blood

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Focal deletions of a promoter tether activate the IRX3 oncogene in T cell acute lymphoblastic leukemia

Tracking no: BLD-2024-024300R1

Sunniyat Rahman (University College London, United Kingdom) Gianna Bloye (University College London, United Kingdom) Nadine Farah (University College London, United Kingdom) Jonas Demeulemeester (VIB - KU Leuven, Belgium) Joana R Costa (University College London, United Kingdom) David O'Connor (University College London, United Kingdom) Rachael Pocock (University College London, United Kingdom) Tanya Rapoz-D'Silva (University College London, United Kingdom) Adam Turna (Lister Hospital, United Kingdom) Lingyi Wang (,) Soo Wah Lee (University College London, United Kingdom) Adele Fielding (UCL Cancer Institute, United Kingdom) Juliette Roels (Genentech, United States) Roman Jaksik (Silesian University of Technology, Poland) Małgorzata Dawidowska (Institute of Human Genetics Polish Academy of Sciences, Poland) Pieter Van Vlierberghe (Ghent University, Belgium) Suzana Hadjur (University College London, United Kingdom) Jim Hughes (Oxford University, United Kingdom) James Davies (University of Oxford, United Kingdom) Alejandro Gutierrez (St. Jude Children's Research Hospital, United States) Michelle Kelliher (UMass Chan Medical School, United States) Peter Van Loo (The University of Texas MD Anderson Cancer Center, United States) Mark Dawson (Peter MacCallum Cancer Centre, Australia) Marc Mansour (University College London, United Kingdom)

Abstract:
Oncogenes can be activated in cis through multiple mechanisms including enhancer hijacking events and noncoding mutations that create enhancers or promoters de novo. These paradigms have helped parse somatic variation of noncoding cancer genomes, thereby providing a rationale to identify noncanonical mechanisms of gene activation. Here we describe a novel mechanism of oncogene activation whereby focal copy number loss of an intronic element within the FTO gene leads to aberrant expression of IRX3, an oncogene in T cell acute lymphoblastic leukemia (T-ALL). Loss of this CTCF bound element downstream to IRX3 (+224 kb) leads to enhancer hijack of an upstream developmentally active super-enhancer of the CRNDE long noncoding RNA (-644 kb). Unexpectedly, the CRNDE super-enhancer interacts with the IRX3 promoter with no transcriptional output until it is untethered from the FTO intronic site. We propose that 'promoter tethering' of oncogenes to inert regions of the genome is a previously unappreciated biological mechanism preventing tumorigenesis.

Conflict of interest: COI declared - see note

COI notes: J.D. and J.H. are a co-founders and share holders of Nucleome Therapeutics and provide consultancy to the company. J.D. has intellectual property licensed to BEAM therapeutics and he receives revenue from this licence and holds personal shares.

Preprint server: Yes; BioRxiv https://doi.org/10.1101/2024.02.06.579027

Author contributions and disclosures: S.R. and M.R.M. designed the study and wrote the manuscript. S.R., G.B., N.F., J.R.C., D.O'C., R.P., T.R., A.T. conducted wet-lab experiments. J.D., L.W., R.J., P.V.L. performed bioinformatic analysis. S.L. and A.K.F. were responsible for sample collection, processing and storage of UKALL2003 and UKALL14 samples. J.R., P.V.V., M.D., J.R.H., J.O.J.D., A.G., M.A.K., and M.A.D., were involved in data interpretation, provided additional samples and resources to complete this study. S.H. provided training for the UMI-4C technique used in this study. All authors revised and agreed to the final version of the manuscript.

Non-author contributions and disclosures: No;

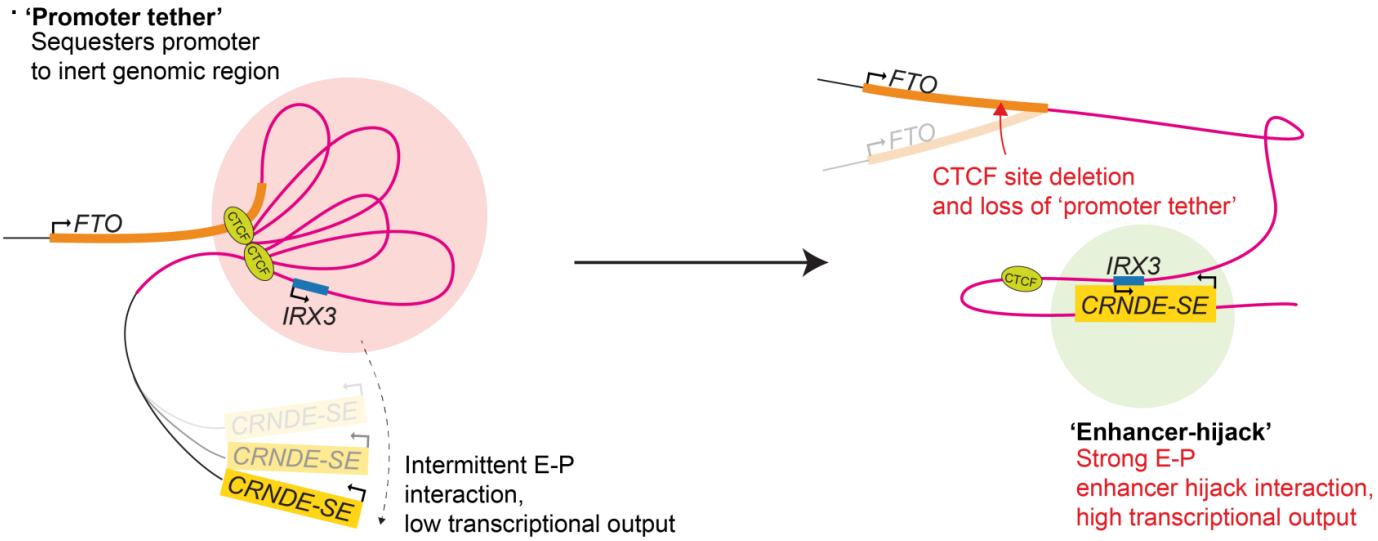
Agreement to Share Publication-Related Data and Data Sharing Statement: The authors will share data through e-mails to the corresponding author and public deposit where this has not been done previously. Data analysed in this study includes: HiChiP data analysed in this study- Kloetgen, A., Thandapani, P., Ntziachristos, P., Ghebrechristos, Y., Nomikou, S., Lazaris, C., Chen, X., Hu, H., Bakogianni, S., Wang, J., Fu, Y., Boccalatte, F., Zhong, H., Paietta, E., Trimarchi, T., Zhu, Y., Van Vlierberghe, P., Inghirami, G.G., Lionnet, T., Aifantis, I., Tsirigos, A., 2020. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. Nat. Genet. 52, 388-400. St Jude's T-ALL Cohort- Liu, Y., Easton, J., Shao, Y., Maciaszek, J., Wang, Z., Wilkinson, M.R., McCastlain, K., Edmonson, M., Pounds, S.B., Shi, L., Zhou, X., Ma, X., Sioson, E., Li, Y., Rusch, M., Gupta, P., Pei, D., Cheng, C., Smith, M.A., Auvil, J.G., Gerhard, D.S., Relling, M.V., Winick, N.J., Carroll, A.J., Heerema, N.A., Raetz, E., Devidas, M., Willman, C.L., Harvey, R.C., Carroll, W.L., Dunsmore, K.P., Winter, S.S., Wood, B.L., Sorrentino, B.P., Downing, J.R., Loh, M.L., Hunger, S.P., Zhang, J., Mullighan, C.G., 2017. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat. Genet. 373, 1541. CCLE dataset- Copy number analysis of cancer cell lines available at https://sites.broadinstitute.org/ccle/datasets

