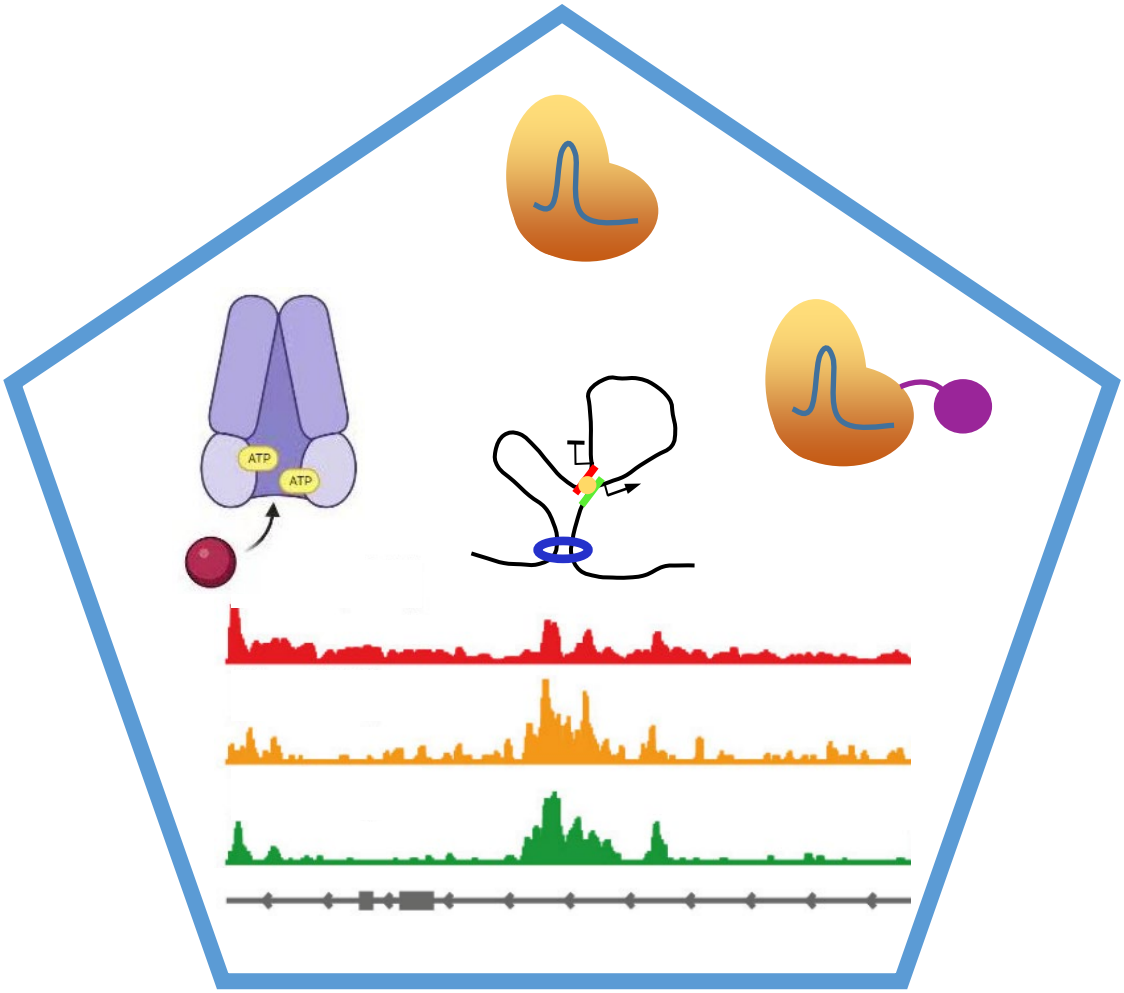
Functional study of the human genome

Li, Y.



Functional study of the human genome

Yufeng Li

李豫丰

Functional study of the human genome

Yufeng Li

Functional study of the human genome

ISBN:

Cover design: Yufeng Li
Layout: Yufeng Li, Baoxu Pang, Shengnan Sun

Functional study of the human genome

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op maandag 02 oktober 2023

klokke 15.30 uur

door

Yufeng Li

geboren te Tianmen, China

in 1992

Promotor: Prof.dr. J.J.C. Neefjes

Co-promotor: Dr. B. Pang

Promotiecommissie:

Prof.dr. R.C. Hoeben
Prof.dr. P. ten Dijke
Prof.dr. M.J.T.H. Goumans
Prof.dr.ir. S.M. van der Maarel
Dr. S.M. Noordermeer

The work described in this thesis was performed at the Department of Cell and Chemical Biology of the Leiden University Medical Center, Leiden, the Netherlands.

The research described in this thesis was supported by a personal grant of the China Scholarship Council to Yufeng Li (No. 201806990016).

Table of Contents

Chapter 1

Introduction

Introduction

The human genome contains 3 billion base pairs. Since the release of the first draft human reference genome in 2000, the details of our own blueprint have been getting more and more accurate[1,2]. According to the most recent telomere-to-telomere sequencing release, one human genome contains 3,054,815,472 bp of nuclear DNA, plus a 16,569-bp mitochondrial genome with a total of 63,494 genes, of which 19,969 are predicted to code proteins[3]. If we consider the human genome as an instruction book to make a human being, we have a reasonably good grasp of the letters in this book. However, we are still far from understanding the content of this instruction book. Proteins are the main building blocks of cells and the body. Even for the proteins, new functions are constantly updated. Yet the sequence coding the proteins only represents less than 2% of the human genome[3].

The rest of the genome is the protein non-coding genome, which contains mainly repeat regions, transposons, gene introns, non-coding RNAs or other intergenic regions[6]. In fact, although the majority of the human genome contains non-coding sequences, many of them have certain functions and play important roles in regulating the transcription of genes. Depending on the function of the non-coding regulatory regions (NCREs), they can be categorized into a few major classes, such as promoters (where transcriptional machinery assemble and drive the transcription), enhancers (which boost transcription), insulators (which bridge distal NCREs with promoters or serve as the barriers), or silencers (which repress transcription)[7-11]. Following the human genome project, we also started to catalog our genome, not only based on its intrinsic genomic information but also on epigenomic build-ups. International consortium projects like ENCODE and Roadmap Epigenomics have made enormous contributions to profile the non-coding part of the genome based on their epigenetic modifications[12-17], hoping to locate and define the potential functions of the NCREs. Besides computational prediction, recent developments of massively parallel reporter systems also facilitate the direct biological activity measurement and identifications of promoters, enhancers and silencers[6,18,19].

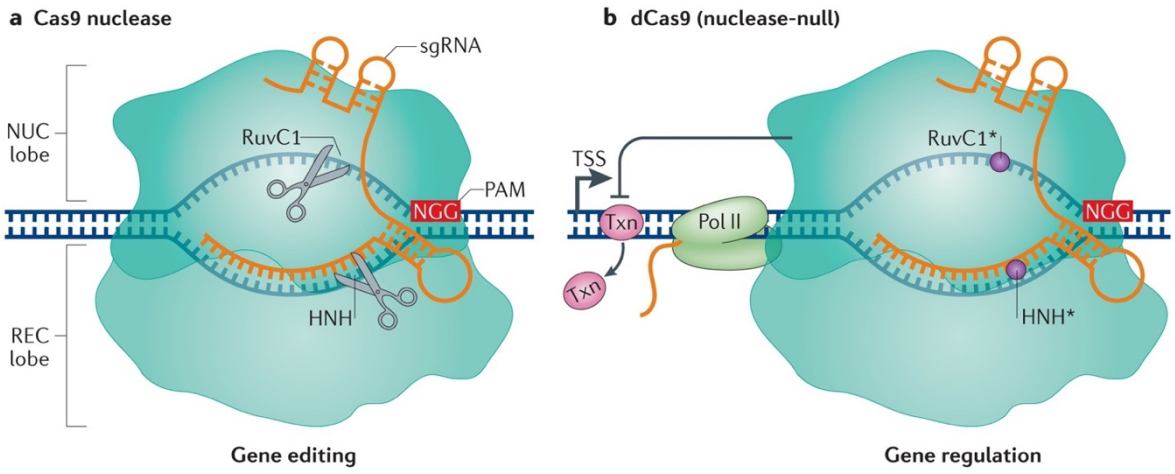


Figure 1: Mechanisms of Streptococcus pyogenes Cas9 and dCas9
a. The *S. pyogenes* Cas9 endonuclease consists of a nuclease (NUC) domain and a recognition (REC) domain. Cas9 is targeted to specific DNA sequences by direct pairing of the chimeric single guide RNA (sgRNA) with the target DNA. **b.** The *S. pyogenes* dCas9 protein contains mutations in its RuvC1 (D10A) and HNH (H841A) domains, which inactivate its nuclease function (circles). dCas9 retains the ability to target specific sequences through the sgRNA and PAM[5].

To understand the content of the human genome, it is key first to understand the function of their key elements—either genes or NCREs, before diving into their complex interactions to form the complicated human bodies. To study the function of genes, many tools have been developed during the past few decades to study the function of genes. The cloning of genes in the 70s led to the explosion of our understanding of gene functions[20]. In the following years, the discovery of siRNA and shRNAs provided a new way to study the functions of genes by reducing the expression levels of genes in cells[21]. Further development of these techniques allows for studying all potential genes in a genome-wide and systematic fashion focusing on distinct phenotypes[22-24]. The past decade witnessed the application of another revolutionary genetic tool to study the function of the genome, as exemplified by the CRISPR-Cas9-mediated genome editing systems[25-27] (Figure 1). Modification of the CRISPR-Cas9 systems not only allows the study of functions genes (in terms of silencing, upregulation and repression), but for the first time also allows the study of NCREs in their endogenous loci[5,28-31] (Figure 2).

CRISPR–Cas9 recognizes 20-bp genomic regions followed by PAM (5'-NGG-3') and typically introduces genetic changes of a few nucleotide deletions or insertions around the targeting sites in the genome[32]. When these mutations are introduced in the gene coding regions, it would render the proteins not functional anymore. On the other hand, NCREs often range from 50-200 bp in length, with multiple transcription factor (TF) binding sites[33], a single guide RNA mediated CRISPR editing may not completely destroy the TF binding sites[34,35]. Due to such limitations, single guide RNAs tiling an entire testing region were often used to study NCREs regulating a few important genes[28]. The modified CRISPR system that uses catalytic inactive Cas9 proteins (dCas9), either linked to a transcription activation or repression system, can also be used to study the functions of enhancers and insulators[35,36]. However, the design of tiling RNAs in selected few regions is also required. Therefore, systematic and genome-wide studies on the NCREs have not been performed yet.

