

## ARTICLE OPEN



## STEM CELL BIOLOGY

# A primary hierarchically organized patient-derived model enables in depth interrogation of stemness driven by the coding and non-coding genome

Hélène Boutzen<sup>1</sup>, Seyed Ali Madani Tonekaboni<sup>2</sup>, Michelle Chan-Seng-Yue <sup>1,3</sup>, Alex Murison<sup>1</sup>, Naoya Takayama<sup>1,4</sup>, Nathan Mbong<sup>1</sup>, Elvin Wagenblast <sup>1</sup>, Elias Orouji<sup>5</sup>, Andrea Arruda<sup>1,6</sup>, Amanda Mitchell<sup>1</sup>, Faiyaz Notta<sup>1,3</sup>, Mark D. Minden<sup>1,6,7</sup>, Mathieu Lupien<sup>1,2,8</sup>, Kerstin B. Kaufmann<sup>1</sup> and John E. Dick <sup>1,9</sup>✉

© The Author(s) 2022

Many cancers are organized as cellular hierarchies sustained by cancer stem cells (CSC), whose eradication is crucial for achieving long-term remission. Difficulties to isolate and undertake in vitro and in vivo experimental studies of rare CSC under conditions that preserve their original properties currently constitute a bottleneck for identifying molecular mechanisms involving coding and non-coding genomic regions that govern stemness. We focussed on acute myeloid leukemia (AML) as a paradigm of the CSC model and developed a patient-derived system termed OCI-AML22 that recapitulates the cellular hierarchy driven by leukemia stem cells (LSC). Through classical flow sorting and functional analyses, we established that a single phenotypic population is highly enriched for LSC. The LSC fraction can be easily isolated and serially expanded in culture or in xenografts while faithfully recapitulating functional, transcriptional and epigenetic features of primary LSCs. A novel non-coding regulatory element was identified with a new computational approach using functionally validated primary AML LSC fractions and its role in LSC stemness validated through efficient CRISPR editing using methods optimized for OCI-AML22 LSC. Collectively, OCI-AML22 constitutes a valuable resource to uncover mechanisms governing CSC driven malignancies.

*Leukemia* (2022) 36:2690–2704; <https://doi.org/10.1038/s41375-022-01697-9>

## INTRODUCTION

Acute myeloid leukemia is a heterogeneous disease [1–6] driven by leukemic stem cells (LSCs) [7–16]. Some AML patients achieve durable remission, but the majority relapse and die of their disease within 2 years. There is strong evidence that relapse arises from LSCs capable of surviving chemotherapy and initiating relapse [7–16]. A better understanding of mechanisms fueling stemness and chemoresistance in LSC is required to design efficient therapies. LSCs from across the spectrum of AML patients are heterogeneous in frequency, immunophenotype and genetic profile [15]. Despite this heterogeneity, all LSCs share the same functional stem cell hallmark that fosters long-term disease propagation: the capacity for self-renewal. The functional properties of LSC are directly linked to relapse [15] and poor prognosis [8, 9] establishing their clinical relevance. Numerous studies have established that the intrinsic stemness properties of LSCs from across diverse AML cohorts can be captured using gene

expression signatures, many of which are shared with normal HSC [8, 9]. These LSC signatures are highly prognostic, demonstrating that stemness is capturing a shared property linked to clinical outcome within heterogeneous AML cohorts. The LSC17 score represents the most recent and validated proof of this concept. Indeed, this score was highly prognostic in multiple independent adult or pediatric AML datasets spanning >1000 AML samples [9, 17]. Thus regardless of the diverse paths taken during leukemogenesis, the high prognostic power of the LSC17 score suggests that many leukemogenic pathways converge onto their impact on stemness properties. The convergence of intrinsic stemness properties between AML patients has enabled the deployment of the LSC17 score into the clinic [18]. Collectively, these studies highlight the need to better understand the determinants that drive stemness in AML.

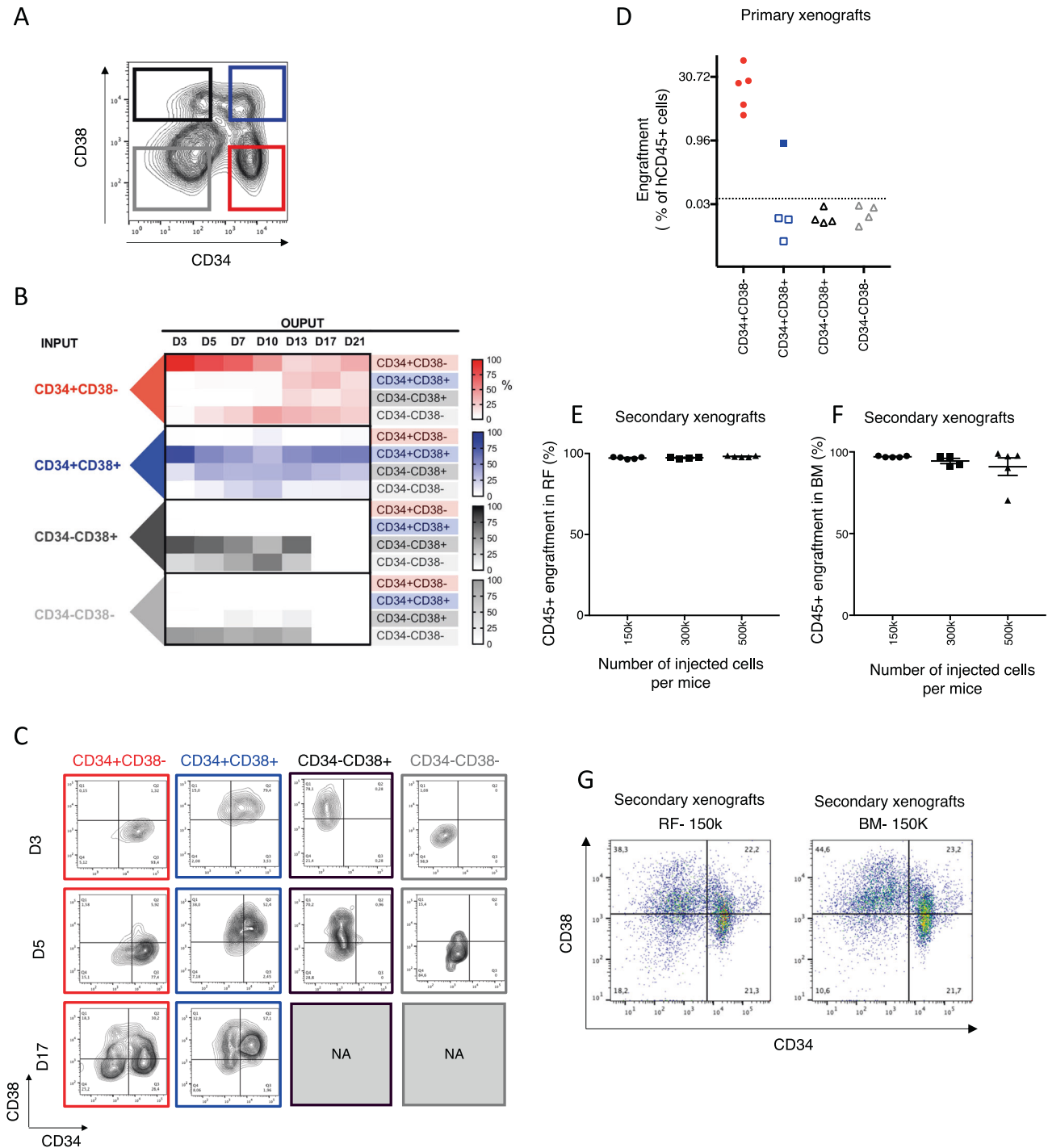
A number of key biological processes distinguish LSC from leukemia cells lacking stem cell activity. These include proteostatic

<sup>1</sup>Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 0A3, Canada. <sup>2</sup>Department of Medical Biophysics, University of Toronto, Toronto, ON M5S 1A4, Canada. <sup>3</sup>PanCuRx Translational Research Initiative, Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada. <sup>4</sup>Department of Regenerative Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan. <sup>5</sup>Princess Margaret Genomics Centre, University Health Network, Toronto, ON M5G 0A3, Canada. <sup>6</sup>Division of Medical Oncology and Hematology, Department of Medicine, University Health Network, Toronto, ON, Canada. <sup>7</sup>Department of Medicine, University of Toronto, Toronto, ON, Canada. <sup>8</sup>Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada. <sup>9</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada. This work is dedicated to the memory of A-MW. ✉email: john.dick@uhnresearch.ca

Received: 1 June 2022 Revised: 23 August 2022 Accepted: 1 September 2022  
Published online: 21 September 2022

