms display aberrant DUX4 expression or activity. Modified from Mocciaro and Runfola et al.2021.*

1.2.4 Transcriptional Activities of DUX4

Gene ontology analyses of the transcripts modulated by DUX4 in the different contexts described above highlight several processes including cell differentiation, proliferation, RNA transcription, RNA processing, cytoskeleton organization, immune response and viral response (Campbell *et al*, 2021; Jagannathan *et al*, 2016; Mocciaro *et al*, 2021)(Figure 3). Overall, the DUX4-associated transcriptional signature is in line with its role in establishing an early embryonic program through negative regulation of cell differentiation and positive regulation of cell proliferation. DUX4 direct transcriptional targets include several transcription factors and transcriptional regulators which are almost exclusively expressed at the cleavage-stage. Some of these targets have been shown to regulate each other, possibly amplifying the DUX4 transcriptional cascade. Among DUX4 targets are also the histone variants H3.X and H3.Y, which have been shown to be incorporated into the genomic regions of DUX4 target genes, facilitating the expression of DUX4 targets after a brief pulse of DUX4 expression (Resnick *et al*, 2019). Strikingly, functional studies based on DUX4 overexpression in different human cells highlighted its ability to activate target genes even if they are in a non-accessible chromatin environment. More in specific, genome-wide studies indicated that DUX4 binds to inaccessible chromatin in 60% of cases, while in the other 40% it binds to previously accessible genomic loci (Choi *et al*, 2016; Darko *et al*, 2021; Vuoristo *et al*, 2019; Mocciaro *et al*, 2021). Consistent with its ability to bind inaccessible genomic regions, DUX4 has been shown to activate the expression of several classes of repetitive elements including *Alu*, *LINE-1*, mammalian apparent LTR-retrotransposons (*MaLR*), endogenous retrovirus elements (*ERVs*) and pericentric human satellite II (*HSATII*) repeats (Hendrickson *et al*, 2017; Whiddon *et al*, 2017; Geng *et al*, 2012; Young *et al*, 2013; Dmitriev *et al*, 2016), elements mostly found in inaccessible chromatin regions. Some DUX4-activated *MaLR* and *ERV* elements give rise to retrotransposon transcripts, long non-coding RNAs and antisense transcripts. In other cases, binding of DUX4 to repeats drives transcription of neighboring protein-coding genes through novel promoters (Geng *et al*, 2012; Young *et al*, 2013). In line with this data, recent reports showed that DUX4 can recruit, via its C-terminal TAD, the transcriptional co-activators cAMP-response element binding protein (CBP) and E1A binding protein P300 (p300) (Choi *et al*, 2016). In line with their protein-lysine

acetyltransferases activity, binding of DUX4 to target gene region is associated with a significant increase in histone H3 lysine 27 acetylation (H3K27Ac), suggesting that DUX4 could associate to inaccessible chromatin regions and, by recruiting CBP/p300 factors, could promote local chromatin relaxation and gene activation (Vuoristo *et al*, 2019; Choi *et al*, 2016; Mocciaro *et al*, 2021). Consistently, similar results have been reported for mouse Dux (Hendrickson *et al*, 2017; Darko *et al*, 2021), suggesting that these activities might contribute to Dux/DUX4 physiological function in ZGA. Notably, the C-terminal domain of DUX4 plays a key role in the transcriptional activity of the pioneer factor. It was recently shown that the interaction of DUX4 C-terminus with transcriptional co-activators CBP and p300 is required for the activity of the

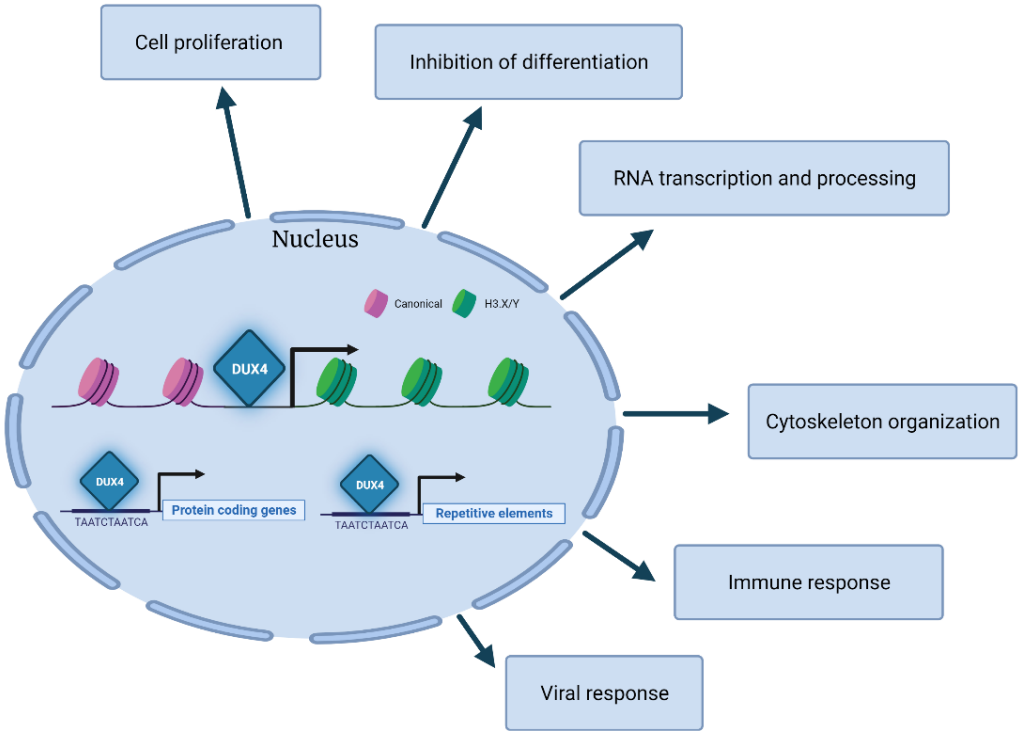


Figure 4 Transcriptional activities of DUX4. *DUX4 modulates the activity of several cellular pathways involved in cell proliferation, inhibition of cell differentiation, RNA transcription and processing, cytoskeleton organization, immune response and viral response. Modified from Mocciaro & Runfola et al. 2021.*

transcription factor, since a DUX4 version lacking the last 50 amino acids is completely unable to activate transcription of its target genes (Bosnakovski *et al*, 2008a; Choi *et al*, 2016). In conclusion, in recent years DUX4 has gone from being the "most wanted" gene in the context of FSHD to being a key factor in important physiological and pathological processes. DUX4 physiological role is confined to early embryonic development and its aberrant expression in various contexts recapitulates such activity

(Figure 4). For this reason, DUX4 expression in most somatic tissues is highly toxic, inducing cell apoptosis. Noteworthy, the DNA-binding domain of DUX4 alone is not sufficient for the TF to perform the above functions. Only full-length DUX4 containing an intact C-terminal TAD activates transcription of its targets leading to the above functional outcomes.

1.3 DUX4 FUSIONS IN CANCER

Chromosomal translocations are recurrent events observed among malignancies and many of these involve genes encoding transcriptional regulators (Rabbitts, 2001; Crans & Sakamoto, 2001). Through the fusion with their partners on different chromosomes, these genes often acquire novel functions (Look, 1997).

1.3.1 CIC-DUX4

In a highly aggressive subgroup of small round cell sarcoma, affecting predominantly children and young adults, the initiating and causative event is a fusion between the high mobility group (HMG) box containing protein Capicua (CIC) and DUX4. CIC is an evolutionarily conserved transcriptional repressor with key roles in several developmental and physiological processes(Lee, 2020). At the structural level, CIC is composed by an N-terminal HMG box and a C-terminal C1 domain, which cooperatively recognize specific DNA-sequences (Forés *et al*, 2017). Of importance, CIC is believed to act as a tumor suppressor in a number of malignancies, also through its ability to repress target gene transcription (Bunda *et al*, 2019; Lee *et al*, 2020; Okimoto *et al*, 2017; Chan *et al*, 2014; Simón-Carrasco *et al*, 2017). Interestingly, the chimeric protein resulting from the rearrangement involving DUX4 possesses intact HMG box and C1 domains from CIC, thus retaining its DNA-binding activity, but carries the C-terminal transactivation domain of DUX4. As a result, CIC is converted into a strong transcriptional activator (CIC-DUX4) which acts as a dominant oncogenic driver (Yoshimoto *et al*, 2017; Okimoto *et al*, 2017; Oyama *et al*, 2017; Nakai *et al*, 2019). The best known direct targets of mammalian CIC are cyclins and the oncogenic E-Twenty six Variant (ETV) transcription factors *ETV1/4/5*, which in CIC cancers are mediators of cell cycle, cell growth, proliferation, metastasis and treatment resistance (Wong & Yip, 2020). While transcriptionally silent due to wt CIC activity, the CIC-DUX4 fusion oncogene is believed to work mostly by aberrant activation of such target genes. In recent years, however, it has been reported that CIC-DUX4 expression directly regulates the transcription of *CCNE1* via distinct regulatory pathways, suggesting new roles for the oncogenic fusion which could also confer therapeutic opportunity. Noteworthy, the various studies on CIC-DUX4 transcriptional proficiency

failed to reveal significant upregulation of potential DUX4 target genes (Kao *et al*, 2017), consistent with the absence of the DUX4 DNA-binding domain in the fusion protein. As a matter of fact, this further confirms how DUX4 necessarily needs both the DNA-binding domain and the C-terminal TAD for the correct activation of its transcriptional signature.

1.3.2 DUX4-IGH

Historically, a novel B-ALL subtype was identified as characterized by a very distinctive microarray gene expression profile (GEP). However, since cytogenetic analyses failed to reveal any recurrent genomic alteration, this subtype was incorporated in the "other" B-ALL subgroup, remaining cytogenetically unclassified and thus lacking genetic information as support for treatment decisions (Yeoh *et al*, 2002). Follow-up studies involving copy number variation (CNV) analysis revealed that many of the patients within this group were characterized by deletions of the *ERG* gene, which encodes for the ETS-family transcription factor v-ets avian erythroblastosis virus E26 oncogene playing key roles in hematopoietic development (Loughran *et al*, 2008). Importantly, such genomic alteration was lacking in almost all other ALL subtypes. For this reason, *ERG* deletion has been initially proposed as the driving lesion in the novel subtype (Mullighan *et al*, 2007b; Harvey *et al*, 2010). On the other hand, further studies demonstrated that monoallelic deletion of *ERG* is observed in only a subset of patients of this B-ALL subgroup. Furthermore, *ERG* deletions were subclonal in several patients at diagnosis and either altered or absent at relapse (Potuckova *et al*, 2016; Zaliova *et al*, 2019; Zhang *et al*, 2016a), suggesting that ERG deregulation does not drive leukemogenesis by itself. More recently, thanks to the use of whole transcriptome sequencing (WTS), precise and sensitive detection of fusion genes enabled the discovery of *DUX4-IGH* as the cryptic genetic lesion driving this subtype of B-ALL (Marincevic-Zuniga *et al*, 2017; Qian *et al*, 2017; Yasuda *et al*, 2016; Liu *et al*, 2016; Zhang *et al*, 2016a; Lilljebjörn *et al*, 2016).

1.3.2.1 Clinical Aspects of DUX4-IGH ALL Subtype

DUX4-IGH driven leukemia is observed in up to 10% of B-ALL patients, with a slight peak in adolescents and young adults (15-39 y.o.). Nonetheless, such lesion is observed across all age groups. In pediatric patients, for example, the DUX4-IGH subtype accounts for 5-7% of diagnosed B-ALL and is characterized by an older age of onset, being more common in patients older than age 10. The *DUX4-IGH* rearrangement is associated exclusively with an early B-cell immunophenotype, even though few reports indicate that CD2, a marker restricted to the T-cell setting, is expressed in a proportion of DUX4-IGH B-ALL cases (Slamova *et al*, 2014). Moreover, recent reports describe surface expression of CD371 as a unique feature of this B-ALL subtype, which now represent a highly specific surrogate marker used to detect DUX4-IGH ALL cases, whose detection otherwise relies exclusively upon NGS (Schinnerl *et al*, 2019). Of importance, this subtype displays enriched lineage switching (swALL) with the respect to other B-ALL subtypes, which is characterized by co-expression of B lymphocyte and monocyte lineage markers at the same time (Novakova *et al*, 2018). More specifically, co-expression of CD19 and CD34 as well as CD33 and CD14 have been reported in swALL DUX4-IGH, which is associated with poorer treatment response and higher relapse (Novakova *et al*, 2018; Schroeder *et al*, 2019). From a clinical perspective, the DUX4-IGH subtype is classified as intermediate or high risk, based on clinical studies showing an higher MRD throughout induction therapy (Marketa Zaliova *et al*, 2019; Zaliova *et al*, 2014) and a slower treatment response (Clappier *et al*, 2014). Despite this, this subtype is associated with excellent prognosis, particularly in pediatric patients who received intensive chemotherapy for remission induction(Yeoh *et al*, 2002; Harvey *et al*, 2010). Importantly, recent reports highlight differences in prognosis between pediatric and adult patients(Liu *et al*, 2016), since DUX4-IGH B-ALL is associated with 93% event-free survival (EFS) and overall survival (OS) in pediatric patients (<18 years old) but 86% EFS and 84% OS in adults. Consistently, longer disease-free survival after complete remission (CR) was achieved in adolescent and young adult patients, though in the context of intensive chemotherapy given for high-risk stratification, highlighting the importance of proper risk stratification of different B-ALL subtype to guide therapeutic intervention.

1.3.2.2 Genetics of DUX4-IGH Rearrangement

At the structural level, rearrangements giving rise to the DUX4-IGH fusion proteins are usually not generated by balanced chromosome translocation or inversion, as most of the rearrangements found in ALL (Yasuda *et al*, 2016). Instead, most analyses reported insertions of small segments of the *D4Z4* array, either from 4q or the homologous region on 10q. These are mostly inserted into the *immunoglobulin heavy chain* (*IGH*) locus located in chromosome 14q, placing them in close proximity to the *IGH* enhancer (Yasuda *et al*, 2016; Marincevic-Zuniga *et al*, 2017; Lilljebjörn *et al*, 2016). Recently, published Hi-C data performed on the NALM6 cell line, in which DUX4-IGH represents the driver lesion(Yasuda *et al*, 2016; Zhang *et al*, 2016a), suggest that a reciprocal translocation could occur in which the telomeric ends of 4/10q are exchanged with 14q(Tian *et al*, 2019). Nonetheless, such structural event has been reported only in this case concerning the NALM6 cell line and future work could address whether such reciprocal chromosomal translocation is observed also in B-ALL patients. Despite the difficulty in defining breakpoint locations within the *DUX4* gene, given the repetitive nature of the locus, most of them have been shown to occur within the 5' region upstream of *DUX4* and within the 3' coding region in exon 1, which results in the translocation of either a partial copy of *DUX4* or one complete and one

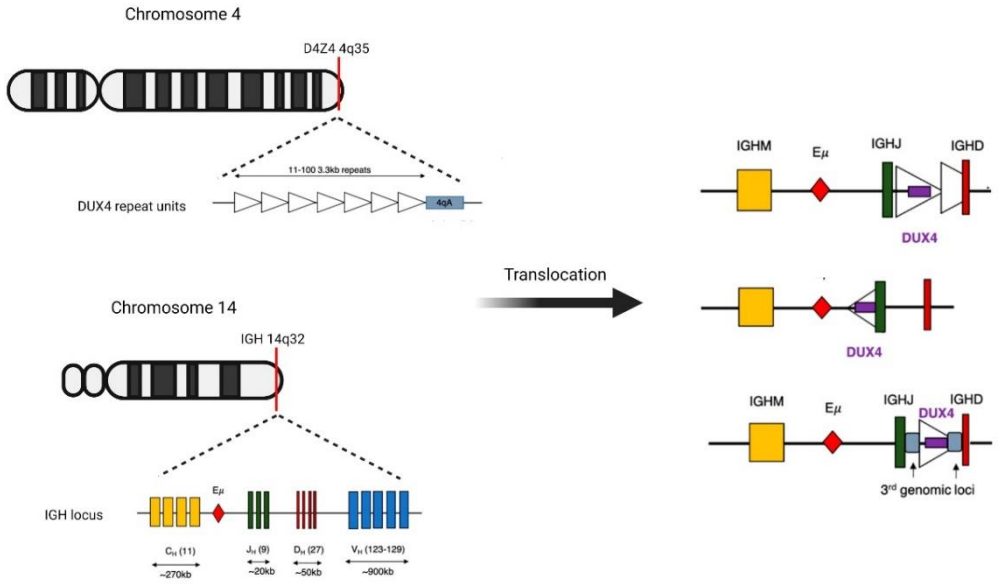


Figure 5. Schematic representation of the t4;14 DUX4-IGH rearrangement in B-ALL. *Upon recombination, at least 1 copy of DUX4 is translocated from the D4Z4 array into the IGH locus, mostly within the IGM constant allele (in yellow) and the IGH D-J junctions (red and green, respectively). In all cases, DUX4 falls under the control of the IGH enhancer (Eμ, in red), which drive its expression specifically in B-cell precursors. Image created with Biorender.*

partial *D4Z4* repeat. On the other hand, breakpoints in the *IGH* locus are enriched in the 3.5 kb region preceding the *IGHM* constant allele and overlapping the *IGH D-J* junctions but can occur throughout the locus(Yasuda *et al*, 2016; Zhang *et al*, 2016a; Tian *et al*, 2019) (Figure 5).

Of importance, all translocations maintain the region encoding for the DNA-binding domain of DUX4 but result in (apparently) random truncation of the transcription activation domain which, in most of cases, is substituted by (apparently) random amino acids encoded by the *IGH* locus (Yasuda *et al*, 2016; Zhang *et al*, 2016a; Lilljebjörn *et al*, 2016; Gu *et al*, 2019). The first important consequence of the translocations is represented by the fact that the *IGH* enhancer drives expression of the rearranged *DUX4* (*DUX4-r*) specifically in B-cell precursors (Yasuda *et al*, 2016; Zhang *et al*, 2016a; Lilljebjörn *et al*, 2016; Dyer *et al*, 2010), in which wt *DUX4* is never expressed (Figure 6). Given the high variability in the genomic breakpoints, as well as the randomness of

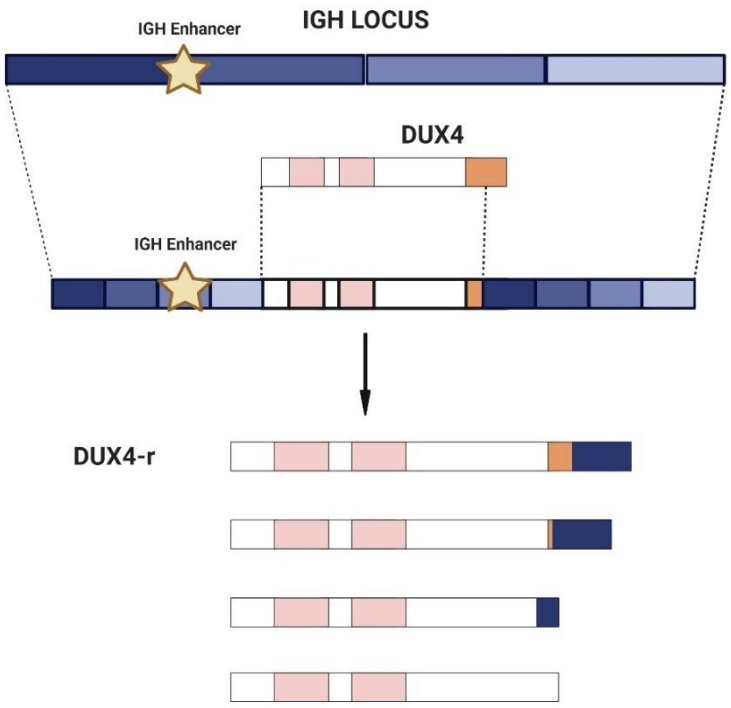


Figure 6. DUX4-IGH variants*. The rearrangement causes the insertion of DUX4 in the IGH locus, under the control of the IGH enhancer (yellow star). Due to this event, the C-terminal TAD of DUX4 is always truncated and, in most of the cases, random amino acids of the locus are appended. Given the randomness of the event, DUX4-IGH variants display highly variable C-termini, from longer ones to complete truncation of the C-terminal domain. Instead, the DNA-binding domain of DUX4 (pink boxes) is always maintained in all rearranged DUX4 in ALL. Image created with Biorender.*

the read-through appendage, the sequence and length of the chimeric DUX4-IGH C-terminal region varies considerably between patients (Figure 6). Nevertheless, since DUX4-r consistently retains an intact DNA binding homeodomains, it could potentially bind the same genomic targets of DUX4 (Figure 4) (Yasuda *et al*, 2016; Zhang *et al*, 2016a; Tian *et al*, 2019).

For the remaining of my thesis, I will refer to wt DUX4 as DUX4, to the chimeric proteins generated by substitution of the DUX4 TAD with IGH-encoded amino acids as DUX4-IGH, and to all the rearranged DUX4 versions in ALL (whether just deleted of the TAD or with the TAD substituted by IGH-encoded amino acids) as DUX4-r.

1.3.2.3 DUX4 and ERG

Before the discovery of the *DUX4* rearrangements, *ERG* deletions and its following transcriptional deregulation were proposed to be the key leukemogenic event driving this B-ALL subtype. ERG plays a key role in hematopoietic differentiation, megakaryopoiesis, and megakaryoblastic leukemia, through the direct activation of several genes (Tsuzuki *et al*, 2011; Taoudi *et al*, 2011; Rainis *et al*, 2005). Moreover, *ERG* is frequently rearranged in carcinoma of the prostate and rarely in acute leukemia, and ERG overexpression is associated with poor outcome in acute myeloid leukemia, consistent with a possible role in leukemogenesis (A. *et al*, 2005; Prasad *et al*, 1994; Marcucci *et al*, 2005). In the case of B-ALL, monoallelic intragenic deletions of *ERG* commonly involve deletion of several exons of the gene, and are observed in 3–5% of B-ALL cases, almost exclusively in the DUX4-r subtype (Mullighan *et al*, 2007b; Clappier *et al*, 2014; Zaliova *et al*, 2014). On the other hand, these alterations are absent in up to 40% of DUX4-r patients, are often subclonal and inconsistent between diagnosis and relapse, thus indicating that *ERG* deletions represent a secondary event in disease progression (Clappier *et al*, 2014; Marketa Zaliova *et al*, 2019). With the discovery of *DUX4* genetic alterations in ALL, efforts in deciphering the activity of DUX4-r highlighted its ability to initiate transcription from a non-canonical exon (exon 6alt) present in *ERG* intron 6, generating an alternative transcript, named *ERGalt*, which lacks the N-terminal and central regulatory domains but retains the DNA-binding ETS and transactivation domains of the wt ERG protein (Zhang *et al*, 2016a; Dong *et al*, 2018). Of importance, ERGalt displayed a lower transactivating activity with respect to

ERG thus acting as a dominant negative inhibitor of the wild type transcription factor (Zhang *et al*, 2016a). Notably, introduction of ERGalt in lineage-negative (lin-) bone marrow cells derived from tumor prone Arf-null mice, following their transplantation in lethally irradiated normal mice sustained the development of lymphoid precursor leukemia (Zhang *et al*, 2016a), consistent with a possible role of ERGalt in leukemia development. On the other hand, the fact that *ERG* exon 6alt is not conserved in the mouse genome (Zhang *et al*, 2016a), and DUX4-IGH expression alone in wt murine pro-B cells is sufficient to give rise to mouse leukemia (Yasuda *et al*, 2016) indicate that the expression of DUX4-r represents an early, initiating event in leukemogenesis, and that the activity of DUX4-r in driving leukemia goes beyond the deregulation of *ERG*. While the exact contribution of *ERG* transcriptional deregulation in the progression of DUX4-r ALL disease is not fully understood, in AML and T-ALL, high levels of wt ERG are associated with a poorer prognosis (Nibourel *et al*, 2017; Marcucci *et al*, 2005; Thoms *et al*, 2011) and shRNA-mediated *ERG* knockdown in the latter system is associated with inhibition of cell growth (Thoms *et al*, 2011). Since *ERG* deletions in B-ALL patients is associated with a favorable prognosis, it is tempting to speculate that transcriptional deregulation of *ERG* may offer a protective effect which results in improved DUX4-r patient outcome.

1.3.2.4 Transcriptional Activities of DUX4-IGH

B-ALL patients associated to the expression of DUX4-r display widespread hypomethylation and a distinct gene expression profile compared with normal B cells and other ALL subtype (Marincevic-Zuniga *et al*, 2017; Yasuda *et al*, 2016; Lilljebjörn *et al*, 2016; Liu *et al*, 2016). Several genes were found to be consistently and selectively upregulated in DUX4-r patients in different datasets, including: *CD34*, *ITGA6*, *CCNJ*, *CLEC12A* and *PTPRM* (Lilljebjörn *et al*, 2016; Marincevic-Zuniga *et al*, 2017). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) analyses supported direct regulation of some of these genes by of DUX4-IGH (Yosuke Tanaka *et al*, 2018). Direct DUX4-IGH targets include *STAP1*, *CCNJ*, and *ITGA6*, whereas indirect targets include *CLEC12A*, *PTPRM*, and *DDIT4L* (Yosuke Tanaka *et al*, 2018). Intriguingly, *CLEC12A* encodes for the surface marker CD371, which has been regarded to as a unique feature of DUX4-r ALL (Schinnerl *et al*, 2019), suggesting a possible activity downstream of DUX4-r which converges on the surface expression of this marker. In

addition, while gene-set enrichment analysis performed on DUX4-r ALL patient-derived datasets failed to uncover up and/or downregulation of specific pathways (Hystad *et al*, 2007; Konermann *et al*, 2015; Yosuke Tanaka *et al*, 2018), possibly due to the heterogeneity of their genetic background, similar studies performed in NALM6 cells show that DUX4-IGH knockdown pushes cells toward B-cell differentiation, as indicated by enrichment in a pro/pre-B cell transcriptional signature (Yosuke Tanaka *et al*, 2018). These results are consistent with the fact that DUX4-IGH expression abrogates murine B-cell differentiation, which imply an ability of DUX4-IGH to retain progenitor cells into an immature status (Yosuke Tanaka *et al*, 2018; Zhang *et al*, 2016a). Nonetheless, as will be discussed later, the above transcriptomic and genomic studies have several caveats and a clear idea regarding the pathways primarily regulated by DUX4-r is currently lacking, consequently hampering the possibility of targeted therapeutic interventions.

1.3.2.5 Currently Proposed Disease Model

Being discovered only very recently, DUX4-r driven B-ALL still lacks a unifying mechanism of disease initiation and progression. Very recently, efforts have been made with the aim of deciphering the activity of the oncogenic transcription factor, particularly using the NALM6 cell line as a model. By using various genome-wide approaches including long-read RNA-seq, ChIP-seq and ATAC-seq, it was found that the rearrangement of *DUX4* has happened in the silenced *IGH* allele (Tian *et al*, 2019). Consequently, DUX4 fusions are expressed at much lower levels than the immunoglobulin heavy chain contained in the other *IGHM* allele (Igμ) (Tian *et al*, 2019). Furthermore, following DUX4 overexpression into NALM6 cells, an increased apoptosis was observed. Based on the above results, it was hypothesized that *DUX4* translocation into the silenced *IGH* allele would avoid high-levels of DUX4-r, which would not be tolerated leading to cell death. Based on this, it has been proposed that DUX4 and DUX4-r function interchangeably and that leukemia development requires "just right" DUX4 or DUX4-r levels, irrespectively of alterations of DUX4 TAD. Specifically, while too much DUX4/DUX4-r would result in cell toxicity and apoptosis, as in the case of DUX4 overexpression in NALM6 cells, too little DUX4/DUX4-r would just block cellular proliferation, as suggested by the reduced proliferation

observed upon shRNA-mediated DUX4-IGH KD in NALM6 cells (Yasuda *et al*, 2016; Liu *et al*, 2019; Seong *et al*, 2021; Yosuke Tanaka *et al*, 2018) (Figure 7).

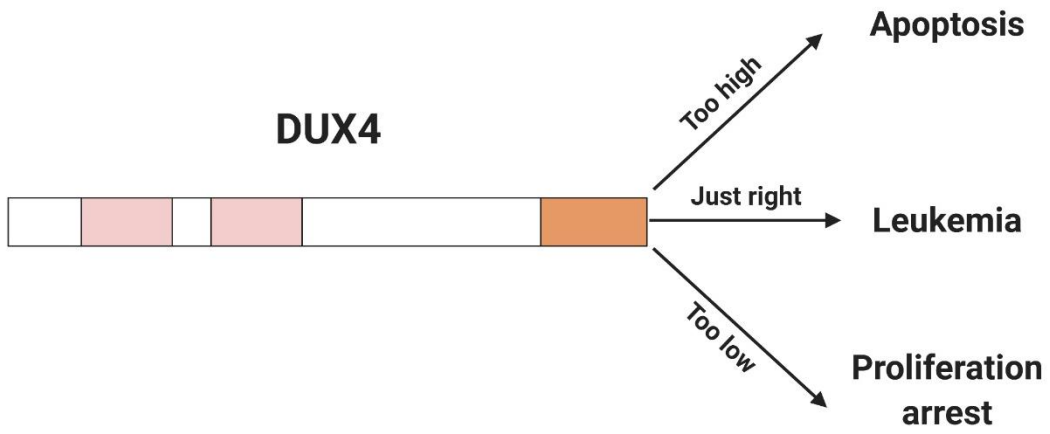


Figure 7. ***Currently proposed leukemic model of DUX4/DUX4-IGH.*** *While high DUX4 expression is toxic, causing cell death, low DUX4/DUX4-IGH expression levels is associated with proliferation arrest. Only "just right" DUX4/DUX4-IGH levels would promote the leukemic transformation of cells. Image created with Biorender.*

This model for the leukemia caused by DUX4 rearrangements, however, does not fit with several observations reported by other authors. In B-ALL, DUX4-r is expressed at much higher levels compared with those of DUX4 in human cleavage stage embryos: 145,4 vs 6,65 FPKM [fragments per kilobase of transcript, per million mapped reads] (Tian *et al*, 2019; Hendrickson *et al*, 2017; Zhang *et al*, 2016a). Despite that, while ectopic DUX4 expression is toxic to somatic cells, DUX4-IGH overexpression transform normal murine pro-B cells, which ultimately generates B-cell leukemia in transplanted NSG mice (Yasuda *et al*, 2016). Consistently, DUX4-IGH fusions but not DUX4 induce cellular transformation of NIH3T3 fibroblasts, demonstrating the oncogenic activity of the fusion proteins (Yasuda *et al*, 2016). Finally, previous reports showed that NALM6 cells ectopically overexpressing DUX4 or DUX4-IGH are characterized by distinct transcriptional signatures (Yosuke Tanaka *et al*, 2018).