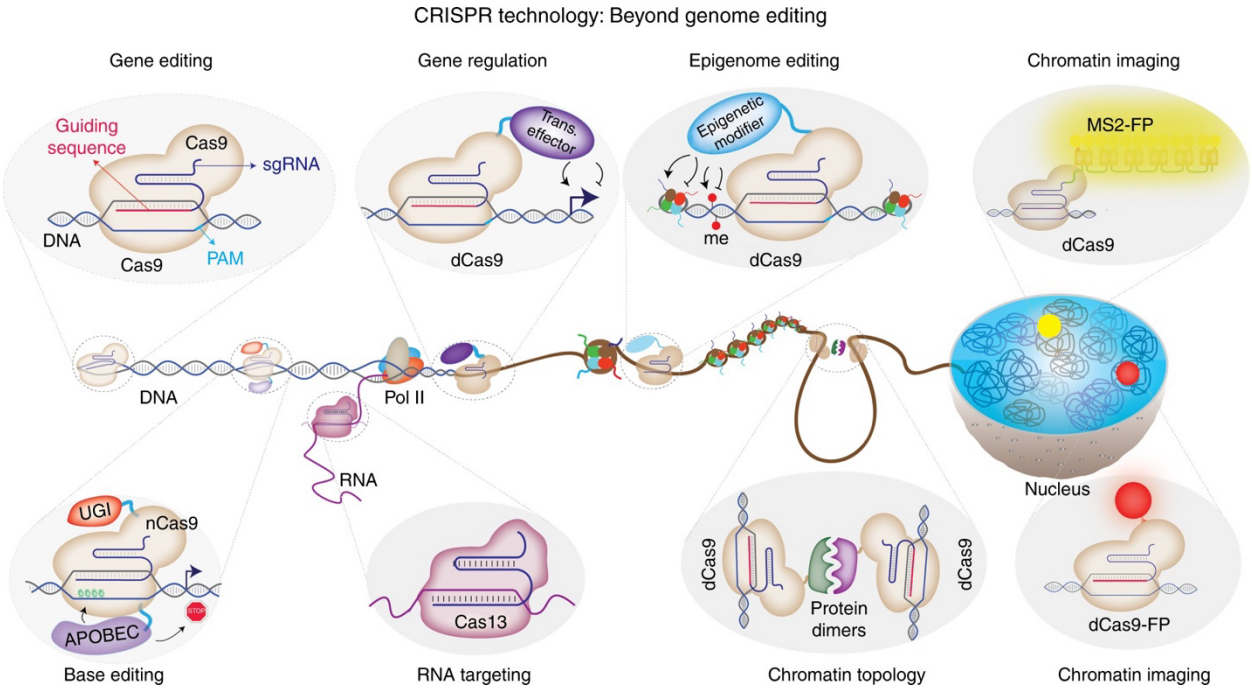


Figure 2: New applications based on CRISPR-Cas systems.
The CRISPR-Cas systems can be repurposed to achieve targeted gene regulation, epigenome editing, chromatin imaging, and chromatin topology manipulations[4].

In this thesis, research was performed to study the function of both the genes and the NCREs of the human genome in a systematic way, using different custom-designed CRISPR screening systems. In **Chapter 2**, aiming to understand the drug importing machinery, custom CRISPR CRISPR–Cas9 knockout (CRISPRko, to induce loss of function of endogenous genes) and CRISPR/dCas9 activation (CRISPRa, to up-regulate endogenous genes) libraries targeting all known transporters from the human genome were made to identify transporters of doxorubicin, one of the most commonly used chemotherapeutic drugs in the clinic. By directly monitoring the doxorubicin drug uptake as the screening readout, both previously confirmed drug exporters of doxorubicin such as ABCB1 and ABCG2 genes, and a novel doxorubicin importer SLC2A3 (GLUT3). The SLC2A3 gene could potentially serve as a new marker for doxorubicin response and for identifying subtypes of patients with tumors that could benefit from doxorubicin treatment.

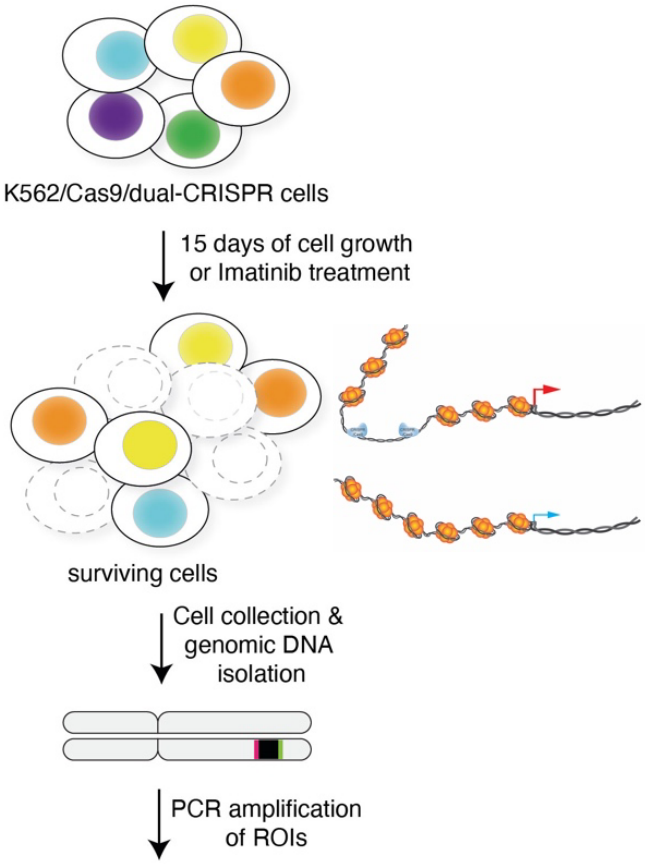


Figure 3: The dual-CRISPR libraries could be used to study different biological pathways.
A pair of guide RNAs within each individual cell will remove one target NCRE. Genomic DNA from different cell populations was then isolated, and the dual-CRISPR libraries could be amplified by direct PCR reactions, which would be ready for next-generation sequencing (NGS). The change in abundance of dual-guide-RNAs will then be calculated to identify potential hits according to the screening phenotype.

In **Chapter 3**, using a new dual-CRISPR screening system developed in our own group (Figure 3), we aim to understand the function of NCREs in cell survival, drug response, and cell differentiation. This system is able to introduce a designed pair of guide RNAs to remove an NCRE in the human genome in a single cell to study the function of the DNA fragment, and tests thousands of such NCREs in one screening. Custom-designed dual-CRISPR libraries targeting all ultraconserved elements in the human genome and all potential enhancer regions in K562 cells were made and used to study these NCREs in a systematic and genome-wide fashion for the first time. Multiple NCREs were identified and validated to affect cell survival or response to imatinib treatment in K562 cells. One ultraconserved element PAX6_Tarzan was also shown to affect cardiomyocyte differentiation using a human embryonic stem cell differentiation model, indicating a potential key function of this ultraconserved element in the evolution. The results from Chapter 3 also indicate that the dual-CRISPR system we developed could be a useful tool for the broad research community to study the functions of the non-coding genome.

References

1 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

2 Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304-1351 (2001).

3 Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022).

4 Adli, M. The CRISPR tool kit for genome editing and beyond. *Nature Communications* **9**, 1911 (2018).

5 Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol* **17**, 5-15 (2016).

6 Pang, B. & Snyder, M. P. Systematic identification of silencers in human cells. *Nature Genetics* **52**, 254-263 (2020).

7 Gregory, T. R. Synergy between sequence and size in Large-scale genomics. *Nature Reviews Genetics* **6**, 699 (2005).

8 Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M. B. Annotating non-coding regions of the genome. *Nature Reviews Genetics* **11**, 559 (2010).

9 Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics* **15**, 234 (2014).

10 Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5**, 1-1 (2012).

11 Pang, B., van Weerd, J. H., Hamoen, F. L. & Snyder, M. P. Identification of non-coding silencer elements and their regulation of gene expression. *Nature Reviews Molecular Cell Biology* **24**, 383-395 (2023).

12 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

13 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489** (2012).

14 Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90 (2012).

15 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108 (2012).

16 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).

17 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49 (2011).

18 Arnold, C. D. *et al.* Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* **339**, 1074-1077 (2013).

19 van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotech* **35**, 145-153 (2017).

20 Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A* **70**, 3240-3244 (1973).

21 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).

22 Johannessen, C. M. *et al.* A melanocyte lineage program confers resistance to MAP kinase pathway inhibition. *Nature* **504**, 138-142 (2013).
23 Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550-553 (2002).
24 Paul, P. *et al.* A Genome-wide Multidimensional RNAi Screen Reveals Pathways Controlling MHC Class II Antigen Presentation. *Cell* **145**, 268-283 (2011).
25 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011).
26 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
27 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
28 Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nature biotechnology* **34**, 192 (2016).
29 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84 (2014).
30 Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509** (2014).
31 Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299-311 (2015).
32 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology* **32**, 347-355 (2014).
33 Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M. B. Annotating non-coding regions of the genome. *Nat Rev Genet* **11**, 559-571 (2010).
34 Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* **34**, 192-198 (2016).
35 Tycko, J. *et al.* Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat Commun* **10**, 4063 (2019).
36 Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**, 510-517 (2015).

Chapter 2

Targeted CRISPR activation and knockout screenings identify novel doxorubicin transporters

Accepted by ***Cellular Oncology***

Targeted CRISPR activation and knockout screenings identify novel doxorubicin transporters

Yufeng Li[1], Minkang Tan[1], Shengnan Sun[1], Elena Stea[1], Baoxu Pang[1]*

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, Netherlands

*Correspondence: b.pang@lumc.nl (B.P.)

Abstract

Tissue-specific drug uptake has not been well studied, compared to the deeper understanding of drug resistance mediated by the cellular efflux system such as MDR1 proteins. It has been suggested that many drugs need active or defined transporters to pass the cell membrane. In contrast to efflux components induced after anti-cancer drugs reach the intracellular compartment, drug importers are required for initial drug responses. Furthermore, tissue-specific uptake of anti-cancer drugs may directly impact the side effects of many drugs when they accumulate in healthy tissues. Therefore, linking anti-cancer drugs to their respective drug import transporters would directly help to predict drug responses, whilst minimizing side effects. We designed and applied custom CRISPR activation and knockout libraries targeting all potential human transporters to identify drug transporters of the commonly used anti-cancer drug doxorubicin. Integrating the data from these comprehensive CRISPR screenings, we confirmed previously indicated doxorubicin exporters such as ABCB1 and ABCG2 genes, and identified novel doxorubicin importer gene SLC2A3 (GLUT3). The newly identified importers may have direct clinical implications for the personalized application of doxorubicin in treating distinct tumors.