**Table 1.** clinical characteristics of AML patient samples tested for in vitro expansion from Supplementary Fig. S1A.

Sample ID	Sex	Age at AML Dx	De novo vs. Secondary	Type of samples	Sample material	FAB	WBC at AML Dx	BM blast count at AML dx	MRC cytogenetics class at AML Dx	Initial Tx	NPM1 at dx	FLT3 ITD at Dx	FLT3 TKD at Dx
646	F	83.7	De novo	Diagnosis	PB	M5b	299	70	Intermediate (abnormal)	Supportive	nd	nd	nd
90240	F	54.2	De novo	Relapse3	Peritoneal fluid	M1	3	nd	Adverse	Induction	nd	nd	nd
80401	F	54	Secondary (MPN)	Diagnosis	PB	M1	24	60	Intermediate (abnormal)	Induction	nd	nd	nd
9682	F	76.6	De novo	Diagnosis	PB	Unclassified	34	nd	nd	Supportive	nd	nd	nd
80529	F	29.4	De novo	Diagnosis	PB	M4	21.1	60	Intermediate (normal)	Induction	Positive	Intermediate	Negative
90185	F	35.7	De novo	Diagnosis	PB	Unclassified	46	nd	Intermediate (normal)	Induction	Positive	Intermediate	Negative
80534	M	20.1	De novo	Relapse1—persistent	PB	M5a	54	95	Intermediate (normal)	Induction	nd	nd	nd
90736	F	49.9	De novo	Diagnosis	PB	Unclassified	41	nd	nd	Induction	nd	nd	nd
90707	M	59	De novo	Diagnosis	PB	M4	25	25	Intermediate (normal)	Induction	Positive	Negative	Negative
90784	F	61.9	De novo	Diagnosis	PB	M5a	59	90	Intermediate (normal)	Positive	High	Negative	Negative
858	M	55.9	Secondary (MPN)	Diagnosis	PB	M1	33	70	Intermediate (normal)	Induction	Positive	Negative	Negative
100052	M	79.2	Secondary (MDS)	Diagnosis	PB	M4	38	60	Adverse	Supportive	nd	nd	nd
100016	M	73.2	De novo	Diagnosis	PB	M5a	20	80	Intermediate (normal)	Induction	Negative (K)	Negative (K)	Negative (K)
100006	F	63.3	De novo	Diagnosis	PB	M5a	25	80	Intermediate (abnormal)	Induction	nd	nd	nd
100112	M	69.5	Secondary (MDS)	Persistent disease	PB	nd	16	80	Intermediate (normal)	Induction	Negative	Negative	Negative
100116	F	65.3	De novo	Diagnosis	PB	M4	46	nd	nd	Supportive	nd	nd	nd
100118	M	61.8	De novo	Persistent disease	PB	M1	204	90	Intermediate (normal)	Induction	Negative	Intermediate	Negative
100127	F	29.2	De novo	Relapse1	PB	Unclassified	10	nd	nd	Induction	nd	nd	nd
100183	M	69.8	De novo	Diagnosis	PB	M5a	88	85	Intermediate (normal)	Induction	Positive	Negative	Negative
100207	M	71.9	De novo	Diagnosis	PB	M3	62	nd	Favorable	Induction	nd	nd	nd
100249	F	67.1	De novo	Diagnosis	PB	Unclassified	282	95	nd	Induction	Positive	Intermediate	Negative
100255	M	59.6	De novo	Relapse1	PB	M4	108	50	Intermediate (normal)	Induction	Positive	Negative	Positive
80014	M	52	Secondary (rads)	Diagnosis	PB	M4Eo	35.9	90	Favorable	Induction	nd	nd	nd
90520	M	61.7	De novo	Diagnosis	PB	Unclassified	103	39	Intermediate (normal)	Induction	Positive	Negative	Positive
100274	M	59.6	De novo	Relapse1	PB	M4	108	50	Intermediate (normal)	Induction	Positive	Negative	Positive
100286	M	50.5	De novo	Diagnosis	PB	M1	77.2	90	Intermediate (abnormal)	Induction	nd	nd	nd
90272	F	29.2	De novo	Diagnosis	PB	Unclassified	10	nd	nd	Induction	nd	nd	nd
90517	M	61.8	De novo	Persistent disease	PB	M1	204	90	Intermediate (normal)	Induction	Negative	Intermediate	Negative
90047	F	58.3	De novo	Diagnosis	PB	M4	46.9	40	Intermediate (normal)	Induction	Positive	Negative	Negative
100347	F	63.3	De novo	Diagnosis	PB	M1	89.9	90	Intermediate (normal)	Induction	Negative	Negative	Negative
100454	M	62.8	Secondary (MPN)	Diagnosis	PB	Unclassified	27	50	Adverse	Supportive	nd	nd	nd



**Fig. 1** OCI-AML22 models the phenotypic and functional hierarchy found in primary AML. **A** Immunophenotypic profile of OCI-AML22 with the four fractions sorted as indicated based on CD34 and CD38 cell surface expression. **B** Immunophenotypic characterization of cellular outputs over time generated by each independently sorted and cultured fractions following in vitro expansion as assessed by flow cytometry. **C** Representative flow cytometric profiles of the independently sorted and cultured fractions over time; related to **B**. NA not available due to no viable cells remaining. **D** NSG mice were injected with sorted populations (100,000 cells per mouse) as indicated. Engraftment level was assessed 8 weeks after injection by flow cytometry measuring the percentage of human CD45+ (hCD45) cells in the injected femur. Each dot on the graph represents a mouse. **E–G** Cells collected from the xenografts generated after injection of OCI-AML22 fractions from **D** were pooled and sorted for human cells then injected into NSG-SGM3 mice at the indicated cell dose per mouse. Engraftment level (AnnexinV<sup>-</sup>, 7AAD<sup>-</sup>, hCD45<sup>+</sup>) was assessed 8 weeks later in the injected bone (RF). Each point represents a mice. **E** Non-injected bone, referred to as bone marrow (BM) (**F**). Representative FACS profiles of grafts in the right femur (left) or the bone marrow (right) are represented (**G**).

**Table 2.** Limiting dilution assay at 12 weeks.

Fraction	Dose	Tested	Response	Lower frequency	Estimated frequency	Upper frequency
CD34+CD38–	30,000	3	3	1/803.9	1/286.4	1/102
	7500	4	4			
	1875	6	6			
	469	5	4			

OCI-AML22 cells were expanded in culture for 4 months before CD34+CD38– cells were isolated and injected into NSG-SGM3 mice at the indicated cell doses. The number of injected mice (tested) and engrafted (response) as well as lower, estimated and upper LSC frequencies are indicated 12 weeks after injection.

responses, epigenetic pathways, immune escape ability and metabolism [5, 19–23] with a number translated into therapies [5, 6, 24–27]. However, difficulty in studying rare LSC populations represents a bottleneck for defining the regulatory processes that govern the stemness state. First, no markers exist to purify rare LSC populations to homogeneity [28]. This forces reliance on cumbersome, expensive and time-consuming xenograft assays for LSC detection for each patient sample [28, 29]. Second, each patient sample is limited in cell number and the rare LSC cannot be expanded since stemness is lost in culture [30]. Unfortunately, traditional human AML cell lines that have been grown for decades are of little use since they have lost many features of primary AML samples including a phenotypic and functional LSC-driven hierarchy [30]. Third, mechanistic interrogation of primary human LSC using standard genetic tools such as lentiviral transduction is possible but not efficient and non-lentiviral CRISPR editing of LSC has not yet been demonstrated. The reliance on lentivectors currently hinders functional validation of potential mechanisms driving stemness because it restricts identification of fundamental LSC features mostly to coding regions that represent only 2% of the genome [31]. There is increasing recognition that non-coding regulatory elements play important roles in controlling cell identities [16, 32]. Hence, it is crucial to identify human cellular models that recapitulate features of LSC-driven hierarchies characteristic of primary AML samples. Such models need to be easily genetically modified in ways that allow for interrogation of the entire coding and non-coding genomic landscape.

Here, we report the development of a patient-derived primary AML model, called OCI-AML22, that reflects the functional, transcriptional and epigenetic cellular hierarchy common to most primary AML samples, providing a powerful resource that permits investigation of the molecular basis of the stemness in LSC. OCI-AML22 can be efficiently modified with lentivectors and CRISPR-based methods, yielding a novel model that allows interrogation of stemness features that are critical for the future development of more efficient LSC-targeted therapies.

## MATERIAL AND METHODS

Please, refer to Supplementary Material for detailed methods

## RESULTS

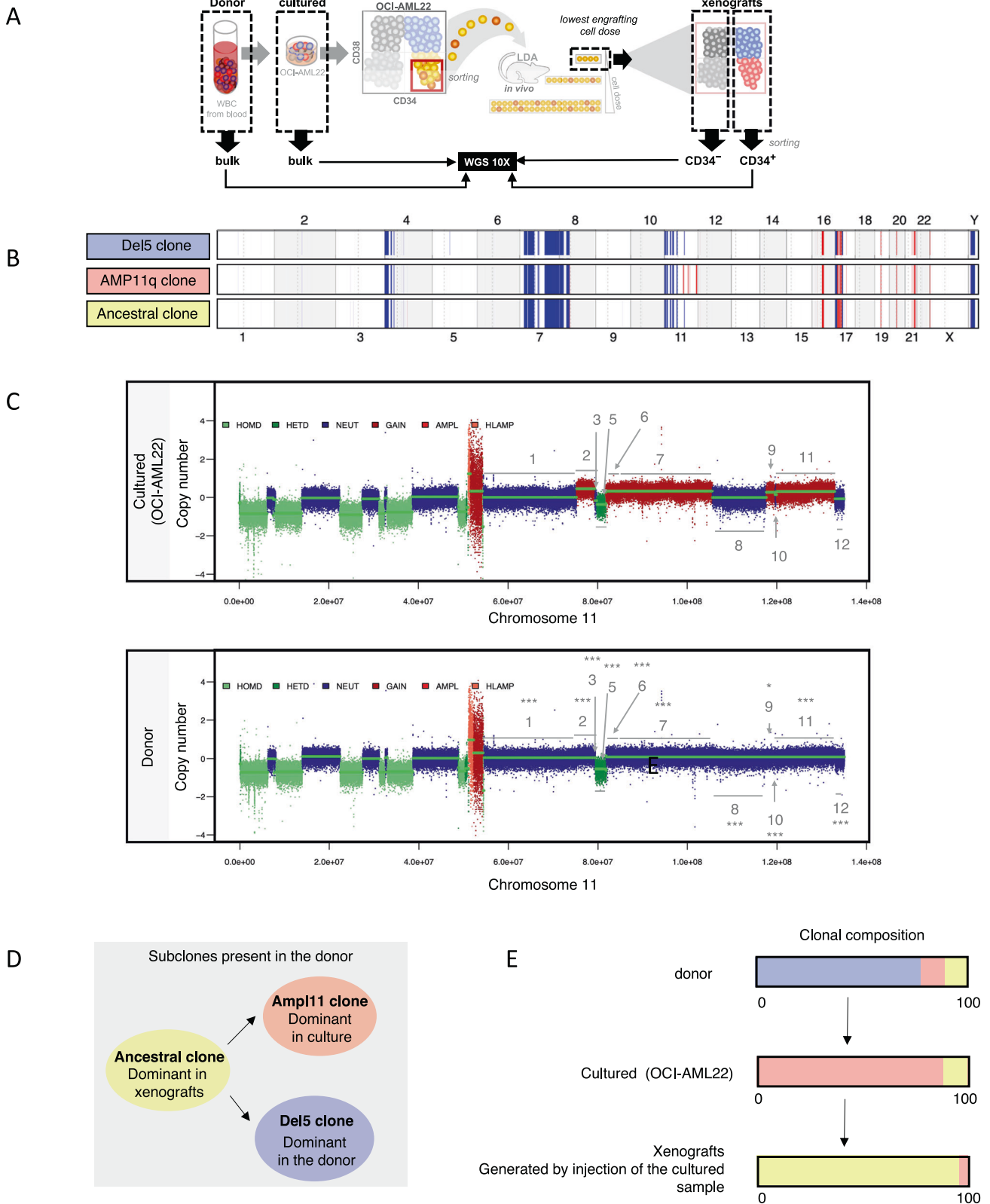
### OCI-AML22 models the phenotypic and functional hierarchy of primary AMLs