Clinical trial registration information (if any):

Focal Deletions of a Promoter Tether Activate the *IRX3* Oncogene in T Cell Acute Lymphoblastic Leukemia

Context of research: *IRX3* is an aberrantly expressed oncogene in T Cell Acute Lymphoblastic Leukemia with no known genetic mechanism of activation.

Main findings: Here we describe a novel mechanism of oncogene activation whereby focal deletion of a promoter tether situated within the *FTO* gene leads to aberrant expression of *IRX3* via hijack of a developmental super-enhancer.



Recurrent focal deletions of a CTCF binding site within *FTO* intron 8 occur in 1.5% of adult and 6.2% of pediatric T-ALL patients.

Focal deletions eliminate a tumor suppressor promoter tether to *IRX3* enabling enhancer hijack and transcriptional activation of *IRX3*.

Rahman et al

Focal deletions of a promoter tether activate the *IRX3* oncogene in T cell acute lymphoblastic leukemia

Sunniyat Rahman[1,2,3], Gianna Bloye[1], Nadine Farah[1], Jonas Demeulemeester[4], Joana R. Costa[1], David O'Connor[1], Rachael Pocock[1], Tanya Rapoz-D'Silva[1], Adam Turna[1], Lingyi Wang[5], SooWah Lee[1], Adele K. Fielding[6], Juliette Roels[7], Roman Jaksik[8], Malgorzata Dawidowska[9], Pieter Van Vlierberghe[7], Suzana Hadjur[1], Jim R. Hughes[10], James O.J. Davies[10], Alejandro Gutierrez[11], Michelle A Kelliher[12], Peter Van Loo[13], Mark A. Dawson[2,3], Marc R. Mansour[1, 5]

Affiliations

1. University College London Cancer Institute, Department of Haematology, London, UK.
2. Peter MacCallum Cancer Centre, Melbourne, Victoria 3000, Australia.
3. Sir Peter MacCallum Department of Oncology, The University of Melbourne, Victoria 3010, Australia.
4. VIB KU Leuven Centre for Cancer Biology, 3000 Leuven, Belgium.
5. Department of Developmental Biology and Cancer, GOS Institute of Child Health, UCL
6. University of York, Hull York Medical School, UK.
7. Ghent University, Department of Biomolecular Medicine, Ghent, Belgium.
8. Department of Systems Biology and Engineering and Biotechnology Centre, Silesian University of Technology, Gliwice, Poland
9. Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland.
10. University of Oxford, Department of Medicine, Medical Research Council Weatherall Institute of Molecular Medicine Centre for Computational Biology, Oxford, UK.
11. Dana-Farber/Harvard Cancer Centre, Boston, MA 02215, USA.
12. UMass Chan Medical School, MA 01605, USA
13. MD Anderson Cancer Center, The University of Texas, Department of Genetics, TX 77030, USA.

Correspondence (e-mail)
Dr. Sunniyat Rahman: sunniyat.rahman@ucl.ac.uk
Prof. Marc R. Mansour: m.mansour@ucl.ac.uk

The authors will share data through e-mails to the corresponding author and public deposit where this has not been done previously. Data analysed in this

study includes: HiChiP data analysed in this study- Kloetgen, A., Thandapani, P., Ntziachristos, P., Ghebrechristos, Y., Nomikou, S., Lazaris, C., Chen, X., Hu, H., Bakogianni, S., Wang, J., Fu, Y., Boccalatte, F., Zhong, H., Paietta, E., Trimarchi, T., Zhu, Y., Van Vlierberghe, P., Inghirami, G.G., Lionnet, T., Aifantis, I., Tsirigos, A., 2020. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. Nat. Genet. 52, 388-400. St Jude's T-ALL Cohort- Liu, Y., Easton, J., Shao, Y., Maciaszek, J., Wang, Z., Wilkinson, M.R., McCastlain, K., Edmonson, M., Pounds, S.B., Shi, L., Zhou, X., Ma, X., Sioson, E., Li, Y., Rusch, M., Gupta, P., Pei, D., Cheng, C., Smith, M.A., Auvil, J.G., Gerhard, D.S., Relling, M.V., Winick, N.J., Carroll, A.J., Heerema, N.A., Raetz, E., Devidas, M., Willman, C.L., Harvey, R.C., Carroll, W.L., Dunsmore, K.P., Winter, S.S., Wood, B.L., Sorrentino, B.P., Downing, J.R., Loh, M.L., Hunger, S.P., Zhang, J., Mullighan, C.G., 2017. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat. Genet. 373, 1541. CCLE dataset- Copy number analysis of cancer cell lines available at https://sites.broadinstitute.org/ccle/datasets

Key points

- Recurrent focal deletions of a CTCF binding site within *FTO* intron 8 occur in 1.5% of adult and 6.2% of paediatric T-ALL patients.
- Focal deletions eliminate a tumour suppressor promoter tether to *IRX3* enabling enhancer hijack and transcriptional activation of *IRX3*.

Abstract

Oncogenes can be activated in *cis* through multiple mechanisms including enhancer hijacking events and noncoding mutations that create enhancers or promoters *de novo*. These paradigms have helped parse somatic variation of noncoding cancer genomes, thereby providing a rationale to identify noncanonical mechanisms of gene activation. Here we describe a novel mechanism of oncogene activation whereby focal copy number loss of an intronic element within the *FTO* gene leads to aberrant expression of *IRX3*, an oncogene in T cell acute lymphoblastic leukemia (T-ALL). Loss of this CTCF bound element downstream to *IRX3* (+224 kb) leads to enhancer hijack of an upstream developmentally active super-enhancer of the *CRNDE* long noncoding RNA (-644 kb). Unexpectedly, the *CRNDE* super-enhancer interacts with the *IRX3* promoter with no transcriptional output until it is untethered from the *FTO* intronic site. We propose that 'promoter tethering' of oncogenes to inert regions of the genome is a previously unappreciated biological mechanism preventing tumorigenesis.

Introduction

The noncoding genome harbors differing classes of *cis*-reglatory elements including distal enhancers, poised promoters, and insulators that ensure precise control of gene expression across specialised tissues[1]. In cancer, somatically acquired mutations of the noncoding genome can transcriptionally activate oncogenes through indels that generate *de novo* enhancers and promoters[2–9], by focal amplification of long-range enhancers[10,11], through deletion of boundary elements [12–15], and structural rearrangements that lead to enhancer hijack[16–18].