2. AIM OF THE WORK

The proposed models for DUX4-r caused ALL, ERGalt activation or just right DUX4 levels, are not fully supported by the evidence in the literature, which reflects the shortage of knowledge concerning the cellular effects of DUX4-r expression. More specifically, while transcriptomic analyses performed in patients-derived samples support a novel activity of the DUX4-IGH oncogene (Liu *et al*, 2016; Marincevic-Zuniga *et al*, 2017; Lilljebjörn *et al*, 2016), these data are biased by the very different genetic background characterizing different B-ALL patients or cellular systems used. This is also reflected from functional enrichment analyses performed on such transcriptomic data, which failed to provide insights on specifically dysregulated pathways in the DUX4-r B-ALL subgroup (Hystad *et al*, 2007; Laurenti *et al*, 2013; Yosuke Tanaka *et al*, 2018). While *in vitro* work performed on B-ALL cell lines provided initial insights on the transcriptional activity of DUX4-IGH, the heterogeneity of the contexts in which these experiments have been performed complicates the dissection of the primary effect of DUX4-r expression. For instance, RNA-sequencing following DUX4-IGH knockdown in NALM6 cells has been performed at late time points (72h) (Yosuke Tanaka *et al*, 2018), thus making it difficult to discern primary DUX4-IGH transcriptional targets from secondary events. Likewise, overexpression of the fusion gene in the same cellular context cannot reflect the first transcriptional changes induced by DUX4-IGH, since an endogenous DUX4-IGH fusion oncogene is already present and active in NALM6 cells. Finally, while ChIP-seq data for DUX4-IGH have been obtained in relevant cell lines (NALM6 and REH) (Yosuke Tanaka *et al*, 2018; Zhang *et al*, 2016a; Dong *et al*, 2018), DUX4-related data have been generated in different contexts, such as induced pluripotent stem cells (iPS) (Hendrickson *et al*, 2017) or human myoblasts (Young *et al*, 2013).

All in all, there is an urgent need to compare the activity of DUX4 and representative DUX4-r variants in a controlled, relevant system in order to define the molecular mechanism through which DUX4-r promotes and maintains ALL.

3. RESULTS

The lack of knowledge concerning the molecular consequences of DUX4-r expression in B-ALL strongly impairs the possibility to develop therapeutic interventions tailored to block the primary activity of the oncogenic protein. To address this issue, there is a strong need to compare the activity of the DUX4 and its variants rearranged in ALL in a controlled, disease-relevant setting.

3.1 IDENTIFICATION OF DUX4-r SPECIFIC TARGET GENES

3.1.1 DUX4 and DUX4-r activity in B-ALL cells

To generate a cellular model for the study of DUX4-r, I decided to use REH cells. REH is a near-diploid B-cell precursor ALL cell line which do not express either DUX4 or DUX4-r. The cell line was established from an adolescent at initial diagnosis, with cells showing exclusively a B-cell immunophenotype and thus representing a good system to test the transcriptional activity of DUX4 and DUX4-r in a disease-relevant setting. Regarding DUX4-r, I decided to focus on two variants representatives of what has been reported in ALL: one variant representative of the DUX4-IGH chimeras and one variant representative of the rare ALL cases displaying DUX4 CTD deletion without addition to extra amino acids. To do this, I generated stable cell lines by introducing in REH cells lentiviral constructs encoding for: I. an Empty Vector (EV) negative control; II. DUX4; III. a DUX4-IGH lacking most of DUX4 CTD and carrying 32 amino acids encoded from the IGH locus (DUX4-IGH); IV. a DUX4-r version lacking the last 50 amino acids and thus the whole DUX4 CTD (DUX4-del50). The DUX4-IGH construct is derived from an actual ALL patient and is the construct mostly used in the literature (Dong *et al*, 2018; Yasuda *et al*, 2016; Zhang *et al*, 2016a). The DUX4-del50 is nearly identical to the SJBALL021405_D1 rearrangement described in Zhang *et al*, 2016a. Importantly, I decided to use doxycycline (dox)-inducible constructs to: I. avoid the toxic effect of prolonged DUX4 expression in somatic cells; II. enrich for primary targets of the transcription factors while avoiding the secondary effects of their long-term overexpression. In order to determine the best experimental conditions to test the transcriptional activities of DUX4, DUX4-IGH and DUX4-del50 in REH cells, I performed several transgene induction trials using different doxycycline concentrations and performed gene expression analyses at different time points.

Through this approach, I established that 12h of induction with 1ug/mL dox was the best condition.

While DUX4-IGH and DUX4-del50 were expressed at comparable levels upon dox induction, DUX4 was expressed at much lower RNA and protein levels (Figure 8A-B), possibly due to transgene leakiness and counter-selection. Nonetheless, despite its lower expression, only DUX4, and not DUX4-IGH or DUX4-del50, was able to induce apoptosis of REH cells (Figure 8C).

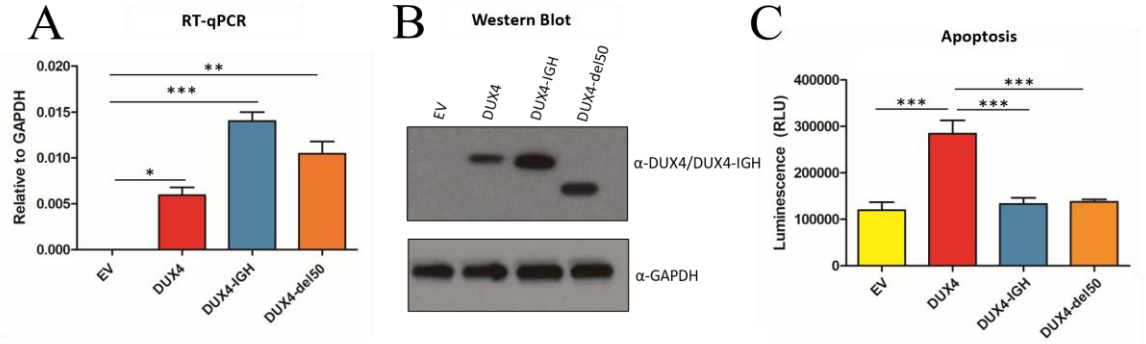


Figure 8. Comparison of DUX4 and DUX4-r activity in REH cells. *A. RT-qPCR analysis of DUX4, DUX4-IGH and DUX4-del50 mRNA levels in REH cells upon 12h dox induction (One-Way ANOVA Analysis of Variance with Tukey's correction. * p≤0,05, ** p≤0,01, *** p≤0,001). B. Western Blot analysis of DUX4, DUX4-IGH and DUX4-del50 protein levels upon 12h dox induction. C. Apoptosis Assay monitoring the activity of caspase performed upon 18h of transgene induction in REH cells (*One-Way ANOVA Analysis of Variance with Tukey's correction. *** *p≤0,001).*

While it is consistent with the fact that DUX4 ectopic expression is toxic to somatic cells, this result is in contrast with the proposed model (Tian *et al*, 2019) that only "just right" DUX4/DUX4-r levels would promote leukemia development while too high DUX4/DUX4-r levels would induce apoptosis. On the contrary, my results indicate that the activity of the rearranged DUX4 versions is different from the one of wt DUX4. Hence, to explore these differences at a genome-wide scale, I performed RNA-sequencing (RNA-seq).

3.1.2 DUX4 and DUX4-r variants activate different gene sets

Through our transcriptomics approach we aimed at dissecting the differences in the activity between DUX4 and its leukemic variants, while focusing on the primary effect of DUX4/DUX4-r expression. To do so, I isolated RNA from four replicates of EV, DUX4, DUX4-IGH and DUX4-del50 REH cells after 12h of induction. The principal

component analysis (PCA) of the RNA-seq results shows very tight clustering of group replicates indicating minimal inter-replicate variance (Figure 9A). In the PCA, the gene expression profiles of cells expressing DUX4-IGH and DUX4-del50 cluster nearby and distantly from those of EV or DUX4 samples (Figure 9A). Analysis of the most deregulated genes in my datasets show that induction of DUX4 or DUX4-r variants is mainly associated to activation of gene expression with respect to EV control. Nevertheless, DUX4 activates very different set of genes compared to DUX4-r variants. Instead, DUX4-IGH and DUX4-del50 activated genes are qualitatively similar, even if DUX4-IGH appears to have a stronger transactivation ability with respect to and DUX4-del50 (Figure 9B).

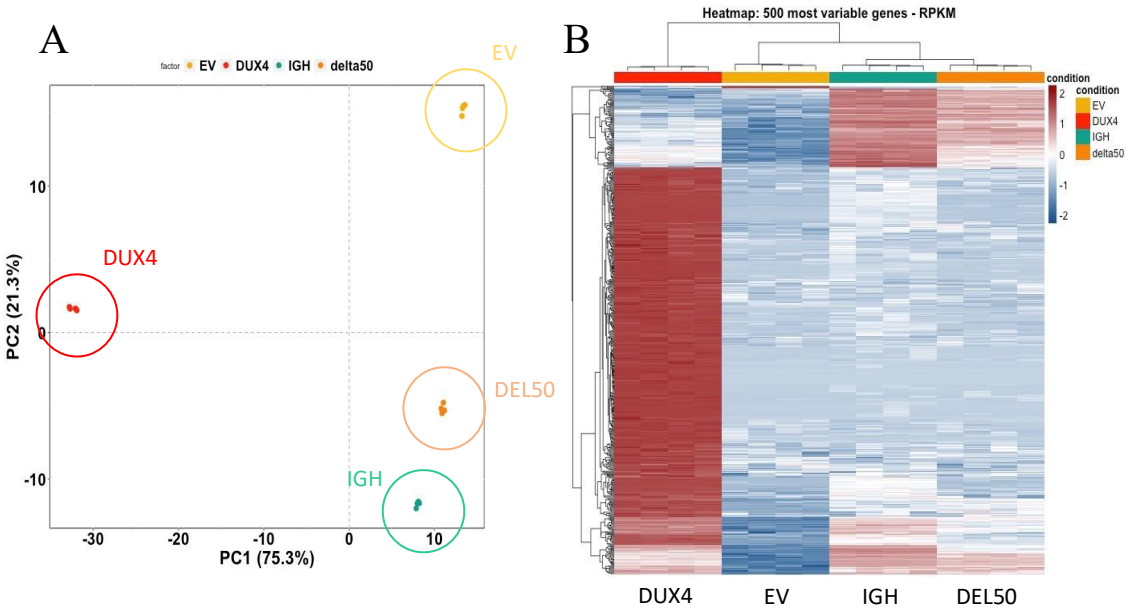


Figure 9. DUX4 and DUX4-r activate different gene sets in B-ALL cells. *A. Principal component analysis of the transcriptome of REH cells after 12h of induction of EV, DUX4, DUX4-IGH and DUX4-DEL50. B. Heatmap of the 500 hundred most variable genes identified in our RNA-seq experiment.*

DUX4 is known to activate genes whose expression is restricted to the cleavage stage embryo (Hendrickson *et al*, 2017; De Iaco *et al*, 2017). Many of these genes are silent or expressed at very low levels (0-5 counts per million, CPM) in EV control REH cells, reaching high levels (more than 10000 CPM) in DUX4 REH cells. For these genes, a very small increase of expression will score as significant despite in absolute level is likely of no biological consequence. For example, the DUX4 target ZSCAN4 displays an average of 3 CPM in EV cells and 28000 CPM in DUX4 cells. Given that DUX4-

IGH and DUX4-del50 cells display 30 and 20 CPM respectively, at first approximation ZSCAN4 would score as a DUX4/DUX4-r common target. Nevertheless, the biological effects caused by ZSCAN4 expression at high levels in DUX4 cells is likely very different from that caused by the minimal expression it reaches in DUX4-r cells. Therefore, to highlight those genes which are more likely to contribute to the biology of DUX4 or DUX4-r, I performed differential gene expression analysis (DGE) applying stringent filters on both statistical significance and magnitude of gene activation in the different samples, as described more in detail in the "Materials and Methods" section.

By applying the above filters, I found 1246, 672 and 499 upregulated and 207, 203 and 143 downregulated genes in the DUX4, DUX4-IGH and DUX4-del50 groups, respectively. In line with the results of Figure 9, my analysis shows a nearly complete overlap for both up- and downregulated genes in DUX4-IGH or DUX4-del50 groups. Instead, I found a minimal (less than 10%) overlap between genes deregulated following DUX4 or DUX4-r expression (Figure 10).

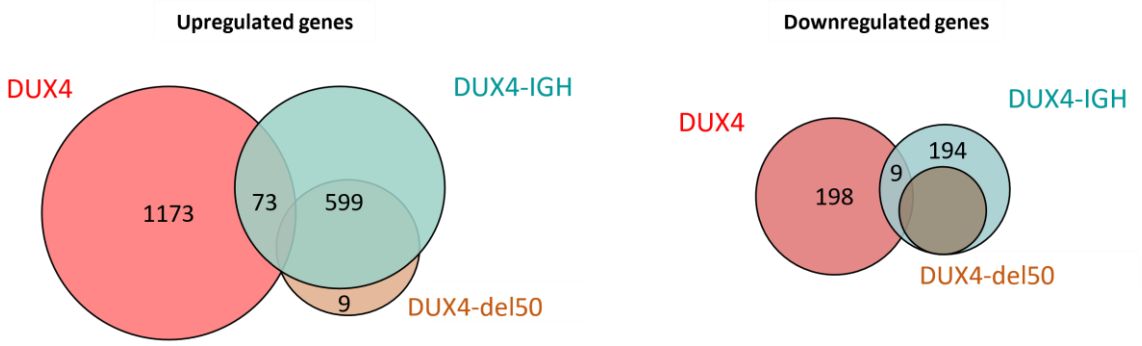


Figure 10. Overlap between DUX, DUX4-IGH and DUX4-del50 datasets. *Venn diagram representation of genes upregulated (left) and downregulated (right) upon DUX4 (red) DUX4-IGH (green) and DUX4-del50 (orange) overexpression in REH cells.*

I then addressed the similarity between the transcriptome of DUX4 or DUX4-r cells with that of DUX4-r ALL patients. By computing the overlap between the lists of genes selectively upregulated and downregulated by DUX4 in our RNA-seq and publicly available lists of genes up- and down-regulated in DUX4-r patients, I found minimal (less than 5%) commonly deregulated genes (Figure 11A). On the other hand, when performing the same comparison using genes differentially expressed in DUX4-IGH or DUX4-del50, I found highly significant overlaps for up to 27% of the genes (Figure 11B). While DUX4-IGH and DUX4-del50 activate the same genes, DUX4-del50

displays a milder effect compared to DUX4-IGH (Figure 9B). I thus hypothesized that the transactivation activity of DUX4-r could correlate with the presence/absence of a chimeric portion, the size of the chimeric portion and/or the size of the final DUX4-r protein.

To test this hypothesis, I re-analyzed publicly available RNA-seq datasets of DUX4-r B-ALL patients to determine the specific DUX4-r variant and the level of activation of DUX4-r targets present in each patient. Strikingly, I found a significant positive correlation between the size of DUX4-r and the number and the level of activation of those genes overlapping between my DUX4-r datasets and DUX4-r B-ALL datasets (Figure 11C-D).

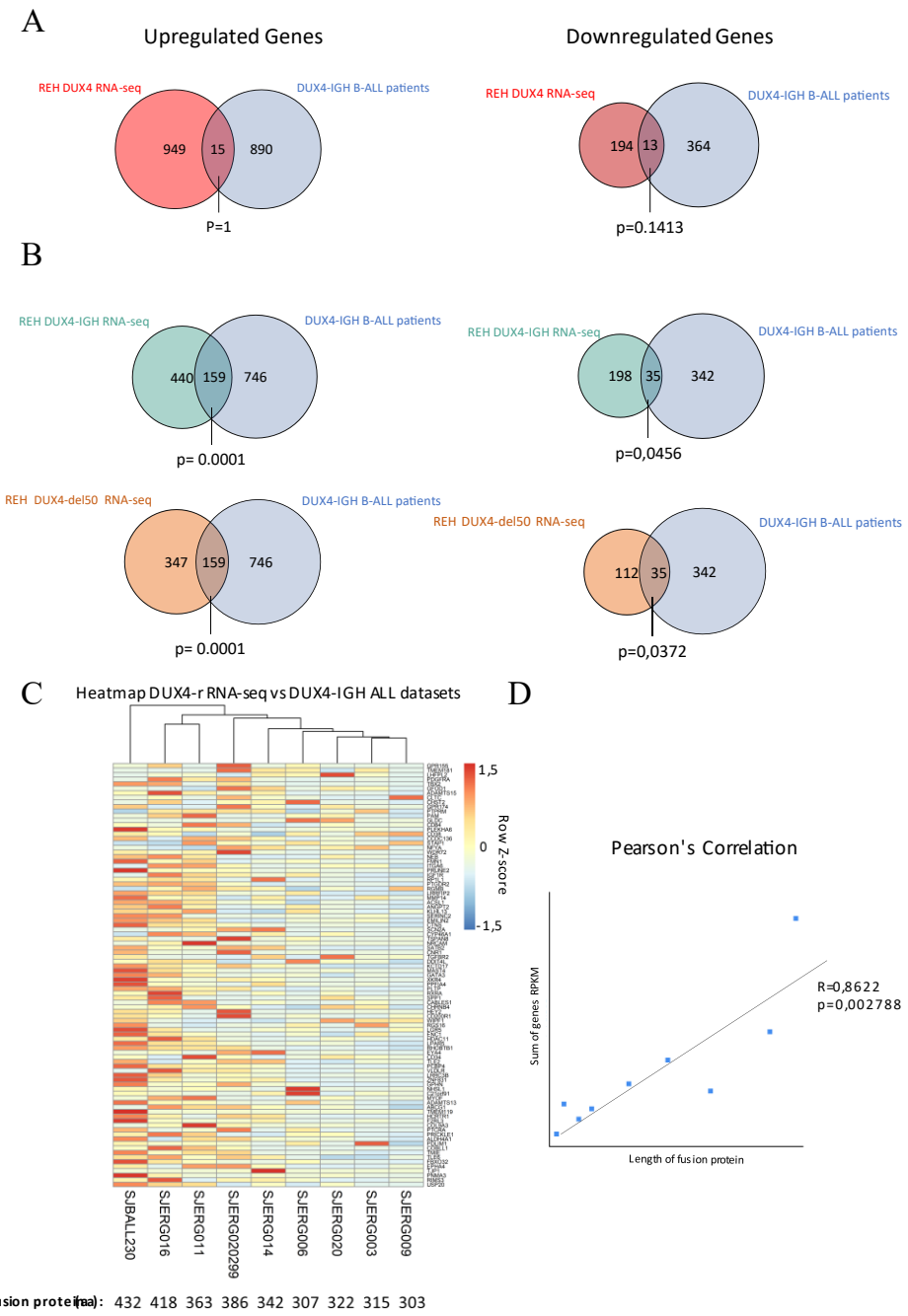


Figure 11. Overlap between DUX4 and DUX4-r datasets and the DUX4-IGH B-ALL signature. *A. Venn diagrams representation of the overlap between genes upregulated (left) or downregulated (right) by DUX4 in REH cells (red) and in the DUX4-IGH B-ALL signature (light blue) (Two-tailed Fisher's exact test). B. Venn diagrams representing the overlap between genes upregulated (left) or downregulated (right) by DUX4-IGH (green) or DUX4-del50 (orange) in REH cells and in the DUX4-IGH B-ALL signature (light blue) (Two-tailed Fisher's exact test). C. Heatmap showing the relative expression levels of the genes overlapping between DUX4-r REH RNA-seq and individual DUX4-r ALL patients. The patient code and the DUX4-r size in each patient are indicated.D. Pearson correlation plot between DUX4-r size and total log2fold gene activation in each patient. The p-value has been calculated for the R coefficient using the t-distribution and assigning a significance level of 0.05*

Altogether, my initial comparison did not support an interchangeable activity between wt DUX4 and DUX4 variants rearranged in ALL. Two extreme versions of DUX4-r variants do not induce apoptosis in REH cells despite being expressed at higher levels than DUX4. These variants display qualitatively nearly identical transcriptional activity, regulating a set of genes significantly overlapping the gene signature of DUX4-r ALL patients. In individual DUX4-r ALL patients, these DUX4-r specific targets display a level of expression positively correlated to DUX4-r size and may include key genes required for leukemia development.

3.1.3 Functional enrichment analyses

Previous works reported that functional enrichment analyses on the gene signature of DUX4-r ALL patients failed to reveal significantly deregulated pathways (Hystad *et al*, 2007; Laurenti *et al*, 2013). I performed gene-set enrichment analysis (GSEA) on the genes commonly deregulated in cells expressing DUX4-IGH or DUX4-del50 and DUX4-r ALL patients (Figures 10 and 11). This analysis revealed that upregulated DUX4-r specific targets are significantly enriched in integrins and cell-adhesion molecules involved to cell-cell/cell-matrix interactions, cell migration as well as a number of signaling pathways, including the RAS and PI3K-Akt pathways (Figure 12A). Downregulated DUX4-r specific targets are almost uniquely involved in functional categories related to the B-cell identity, particularly concerning the activity of the B-cell receptor (Figure 12B). Hence, focusing specifically on DUX4-r specific targets highlight pathways with a clear relevance in B-ALL.).

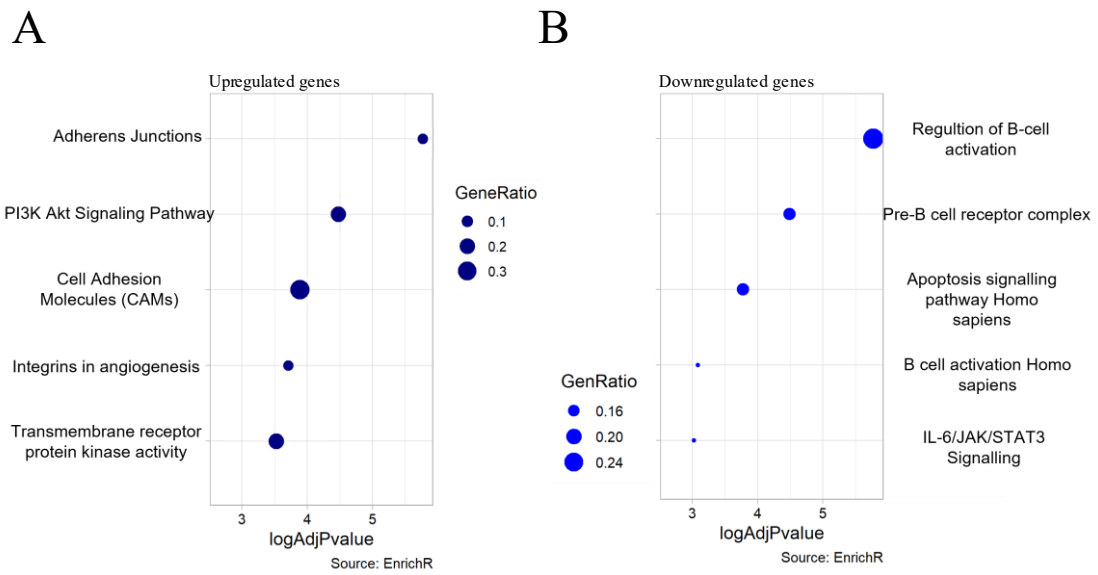


Figure 12. Functional Enrichment analyses of the overlap between DUX4-IGH signature and DUX4-IGH B-ALL patients. *A. Analysis of the functional families enriched in DUX4-IGH positively regulated target genes. B. Analysis of the functional families enriched in DUX4-IGH negatively regulated target genes*

To validate my results, I selected genes which: I. show highest magnitude of deregulation; are consistently regulated in DUX4-r patients; II. are regulated in the opposite direction in publicly available datasets of NALM6 knockdown for DUX4-r (Yosuke Tanaka *et al*, 2018; Yasuda *et al*, 2016); and III. are enriched within the top deregulated pathways in the functional enrichment analyses. Using retro-transcription followed by quantitative PCR (RT-qPCR), in cells expressing DUX4-IGH or DUX4-del50 I confirmed a selective upregulation of *ITGA6*, *CD34*, *PTPRM*, *NR3C2*, *ABCG1* and *DDIT4L* (Figure 13A). *ITGA6*, *CD34* and *PTPRM* encode for cell-surface molecules involved in cell-cell/cell-matrix adhesion as well as transduction of extracellular signals. *ABCG1*, *NR3C2* and *DDIT4L* play roles in signaling downstream of the AKT1/2, congruent with the observed deregulation of this pathway in DUX4-r transcriptomes. Concerning genes negatively regulated upon DUX4-r expression, I validated downregulation of several genes encoding for essential structural components of the pre-BCR, such as *VPREB3* and *CD79b*, as well as the transcriptional repressor *BCL6*, which conveys into the cell nucleus pathways downstream the pre-BCR and BCR signaling, among others (Figure 13B). Core DUX4 targets encoding for genes restricted to pre-implantation embryo development were selectively deregulated in response to DUX4 expression (Figure 13C).

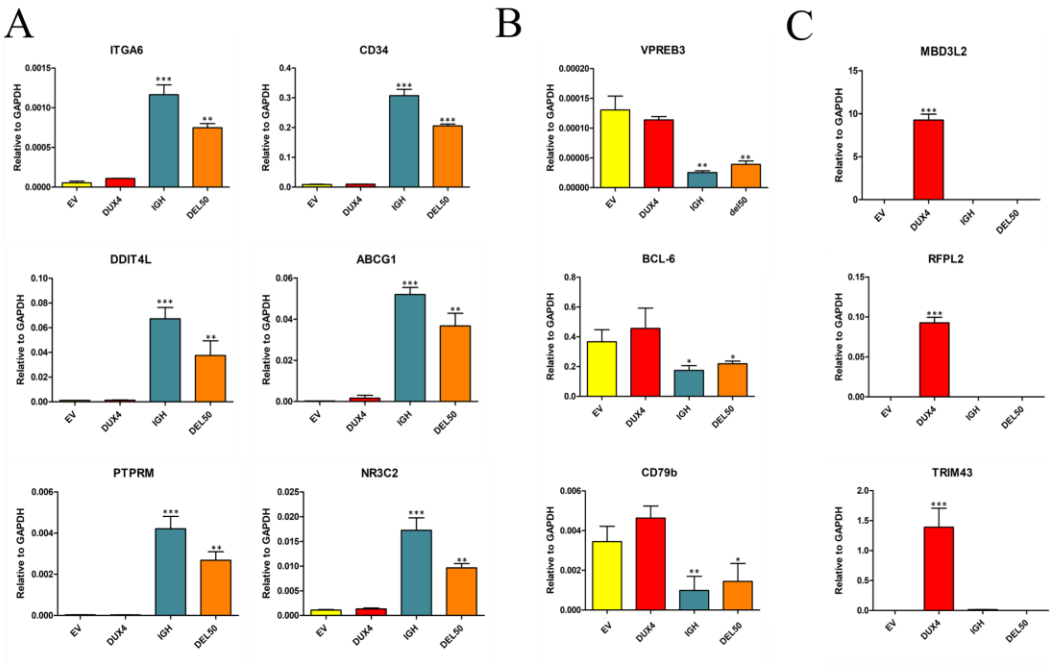


Figure 13. Validation of RNA-seq results. *A. RT-qPCR analysis of genes consistently upregulated in DUX4-r expressing REH cells and regulated in opposite direction in publicly available datasets of NALM6 knockdown for DUX4-r (One-Way ANOVA Analysis of Variance with Tukey's correction. ** p≤0,01 *** p≤0,001). B. RT-qPCR analysis of genes downregulated in DUX4-r expressing REH cells and regulated in opposite direction in publicly available datasets of NALM6 knockdown for DUX4-r (One-Way ANOVA Analysis of Variance with Tukey's correction. * p≤0,05 ** p≤0,01). C. RT-qPCR analysis of core DUX4 target genes (One-Way ANOVA Analysis of Variance with Tukey's correction. *** p≤0,001).*

All in all, DUX4-r variants induce a unique gene signature, very different from that of DUX4 but significantly overlapping the one of DUX4-r ALL patients and enriched for protein-coding genes encoding for factors involved in cellular pathways which are possibly relevant for leukemia development.

3.1.4 Repetitive DNA elements

As previously discussed, many DUX4 transcriptional targets are repetitive elements (REs). The ability of DUX4-r to activate RE expression has never been addressed. By analyzing my RNA-seq results focusing on human REs I found that, while DUX4 induction is associated with the upregulation of hundreds of REs, DUX4-r variants have a minimal impact on the RE expression (Figure 14A). DGE analysis on the repetitive transcriptome shows that DUX4 activates 356 REs and downregulates 73 REs. In contrast, DUX4-IGH and DUX4-del50 activate 29 and 26 REs and downregulate 14 and 5 REs, respectively, all of which are regulated also by DUX4. Using RT-qPCR, I confirmed the exclusive ability of DUX4 to activate a number of RE targets (Figure 14B).

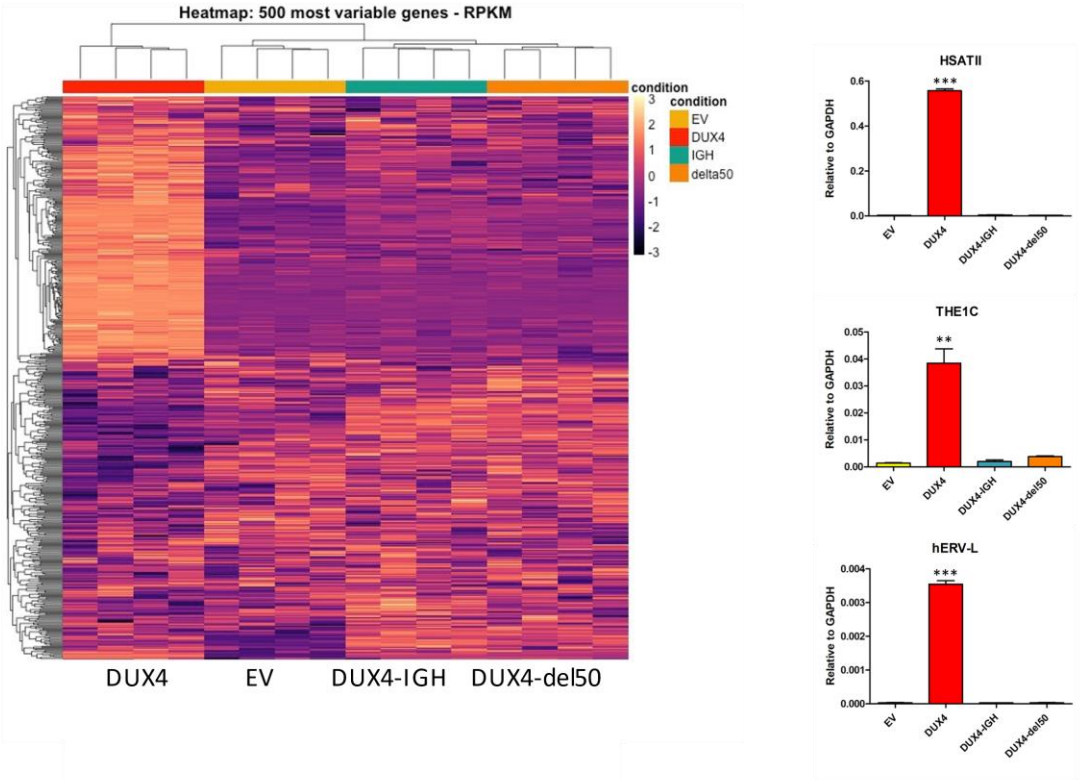


Figure 14. DUX4 but not DUX4-r activate the expression of repetitive DNA elements. *Heatmap (left) of the 500 most deregulated repetitive elements upon DUX4, DUX4-IGH and DUX4-del50 overexpression in REH cells. RT-qPCR analysis (right) of three representative repetitive elements activated by DUX4 in REH cells (One-Way ANOVA Analysis of Variance with Tukey's correction. ** p≤0,01 *** p≤0,001).*

Taken together, these results are in line with the known transcriptional activities of DUX4 and emphasizes once more the very different transactivation ability of DUX4-r variants.