To establish an AML model that recapitulates a cellular hierarchy with LSCs at the apex, we screened AML patient samples ( $n = 34$ ) for their ability to grow and expand in culture (Supplementary Fig. S1A and Table 1). Most AML samples could not be maintained for more than a few weeks and no sample consistently expanded in culture (Supplementary Fig. S1A), which is aligned with previous observations [30]. However, we identified one sample, from a relapse patient that could be expanded long term in vitro (Supplementary Fig. S1A, B) after optimization of culturing conditions. This cultured AML model is termed OCI-AML22. In

vitro expansion resulted in an immunophenotypic hierarchy of four fractions defined by CD34 and CD38 cell surface expression (Fig. 1A and Supplementary Fig. S1B). The CD34+CD38– fraction was the only fraction able to maintain the entire phenotypic hierarchy in culture re-establishment assays (Fig. 1B, C). In parallel, functional xenotransplant LSC assays on each of the four CD34/CD38 subpopulations, sorted after 3–4 months of ex vivo expansion, established that only CD34+CD38– fractions contained LSC (Fig. 1D). All NSG mice (5/5) injected with 100,000 CD34+CD38– cells were engrafted at 8 weeks, while the same number of CD34+CD38+ or CD34– cells resulted either only in a single engrafted mouse (1 of 4 injected mice, 0.8% engraftment) or in no detectable engraftment, respectively (Fig. 1D). Robust engraftment following serial transplantation, that regenerated a full phenotypic hierarchy, conclusively established that CD34+CD38– LSC from primary mice had self-renewal potential [33] (Fig. 1E–G). The absolute LSC content within the CD34+CD38– fraction, quantified using limiting dilution xenograft assays, established that the LSC frequency was high (1/286; range: 1/102–1/804) (Table 2) when compared to the spectrum of LSC frequencies from fractions obtained from primary AML samples [9, 34] (Supplementary Fig. S2A). Collectively, these data establish that OCI-AML22 exhibits a phenotypic and functional hierarchy driven from LSC at the apex and has the capacity for both in vitro and in vivo long-term propagation.

### OCI-AML22 maintains the polyclonal genetic architecture of the primary AML sample

Since the donor sample from which OCI-AML22 was derived had complex cytogenetics, we assessed its genetic stability during in vitro and in vivo expansion. Whole genome sequencing (WGS) was undertaken on the primary donor sample (day 0), the cultured OCI-AML22 model (~day 30), and on xenografts generated from CD34+CD38– cells sorted from the cultured sample (~day 120) (Fig. 2A). Genomic analysis showed that cultured and donor samples were highly similar with 99% conservation. The CD34+ fraction from xenografts showed an even higher conservation (99.8%). Detailed analysis further revealed that the genetic differences between the dominant clones of the cultured, xenografted and donor samples were reflective of subclonal genetic diversity, where the dominant populations of the culture and xenografts arose from rare preexisting subclones present in the donor sample (Fig. 2B–D and Supplementary Fig. S3A and Tables 3 and 4). WGS showed that the dominant clone in the OCI-AML22 cultured sample exhibits a series of amplifications on chromosome 11 (Supplementary Fig. S3A, arrow 1). This clone was termed clone AMP11 (Supplementary Fig. S3A, line 2 and Fig. 2B, line 2, in red). This series of amplifications on chromosome 11 is not present in the dominant clone of the donor sample (Supplementary Fig. S3, arrow 1, line 1). However, a focused copy number analysis of this region on chromosome 11 in the donor sample showed that these alterations were already present at a subclonal level (Fig. 2C and Tables 3 and 4). Thus the series of 11q amplifications were not generated by a genetic drift due to culture, but arose by the selective amplification of a minor clone



preexisting in the donor sample. Of note, the AMP11 clone could also be detected at the subclonal level in xenografts generated by the cultured sample (Tables 3 and 4). WGS also revealed that the dominant clone in the donor (Supplementary Fig. S3A, arrow 1) contains a small deletion on chromosome 5 (and lacking the 11q

amplification) (Supplementary Fig. S3A, arrow 2); termed clone del5 (Supplementary Fig. S3A, line1 and Fig. 2B, line 1, in blue). By contrast, the dominant clone in xenografts, generated by the cultured sample, lacked both the chromosome 5 deletion and the chromosome 11q amplifications (Supplementary Fig. S3A, line 3).

**Fig. 2 OCI-AML22 maintains the polyclonal genetic architecture of the primary AML sample. A** Schematic representation of the samples sequenced for WGS. **B** Alterations present in each of the dominant clones are displayed. Copy number losses are defined as CN less than 1.5 (shown in blue) while gains have CN greater than 2.5 (shown in red). The width of the colored region corresponds to the size of the modified region in the genome. **C** Representation of the copy number obtained on chromosome 11 for cultured and donor samples. Segment regions that were detected using Hmcopy in the cultured sample (top) were used to break down the chr11 arm q of the donor sample (bottom) into matching regions. Within these segments, each dot (at 1 kb intervals) was taken to compare the average copy number to the adjacent segments and thus determine if the average copy numbers were different between the indicated adjacent regions. Wilcoxon tests were run. Each start on top of each region shows the existence of amplifications that can be significantly detected at a subclonal level in the donor sample. **D** Evolutionary relationships of subclones present in the donor sample. The sample where the clone is dominant is indicated. **E** Clonal composition of the different sequenced samples using the same clones color coding as in **B** and **D**.

**Table 3.** Chromosome 11q fragments being compared coordinates.

Fragments being compared	Coordinates
1	chr11:54526001-75256001
2	chr11:75256001-79404001
3	chr11:79404001-79816001
4	chr11:79816001-81783001
5	chr11:81783001-81826001
6	chr11:81826001-81900001
7	chr11:81900001-105587001
8	chr11:105587001-117653001
9	chr11:117653001-119635001
10	chr11:119635001-119932001
11	chr11:119932001-132956001
12	chr11:132956001-135087001

Segment regions that were detected using Hmcopy in the cultured sample and were used to break down the chr11 arm q of the donor sample or the fractions sorted from xenografts.

Since a deletion cannot be spontaneously restored to normal, the clone that preferentially expands in vivo must be ancestral to both the del5 and the AMP11 clones. This clone is termed the ancestral clone (Supplementary Fig. S3A, line 3 and Fig. 2B, line 3 in yellow). Thus, in vitro culture, followed by in vivo propagation of the donor sample, provided insight into the evolutionary relationships of the three genetic subclones present at different levels in the original donor sample (Fig. 2D). These include the del5 clone, dominant in the donor sample (Fig. 2B, D, E, in blue), the minor AMP11 clones, and the ancestral clone able to generate both del5 and AMP11 subclones (Fig. 2B, D, E, in yellow). The ancestral clone must be present in both the donor and in culture at subclonal levels since it becomes detectable following xenotransplantation of the cultured OCI-AML22 cells (Fig. 2E, in yellow). The AMP11 clone, present at the subclonal level in the donor, preferentially expands in culture and remains present but at the subclonal levels in xenografts generated by the cultured sample (Fig. 2B, D, E, in red). Altogether, the WGS analysis revealed that the cultured OCI-AML22 model is genetically stable during prolonged in vitro and in vivo propagation, while maintaining at least two of the clones originally present in the donor sample: the ancestral clone and the AMP11q clone. These results are reminiscent of our findings documenting leukemia evolution using xenografts [15, 35, 36] and highlight that the OCI-AML22 model maintains the polyclonal genetic architecture characteristic of primary AML samples in contrast with traditional AML cell lines that are clonal.

### OCI-AML22 preserves a transcriptional and epigenetic landscape of primary stem cells throughout ex vivo expansion

To determine if the transcriptional signature of the CD34+CD38- fraction was maintained over time, we performed deep RNA sequencing (RNA-seq) of CD34+CD38- cells from the donor sample and four immunophenotypic fractions harvested at

multiple time points of culture (12, 60, 90 days) (Fig. 3A). Principal component analysis (PCA) showed that even after 2 to 3 months of in vitro culture, the CD34+CD38- fractions of OCI-AML22 and the donor sample clustered together indicating that ex vivo expanded cells preserved the global transcriptomic landscape of its original donor (Fig. 3B). In accordance with our functional data (Fig. 1), CD34+CD38+ populations were positioned in the PCA space between the CD34+CD38- engrafting and non-engrafting CD34- fractions. CD34- fractions clustered together and were the most distinct from CD34+CD38- populations. The functional hierarchical organization described in Fig. 1 prompted us to investigate whether OCI-AML22 recapitulates LSC and non-LSC features extracted from heterogeneous primary AML cohorts, as well as the diverse AML cellular states recently extracted from scRNA-seq of primary AML samples [10]. Gene set variation analysis (GSVA) on each of the OCI-AML22 fractions showed that the OCI-AML22 model recapitulates the various cellular states extracted from primary AML samples with enrichment of primitive states in the functional LSC fractions (HSC, Progenitor, summarized as HSC.Prog) and progressive enrichment for mature states in the non-LSC containing fractions (GMP, Pro-mono, Monocyte, summarized as Myeloid) (Fig. 3C, D and Supplementary Fig. S4A-E). Similarly, using chromatin accessibility signatures generated from highly purified fractions obtained from the normal hematopoietic system [32], we show that the OCI-AML22 CD34+ fraction was enriched in stem cell signatures (LT/HSPC and Act/HSPC), while the CD34- non-engrafting fraction was enriched for signatures of mature myeloid populations (granulocyte, monocyte signatures) (Supplementary Fig. S4F). All OCI-AML22 fractions showed the lowest concordance with the erythroid and lymphocyte (T cells and B cells) signatures as compared to mature myeloid populations (granulocyte, monocyte). This result is concordant with a block of differentiation and a shift toward myeloid pathways characteristic for AML (Supplementary Fig. S4F).

To determine whether OCI-AML22 CD34+CD38- fraction recapitulates the stem cell transcriptional programs shared with primary LSCs, we used the LSC104 signature, a stem cell tool whose clinical relevance has been demonstrated in multiple cohorts [9]. The stemness properties of any group of samples can be determined on the basis of a Spearman correlation coefficient to the average gene expression of each of these 104 LSC+ genes [9]. First we compared the gene expression values of the reference set of 104 LSC-specific genes generated on a group of 138 primary LSC+ fractions (LSC+ reference), to LSC104 gene expression values calculated for each of the OCI-AML22 fractions obtained at different time points. The resulting correlation score was the highest when comparing the OCI-AML22 CD34+CD38- fractions to the LSC+ reference; the score gradually decreased alongside the reduced engraftment ability of the fractions with the lowest coming from the non-engrafting CD34-CD38- fractions (Fig. 3E, left). These results remained consistent, independent of the time the OCI-AML22 model was maintained in culture (Fig. 3E, right). The same conclusions were validated using three independent approaches: LSC17 scoring [9] across all fractions (Fig. 3F); GSVA scoring using the LSC-R signature [8], a previous stemness

**Table 4.** Comparisons of fragments between the donor or xenografts fractions (CD34+ or CD34-) and the cultured sample.