Previously, first-in-class mechanisms of oncogene activation following somatic mutation of the noncoding genome were discovered in T cell acute lymphoblastic leukaemia (T-ALL), including *cis*-acting mutations that create a neomorphic enhancers, and recurrent focal deletions that disrupt boundaries between insulated neighbourhoods to activate *TAL1* and *LMO2* oncogenes respectively[5,12]. Although most T-ALL oncogenes *TAL1*, *LMO2*, *TLX1*, *TLX3*, *NKX2-1* and *MYB* are ectopically expressed by well-characterized mechanisms, some have no known genetic lesion, and thus provide an opportunity for the discovery of novel mechanisms of oncogene activation[19,20].

In this study, we identify recurrent deletions of *FTO* intron 8 (*FTO*^{int8del}) in a subgroup of T-ALL patients with aberrant expression of *IRX3*, a putative oncogene in T-ALL with no known genetic driver[21]. Mechanistically, we show that *IRX3* is normally tethered to a CTCF site within *FTO* intron 8, a transcriptionally inert region with minimal transcriptional output. Deletion of *FTO* intron 8 releases this 'tether', enabling *IRX3* to be hijacked by a distal highly active developmental super-enhancer of *CRNDE,* resulting in *IRX3* transcription. We posit that 'promoter tethering' to inert regions of the genome is a previously unappreciated

tumor suppressor mechanism, ensuring proto-oncogenes remain protected from activation by distal developmental super-enhancers.

Methods

Detailed methods are described in the supplemental information section. Recurrent *FTO* intron 8 deletions were discovered by analysis of St. Jude (Liu et al., 2017), UKALL2003 and ICGZ Poznan T-ALL primary patient cohorts. *IRX3* expression data from available patient samples were determined by RNA-seq. Frequency of *FTO* intron 8 CTCF and MYB site deletions were identified by ddPCR on an unselected cohort of T-ALL patient samples. CRISPR/Cas9 was used to make *FTO* intron 8 CTCF site and MYB site deletions in PF-382 cells. *IRX3* expression was determined by qPCR. CTCF knockdown was achieved by electroporation of siRNA pools and knockdown was validated by qPCR. Looping interactions from *FTO* intron 8 CTCF site and *IRX3* promoter were identified by UMI-4C. Enhancer-promoter interactions in ALL-SIL cells were identified by HiChIP with immunoprecipitation for H3K27ac. ATAC-seq was used to characterise the *CRNDE* super-enhancer in *FTO* wild-type and *FTO* intron 8 CTCF site deleted cells. CRISPR/Cas9 was used to delete the *CRNDE* super-enhancer in ALL-SIL cells. All CRISPR/Cas9 edits were confirmed by flanking PCR. Public datasets used are described throughout the manuscript and referenced accordingly.

Ethical approval for UKALL2003 obtained from Scottish Multi-Centre Research Ethics Committee on 25/02/2003, ref: 02/10/952. Ethical approval for UKALL14 obtained from London-Fulham Research Ethics Committee ref: 09/H0711/90. All samples were collected from patients with informed consent according to the Declaration of Helsinki.

Results and Discussion

IRX3 is aberrantly expressed in T-ALL with enhancer-promoter contacts to neighbouring genes _FTO_, _CRNDE_ and _IRX5_

We hypothesized that genes aberrantly expressed in T-ALL compared to their developmentally-matched normal cellular counterpart might uncover previously unrecognized oncogenes, enabling us to explore novel mechanisms of oncogene activation. We thus generated a list of genes that were not expressed in normal thymic subsets (FPKM <0.125; n=2,468) and ranked their mean expression in 264 childhood cases of T-ALL (Fig. 1a)[20,22]. Within the top 50 aberrantly expressed gene list were well-characterized T-ALL oncogenes such as *TLX3*, *TLX1* and *NKX3-2*, validating this explorative approach (Fig. 1b, Supplementary Table 1). Our attention was drawn to *IRX3* (rank 20), given recent reports licencing it as an oncogene through its ability to immortalise HSCs and induce T-lymphoid leukemias *in vivo*[21]. *IRX3* encodes for an Iroquois-family homeobox transcription factor essential for limb bud pattern formation, nephron segmentation, and cardiac function, but with no known role in normal hematopoiesis[23–25].

In contrast to the developmental oncogenes *TAL1* and *LMO2*, *IRX3* is not expressed in any normal T-cell precursor (Fig. 1c). Further analysis of the pediatric T-ALL cohort separated according to their class-defining oncogenic subtypes showed that a greater proportion of the patients in the TLX3 and HOXA subgroups had aberrant *IRX3* expression (defined as FPKM >1) when compared to the other subgroups (86% vs 23%; Fisher exact test statistic <0.00001) (Fig. 1d).[20] Overall, aberrant *IRX3* expression is observed in 49% (22/45) of adults and 42% (111/264) of paediatric patients with T-ALL (Supplementary Fig. 1a). IRX3 positive patients had a higher incidence of NOTCH1 mutations (88% vs 65%; P<0.0001), CTCF mutations (12% vs 2%; P<0.001), PHF6 mutations (38% vs 14%; P<0.0001), JAK-STAT pathway mutations (37% vs 16%; P<0.0001), NOTCH-MYC enhancer amplification (21% vs 5%;

P<0.0001), but a lower incidence of PI3K pathway mutations (15% vs 39%; P<0.0001; Supplementary Fig. 2; Supplementary Table 2). Furthermore, 74% (17/23) of T-ALL cell lines exhibited aberrant *IRX3* expression (Fig. 1e, Supplementary Fig. 1b).

Analysis of published HiC data from normal human thymic tissue identified *IRX3* within a single TAD shared with *FTO*, *IRX5* and *CRNDE* encompassing ~1.3 Mb (Supplementary Fig. 3, 4). To identify enhancer-promoter loops from *IRX3*, we next examined HiChIP data from the IRX3 positive CUTLL-1 T-ALL cell line and identified two candidate intra-TAD *cis*-regulatory elements, which we named cCRE_1 (within *FTO*) and cCRE_2 (at *CRNDE*/*IRX5*) (Fig. 1f). Additionally, generation of a rank ordered gene list by comparing IRX3-negative and IRX3-positive primary T-ALL samples (n=118) by RNA-seq, identified the expression of multiple genes positively correlated with *IRX3* mRNA levels including *FTO* at rank 64 and *IRX5* at rank 109 out of 19,464 total genes (Fig. 1g), suggesting co-ordinated long-range intra-TAD interactions.