3.1.5 Genome-wide DNA-binding analysis by CUT&Tag

My transcriptomic analyses demonstrate that DUX4 and DUX4-r variants regulate exclusive gene sets. This is very surprising in light of the fact that all DUX4-r variants associated to ALL described so far maintain a DNA-binding domain identical to that of DUX4 wt (Yosuke Tanaka *et al*, 2018; Yasuda *et al*, 2016; Zhang *et al*, 2016a). Moreover, published chromatin immunoprecipitation followed by sequencing (ChIP-seq) indicate an extensive overlap between DUX4 and DUX4-r genomic targets (Yosuke Tanaka *et al*, 2018; Liu *et al*, 2019; Zhang *et al*, 2016a). Nevertheless, these analyses suffer from several caveats. For example, in some study DUX4-r ChIP-seq in ALL cell lines was compared with previously published DUX4 ChIP-seq datasets in skeletal muscle cells or embryonic stem cells (Zhang *et al*, 2016b; De Iaco *et al*, 2017). In other studies, DUX4 and DUX4-r variants have been overexpressed at very high levels and ChIP-seq has been performed several days after DUX4/DUX4-r overexpression (Yosuke Tanaka *et al*, 2018; Yasuda *et al*, 2016).

ChIP-seq produces extensive protein-nucleic acids and protein-protein crosslinking due to relatively high concentration (10%) of formaldehyde for a long time (10 minutes). For transcription factors (TFs) this might lead to low signal, high background and epitope masking, and low yields requiring the use of large numbers of cells (Teytelman *et al*, 2013). Therefore, I decided to perform a side-by-side comparison of DUX4 and DUX4-r variants genomic targets upon minimal (12h) expression in REH cells using a more stringent approach. Cleavage Under Targets and Tagmentation (CUT&Tag) offers several advantages over traditional ChIP-seq. First, the protocol works with intact nuclei and no sonication allowing to better preserve chromatin architecture. Second, the Tn5 transposase is directly recruited to the site of TF binding leading to specific cleavage and release only of the DNA directly underneath the site of TF binding. Third, tagmentation generates fragments ready for PCR enrichment and provides amplified sequence-ready libraries in a day. Fourth, CUT&Tag provides exceptionally low backgrounds/high resolution results using low cell numbers or even single cells. Fifth, CUT&Tag requires low sequencing depth, translating in low-cost and simpler data analysis (Kaya-Okur *et al*, 2019; Henikoff *et al*, 2020).

The CUT&Tag protocol up to tagmentation is very similar to an immunofluorescence (IF). Therefore, to set up the best conditions, I used IF to compared four different anti-

DUX4 antibodies on REH upon DUX4/DUX4-r induction. Next, I compared different conditions for efficient nuclei isolation. Finally, I performed CUT&Tag trials followed by qPCR with primers for the genomic regions of known DUX4 targets to compare native to light cross-linking (0.1% formaldehyde for 2 min) and different amounts/ratio of nuclei/concanavalin A beads/buffers. Using the defined conditions (see Materials and Methods) I performed CUT&Tag on one replicate of control EV REH cells and two replicates of DUX4, DUX4-IGH and DUX4-del50 expressing REH cells. Overall, I obtained similar sequencing coverage in all samples expressing the transgene, while control EV REH cells showed less than 106 reads, highlighting the specificity of DNA tagmentation and library preparation (Figure 15A). I used the sequencing data from the EV control sample for the normalization of control to target data, as well as to generate an empirical threshold for peak calling. This identified 2228, 2020 and 1883 CUT&Tag peaks in the DUX4, DUX4-IGH and DUX4-del50 groups, respectively (Figure 15B). Looking at the genomic features which are more enriched underneath CUT&Tag peaks, I found similar distributions for DUX4 and DUX4-r over different genomic features (Figure 15).

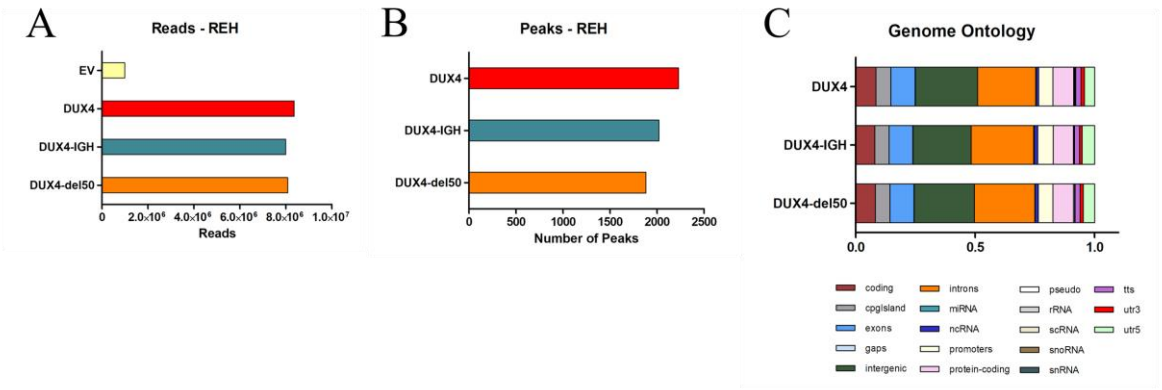


Figure 15. Overview of CUT&Tag experiment. *A. Barplot indicating the number of reads obtained within the different samples. B. Number of peaks identified through our Cut&Tag in the DUX4, DUX4-IGH and DUX4-del50 datasets. C. Annotation of the genomic features overrepresented under DUX4, DUX4-IGH and DUX4-del50 peaks identified by Cut&Tag.*

Surprisingly, despite sharing the same DNA-binding domain, I found that DUX4 and DUX4-r variants associate to largely non-overlapping genomic regions (Figure 16A) while, consistent with what observed at the transcriptomic level, DUX4-IGH and DUX4-del50 mostly bind to the same genomic loci. Only 160 (<10%) peaks were presented in all three groups, but expression of genes associated to those peaks was not

altered by DUX4 or DUX4-r versions in my RNA-seq, as it will be discussed more in detail later. These results are even more surprising considering that, by performing motif calling on the sequences falling under the respective peaks, I found that the DUX4 binding site is the most significantly enriched consensus sequence for both DUX4 and DUX4-r variants (Figure 16B).

Focusing on the RE classes, I found more DUX4 peaks falling in the REs genome

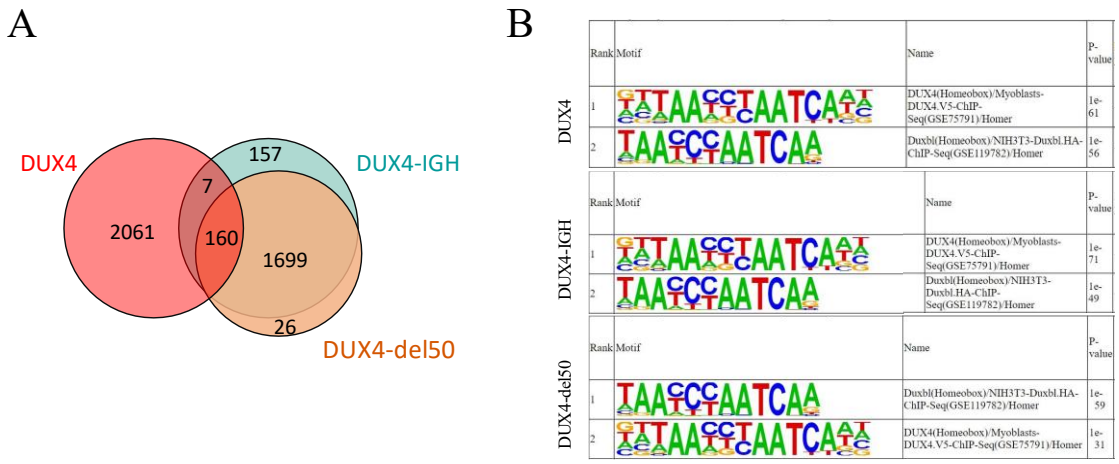


Figure 16. DUX4 and DUX4-r versions bind to distinct genomic regions. *A. Venn diagram showing the overlap between peaks identified in DUX4, DUX4-IGH and DUX4-del50 expressing REH cells. B. Motif calling analyses of peaks specifically associated to DUX4, DUX4-IGH and DUX4-del50.*

compared to DUX4-IGH. Moreover, the DUX4 and DUX4-IGH display enrichment for very different RE classes, consistent with the different ability of the two transcription factors to regulate the expression of repetitive elements (Figure 17).

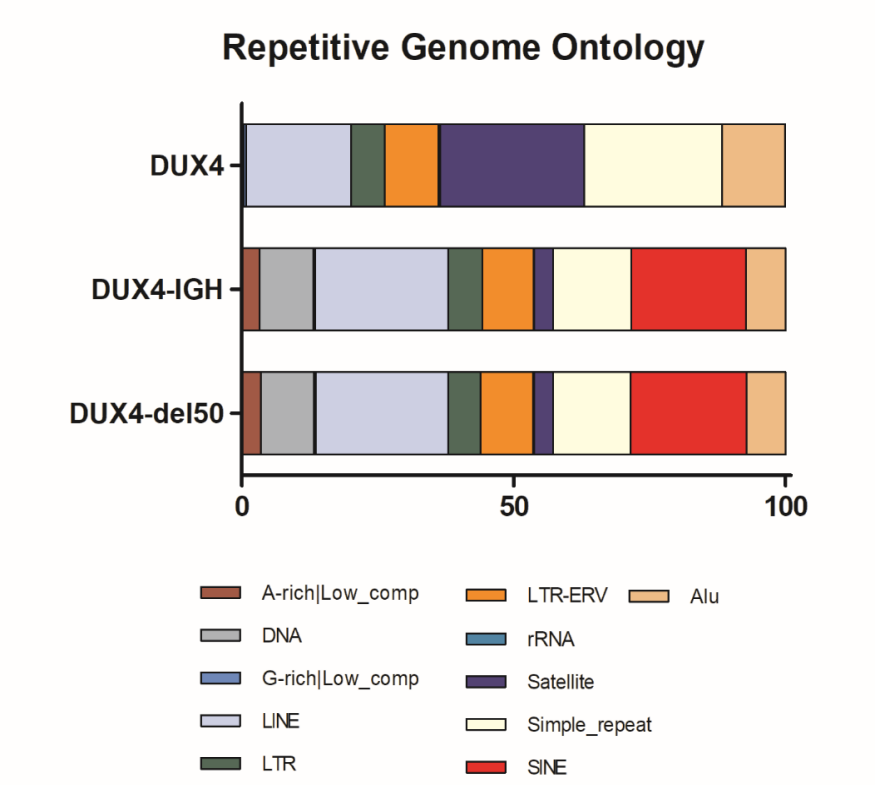


Figure 17. Annotation of the repetitive genome. *Barplot showing the nature of the repetitive element features overrepresented under DUX4, DUX4-IGH and DUX4-del50 peaks in REH cells.*

To identify direct DUX4 and DUX4-r protein-coding gene targets, I computed the overlap between those genes associated to the genomic regions selectively enriched for either TF and the list of its core targets defined earlier. This analysis identified 145 direct DUX4 and 154 direct DUX4-r high confidence target genes, respectively (Figure 18A).

In line with the literature (Snider *et al*, 2010; Geng *et al*, 2012; De Iaco *et al*, 2017), DUX4 direct targets in REH cells are mostly involved in the regulation of gene expression and pre-implantation development (Figure 18A). In contrast, direct DUX4-r targets are enriched in genes involved in cell adhesion, migration and cancer pathways (Figure 18B).

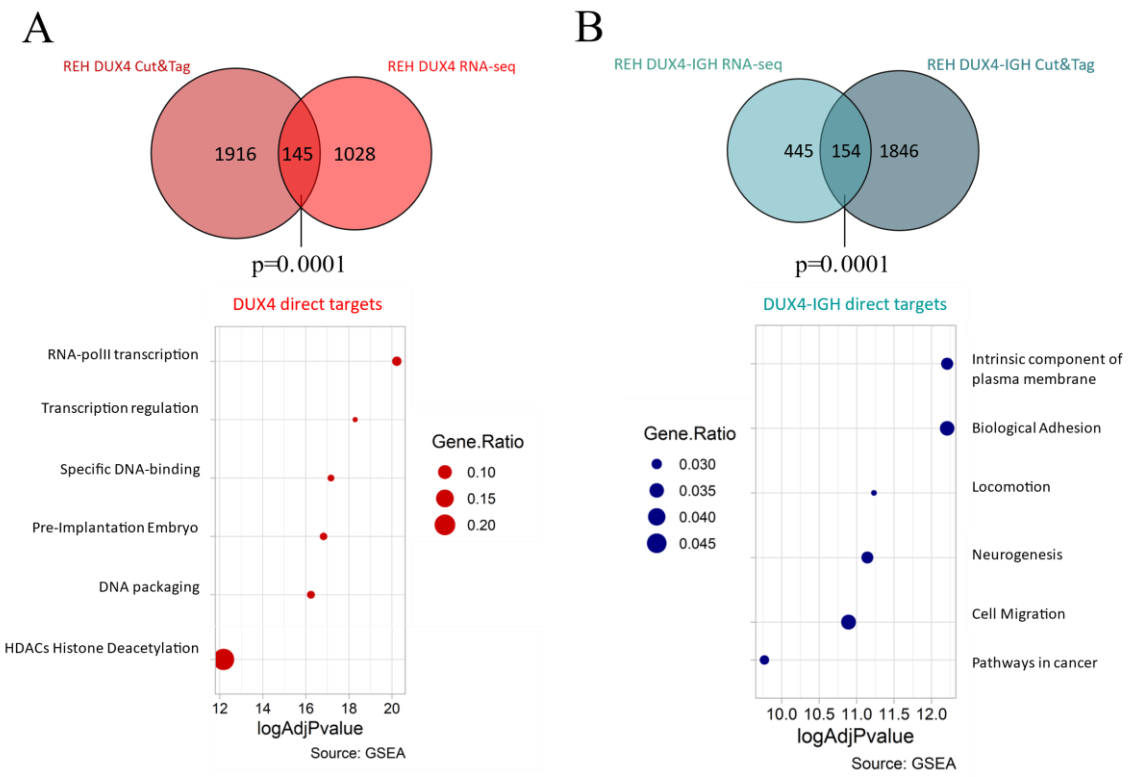


Figure 18. Functional enrichment analysis of DUX4 and DUX4-IGH direct transcriptional targets. *A. Venn diagram representation (top) of the overlap between DUX4 CUT&Tag and its core transcriptional targets. Enriched functional categories in the overlap between the list (bottom). B. Venn diagram representation (top) of the overlap between DUX4-IGH CUT&Tag and the transcriptional targets consistently deregulated in DUX4-IGH B-ALL patients datasets. Functional categories enriched in the overlap between the lists (bottom).*

ERGalt, the alternative *ERG* transcript previously involved in DUX4-r leukemogenesis provides an interesting example to compare my to previously reported results. Zhang *et al.* reported that DUX4 and DUX4-r variants are equally able to associate to exon 6alt and drive the expression of *ERGalt* (Zhang *et al*, 2016a). To do this, they compared publicly available results of RNA-seq and ChIP-seq in human muscle cells ectopically overexpressing high levels of DUX4 for several days to their own RNA-seq and ChIP-seq performed with the endogenous DUX4-IGH expressed in the ALL cell line NALM6. Instead, by comparing DUX4 and DUX4-r expressed for a minimal time in REH cells using CUT&Tag, I found that only DUX4-r variants are able to associate to exon 6alt to activate *ERGalt* expression (Figure 19A). Nonetheless, in my system DUX4-r-mediated *ERGalt* biogenesis does not involve massive deregulation of wt *ERG*, since its overall transcript level are comparable in all conditions (Figure 19B), which explains why *ERG*/*ERGalt* does not fall into the DUX4-r core targets.

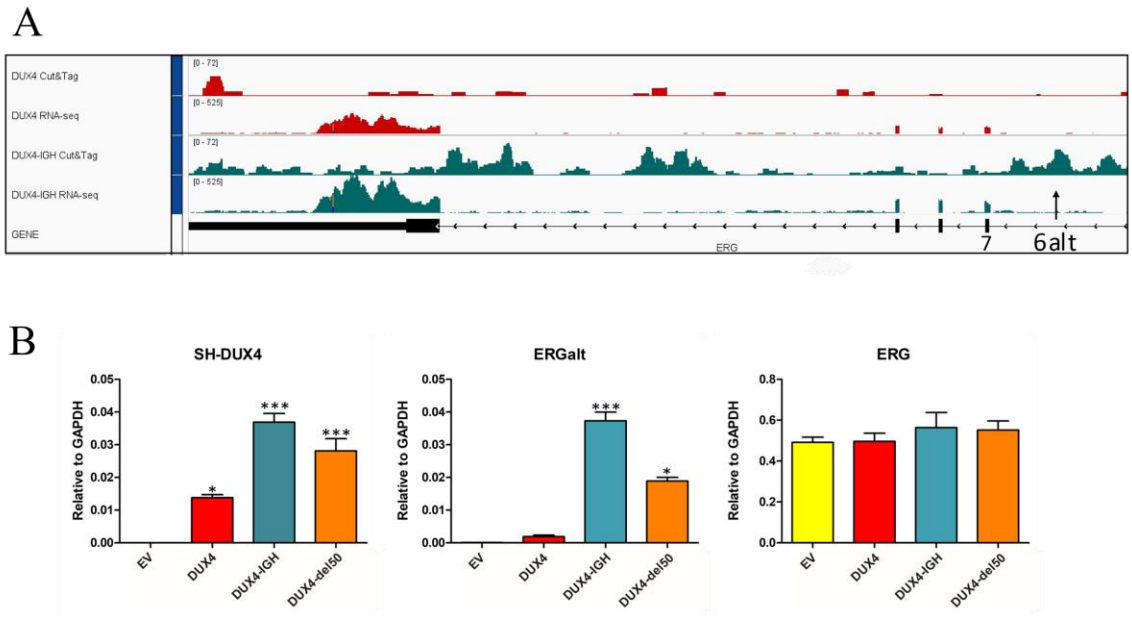


Figure 19. DUX4-r but not DUX4 binds to and activate expression of ERGalt. *A. IGV tracks showing the sequencing profile of our RNA-sequencing (upper lane) and CUT&Tag (lower lane) in DUX4 (red) and DUX4-IGH (green) expressing cells. B. RT-qPCR analysis of DUX4, ERGalt and total ERG expression levels in DUX4, DUX4-IGH and DUX4-del50 expressing cells (One-Way ANOVA Analysis of Variance with Tukey's correction. * p≤0,05 *** p≤0,001).*

In summary, for the first time, I discover that DUX4 and DUX4-r variants display very different transactivation abilities and I identified a set of genes that might confer unique functional properties to DUX4-r expressing cells and might be relevant for leukemogenesis.

3.1.6 DUX4-IGH activates the expression of a novel CD34 isoform

CD34 is a single-pass membrane protein involved in cell-cell adhesion, stem cell attachment to the bone-marrow matrix and cellular migration (AbuSamra *et al*, 2017; Healy *et al*, 1995; Nielsen & McNagny, 2008). Being an ubiquitous marker of progenitor cells, especially hematopoietic stem cells (HSC), CD34 selection is used clinically to enrich hematopoietic stem cells. Based on my analyses, *CD34* is one of the most consistent DUX4-r direct target genes. Nevertheless, DUX4-r ALL patient immunophenotypic data fail to detect CD34 positivity. There are two CD34 isoforms known, which are encoded by alternatively splicing transcripts each composed of 8 exons. The longer isoform contains the signal peptide (exon 1), mucin-like domain (exons 2 and 3), globular domain (exons 4 and beginning of exon 5), stalk domain (end

of exon 5, exon 6 and beginning of exon 7), single pass transmembrane domain (middle of exon 7) and cytoplasmic tail (end of exon 7 and exon 8). One alternative exon 8 introduces an earlier stop codon giving rise to an isoform lacking most of the intracellular domain. By integrating my RNA-seq and CUT&Tag data, I found that DUX4-IGH drives the expression of a novel *CD34* isoform produced from an alternative promoter located in intron 4 (Figure 20A). This isoform, which I named CD34-short (CD34-s), is present as uncharacterized CD34 isoform in UCSC, Ensembl (ENST00000367036.7) and UniProt (Q5JTA5) databases. CD34-s is thus encoded by a novel exon 1 joined to exons 5-6-7-8. Hence, CD34-s is predicted to lack the mucin-like domain, presenting a partially different globular domain and maintaining the stalk, transmembrane and cytoplasmic tail of full-length CD34 (Figure 20B). CD34-s lack of N-terminal extracellular domain is in line with the inability to detect CD34 by FACS immunophenotyping of DUX4-r ALL patients. Using RT-qPCR with isoform-specific primers, I confirmed that DUX4-IGH selectively activates the expression of *CD34-s*, which is the most abundant isoform in ALL cells endogenously expressing DUX4-IGH (Figure 20C). Using antibodies recognizing all CD34 isoforms, I confirmed my results also at protein level (Figure 20D). Given the important roles of CD34 full-length isoform in mediating cellular adhesion and migration in multiple contexts, it tempting to speculate that CD34-s could have a significant impact on the biology of expressing cells.

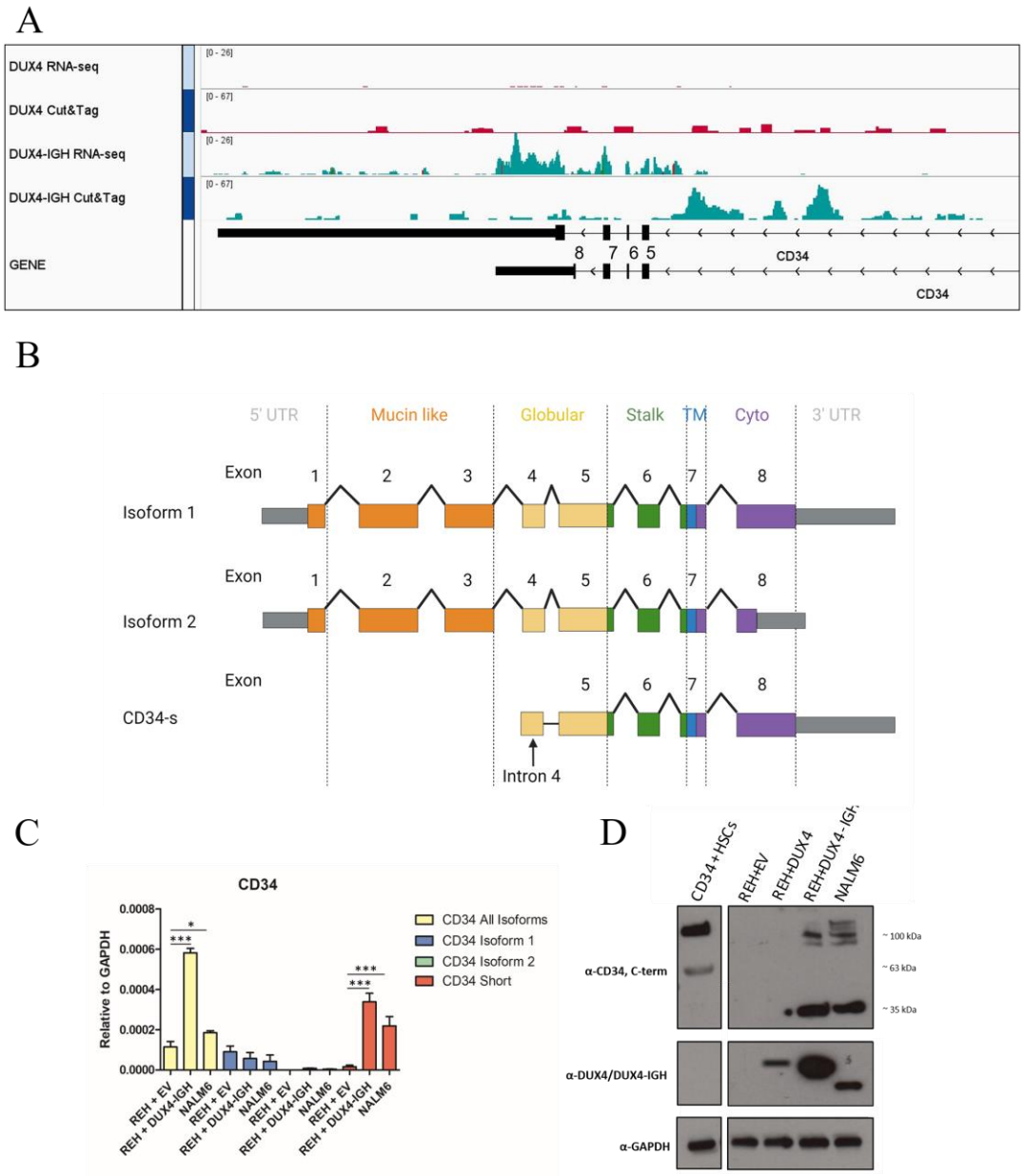


Figure 20. DUX4-IGH activates the expression of a novel CD34 isoform. *A. IGV tracks showing the sequencing profile of our RNA-sequencing (upper lane) IGV tracks showing the sequencing profile of our RNA-sequencing (upper lane) and Cut&Tag (lower lane) in DUX4 (red) and DUX4-IGH (green) expressing cells. B. Schematic representation of CD34-s transcript variant, as compared to the canonical isoforms 1 and 2. Image created with BioRender. C. RT-qPCR analyses of the expression levels of different CD34 isoforms in EV, DUX4, DUX4-IGH expressing REH cells as well as NALM6 cells (One-Way ANOVA Analysis of Variance with Tukey's correction. * p≤0,05 *** p≤0,001) D. Western Blot analysis of the protein levels of different CD34 isoforms in EV, DUX4, DUX4-IGH expressing REH cells, NALM6 cells and CD34+ve HSCs.*

3.1.7 DUX4-IGH induces migration of B-ALL cells

Functional enrichment analysis on direct DUX4-r target genes highlighted activation of cell adhesion as well as migration pathways. A high migratory behavior may result in increased dissemination of leukemic blasts, via the bloodstream, into secondary sites which are commonly the spleen and CNS (Jabbour *et al*, 2015; Gu *et al*, 2016). Furthermore, an higher adhesion rate of ALL cells is frequently associated with protection from chemotherapy and, together with deregulated migration, results in persistence of residual disease, which is at the basis of subsequent relapses(Fonseca *et al*, 2018; Sharma *et al*, 2021). Given the possible relevance of these aspects in the oncogenic activity exerted by DUX4-r, I performed a functional characterization DUX4-r variants expressing cells.

To test the migratory activity, I used NALM6 cells, which endogenously express DUX4-IGH. I generated NALM6 cells constitutively expressing the green fluorescent protein (GFP) marker (NALM6-GFP). Moreover, I stably introduced in these cells inducible short-hairpin vectors targeting *DUX4-IGH* (sh*DUX4-IGH*) or a non-targeting shRNA as control (sh*CTRL*). I used these cells to perform transwell migration assays. Briefly, human bone-marrow stromal (HS-5) cells were seeded in the bottom compartment of the transwell, while sh*CTRL* or sh*DUX4-IGH* cells were incubated on the upper compartment following induction of the shRNA construct. Importantly, I used transwells with a 5um permeable membrane separating the compartments to only allow passage of nutrients/cytokines while preventing free flow of NALM6 cells due to gravity. With this set up, I aimed at testing whether cells in the upper compartment actively migrate toward stroma cells located in the bottom well, possibly attracted by stroma-released cytokines. Intriguingly, I found that sh*CTRL* cells migrated toward the lower compartment of the well three times more with the respect to sh*DUX4-IGH* cells, in line with the possibility that DUX4-IGH promotes cell migration (Figure 21A). Furthermore, by performing a similar experiment, in which in the lower compartment stromal cells are co-cultured with either unmodified NALM6 cells or REH cells expressing EV or DUX4-IGH, I found that NALM6-GFP cells are more proficient in migrating toward cells expressing DUX4-IGH (Figure 21A). Hence, DUX4-IGH expressing cells are attracted by bone-marrow stromal cells and even more by other DUX4-IGH expressing leukemia cells. While this aspect may have consequences on the

adhesive properties of DUX4-IGH expressing cells, it remains to be established its relevance in the biology of DUX4-r ALL subgroup.

3.1.8 DUX4-IGH stimulates adhesion of ALL cells

To test whether DUX4-IGH confers adhesion properties to expressing cells, I generated inducible REH cells, which also constitutively express GFP (REH-GFP) to facilitate their visualization. Next, EV- or DUX4-IGH-expressing REH-GFP cells or NALM6-GFP cells were co-cultured with HS-5 stroma cells for 16h, before performing washes with PBS to remove non-adherent cells. After trypsinization, GFP-positive adherent cells were quantified by flow cytometry. Figure 21B shows that NALM6 and DUX4-IGH expressing REH cells adhere to the BM stroma cells significantly more that EV expressing REH control cells. Importantly, DUX4-IGH knockdown abrogates the increased adhesive properties of NALM6 or REH-DUX4-IGH cells, in line with a key role of DUX4-IGH in driving cell adhesion (Figure 21B).