Keywords

CRISPR Screening, Anti-cancer, Doxorubicin, Transporters

1 Introduction

Precisely treating the diseased cells or tissues whilst exhibiting minimal side effects on healthy bystander cells has always been the ultimate goal of medicine. This is especially challenging for chemotherapeutic drugs that treat cancer, as these drugs' effectiveness often coincides with notorious side effects. There can be several ways to reach this therapeutic window. First, by selecting drugs that could reach the target cells better, the total dose could be reduced to minimize side effects. Second, choosing drugs that do not affect normal tissues but only affect the target cells, exemplified by modern targeted therapy drugs like imatinib (Gleevec), which only targets a specific protein kinase in tumor cells [1], should decrease toxicity.

Although many drugs are very effective for different tumors, de novo resistance does exist. Much research has been done since the realization of drug resistance, and there are many mechanisms proposed to explain the drug resistance, such as up-regulation of drug transporters to pump out drugs, bypassing the original drug target by the tumor cells, or boosting the cell survival pathways [2-5]. However, to exert their effects, most drugs (except for antibody-based drugs) first need to reach the inside of the cell. It is generally assumed that most drugs pass the cell membrane by passive diffusion [6, 7]. However, recent research indicates that carrier-mediated uptake of drugs may be more common and important than we thought [6, 7]. But only limited systematic research has been done to define the uptake mechanisms of chemotherapeutic drugs, where SLC-specific CRISPR–Cas9 knockout genetic screening has been used to identify potential drug transporters [8]. However, many potential drug transporters are lowly expressed or not even expressed at all in a given cell type, where deleting these genes by CRISPR–Cas9 knockout genetic screenings will not be able to show a phenotype, therefore making it hard to evaluate potential drug transporters comprehensively.

To evaluate all potential drug transporters in a comprehensive way, we created focused CRISPR–Cas9 knockout (CRISPRko, to induce loss of function of endogenous genes) [9-12] and CRISPR/dCas9 activation (CRISPRa, to up-regulate endogenous genes) [13-16] libraries targeting all known transporters from the human genome. We integrated the results from the two screening systems to identify new

transporters for different drugs. Anthracycline family drugs are amongst the most widely used anti-cancer drugs and include several analogs that exert different effectivity in solid and non-solid tumors [17, 18]. The tissue distribution of these drugs is very different [19, 20], which often coincides with side effects observed in the clinic; therefore it is possible that different drug importers might play important roles in the efficacy and side effects of these drugs.

We performed CRISPRko and CRISPRa genetic screens to identify the potential transporters for doxorubicin, the most widely used anthracycline drug with strong side effects such as cardiotoxicity [18, 21-23]. Integrating the data from these comprehensive CRISPR screenings, we confirmed drug exporters for doxorubicin such as ABCB1 and ABCG2 genes, and identified novel doxorubicin importer gene SLC2A3 based on the CRISPRa genetic screens. Upregulation of SLC2A3 led to higher drug uptake and better cell killing, indicating SLC2A3 could be a new marker to predict doxorubicin drug response and minimize side effects for the personalized application of this conventional chemotherapeutic drug. These results also highlight the necessity of combining both CRISPRko and CRISPRa genetic screens for the identification of drug transporters.

2 Materials and methods

2.1 Cell lines and cell culture

K562 cells were cultured in RPMI 1640 (Gibco) supplemented with 10% fetal bovine serum (Biowest), L-Glutamine (Gibco), and 1% Penicillin-Streptomycin (Gibco), at 5% CO2 in a humidified 37°C incubator. 293T cells were cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (Biowest), L-Glutamine (Gibco), and 1% Penicillin-Streptomycin (Gibco). For stable dCas9-VP64-MS2-P65-HSF (K562/SAM) cell line generation, the wild-type K562 cell line was transduced with lentivirus generated using the Lenti MS2-P65-HSF1-Hygro plasmid at an MOI of 0.3 (Addgene #61426) and selected with 20 μg ml^{-1} hygromycin (InvivoGen) for at least 1 week. Then dCas-VP64 was transduced into K562-MS2-P65-HSF1 cell line, and selected with blasticidine at 8 μg ml^{-1}. Single cells were seeded in 96-well plates and expanded.

2.2 Reagents

Doxorubicin was obtained from Pharmachemie (the Netherlands). Aclarubicin was obtained from Santa Cruz (USA). Daunorubicin was obtained from Sanofi-Aventis (the Netherlands). Epirubicin was obtained from Accord Healthcare Limited (UK). Idarubicin was obtained from Pfizer (USA). All of these drugs were dissolved according to the manufacturer's formulation.

2.3 Design of the libraries

The RefSeq IDs of potential transporter genes were extracted from TransportDB 2.0 [24] and converted into gene symbols with DAVID [25].

For CRISPRko sgRNA library design, GUIDES web tool was used to target 952 potential transporter genes in the list, and on average 12 individual guides were prioritized by GUIDES for each gene [26]. Furthermore, 10 individual guides were designed to target 58 essential genes from K562 as positive controls and 400 human genome non-targeting sgRNAs from the GeCKO v2.0 library were added as negative controls [12, 27]. The designed CRISPR knockout library sequences (20nt protospacer) were inserted into the cloning sequence backbone: GGAAAGGACGAAACACCG-[20nt protospacer]-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC and sent for oligo pool synthesis.

For CRISPRa sgRNA library design, the previous guide designs targeting 300 bp upstream regions from the transcription start sites of 977 potential transporter genes in the list were selected [15]. On average, 7 individual guides were included for each gene. 400 human genome non-targeting sgRNAs from the GeCKO v2.0 library were added as negative controls. The designed CRISPR activation library sequences (20nt protospacer) were inserted into the cloning sequence backbone: GGAAAGGACGAAACACCg-[20nt protospacer]-GTTTTAGAGCTAGGCCAACATGAGGATCACCCATG and sent for oligo pool synthesis.

2.4 Construction of the CRISPRko and CRISPRa libraries

Construction of the libraries was performed as previously described [28]. Oligos with the CRISPRko and CRISPRa sgRNA (20nt protospacer) sequences and flanking cloning sequences were ordered as oligo pools from Genscript.

For the CRISPRko library, the synthesized single-strand oligo pool was directly assembled with the digested vector products by Gibson Assembly. The lentiCRISPRv2-puro plasmid (Addgene, Plasmid #98290) was digested with Esp3I (Anza™ 13) and dephosphorylated using FastAP (Thermo Scientific™, EF0651) for generating the CRISPRko library. The resulting linearized vector products were further size-selected and gel-purified using QIAGEN MinElute column and used for library cloning using Gibson Assembly. The assembly mix was made using 200 ng of digested lentiCRISPRv2-puro vector, 6 ng CRISPRko oligo pool (at molar ratio 1:10) and 10 µl of 2× Gibson Assembly Master Mix for a final volume of 20 µl. The assembly mix was incubated at 50 °C for 60 min, and in total 10 reactions were pooled for making the full library with good coverage. The Gibson Assembly mix was purified by isopropanol precipitation and resuspended in 5 µl water, from which 1 µl of the products were electroporated into 25 µl of Endura electrocompetent cells (Endura, 60242). Five individual electroporation reactions were pooled and grown in 5 ml recovery medium for 1 h. Then 5 µl media from the 5 ml LB culture was used to perform a serial dilution to determine the electroporation efficiency and thus the library coverage, which was aimed to be at least 300-fold. The rest was further cultured in 1000 ml LB medium with 100 µg ml^{-1} carbenicillin overnight. The plasmid libraries were extracted using the QIAGEN Maxiprep Kit and verified by next-generation sequencing.

For the CRISPRa library, the synthesized oligo library was first amplified by PCR using the following primer:

CRISPRa- Forward primer: GTAACTTGAAAGTATTTCGATTTCTT GGCTTTATATATCTTGTGGAAAGGAC GAAACACC;

CRISPRa- Reverse primer: ATTTTAACTTGCTAGGCCCTGCAGAC ATGGGTGATCCTCATGTTGGCCTAGC TCTAAAAC;

PCR procedures using NEBNext High-Fidelity 2X PCR Master Mix were: 98 °C for 30 s, 15 cycles of 98 °C for 10 s, 68 °C for 30 s and 72 °C for 30 s, and 72 °C for 5 min. For each reaction, 10 ng of the oligo pool was used for a 100 µl PCR reaction, and 20 reactions per library were pooled. The pooled PCR products were further size-selected

and gel-purified using QIAGEN MinElute column. The lenti-sgRNA(MS2)-Puro plasmid (Addgene, Plasmid #73797) was digested with Esp3I (Anza™ 13) and dephosphorylated using FastAP (Thermo Scientific™, EF0651). The resulting linearized vector products were further size-selected and gel-purified using QIAGEN MinElute column and used for library cloning using Gibson Assembly. The assembly mix was made using 200 ng of digested lenti sgRNA(MS2)-Puro plasmid vector, 11.5 ng insert DNA (at molar ratio 1:10) and 10 µl of 2× Gibson Assembly Master Mix for a final volume of 20 µl. The downstream library construction and production steps were the same as for the CRISPRko library.

2.5 Lentivirus production and MOI determination

For generating MS2-P65-HSF1 (MPH), dCas9-VP64 and lentiSAMv2-Puro virus, the 293T cells were grown in 6 well plates. When cells reached 50% confluency, they were transfected with the mix of 1 µg of Lenti MS2-P65-HSF1-Hygro plasmid or dCas9-VP64-blast plasmid or lentiSAMv2-Puro plasmid, 0.6 µg of psPAX2, 0.4 µg of pCMV-VSV-G, and 6 µg PEI (Polyscience, 23966) in 50 µl of serum-free medium. Media supernatant containing the virus particles was collected on the third day after transfection.

For packaging lentivirus of the screening libraries, the 293T cells were grown in 15 plates of 15 cm^2 dishes for each library. For each dish, cells of 50% confluency were transfected with the mix of 30 µg of library plasmids, 20 µg of psPAX2, 10 µg of pCMV-VSV-G, and 180 µg PEI in 1 ml of serum-free medium. Media were refreshed the day following transfection. The medium supernatant containing virus particles was collected on the second and third days after transfection. Pooled medium supernatant was further concentrated using ultracentrifugation (4°C, 10,000g, 2 h). The virus titer was determined by making serial dilutions (10^{-3} to 10^{-10}) of 10 µl of frozen virus supernatant to infect 293T cells. Two days after infection, cells were selected with 2 µg ml^{-1} puromycin for an additional 7 days. The virus titer was then calculated based on the survival colonies and the related dilution.