FTO intron 8 is recurrently deleted in patients with T-ALL

We next explored whether copy number aberrations (CNAs) affected the *IRX3* CREs by examining copy number calls from published datasets of primary T-ALL patient samples and T-ALL cell lines. This analysis revealed 13 T-ALL genomes (12 of patient origin and 1 cell line – ALL-SIL) with heterozygous copy number losses impinging on the *FTO* gene, and notably all intersected with cCRE_1 (Fig. 2a, Supplementary Table 3). Expression data was available for 6/13 T-ALL samples and all exhibited aberrant expression of *IRX3* mRNA (Fig. 2b). Due to relatively small size of *IRX3* coding sequences (~2.6 kb), the gene often lacks informative heterozygous SNPs to make consistent allele specific expression (ASE) calls. However, we were able to confirm ASE in one patient sample, while ALL-SIL had evidence

of promoter methylation allelic imbalance, a proxy for ASE (Supplementary Fig. 5), indicative of a heterozygous *cis*-acting genetic lesion (Fig. 2b).

We next analyzed potential regulatory elements within the minimally deleted region (~155 kb) of *FTO* intron 8 by examining ChIP-Seq datasets from the *FTO* wild-type Jurkat T-ALL cell line (Fig. 2c). This was largely devoid of high-amplitude ChIP-Seq peaks, except for a single CTCF binding peak, and a single MYB peak enriched with H3K27ac. From this we hypothesized that loss of CTCF and/or MYB binding within *FTO* intron 8 may be involved in dysregulated *IRX3* expression (Fig. 2c). To explore this hypothesis further, we developed a digital droplet PCR (ddPCR) assay capable of distinguishing CNA at the CTCF and MYB binding sites (Supplementary Fig. 6). This had two aims, first to ascertain the frequency of *FTO* intron 8 copy number aberrations (CNAs) in a larger cohort of primary T-ALL samples, and secondly to allow for independent copy number calls at both loci where differences between the calls may provide mechanistic insight. Among 298 unselected primary T-ALL samples collected at diagnosis, CNAs within *FTO* intron 8 were more common in pediatric than adult T-ALL (10/161, 6.2% versus 2/137, 1.4% respectively; $p = 0.04$ by Chi-Squared test, Fig. 2d). While 8/12 had copy number loss at both the CTCF and MYB sites, notably 4/12 had heterozygous copy number loss of the CTCF site alone, suggesting that loss of this CTCF site is most likely to be functionally relevant. Furthemore, the three patients with this lesion and available RNA-seq exhibited aberrant *IRX3* expression (Fig. 2d). The average fractional abundance of these heterozygous mutations was 35% across the 12 patient samples, suggesting these mutations are clonal and present in >70% T-ALL blasts (Supplementary Fig. 7).

CRISPR/Cas9 mediated disruption of *FTO* intron 8 CTCF site transcriptionally activates *IRX3*

To functionally validate whether loss of the CTCF and/or MYB site was capable of upregulating *IRX3* expression, we utilized CRISPR/Cas9 editing in the *FTO* wild-type, IRX3 negative PF-382 T-ALL cell line (Fig. 2e; Supplementary Fig. 8, Supplementary Table. 4). Single cell clones with deletions of ~12kb, impinging on both CTCF and MYB binding sites, thus mimicking the copy number losses observed in the primary patient samples, led to significant upregulation of *IRX3* to levels comparible to $FTO^{int8del}$ ALL-SIL cells (Fig. 2f; Supplementary Fig. 9). Crucially, similar upregulation was observed in clones with sole disruptive indels of the CTCF binding site, but not those affecting the MYB site alone. Furthermore, deletion of the CTCF binding site in a polyclonal population led to a significant increase of *IRX3* mRNA relative to unedited controls (Fig. 2g). These data strongly implicate loss of the CTCF site as the key drivers of aberrant *IRX3* expression in patients with $FTO^{int8del}$.

Given many *IRX3* positive T-ALL samples do not harbor $FTO^{int8del}$ (106/109 with FPKM>1, St. Jude cohort), we postulated that mutations of *CTCF* itself, which are recurrent in T-ALL genomes may create the same phenotype. To explore this, we compared RNA-seq data from CTCF wild-type and CTCF mutant T-ALL samples and found significantly higher *IRX3* expression in the CTCF mutant group (Fig. 2h, Supplementary Table 5). To address this functionally, we knocked down CTCF with siRNAs in a panel of T-ALL cell lines and observed significant *IRX3* upregulation in PF-382, PEER, Loucy and HPB-ALL cells but not other cell lines (Fig. 2i, Supplementary Fig.10), suggesting the CTCF-IRX3 axis is influenced by variations in the genomic context.

Focal deletion of the *IRX3* 'promoter tether' within *FTO* intron 8 enables enhancer hijack by the *CRNDE* developmental super-enhancer

Given the importance of CTCF as an architectural protein, we quantified interactions from the *FTO* intron 8 CTCF site and *IRX3* promoter by UMI-4C. Baiting the *FTO* intron 8 CTCF site in PF-382 (IRX3 negative) cells revealed a dense cluster of looping interactions between the CTCF site and *IRX3*, suggesting that the proximal promoter of *IRX3* is tethered to the *FTO* intron 8 in the wild-type setting (Fig. 3a, Supplementary Fig. 11). By comparing ALL-SIL cells (harboring $FTO^{int8del}$) and PF-382 cells (IRX3 negative, FTO^{wt}), we observed a marked increase in the number of interactions between the *IRX3* promoter bait and the *CRNDE* locus in ALL-SIL cells, previously identified as cCRE_2 (Fig. 3a,b). The same was observed in PF-382 clones with CRISPR/Cas9 induced disruption of the CTCF binding site in *FTO* intron 8, providing strong evidence for causality (Supplementary Fig. 12). Interestingly, *CRNDE* is expressed at high levels (FPKM>20) in developing CD3- double positive (DP) thymocytes and downregulated through normal T cell differentiation (Fig. 3c). HiChIP for H3K27ac in ALL-SIL ($FTO^{int8del}$) also showed enhancer loops originating from the *CRNDE/IRX5* locus to *IRX3* (Fig. 3d). Further examination of the *CRNDE* locus by H3K27ac ChIP-Seq in PF-382 cells classified this locus as a super-enhancer (113^{th} out of 23,737 total enhancers by ROSE), with broad H3K27ac marks covering the *IRX5* and *CRNDE* promoters and gene bodies (Fig. 3e,f). The *CRNDE* enhancer does not appear to be regulated by the *FTO* intron 8 locus, since PF-382 cells harboring CTCF binding-site deletions did not alter the ATAC-seq signal intensity (Fig. 3f). In contrast, disruption of the *CRNDE* enhancer in ALL-SIL ($FTO^{int8del}$) cells using CRISPR/Cas9 resulted in significant downregulation of *IRX3* expression (p-value = 0.01; Fig 3g, Supplementary Fig. 13), consistent with the hypothesis aberrant expression of *IRX3* is driven by increased interaction with the *CRNDE* super-enhancer.