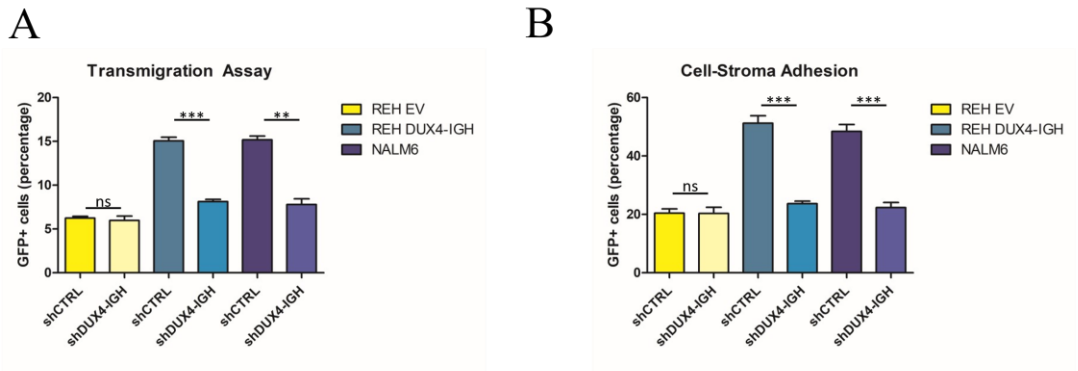


Figure 21. DUX4-IGH stimulates adhesion and migration of B-ALL cells. *A. Transmigration Assay showing that NALM6 cells are characterized by an enhanced ability to migrate toward BM stromal cells, particularly when co-cultured with DUX4-IGH expressing cells (One-Way ANOVA Analysis of Variance with Tukey's correction. **p≤0,01 *** p≤0,001). B. Cell-Stroma adhesion assay showing a significantly higher tendency of DUX4-IGH-expressing cells in adhering to bone marrow stromal cells, while such activity is completely impaired by shRNA-mediated KD of DUX4-IGH. (One-Way ANOVA Analysis of Variance with Tukey's correction. *** p≤0,001).*

Given their increased adhesive ability and their tendency to preferentially migrate toward each other, I wondered whether DUX4-r expressing cells could have an increased tendency to aggregate. To test this, I performed a spheroid assay. In brief, EV, DUX4, DUX4-IGH or DUX4-del50 REH cells were seeded at low density, their

transgene was subsequently induced and spheroid formation was monitored and quantified in real time with the fully automated Incucyte Live-Cell analysis system. As shown in Figure 22A and quantified in Figure 22B, I found that only DUX4-r expressing cells migrate toward each other generating areas of high cellular density, which result in the formation of large spheroids.

3.1.9 DUX4-IGH confers resistance to serum deprivation and promotes cellular proliferation

The peculiar migratory and adhesive behavior of DUX4-IGH expressing cells could protect them by unfavorable environmental conditions. To test this hypothesis, EV, DUX4 or DUX4-IGH REH cells were seeded at low density and serum-starving conditions and cell viability was scored during prolonged induction of respective transgenes (detail in the “Materials and Methods” section). As shown in Figure 22C, REH cells expressing DUX4-IGH display higher cellular fitness, proliferating significantly faster than the other two.

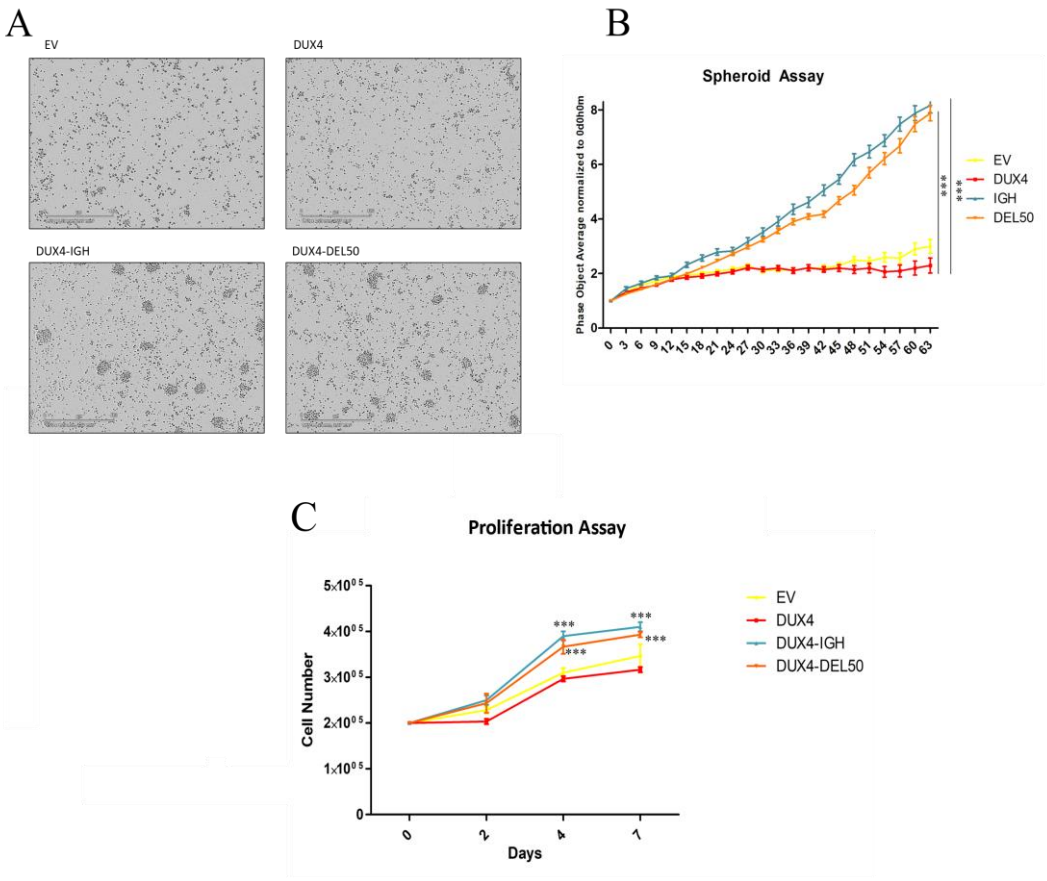


Figure 22. DUX4-IGH stimulates cellular aggregation and resistance to serum deprivation of B-ALL cells. *Representative phase microscopy images of large spheroids generated in liquid cell culture specifically by DUX4-r expressing REH cells. B. Real-time quantification of spheroid with the fully automated Incucyte Live-Cell analysis system, performed by computing the phase object average as normalized to T0 (Two-Way ANOVA Analysis of Variance with Bonferroni's correction to compare all columns. *** p≤0,001).*

Collectively, my results indicate that DUX4-IGH directly activates the expression of cell-adhesion molecules and mitogens/mitogens receptors, which mediate the ability of DUX4-IGH expressing cells to migrate, perform homotypic adhesion and growth in poor nutrients conditions.

3.1.10 DUX4-IGH fusions show a B-cell restricted activity

By coupling RNA-seq with CUT&Tag, I found out that at the basis of the divergent transcriptional activities of DUX4 and DUX4-IGH lies a differential ability of the two to bind respective genomic targets. However, the notion that the two proteins possess the same DNA-binding domain and the same nucleotide specificity hints to the fact that the DNA-binding activity of the two is not mediated uniquely by the DNA-binding domain, and that additional factors may be required for correct binding of the proteins to their respective genomic targets. To extend the validation of my RNA-seq and CUT&Tag results, I took advantage of HEK cells stably expressing inducible DUX4, DUX4-IGH as well as an EV control, which were previously established in the lab. Surprisingly, by performing RT-qPCR in this setting, I found that DUX4-IGH is unable to activate the expression of its direct target genes. Instead, DUX4 efficiently drove expression of its specific targets (Figure 23A). I obtained similar data in Jurkat (T-ALL) cells transduced with inducible DUX4, DUX4-IGH or EV vectors (Figure 23A). To further compare the activity of DUX4 and DUX4-IGH, I used a reporter plasmid in which nuclear GFP expression is regulated by multiple DUX4/DUX4-IGH binding sites placed upstream of a TATA box (Rickard *et al*, 2015). Despite small differences in the magnitude of their effect, in a transient transfection with a plasmid, DUX4 and DUX4-IGH were both able to induce the expression of the GFP reporter in HEK cells (Figure 23B).

Thus, despite having the same DNA-binding domain of DUX4 and being able to activate a DUX4-dependent reporter cloned in a plasmid, DUX4-IGH is unable to activate its genomic targets in a non-B-cell setting.

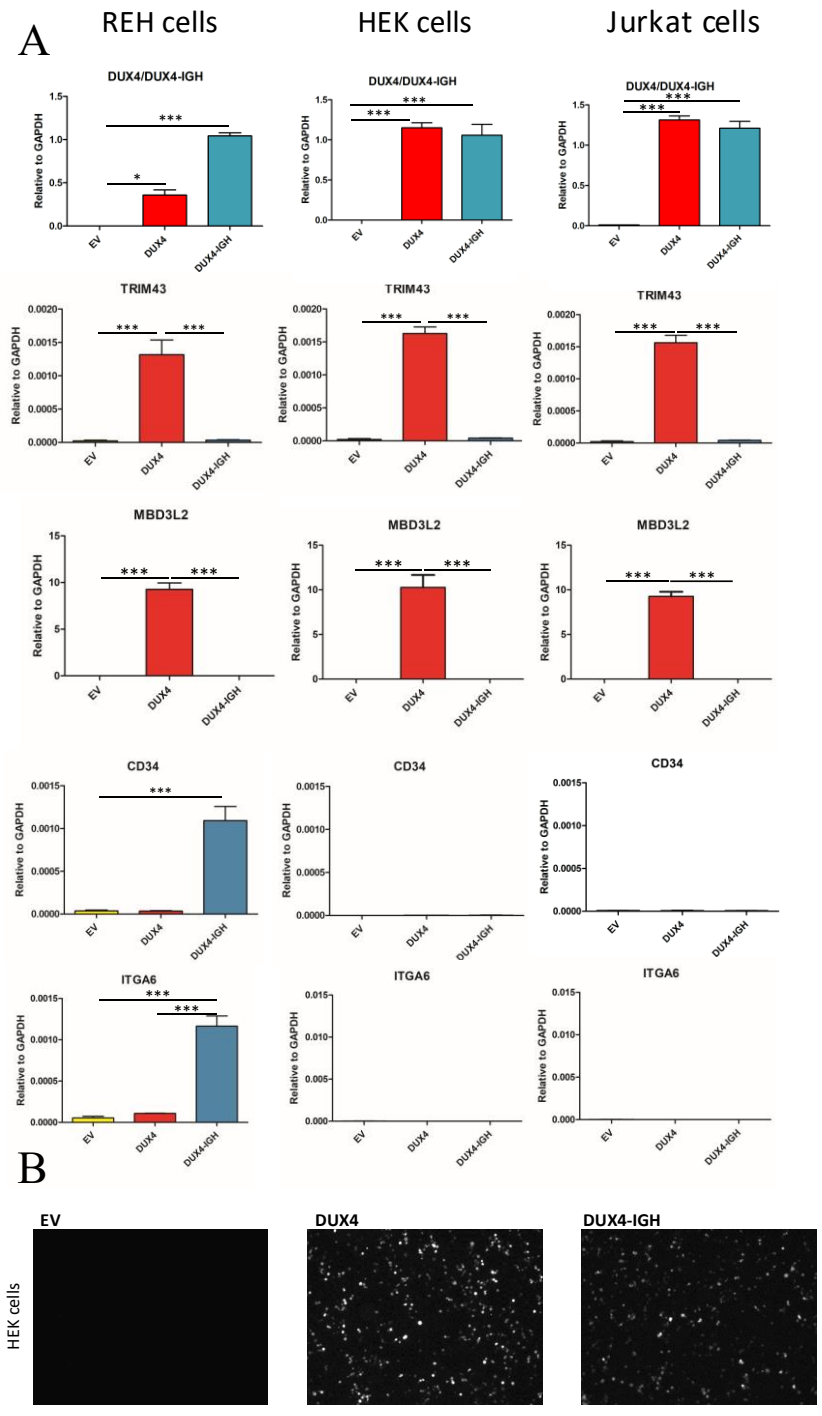


Figure 23. DUX4-IGH show a B-cell restricted transcriptional activity. *A. RT-qPCR analyses of DUX4 and DUX4-IGH direct target genes performed on REH, HEK and JURKAT cells expressing DUX4 or DUX4-IGH indicate that DUX4-IGH is only capable of activating its direct transcriptional targets only in a B-cell setting (One-Way ANOVA Analysis of Variance with Tukey's correction. * *p≤0,05, ** p≤0,01 *** p≤0,001). B. Epifluorescence microscopy images showing that in HEK cells both DUX4 and DUX4-IGH are able to activate the expression of an exogenous GFP reporter gene controlled by multiple DUX4 binding sites.*

3.1.11 DUX4-IGH binds to its target gene loci in HEK cells

To test whether DUX4-IGH is equelly able to bind to its direct transcriptional targets in a non-B cellular setting, I performed CUT&Tag on inducible HEK cells expressing DUX4 or DUX4-IGH, as well as EV control cells. Also in this case, while I obtained less than $2x10^6$ reads in the EV control sample, DUX4 and DUX4-IGH expressing HEK cells show comparable read coverage with more than 8 $x10^6$ reads (Figure 24A). As described previously, peak calling was performed through normalization of the target signal as compared to the EV control, and this approach identified 3177 and 2411 peaks in DUX4 and DUX4-IGH expressing cells respectively (Figure 24B). Again, I found minimal overlap between DUX4 and DUX4-IGH associated peaks (Figure 24B). Also in this case, motif calling on the genomic regions specifically bound by either of the two indicated the DUX4 consensus motif as the most enriched one (Figure 25A). Importantly, by performing differential peak calling, I found that, while DUX4 binds to its targets with comparable magnitude in both cell lines, DUX4-IGH displays significantly higher CUT&Tag signal at its target genes in REH versus HEK cells (Figure 25B).

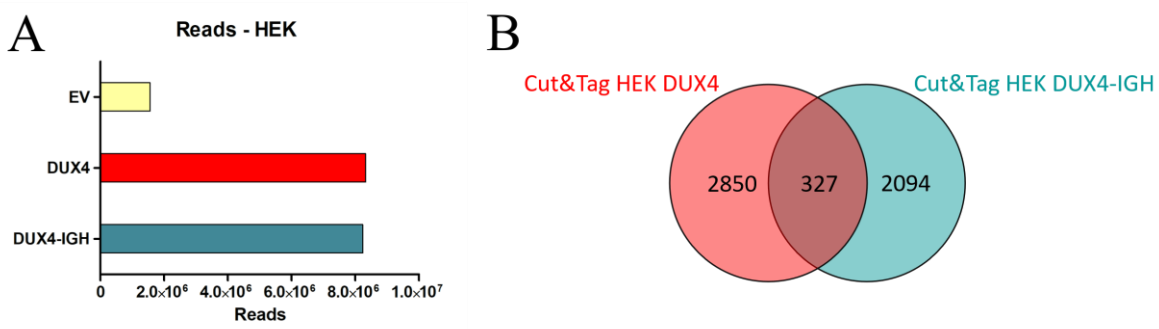


Figure 24. CUT&Tag analyses of HEK cells expressing inducible EV, DUX4 or DUX4-IGH. *A. Barplot showing the relative sequencing coverage of EV, DUX4 or DUX4-IGH HEK cells. B. Venn diagram showing the overlap between genes bound by DUX4 and DUX4-IGH in HEK cells.*

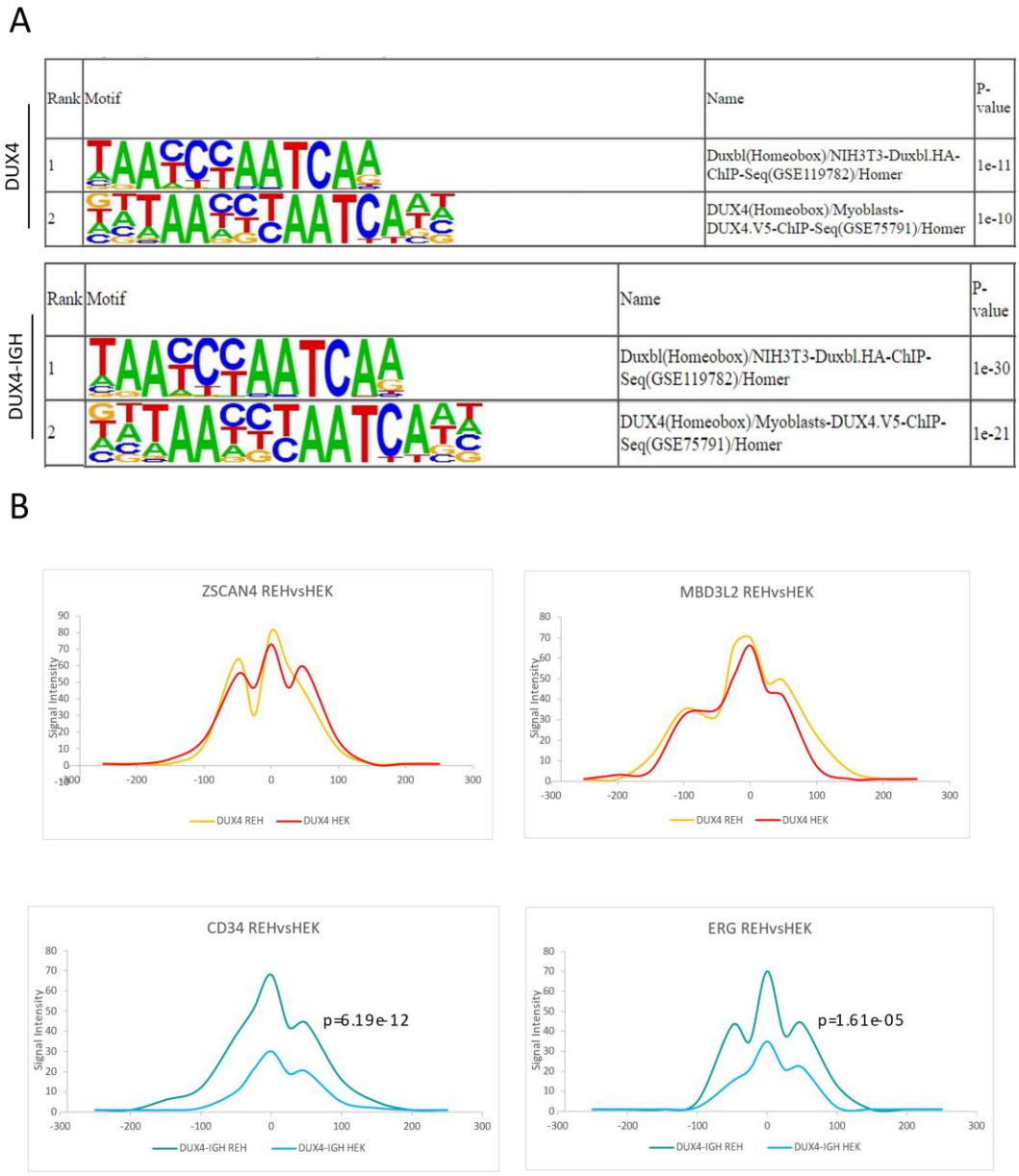


Figure 25. DUX4-IGH displays reduced association to its target loci in HEK cells. *A. Motif calling analysis performed on DUX4 and DUX4-IGH specific peaks in HEK cells. The two most enriched motives are reported. B. CUT&Tag signal intensity for DUX4 in HEK (red) vs REH (orange) cells (left), DUX4-IGH in HEK (cyan) vs REH (green) cells (right). The result for two representative DUX4/DUX4-IGH targets are shown.*

Taken together, my data suggest that DUX4-IGH inability to activate its targets in HEK cells is due to inefficient genomic association. Based on these results, I hypothesized that for its activity DUX4-IGH requires a co-factor enriched in B-cells.

3.2 QUANTITATIVE PROTEOMICS FOR SELECTIVE DUX4 IGH INTERACTORS

My results so far indicate that DUX4-r displays a very different activity with respect to DUX4, which could be mediated by its ability to interact with co-factors preferentially expressed in B cells. To test this hypothesis, I took advantage of the fact that the inducible DUX4 and DUX4-r versions that I have generated are fused to a streptavidin-binding peptide and a hemagglutinin tag (SH-tag). Therefore, I performed tandem affinity purification under high stringency prior to mass spectrometry (TAP-MS), to reduce nonspecific background binding and identify tight interactors (Rigaut *et al*, 1999; Köcher & Superti-Furga, 2007; Glatter *et al*, 2009; Li, 2011). TAP was performed using nuclease-treated and pre-cleared nuclear extracts from dox-induced EV, DUX4, DUX4-IGH and DUX4-del50 REH cells in quadruplicate followed by quantitative mass spectrometry. Furthermore, to determine whether the identified interactors could be enriched in a B-cell setting, I also analyzed the nuclear proteomes of REH, HEK and JURKAT cells to possibly fish out DUX4-IGH interactors more enriched in B-cells. I sorted the MS results based on: I. significant enrichment of DUX4 or DUX4-r associated proteins over the EV control; II. significant enrichment of proteins in DUX4-r over DUX4; III. DUX4-r interactors enrichment in REH over Jurkat and HEK cells. My analysis revealed 52, 50 and 47 proteins interacting with DUX4, DUX-IGH and DUX4-del50, respectively (Figure 26A). Importantly, the DUX4 and DUX4-r signals were comparable across the different samples indicating that the differential protein association with DUX4 or DUX4-r variants is not due to different efficiency of DUX4 or DUX4-r affinity purification (Figure 26A). Next, I compared the enrichment of DUX4, DUX4-IGH and DUX4-del50 interactors using the nuclear proteomes of HEK and REH cells (Figure 26B). Through this analysis, I identified 13 proteins selectively associated with DUX4-IGH and DUX4-del50 that are more abundant in REH compared to HEK cells (Figure 26C).

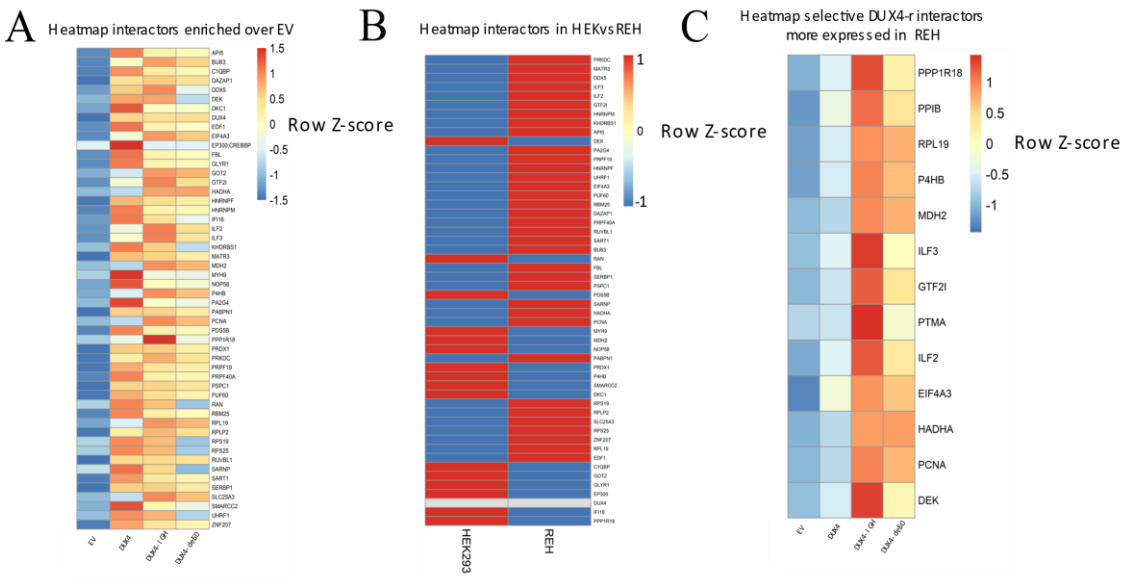


Figure 26 Proteomics analyses. *A. Heatmap showing the 52 interactors enriched in DUX4, DUX4-IGH and DUX4-del50 expressing REH cells. Relative protein enrichment per sample has been normalized by row to allow clear visualization of the enrichment. B. Heatmap of the nuclear protein levels of the 52 interactors in HEK and REH cells. C. Heatmap showing DUX4-r selective interactors more expressed in REH cells. Relative protein enrichment per sample has been normalized by row to allow clear visualization of the enrichment.*

As expected, the p300 transcriptional co-activator was identified as DUX4 selective interactor in line with the fact that the p300 interaction domain is missing in DUX4-r variants (Figure 27A). Among the selective DUX4-r interactors preferentially expressed in REH cells, the most enriched in both DUX4-IGH and DUX4-del50 groups was the General Transcription Factor 2 I (GTF2I/TFII-I) (Figure 27B). Despite its name, GTF2I is a sequence specific and signal-induced transcription factor involved in a variety of gene regulatory processes which, depending on the cellular context, can act as transcriptional activator or repressor (Roy, 2012). GTF2I activity is regulated through phosphorylation by tyrosine kinases downstream several signaling pathways, including the pre-BCR (Chailangkarn *et al*, 2018; Roy, 2017). *GTF2I* translocations have been associated to acute lymphoblastic and acute promyelocytic leukemia (Li *et al*, 2015; Panagopoulos *et al*, 2019). Moreover, a recurrent hotspot *GTF2I* mutation is the most frequent oncogenic driver in thymic epithelial tumors (Feng *et al*, 2017; Petrini *et al*, 2014). I hence decided to focus on GTF2I for the remaining part of my PhD.

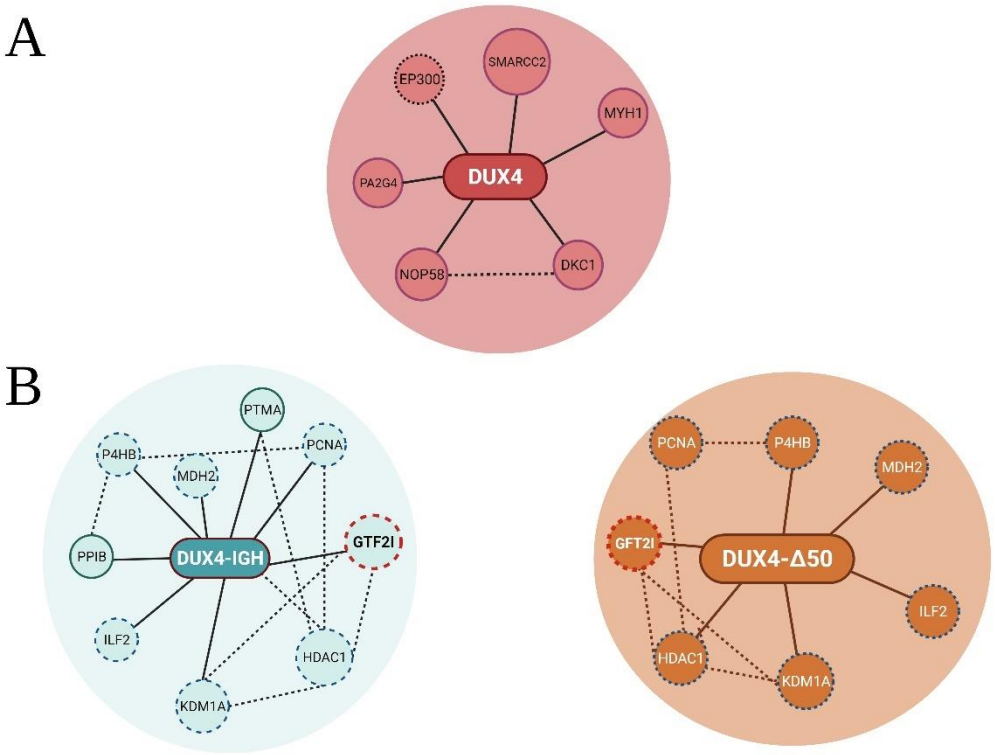


Figure 27. Network analysis of the DUX4, DUX4-IGH and DUX4-DEL50 interactomes. *Protein-protein ranked interaction networks have been generated using the SAINT package, by performing the empirical fold change score and saint (SP) score as compared to the EV control sample to compute the enrichment of DUX4, DUX4-IGH and DUX4-DEL50 with respect to the background. A. Selective DUX4 interactors having a Log2FC enrichment >1,5 over EV and DUX4-r. EP300, which is more expressed in HEK cells, is circled in black. B. Selective DUX4-IGH/DUX4-DEL50 interactors having a Log2FC enrichment >1,5 over EV and DUX4. Common interactors are circled in blue, the most enriched interactor (GTF2I) in red. Image created with BioRender.*

3.2.1 Validation of GTF2I interaction with DUX4-r variants

To validate GTF2I as a selective DUX4-r interactor, I performed new Strep-HA tandem affinity purifications using nuclear extracts of REH cells expressing EV, DUX4, DUX4-IGH or DUX4-del50. As shown in Figure 28A, GTF2I selectively copurified with DUX4-r variants. Moreover, immunoblotting on the total protein extracts confirmed that GTF2I is significatively more abundant in REH compared HEK cells (Figure 28B).

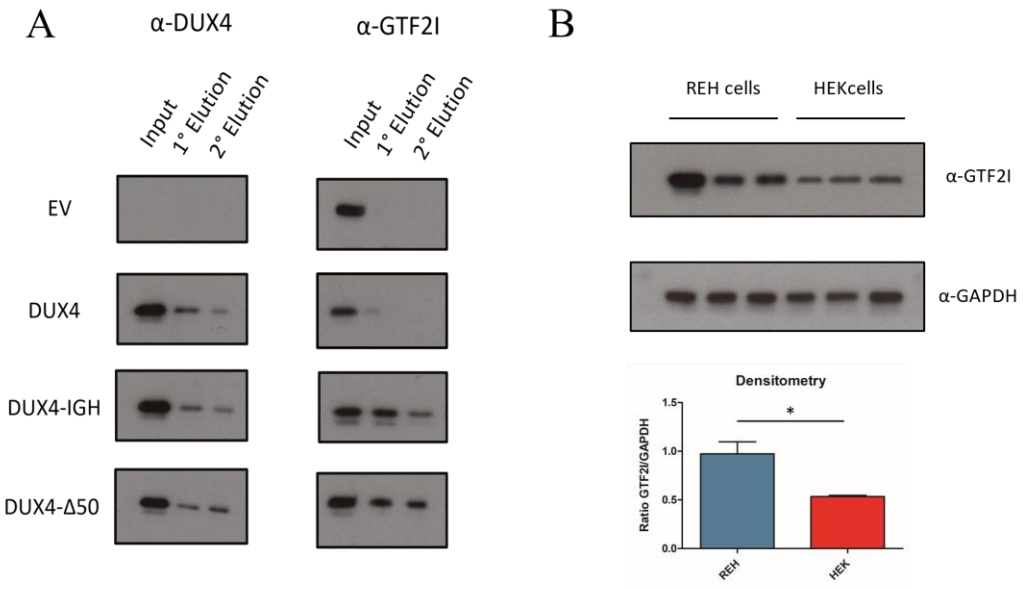


Figure 28. GTF2I specifically interacts with DUX4-IGH and is expressed at higher levels in REH cells as compared to HEK cells. *A. Tandem affinity purification followed by immunoblotting was performed to validate the specific interaction of GTF2I with DUX4-r, but not with wt DUX4. B. Western blot analyses of whole cells extract from REH and HEK cells showing a significantly higher abundance of GTFI in REH cells (Student's T-test, * P≤0,05).*

3.2.2 GTF2I binds to DUX4-r target gene loci.

Despite its name, GTF2I is a sequence-specific transcriptional activator or repressor, which work through the binding to initiator elements of its target genes or other cis-regulatory DNA-binding sequences. Given the fact there are no publicly available ChIP-seq or CUT&Tag datasets for GTF2I in ALL, I performed CUT&Tag for GTF2I and identified 4390, 4489 and 4552 peaks in EV, DUX4 and DUX4-IGH REH cells, respectively (Figure 29A). More than 90% of them were shared between all the

samples, leaving only 37, 62 and 125 peaks selectively associated with EV, DUX4 and DUX4-IGH REH cells, respectively (Figure 29B).