Comparison	In the donor sample		CD34 positive from xenografts		CD34 minus from xenograft		In either 34p or cd34m
	Wilcox.p value	Significance	Wilcox.p value	Significance	Wilcox.p value	Significance	
1 to 2	1.11E-267	***	0.523362	ns	0.087603	ns	ns
2 to 3	3.80E-218	***	3.93E-239	***	1.47E-240	***	***
3 to 4	7.64E-06	***	0.461146	ns	0.714475	ns	ns
4 to 5	2.01E-25	***	3.27E-28	***	3.92E-27	***	***
5 to 6	1.88E-17	***	1.56E-18	***	1.03E-17	***	***
6 to 7	4.44E-46	***	1.58E-48	***	2.78E-48	***	***
7 to 8	0	***	0.722962	ns	0.004922	***	***
8 to 9	3.96E-09	***	0.000187	***	9.39E-16	***	***
9 to 10	0.0387516	*	2.05E-06	***	5.40E-05	***	***
10 to 11	6.89E-12	***	0.000777	***	0.25894	ns	***
11 to 12	4.85E-125	***	0.000819	***	5.27E-14	***	***

Segment regions that were detected using hmmcopy (Table 3) in the cultured sample were used to break down the chr11 arm q of the donor sample, or xenografts into matching regions. Within these segments, each dot (at 1kb intervals) was taken to compare the average copy number to the adjacent segments and thus determine if the average copy numbers were different between the indicated adjacent regions. Wilcox tests were run. Significance shows the existence of amplifications that can be detected at a subclonal level in the interrogated sample: donor sample or xenografts (CD34+ fraction, CD34- fraction or both).  $p < 0.05$ : \*,  $p < 0.01$ : \*\*,  $p < 0.001$ : \*\*\*.

signature (Fig. 3G); and GSVA scoring using the HSC-R signature, a stemness signature generated from normal hematopoietic stem cells [8] (Supplementary Fig. S4G). This demonstrates that the OCI-AML22 hierarchy recapitulates the stemness features of a diverse cohort of primary samples.

Although the OCI-AML22 model is derived from a complex cytogenetic sample, we validated that LSC from OCI-AML22 are representative of stemness properties exhibited by primary AML samples regardless of karyotype. Indeed, correlation of the LSC104 gene expression values generated from the CD34+CD38- OCI-AML22 fraction was similar when compared to the LSC104 average values generated from either LSC+ fractions of normal karyotype ( $n = 60$ ) or abnormal karyotype ( $n = 55$ ) samples (Fig. 3H). Of note, both normal and abnormal karyotype spanned similar LSC frequency ranges (Supplementary Fig. S5A). Finally, since the LSC+ fractions obtained from primary AML samples present a broad range of LSC frequencies, we generated LSC104 correlation data of four groups of primary AML fractions, based on their LSC frequency: none, low, medium, and high, to determine which group is the most reflective of the OCI-AML22 LSC population. The LSC104 gene expression values of the the OCI-AML22 CD34+CD38- fraction correlated best with the LSC104 values of the highest LSC frequency group (Fig. 3I, left), and this was independent of the culture time (Fig. 3I, right), confirming the strong LSC enrichment in the OCI-AML22 LSC fraction. Altogether, our data establishes that the OCI-AML22 LSC fraction captures the transcriptional and epigenetic stemness programs of highly purified LSC fractions obtained from patients across a wide spectrum of clinical and genetic properties (Supplementary Fig. S6) and these are preserved during culture.

#### OCI-AML22 enables functional interrogation of leukemia stemness properties

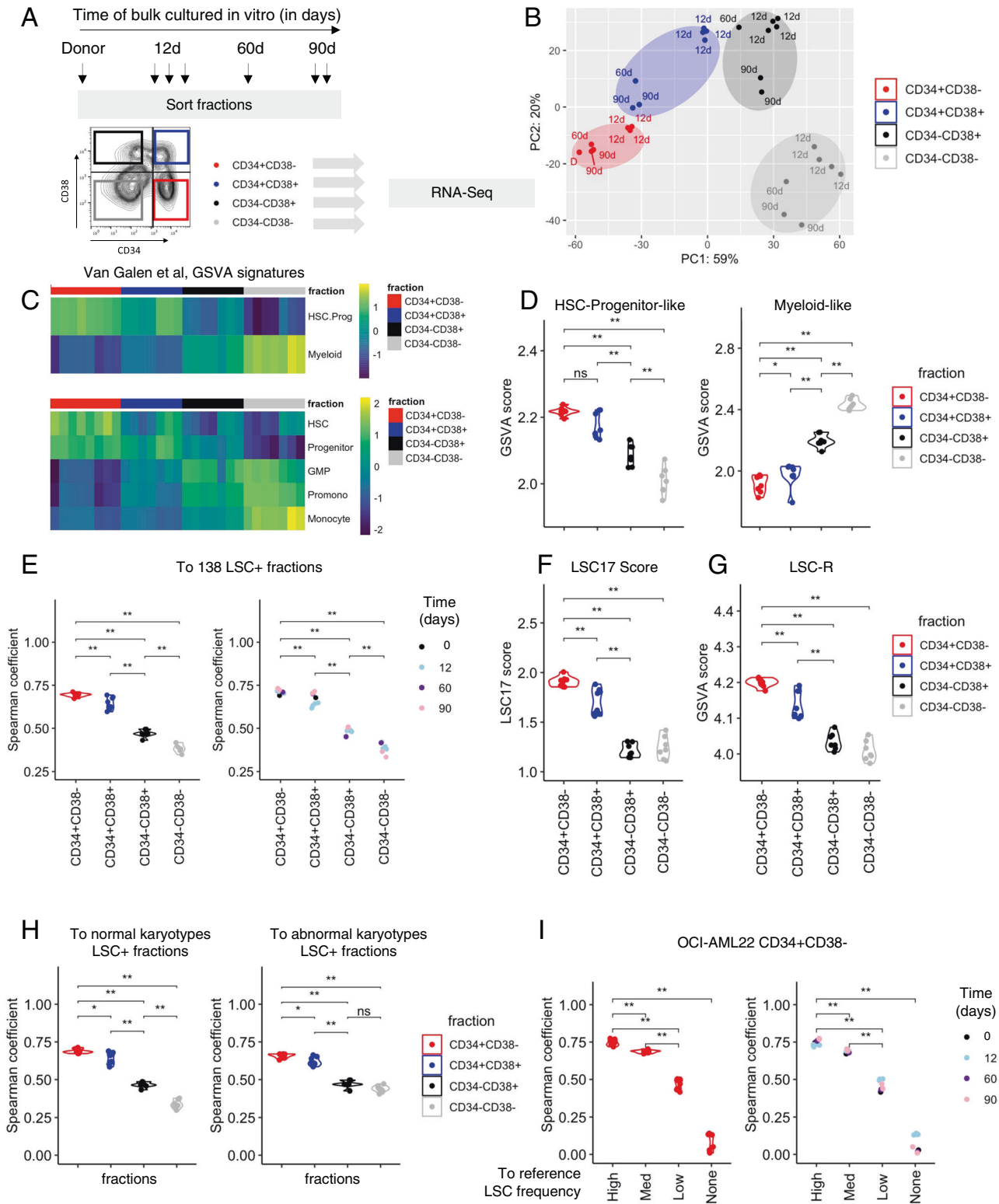
To determine if the OCI-AML22 model can be used to interrogate molecular determinants driving stemness, we tested a variety of biological properties that were previously reported to discriminate populations within the cellular hierarchy of primary AML samples. We first tested the potential of the OCI-AML22 model to give insights into LSC-specific immune evasion properties since recent studies of 177 primary AML samples showed that LSC, but

not their non-stem cell progeny, are able to evade Natural Killer (NK)-driven anti-tumor immunity. LSCs upregulate PARP1 whose overexpression downregulates the activity of NKG2D ligands that are recognized by NK cells [21]. As expected, PARP1 expression was higher in engrafting (CD34+) vs. non-engrafting (CD34-) OCI-AML22 fractions (Fig. 4A). GSVA showed that genes known to be upregulated by NKG2D ligands were less enriched in CD34+ compared to CD34- fractions (Fig. 4B). Inversely, genes downregulated by NKG2D ligands were more enriched in CD34+ compared to CD34- fractions (Fig. 4C, D). Second, GSVA on each OCI-AML22 fraction showed that ATF4 upregulated genes were significantly enriched in CD34+CD38- fractions as compared to other non-engrafting fractions (Fig. 4E and Supplementary Fig. S7A). This was also validated in vivo, using the ATF4 biosensor (Fig. 4F, G and Supplementary Fig. S7B-E). These data confirm that the integrated stress response (ISR) is regulated across the OCI-AML22 hierarchy in the same way as previously described across primary normal or LSC hierarchies, where the ISR is closely associated with LSC function [20, 37]. Moreover, HOXA9 and MEIS1 are key transcription factors well known to be overexpressed in LSC compared to non-LSC [11] and to be correlated to poor prognosis [38], which we have confirmed across the OCI-AML22 hierarchy (Fig. 4H, I). Additionally, many studies have shown that LSCs harbor a characteristic energy metabolism centering on a high oxidative phosphorylation signature [39-41]. We confirmed that this pathway is also upregulated in the LSC+ OCIAML22 fractions compared to the non-LSC fractions (Fig. 4J). Finally, we confirmed that the ROSLow signature [42] that discriminates LSC+ from LSC- fractions in primary AML samples also discriminates LSC+ from LSC- fractions in OCI-AML22 (Fig. 4K).

Collectively, these data validate that the OCI-AML22 hierarchical structure recapitulates key hallmarks previously reported for primary AML samples, both in vitro and in vivo. Thus, OCI-AML22 represents a good model to study LSCs and interrogate stemness properties.