In this study, we discovered recurrent focal deletions of *FTO* that explains a subgroup of T-ALL patients with aberrant *IRX3* expression. Although CTCF mutations may have the same consequence, the mechanism for aberrant *IRX3* expression in other T-ALL patients currently remains unclear. From a clinical perspective, it is also uncertain if targeting IRX3, or its pathway, offers a feasible therapeutic approach, or if the *FTO* deletions can be used to track minimal residual disease. Although situated downstream to *IRX3*, deletion of this long-range insulator counterintuitively permits enhancer hijack of an upstream super-enhancer. This is distinct from canonical TAD fusion events in cancer whereby focal deletions or methylation disrupts boundary elements positioned between the oncogene and *cis*-regulatory effector [12,13,15,26]. In contrast to previously discovered enhancer hijack events, where the enhancer-promoter (E-P) interaction remains naïve until the structural rearrangement occurs, we reveal interactions between the *IRX3* promoter and *CRNDE* super-enhancer in IRX3 negative cells [16–18].We posit that the *IRX3* promoter is sequestered to a relatively inert region of *FTO* intron 8, yielding minimal transcriptional output, despite residual interactions with the *CRNDE* super-enhancer (Fig. 3h). This sequestration is facilitated by CTCF binding at the *FTO* intron leading to the formation of a 'promoter tether'. Therefore, loss of this CTCF site by focal deletion untethers the *IRX3* promoter, providing an example of enhancer-promoter competition occurring within the same TAD (Fig. 3h). This is distinct from recent findings in acute myeloid leukemia (AML) describing an intronic lncRNA in *FTO* that regulates *IRX3* expression[27]. Together, these findings add to the complex regulatory relationship between the *FTO* and *IRX3* genes first identified through the discovery of obesity-associated germline variants[28,29]. We further speculate that promoter tethering to inert regions of the genome is a previously unappreciated tumor suppressor mechanism through which potent oncogenes are protected from activation of nearby developmental super-enhancers. Integrating 3-

dimensional promoter interactions with copy number data may highlight further examples of this phenomenon, potentially explaining the functional consequence of recurrent focal deletions in noncoding genomes of cancers that as yet remain unexplained.

Acknowledgements

The authors thank the patients, families, and clinical teams who have been involved in all trials. Primary childhood leukemia samples used in this study were provided by VIVO Biobank, supported by Cancer Research UK & Blood Cancer UK (Grant no. CRCPSC-Dec21\100003). Samples were acquired with support from laboratory teams in the Bristol Genetics Laboratory, Southmead Hospital, Bristol, United Kingdom; Molecular Biology Laboratory, Royal Hospital for Sick Children, Glasgow, United Kingdom; Molecular Haematology Laboratory, Royal London Hospital, London, United Kingdom; and Molecular Genetics Service and Sheffield Children's Hospital, Sheffield, United Kingdom. This work was supported in part by Cancer Research UK CRUK/A13920 to A.K.F for UKALL14 trial and CRUK/A21019 to A.K.F for UKALL14 Biobank. M.R.M. is supported through a GOSH Children's Charity Professorship. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (CC2008), the UK Medical Research Council (CC2008), and the Wellcome Trust (CC2008). For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. P.V.L. is a CPRIT Scholar in Cancer Research and acknowledges CPRIT grant support (RR210006). M.D and P.V.V were supported by the European Union's Horizon 2020 research and innovation program under Grant agreement no. 952304. S.R. was supported by a John Goldman Fellowship from Leukaemia UK. This project was supported by funding from The National Centre for Research and Development: STRATEGMED3/304586/5/NCBR/2017. S.R. would like to thank A. Motazedian, A. Das, E. Wainwright, D. Vassiliadis and J. Balic for academic discussions pertaining to this study and Lorna Neal for invaluable non-academic discussions.

Authorship contributions

S.R. and M.R.M. designed the study and wrote the manuscript. S.R., G.B., N.F., J.R.C., D.O'C., R.P., T.R., A.T. conducted wet-lab experiments. J.D., L.W., R.J., P.V.L. performed bioinformatic analysis. S.L. and A.K.F. were responsible for sample collection, processing and storage of UKALL2003 and UKALL14 samples. J.R., P.V.V., M.D., J.R.H., J.O.J.D., A.G., M.A.K., and M.A.D., were involved in data interpretation, provided additional samples and resources to complete this study. S.H. provided training for the UMI-4C technique used in this study. All authors revised and agreed to the final version of the manuscript.

Disclosure of Conflicts of Interest

J.D. and J.H. are a co-founders and share holders of Nucleome Therapeutics and provide consultancy to the company. J.D. has intellectual property licensed to BEAM therapeutics and he receives revenue from this licence and holds personal shares.

References

1. ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
2. Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
3. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
4. Vinagre, J. *et al.* Frequency of TERT promoter mutations in human cancers. *Nat. Commun.* **4**, 2185 (2013).
5. Mansour, M. R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
6. Abraham, B. J. *et al.* Small genomic insertions form enhancers that misregulate oncogenes. *Nat. Commun.* **8**, 14385 (2017).
7. Rahman, S. *et al.* Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. *Blood* **129**, 3221–3226 (2017).
8. Li, Z. *et al.* APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. *Leukemia* **31**, 2057–2064 (2017).
9. Rheinbay, E. *et al.* Recurrent and functional regulatory mutations in breast cancer. *Nature* **547**, 55–60 (2017).
10. Herranz, D. *et al.* A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137 (2014).

11. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).
12. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
13. Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).
14. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
15. Yang, M. *et al.* 13q12.2 deletions in acute lymphoblastic leukemia lead to upregulation of FLT3 through enhancer hijacking. *Blood* **136**, 946–956 (2020).
16. Montefiori, L. E. *et al.* Enhancer Hijacking Drives Oncogenic BCL11B Expression in Lineage-Ambiguous Stem Cell Leukemia. *Cancer Discov.* **11**, 2846–2867 (2021).
17. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369–381 (2014).
18. Drier, Y. *et al.* An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nat. Genet.* **48**, 265–272 (2016).
19. Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).
20. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **373**, 1541 (2017).
21. Somerville, T. D. D. *et al.* Derepression of the Iroquois Homeodomain Transcription Factor Gene IRX3 Confers Differentiation Block in Acute Leukemia. *Cell Rep.* **22**, 638–652 (2018).

22. Casero, D. *et al.* Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat. Immunol.* **16**, 1282–1291 (2015).
23. Tao, H. *et al.* IRX3/5 regulate mitotic chromatid segregation and limb bud shape. *Development* **147**, (2020).
24. Koizumi, A. *et al.* Genetic defects in a His-Purkinje system transcription factor, IRX3, cause lethal cardiac arrhythmias. *Eur. Heart J.* **37**, 1469–1475 (2016).
25. Reggiani, L., Raciti, D., Airik, R., Kispert, A. & Brändli, A. W. The prepattern transcription factor Irx3 directs nephron segment identity. *Genes Dev.* **21**, 2358–2370 (2007).
26. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
27. Camera, F. *et al.* Differentiation block in acute myeloid leukemia regulated by intronic sequences of FTO. *iScience* **26**, 107319 (2023).
28. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
29. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).