2.6 Pooled CRISPRko and CRISPRa Screen

The K562 cells were transduced with the lentiviral CRISPRko library, and K562/SAM stable cells were transduced with the lentiviral CRISPRa libraries respectively by spin

infection to achieve an initial library coverage of at least 500-fold at an MOI of ~ 0.3. For spin-infection, 3×10^6 cells in each well of a 12-well plate were infected in 1 ml of medium containing 8 μg ml^{-1} of polybrene and the virus. In total, four 12-well plates were used for each screening to infect a total of 1.5×10^8 cells. Then, 48 h post-infection, successfully transduced cells were selected with 2 μg ml^{-1} puromycin (Biomol) for a further 3 days. Cells were harvested and dead cells were then removed with Histopaque-1077 (Sigma) by centrifuging cells at 400g for 30 minutes at room temperature. After live cell enrichment, an aliquot of 10^8 cells infected with either the CRISPRko or CRISPRa libraries were frozen as the control population. From the same cell population, another aliquot of 10^8 K562 cells was treated with 2 μM doxorubicin for 2 h. Then cells were sorted based on their fluorescence intensity as a surrogate for direct doxorubicin uptake. In total 2×10^6 cells that fell in the lowest 10% fluorescence intensity (L10) and highest 10% fluorescence intensity (H10) were FACS sorted respectively, and were frozen for downstream processing. For each biological replicate experiment, lentivirus was made freshly, and the infection and doxorubicin treatment were repeated.

2.7 Genomic DNA extraction, library preparation, and sequencing

Genomic DNA was isolated from frozen cell pellets using the QIAmp DNA Blood and Tissue Maxi Kit (Qiagen). The sgRNA spanning region was amplified from purified genomic DNA, with primers containing the Illumina adapters and indices using the one-step PCR reaction. For each 100 μl PCR reaction, 10 μg of genomic DNA, 50 μl of 2× NEBNext High-Fidelity master mix, 2 μl of 10 μM forward and reverse primers, respectively were used [15]. In total 5 PCR reactions with different pairs of primer pairs were used and pooled, assaying in total of 50 μg of genomic DNA. PCR procedures were: 98 °C for 30 s, 19 cycles of 98 °C for 10 s, 68 °C for 30 s and 72 °C for 30 s, and 72 °C for 5 min. The PCR products were further size-selected and gel-purified using the QIAGEN MinElute column and confirmed using a High Sensitivity Bioanalyzer DNA Kit (Agilent). Then the libraries were sequenced on the Illumina HiSeq4000 platform.

2.8 Pooled CRISPRko and CRISPRa Screen analysis

Cutadapt 3.4 was used to extract the unique 20nt protospacer sequences from Illumina paired-end reads by trimming out the U6 promoter overlapping sequence from the 5' end and the scaffold sequence from the 3' end. For the CRISPRko sgRNA

library, the following sequences were used U6 promoter: ATCTTGTGGAAAGGACGAAACACCG, scaffold: GTTTTAGAGCTAGAAATAGC. For the CRISPRa sgRNA library, the following sequences were used U6 promoter: ATCTTGTGGAAAGGACGAAACACCG, scaffold: GTTTTAGAGCTAGGCCAACA. Trimmed reads were then aligned to the custom screen libraries using BWA [29]. To quantify each sgRNA abundance, from the same pair-end read, read with MAPQ score over 30 on one end and MAPQ score at least over 10 on the other end was accounted as valid. MAGeCK RRA was used to identify the sgRNAs of enriched/depleted significantly from comparisons between the FACS sorted highest 10% fluorescence intensity (H10) cells or lowest 10% fluorescence intensity (L10) cells and unsorted control cell population [30]. The sgRNA enrichment/depletion performance was further aggregated to identify positively/negatively selected genes in the comparison robustly.

2.9 Generation of clones

Single-guide RNAs targeting the promoter regions of ABCB1 and SLC2A3 genes were selected from the pooled CRISPRa screen analysis. The guide RNA sequence was cloned into lentiSAMv2-Puro plasmid containing the gRNA scaffold and dCas9 sequence, and lentivirus was made as previously described. Then K562/SAM stable cells were transduced with the virus containing the respective guide RNAs and then selected using puromycin (2 μg ml^{-1}). Single clones of cells were picked and verified using PCR and Sanger sequencing. The sequences of the gRNAs are listed in Supplementary Table 1.

2.10 FACS analysis of drug uptake

The K562 ABCB1-CRISPRa, K562 SLC2A3-CRISPRa clones and 293T SLC2A3-CRISPRa bulk cells were treated with the respective drugs at 2 uM for 2 h, and then analyzed by FACS as the indication of drug uptake.

2.11 Cell viability assay

The K562 ABCB1-CRISPRa and K562 SLC2A3-CRISPRa clones were seeded in 96-well plates at the density of 5×10^4 cells per well. Then, doxorubicin or other anthracycline drugs were added to the cells at a series of concentrations indicated. After 72 h of treatment, CellTiter-Blue (Promega) was used to quantify the cell viability according to the manufacturer's protocol.

2.12 Western blot

Cells were washed with cold PBS and lysed in 1X RIPA buffer (Thermo Scientific) containing the protease inhibitor cocktail (Roche). Then the cell lysate was sonicated and the protein concentration in the supernatant was determined using the Pierce BCA protein assay kit (Thermo Scientific). Equal amounts of proteins were separated via 8% SDS-PAGE. Proteins were transferred to PVDF membranes, blocked with 5% nonfat milk in PBST for 1 h at room temperature, probed with primary antibodies against ABCB1 (CST 13342S, 1:1000), SLC2A3 (Abcam ab191071, 1:1000), and Vinculin (Merck V9131, 1:5000; as the loading control) overnight at 4°C. Membranes were washed with PBST three times, incubated with the appropriate secondary antibodies (1:10000 dilution) for 1 h at room temperature, and then washed with PBST three times. The resulting signal was visualized by Bio-Rad ChemiDoc Imaging System using the Pierce enhanced chemiluminescence (ECL) Plus Western Blotting Substrate (Thermo Scientific).

2.13 Quantitative PCR

Total RNA was isolated from 10^6 cells using ISOLATE II RNA Mini Kit (Bioline, BIO-52073) according to the manufacturer's instructions. The cDNAs were synthesized from 500 ng of total RNA using a SuperScript™ IV VILO™ Master Mix (Invitrogen, 11756050), and were analyzed using SensiFAST SYBR No-ROX Kit (Bioline, BIO-98020) with primers for ABCB1, SLC2A3, and GAPDH (a housekeeping gene) on Biorad CFX Opus 384 Real-Time PCR Systems system. The sequences of the primers are listed in the Supplementary Table 1.

2.14 Statistical analyses

For determining the significance of differences in the experimental data, the Student's t-test (two-tailed), one-way analysis of variance, and the Mann–Whitney U test were performed using GraphPad Prism version 8 software. P values less than 0.05 were considered to be significant.

3 Results

3.1 Construction of custom CRISPRko and CRISPRa libraries

An unbiased way to search for drug transporters in a systematic and comprehensive way is to use a genome-wide screening system, such as insertional mutagenesis, CRISPRko, or CRISPRa genetic screenings [9, 10, 31, 32]. However, such screenings are intrinsically noisy [33]. If other confounding phenotypes are dominant in the screen readout, the drug transporters may not be identified during such screenings, since the drug might also be able to diffuse into the cells [10, 34]. As an alternative, we constructed a custom and dedicated CRISPRko library (for gene silencing, which would help to identify expressed transporters that determine the drug uptake in different cell types), as well as CRISPRa library (for gene activation, which would identify both expressed and not expressed potential transporters that may be involved in the drug uptake), specifically targeting the known transporters and transporter associated proteins (TransportDB 2.0) [24]. In this way, focused screening systems were made to study all potential transporters, irrespective of their expression level in certain cell lines (Supplementary Fig. 1). In total 1313 potential importers from TransportDB 2.0 were targeted for design, and the respective CRISPR knockout (952 genes were targeted) and activation libraries (977 genes were targeted) were generated. For the transporter CRISPRko library, 12 guides per gene were designed to inactivate the targeting transporters based on published algorism [26]. For the transporter CRISPRa library, the 300 bp upstream regions of the transcription start sites of the 977 potential importers were used to design the CRISPR guide RNAs [15]. The two libraries were then assembled according to the protocol [33] (Fig. 1A).

Doxorubicin, like many other anthracycline drugs, is autofluorescent [17], which allows the monitoring of direct drug transport by fluorescence-activated cell sorting (FACS). K562 cells either infected with the transporter CRISPRko or CRISPRa libraries were treated with 2 uM doxorubicin for 2 hours. Then cells were sorted based on their fluorescence intensity as a surrogate for direct doxorubicin uptake. Cells that fell in the lowest 10% fluorescence intensity (L10) and highest 10% fluorescence intensity (H10) were FACS sorted, respectively, and subjected to downstream processing (Fig. 1B). Genomic DNA from the respective populations was isolated, and CRISPR guide RNA sequences were PCR amplified and subjected to next-generation sequencing. The

diversity and abundance of the guide RNAs in different FACS-sorted populations were compared to the initial population before the FACS sort, and enriched and depleted guide RNAs were calculated using MAGeCK (Fig. 1C) [35].

Designed CRISPRko and CRISPRa libraries were first verified by sequencing. We recovered 98.9% of the designed guide RNAs from CRISPRko library (Fig. 2A), and 100% of the designed guide RNAs from CRISPRa library (Fig. 2B). In the CRISPRko library, on average 12 guides per gene were recovered (Fig. 2C), and in the CRISPRa library at least 5 guides were recovered to target each gene promoter for the majority of the target genes (Fig. 2D). For each sequencing sample, an average 500X and 200X sequence depth per guide were reached for CRISPRko (Fig. 2E) and CRISPRa (Fig. 2F) screenings respectively. The correlations of all the samples from both CRISPRko and CRISPRa screenings were good (Supplementary Fig. 2).

3.2 Screening direct doxorubicin transport using CRISPRko and CRISPRa libraries

In the CRISPRko screening, the enriched genes in the L10 population represent the potential membrane proteins involved in drug importing (Fig. 3A, top-right corner in red and Supplementary Table 2). More than 10 potential membrane proteins involved in drug import were identified. The top hit is gene ASNA1, an ATPase and a component of transmembrane domain (TMD) recognition complex (TRC) that is involved in the post-translational delivery of tail-anchored (TA) proteins from the cytosol to the endoplasmic reticulum (ER) [36]. On the contrary, the depleted genes in the L10 population represent the 10 potential membrane proteins involved in drug exporting (Fig. 3A, lower-left corner in blue and Supplementary Table 3). Multiple ATPases were among the top hits. In addition, the well-known multi-drug resistance exporter gene ABCB1 which is frequently identified in different genetic screenings for drug resistance [2, 34], was also identified. The ABCB1 gene was not ranked the highest, possibly because the screening was done under a short period of drug exposure and the ABCB1 gene needs to be upregulated in response to drug exposure.