#### Clusters of cis-regulatory elements (CORE) discriminate LSC from non-LSC fractions in primary AML samples

We next set out to uncover potential new stemness regulators associated with LSC in primary AML, as a basis to demonstrate the



power of OCI-AML22 to provide new insights into LSC biology. Chromatin accessibility of non-coding genomic regions has greater power in predicting cell state identity as compared to gene expression [16]. We established a new discovery approach to identify functionally relevant non-coding regions that define the LSC state as compared to non-LSC populations. In the human genome, 20–40% of the non-coding regions are predicted to be

covered by cis-regulatory elements (CRE) [38]. These elements are not evenly spread across the human genome and can be either isolated or clustered together, suggesting that coordinated, fine-tuned regulation occurs at these clusters and they are acting as larger entities of biological relevance. Clusters of cis-regulatory elements (COREs) are regions of high CRE density defined by chromatin accessibility which we identified using the machine



**Fig. 3 The OCI-AML22 CD34+CD38- fraction preserves the transcriptional and epigenetic landscape of primary stem cells throughout ex vivo expansion.** **A** Schematic representation of OCI-AML22 sorting strategy applied for RNA-Seq. Each arrow indicates an independently expanded culture. These fractions are used throughout the RNA-Seq analysis. **B** Principal component analysis (PCA) of RNA-seq data generated from CD34+CD38- (red), CD34+CD38+ (blue), CD34-CD38+ (black) and CD34-CD38- (gray) subpopulations. **C** Supervised heatmap clustering for GSVA scores calculated for the van Galen signatures [10] and organized based on OCI-AML22 sorted fraction. Pathways indicative of primitive-like AML cells (HSC-like, Progenitor-like are combined into the HSC-Prog-like signature) while the others (GMP, Promo and Monocyte) are combined in the Myeloid signature. **D** GSVA score of genes present in the HSC-Progenitor-like (left), Myeloid-like (right) signatures (Van Galen et al.) across OCI-AML22 fractions. **E** Spearman correlation calculated for the gene expression of the LSC104 genes [9] of each OCI-AML22 fractions compared to the average gene expression of these genes across 138 primary AML fractions, enriched for functional LSC described in ref. [9]. Fractions are colored depending of the fraction type (left) or the time OCI-AML22 has been maintained in culture before the sort (right). **F** LSC17 score calculated for each OCI-AML22 fraction. **G** Gene set variation analysis (GSVA) score of genes present in the LSC-R signature [8] across OCI-AML22 fractions. **H** Spearman correlation coefficient calculated between the LSC104 signature of each of the OCI-AML22 indicated fraction at the bottom, and the LSC104 signature from a group of normal karyotype LSC+ (60 fractions) (left) or a group of abnormal karyotype LSC+ fractions (55 fractions) (right). **I** Spearman correlation coefficient comparing the LSC104 signature of the OCI-AML22 CD34+CD38- fractions, to each of the LSC104 signatures for groups of samples indicated on the bottom (LSC frequency High, Medium (Med), low or no detectable LSC activity (LSC neg). Points are colored according to the OCI-AML22 fraction sorted (left) or depending on the time OCI-AML22 has been kept in culture before sorting the CD34+CD38- fraction; 12d : 12 days, 60d : 60 days, 90d : 90 days, D: primary donor(right).

learning algorithm CREAM [39]. They have been shown to reliably discriminate cell identities across multiple cancer cell types. Hence, we hypothesized that CREAM could identify COREs able to distinguish LSC from their non-LSC counterparts thereby uncovering new regulators of stemness governing LSC biology. CREAM was applied to chromatin accessibility data generated on LSC containing (LSC+;  $n = 41$ ) and LSC depleted (LSC-;  $n = 52$ ) fractions that were isolated from 25 uncultured AML patient samples and functionally validated as described previously [9] (Fig. 5A). According to the calculated predictability coefficient, the CORE with the highest potential to distinguish LSC+ from LSC- fractions was located within intronic regions of chromosome 9 (CORE-chr9-2014811-2032652) (Fig. 5B) and was detected in the majority (>75%) of the LSC+ fractions but only in 20% of the LSC- fractions (Fig. 5C). CREAM identified this CORE as encompassing seven individual CREs located within intronic regions of the SMARCA2 locus (Fig. 5D), and that were accessible in different proportions of LSC+ compared to LSC- fractions (Fig. 5E). Overall, the recurrent pattern defined by CORE-chr9-2014811-2032652 (LSC + CORE) across LSC+ vs. LSC- suggests that individual CREs of this CORE might be particularly relevant for LSC function, but functional validation is required.

#### Functional interrogation of individual CRE located within CORE-chr9-2014811-2032652 through OCI-AML22 LSC CRISPR/Cas9 editing

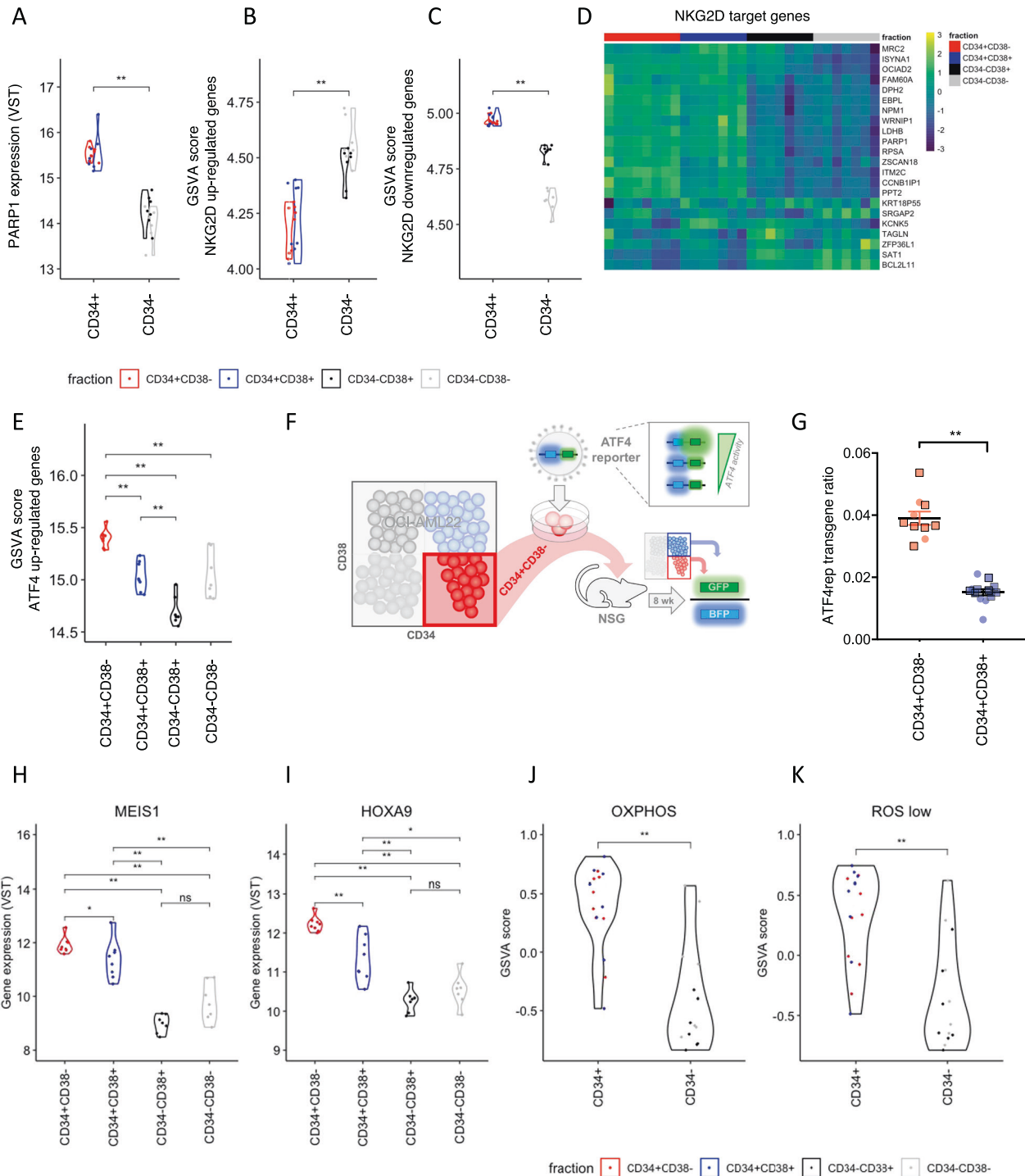
Functional studies require an experimental model, so we first mined ENCODE. No cell lines showed a similar pattern of accessibility peaks as found in our set of 93 primary AML sorted fractions (Fig. 6A, B). By contrast, there was a highly concordant accessibility pattern of the LSC+ CORE region between OCI-AML22 fractions and the catalog of peaks obtained from 93 primary AML fractions (Fig. 6A, B). These results demonstrate the superiority of OCI-AML22, compared to any other AML cell lines, at recapitulating chromatin accessibility features characteristic of LSCs.

To functionally test if LSCs depend on individual CREs found within the LSC+ CORE locus, we specifically targeted 2 CREs in the LSC fractions of OCI-AML22 using CRISPR/Cas9 methods we optimized. The OCI-AML-22 LSC fraction was used to either individually knock-out (KO) CRE3 or CRE6 or delete the entire flanking region (Fig. 7A, B and Table 5). Effective KO of each of the regions of interest was confirmed by PCR/gel electrophoresis and drop out of either CRE3 or CRE6 was observed for any of the gRNA pairs selected (Fig. 7B). Of note, control and CRE3 KO cells showed no adverse short-term effects mediated by CRISPR/Cas9 editing, as they expanded and maintained their ability to