Figure legends

Fig. 1: *IRX3* is aberrantly expressed in T-ALL and resides in a shared topologically associated domain with enhancer-promoter contacts to neighbouring genes *FTO*, *CRNDE* and *IRX5*

a, Schematic outlining the comparative analysis conducted on bulk RNA-seq from developing T cell subsets (NCBI GEO accession #GSE69239) and bulk RNA-seq from St. Jude's paediatric T-ALL cohort (Liu et al., n=264) to identify aberrantly expressed genes. **b**, Box and whisker plot showing expression of top 50 genes aberrantly expressed in the St. Jude's paediatric T-ALL cohort (n=264) when compared to normal hematopoietic progenitors (NCBI GEO accession #GSE69239); haematopoetic stem cells (HSCs), lymphoid primed multipotent progenitors (LMPPs), common lymphoid progenitors (CLPs) and T cell subsets (Thy1-6). Expression values for T cell subsets are from 2 replicates for each population and expression is FPKM avareaged for each cell type. T-ALL cohort genes are ranked along the x-axis by mean expression. **c**, Line graph tracking *IRX3*, *LMO2*, and *TAL1* expression by RNA-seq across hematopoietic and thymic progenitors. RNA-Seq from NCBI GEO accession #GSE69239, with the following immunophenotypic definitions: from BM $CD34^+$ cells, $CD34^+CD38^-Lin^-$ (HSCs), $CD34^+CD45RA^+CD38^+CD10^-CD62L^{hi}Lin^-$ (LMPPs), $CD34^+CD38^+CD10^+CD45RA^+Lin^-$ (CLPs); from thymic CD34+ cells, $CD34^+CD7^-CD1a^-CD4^-CD8^-$ (Thy1), $CD34^+CD7^+CD1a^-CD4^-CD8^-$ (Thy2), and $CD34^+CD7^+CD1a^+CD4^-CD8^-$ (Thy3); from thymic $CD34^-$ cells, $CD4^+CD8^+$ (Thy4), $CD3^+CD4^+CD8^-$ (Thy5), and $CD3^+CD4^-CD8^+$ (Thy6). **d**, Violin plot showing *IRX3* expression (FPKM) by RNA-seq from the St. Jude primary T-ALL cohort separated by class defining oncogenic subtypes (n=264). **e**, Bar chart showing *IRX3* expression (FPKM) by RNA-seq from T-ALL cell lines (n=24) and labelled by oncogenic subtype. **f,** Enhancer-

promoter interactions about the *IRX3* locus mapped by HiChIP following pull-down for H3K27ac from the IRX3 positive CUTTL1 T-ALL cell line (NCBI GEO accession #GSE115896) and ChIP-seq for H3K27ac in CUTTL1 cells. Loops between IRX3 and candidate cis-regulatory elements (cCREs) for *IRX3* are highlighted in red and indicated with arrows. **g,** Ranked gene list by comparing IRX3 positive (top, n=59) versus IRX3 negative (bottom, n=59) T-ALL samples by RNA-seq from the St. Jude cohort. The y-axis ranking score metric for each gene was calculated by the GSEA 'Signal2Noise' computational method for categorical phenotypes. Genes are listed along the x-axis in order of the ranked score.

Fig. 2: Recurrent deletions of *FTO* intron 8 in T-ALL patients impinge on a CTCF binding site

a, Focal heterozygous deletions identified in T-ALL genomes at *FTO*. 12 originate from primary patient samples and 1 from the ALL-SIL T-ALL cell line. Deletions identified in patient samples are from multiple cohorts including St. Jude (n=264), UKALL2003(n=148), and ICGZ Poznan (n=63) **b,** *IRX3* expression (FPKM) by RNA-seq from matched primary patient samples and the ALL-SIL T-ALL cell line harbouring *FTO* intron 8 deletions. Allele-specific expression identified in patient 11 was determined by St. Jude's prior analyses, and unbalanced promoter methylation for the ALL-SIL cell line was ascertained by analysis of the CCLE Promoter Methylation dataset. **c,** ChIP-Seq for CTCF (NCBI GEO accession #GSE68976), MYB and H3K27ac (NCBI GEO accession #GSE76783) in the Jurkat T-ALL cell line centred on the minimally deleted region within *FTO* intron 8. **d,** Pie charts showing the frequency of *FTO* intron 8 deletions determined by ddPCR in adult (n=137) and paediatric (n=161) T-ALL cohorts. Stacked boxes summarise CTCF or MYB site-specific copy number calls for each patient identified with *FTO* intron 8 deletions and a bar chart

showing *IRX3* expression (FPKM) is shown for patient samples where matched RNA was available for sequencing. Matched *IRX3* expression by RNA-seq for samples that exhibited normal copy number (CN normal, n=9, UKALL14 cohort) are also plotted. **e,** Experimental outline for CRISPR/Cas9-mediated disruption of *FTO* intron 8 CTCF and MYB sites in the PF-382 (FTO^{wt}) and IRX3 negative (FPKM<1) T-ALL cell line. **f,** Bar chart showing *IRX3* expression determined by qPCR for PF-382 (FTO^{wt}) and ALL-SIL ($FTO^{int8del}$) T-ALL cell lines and unedited (wt) and edited (mut) clones. Data presented are 3 technical replicates ± SD for each biological sample. **g,** Bar chart showing *IRX3* expression determined by qPCR for PF-382 polyclonal edited cells following CRISPR/Cas9 mediated disruption of the *FTO* intron 8 CTCF site. Technical replicates from 3 independent experiments are shown. $p<0.0001$ from a two-tailed T test. **h,** Violin plot showing *IRX3* expression (FPKM) of primary T-ALL samples with (n=16) and without (n=248) CTCF mutations from the St. Jude T-ALL cohort. $p=0.0007$ from two-tailed T test. **i**, Bar chart showing *CTCF* and *IRX3* expression by qPCR following CTCF knockdown in PF-382, PEER, Loucy and HPB-ALL T-ALL cell lines. Final CTCF siRNA pool concentrations used are shown on the x-axis, where NT is a negative control non-targeting siRNA pool. Technical replicates from 2 independent experiments are shown. Statistical comparisons were made to NT groups by a two-tailed T test where $p<0.05$ (*), $p<0.01$(**), $p<0.001$(***) and ns = not significant.

Fig. 3 Transcriptional activation of *IRX3* in T-ALL by loss of a promoter tether and enhancer hijack from the *CRNDE* locus

a, (Top panel) UMI-4C contact profile generated by baiting the *FTO* intron 8 CTCF site in PF382 (FTO^{wt}) cells. A green bar highlights the contacts that form a promoter tether between *IRX3* and *FTO* intron 8. (Bottom panel) UMI-4C contact profile generated by baiting the *IRX3* proximal promoter in the PF382 (FTO^{wt}) and ALL-SIL ($FTO^{int8del}$) T-ALL cell line.