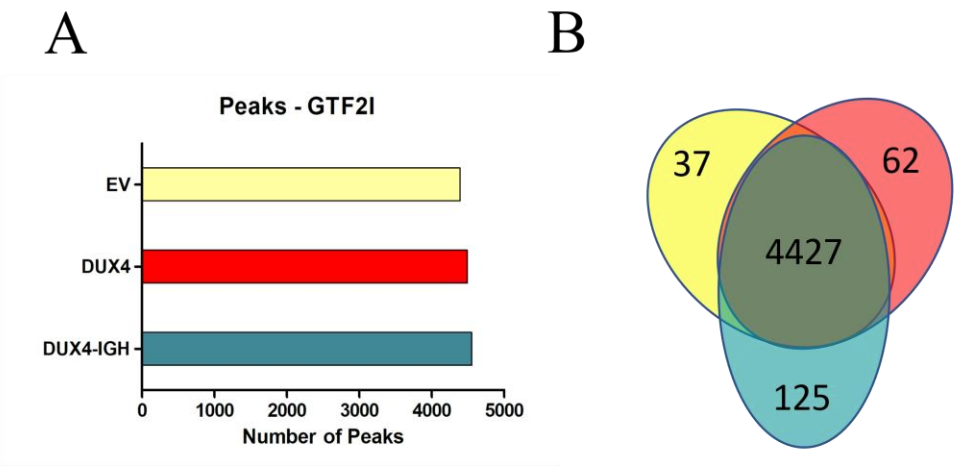
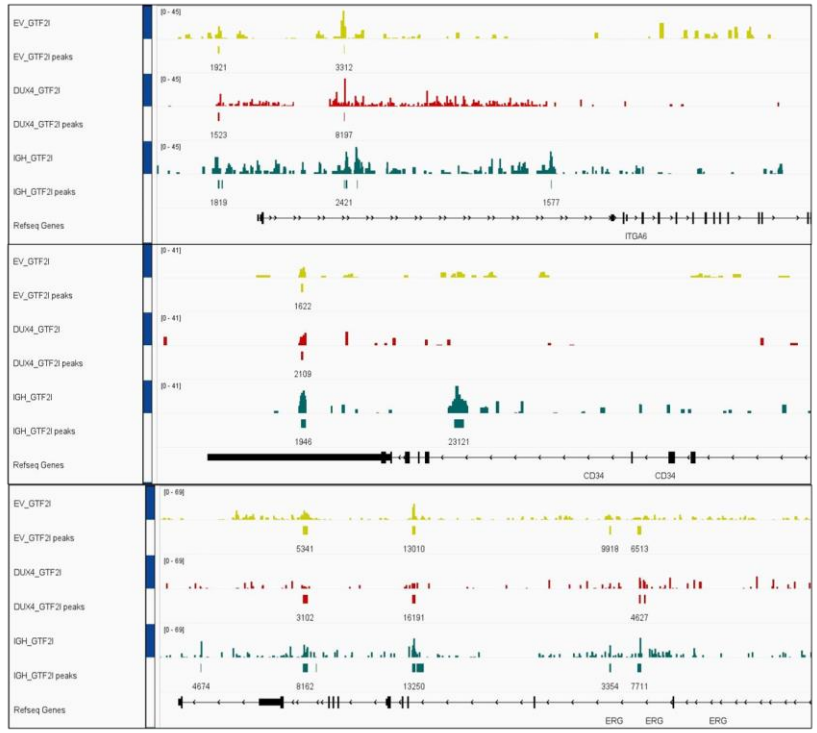


Figure 29. Genomic analysis of GTF2I association to DNA. *A. Barplot showing the relative sequencing coverage of EV, DUX4 and DUX4-IGH REH cells while targeting GTF2I. B. Venn diagram showing the overlap between genomic regions associated to GTF2I in EV (yellow), DUX4 (red) and DUX4-IGH (green) expressing REH cells.*

Despite the comparable number of GTF2I peaks in EV, DUX4 and DUX4-IGH REH cells, by performing differential peak calling analysis, I found that DUX4-r direct targets selectively display a significantly higher GTF2I signal in DUX4-IGH cells compared to both EV and DUX4 cells (Figure 30A). Accordingly, by performing motif calling analysis, I found significant enrichment for the DUX4 consensus motif exclusively in DUX4-r associated peaks (Figure 30B).

A



B

Rank	Motif	Name	P-value
76		Duxbl(Homeobox)/NIH3T3-Duxbl.HA-ChIP-Seq(GSE119782)/Homer	1e-7
108		DUX4(Homeobox)/Myoblasts-DUX4.V5-ChIP-Seq(GSE75791)/Homer	1e-4

Figure 30. GTF2I is enriched in DUX4-r target genes loci. *A. IGV genomic tracks (top panel) and differential peak calling (bottom panel) showing GTF2I association to three direct DUX4-r targets, ERGalt, CD34 and ITGA6, as well as to non DUX4-r targets DHFR, ARGHEF1 and AKAP6 in EV (yellow), DUX4 (red) and DUX4-IGH (green) expressing REH cells. Both the sequencing (upper lane) and peak (lower lane) profiles are shown. B. Motif calling analysis on GTF2I peaks identified in EV, DUX4 and DUX4-IGH expressing cells show significant enrichment in the DUX4 consensus motif only in DUX4-IGH expressing cells.*

Altogether, these results suggest that DUX4-r recruits GTF2I to their target loci, which prompted me to test the relevance of GTF2I in the transactivation ability of DUX4-IGH in B-ALL cells.

3.3 EVALUATE THE BIOLOGICAL RELEVANCE OF FINDINGS IN PRE-CLINICAL SETTINGS

3.3.1 GT2FI downregulation by SAHA

Haploinsufficiency or duplication of the genomic region encompassing *GTF2I* in 7q11.23, and its subsequent transcriptional deregulation, is implicated in neurological conditions such as Williams-Beuren Syndrome (WBS) and autism spectrum disorder (7dup). Mouse models and human studies showed that GTF2I plays a critical for role in the pathogenesis of WBS and 7Dup (Malenfant *et al*, 2012). In a recently published paper, an high-throughput screening looking for molecules able to restore the correct expression levels of WBS/7dup genes identified vorinostat/suberoylanilide hydroxamic acid (SAHA), a small-molecule inhibitor of histone deacetylases (HDACi), as a selective repressor of GTF2I expression at RNA and protein levels (Cavallo *et al*, 2020).

SAHA treatment has been shown to be effective against a number of hematological malignancies, including leukemia, by promoting cell-cycle arrest and apoptosis through regulation of gene expression. Importantly, it has been approved by the U.S. Food and Drug Administration for the treatment of Cutaneous T-Cell Lymphoma (CTCL) and is currently being evaluated for use against acute myeloid leukemia.

First, I tested whether, also in our system, SAHA treatment was able to downregulate the mRNA level of GTF2I. Indeed, as shown in Figure 31, a non-toxic concentration (1 uM) of SAHA was able to significantly downregulate GTF2I expression levels. Based on this, I decided to test whether SAHA-mediated downregulation of GTF2I could impact the ability of DUX4-IGH to activate its target genes.

3.3.2 GTF2I downregulation selectively blocks DUX4-IGH transcriptional activity

After validating the effect of SAHA in my system, I tested whether SAHA treatment could interfere with the transcriptional activity of DUX4-IGH. At 1 uM, SAHA had no significant effect on DUX4 or DUX4-IGH levels. Instead, SAHA treatment almost completely ablate activation of DUX4-IGH target genes, with minimal effect on DUX4 targets (Figure 31). This result is in line with a requirement for GTF2I selectively for the activation of DUX4-IGH targets.

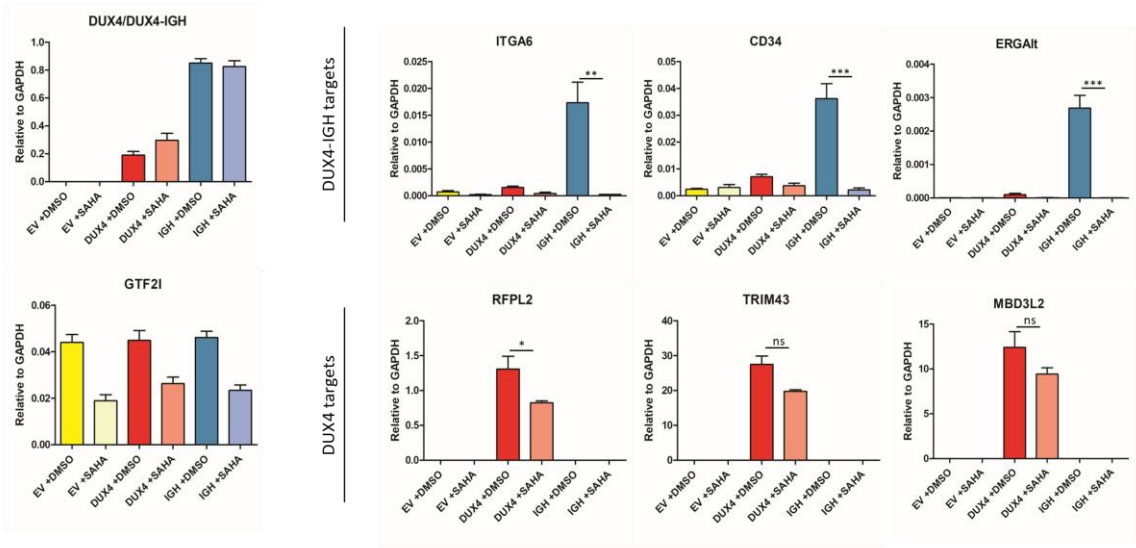


Figure 31. GTF2I downregulation selectively impairs DUX4-IGH transcriptional activity. *RT-qPCR analysis of the effects of GTF2I downregulation by 1uM SAHA treatment for 24hon the transcriptional activity of DUX4 and DUX4-IGH in REH cells (Two-Way ANOVA Analysis of Variance with Bonferroni's correction to compare all columns. * p≤0,05, **p≤0,01, *** p≤0,001).*

3.3.3 GTF2I downregulation blocks DUX4-IGH-induced cell adhesion and migration

The selective impairment of DUX4-IGH transcriptional activity observed upon SAHA treatment, prompted me to investigate whether such effects at the transcriptional level could also translate in functional impairment of DUX4-IGH expressing cells. Intriguingly, I found that upon SAHA treatment DUX4-IGH expressing REH cells lose the ability to perform homotypic adhesion and generate spheroids, becoming similar to EV control REH cells (Figure 32A). SAHA treatment also significantly reduced the ability of NALM6-GFP cells to transmigrate toward BM-stromal cells (Figure 32B).

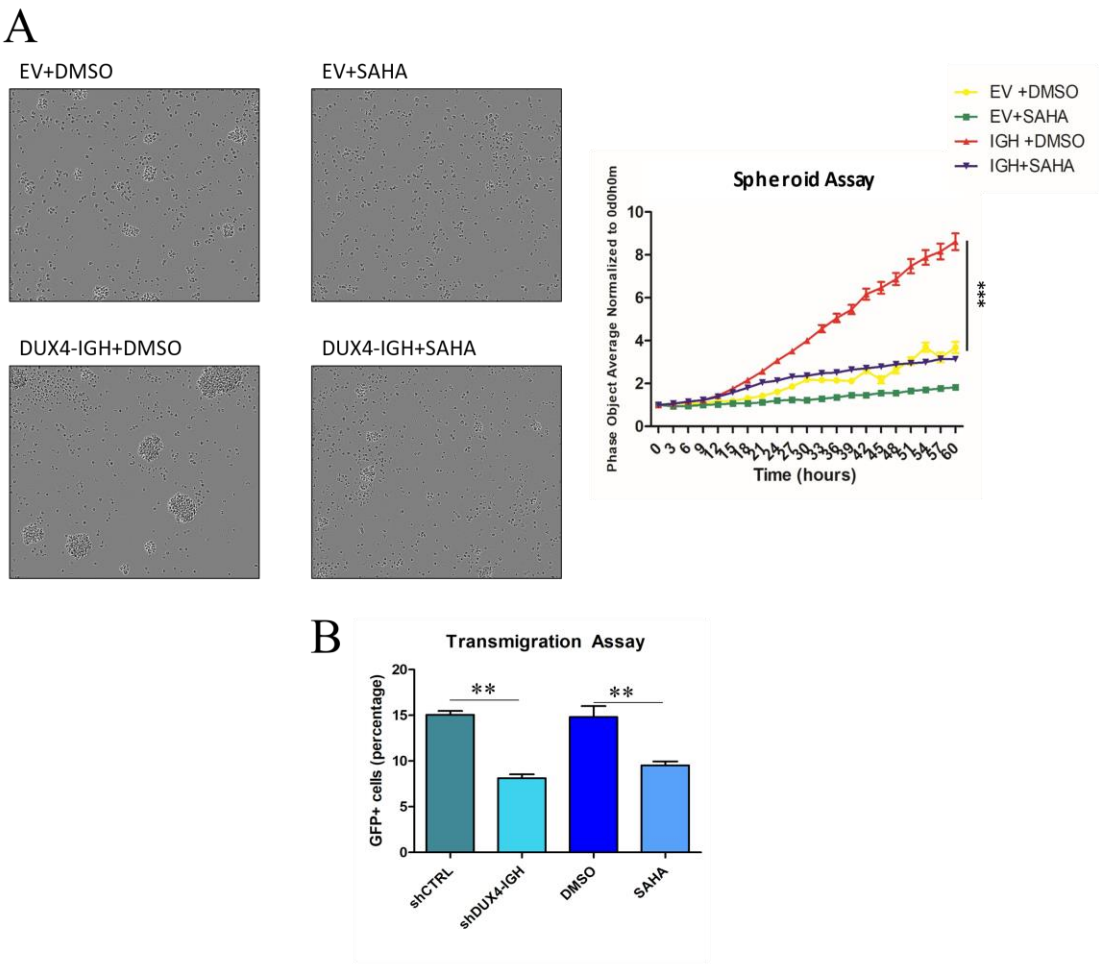


Figure 32. SAHA treatment blocks cellular adhesion and migration induced by DUX4-IGH. *A. Representative light microscopy images (left) and phase-object average quantification showing how aggregation of DUX4-IGH-expressing REH cells is impaired by 1uM SAHA treatment (Two-Way ANOVA Analysis of Variance with Bonferroni's correction. *** p≤0,001). B. Transmigration assay showing a comparable effect of SAHA treatment and shRNA-mediated DUX4-IGH knockdown in impairing migration of NALM6 cells (Student's T-test, ** p≤0,01).*

3.3.4 GTF2I downregulation abrogates proliferation and causes apoptosis selectively of DUX4-IGH expressing cells

I found that DUX4-IGH expressing are exquisitely sensitive to SAHA treatment. At 1 uM, SAHA treatment significantly reduced proliferation of both DUX4-IGH expressing REH and NALM6 cells, while EV or DUX4 expressing REH cells were unaffected. In the case of inducible DUX4-IGH REH cells, SAHA brought back proliferation to the level EV control REH cells (Figure 33A). In NALM6 cells, the reduction of proliferation caused by SAHA treatment was comparable to that caused by DUX4-IGH knockdown (Figure 33A).

Using 10 uM SAHA, I observed a minimal effect on the viability on control REH cells, while apoptosis was significantly increased in both NALM6 and REH cells expressing DUX4-IGH (Figure 33B)

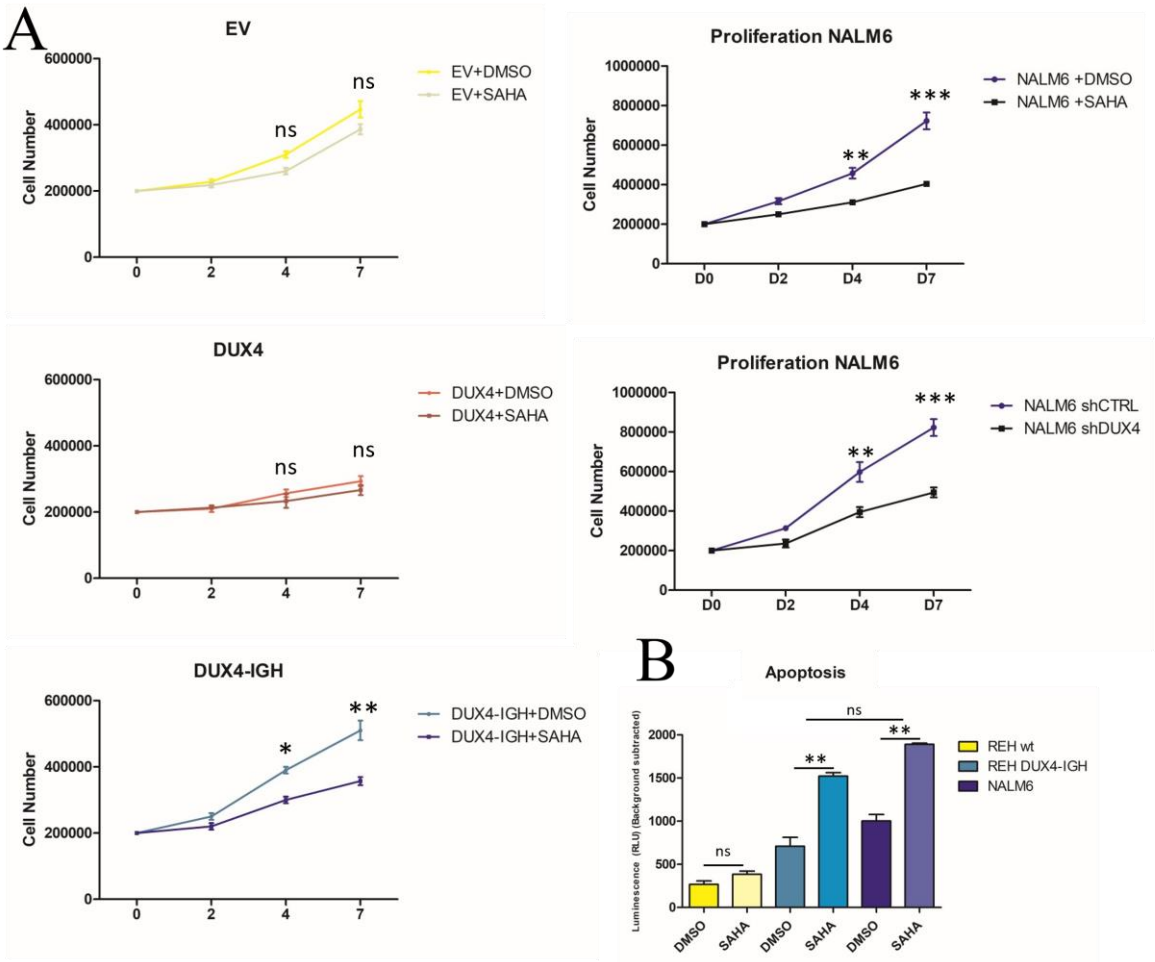


Figure 33. SAHA-mediated GTF2I downregulation blocks cell proliferation and induces apoptosis selectively to DUX4-IGH expressing B-ALL cells. *A. Proliferation assays in serum-starving conditions of EV, DUX4 and DUX4-IGH expressing REH cells and NALM6 cells treated with 1uM SAHA or, in the case of NALM6, also in with shRNA targeting DUX4-IGH (Two-Way ANOVA Analysis of Variance with Bonferroni's correction. * p≤0,05, ** p≤0,01, *** p≤0,001. B. Apoptosis/Caspase assay in wt REH cells, REH cells expressing DUX4-IGH, NALM6 cells upon 10uM SAHA treatment for 48h One-Way ANOVA Analysis of Variance with Tukey's correction. * *p≤0,05, ** p≤0,01 *** p≤0,001.*

3.3.5 Forced expression of GTF2I in HEK cells endows DUX4-IGH activity

As shown in Figure 23, DUX4-IGH is unable to activate its target genes in HEK cells. Since I found that GTF2I is required for DUX4-IGH activity and that GTF2I is expressed at significantly lower level in HEK compared to REH cells, I hypothesized that in HEK cells there is not enough GTF2I to interact with DUX4-IGH and allow for its activity. Hence, I wondered if increasing GTF2I level in HEK cells could make them responsive to DUX4-IGH. To test this, I transiently co-transfected in HEK cells GTF2I (or the empty vector as control) in combination with either inducible vectors encoding EV, DUX4 or DUX4-IGH. GTF2I was expressed at similar levels in all conditions and did not significantly altered DUX4 or DUX4-IGH levels (Figure 34A). While it did not significantly altered the ability of DUX4 to activate its target genes, GTF2I overexpression together with DUX4-IGH induction was associated to a significant upregulation with respect to DUX4-IGH induction in cells transfected with EV (Figure 34B). Hence, GTF2I is sufficient to make HEK cells permissive for DUX4-IGH transcriptional activity further supporting its role in transcriptional regulation by leukemogenic DUX4-r variants.in transcriptional regulation by leukemogenic DUX4-r variants.

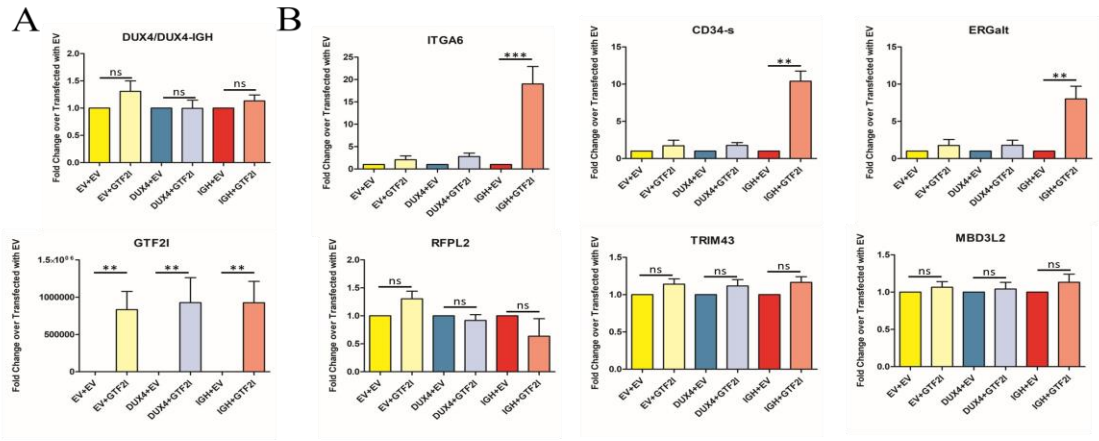


Figure 34. Transient GTF2I expression in HEK cells endows DUX4-IGH transcriptional activity. *RT-qPCR analysis of DUX4/DUX4-IGH and GTF2I expression levels in HEK cells co-transfected with EV, DUX4 or DUX4-IGH in combination with an EV control or a GTF2I-encoding plasmid (One-Way ANOVA Analysis of Variance with Tukey's correction. * *p≤0,05, ** p≤0,01 *** p≤0,001). B. RT-qPCR analysis of DUX4/DUX4-IGH transcriptional targets upon forced GTF2I expression in HEK cells (One-Way ANOVA Analysis of Variance with*

Collectively, my results strongly suggest that DUX4-r variants in ALL acquire their leukemogenic ability through a gain of interaction with GTF2I.

4. DISCUSSION

Altogether, results from my PhD project support a novel mechanism through which DUX4 rearrangements drive leukemia development. I hypothesize that the rearrangements lead to both a loss and a gain of function. Deletion of its natural C-terminal transactivation domain, lead to DUX4 loss of interaction with CBP/p300 transcriptional co-activators. On the other hand, this genetic event represents a gain of interaction with the transcription factor GTF2I. I believe that the co-activator switch is key for the acquisition of DUX4-r leukemogenic activity.

My studies revealed that DUX4 and DUX4-r bind to and activate highly different genes sets, even though they share the very same DNA-binding domain. Strikingly, the most enriched consensus motif underneath the respective peaks is still the one of DUX4, indicating that the DNA-binding activity of DUX4 and DUX4-r is not solely determined by the DNA binding domain, and that additional players could play a role in target gene definition. While DUX4 shows a similar transcriptional activity in every cell type that I have tested, the activity of DUX4-r is intimately linked to GTF2I cellular availability. DUX4-r ability to activate its targets is impaired by GTF2I downregulation and enhanced by its overexpression. At genomic level, DUX4-r enrichment at its targets is detectable but relatively low in HEK cells while it is significantly higher in REH cells, which express significantly more GTF2I compared to HEK cells. In REH cells, I found that GTF2I is already associated to DUX4-r targets in its absence. It is tempting to speculate that GTF2I pre-marking in the vicinity of DUX4 binding sites contribute to DUX4-r target gene definition. To test this, in the future I plan to: i. perform DUX4-r CUT&Tag in REH cells knockdown for GTF2I or treated with SAHA; compare GTF2I CUT&Tag signal in REH and HEK cells; perform DUX4-r CUT&Tag in HEK cells overexpressing GTF2I. I expect to find that GTF2I modulation affects DUX4-r recruitment to its genomic targets.

Interestingly, I found that also GTF2I is affected by DUX4-r. Indeed, the relatively low GTF2I enrichment at DUX4-r targets is significantly enhanced upon DUX4-r expression. While the precise mechanism at the basis of the synergy between DUX4-r and GTF2I remains to be determined, it is tempting to speculate that DUX4-r binding tethers GTF2I to target loci and/or that the two factors stabilize each other chromatin association.

Several DUX4-r target genes display absent or negligent transcript levels in the absence of DUX4-r. GTF2I acts as transcriptional activator or repressor according to the interaction partner or the genomic context Given that GTF2I shows a basal enrichment at DUX4-r targets in its absence, GTF2I could normally repress those genes. DUX4-r association to nearby sites could directly promote dissociation of GTF2I from its repressive interactors, forcing it to work as a DUX4-r co-activator. A non-mutually exclusive hypothesis is that DUX4-r could induces chromatin conformational changes indirectly favoring GTF2I switch to co-activator. To start addressing this, I plan to perform CUTAC (Henikoff *et al*, 2020) experiments to determine whether DUX4-r and/or GTF2I regulate genome-wide chromatin accessibility of DUX4-r target loci. Most ALL cases present DUX4-r variants carrying C-terminal deletions of variable size and fusions with *IGH*-encoded amino acids of random length and sequence. A minority of ALL cases display DUX4-r that are just deleted at the C-terminus. For my thesis, I decided to focus on the two extremes of the DUX4-r spectra by comparing the DUX4-IGH variant mostly used in the literature to a C-terminal deletion only variant. While the two variants behave qualitatively the same in all molecular and functional analyses performed, the DUX4-IGH variant caused quantitatively stronger effects. The two variants regulate the same genes, with stronger activation by DUX4-IGH respect to DUX4-del50. The direct DUX4-IGH/DUX4-del50 target genes are significantly overlapping the DUX4-r ALL patients selective gene signature. Intriguingly, by re-analyzing DUX4-r ALL patient datasets, I found that the expression level of these targets is correlated to the size of DUX4 C-terminal deletion and the presence and size of the *IGH*-encoded amino acids. To determine the clinical relevance of my findings, in collaboration with the Italian Association of Pediatric Hematology and Oncology ALL working group, we plan to determine if the expression level of DUX4-r targets correlate with the size of C-terminal DUX4-deletion and IGH amino acids appendage. By re-analyzing DUX4-r ALL patient datasets, I found that the minimal DUX4 portion maintained in the patients contains the first 301 amino acids. This provides the starting point to map the DUX4-r region interacting with GTF2I. To this aim, I plan to use co-IP to test the interaction of GTF2I with DUX4 deletion mutants. I will start by comparing the DUX4 dbd ($DUX4_{1\text{-}155}$) to the smallest DUX4 deleted variant in ALL ($DUX4_{1\text{-}301}$), since I believe it is unlikely that the DUX4 dbd is involved in GTF2I interaction.

Pulldown with recombinant proteins will be also used to determine if DUX4-r and GTF2I interaction is direct. Results from these analyses should help elucidating the molecular consequences of DUX4 rearrangements and DUX4-r mechanism of action.

Previous ChIP-seq performed on NALM6 cells and REH cells transduced with patient-derived DUX4-IGH constructs showed that DUX4-IGH binds to 97% of the same genomic regions bound by DUX4 (Yosuke Tanaka *et al*, 2018). Nevertheless, such analyses have been performed by comparing DUX4-IGH datasets from NALM6 or REH cells to DUX4 datasets obtained in hiPSC or muscle cells. Moreover, while in my inducible systems DUX4 and DUX4-r are expressed at levels similar to those of ALL patients, in previous studies DUX4 and DUX4-IGH were ectopically expressed a much higher level. In addition, ChIP-seq uses formaldehyde crosslinking at much higher concentration (10% vs 0.1%) and for longer time (10 min vs 2 min) compared to CUT&Tag. It is thus possible that many previously reported DUX4/DUX4-r genomic targets are false positive.

GTF2I activity and cellular localization is regulated through phosphorylation by several tyrosine kinases, which respond to various extracellular stimuli and convey them in the nucleus though shifts in GTF2I subcellular localization (Shirai *et al*, 2017; Chailangkarn *et al*, 2018). B cells, as in the case of REH, rely on extracellular stimuli (eg. binding of ligands and/or cell-matrix contact) to take cell fate, proliferation and maturation decisions (Glodek *et al*, 2003). Self-renewal and differentiation of B-cell precursors depends on their interaction with bone marrow (BM) stromal cells and the associated extracellular matrix (Dittel *et al*, 1993). It is thus possible that, upon activation of such pathways during early B-cell development, GTF2I subcellular localization shifts from cytoplasmic to nuclear, facilitating its interaction with DUX4-r and the transactivation of leukemogenic genes. Future GTF2I immunofluorescence and subcellular fractionation studies in HEK and REH cells in the presence or absence of DUX4-r could provide information concerning this aspect.