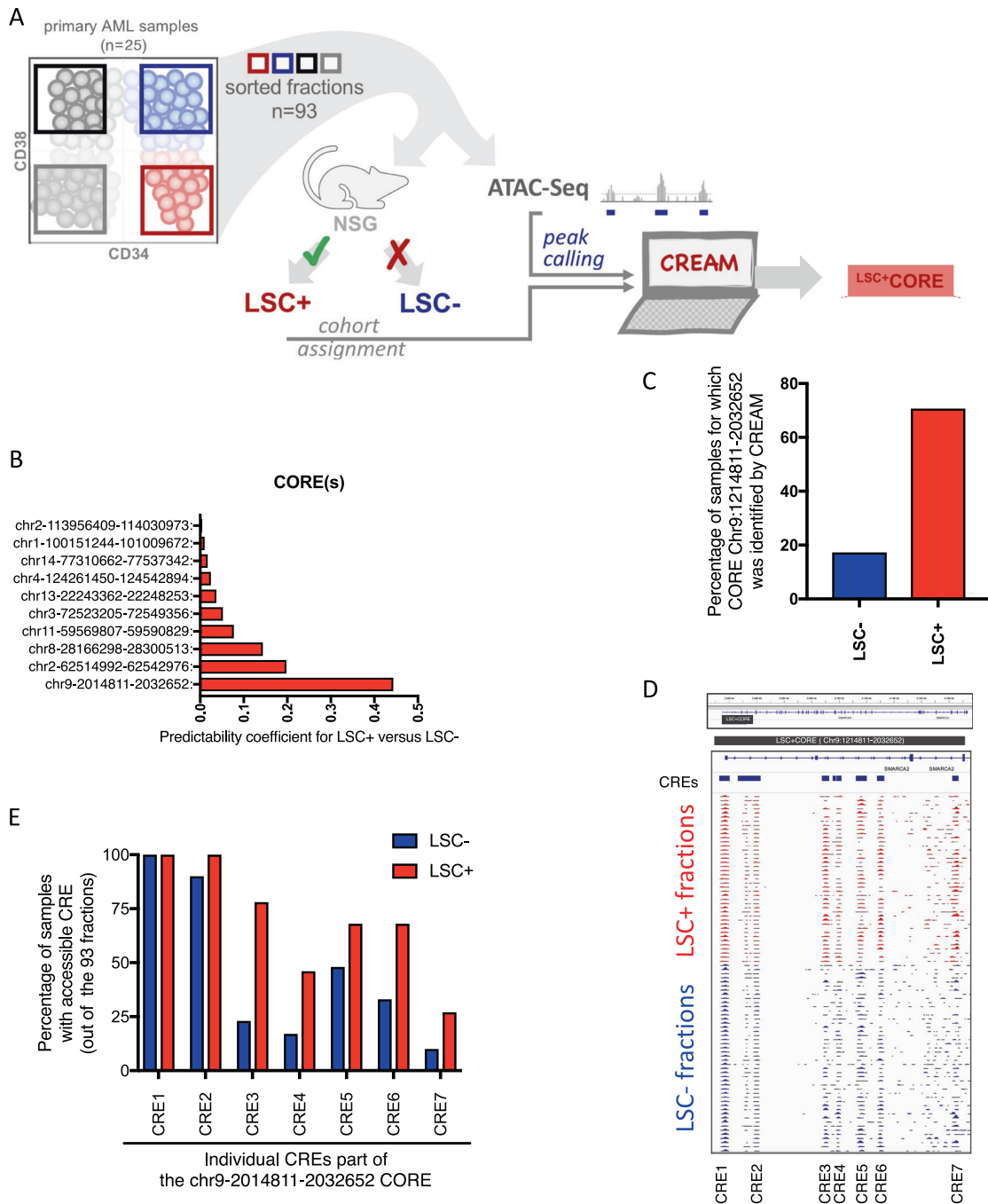
regenerate all four subpopulations (Fig. 7C and Supplementary Fig. S8A–J). In contrast, KO of either CRE6 or of the entire region from CRE3 to CRE6 in the LSC fraction significantly diminished in vitro growth potential, leading to an overall reduction of the number of cells (Fig. 7C), as well as reduction of the four fractions generated by the initially edited CD34+CD38- cells (Supplementary Fig. S8A–J). We additionally performed in vivo functional assays by injecting control and KO cells in NSG-SGM3 mice. All control mice were engrafted 21 weeks after cell injection whereas none of the KO CRE3 or CRE3-6 mice were engrafted (Fig. 7D–F). Taken together, these results establish that KO of CRE3 or KO of the CRE3 to CRE6 region (CRE3-6) caused significant reduction in repopulation potential of the transplanted cells (Fig. 7E, F). Since none of the CRE3 nor CRE3-6 mice were engrafted, this formally established that the KO eradicated all serially transplantable LSC. Overall, these data report the discovery of a new class of stemness regulators within specific non-coding regions of the genome; a result of great value to the stem cell and cancer stem cell community, that exemplifies the utility of OCI-AML22.

#### DISCUSSION

Here, we address the long-standing need for a tool to overcome the challenges of identifying, extracting, and obtaining high numbers of LSC from primary AML samples to enable functional, genetic/epigenetic, biochemical and metabolic studies. OCI-AML22 is the first human AML model that not only faithfully recapitulates the hierarchically organized LSC and non-LSC hallmark states of primary AML samples [8, 9, 20, 21], but also whose LSC fraction can be efficiently identified, isolated and edited using CRISPR editing technology. This sets the stage for mechanistic, functional, and translational studies in the accurate cellular context of the CSC state. Additionally, our study shows how it is now feasible to extend mechanistic studies to uncover stemness regulators lying outside of the coding genome, as demonstrated with functional validation of CRE6 within the LSC +CORE. The OCI-AML22 model shows broad clinical relevance, not just to the relapse sample from which it was derived. The stem cell properties present in the OCI-AML22 LSC population are highly reflective of the stemness properties that are common and highly prognostic from across many highly diverse independent cohorts of primary AML samples [9, 17, 40]. We foresee that the OCI-AML22 model, as well as the extensive dataset generated on the functionally characterized LSC and non-LSC fractions, will be extensively used to interrogate and functionally validate features driving stemness.



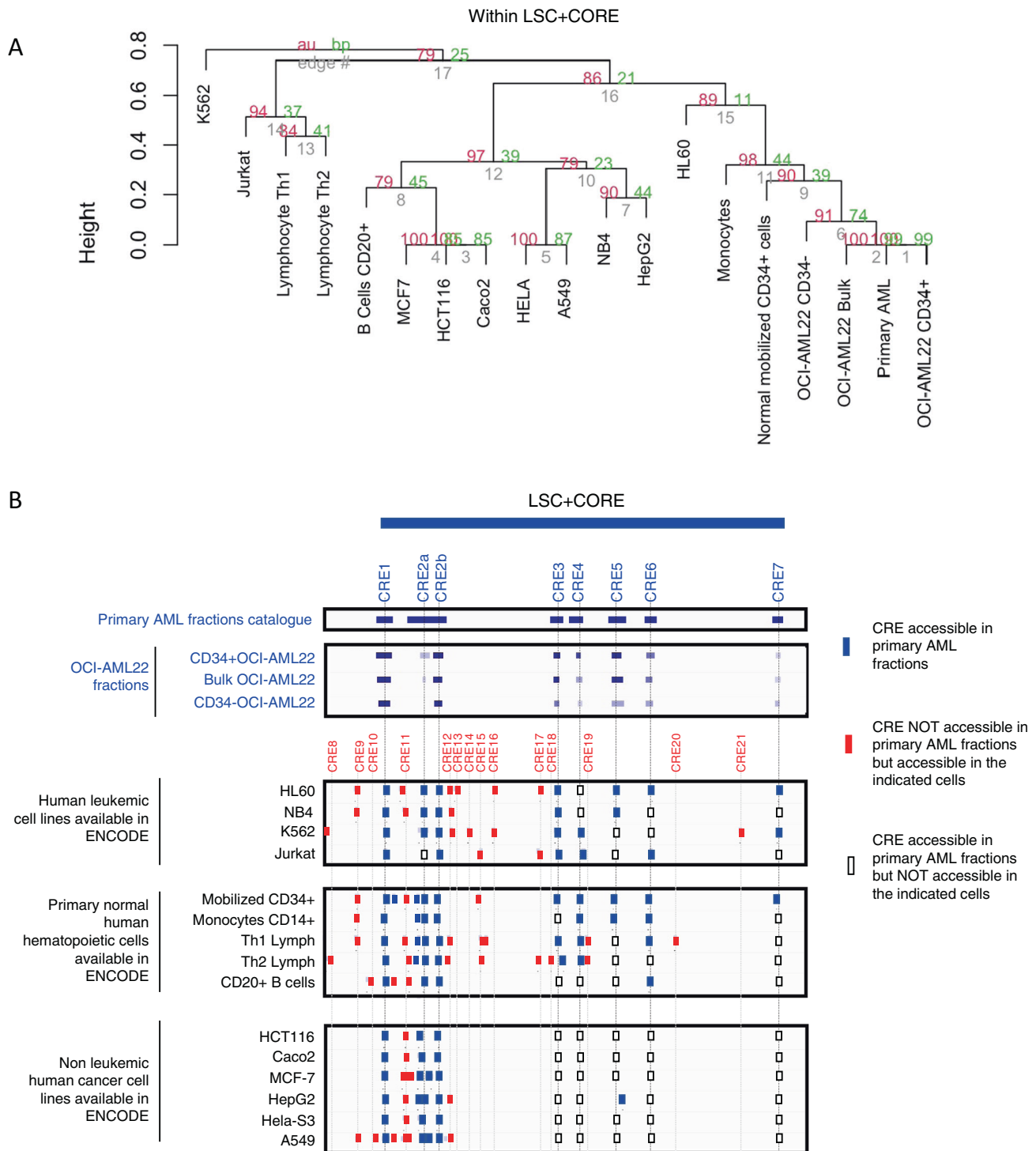
**Fig. 4** OCI-AML22 enables functional interrogation of leukemia stemness properties. **A** PARP1 expression of OCI-AML22 fractions was determined by RNA-Seq. GSVAs for NKG2D upregulated (**B**) or downregulated (**C**) genes as described in ref. [21] across OCI-AML22 sorted fractions as indicated in Fig. 2A. **D** Supervised heatmap clustering of NKG2D target gene expressions of sorted OCI-AML22 fractions. **E** GSVAs score of ATF4 upregulated genes as described in ref. [20] across individual OCI-AML22 fractions as extracted in Fig. 2B. **F** Experimental scheme. **G** OCI-AML22 CD34+CD38- cells were transduced with ATF4 lentiviral fluorescent reporter (ATF4rep) to track ATF4 transcriptional activity. BFP and GFP were assessed on viable human CD45+ cells extracted from the injected femora of NSG mice injected with 100k (round) or 200k (square) transduced cells, 8 weeks post xenotransplantation. The ATF4rep transgene ratio (GFP/BFP; BFP as internal transduction control) was then calculated for each indicated fractions of the 7AAD-AnnexinV- hCD45+ population from xenografts. Error bars (s.e.m) are indicated. The expression of MEIS1 (**H**) and HOXA9 (**I**) was assessed across OCI-AML22 sorted fractions as indicated in Fig. 2A. **J** GSVAs score for genes that are part of the OXPPOS High signature [41], across OCI-AML22 sorted fractions as indicated in Fig. 2A. **K** GSVAs score for genes that are part of the ROSLow signature [42], across OCI-AML22 sorted fractions as indicated in Fig. 2A.