CTCF/CTCFL binding sites across the *FTO*, *IRX3* and *CRNDE*/*IRX5* locus were *in silico* predicted by HOCOMOCO v11. **b,** UMI count of interactions between the *IRX3* proximal promoter and the *CRNDE* locus (p = 0.0005; Fisher's exact test; UMI4Cats analysis package). **c**, Expression (FPKM) of *CRNDE*, *IRX5* and *IRX3* across T cell subsets by RNA-seq from the Blueprint Epigenome Project, sample TH91. **d,** ChIP-seq for H3K27 acetylation (NCBI GEO accession #GSM1816978) and enhancer-promoter loops generated by HiChIP for H3K27 acetylation in the ALL-SIL ($FTO^{int8del}$) T-ALL cell line about the *FTO*, *IRX3* and *CRNDE/IRX5* locus. Loops shown passed a q-value threshold of 0.01 at a bin size of 40 kb and were called by FitHiChIP. **e**, Rank ordering of super-enhancers (ROSE analysis) in the PF-382 (FTO^{wt}) T-ALL cell line by using H3K27ac ChIP-Seq (NCBI GEO accession #GSE76783). Red points indicate genomic regions classed as super-enhancers which includes the *CRNDE* locus. **f**, ATAC-seq at the *CRNDE*/*IRX5* locus in PF-382 (FTO^{wt}) and PF-382 (*FTO* intron 8 CTCF site deleted) T-ALL cell lines, in addition to ChIP-seq for MYB (NCBI GEO accession #GSM2466687) and H3K27ac (NCBI GEO accession #GSM2037796) in PF-382 (FTO^{wt}) T-ALL cells. Shaded box represents peak region targeted to disrupt the *CRNDE* super-enhancer by CRISPR/Cas9 excision. **g**, Expression level of *IRX3* mRNA as determined by qPCR from ALL-SIL ($FTO^{int8del}$) T-ALL cell line following CRISPR/Cas9 mediated deletion of the *CRNDE* super-enhancer locus. **h**, Proposed mechanism of action whereby promoter tethering of *IRX3* to the relatively inert *FTO* intron 8 locus by CTCF binding allows infrequent enhancer-promter interactions and low transcriptional output. Subsequent focal deletion of this intronic CTCF site leads to loss of the promoter tether and enhancer hijack of the *CRNDE* super-enhancer.