Previous work reported that treatment with the HDAC inhibitor SAHA downregulate GTF2I mRNA and protein levels. Notably, SAHA treatment blocks activation of DUX4-IGH targets and the downstream phenotypic effects as effectively as direct DUX4-r knockdown further strengthening the relevance of the interplay between DUX4-IGH and GTF2I in leukemia. Moreover, DUX4-IGH makes expressing cells

particularly sensitive to SAHA treatment. This result is highly relevant with a view to develop therapeutic strategies tailored to DUX4-r B-ALL. While SAHA is already approved for the treatment of cutaneous T-cell lymphoma and is in clinical trial for several other types of cancer, further studies concerning the molecular mechanisms of GTF2I downregulation by SAHA could identify additional molecules able to specifically regulate its expression levels, which could be used alone or in combination with other therapeutic approaches to target DUX4-r/GTF2I activity in B-ALL.

DUX4-r activates the expression of cell adhesion molecules, integrins, mitogen and mitogen receptor-encoding genes, which converge in signaling pathways involved in cell-cell, cell-matrix adhesion and cell migration. Accordingly, DUX4-r-expressing B-ALL cells display an enhanced adhesive and migratory behavior with the respect to control cells. Among DUX4-r direct targets, which are consistently upregulated in DUX4-r B-ALL patients and could be mediating DUX4-r cell adhesion/migration, are *ITGA6* and *CD34*.

ITGA6 encodes for integrin alpha 6 (also known as CD49f), which is involved cell adhesion and migration, and, through the activation of PI3K/Akt and MEK/Erk signaling, controls stem-like cell phenotypes and invasion in a variety of tumors (Hu *et al*, 2016; Kwon *et al*, 2013; Brooks *et al*, 2016). ITGA6 has been implicated in the migration of ALL cells to the central nervous system (CNS) and in ALL minimal residual disease (Coustan-Smith *et al*, 2011; DiGiuseppe *et al*, 2009). Intriguingly, ITGA6 expression levels are positively correlated to the incidence of CNS disease in ALL xenograft models and in patients with ALL who had CNS relapse. Notably, treatment with ITGA6-neutralizing antibodies or with a PI3K inhibitor, which decreases ITGA6 expression, significantly reduce migration toward the cerebrospinal fluid *in vitro*, CNS disease tumor burden and prolong survival of mice engrafted with DUX4-r-depending NALM6 cells (Yao *et al*, 2018). B-ALL patients uniformly treated according to the same chemotherapy regimen which display $ITGA6^{high}$ blasts at diagnosis are significantly less likely to achieve deep remission compared to $ITGA6^{low}$ patients. Moreover, anti-human ITGA6-blocking antibodies or *ITGA6* knockout induce primary ALL cell apoptosis and sensitizes them to chemotherapy or tyrosine kinase inhibition *in vitro* and *in vivo* (Gang *et al*, 2020). It is thus tempting to speculate that ITGA6 could represent a key oncogenic target of DUX4-r, and testing the effect of targeting ITGA6

alone or in combination with SAHA, PI3K/tyrosine kinase inhibitors or chemotherapy is among our future plans.

CD34 is a transmembrane phosphor-glycoprotein whose function, despite its widespread use as a stem cell marker clinically as well as in research, remains enigmatic. CD34 has been implicated in cell proliferation, differentiation, adhesion, migration, signal transduction, and maintenance of progenitor phenotype (Kapoor *et al*, 2020; Furness & McNagny, 2006; Ohnishi *et al*, 2013).

CD34 expression level is positively correlated to multidrug resistance and is considered a prognostic biomarker in B-ALL (Jiang *et al*, 2016; Zhou *et al*, 2022).

Of the two CD34 isoforms known before my thesis, only forced expression of the full-length isoform, but not the C-terminally truncated one, is associated with block of differentiation, supporting the functional relevance of the highly conserved CD34 cytoplasmic tail (Suda *et al*, 1992; Fackler *et al*, 1995). The only known interactor of the CD34 cytoplasmic tail is CRKL, a member of the Crk family of adaptor proteins (Felschow *et al*, 2001).

Crk proteins contain one Src-homology 2 (SH2) domain (phosphotyrosine-binding) and two Src-homology 3 (SH3) domains (proline-rich-sequence-binding), and link proteins that do not possess intrinsic kinase activity to intracellular signaling cascades, thereby enabling them to transmit signals indirectly. CRKL is a BCR/ABL oncogene target (Uemura *et al*, 1997) and its knockdown enhances sensitivity to chemotherapy in Ph-like ALL (Sasaki *et al*, 2022). By combining RNA-seq and CUT&Tag, I found that DUX4-r directly activates the expression of a novel CD34 isoform (CD34-s) from an internal promoter in intron 4. CD34-s is predicted to lack the mucin-like domain, presenting a partially different globular domain and maintaining the stalk, transmembrane and cytoplasmic tail of full-length CD34. It is tempting to speculate that CD34-s is able to constitutively activate signal transduction pathways stimulating proliferation, blocking differentiation and promoting adhesion of B-cell precursors. In the future, I plan to test my hypothesis by comparing the activity of the three CD34 isoforms and by selective CD34-s knockdown in leukemia settings.

DUX4-r expressing B-ALL cells are characterized by an early block of differentiation at the Pro-B cell stage and downregulation genes involved with B-cell identity and regulation of the pre-BCR activity. An interplay between ITGA6 and

CD34-s at the cell periphery may hijack the pre-BCR checkpoint in B-ALL cells, which fail to undergo apoptosis and proliferate as immature B-cell precursors giving rise to leukemia development.

All in all, results from my PhD project establish a novel model of leukemia initiation by DUX4-r and lay the basis for the development of new therapeutic approaches aiming to specifically target the activity of the fusion transcription factor in B-ALL.

5. MATERIALS AND METHODS

Plasmids and Cloning

Lentiviral vectors encoding inducible, N-terminally tagged DUX4, DUX4-IGH and DUX4-del50 was generated through the use of the Gateway Technology (Invitrogen) starting from the pCW57.1 (Addgene #41393) destination vector. The N-terminal Twin-Strep-2xHA (Strep-HA) double affinity tag was recovered by PCR (table 1) from pTO Strep-HA plasmid and inserted in the EcoRI site of pCW57.1 generating the vector, which was subsequently used as EV control. To insert DUX4 variants into the EV, *DUX4* and *DUX4-IGH* open reading frames (ORFs) were recovered by PCR (table 1) from vectors XX and YY, respectively, using the same 5'-primer but different 3' ones. Primers were designed to carry attB1/2 recombination sites. DUX4-del50 was generated through site-directed mutagenesis starting from the wt *DUX4* entry clone, by introducing a premature stop codon 150bp upstream of *DUX4* wild type stop codon.

Constitutive and inducible shRNA vectors targeting *DUX4-IGH* were generated by the GeneScript gene-synthesis service by introducing the *shDUX4-IGH* sequence described in (Carlet *et al*, 2021) into pLKO.1-hPGK-Puro (Addgene #8453) and pLKO.1-Tet-On-Puro (Addgene #21915) lentiviral vectors.

The lentiviral expression vector encoding the GFP reporter controlled by multiple DUX4 binding sites (pLentiDUX4-BS-GFP-TKneo) was obtained from the Miller's group (doi: 10.1093/hmg/ddv315).

The GTF2i mammalian expression vector pEBB-GFP-GT2FI-gamma was obtained from Addgene (Cat #22148).

The lentiviral vector encoding for the constitutively expressed GFP reporter was obtained from Addgene pLenti-CMV-GFP-Neo (Addgene # 17447).

Primer name	**Sequence**
EcoRI-Strep-HA cloning Fw	GCGGCTAGCATGTACCCATACGATGTTCCCGACTACGCCGGTACC GAGCTCGGATCCACCATGGCAAGCTGGAGCCACCCGCAGTTCGAG
EcoRI-Strep-HA cloning Rv	CGCGCTAGCAGTTTTTCGAACTGCGGGTGGCTCCACGATCCACCT CCCGATCCACCTCCGGAACCTCCACCTTTCTCGAACTGCGGGTGG
attB DUX4 Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGGCGGAGGCGGAG GCGCCCTCCCGACACCCTCGGACAGC
attB DUX4-Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAAGTTCTTCCAGA AGGGCTCT
attB DUX4-IGH Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAGGCACCC-CAGTGCCCGTC

DUX4-del50 Mutagenesis Fw	GCGGCCTGCTGCTG TAG GAGCTCCTGGCG
DUX4-del50 Mutagenesis Rv	CGCCAGGAGCTC CTA CAGCAGCAGGCCGC

Table 1. Sets of primers used for cloning.

Lentiviral preparation

Preparation of lentiviral particles was performed by calcium phosphate co-transfection of HEK293T cells (ATCC-crl321) with the lentiviral vector of interest in combination with a pCMV-VSV-G lentiviral envelop (Addgene#8454) and the pCMV-dR8.2-dvpr lentiviral packaging plasmids (Addgene#8455). For each transfection, 9x10^6 HEK293T cells were plated 18h prior to transfection. Culture medium was replaced with fresh IMDM medium(Lonza-LOBE12726F), supplemented with 10% fetal bovine serum (FBS), 5mM L-Glutamine and 1% Penicillin/Streptomycin antibiotic cocktail (P/S), 2h before transfection. Respectively, 32ug, 9ug and 12,5ug of lentiviral vector, VSV-G and packaging vectors were diluted in 0.06X TE and brought to a final volume of 1,125mL. 125uL of ice-cold CaCl2 was added and the mix was allowed to incubate at RT for 10', following drop-wise addition of 1,250mL of 2X HBSS while vortexing the mix at full-speed. 2,5mL of the transfection mix was gently added dropwise on HEK293T cells. After 16h, the medium was replaced with fresh IMDM medium and cells were allowed in the incubator for 30h before collection of the viral supernatant. Medium was then replaced and, after 24h of further incubation, a second round of virus collection was performed. Viral collections were then pooled and ultracentrifuged at 20000xg for 2h at 4°C. Viral particles were then resuspended in 75uL of Optimem (ThermoFisher-31985-047) for each tube and stored at -80°C.

Cell culture, transfection and transduction

REH (DSMZ-ACC22) and NALM6 (DSMZ-ACC128) were maintained in RPMI-1640 (Euroclone-ECM2001L) medium supplemented with 20% and 10% FBS, respectively and 1% P/S. HEK293 (ATCC-crl1573), HEK293T and HS-5 (ATCC-crl11882) cells were maintained in DMEM medium supplemented with 10% FBS and 1% P/S.

Transfection of DUX4-GFP reporter in inducible HEK cells has been performed using the Lipofectamine LTX reagents (ThermoFisher Scientific), according to manufacturer's instructions. 1 ug of control or DUX4-GFP reporter plasmids were transfected, and cell were visualized at the microscope 8h after induction of transgenes.

DUX4, DUX4-IGH and GTF2I-encoding plasmids were transfected into HEK cells, using 1ug of control or gene-of-interest (DUX4 or DUX4-IGH) vector and 200ng of GTF2I vector + 800ng of control vector to balance the transfection reaction. RNA and proteins were collected 24h post transfection.

Transduction of suspension cells was performed by spinoculation of 1x10^6 cells in 1mL of RPMI medium containing 8ug/mL Polybrene (SigmaAldrich-TR-1003-G), to increase transduction efficiency. Spinoculation performed at 1290xg, for 1.30h at 34°C and cells were allowed in the cell culture incubator for 48h before changing the medium.

REH cells expressing inducible, Strep-HA-tagged DUX4, DUX4-IGH or DUX4-del50 were generated by transduction of REH cells with the lentiviral vectors as described above, followed by selection with 0.25ug/uL puromycin (ThermoFisher-A1113803) for 2 weeks. Then, cells were maintained at 0.1ug/uL of puromycin.

Inducible-*shDUX4-IGH* NALM6 expressing cells were generated by transduction with the inducible vector described above. Cells were selected with 1ug/mL puromycin for 2 weeks and then maintained at 0,2ug/mL of the antibiotic.

GFP-inducible REH cells and *shCTRL/DUX4-IGH* NALM6 cells have been generated by transduction with the pCMV-GFP-Neo vector and selected with 500ug/mL of G418-Sodium Salt (SigmaAldrich-A1720-5G) dissolved in water for 1 month. Then, cells were maintained at 100ug/mL G418 and 0,2ug/mL puromycin.

Induction of transgene expression was performed by adding 1ug/mL Doxycycline (SigmaAldrich-D9891-10G).

RNA extraction and RT-qPCR analyses

For >2x10^6 cells, RNA was extracted using the PureLink RNA Mini Kit (LifeTechnologies-12183025), following the manufacturer's instructions. For

processing lower amount of cells, RNA was extracted using the NucleoSpin RNA XS kit (Macherey-Nagel-FC140955N).

Reverse transcription was performed with SuperScript™ III First-Strand Synthesis SuperMix (LifeTechnologies-11752-250), following the manufacturer's instructions. As input, 1ug of DNA-free RNA was used, or the same amount of RNA per sample in case of lower concentrations.

Quantitative real-time PCR was performed with the iTaq Universal SYBR Green Supermix (Bio-Rad-1725122), using 25ng or 15ng of cDNA and 0,4uM or 0,2uM concentration of primers in the 96 or 384 well plates, respectively. The real-time PCR program used was: 10' @ 95°C, 30'' @ 95°C, 30'' @ 58°C, 30'' @ 72°C, repeated 39 times, 5'' @ 65°C and 30'' @ 65°C to terminate the reaction.

Target gene name	**Sequence**
GAPDH Fw	TCAAGAAGGTGGTGAAGCAGG
GAPDH Rv	ACCAGGAAATGAGCTTGACAAA
DUX4/DUX4-IGH Fw	CGATGTTCCCGACTACGC
DUX4/DUX4-IGH Rv	GCCGGAGCCTGCTTT
DUX4-IGH Fw	ATCCCAGACCGCCCTGC
DUX4-IGH Rv	CTGGCCCTTCGATTCTG
ERGalt Fw	GAGTTGGAAAGCTCTGGGAGAATCTG
ERGalt Rv	GGTGAATGCACGCTGATGGGAAAAG
ERG Fw	CTTGATCGCATTATGGCCAGCACT
ERG Rv	TTGGATCTCTTCCCCGGCTTCCTTC
RFPL2 Fw	CCCACATCAAGGAACTGGAG
RFPL2 Rv	TGTTGGCATCCAAGGTCATA
TRIM43 Fw	ACCCATCACTGGACTGGTGT
TRIM43 Rv	CACATCCTCAAAGAGCCTGA
MBD3L2 Fw	GCGTTCACCTCTTTTCCAAG
MBD3L2 Rv	GCCATGTGGATTTCTCGTTT
ITGA6 Fw	AGCTGTGCTTGCTCTACCTG
ITGA6 Rv	GAGCAACAGCCGCTTGTCC
DDIT4L Fw	AGGAAACAGAGCCGTTGACC
DDIT4L Rv	GTCAAAATCACTTAGCAGGCTCT
ABCG1 Fw	GAAAACACTGGTCAAGAACCCA
ABCG1 Rv	TTCTGGTGTTCAGGTGAATGACT
PTPRM Fw	CTCAGCACCATGAGGGGACT
PTPRM Rv	AAAGAGGCAGCCACCTGAGAAC
NR3C2 Fw	ATACTGCTGGCGATGGTAGC

NR3C2 Rv	TTCGCTGCTTTCATCCACCT
CD34 All isoforms Fw	GCGGAGTTTAAGAAGGACA
CD34 All isoforms Fw	TCAGGTCAGATTGGTGCT
VPREB3 Fw	CTTCCTGTCAGGCCAAGT
VPREB3 Rv	TTGGCTGCCGAGAATC
BCL6 Fw	AGGCATTGGTGAAGACAA
BCL6 Rv	GCCATGAGGACCGTTTT
CD79b Fw	AATGCCAACCTCAGCAC
CD79b Rv	CTGTGAGTTGATGAGCTTCTT
CD34 Isoform 1 Fw	CCAGGTCCTTGTTTGCT
CD34 Isoform 1 Rv	CCCAGCTAAGACAGAGTCAC
CD34 Isoform 2 Fw	CTGGGGATCCTAGATTTCA
CD34 Isoform 2 Rv	AGCTGCATGTGCAGACTC
CD34-s Fw	GCAATGCTTCAACGAGAC
CD34-s Rv	TGTCCTTCTTAAACTCCGC
HSATII Fw	TTGGTGATTCCACTGGATTTCT
HSATII Rv	TCGGATGGAATCAATGAAGGGA
THE1C Fw	TACCCAAAAATGTGGAAGCGA
THE1C Rv	AGCTAGAGTGACTGGGATGC
hERV-L Fw	ATATCCTGCCTGGATGGGGT
hERV-L Rv	GAGCTTCTTAGTCCTCCTGTGT
GTF2I Fw	AATCTACAACCCAGGCAAA
GTF2I Rv	GAAGATGCTGCGAGACC

Table 2. Sets of primers for qRT-PCR experiments.

Protein extraction and immunoblotting

Total cell proteins were extracted in RIPA buffer (25mM Tri-HCl pH8, 150mM NaCl, 10mM NaF, 20mM Na3VO4, 0,1% SDS, 1% NP-40) by incubation for 10’ in ice, followed by centrifugation of the extract at 16000xg, for 10’ at 4°C, and supernatant was collected. Proteins were quantified with the Protein Assay Dye Reagent Concentrate (Bio-Rad-#5000006) using 1ug/uL gamma-globin as standard. Absorbance was read at 595nm using a spectrophotometer. For immunoblotting analyses, 20ug of proteins were diluted in 4X Laemmli buffer (Bio-Rad-#1610747) to a concentration of 1X and boiled for 5’ at 95°C. Standard SDS-PAGE was performed at 10/12% polyacrylamide concentration, followed by wet-transferred onto a 0.45um nitrocellulose membrane using an 80% Tris-HCl, 10% Methanol solution for 1.30h. Membranes were then stained with ponceau solution and then incubated with a 5% Milk blocking solution

for >30'. Primary antibodies (table 3) were incubated overnight, oscillating at 4°C and secondary, HRP-conjugated antibodies were incubated for 1hr at room temperature. The SuperSignal West Pico PLUS Chemiluminescent Substrate (LifeTechnologies-34580) was used for the visualization of proteins.

Antibody (ref. Number)	**Producer**	**Lot**
α-DUX4 e55 (ab124699)	Abcam	GR3191135
α-hDUX4-C (MAB95351)	Bio-Techne	CLPV0119111
α-GAPDH (G9545)	Sigma Aldrich	099M4801V
α-CD34 (SAB4300690)	Sigma Aldrich	371521469
α-GTF2I	Bethyl	#1

Table 3. Sets of antibodies for Western Blot analysis.

Cell-stroma adhesion assay

HS-5 cells were maintained in fully supplemented RPMI medium for at least 72h prior to starting the experiment. Next, 300000 HS-5 cells were plated into a 12 well plate 6-8h before co-culturing with 50000 REH-GFP or NALM6-GFP cells for 18h. Cells were then washed twice with PBS. Susequently, adhering cells were recovered with trypsin, washed, resuspended in PBS and analyzed by FACS for the percentage GFP-expressing cells.

Transwell migration assay

HS-5 cells were maintained in fully supplemented RPMI medium for at least 72h prior to starting the experiment. 120000 HS-5 cells were plated into the bottom compartment of a 24 Transwell plate 24h before the start of the co-culture. Once HS-5 were fully attached and well-elongated, 25000 REH (EV or DUX4-IGH) or NALM6 cells were added to the bottom well to obtain HS-5-only, HS-5/REH, HS-5/NALM6 combinations, which were brought to 600uL with fully supplemented RPMI and treated with 1ug/mL Doxycyline, before insertion of the 5um polycarbonate membrane cell culture inserts followed by incubation for 4-6h prior starting the experiment. Next, 10000 *shCTRL* or *shDUX4-IGH* NALM6-GFP cells were seeded in the upper compartment and treated with 1ug/mL of Doxycycline. After 16-18h incubation, cells

in the bottom compartment were trypsinized and the amount of GFP+ (GFP-positive) cells that had migrated from the upper compartment was determined by flow cytometry.

Cell proliferation in low serum

To avoid extensive cell death caused by severe serum starvation, different concentrations of FBS were tested, and 5% FBS gave the best result. 200000 REH or NALM6 cells were seeded in 2mL of RPMI medium supplemented with 5% of FBS in a 6 well plate to get a final concentration of cells of 100000/mL. Transgenes were induced with 1ug/mL of Doxycycline and cell proliferation was estimated by automatic cell counting (Bio-Rad, TC20-1450102) every 48h, accompanied by replacement of the medium with fresh Doxycycline, up to 7 days.

Live-cell spheroid assay

50000 inducible REH cells were seeded in 1mL (50000/mL) of RPMI medium containing 5% of FBS and 1 ug/mL of Doxycyline into a 24 well plate. Cells were then incubated for 72h into the Incucyte Live-Cell Analysis Systems (Satorius) for real-time analysis of cellular aggregation into spheroids. Clustering analysis was performed according to the Incucyte Handbook, Chapter 5a, “Kinetic Assays for Immune Cell Activation and Proliferation”. Label-free, phase-contrast imaging was used to detect morphological differences between cells expressing different DUX4 variants. Quantification of the average spheroid area was calculated using the following parameters: Analysis of Phase Area Object Average normalized to T0, cell/background ratio: -1, fill holes: -5.

RNA-sequencing

Total RNA was extracted as described above from ~3x10^6 inducible REH cells per sample, after 12h of doxycycline induction. RNA concentration, purity and integrity were determined by capillary gel electrophoresis through the BioAnalyzer2100 system. All samples used for RNA-seq library preparation showed an RNA integrity number (RIN) greater than 9. RNA-seq libraries were generated from biological quadruplicates

using the TruSeq Stranded mRNA library prep kit, using 100ng of total RNA as input. Libraries were run with a NovaSeq instrument and sequenced in paired end (PE) generating 2x50bp reads, for a total number of 50M (million) reads per sample.

Transcriptomic analyses

Sequencing adapters were removed using trimmomatic (v0.39) (Bolger *et al*, 2014) and fastq files were then aligned to the human genome assembly GRCh38 (hg38) using the STAR aligner (v2.5.3a) (Dobin *et al*, 2013). Annotation of genomic features was performed using the featureCounts tool (v1.6.4) (Liao *et al*, 2014), using the GENCODE v31 Gene transfer format (GTF). Differential gene expression analysis was carried out with DeSeq2 package (Bioconductor) (Costa-Silva *et al*, 2017). As previously described, several genes activated by DUX4 are restricted to the cleavage stage embryo and, for this reason, are silent or expressed at very low levels in REH cells. As a consequence, even a very small increase of expression of these genes will score as significant, even though the absolute abundance of some of these transcripts is very low and likely does not functionally impact the cells, potentially masking those genes which are more likely to contribute to the biology of DUX4 or DUX4-r. In order to dissect such genes which are selectively activated by DUX4 or DUX4-r, after filtering for a Log2FC>=1 and adjP-value<=0,01 over the EV control, filters on the magnitude of Log2FC in the different conditions were applied as follows:

- DUX4 selective target genes were obtained by filtering for a Log2FC>=2 and adjP-value<=0,05 over either DUX4-IGH or DUX4-del50; DUX4 and DUX4-IGH commonly upregulated genes by Log2FC=<2 and an adjP-value>=0,05 over DUX4 and DUX4-IGH; DUX4 and DUX4-del50 commonly upregulated genes by a Log2FC=<2 and an adjP-value>=0,05 over DUX4 and DUX4-del50; genes commonly regulated by the three DUX4 versions by Log2FC=<2 and an adjP-value>=0,05 over both DUX4-IGH and DUX4-del50.
- DUX4-IGH selective target genes were obtained by filtering for a Log2FC>=2 and adjP-value<=0,05 as compared to either DUX4 or DUX4-del50; DUX4-IGH and DUX4 commonly upregulated genes by a Log2FC=<2 and an adjP-value>=0,05 by comparing DUX4-IGH and DUX4; DUX4-IGH and DUX4-

del50 commonly upregulated genes by a Log2FC=<2 and an adjP-value>=0,05 over DUX4-IGH and DUX4-del50; genes commonly regulated by the three by Log2FC=<2 and an adjP-value>=0,05 over both DUX4 and DUX4-del50.

- DUX4-del50 were obtained by filtering for a Log2FC>=2 and adjP-value<=0,05 over either DUX4 or DUX4-IGH; DUX4-del50 and DUX4 commonly upregulated genes by a Log2FC=<2 and an adjP-value>=0,05 over DUX4-del50 and DUX4, DUX4-del50 and DUX4-IGH commonly upregulated genes by a Log2FC=<2 and an adjP-value>=0,05 over DUX4-del50 and DUX4-IGH; genes commonly regulated by the three by Log2FC=<2 and an adjP-value>=0,05 over both DUX4 and DUX4-IGH.

To automatize the above analysis, I wrote the following MatLab script:

```
v==(col1)
EV=c==(col2)
DUX4=a==(col3)
DUX4-IGH=b==(col4)
DUX4-DEL50=d==(col5)

for v=(a/b, a/d) &&
for a/c>=1
if DUX4/DUX4-IGH>2 && DUX4/DUX4-DEL50>2 then v=A elseif DUX4/DUX4-IGH >2 &&
DUX4/DUX4-DEL50 <2 then v=AD elseif DUX4/DUX4-IGH<2 && DUX4/DUX4-DEL50>2 then v=AB
elseif DUX4/DUX4-IGH<2 && DUX4/DUX4-DEL50<2 then v=ABD
End

for v=(b/a,b/d) &&
for b/c>=1
DUX4-IGH/DUX4>2 && DUX4-IGH/DUX4-DEL50>2 then v=B elseif DUX4-IGH/DUX4 >2 &&
DUX4-IGH/DUX4-DEL50 <2 then v=BD elseif DUX4-IGH/DUX4<2 && DUX4-IGH/DUX4-DEL50>2
then v=AB elseif DUX4-IGH/DUX4<2 && DUX4-IGH/DUX4-DEL50<2 then v=ABD
end

for v=(d/a,d/b)
for d/c>=1
DUX4-DEL50/DUX4 >2 && DUX4-DEL50/DUX4-IGH>2 then v=D elseif DUX4-DEL50/DUX4 >2 &&
DUX4-DEL50/DUX4-IGH <2 then v=BD elseif DUX4/DUX4-IGH<2 && DUX4/DUX4-DEL50>2 then
v=AD elseif DUX4-DEL50/DUX4<2 && DUX4-DEL50/DUX4-IGH<2 then v=ABD
    end
```

Gene-Set enrichment analysis was performed using the GSeABase software package (1.57.0) package (Bioconductor) (Morgan M, Falcon S, 2021), using the Canonical Pathways (CP), Gene Ontology (GO), ChIP-X Enrichment Analysis 3 (ChEA3) from the Molecular Signatures Database MSigDB (v7.5.1) (Subramanian *et al*, 2005) and setting a False Discovery Rate value to 0,05.

CUT&Tag

Starting from >5x10^6 REH cells induced with doxycycline for 12h, nuclei were prepared by resuspending cells in NE1 buffer (20mM HEPES-KOH pH7.9, 10mM KCl, 0,5mM Spermidine, 1% Triton X-100) at a cellular density of 10^6 cells/mL, followed by 10' incubation in ice. Samples were spun for 4' at 600xg at 4°C and the nuclear pellet was resuspended in 1 volume (V) of PBS. A light cross-linking was performed by addition formaldehyde to 0,1% final concentration for 2', followed by quenching with a double molar concentration of glycine. Nuclei were spun and resuspended in ½ V of Wash Buffer 150 (20mM HEPES-KOH pH7.9, 150mM NaCl, 0,5mM Spermidine). Nuclei were then manually counted using the trypan blue exclusion assay, as described (Kaya-Okur *et al*, 2019). Next, for each CUT&Tag assay, 100000 nuclei were immobilized to 4uL of Concanavalin A (ConA-Epycypher-21-1401)-conjugated beads for 10' at RT. The primary (1^{st}) antibody (Ab) was diluted 1:100 in antibody buffer (20mM HEPES-KOH pH7.9, 150mM NaCl, 0,5mM Spermidine, 0,5mg/mL BSA, 0,4mM EDTA pH7.5) and incubated with the ConA-bound nuclei ON at 4°C while rotating. The following day, ConA-nuclei-1^{st}Ab were spun 100xg, 1'', the supernatant was removed, nuclei were resuspended in 100ul of Wash Buffer 150 and incubated with a secondary antibody (2^{nd}) diluted 1:100 in Wash Buffer 150 for 1h at RT. After removal of the 2^{nd} Ab, nuclei were washed once with X ul of Wash Buffer 150 and incubated with the protein AG-conjugated-Transposase (pAG-Tn5, Epycypher- 23615-1117) preloaded with sequencing adapters, diluted 1:20 in Wash Buffer 300 (20mM HEPES-KOH pH7.9, 300mM NaCl, 0,5mM Spermidine) for 1h at RT. Nuclei were then washed twice in Wash Buffer 300 and tagmentation reaction was performed by addition of 10mM MgCl2 following incubation at 37°C for 1h. After a wash with TAPS buffer (10mM TAPS in H_2O), tagmented DNA was released by addition of 5uL 0.1% SDS release buffer (0,1% SDS, 10mM TAPS in H_2O) following incubation at 58°C for 1h. Then, 15uL of 0.67% Triton-X-100 was added to allow PCR for library preparation. For the library prep, 2uL of 10uM i5 universal adapter primer (Nextera) and 2uL of 10uM i7 uniquely barcoded adapter primers were added to released, tagmented DNA, followed by addition of 25uL of the Q5 High-Fidelity 2X Master Mix (New England Biolabs, M0492L). PCR was performed as follows: Cycle 1: 72 °C for 5' (gap filling), Cycle 2: 98 °C for 30'', Cycle 3: 98 °C for 10'' Cycle 4: 63 °C for 10''. Repeat Cycles

3-4 13 times. 72°C for 1’ and hold at 8°C. Amplified DNA was then recovered through size exclusion purification by addition of 1.3V of AMPureXP magnetic beads (Beckman-A63881), followed by two washes in 80% ethanol and elution in 22uL of 10mM Tris-HCl (pH8). Quantification and average fragment size of libraries was performed via the 4200 TapeStation system (Agilent), using the D5000 high sensitivity kit (Agilent 5067-5592/5067-5593). 450bp of average library fragment size was selected for library quantification. A 5nM pool of the libraries was then run with the NovaSeq instrument and sequenced in paired end (PE), generating 2x75bp reads, for a total number of 10M (million) reads per samples.