When the H10 population from the CRISPRko screening was analyzed, similar groups of genes were identified. Here the enriched genes represent 44 potential membrane proteins involved in drug exporting (Fig. 3B, top-right corner in red and Supplementary

Table 3). It is interesting to see that gene ATP8B2 ranked amongst the top hits in this population, whilst also ranked within the top hits of the potential membrane proteins involved in drug exporting in L10 population (Figure 3A, lower-left corner in blue). The depleted genes in the H10 population represent 42 potential membrane proteins involved in drug importing (Fig. 3B, lower-left corner in blue and Supplementary Table 2), which also includes gene ASNA1 among the top hits. This gene was also enriched as the top hit involved in drug importing in the L10 group (Figure 3A, top-right corner in red). The fact that many same gene hits were indicated for the same potential function from different populations suggests that the CRISPRko screening was robust and would reliably identify potential proteins that are involved in doxorubicin transport.

Lowly expressed genes are not feasible to be studied using CRISPR knockout. On the contrary, CRISPR activation would allow the identification of lowly expressed genes which would play a role in drug transport; therefore we expect that additional novel genes would be identified compared to the CRISPRko screening. From the CRISPRa screenings with a similar experimental set-up, the enriched genes in the L10 population represent the 4 potential membrane proteins involved in drug exporting (Fig. 3C, top-right corner in red and Supplementary Table 3). Again the well-known drug transporter genes ABCG2 and ABCB1 were significantly enriched [2, 34, 37], among some other genes such as genes SCL35G5 and ATP8B4, indicating upregulation of these genes led to less drug accumulation. On the other side, the SLC2A3 gene, among some other solute carrier transporter genes, were depleted in the L10 population, suggesting these may be the drug importers (Fig. 3C, lower-left corner in blue and Supplementary Table 2).

From the CRISPRa screenings, the enriched genes in the H10 population would indicate 4 potential genes involved in drug importing. The top hit, among some other solute carriers, was the SLC2A3 gene (Figure 3D, top-right corner in red and Supplementary Table 2), confirming the results from similar analyses in a different population (Fig. 3C, lower-left corner in blue). The top depleted hits from H10 population, which indicates drug exporting roles, were the ABCG2 and ABCB1 genes (Fig. 3D, lower-left corner in blue and Supplementary Table 3), which were also seen in similar analyses in the L10 population (Fig. 3C, top-right corner in red). All these

mutually confirming data indicate that the CRISPRa screening identified reliable hits involved in drug exporting and importing.

We further grouped all the potential genes involved in drug exporting identified from CRISPRko and CRISPRa screenings respectively, and compared them together. Only the ABCB1 gene was jointly identified from the two different types of screenings (Fig. 3E). A similar analysis was also performed integrating the potential genes involved in drug importing from both screenings. Only two solute carrier genes SLC16A1 and SLC18B1 appeared from both screening methods (Fig. 3F). These data indicate that the CRISPRko and CRISPRa screening methods complement each other and would identify different hits in the drug transport setup.

3.3 Drug accumulation regulated by the ABCB1 gene during transient drug exposure

To show that our proposed experimental setup would reliably identify players in drug transport, the ABCB1 gene, one of the top hits from both screenings, was chosen for further analyses. Despite the fact that the ABCB1 gene has been proposed to play a role in drug exporting, the majority of the genetic perturbation screenings were done during a long period of drug exposure [34], making it interesting to see the effect of ABCB1 manipulation during transient drug exposure. As the ABCB1 gene is lowly expressed before drug exposure, we chose to use the CRISPRa system to upregulate the gene expression (Fig. 4A). Two guide RNAs were designed to target the promoter region of the ABCB1 gene, and more than 15-fold upregulation of gene expression was achieved as measured by qPCR (Fig. 4A). We also confirmed that the protein level of the ABCB1 gene was elevated by the two respective CRISPRa guide RNAs (Fig. 4B).

As the result of the upregulation of the ABCB1 gene, the direct uptake of doxorubicin monitored by FACS was reduced by 50% in the two respective clones (Fig. 4C). It is interesting to see that not only the direct uptake of doxorubicin, but that of other clinically used anthracycline drugs such as daunorubicin, epirubicin and idarubicin was also reduced, except for aclarubicin (Fig. 4C and Supplementary Fig. 3). These data indicate that aclarubicin would serve as an alternative anthracycline to overcome the drug resistance caused by the upregulation of the ABCB1 gene. To further confirm

that the reduced uptake of doxorubicin measured by FACS was directly caused by CRISPRa-mediated upregulation of ABCB1, a specific ABCB1 inhibitor tariquidar was used to treat the CRISPRa clones. No reduction of doxorubicin uptake was observed in tariquidar-treated clones expressing CRISPRa activating the ABCB1 gene (Fig. 4D), suggesting that the change of doxorubicin transport was not induced by the off-target effects of CRISPRa system. As a result, K562 clones with CRISPRa-mediated upregulation of ABCB1 gene became more resistant to doxorubicin and other anthracycline drugs compared to the parental K562 cells (Fig. 4E and Supplementary Fig. 4).

3.4 SLC2A3 gene serving as a novel doxorubicin importer and response marker

Only limited research has been done to identify drug importers of doxorubicin over the past several decades, and the contribution of these factors to drug response is still obscure [38]. Recently, a CRISPRko screening effort has been attempted to identify potential importers for many drugs, including doxorubicin; however the screening readout was based on cell survival but not direct drug uptake, and the CRISPRko screening system lacks the resolution to identify lowly expressed genes as potential functional hits[8]. Therefore, CRISPRa screenings that upregulate potential genes may identify additional and novel hits that were overlooked before, as we also observed (Fig. 3F). SLC2A3 gene was the top hit as a drug importer of doxorubicin from our CRISPRa screenings, which studied doxorubicin uptake directly (Fig. 3C and D). SLC2A3 encodes the solute carrier transmembrane glucose transporter 3 (GLUT3) that binds glucose and facilitates glucose uptake [39]. GLUT3 is expressed in neuronal tissues, as well as heart and white blood tissues [40]. Some tumors such as glioblastoma and triple-negative breast cancer may rely on the overexpression of GLUT3 and hence have an addition to it, making it a vulnerability drug target [41, 42].

To confirm the role of SLC2A3 in mediating doxorubicin uptake, we generated 4 independent clones using 2 different guide RNAs targeting the promoter region of SLC2A3 using CRISPRa system. A robust upregulation of SLC2A3 was achieved from all these clones, as measured by qPCR (Fig. 5A). As a result, the protein levels of SLC2A3 were also increased (Fig. 5B). It is noteworthy that the protein level of SLC2A3 was very low in the parental K562 cells, pointing out that screening strategies

to knockdown or knockout this gene would not be possible to identify this gene as a potential doxorubicin importer.

We then monitored the direct uptake of doxorubicin of these clones, as measured by FACS. Up to 50% increase in doxorubicin accumulation was observed in these clones compared to the parental K562 cells when cells were transiently exposed to doxorubicin (Fig. 5C and Supplementary Fig. 5). SLC2A3-mediated doxorubicin uptake was also observed in 293T cells, indicating a general role of this gene in doxorubicin importing (Supplementary Fig. 6). We further determined the effects of the increase of drug accumulation, and observed a significantly elevated cell killing in the SLC2A3 upregulating clones compared to the parental K562 cells when exposed to doxorubicin (Fig. 5D), suggesting a potentially important role of SLC2A3 in tumors responding to doxorubicin treatment. Indeed, in some of the AML patients, the expression level of SLC2A3 is higher, and when the contribution of SLC2A3 was considered in the overall survival of AML patients from the TARGET study, potential better survival effects were seen in patients with higher SLC2A3 expression (Fig. 5E) [43]. Because only a subgroup of these AML patients might have received doxorubicin treatment, it is possible that the contribution of SLC2A3 benefiting patients' response to the doxorubicin-containing regimen is even stronger. At the same time, these data also suggest that the SLC2A3 gene could be a predictive marker of tumors responding to doxorubicin treatment. The expression level of SLC2A3 is higher in tumors such as kidney renal clear cell carcinoma (KIRC), pancreatic adenocarcinoma (PAAD), and testicular germ cell tumors (TGCT), compared to the matching healthy tissues (Fig. 5F) [44], indicating these patients with higher SLC2A3 expression might also benefit from doxorubicin containing regimen. Indeed, previous studies showed that patients with testicular germ cell tumors may benefit from doxorubicin treatment in clinical trials [45], and future clinical trials may be warranted to test doxorubicin in treating subgroups of patients with tumors expressing elevated SLC2A3.

4 Discussion

The uptake and export of drugs by different cells and tissues represent one of the first factors affecting drug responses. The contribution of drug exporters, exemplified by ABCB1 in multi-drug resistance, has been characterized for a long time. In preclinical studies for new drugs, testing the effect of these exporters on drugs is also one of the requirements for subsequent filing of clinical trials, highlighting the importance of drug transporters in drug development. However, so far, systematic studies on the uptake of different drugs have been very limited, and often cell survival rather than direct uptake was used as the readout [8]. Furthermore, only CRISPR-mediated knockout was used to identify potential transporters, limiting the resolution to identify potential drug importers with lower expression in a defined cell type. We designed custom CRISPRko (for gene knockout) and CRISPRa (for gene activation) libraries targeting all potential membrane-associated transporters and proteins; furthermore, we monitored the direct uptake of drugs by FACS as the screening readout for the identification of transporters/proteins directly involved in drug uptake. Using this rationale, we were able to identify both known drug export transporter genes such as ABCB1 and ABCG2, and novel doxorubicin import transporter gene SLC2A3 for a broadly used chemotherapeutic drug doxorubicin. In addition, we found that CRISPRa screening would identify novel drug transporters, which would be missed by CRISPRko methods. This is especially important, as often genetic screenings are performed in one or limited cell or tissue types, in which the expression level of many genes is low (Supplementary Fig. 1). The upregulation of the novel doxorubicin importer SLC2A3 would directly increase the uptake of doxorubicin in K562 and 293T cells (Fig. 5C and Supplementary Fig. 6), and result in the better killing of these tumor cells. These data indicate that SLC2A3 could be a potential marker for identifying patients who may benefit from doxorubicin treatment. Furthermore, as GLUT3, the protein product of SLC2A3, is also present in heart tissues [46], and heart damage is still one of the major side effects associated with doxorubicin [19], SLC2A3 (GLUT3) may also contribute to doxorubicin-mediated cardiotoxicity. GLUT3 could be a novel target to reduce cardiotoxicity.

Contributions

B.P. conceived and supervised the study. Y.L. designed and performed the experiments, and analyzed the data, with the help of S.S and E.S. M.T. performed the computational analysis. Y.L., M.T., S.S. and B.P. wrote the manuscript.

Funding

Y.L. and S.S. were supported by the CSC scholarships. The research work was supported by the Gisela Their fellowship and the KWF Young Investigator Grant from the Dutch Cancer Society (11707) awarded to B.P.

Competing interests

The authors declare that they have no competing interests.