Figure 1

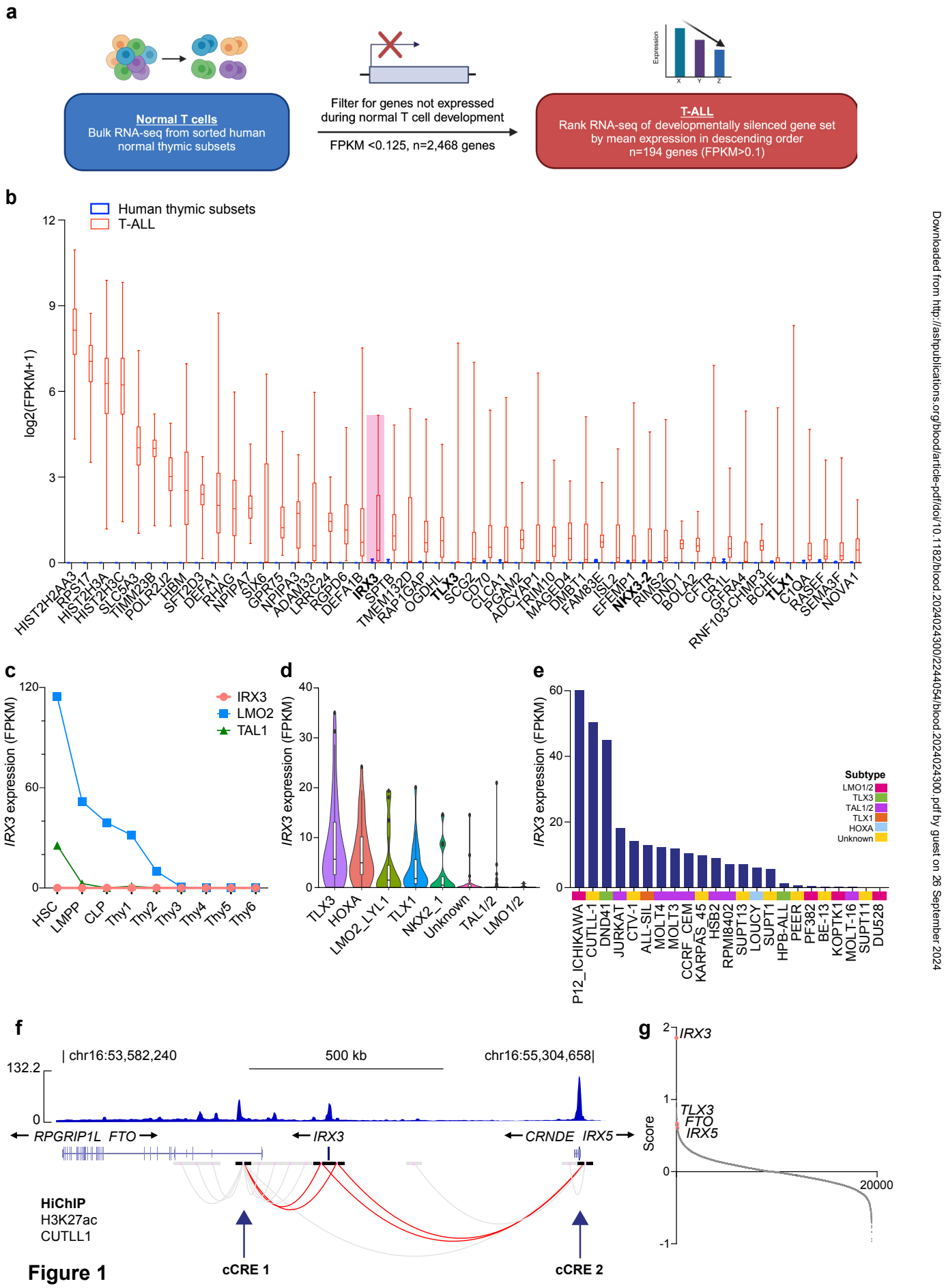


Figure 1

Figure 2

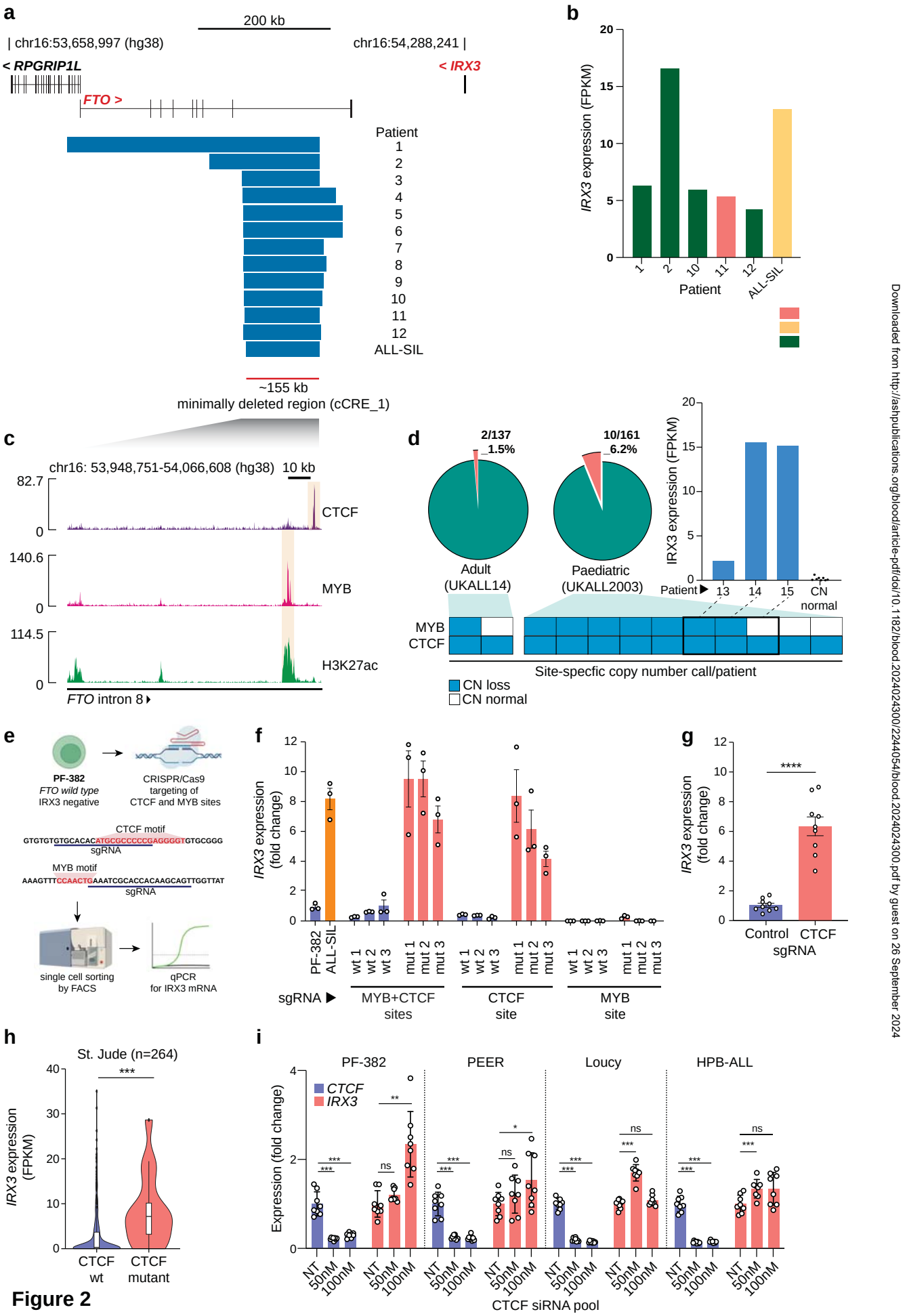


Figure 2

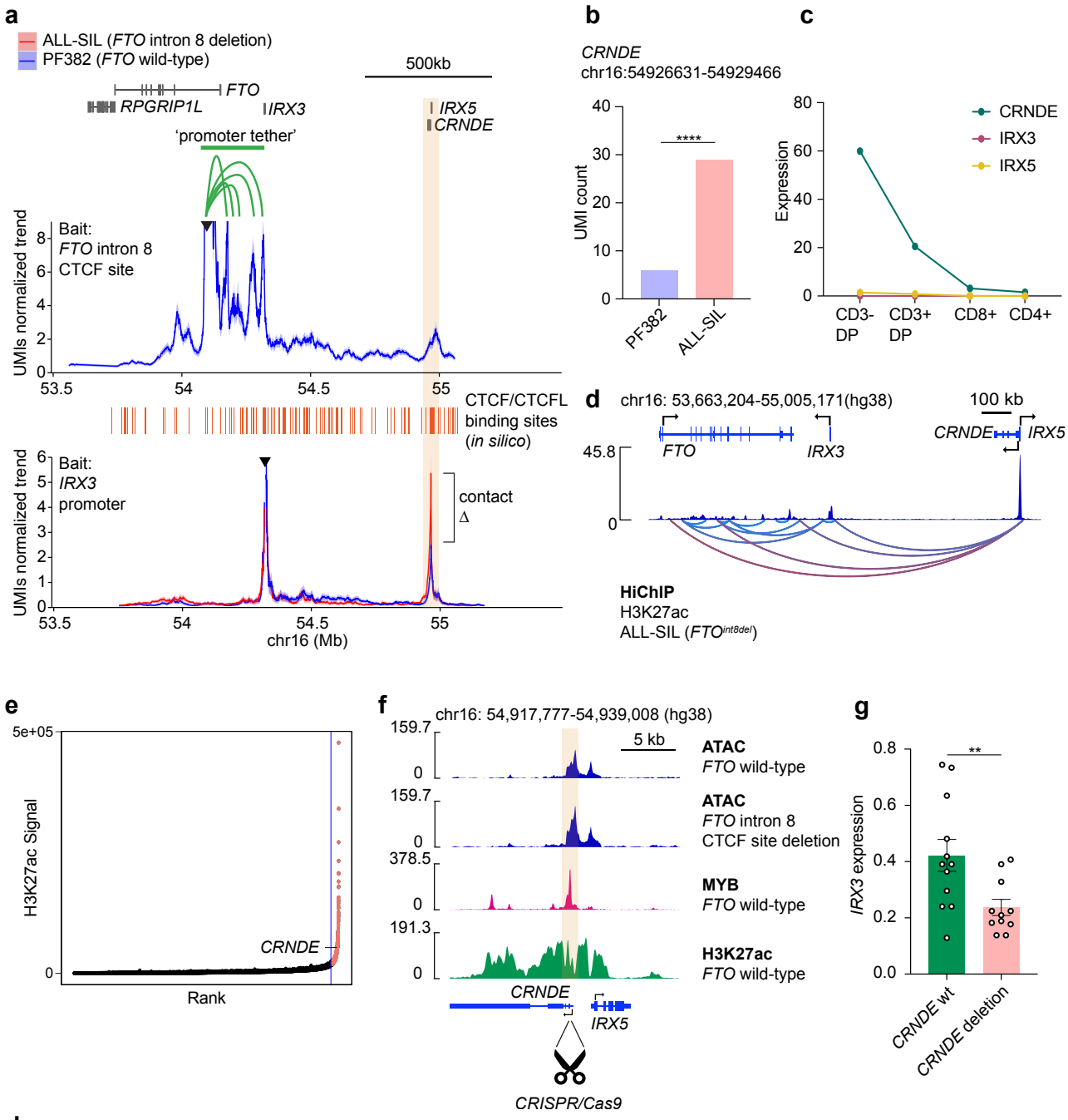


Figure 3