**Fig. 5 Chromatin accessibility profile of LSC enriched and depleted fractions identifies COREs discriminating LSC+ from non-LSC fractions.** **A** Schematic representation of the experimental strategy, **B** coordinates and predictability coefficient for LSC+ vs. LSC- for CORE that are detected in either LSC+ or LSC- fractions. **C** Percentage of LSC- and LSC+ fraction for which CREAM algorithm detects the CORE-chr9 2014811-2032652, out of the 93 fractions sequenced, **D** IGV representation of the CORE-chr9 2014811-2032652 for primary AML LSC+ and LSC- fractions, **E** Percentage of LSC+ or LSC- primary AML fractions for which CRE part of the CORE of interest are accessible out of the 93 fractions sequenced.

CRISPR editing strategies have opened new avenues beyond gene interrogation including editing non-coding elements, which are increasingly being recognized as being regulators of stemness and/or cancer [41]. Non-coding regions are known to harbor cell type specific determinants such as transposable elements or cis-regulatory elements [42]. The importance of these regions in stem cell driven cancers including AML, has been postulated via computational exploration of primary AML sample landscapes [43]. However, these studies did not focus on the rare stem cell

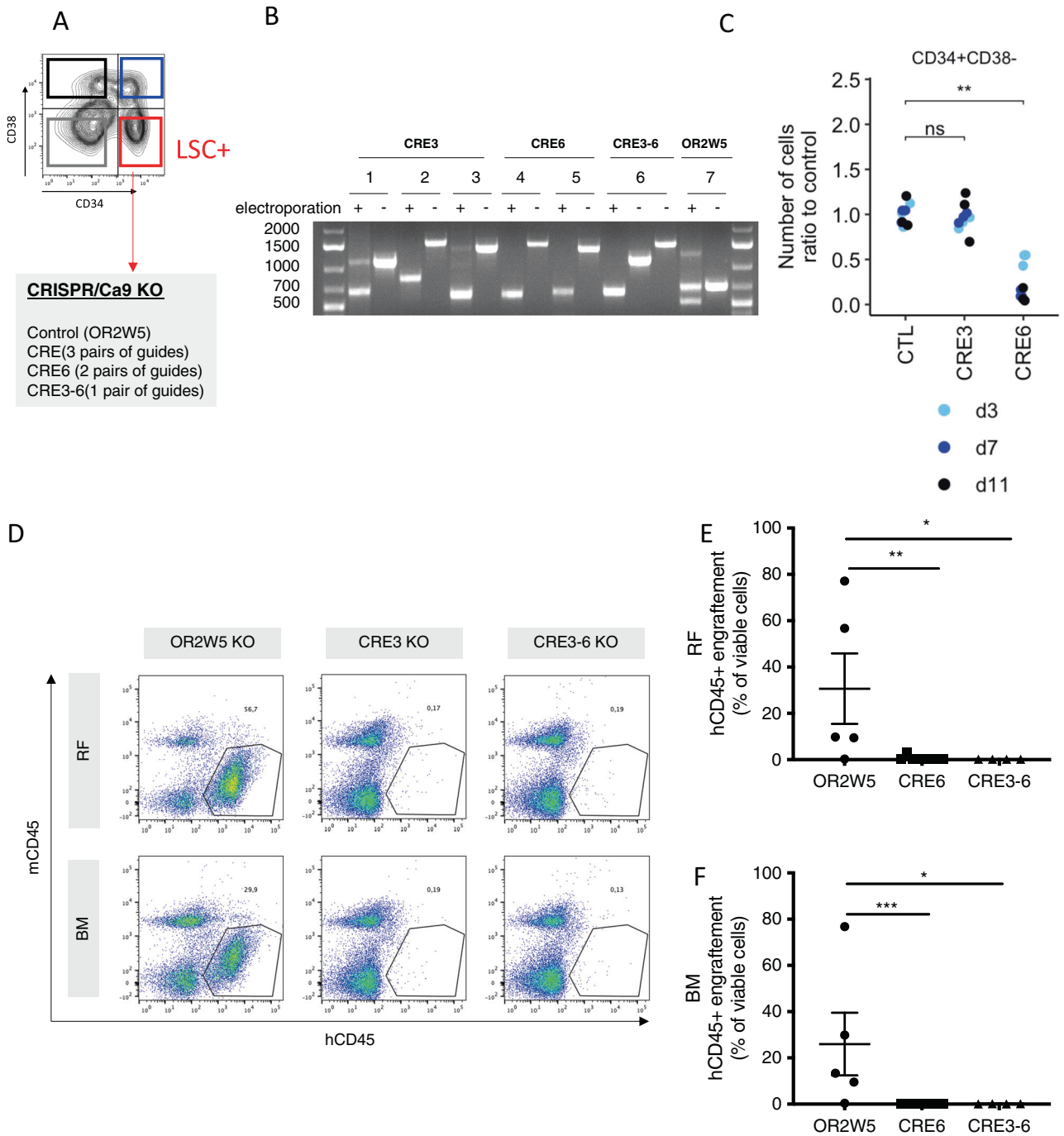
population that drives the disease nor did they provide functional validation in primary LSC. Functional validation of non-coding elements has been performed via CRISPR editing but only on AML cell lines [44] that can be easily expanded and genetically modified. However, contrary to the OCI-AML22 model, none of these cell lines reflects an LSC-driven hierarchical structure that mimics primary AML samples. More concerning, after decades in culture, their epigenetic landscape has lost the epigenetic architecture of primary AML samples, limiting the clinical



**Fig. 6 OCIAML22 recapitulates chromatin accessibility exhibited by primary AML samples within the CORE-chr9 2014811-2032652 locus.** **A** Hierarchical clustering using CRE elements within the CORE of interest (CORE-chr9-2014811\_2032652) across the indicated samples. **B** Chromatin accessibility profiles within the CORE of interest, for the indicated samples. CRE accessible in primary AML fractions have been highlighted in dark blue, CRE not accessible in primary AML fractions but present in the indicated cells have been highlighted in red, CRE accessible in primary AML fractions but not accessible in the indicated cells are surrounded with a black line as indicated in the legend.

transferability of functional validation performed on these models. Developing and implementing CRISPR/Cas9 editing strategies has not yet been reported in primary AML samples, let alone in rare human LSC, thus the functional proof of the role of CRE in a primary human LSC context with the precision of CRISPR-

mediated KO was still missing. Our study provides strong evidence for the functional relevance of non-coding regions that have been predicted through machine learning analysis of matched primary LSC and non-LSC fractions as determinants of stemness. We foresee that leveraging the power of CRISPR with the ability of



**Fig. 7 CRISPR/Cas9 in OCIAML22 LSC fraction allows to functionally interrogate individual CRE located within the CORE of interest.** **A** Schematic representation of the CRISPR editing strategy. **B** Fragments generated after PCR of the CRISPR edited (electroporated) or non CRISPR edited (not electroporated) after KO of the indicated elements: CRE3, CRE6 or the entire region from CRE3 to CRE6 (CRE3-6) or control OR2W5, using PCR guides that are flanking the region to KO. The shorter band represents the fraction of KO cells while the longer band represents the proportion of cells not KO. **C** Total cells expanded in culture at the indicated time points after CRISPR editing of the regions of interest in the OCIAML22 LSC fraction. **D–F** The OCIAML22 LSC fraction was CRISPR edited then cells were expanded for 22 days before injection in NSG mice. After 21 weeks, the engraftment was assessed. representative FACS profile (**D**) and engraftment level in either the injected femur (right femur, RF) (**E**) or Bone marrow (BM) (**F**) are represented. Error bars (s.e.m) are indicated.

**Table 5.** Size of the expected amplicons and additional information for CRISPR editing methods.

Portion to KO	gRNA pairs used	Guides sequence	Expected fragment size if portion KO (bp)	Expected fragment size if portion is NOT KO (bp)	F primer sequence	R primer sequence
CRE3-1	3' guide	TCAAATCATGACCCCTGAA	659	1432	GAAAGTCAGGCCAGTTGAC	AAGGTTGTCGGGATAAGCGG
	5' guide	CTGGCAAGGACCATCCCA				
CRE3-2	3' guide	TCAAATCATGACCCCTGAA	836	1653	GAAAGTCAGGCCAGTTGAC	TCTTGGATCCTCCAGCCTGTC
	5' guide	TGTCAGATACAAATCACACC				
CRE3-3	3' guide	TCAAATCATGACCCCTGAA	608	1690	GAAAGTCAGGCCAGTTGAC	GGAGAACTGCAAGGTTTCAGGA
	5' guide	TAAGCGGTGAAACATTACCA				
CRE6-1	3' guide	ACATCAATTATTAAACGT	608	1773	CGTCACGCTTCTATGAGAGTCC	ACTCTAGGACCCCATCACT
	5' guide	caagtagctaggactacagg				
CRE6-2	3' guide	TTCCTTTAGCaAAAAaCG	642	1677	AGTCCAGGCATCTGAGGTTTC	ACTCTAGGACCCCATCACT
	5' guide	caagtagctaggactacagg				
CRE 3 to 6	3' guide	TCAAATCATGACCCCTGAA	637	Portion 1 <sup>a</sup> : 1432	GAAAGTCAGGCCAGTTGAC	AAGGTTGTCGGGATAAGCGG
	5' guide	caagtagctaggactacagg		Portion 2 <sup>a</sup> : 1773	CGTCACGCTTCTATGAGAGTCC	ACTCTAGGACCCCATCACT
OR2W5	3' guide	GACAACCGAGGAGCGCACT	500	For fragment cut <sup>a</sup>	GAAAGTCAGGCCAGTTGAC	ACTCTAGGACCCCATCACT
	5' guide	CTCCGGTGTGACGTCGCA		700	5'-TCGGCCTGGACTGGAGAAAA-3'	5'-GAGACCACACTGTGAGGTGAGA-3'

<sup>a</sup>Given that the region from CRE3 to CRE6 was too long to amplify with classic PCR, only a portion of the region was sequenced using the indicated primers, starting from one side (primers F and R1) or the other side (portion 1) or the other side (portion 2), or the full region in the case same has been KO (for fragment cut), leading to the following expected fragment sizes.

OCI-AML22 to model stemness determinants that can be extracted from primary AML samples will allow functional interrogations of many more aspects of human LSC features without being limited to the coding genome.