References

1. Cohen, P., D. Cross, and P.A. Jänne, *Kinase drug discovery 20 years after imatinib: progress and future directions.* Nature Reviews Drug Discovery, 2021. **20**(7): p. 551-569.
2. Wijdeven, R.H., et al., *Old drugs, novel ways out: Drug resistance toward cytotoxic chemotherapeutics.* Drug Resistance Updates, 2016. **28**(Supplement C): p. 65-81.
3. Borst, P., et al., *The multidrug resistance protein family.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 1999. **1461**(2): p. 347-357.
4. Prahallad, A., et al., *Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR.* Nature, 2012. **483**.
5. Sun, C., et al., *Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma.* Nature, 2014. **508**(7494): p. 118-122.
6. Sugano, K., et al., *Coexistence of passive and carrier-mediated processes in drug transport.* Nat Rev Drug Discov, 2010. **9**(8): p. 597-614.
7. Kell, D.B. and S.G. Oliver, *How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion.* Frontiers in Pharmacology, 2014. **5**.
8. Girardi, E., et al., *A widespread role for SLC transmembrane transporters in resistance to cytotoxic drugs.* Nature Chemical Biology, 2020. **16**(4): p. 469-478.
9. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-87.
10. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-4.
11. Zhou, Y., et al., *High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells.* Nature, 2014. **509**.
12. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-784.

13. Gilbert, Luke A., et al., *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-661.
14. Qi, L.S., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**: p. 1173-1183.
15. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-588.
16. Maeder, M.L., et al., *CRISPR RNA-guided activation of endogenous human genes.* Nat Methods, 2013. **10**(10): p. 977-9.
17. Pang, B., et al., *Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin.* Nature Communications, 2013. **4**: p. 1908.
18. Pang, B., et al., *Chemical profiling of the genome with anti-cancer drugs defines target specificities.* Nat Chem Biol, 2015. **11**(7): p. 472-480.
19. Qiao, X., et al., *Uncoupling DNA damage from chromatin damage to detoxify doxorubicin.* Proceedings of the National Academy of Sciences, 2020: p. 201922072.
20. Robert, J. and L. Gianni, *Pharmacokinetics and metabolism of anthracyclines.* Cancer Surv, 1993. **17**: p. 219-52.
21. Volkova, M. and R. Russell, 3rd, *Anthracycline cardiotoxicity: prevalence, pathogenesis and treatment.* Curr Cardiol Rev, 2011. **7**(4): p. 214-20.
22. Shan, K., A.M. Lincoff, and J.B. Young, *Anthracycline-induced cardiotoxicity.* Ann Intern Med, 1996. **125**(1): p. 47-58.
23. Pai, V.B. and M.C. Nahata, *Cardiotoxicity of chemotherapeutic agents: incidence, treatment and prevention.* Drug Saf, 2000. **22**(4): p. 263-302.
24. Elbourne, Liam D H., et al., *TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life.* Nucleic Acids Research, 2017. **45**(Database issue): p. D320-D324.
25. Sherman, B.T., et al., *DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update).* Nucleic Acids Res, 2022. **50**(W1): p. W216-w221.
26. Meier, J.A., F. Zhang, and N.E. Sanjana, *GUIDES: sgRNA design for loss-of-function screens.* 2017. **14**(9): p. 831-832.
27. Wang, T., et al., *Identification and characterization of essential genes in the human genome.* Science, 2015. **350**(6264): p. 1096-101.
28. Joung, J., et al., *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening.* Nature Protocols, 2017. **12**(4): p. 828-863.
29. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
30. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biol, 2014. **15**(12): p. 554.
31. Carette, J.E., et al., *Global gene disruption in human cells to assign genes to phenotypes by deep sequencing.* Nat Biotech, 2011. **29**(6): p. 542-546.
32. Shalem, O., N.E. Sanjana, and F. Zhang, *High-throughput functional genomics using CRISPR-Cas9.* Nat. Rev. Genet., 2015. **16**: p. 299-311.
33. Joung, J., et al., *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening.* Nat. Protocols, 2017. **12**(4): p. 828-863.
34. Wijdeven, R.H., et al., *Genome-Wide Identification and Characterization of Novel Factors Conferring Resistance to Topoisomerase II Poisons in Cancer.* Cancer Research, 2015. **75**(19): p. 4176-4187.

35. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biology, 2014. **15**(12): p. 554.
36. Stefanovic, S. and R.S. Hegde, *Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER.* Cell, 2007. **128**(6): p. 1147-1159.
37. Pang, B. and M.P. Snyder, *Systematic identification of silencers in human cells.* Nature Genetics, 2020. **52**(3): p. 254-263.
38. Okabe, M., et al., *Characterization of the organic cation transporter SLC22A16: a doxorubicin importer.* Biochem Biophys Res Commun, 2005. **333**(3): p. 754-62.
39. Simpson, I.A., et al., *The facilitative glucose transporter GLUT3: 20 years of distinction.* Am J Physiol Endocrinol Metab, 2008. **295**(2): p. E242-53.
40. Ziegler, G.C., et al., *Cellular effects and clinical implications of SLC2A3 copy number variation.* Journal of Cellular Physiology, 2020. **235**(12): p. 9021-9036.
41. Cosset, É., et al., *Glut3 Addiction Is a Druggable Vulnerability for a Molecularly Defined Subpopulation of Glioblastoma.* Cancer Cell, 2017. **32**(6): p. 856-868.e5.
42. Tsai, T.H., et al., *Overexpression of GLUT3 promotes metastasis of triple-negative breast cancer by modulating the inflammatory tumor microenvironment.* J Cell Physiol, 2021. **236**(6): p. 4669-4680.
43. Dwivedi, B., et al., *Survival Genie, a web platform for survival analysis across pediatric and adult cancers.* Scientific Reports, 2022. **12**(1): p. 3069.
44. Tang, Z., et al., *GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis.* Nucleic Acids Res, 2019. **47**(W1): p. W556-w560.
45. Lederman, G.S. and M.B. Garnick, *Possible benefit of doxorubicin treatment in patients with refractory germ cell cancer.* Cancer, 1986. **58**(11): p. 2393-8.
46. Grover-McKay, M., S.A. Walsh, and S.A. Thompson, *Glucose transporter 3 (GLUT3) protein is present in human myocardium.* Biochim Biophys Acta, 1999. **1416**(1-2): p. 145-54.

Fig.1

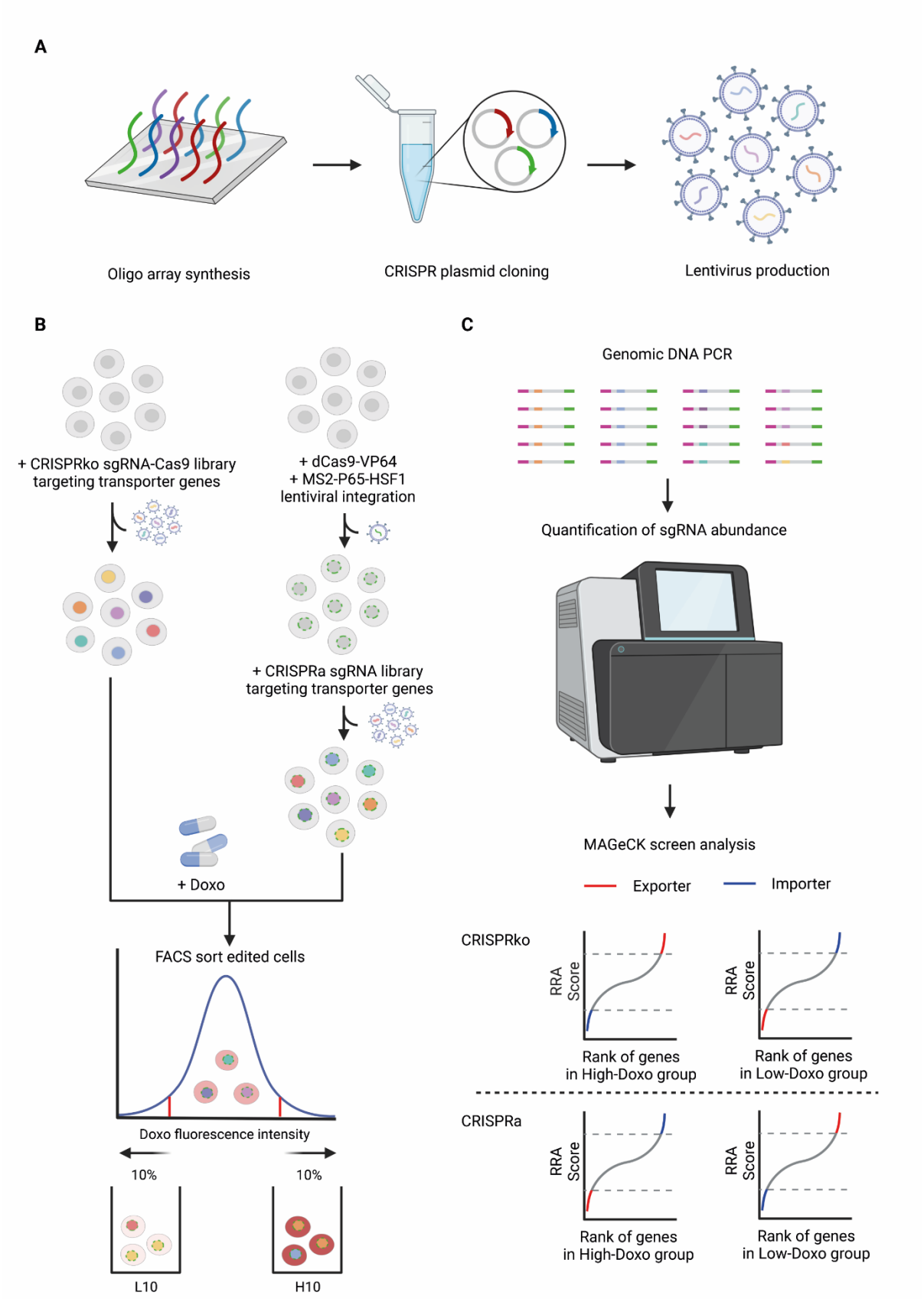

A
Oligo array synthesis
CRISPR plasmid cloning
Lentivirus production
B
+ CRISPRko sgRNA-Cas9 library targeting transporter genes
+ dCas9-VP64 + MS2-P65-HSF1 lentiviral integration
+ CRISPRa sgRNA library targeting transporter genes
+ Doxo
FACS sort edited cells
Doxo fluorescence intensity
10%
10%
L10
H10
C
Genomic DNA PCR
Quantification of sgRNA abundance
MAGeCK screen analysis
Exporter
Importer
CRISPRko
RRA Score
Rank of genes in High-Doxo group
RRA Score
Rank of genes in Low-Doxo group
CRISPRa
RRA Score
Rank of genes in High-Doxo group
RRA Score
Rank of genes in Low-Doxo group