Universal Primer (5')	i5-1	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATGTGTAT
	i5-2	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTAT
Uniquely barcoded primer (3')	i7-1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGTG
	i7-2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGTG
	i7-3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGTG
	i7-4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGTG
	i7-5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTG
	i7-6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTG
	i7-7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTG
	i7-8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGTG
	i7-9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGTG
	i7-10	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGTG
	i7-11	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGTG
	i7-12	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGTG
	i7-13	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGTG
	i7-14	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGTG
	i7-15	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGTG
	i7-16	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGTG
	i7-17	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGTG
	i7-18	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGTG
	i7-19	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGTG
	i7-20	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGTG
	i7-21	CAAGCAGAAGACGGCATACGAGATTGTCCGCGGTCTCGTGGGCTCGGAGATGTG
	i7-22	CAAGCAGAAGACGGCATACGAGATATATGGAAGTCTCGTGGGCTCGGAGATGTG
	i7-23	CAAGCAGAAGACGGCATACGAGATAACGAATTGTCTCGTGGGCTCGGAGATGTG
	i7-24	CAAGCAGAAGACGGCATACGAGATTCGACGCCGTCTCGTGGGCTCGGAGATGTG

Table 4. Sets of primers used for CUT&Tag.

CUT&Tag analyses

Sequencing adapters were removed via the Cutadapt tool.and then trimmed fastq files were aligned with Bowtie2 (v2.2.1)(Langmead & Salzberg, 2012) to the human

genome assembly GRCh38 (hg38) from Genome Reference Consortium, using the following parameters: --local --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700. File format conversion was required for peak calling, and it was performed using the bedtools utilities. The sparse enrichment analysis for CUT & RRU (SEACR) package is designed to call peaks and enriched regions from chromatin profiling data with very low backgrounds, which is typical for CUT&Tag data. In case of absent IgG control data, peak calling with SEACR was performed using a 0.001 FDR, -non parameter for the automated normalization of fragment counts with the E. coli read count and the -stringent parameter for thresholding the signal using the peak of the signal curve. Annotation of genomic features enriched within CUT&Tag peaks was performed using the AnnotatePeaks.pl tool from the HOMER (Hypergeometric Optimization of Motif EnRichment) package and the FindMotifsGenome.pl command was used to search for overrepresented DNA motifs within CUT&Tag peak sets, by setting the region size to 75bp and motif length to 15bp. Overlapping and differential peaks from DUX4, DUX4-IGH and DUX4-del50 peak sets were evaluated using the getDifferentialPeaksReplicates.pl tool from the HOMER package, using the default settings.

Tandem affinity purification

Tandem affinity purification from REH cells was performed 12h after doxycycline induction. 2.5x10^8 cells were resuspended in 20mL of Buffer N (10mM HEPES-KOH pH7.9, 300mM sucrose, 10mM KCl, 0,1mM EDTA, 0,1mM EGTA, 0,1mM DTT, 0,75mM spermidine, 0,15mM spermine, 0,1% NP-40, 50mM NaF) and incubated in ice for 10’ to isolate intact nuclei. After one wash in the same volume of Buffer N, nuclei were resuspended in the same volume of high-salt nuclear extraction buffer C420 (20mM HEPES-KOH pH7.9, 420mM NaCl, 25% Glycerol, 1mM EGTA, 0,1 mM DTT 50mM NaF), and vigorously shaken in a thermomixer at 1400rpm for 30’ at 4°C in order to extract nuclear proteins. Samples where then centrifuged at 16000xg for 10’ at 4°C, and the supernatant, which contains nuclear proteins, was collected. Nuclear extracts were then adjusted to 150mM salt by addition of HEPES buffer (20mM HEPES-KOH pH7.9, 50mM NaF), and incubated at 4°C while rotating

with 1ug/mL of avidin (IBA-2-0204-015), in order to chelated biotinylated proteins, which would compete with Strep-HA tagged proteins for binding to StrepTactin beads (IBA- 2-1201-010), 20U/mL of Benzonase nuclease (Sigma Aldrich- E8263-5KU) and 50ng/mL of RNase A (ThermoFisher- EN0531) to remove any remaining nucleic acids. After clearing at 16000xg for 10' at 4°C, supernatants were pre-cleared with 2,5mL (bead volume) of cross-linked Sepharose (SigmaAldrich) for 1h while rotating at 4°C. Extracts where then quantified with the Protein Assay Dye Reagent Concentrate (Bio-Rad-#5000006) using 1ug/uL gamma-globin as standard. Absorbance was read at 595nm using a spectrophotometer. For each purification, 80mg of pre-cleared nuclear extracts were incubated with 2mL (bead volume) of StrepTactin sepharose beads (IBA) for 4h at 4°C while rotating. Then, three washes in TNN-HS (10mM HEPES-KOH pH7.9, 150mM NaCl, 0,1% NP-40, 50mM NaF) buffer were performed, and beads-bound proteins were specifically eluted with three subsequent incubations with 2V of 2.5mM biotin-containing TNN-HS buffer. Next, eluted proteins were incubated with 2mL (bead volume) of anti-HA agarose beads (SigmaAldrich-A2095) while rotating at 4°C overnight. The following day, two washes in TNN-HS were performed, followed by two additional washes in TNN buffer (10mM HEPES-KOH pH7.9, 150mM NaCl, 50mM NaF) in order to get the rid of any leftover NP-40 detergent in the buffer. Finally, HA-beads-bound proteins were eluted with an SDS-containing elution buffer (50mM HEPES-KOH pH7.9, 150mM NaCl, 1% SDS) and precipitated with Acetone. For this, 4V of ice-cold 100% acetone were added to 1V of eluted proteins, followed by vortexing and incubation at -20°C for 1h. Samples were then centrifuged at 15000xg for 10'. The whole procedure was repeated twice to get the rid of all the remaining SDS. One aliquot of each protein sample was kept for quantification via silver staining.

Total protein extracts for mass spectrometry

Total protein extracts from HEK293, REH and JURKAT cells were performed starting from 2x10^7 cells per sample as previously described and precipitated in acetone as above.

Mass spectrometry analyses

Precipitated proteins were dissolved in 8M Urea, 50mM TrisHCl (pH8), 5mM DTT buffer at 37°C for 1h, to help solubilization of the protein precipitate. Then, iodoacetamide was added to dissolved proteins to a final concentration of 15mM, and incubated 30' at RT in the dark. The sample was then diluted with 50mM Tris-HCl (pH8) to bring Urea to a 2M final concentration and Trypsin Gold, Mass Spectrometry Grade (Promega-V5280) was added to a final protease:protein ratio of 1:50 and the sample was incubated overnight at 37°C for protease digestion. Samples were then run int the Agilent 6520 q-tof mass spectrometer and proteomics datasets were analyzed using the MaxQuant doi.org/10.1002/pmic.201400449 proteomics software package to generate MS spectra. Spectral counts of proteins in the DUX4, DUX4-IGH and DUX4-del50 groups were confronted with those of the EV control sample to exclude non-specific interactors. The list of specific interactors was then analyzed for their expression levels in HEK293, REH and JURKAT cells to determine interactors preferentially expressed in REH cells. Differential protein-protein interaction within our datasets of B-cell enriched interactors were then determined with the SAINT package, using the following parameters: Empirical Fold Change Score (FC)=2.5, Significant analysis of interactome (SAINT, SP) = default (0,95).

SAHA treatments

SAHA (SelleckChem-S1047-200MG) was dissolved in DMSO to a final concentration of 100mM and working aliquots were kept at -80° protected from light. Cells were treated with a 1uM dose from 24h and up to 3d for either gene expression analysis or functional studies. In the case of proliferation assays, SAHA was replaced every 3 days upon medium change. An high 10uM dose was used for 48h to perform apoptosis/caspase assays.

REFERENCES

A. TS, R. RD, Sven P, M. DS, Rohit M, Xiao-Wei S, Sooryanarayana V, Xuhong C, Joelle T, Rainer K, *et al* (2005) Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer. *Science (80-)* 310: 644–648

AbuSamra DB, Aleisa FA, Al-Amoodi AS, Jalal Ahmed HM, Chin CJ, Abuelela AF, Bergam P, Sougrat R & Merzaban JS (2017) Not just a marker: CD34 on human hematopoietic stem/progenitor cells dominates vascular selectin binding along with CD44. *Blood Adv* 1: 2799–2816

Andersson AK, Ma J, Wang J, Chen X, Gedman AL, Dang J, Nakitandwe J, Holmfeldt L, Parker M, Easton J, *et al* (2015) The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet* 47: 330–337

Arahata K, Ishihara T, Fukunaga H, Orimo S, Lee JH, Goto K & Nonaka I (1995) Inflammatory response in facioscapulohumeral muscular dystrophy (FSHD): Immunocytochemical and genetic analyses. *Muscle Nerve* 18: S56–S66

Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR & Korsmeyer SJ (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30: 41–47

Banerji CRS & Zammit PS (2021) Pathomechanisms and biomarkers in facioscapulohumeral muscular dystrophy: roles of DUX4 and PAX7. *EMBO Mol Med* 13: e13695

Barber KE, Harrison CJ, Broadfield ZJ, Stewart ARM, Wright SL, Martineau M, Strefford JC & Moorman A V (2007) Molecular cytogenetic characterization of TCF3 (E2A)/19p13.3 rearrangements in B-cell precursor acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 46: 478–486

Berger AM, Mooney K, Alvarez-Perez A, Breitbart WS, Carpenter KM, Cella D, Cleeland C, Dotan E, Eisenberger MA, Escalante CP, *et al* (2015) Cancer-Related Fatigue, Version 2.2015. *J Natl Compr Canc Netw* 13: 1012–1039

Bielorai B, Fisher T, Waldman D, Lerenthal Y, Nissenkorn A, Tohami T, Marek D,

Amariglio N & Toren A (2013) Acute lymphoblastic leukemia in early childhood as the presenting sign of ataxia-telangiectasia variant. *Pediatr Hematol Oncol* 30: 574–582

Boer JM, Valsecchi MG, Hormann FM, Antić Ž, Zaliova M, Schwab C, Cazzaniga G, Arfeuille C, Cavé H, Attarbaschi A, *et al* (2021) Favorable outcome of NUTM1-rearranged infant and pediatric B cell precursor acute lymphoblastic leukemia in a collaborative international study. *Leukemia* 35: 2978–2982

Den Boer ML, van Slegtenhorst M, De Menezes RX, Cheok MH, Buijs-Gladdines JGCAM, Peters STCJM, Van Zutven LJCM, Beverloo HB, Van der Spek PJ, Escherich G, *et al* (2009) A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol* 10: 125–134

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120

Bosnakovski D, Lamb S, Simsek T, Xu Z, Belayew A, Perlingeiro R & Kyba M (2008a) DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp Neurol* 214: 87–96

Bosnakovski D, Toso EA, Hartweck LM, Magli A, Lee HA, Thompson ER, Dandapat A, Perlingeiro RCR & Kyba M (2017) The DUX4 homeodomains mediate inhibition of myogenesis and are functionally exchangeable with the Pax7 homeodomain. *J Cell Sci* 130: 3685–3697

Bosnakovski D, Xu Z, Gang EJ, Galindo CL, Liu M, Simsek T, Garner HR, Agha-Mohammadi S, Tassin A, Coppée F, *et al* (2008b) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* 27: 2766–2779

Brooks DLP, Schwab LP, Krutilina R, Parke DN, Sethuraman A, Hoogewijs D, Schörg A, Gotwald L, Fan M, Wenger RH, *et al* (2016) ITGA6 is directly regulated by hypoxia-inducible factors and enriches for cancer stem cell activity and invasion in metastatic breast cancer models. *Mol Cancer* 15: 26

Buitenkamp TD, Izraeli S, Zimmermann M, Forestier E, Heerema NA, van den Heuvel-Eibrink MM, Pieters R, Korbijn CM, Silverman LB, Schmiegelow K, *et al* (2014) Acute lymphoblastic leukemia in children with Down syndrome: a retrospective analysis from the Ponte di Legno study group. *Blood* 123: 70–77

Bunda S, Heir P, Metcalf J, Li ASC, Agnihotri S, Pusch S, Yasin M, Li M, Burrell K, Mansouri S, *et al* (2019) CIC protein instability contributes to tumorigenesis in glioblastoma. *Nat Commun* 10: 661

Bunin GR, Feuer EJ, Witman PA & Meadows AT (1996) Increasing incidence of childhood cancer: report of 20 years experience from the greater Delaware Valley Pediatric Tumor Registry. *Paediatr Perinat Epidemiol* 10: 319–338

Burmeister T, Gökbuget N, Schwartz S, Fischer L, Hubert D, Sindram A, Hoelzer D & Thiel E (2010) Clinical features and prognostic implications of TCF3-PBX1 and ETV6-RUNX1 in adult acute lymphoblastic leukemia. *Haematologica* 95: 241–246

Bury M, Le Calvé B, Lessard F, Dal Maso T, Saliba J, Michiels C, Ferbeyre G & Blank V (2019) NFE2L3 Controls Colon Cancer Cell Growth through Regulation of DUX4, a CDK1 Inhibitor. *Cell Rep* 29: 1469-1481.e9

C. FC, W. WA, Lara D, J. RA, Marie-Sophie F, Michael K, C. WJ, Somesh S, Oliveira VR, Sascha S, *et al* (2021) Dissecting Herpes Simplex Virus 1-Induced Host Shutoff at the RNA Level. *J Virol* 95: e01399-20

Campbell AE, Dyle MC, Calviello L, Matheny T, Cortazar MA, Forman T, Fu R, Gillen AE, Floor SN & Jagannathan S (2021) The myopathic transcription factor DUX4 induces the production of truncated RNA-binding proteins in human muscle cells. *bioRxiv*: 2021.06.28.450189

Carlet M, Völse K, Vergalli J, Becker M, Herold T, Arner A, Senft D, Jurinovic V, Liu W-H, Gao Y, *et al* (2021) In vivo inducible reverse genetics in patients' tumors to identify individual therapeutic targets. *Nat Commun* 12: 5655

Cavallo F, Troglio F, Fagà G, Fancelli D, Shyti R, Trattaro S, Zanella M, D'Agostino G, Hughes JM, Cera MR, *et al* (2020) High-throughput screening identifies histone

deacetylase inhibitors that modulate GTF2I expression in 7q11.23 microduplication autism spectrum disorder patient-derived cortical neurons. *Mol Autism* 11: 88

El Chaer F, Keng M & Ballen KK (2020) MLL-Rearranged Acute Lymphoblastic Leukemia. *Curr Hematol Malig Rep* 15: 83–89

Chailangkarn T, Noree C & Muotri AR (2018) The contribution of GTF2I haploinsufficiency to Williams syndrome. *Mol Cell Probes* 40: 45–51

Chan AK-Y, Pang JC-S, Chung NY-F, Li KK-W, Poon WS, Chan DT-M, Shi Z, Chen L, Zhou L & Ng H-K (2014) Loss of CIC and FUBP1 expressions are potential markers of shorter time to recurrence in oligodendroglial tumors. *Mod Pathol an Off J United States Can Acad Pathol Inc* 27: 332–342

Chen C-W, Koche RP, Sinha AU, Deshpande AJ, Zhu N, Eng R, Doench JG, Xu H, Chu SH, Qi J, *et al* (2015) DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. *Nat Med* 21: 335–343

Chew G-LL, Campbell AE, De Neef E, Sutliff NA, Shadle SC, Tapscott SJ & Bradley RK (2019) DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Dev Cell* 50: 658-671.e7

Choi SH, Gearhart MD, Cui Z, Bosnakovski D, Kim M, Schennum N & Kyba M (2016) DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res* 44: 5161–5173

Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, Armour JALL & Hewitt JE (2007) Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 81: 264–279

Clappier E, Auclerc MF, Rapion J, Bakkus M, Caye A, Khemiri A, Giroux C, Hernandez L, Kabongo E, Savola S, *et al* (2014) An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia* 28: 70–77

Costa-Silva J, Domingues D & Lopes FM (2017) RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One* 12: 1–18

Coustan-Smith E, Song G, Clark C, Key L, Liu P, Mehrpooya M, Stow P, Su X, Shurtleff S, Pui C-H, *et al* (2011) New markers for minimal residual disease detection in acute lymphoblastic leukemia. *Blood* 117: 6267–6276

Crans HN & Sakamoto KM (2001) Transcription factors and translocations in lymphoid and myeloid leukemia. *Leukemia* 15: 313–331

Darko B, T. da SM, T. SS, T. EE, A. TE, Ce Y, Ziyou C, A. WM, Ajit J, Michael K, *et al* (2021) A novel P300 inhibitor reverses DUX4-mediated global histone H3 hyperacetylation, target gene expression, and cell death. *Sci Adv* 5: eaaw7781

Das S & Chadwick BP (2016) Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One* 11: 1–26

Deenen JCW, Arnts H, van der Maarel SM, Padberg GW, Verschuuren JJGM, Bakker E, Weinreich SS, Verbeek ALM & van Engelen BGM (2014) Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 83: 1056 LP – 1059

DeSalvo J, Ban Y, Li L, Sun X, Jiang Z, Kerr DA, Khanlari M, Boulina M, Capecchi MR, Partanen JM, *et al* (2021) ETV4 and ETV5 drive synovial sarcoma through cell cycle and DUX4 embryonic pathway control. *J Clin Invest* 131

DiGiuseppe JA, Fuller SG & Borowitz MJ (2009) Overexpression of CD49f in precursor B-cell acute lymphoblastic leukemia: potential usefulness in minimal residual disease detection. *Cytometry B Clin Cytom* 76: 150–155

Dittel BN, McCarthy JB, Wayner EA & LeBien TW (1993) Regulation of human B-cell precursor adhesion to bone marrow stromal cells by cytokines that exert opposing effects on the expression of vascular cell adhesion molecule-1 (VCAM-1). *Blood* 81: 2272–2282

Dmitriev P, Bou Saada Y, Dib C, Ansseau E, Barat A, Hamade A, Dessen P, Robert T, Lazar V, Louzada RAN, *et al* (2016) DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD

patients. *Free Radic Biol Med* 99: 244–258

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21

Dong X, Zhang W, Wu H, Huang J, Zhang M, Wang P, Zhang H, Chen Z, Chen SJ & Meng G (2018) Structural basis of DUX4/IGH-driven transactivation. *Leukemia* 32: 1466–1476

Dyer MJS, Akasaka T, Capasso M, Dusanjh P, Lee YF, Karran EL, Nagel I, Vater I, Cario G & Siebert R (2010) Immunoglobulin heavy chain locus chromosomal translocations in B-cell precursor acute lymphoblastic leukemia: rare clinical curios or potent genetic drivers? *Blood* 115: 1490–1499

F. LRJL, J. van der VP, Rinse K, Sabrina S, Pilar C, G. DJ, Lauren S, R. SK, Gert J van O, W. PG, *et al* (2010) A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy. *Science (80-)* 329: 1650–1653

Fackler MJ, Krause DS, Smith OM, Civin CI & May WS (1995) Full-length but not truncated CD34 inhibits hematopoietic cell differentiation of M1 cells. *Blood* 85: 3040–3047

Faham M, Zheng J, Moorhead M, Carlton VEH, Stow P, Coustan-Smith E, Pui C-H & Campana D (2012) Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. *Blood* 120: 5173–5180

Felschow DM, McVeigh ML, Hoehn GT, Civin CI & Fackler MJ (2001) The adapter protein CrkL associates with CD34. *Blood* 97: 3768–3775

Femke M. Hormann, Alex Q. Hoogkamer, H. Berna Beverloo, Aurélie Boeree, Ilse Dingjan, Moniek M. Wattel, Ronald W. Stam, Gabriele Escherich, Rob Pieters, Monique L. den Boer, *et al* (2019) NUTM1 is a recurrent fusion gene partner in B-cell precursor acute lymphoblastic leukemia associated with increased expression of genes on chromosome band 10p12.31-12.2. *Haematologica* 104: e455–e459

Feng Q, Snider L, Jagannathan S, Tawil R, van der Maarel SM, Tapscott SJ & Bradley RK (2015) A feedback loop between nonsense-mediated decay and the retrogene

DUX4 in facioscapulohumeral muscular dystrophy. *Elife* 2015: 1–13

Feng Y, Lei Y, Wu X, Huang Y, Rao H, Zhang Y & Wang F (2017) GTF2I mutation frequently occurs in more indolent thymic epithelial tumors and predicts better prognosis. *Lung Cancer* 110: 48–52

Fielding AK, Rowe JM, Buck G, Foroni L, Gerrard G, Litzow MR, Lazarus H, Luger SM, Marks DI, McMillan AK, *et al* (2014) UKALLXII/ECOG2993: addition of imatinib to a standard treatment regimen enhances long-term outcomes in Philadelphia positive acute lymphoblastic leukemia. *Blood* 123: 843–850

Fischer U, Forster M, Rinaldi A, Risch T, Sungalee S, Warnatz H-J, Bornhauser B, Gombert M, Kratsch C, Stütz AM, *et al* (2015) Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options. *Nat Genet* 47: 1020–1029

Foà R, Bassan R, Vitale A, Elia L, Piciocchi A, Puzzolo M-C, Canichella M, Viero P, Ferrara F, Lunghi M, *et al* (2020) Dasatinib–Blinatumomab for Ph-Positive Acute Lymphoblastic Leukemia in Adults. *N Engl J Med* 383: 1613–1623

Foà R, Vitale A, Vignetti M, Meloni G, Guarini A, De Propris MS, Elia L, Paoloni F, Fazi P, Cimino G, *et al* (2011) Dasatinib as first-line treatment for adult patients with Philadelphia chromosome–positive acute lymphoblastic leukemia. *Blood* 118: 6521–6528

Fonseca ICCFE, da Luz FAC, Uehara IA & Silva MJB (2018) Cell-adhesion molecules and their soluble forms: Promising predictors of 'tumor progression' and relapse in leukemia. *Tumour Biol J Int Soc Oncodevelopmental Biol Med* 40: 1010428318811525

Forés M, Simón-Carrasco L, Ajuria L, Samper N, González-Crespo S, Drosten M, Barbacid M & Jiménez G (2017) A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer. *PLoS Genet* 13: e1006622

Full F, van Gent M, Sparrer KMJ, Chiang C, Zurenski MA, Scherer M, Brockmeyer NH, Heinzerling L, Stürzl M, Korn K, *et al* (2019) Centrosomal protein TRIM43

restricts herpesvirus infection by regulating nuclear lamina integrity. *Nat Microbiol* 4: 164–176

Furness SGB & McNagny K (2006) Beyond mere markers: functions for CD34 family of sialomucins in hematopoiesis. *Immunol Res* 34: 13–32

Gabriëls J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al* (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 236: 25–32

Gang EJ, Kim HN, Hsieh Y-T, Ruan Y, Ogana HA, Lee S, Pham J, Geng H, Park E, Klemm L, *et al* (2020) Integrin α6 mediates the drug resistance of acute lymphoblastic B-cell leukemia. *Blood* 136: 210–223

Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, vanderMaarel SM, Ruzzo WL, Gentleman RC, Tawil R, *et al* (2012) DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. *Dev Cell* 22: 38–51

Glatter T, Wepf A, Aebersold R & Gstaiger M (2009) An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol Syst Biol* 5: 237

Glodek AM, Honczarenko M, Le Y, Campbell JJ & Silberstein LE (2003) Sustained activation of cell adhesion is a differentially regulated process in B lymphopoiesis. *J Exp Med* 197: 461–473

Gocho Y & Yang JJ (2019) Genetic defects in hematopoietic transcription factors and predisposition to acute lymphoblastic leukemia. *Blood* 134: 793–797

Gu Z, Churchman M, Roberts K, Li Y, Liu Y, Harvey RC, McCastlain K, Reshmi SC, Payne-Turner D, Iacobucci I, *et al* (2016) Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat Commun* 7: 13331

Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, Hagiwara K, Pelletier S, Gingras S, Berns H, *et al* (2019) PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet* 51: 296–307

Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–674

Harrison CJ (2015) Blood Spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease. *Blood* 125: 1383–1386

Harvey RC, Mullighan CG, Wang X, Dobbin KK, Davidson GS, Bedrick EJ, Chen I-M, Atlas SR, Kang H, Ar K, *et al* (2010) Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood* 116: 4874–4884

Healy L, May G, Gale K, Grosveld F, Greaves M & Enver T (1995) The stem cell antigen CD34 functions as a regulator of hemopoietic cell adhesion. *Proc Natl Acad Sci* 92: 12240 LP – 12244

Heerema NA, Carroll AJ, Devidas M, Loh ML, Borowitz MJ, Gastier-Foster JM, Larsen EC, Mattano LA, Maloney KW, Willman CL, *et al* (2013) Intrachromosomal Amplification of Chromosome 21 Is Associated With Inferior Outcomes in Children With Acute Lymphoblastic Leukemia Treated in Contemporary Standard-Risk Children's Oncology Group Studies: A Report From the Children's Oncology Group. *J Clin Oncol* 31: 3397–3402

Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL, *et al* (2017) Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* 49: 925–934

Henikoff S, Henikoff JG, Kaya-Okur HS & Ahmad K (2020) Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *Elife* 9

Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, van Deutekom JC, Francis F, Sharpe PT & Hofker M (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3: 1287–1295

Himeda CL & Jones PL (2019) The Genetics and Epigenetics of Facioscapulohumeral

Muscular Dystrophy. *Annu Rev Genomics Hum Genet* 20: 265–291

Homma S, Beermann M Lou, Boyce FM & Miller JB (2015) Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. *Ann Clin Transl Neurol* 2: 151–166

Hu T, Zhou R, Zhao Y & Wu G (2016) Integrin α6/Akt/Erk signaling is essential for human breast cancer resistance to radiotherapy. *Sci Rep* 6: 33376

Hunger SP & Mullighan CG (2015) Acute lymphoblastic leukemia in children. *N Engl J Med* 373: 1541–1552

Hystad ME, Myklebust JH, Bø TH, Sivertsen EA, Rian E, Forfang L, Munthe E, Rosenwald A, Chiorazzi M, Jonassen I, *et al* (2007) Characterization of early stages of human B cell development by gene expression profiling. *J Immunol* 179: 3662–3671

De Iaco A, Planet E, Coluccio A, Verp S, Duc J & Trono D (2017) DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* 49: 941–945

Iacobucci I, Kimura S & Mullighan CG (2021) Biologic and therapeutic implications of genomic alterations in acute lymphoblastic leukemia. *J Clin Med* 10

Iacobucci I & Mullighan CG (2017) Genetic basis of acute lymphoblastic leukemia. *J Clin Oncol* 35: 975–983

Iacobucci I, Storlazzi CT, Cilloni D, Lonetti A, Ottaviani E, Soverini S, Astolfi A, Chiaretti S, Vitale A, Messa F, *et al* (2009) Identification and molecular characterization of recurrent genomic deletions on 7p12 in the IKZF1 gene in a large cohort of BCR-ABL1-positive acute lymphoblastic leukemia patients: on behalf of Gruppo Italiano Malattie Ematologiche dell'Adulto Acute Leuk. *Blood* 114: 2159–2167

Inaba H, Azzato EM & Mullighan CG (2017) Integration of Next-Generation Sequencing to Treat Acute Lymphoblastic Leukemia with Targetable Lesions: The St. Jude Children's Research Hospital Approach. *Front Pediatr* 5: 258

Inaba H, Greaves M & Mullighan CG (2013) Acute lymphoblastic leukaemia. *Lancet* 381: 1943–1955

Inaba H & Mullighan CG (2020) Pediatric acute lymphoblastic leukemia. *Haematologica* 105: 2524–2539

Inaba T, Inukai T, Yoshihara T, Seyschab H, Ashmun RA, Canman CE, Laken SJ, Kastan MB & Look AT (1996) Reversal of apoptosis by the leukaemia-associated E2A–HLF chimaeric transcription factor. *Nature* 382: 541–544

Jabbour E, O'Brien S, Konopleva M & Kantarjian H (2015) New insights into the pathophysiology and therapy of adult acute lymphoblastic leukemia. *Cancer* 121: 2517–2528

Jabbour EJ, Faderl S & Kantarjian HM (2005) Adult acute lymphoblastic leukemia. *Mayo Clin Proc* 80: 1517–1527

Jagannathan S, Ogata Y, Gafken PR, Tapscott SJ & Bradley RK (2019) Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. *Elife* 8: 1–16

Jagannathan S, Shadle SC, Resnick R, Snider L, Tawil RN, van der Maarel SM, Bradley RK & Tapscott SJ (2016) Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum Mol Genet* 25: 4419–4431

Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ & Thun MJ (2004) Cancer statistics, 2004. *CA Cancer J Clin* 54: 8–29

Jiang Z, Wu D, Lin S & Li P (2016) CD34 and CD38 are prognostic biomarkers for acute B lymphoblastic leukemia. *Biomark Res* 4: 23

Kao Y-C, Sung Y-S, Chen C-L, Zhang L, Dickson BC, Swanson D, Vaiyapuri S, Latif F, Alholle A, Huang S-C, *et al* (2017) ETV transcriptional upregulation is more reliable than RNA sequencing algorithms and FISH in diagnosing round cell sarcomas with CIC gene rearrangements. *Genes Chromosomes Cancer* 56: 501–510

Kapoor S, Shenoy SP & Bose B (2020) CD34 cells in somatic, regenerative and cancer

stem cells: Developmental biology, cell therapy, and omics big data perspective. *J Cell Biochem* 121: 3058–3069

Kawazu M, Yamamoto G, Yoshimi M, Yamamoto K, Asai T, Ichikawa M, Seo S, Nakagawa M, Chiba S, Kurokawa M, *et al* (2007a) Expression Profiling of Immature Thymocytes Revealed a Novel Homeobox Gene That Regulates Double-Negative Thymocyte Development. *J Immunol* 179: 5335–5345

Kawazu M, Yamamoto G, Yoshimi M, Yamamoto K, Asai T, Ichikawa M, Seo S, Nakagawa M, Chiba S, Kurokawa M, *et al* (2007b) Expression Profiling of Immature Thymocytes Revealed a Novel Homeobox Gene That Regulates Double-Negative Thymocyte Development. *J Immunol* 179: 5335–5345

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K & Henikoff S (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10: 1930

Kentaro Ohki, Nobutaka Kiyokawa, Yuya Saito, Shinsuke Hirabayashi, Kazuhiko Nakabayashi, Hitoshi Ichikawa, Yukihide Momozawa, Kohji Okamura, Ai Yoshimi, Hiroko Ogata-Kawata, *et al* (2019) Clinical and molecular characteristics of MEF2D fusion-positive B-cell precursor acute lymphoblastic leukemia in childhood, including a novel translocation resulting in MEF2D-HNRNPH1 gene fusion. *Haematologica* 104: 128–137