## DATA AVAILABILITY

Raw data are deposited in the EGA under the series accession number EGAD00001009271 that includes RNA-Seq; EGAS00001006512, ATAC-Seq; EGAS00001006511, WGS; EGAS00001006513. Due to privacy reason, raw data are not available publicly but will be available from the corresponding author on reasonable request. Processed files for RNA-Seq, ATAC-Seq and WGS are deposited under the NCBI's Gene Expression Omnibus, accessible through GEO series accession GSE211596.

## REFERENCES

- Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med.* 2016;374:2209–21.
- Brewin J, Horne G, Chevassut T. Genomic landscapes and clonality of de novo AML. *N Engl J Med.* 2013;369:1472–3.
- Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368:2059–74.
- Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell.* 2010;17:13–27.
- Jones CL, Stevens BM, Pollyea DA, Culp-Hill R, Reisz JA, Nemkov T, et al. Nicotinamide metabolism mediates resistance to venetoclax in relapsed acute myeloid leukemia stem cells. *Cell Stem Cell.* 2020;27:748–64.e4.
- Pollyea DA, Stevens BM, Jones CL, Winters A, Pei S, Minhajuddin M, et al. Venetoclax with azacitidine disrupts energy metabolism and targets leukemia stem cells in patients with acute myeloid leukemia. *Nat Med.* 2018;24:1859–66.
- Laverdière I, Boileau M, Neumann AL, Frison H, Mitchell A, Ng SWK, et al. Leukemic stem cell signatures identify novel therapeutics targeting acute myeloid leukemia. *Blood Cancer J.* 2018;8:1–16.
- Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med.* 2011;17:1086–93.
- Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature.* 2016;540:433–7.
- van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. *Cell.* 2019;176:1265–81.
- Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun.* 2015;6:8489.
- van Rhenen A, Feller N, Kelder A, Westra AH, Rombouts E, Zweegman S, et al. High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. *Clin Cancer Res.* 2005;11:6520–7.
- Misaghian N, Ligresti G, Steelman LS, Bertrand FE, Bäsbecke J, Libra M, et al. Targeting the leukemic stem cell: the Holy Grail of leukemia therapy. *Leukemia.* 2009;23:25–42.
- Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med.* 1997;3:730–7.
- Shlush LI, Mitchell A, Heisler L, Abelson S, Ng SWK, Trotman-Grant A, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature.* 2017;547:104–8.
- Corces MR, Buenostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet.* 2016;48:1193–203.
- Duployez N, Marceau-Renaut A, Villenet C, Petit A, Rousseau A, Ng SWK, et al. The stem cell-associated gene expression signature allows risk stratification in pediatric acute myeloid leukemia. *Leukemia.* 2019;33:348–57.
- Ng SW, Murphy T, King I, Zhang T, Mah M, Lu Z, et al. A clinical laboratory-developed LSC17 stemness score assay for rapid risk assessment of acute myeloid leukemia patients. *Blood Adv.* 2021. <https://doi.org/10.1182/bloodadvances.2021005741>.
- Xie SZ, Kaufmann KB, Wang W, Chan-Seng-Yue M, Gan OI, Laurenti E, et al. Sphingosine-1-phosphate receptor 3 potentiates inflammatory programs in normal and leukemia stem cells to promote differentiation. *Blood Cancer Discov.* 2021;2:32–53.

20. van Galen P, Mbong N, Kreso A, Schoof EM, Wagenblast E, Ng SWK, et al. Integrated stress response activity marks stem cells in normal hematopoiesis and leukemia. *Cell Rep.* 2018;25:1109–17.
21. Paczulla AM, Rothfelder K, Raffel S, Konantz M, Steinbacher J, Wang H, et al. Absence of NKG2D ligands defines leukaemia stem cells and mediates their immune evasion. *Nature.* 2019;572:254–9.
22. Lechman ER, Gentner B, Ng SWK, Schoof EM, van Galen P, Kennedy JA, et al. miR-126 regulates distinct self-renewal outcomes in normal and malignant hematopoietic stem cells. *Cancer Cell.* 2016;29:214–28.
23. Jones CL, Stevens BM, D'Alessandro A, Reisz JA, Culp-Hill R, Nemkov T, et al. Inhibition of amino acid metabolism selectively targets human leukemia stem cells. *Cancer Cell.* 2018;34:724–40.
24. Chao MP, Takimoto CH, Feng DD, McKenna K, Gip P, Liu J, et al. Therapeutic targeting of the macrophage immune checkpoint CD47 in myeloid malignancies. *Front Oncol.* 2020;9. <https://doi.org/10.3389/fonc.2019.01380>.
25. Jin L, Hope KJ, Zhai Q, Smadja-Joffe F, Dick JE. Targeting of CD44 eradicates human acute myeloid leukemic stem cells. *Nat Med.* 2006;12:1167–74.
26. Jin L, Lee EM, Ramshaw HS, Busfield SJ, Peoppl AG, Wilkinson L, et al. Monoclonal antibody-mediated targeting of CD123, IL-3 receptor alpha chain, eliminates human acute myeloid leukemic stem cells. *Cell Stem Cell.* 2009;5:31–42.
27. Majeti R, Chao MP, Alizadeh AA, Pang WW, Jaiswal S, Gibbs KD, et al. CD47 is an adverse prognostic factor and therapeutic antibody target on human acute myeloid leukemia stem cells. *Cell.* 2009;138:286–99.
28. Arnone M, Konantz M, Hanns P, Paczulla Stanger AM, Bertels S, Godavarthy PS, et al. Acute myeloid leukemia stem cells: the challenges of phenotypic heterogeneity. *Cancers.* 2020;12. <https://doi.org/10.3390/cancers12123742>.
29. Pollyea DA, Jordan CT. Therapeutic targeting of acute myeloid leukemia stem cells. *Blood.* 2017;129:1627–35.
30. Pabst C, Krosil J, Fares I, Boucher G, Ruel R, Marinier A, et al. Identification of small molecules that support human leukemia stem cell activity ex vivo. *Nat Methods.* 2014;11:436–42.
31. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860–921.
32. Takayama N, Murison A, Takayanagi S-I, Arlidge C, Zhou S, Garcia-Prat L, et al. The transition from quiescent to activated states in human hematopoietic stem cells is governed by dynamic 3D genome reorganization. *Cell Stem Cell.* 2021;28:488–501.
33. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell.* 2014;14:275–91.
34. Kaufmann KB, Garcia-Prat L, Liu Q, Ng SWK, Takayanagi S-I, Mitchell A, et al. A stemness screen reveals C3orf54/INKA1 as a promoter of human leukemia stem cell latency. *Blood.* 2019;133:2198–211.
35. Notta F, Mullighan CG, Wang JCY, Poepll A, Doulatov S, Phillips LA, et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature.* 2011;469:362–7.
36. Dobson SM, Garcia-Prat L, Vanner RJ, Wintersinger J, Waanders E, Gu Z, et al. Relapse-fated latent diagnosis subclones in acute B lineage leukemia are drug tolerant and possess distinct metabolic programs. *Cancer Discov.* 2020;10:568–87.
37. van Galen P, Kreso A, Mbong N, Kent DG, Fitzmaurice T, Chambers JE, et al. The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature.* 2014;510:268–72.
38. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA.* 2014;111:6131–8.
39. Madani Tonekaboni SA, Mazrooei P, Kofia V, Haibe-Kains B, Lupien M. Identifying clusters of cis-regulatory elements underpinning TAD structures and lineage-specific regulatory networks. *Genome Res.* 2019;29:1733–43.
40. Murphy T, Ng SWK, Zhang T, King I, Arruda A, Claudio JO, et al. Trial in progress: feasibility and validation study of the LSC17 score in acute myeloid leukemia patients. *Blood.* 2019;134:2682.
41. Burns KH. Transposable elements in cancer. *Nat Rev Cancer.* 2017;17:415–24.
42. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet.* 2016;17:93–108.
43. Assi SA, Imperato MR, Coleman DJL, Pickin A, Potluri S, Ptasinaka A, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet.* 2019;51:151–62.
44. Li K, Zhang Y, Liu X, Liu Y, Gu Z, Cao H, et al. Noncoding variants connect enhancer dysregulation with nuclear receptor signaling in hematopoietic malignancies. *Cancer Discov.* 2020;10:724–45.

## ACKNOWLEDGEMENTS

We thank Olga Gan for technical support, Andy G X Zeng for helpful insight, and the Leukemia Tissue Bank at Princess Margaret Cancer Centre/University Health Network

as the source of the primary sample blood. We thank Nicholas Khuu from the Princess Margaret Genomics Centre, Toronto, Canada ([www.pmgenomics.ca](http://www.pmgenomics.ca)) for ATAC-Seq sequencing, Sergio Pereira of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada for assistance with RNA sequencing. The WGS study was conducted with the support of the Ontario Institute for Cancer Research's Genomics & Bioinformatics platform ([genomics.oicr.on.ca](http://genomics.oicr.on.ca)) through funding provided by the Government of Ontario. We thank Sabrina Smith and Gaayatiri Pushparaj for assistance with experiments. We thank Melissa Anders for providing support for writing, Eric Lechman and Jean C Y Wang for critical reading of the manuscript and all members of the Dick lab for critical feedback.

## AUTHOR CONTRIBUTIONS

HB established the model, conceived the study, performed research, performed bioinformatic analysis and wrote the manuscript. SAMT performed chromatin study, AMu and MC-S-Y performed bioinformatic analysis, SAMT and MC-S-Y assisted with bioinformatic analysis. EW helped for methods, NM and AMi performed research, ML provided funding, supervised chromatin studies, MDM and AA provided AML samples, EO and FN supervised part of the study, AMu and MDM edited the manuscript, KBK performed research, wrote the manuscript, provided input to conceive the study. JED wrote the manuscript, secured funding and supervised the study.

## FUNDING

JED is supported by funds from the: Princess Margaret Cancer Centre Foundation, Ontario Institute for Cancer Research through funding provided by the Government of Ontario, Canadian Institutes for Health Research (RN380110–409786), Canadian Cancer Society (grant #703212 (end date 2019), #706662 (end date 2025)), Terry Fox New Frontiers Program Project Grant (Project# 1106), a Canada Research Chair, Princess Margaret Cancer Centre, The Princess Margaret Cancer Foundation (PMCF), and Ontario Ministry of Health. HB was supported by the Helena Lam Fellowship in Cancer Research Fund from the PMCF.

## COMPETING INTERESTS

JED serves on the SAB for Graphite Bio, receives royalties from Trillium Therapeutics Inc/Pfizer and receives a commercial research grant from Celgene/BMS.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41375-022-01697-9>.

**Correspondence** and requests for materials should be addressed to John E. Dick.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022