Fig.2

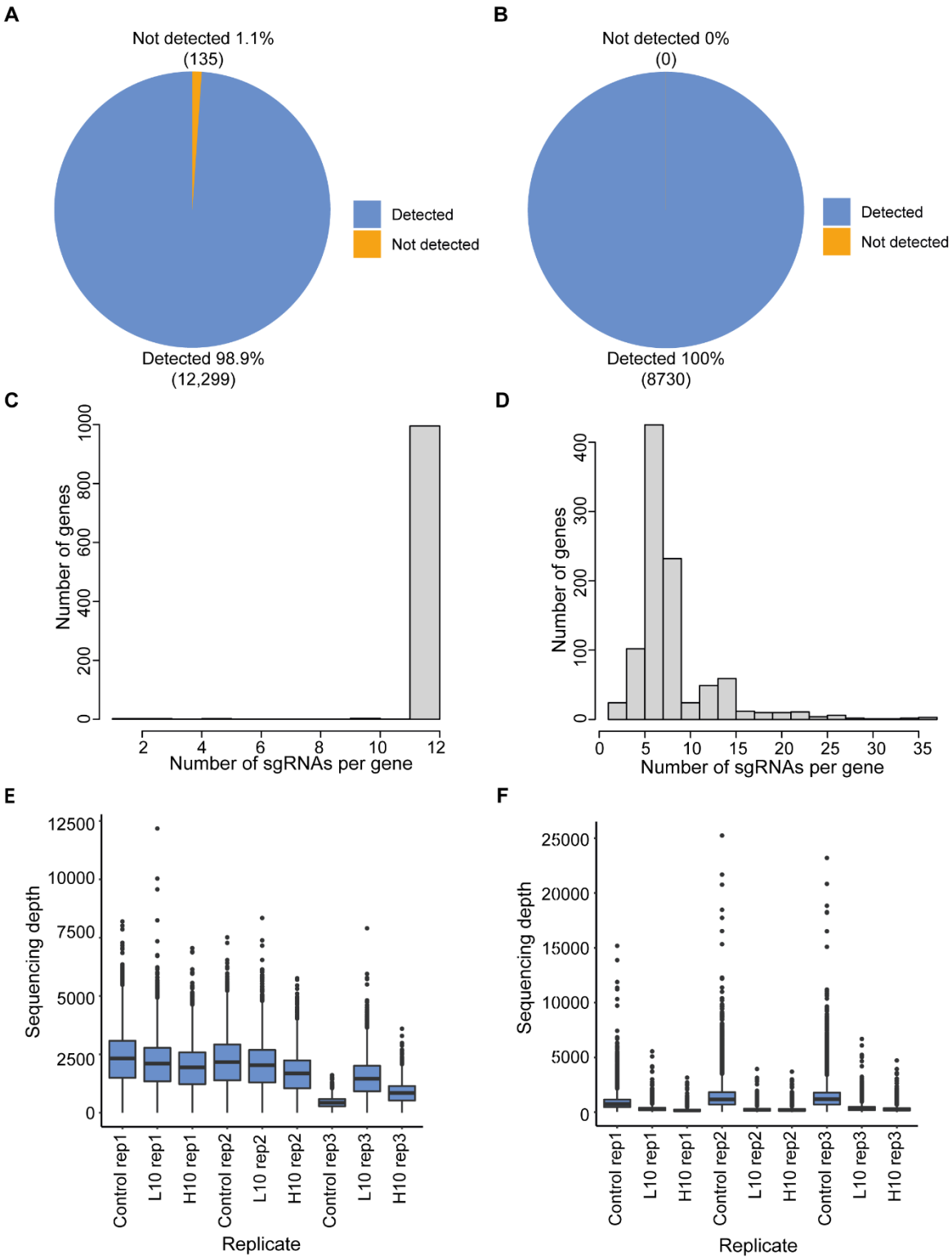

A
Not detected 1.1%
(135)
Detected
Not detected
Detected 98.9%
(12,299)
B
Not detected 0%
(0)
Detected
Not detected
Detected 100%
(8730)
C
Number of genes
Number of sgRNAs per gene
D
Number of genes
Number of sgRNAs per gene
E
Sequencing depth
Control rep1
L10 rep1
H10 rep1
Control rep2
L10 rep2
H10 rep2
Control rep3
L10 rep3
H10 rep3
Replicate
F
Sequencing depth
Control rep1
L10 rep1
H10 rep1
Control rep2
L10 rep2
H10 rep2
Control rep3
L10 rep3
H10 rep3
Replicate

Fig.3

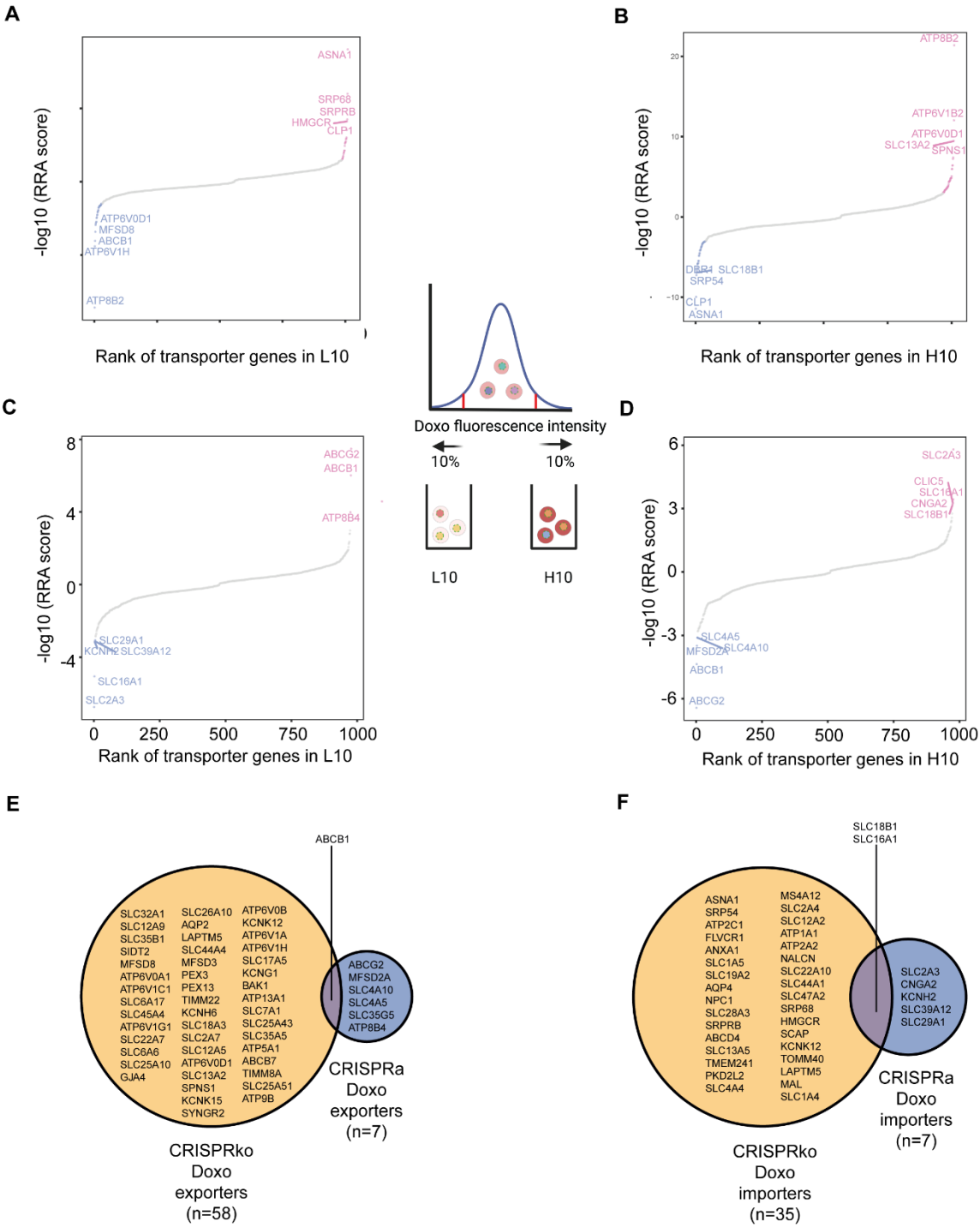

A
-log10 (RRA score)
ASNA1
SRP68
SRPRB
HMGCR
CLP1
ATP6V0D1
MFSD8
ABCB1
ATP6V1H
ATP8B2
Rank of transporter genes in L10
B
-log10 (RRA score)
ATP8B2
ATP6V1B2
ATP6V0D1
SLC13A2
SPNS1
DERL1
SLC18B1
SRP54
CLP1
ASNA1
Rank of transporter genes in H10
Doxo fluorescence intensity
10%
10%
L10
H10
C
-log10 (RRA score)
ABCG2
ABCB1
ATP8B4
SLC29A1
KCNH2
SLC39A12
SLC16A1
SLC2A3
0 250 500 750 1000
Rank of transporter genes in L10
D
-log10 (RRA score)
SLC2A3
CLIC5
SLC16A1
CNGA2
SLC18B1
SLC4A5
MFSD2A
SLC4A10
ABCB1
ABCG2
0 250 500 750 1000
Rank of transporter genes in H10
E
ABCB1
SLC32A1
SLC12A9
SLC35B1
SIDT2
MFSD8
ATP6V0A1
ATP6V1C1
SLC6A17
SLC45A4
ATP6V1G1
SLC22A7
SLC6A6
SLC25A10
GJA4
SLC26A10
AQP2
LAPTM5
SLC44A4
MFSD3
PEX3
PEX13
TIMM22
KCNH6
SLC18A3
SLC2A7
SLC12A5
ATP6V0D1
SLC13A2
SPNS1
KCNK15
SYNGR2
ATP6V0B
KCNK12
ATP6V1A
ATP6V1H
SLC17A5
KCNG1
BAK1
ATP13A1
SLC7A1
SLC25A43
SLC35A5
ATP5A1
ABCB7
TIMM8A
SLC25A51
ATP9B
ABCG2
MFSD2A
SLC4A10
SLC4A5
SLC35G5
ATP8B4
CRISPRa
Doxo
exporters
(n=7)
CRISPRko
Doxo
exporters
(n=58)
F
SLC18B1
SLC16A1
ASNA1
SRP54
ATP2C1
FLVCR1
ANXA1
SLC1A5
SLC19A2
AQP4
NPC1
SLC28A3
SRPRB
ABCD4
SLC13A5
TMEM241
PKD2L2
SLC4A4
MS4A12
SLC2A4
SLC12A2
ATP1A1
ATP2A2
NALCN
SLC22A10
SLC44A1
SLC47A2
SRP68
HMGCR
SCAP
KCNK12
TOMM40
LAPTM5
MAL
SLC1A4
SLC2A3
CNGA2
KCNH2
SLC39A12
SLC29A1
CRISPRa
Doxo
importers
(n=7)
CRISPRko
Doxo
importers
(n=35)