Kimura S & Mullighan CG (2020) Molecular markers in ALL: Clinical implications. *Best Pract Res Clin Haematol* 33: 1–24

Klein F, Mitrovic M, Roux J, Engdahl C, von Muenchow L, Alberti-Servera L, Fehling HJ, Pelczar P, Rolink A & Tsapogas P (2019) The transcription factor Duxbl mediates elimination of pre-T cells that fail β-selection. *J Exp Med* 216: 638–655

Klossowski S, Miao H, Kempinska K, Wu T, Purohit T, Kim E, Linhares BM, Chen D, Jih G, Perkey E, *et al* (2020) Menin inhibitor MI-3454 induces remission in MLL1-rearranged and NPM1-mutated models of leukemia. *J Clin Invest* 130: 981–997

Knopp P, Krom YD, Banerji CRSS, Panamarova M, Moyle LA, den Hamer B, van der

Maarel SMSMSM & Zammit PS (2016) DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. *J Cell Sci* 129: 3816–3831

Köcher T & Superti-Furga G (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* 4: 807–815

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, *et al* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517: 583–588

Kwon J, Lee T-S, Lee Won H, Kang Chul M, Yoon H-J, Kim J-H & Park Ho J (2013) Integrin alpha 6: A novel therapeutic target in esophageal squamous cell carcinoma. *Int J Oncol* 43: 1523–1530

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359

Laurenti E, Doulatov S, Zandi S, Plumb I, Chen J, April C, Fan J-B & Dick JE (2013) The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol* 14: 756–763

Lee J-S, Kim E, Lee J, Kim D, Kim H, Kim C-J, Kim S, Jeong D & Lee Y (2020) Capicua suppresses colorectal cancer progression via repression of ETV4 expression. *Cancer Cell Int* 20: 42

Lee JK, Bosnakovski D, Toso EA, Dinh T, Banerjee S, Bohl TE, Shi K, Orellana K, Kyba M & Aihara H (2018) Crystal Structure of the Double Homeodomain of DUX4 in Complex with DNA. *Cell Rep* 25: 2955-2962.e3

Lee Y (2020) Regulation and function of capicua in mammals. *Exp Mol Med* 52: 531–537

Leidenroth A & Hewitt JE (2010) A family history of DUX4: Phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol Biol* 10: 364

Lek A, Zhang Y, Woodman KG, Huang S, DeSimone AM, Cohen J, Ho V, Conner J,

Mead L, Kodani A, *et al* (2020) Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci Transl Med* 12: 9–11

Lemmers RJLF, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CMM, de Knijff P, Padberg GW, Frants RR & van der Maarel SM (2007) Specific Sequence Variations within the 4q35 Region Are Associated with Facioscapulohumeral Muscular Dystrophy. *Am J Hum Genet* 81: 884–894

Li J, Dai Y, Wu L, Zhang M, Ouyang W, Huang J & Chen S (2021) Emerging molecular subtypes and therapeutic targets in B-cell precursor acute lymphoblastic leukemia. *Front Med* 15: 347–371

Li J, Zhong H-Y, Zhang Y, Xiao L, Bai L-H, Liu S-F, Zhou G-B & Zhang G-S (2015) GTF2I-RARA is a novel fusion transcript in a t(7;17) variant of acute promyelocytic leukaemia with clinical resistance to retinoic acid. *Br J Haematol* 168: 904–908 doi:10.1111/bjh.13157 [PREPRINT]

Li JF, Dai YT, Lilljebjörn H, Shen SH, Cui BW, Bai L, Liu YF, Qian MX, Kubota Y, Kiyoi H, *et al* (2018a) Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc Natl Acad Sci U S A* 115: E11711–E11720

Li Y (2011) The tandem affinity purification technology: an overview. *Biotechnol Lett* 33: 1487–1499

Li Y, Wu B, Liu H, Gao Y, Yang C, Chen X, Zhang J, Chen Y, Gu Y & Li J (2018b) Structural basis for multiple gene regulation by human DUX4. *Biochem Biophys Res Commun* 505: 1161–1167

Liao Y, Smyth GK & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930

Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, Olsson L, Orsmark-Pietras C, Von Palffy S, Askmyr M, Rissler M, Schrappe M, Cario G, *et al* (2016) Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell

precursor acute lymphoblastic leukaemia. *Nat Commun* 7: 11790

Lim KRQ, Nguyen Q & Yokota T (2020) Dux4 signalling in the pathogenesis of facioscapulohumeral muscular dystrophy. *Int J Mol Sci* 21

Linet MS, Ries LA, Smith MA, Tarone RE & Devesa SS (1999) Cancer surveillance series: recent trends in childhood cancer incidence and mortality in the United States. *J Natl Cancer Inst* 91: 1051–1058

Liu Y-FF, Wang B-YY, Zhang WGWNW-GW-NWG, Huang J-YY, Li B-SS, Zhang M, Jiang L, Li JFJMJFJ-MJ-F, Wang M-JJ, Dai Y-JJ, *et al* (2016) Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine* 8: 173–183

Liu YYY-FF, Wang B-YY, Zhang WNW-GWW-NWG, Huang JYJJ-Y, Li B-SSB, Zhang M, Jiang L, Li JMJFJ-MJJJ-F, Wang M-JJ, Dai YTY-JYJ, *et al* (2019) Transcriptional activities of DUX4 fusions in B-cell acute lymphoblastic leukemia. *Nat Genet* 10: 1–13

Look AT (1997) Oncogenic transcription factors in the human acute leukemias. *Science* 278: 1059–1064

Loughran SJ, Kruse EA, Hacking DF, de Graaf CA, Hyland CD, Willson TA, Henley KJ, Ellis S, Voss AK, Metcalf D, *et al* (2008) The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol* 9: 810–819

van der Maarel SM, Frants RR & Padberg GW (2007) Facioscapulohumeral muscular dystrophy. *Biochim Biophys Acta - Mol Basis Dis* 1772: 186–194

Malaguti M, Migueles RP, Blin G, Lin C-Y & Lowell S (2019) Id1 Stabilizes Epiblast Identity by Sensing Delays in Nodal Activation and Adjusting the Timing of Differentiation. *Dev Cell* 50: 462-477.e5

Malenfant P, Liu X, Hudson ML, Qiao Y, Hrynchak M, Riendeau N, Hildebrand MJ, Cohen IL, Chudley AE, Forster-Gibson C, *et al* (2012) Association of GTF2i in the Williams-Beuren syndrome critical region with autism spectrum disorders. *J Autism Dev Disord* 42: 1459–1469

Marcucci G, Baldus CD, Ruppert AS, Radmacher MD, Mrózek K, Whitman SP, Kolitz JE, Edwards CG, Vardiman JW, Powell BL, *et al* (2005) Overexpression of the ETS-Related Gene, ERG, Predicts a Worse Outcome in Acute Myeloid Leukemia With Normal Karyotype: A Cancer and Leukemia Group B Study. *J Clin Oncol* 23: 9234–9242

Marincevic-Zuniga Y, Dahlberg J, Nilsson S, Raine A, Nystedt S, Lindqvist CM, Berglund EC, Abrahamsson J, Cavelier L, Forestier E, *et al* (2017) Transcriptome sequencing in pediatric acute lymphoblastic leukemia identifies fusion genes associated with distinct DNA methylation profiles. *J Hematol Oncol* 10: 148

Marketa Zaliova, Jan Stuchly, Lucie Winkowska, Alena Musilova, Karel Fiser, Martina Slamova, Julia Starkova, Martina Vaskova, Ondrej Hrusak, Lucie Sramkova, *et al* (2019) Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort. *Haematologica* 104: 1396–1406

Martinelli G, Iacobucci I, Storlazzi CT, Vignetti M, Paoloni F, Cilloni D, Soverini S, Vitale A, Chiaretti S, Cimino G, *et al* (2009) IKZF1 (Ikaros) Deletions in BCR-ABL1–Positive Acute Lymphoblastic Leukemia Are Associated With Short Disease-Free Survival and High Rate of Cumulative Incidence of Relapse: A GIMEMA AL WP Report. *J Clin Oncol* 27: 5202–5207

Mocciaro E, Runfola V, Ghezzi P, Pannese M & Gabellini D (2021) DUX4 Role in Normal Physiology and in FSHD Muscular Dystrophy. *Cells* 10 doi:10.3390/cells10123322 [PREPRINT]

Moorman A V, Harrison CJ, Buck GAN, Richards SM, Secker-Walker LM, Martineau M, Vance GH, Cherry AM, Higgins RR, Fielding AK, *et al* (2007) Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood* 109: 3189–3197

Moorman A V, Robinson H, Schwab C, Richards SM, Hancock J, Mitchell CD, Goulden N, Vora A & Harrison CJ (2013) Risk-Directed Treatment Intensification Significantly Reduces the Risk of Relapse Among Children and Adolescents With

Acute Lymphoblastic Leukemia and Intrachromosomal Amplification of Chromosome 21: A Comparison of the MRC ALL97/99 and UKALL2003 Trials. *J Clin Oncol* 31: 3389–3396

Morgan M, Falcon S GR (2021) GSEABase: Gene set enrichment data structures and methods. *R Packag version 1560*

Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M, Yang W, Cheng C, Cao X, Quinn E, *et al* (2015) Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol* 16: 1659–1666

Moyle LA, Blanc E, Jaka O, Prueller J, Banerji CR, Tedesco FS, Harridge SD, Knight RD & Zammit PS (2016) Ret function in muscle stem cells points to tyrosine kinase inhibitor therapy for facioscapulohumeral muscular dystrophy. *Elife* 5

Mullighan CG (2012) Molecular genetics of B-precursor acute lymphoblastic leukemia. *J Clin Invest* 122: 3407–3415

Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, *et al* (2007a) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446: 758–764

Mullighan CG, Miller CB, Su X, Radtke I, Dalton J, Song G, Zhou X, Pui C-H, Shurtleff SA & Downing JR (2007b) ERG Deletions Define a Novel Subtype of B-Progenitor Acute Lymphoblastic Leukemia. *Blood* 110: 691

Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA & Downing JR (2008) Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 322: 1377–1380

Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, Ma J, Liu W, Cheng C, Schulman BA, *et al* (2009) Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* 360: 470–480

Nakai S, Yamada S, Outani H, Nakai T, Yasuda N, Mae H, Imura Y, Wakamatsu T, Tamiya H, Tanaka T, *et al* (2019) Establishment of a novel human CIC-DUX4 sarcoma cell line, Kitra-SRS, with autocrine IGF-1R activation and metastatic

potential to the lungs. *Sci Rep* 9: 15812

Nibourel O, Guihard S, Roumier C, Pottier N, Terre C, Paquet A, Peyrouze P, Geffroy S, Quentin S, Alberdi A, *et al* (2017) Copy-number analysis identified new prognostic marker in acute myeloid leukemia. *Leukemia* 31: 555–564

Nielsen JS & McNagny KM (2008) Novel functions of the CD34 family. *J Cell Sci* 121: 3683–3692

Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A, Rajpurkar M, Jones K, Gowan K, Balduini C, *et al* (2015) Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet* 47: 535–538

Novakova M, Vakrmanova B, Slamova L, Musilova A, Brüggemann M, Ritgen M, Fronkova E, Kalina T, Trka J, Stary J, *et al* (2018) Switching Towards Monocytic Lineage and Discordancy between Flow Cytometric and PCR Minimal Residual Disease Results Is a Hallmark Feature of DUX4 Rearranged B-Cell Precursor Acute Lymphoblastic Leukemia. *Blood* 132: 2825

Ohnishi H, Sasaki H, Nakamura Y, Kato S, Ando K, Narimatsu H & Tachibana K (2013) Regulation of cell shape and adhesion by CD34. *Cell Adh Migr* 7: 426–433

Okimoto RA, Breitenbuecher F, Olivas VR, Wu W, Gini B, Hofree M, Asthana S, Hrustanovic G, Flanagan J, Tulpule A, *et al* (2017) Inactivation of Capicua drives cancer metastasis. *Nat Genet* 49: 87–96

Oyama R, Takahashi M, Yoshida A, Sakumoto M, Takai Y, Kito F, Shiozawa K, Qiao Z, Arai Y, Shibata T, *et al* (2017) Generation of novel patient-derived CIC- DUX4 sarcoma xenografts and cell lines. *Sci Rep* 7: 4712

Padberg GW (2009) Facioscapulohumeral Muscular Dystrophy. *Int Neurol*: 223–224 doi:https://doi.org/10.1002/9781444317008.ch59 [PREPRINT]

Panagopoulos I, Brunetti M, Stoltenberg M, Strandabø RAU, Staurseth J, Andersen K, Kostolomov I, Hveem TS, Lorenz S, Nystad TA, *et al* (2019) Novel GTF2I-PDGFRB and IKZF1-TYW1 fusions in pediatric leukemia with normal karyotype. *Exp Hematol Oncol* 8: 12 doi:10.1186/s40164-019-0136-y [PREPRINT]

Paul S, Kantarjian H & Jabbour EJ (2016) Adult Acute Lymphoblastic Leukemia. *Mayo Clin Proc* 91: 1645–1666

Paulsson K, Lilljebjörn H, Biloglav A, Olsson L, Rissler M, Castor A, Barbany G, Fogelstrand L, Nordgren A, Sjögren H, *et al* (2015) The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Genet* 47: 672–676

Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, Xu H, Gastier-Foster J, E S, Lim JY-S, *et al* (2013) Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet* 45: 1494–1498

Petrini I, Meltzer PS, Kim I-K, Lucchi M, Park K-S, Fontanini G, Gao J, Zucali PA, Calabrese F, Favaretto A, *et al* (2014) A specific missense mutation in GTF2I occurs at high frequency in thymic epithelial tumors. *Nat Genet* 46: 844–849

Potuckova E, Zuna J, Hovorkova L, Starkova J, Stary J, Trka J & Zaliova M (2016) Intragenic ERG Deletions Do Not Explain the Biology of ERG-Related Acute Lymphoblastic Leukemia. *PLoS One* 11: e0160385

Prasad DD, Ouchida M, Lee L, Rao VN & Reddy ES (1994) TLS/FUS fusion domain of TLS/FUS-erg chimeric protein resulting from the t(16;21) chromosomal translocation in human myeloid leukemia functions as a transcriptional activation domain. *Oncogene* 9: 3717–3729

Preussner J, Zhong J, Sreenivasan K, Günther S, Engleitner T, Künne C, Glatzel M, Rad R, Looso M, Braun T, *et al* (2018) Oncogenic Amplification of Zygotic Dux Factors in Regenerating p53-Deficient Muscle Stem Cells Defines a Molecular Cancer Subtype. *Cell Stem Cell* 23: 794-805.e4

Pui C-H, Nichols KE & Yang JJ (2019) Somatic and germline genomics in paediatric acute lymphoblastic leukaemia. *Nat Rev Clin Oncol* 16: 227–240

Qian M, Xu H, Perez-Andreu V, Roberts KG, Zhang H, Yang W, Zhang S, Zhao X, Smith C, Devidas M, *et al* (2019a) Novel susceptibility variants at the ERG locus for childhood acute lymphoblastic leukemia in Hispanics. *Blood* 133: 724–729

Qian M, Zhang H, Kham SK-Y, Liu S, Jiang C, Zhao X, Lu Y, Goodings C, Lin T-N, Zhang R, *et al* (2017) Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. *Genome Res* 27: 185–195

Qian M, Zhao X, Devidas M, Yang W, Gocho Y, Smith C, Gastier-Foster JM, Li Y, Xu H, Zhang S, *et al* (2019b) Genome-Wide Association Study of Susceptibility Loci for T-Cell Acute Lymphoblastic Leukemia in Children. *J Natl Cancer Inst* 111: 1350–1357

Rabbitts TH (2001) Chromosomal translocation master genes, mouse models and experimental therapeutics. *Oncogene* 20: 5763–5777

Rainis L, Toki T, Pimanda JE, Rosenthal E, Machol K, Strehl S, Göttgens B, Ito E & Izraeli S (2005) The Proto-Oncogene ERG in Megakaryoblastic Leukemias. *Cancer Res* 65: 7596 LP – 7602

Reshmi SC, Harvey RC, Roberts KG, Stonerock E, Smith A, Jenkins H, Chen I-M, Valentine M, Liu Y, Li Y, *et al* (2017) Targetable kinase gene fusions in high-risk B-ALL: a study from the Children's Oncology Group. *Blood* 129: 3352–3361

Resnick R, Wong C-JJ, Hamm DC, Bennett SR, Skene PJ, Hake SB, Henikoff S, van der Maarel SM & Tapscott SJ (2019) DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep* 29: 1812-1820.e5

Rickard AM, Petek LM & Miller DG (2015) Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum Mol Genet* 24: 5901–5914

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M & Séraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17: 1030–1032

Roberts KG (2018) Genetics and prognosis of ALL in children vs adults. *Hematol Am Soc Hematol Educ Progr* 2018: 137–145

Roberts KG, Gu Z, Payne-Turner D, McCastlain K, Harvey RC, Chen I-M, Pei D, Iacobucci I, Valentine M, Pounds SB, *et al* (2017) High Frequency and Poor

Outcome of Philadelphia Chromosome-Like Acute Lymphoblastic Leukemia in Adults. *J Clin Oncol Off J Am Soc Clin Oncol* 35: 394–401

Roberts KG, Li Y, Payne-Turner D, Harvey RC, Yang Y-L, Pei D, McCastlain K, Ding L, Lu C, Song G, *et al* (2014) Targetable Kinase-Activating Lesions in Ph-like Acute Lymphoblastic Leukemia. *N Engl J Med* 371: 1005–1015

Roberts KG & Mullighan CG (2020) The biology of B-progenitor acute lymphoblastic leukemia. *Cold Spring Harb Perspect Med* 10: 1–22

Rodriguez I, Ody C, Araki K, Garcia I & Vassalli P (1997) An early and massive wave of germinal cell apoptosis is required for the development of functional spermatogenesis. *EMBO J* 16: 2262–2270

Roy AL (2012) Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10 years later. *Gene* 492: 32–41

Roy AL (2017) Pathophysiology of TFII-I: Old Guard Wearing New Hats. *Trends Mol Med* 23: 501–511

Sacconi S, Salviati L & Desnuelle C (2015) Facioscapulohumeral muscular dystrophy. *Biochim Biophys Acta - Mol Basis Dis* 1852: 607–614

Sasaki K, Yamauchi T, Semba Y, Nogami J, Imanaga H, Terasaki T, Nakao F, Akahane K, Inukai T, Verhoeyen E, *et al* (2022) Genome-wide CRISPR-Cas9 screen identifies rationally designed combination therapies for CRLF2-rearranged Ph-like ALL. *Blood* 139: 748–760

Schätzl T, Kaiser L & Deigner H-P (2021) Facioscapulohumeral muscular dystrophy: genetics, gene activation and downstream signalling with regard to recent therapeutic approaches: an update. *Orphanet J Rare Dis* 16: 129

Schinnerl D, Mejstrikova E, Schumich A, Zaliova M, Fortschegger K, Nebral K, Attarbaschi A, Fiser K, Kauer MO, Popitsch N, *et al* (2019) CD371 cell surface expression: a unique feature of DUX4 -rearranged acute lymphoblastic leukemia. *Haematologica* 104: e352–e355

Schroeder MP, Bastian L, Eckert C, Gökbuget N, James AR, Tanchez JO, Schlee C,

Isaakidis K, Häupl B, Baum K, *et al* (2019) Integrated analysis of relapsed B-cell precursor Acute Lymphoblastic Leukemia identifies subtype-specific cytokine and metabolic signatures. *Sci Rep* 9: 4188

Schwab C & Harrison CJ (2018) Advances in B-cell Precursor Acute Lymphoblastic Leukemia Genomics. *HemaSphere* 2

Schwartz BE, Hofer MD, Lemieux ME, Bauer DE, Cameron MJ, West NH, Agoston ES, Reynoird N, Khochbin S, Ince TA, *et al* (2011) Differentiation of NUT midline carcinoma by epigenomic reprogramming. *Cancer Res* 71: 2686–2696

Seong BKA, Dharia N V, Lin S, Donovan KA, Chong S, Robichaud A, Conway A, Hamze A, Ross L, Alexe G, *et al* (2021) TRIM8 modulates the EWS/FLI oncoprotein to promote survival in Ewing sarcoma. *Cancer Cell* 39: 1262-1278.e7

Shadle SC, Zhong JW, Campbell AE, Conerly ML, Jagannathan S, Wong C-JJ, Morello TD, van der Maarel SM & Tapscott SJ (2017) DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet* 13: 1–25

Shaha C, Tripathi R & Mishra DP (2010) Male germ cell apoptosis: regulation and biology. *Philos Trans R Soc London Ser B, Biol Sci* 365: 1501–1515

Sharma ND, Keewan E & Matlawska-Wasowska K (2021) Metabolic Reprogramming and Cell Adhesion in Acute Leukemia Adaptation to the CNS Niche. *Front Cell Dev Biol* 9

Shirai Y, Li W & Suzuki T (2017) Role of Splice Variants of Gtf2i, a Transcription Factor Localizing at Postsynaptic Sites, and Its Relation to Neuropsychiatric Diseases. *Int J Mol Sci* 18

Simón-Carrasco L, Graña O, Salmón M, Jacob HKC, Gutierrez A, Jiménez G, Drosten M & Barbacid M (2017) Inactivation of Capicua in adult mice causes T-cell lymphoblastic lymphoma. *Genes Dev* 31: 1456–1468

Sive JI, Buck G, Fielding A, Lazarus HM, Litzow MR, Luger S, Marks DI, McMillan A, Moorman A V, Richards SM, *et al* (2012) Outcomes in older adults with acute lymphoblastic leukaemia (ALL): results from the international MRC UKALL

XII/ECOG2993 trial. *Br J Haematol* 157: 463–471

Slamova L, Starkova J, Fronkova E, Zaliova M, Reznickova L, van Delft FW, Vodickova E, Volejnikova J, Zemanova Z, Polgarova K, *et al* (2014) CD2-positive B-cell precursor acute lymphoblastic leukemia with an early switch to the monocytic lineage. *Leukemia* 28: 609–620

Slayton WB, Schultz KR, Kairalla JA, Devidas M, Mi X, Pulsipher MA, Chang BH, Mullighan C, Iacobucci I, Silverman LB, *et al* (2018) Dasatinib Plus Intensive Chemotherapy in Children, Adolescents, and Young Adults With Philadelphia Chromosome–Positive Acute Lymphoblastic Leukemia: Results of Children's Oncology Group Trial AALL0622. *J Clin Oncol* 36: 2306–2314

de Smith AJ, Lavoie G, Walsh KM, Aujla S, Evans E, Hansen HM, Smirnov I, Kang AY, Zenker M, Ceremsak JJ, *et al* (2019) Predisposing germline mutations in high hyperdiploid acute lymphoblastic leukemia in children. *Genes Chromosomes Cancer* 58: 723–730

Snider L, Geng LN, Lemmers RJLFLF, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der Maarel SM, Tapscott SJ, *et al* (2010) Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet* 6: 1–14

Stefania G, Giuseppe A, Paolo A, Sergio B, Vanna Chiarion S, Alessandro C, Ferdinando DV, Lucia DM, Massimo DM, Maria Teresa I, *et al* (2015) I numeri del cancro in Italia. *Rep AIOM-AIRTUM*: 1–232

Stock W, Luger SM, Advani AS, Geyer S, Harvey RC, Mullighan CG, Willman CL, Malnassy G, Parker E, Laumann KM, *et al* (2014) Favorable Outcomes for Older Adolescents and Young Adults (AYA) with Acute Lymphoblastic Leukemia (ALL): Early Results of U.S. Intergroup Trial C10403. *Blood* 124: 796

Stock W, Luger SM, Advani AS, Yin J, Harvey RC, Mullighan CG, Willman CL, Fulton N, Laumann KM, Malnassy G, *et al* (2019) A pediatric regimen for older adolescents and young adults with acute lymphoblastic leukemia: results of CALGB 10403. *Blood* 133: 1548–1559

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA,

Paulovich A, Pomeroy SL, Golub TR, Lander ES, *et al* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 102: 15545 LP – 15550

Suda J, Sudo T, Ito M, Ohno N, Yamaguchi Y & Suda T (1992) Two types of murine CD34 mRNA generated by alternative splicing. *Blood* 79: 2288–2295

Sundaresh A & Williams O (2017) Mechanism of ETV6-RUNX1 Leukemia. In *RUNX Proteins in Development and Cancer*, Groner Y Ito Y Liu P Neil JC Speck NA & van Wijnen A (eds) pp 201–216. Singapore: Springer Singapore

Taoudi S, Bee T, Hilton A, Knezevic K, Scott J, Willson TA, Collin C, Thomas T, Voss AK, Kile BT, *et al* (2011) ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. *Genes Dev* 25: 251–262

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

Teytelman L, Thurtle DM, Rine J & van Oudenaarden A (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci* 110: 18602 LP – 18607

Thoms JAI, Birger Y, Foster S, Knezevic K, Kirschenbaum Y, Chandrakanthan V, Jonquieres G, Spensberger D, Wong JW, Oram SH, *et al* (2011) ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood* 117: 7079–7089

Tian L, Shao Y, Nance S, Dang J, Xu B, Ma X, Li Y, Ju B, Dong L, Newman S, *et al* (2019) Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia. *Nat Commun* 10: 2789

Tsuzuki S, Taguchi O & Seto M (2011) Promotion and maintenance of leukemia by ERG. *Blood* 117: 3858–3868

Tzoneva G, Dieck CL, Oshima K, Ambesi-Impiombato A, Sánchez-Martín M, Madubata CJ, Khiabanian H, Yu J, Waanders E, Iacobucci I, *et al* (2018) Clonal evolution mechanisms in NT5C2 mutant-relapsed acute lymphoblastic leukaemia.

Nature 553: 511–514

Uemura N, Salgia R, Li J-L, Pisick E, Sattler M & Griffin JD (1997) The BCR/ABL oncogene alters interaction of the adapter proteins CRKL and CRK with cellular proteins. *Leukemia* 11: 376–385

Valentine MC, Linabery AM, Chasnoff S, Hughes AEO, Mallaney C, Sanchez N, Giacalone J, Heerema NA, Hilden JM, Spector LG, *et al* (2014) Excess congenital non-synonymous variation in leukemia-associated genes in MLL- infant leukemia: a Children's Oncology Group report. *Leukemia* 28: 1235–1241

Vastenhouw NL, Cao WX & Lipshitz HD (2019) The maternal-to-zygotic transition revisited. *Development* 146: dev161471

Vuoristo S, Hydén-Granskog C, Yoshihara M, Gawriyski L, Damdimopoulos A, Bhagat S, Hashimoto K, Krjutškov K, Ezer S, Paluoja P, *et al* (2019) DUX4 regulates oocyte to embryo transition in human. *bioRxiv*: 732289

Wagner KR (2019) Facioscapulohumeral Muscular Dystrophies. *Continuum (Minneap Minn)* 25: 1662–1681

Wang LH & Tawil R (2016) Facioscapulohumeral Dystrophy. *Curr Neurol Neurosci Rep* 16: 66

Whiddon JL, Langford AT, Wong CJ, Zhong JW & Tapscott SJ (2017) Conservation and innovation in the DUX4-family gene network. *Nat Genet* 49: 935–940

Winters AC & Bernt KM (2017) MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* 5: 4

Wong D & Yip S (2020) Making heads or tails - the emergence of capicua (CIC) as an important multifunctional tumour suppressor. *J Pathol* 250: 532–540

Yao H, Price TT, Cantelli G, Ngo B, Warner MJ, Olivere L, Ridge SM, Jablonski EM, Therrien J, Tannheimer S, *et al* (2018) Leukaemia hijacks a neural mechanism to invade the central nervous system. *Nature* 560: 55–60

Yasuda T, Tsuzuki S, Kawazu M, Hayakawa F, Kojima S, Ueno T, Imoto N, Kohsaka S, Kunita A, Doi K, *et al* (2016) Recurrent DUX4 fusions in B cell acute

lymphoblastic leukemia of adolescents and young adults. *Nat Genet* 48: 569–574

Yeoh E-J, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling M V, Patel A, *et al* (2002) Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1: 133–143

Yoshimoto T, Tanaka M, Homme M, Yamazaki Y, Takazawa Y, Antonescu CR & Nakamura T (2017) CIC-DUX4 Induces Small Round Cell Sarcomas Distinct from Ewing Sarcoma. *Cancer Res* 77: 2927–2937

Yosuke Tanaka, Masahito Kawazu, Takahiko Yasuda, Miki Tamura, Fumihiko Hayakawa, Shinya Kojima, Toshihide Ueno, Hitoshi Kiyoi, Tomoki Naoe, Hiroyuki Mano, *et al* (2018) Transcriptional activities of DUX4 fusions in B-cell acute lymphoblastic leukemia. *Haematologica* 103: e522–e526

Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, Balog J, Tawil R, van der Maarel SM & Tapscott SJ (2013) DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis. *PLoS Genet* 9

Zaliova M, Kotrova M, Bresolin S, Stuchly J, Stary J, Hrusak O, Te Kronnie G, Trka J, Zuna J & Vaskova M (2017) ETV6/RUNX1-like acute lymphoblastic leukemia: A novel B-cell precursor leukemia subtype associated with the CD27/CD44 immunophenotype. *Genes Chromosomes Cancer* 56: 608–616

Zaliova M, Potuckova E, Hovorkova L, Musilova A, Winkowska L, Fiser K, Stuchly J, Mejstrikova E, Starkova J, Zuna J, *et al* (2019) ERG deletions in childhood acute lymphoblastic leukemia with DUX4 rearrangements are mostly polyclonal, prognostically relevant and their detection rate strongly depends on screening method sensitivity. *Haematologica* 104: 1407–1416

Zaliova M, Zimmermannova O, Dörge P, Eckert C, Möricke A, Zimmermann M, Stuchly J, Teigler-Schlegel A, Meissner B, Koehler R, *et al* (2014) ERG deletion is associated with CD2 and attenuates the negative impact of IKZF1 deletion in childhood acute lymphoblastic leukemia. *Leukemia* 28: 182–185

Zhang J, McCastlain K, Yoshihara H, Xu B, Chang Y, Churchman ML, Wu G, Li Y,

Wei L, Iacobucci I, *et al* (2016a) Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat Genet* 48: 1481–1489

Zhang Y, Lee JK, Toso EA, Lee JS, Choi SH, Slattery M, Aihara H & Kyba M (2016b) DNA-binding sequence specificity of DUX4. *Skelet Muscle* 6: 8

Zhou Y, Wai-Choi Tse E, Leung R, Cheung E, Li H & Sun H (2022) Multiplex Single-Cell Analysis of Cancer Cells Enables Unbiased Uncovering Subsets Associated with Cancer Relapse: Heterogeneity of Multidrug Resistance in Precursor B-ALL. *ChemMedChem* 17: e202100638