Fig.4

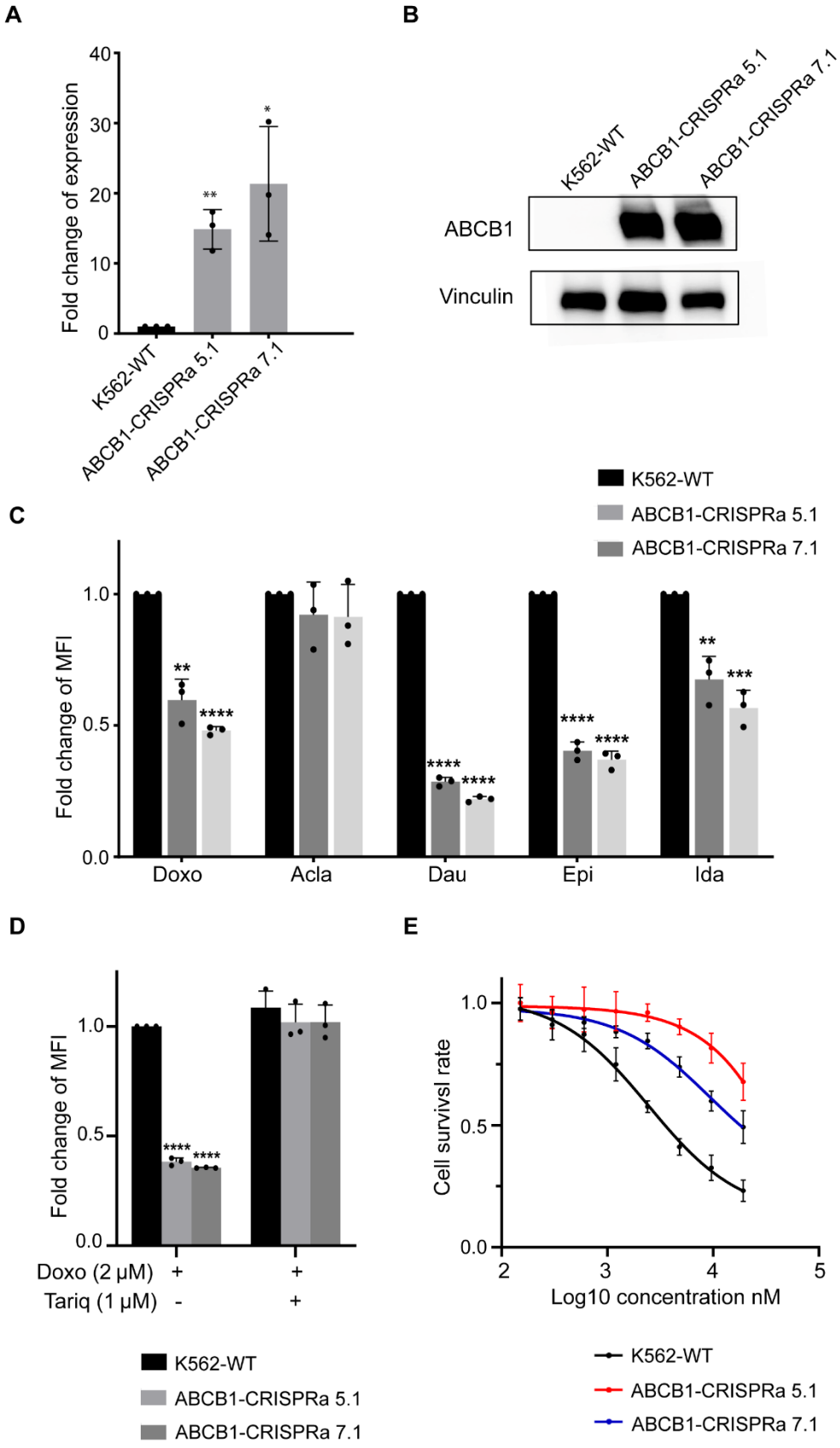

A
Fold change of expression
40
30
20
10
0
**
*
K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
B
K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
ABCB1
Vinculin
C
K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
Fold change of MFI
1.0
0.5
0.0
**
****
****
****
****
****
**
***
Doxo
Acla
Dau
Epi
Ida
D
Fold change of MFI
1.0
0.5
0.0
****
****
Doxo (2 μM) + +
Tariq (1 μM) - +
K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
E
Cell survivsl rate
1.0
0.5
0.0
2
3
4
5
Log10 concentration nM
K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1

Fig.5

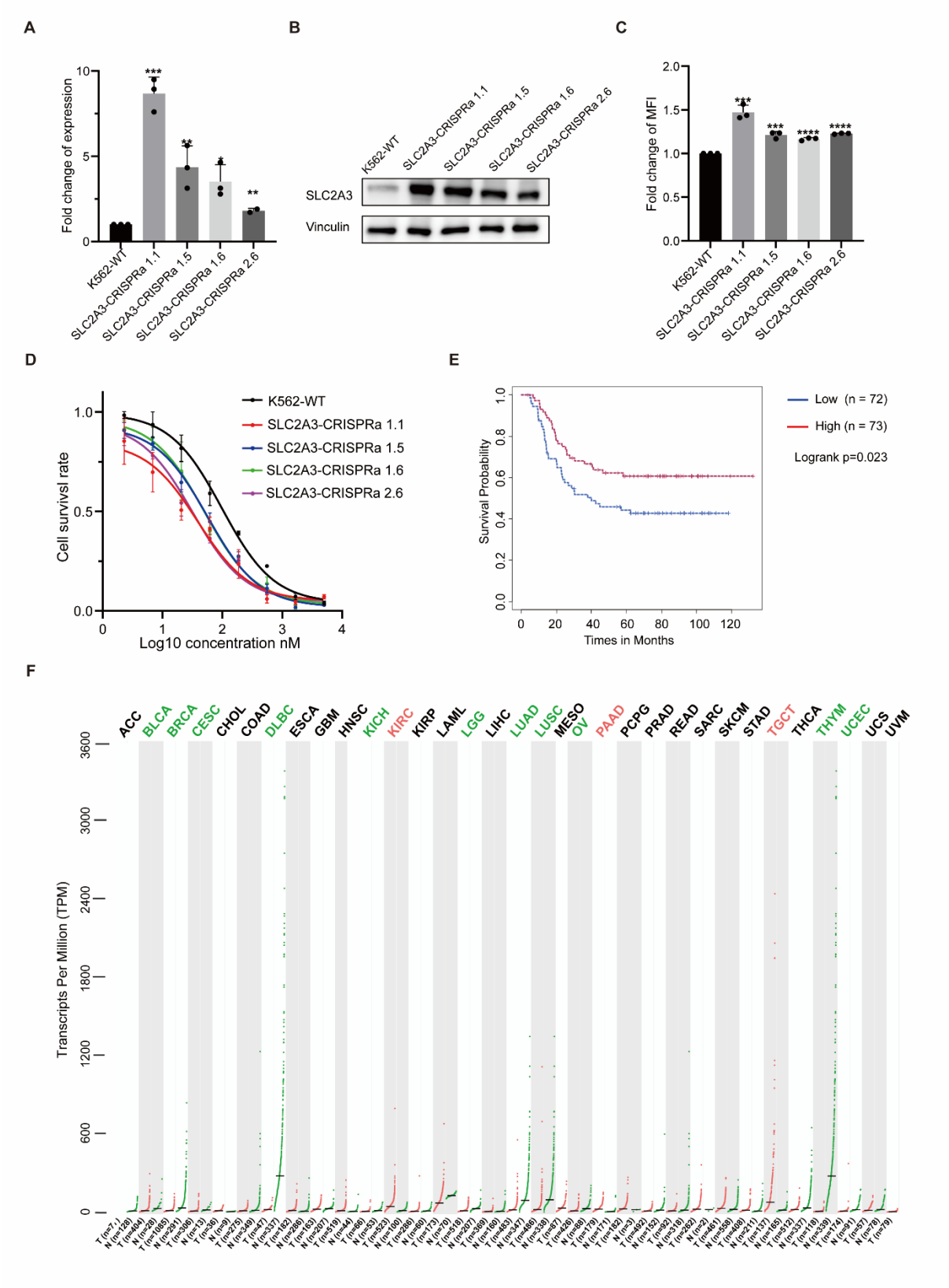

A
Fold change of expression
K562-WT
SLC2A3-CRISPRa 1.1
SLC2A3-CRISPRa 1.5
SLC2A3-CRISPRa 1.6
SLC2A3-CRISPRa 2.6
B
SLC2A3
Vinculin
C
Fold change of MFI
D
Cell survivsl rate
Log10 concentration nM
E
Survival Probability
Times in Months
Low (n = 72)
High (n = 73)
Logrank p=0.023
F
Transcripts Per Million (TPM)
ACC BLCA BRCA CESC CHOL COAD DLBC ESCA GBM HNSC KICH KIRC KIRP LAML LGG LIHC LUAD LUSC MESO OV PAAD PCPG PRAD READ SARC SKCM STAD TGCT THCA THYM UCEC UCS UVM

Figure Legends

Fig 1: Workflow of the transporter-targeting CRISPRko and CRISPRa screenings. **A**. Two custom oligo pools were designed and synthesized for CRISPRko and CRISPRa library plasmid construction. Synthesized oligo pools were cloned into the digested CRISPRko or CRISPRa plasmid backbones. Constructed CRISPRko or CRISPRa plasmid library was transfected into 293T cells together with packaging plasmids to produce lentiviral CRISPRko and CRISPRa library. **B**. Outline of the drug uptake screenings. The K562 cells with CRISPRko or CRISPRa library were treated with 2 μM doxorubicin for 2 h before FACS sorting. The cells falling in the lowest 10% fluorescence intensity (L10) and highest 10% fluorescence intensity (H10) were sorted for downstream processing. **C.** The sgRNA spanning region was PCR amplified from total genomic DNA that was isolated from sorted or control cells for NGS sequencing. A custom computational workflow was used to quantify the sgRNA abundance. MAGeCK was used to identify positively or negatively selected genes for doxorubicin uptake or export with the cutoff of RRA score < 0.001.

Fig 2: Quality control of transporter-targeting CRISPRko and CRISPRa libraries. **A-B**. Piechart of the percentage of detected or undetected sgRNA in cells infected with CRISPRko (A) or CRISPRa libraries (B). The detected or undetected sgRNAs were indicated in different colors. **C-D**. Histogram shows the distribution of sgRNAs per gene from CRISPRko (C) and CRISPRa (D) libraries by sequencing. The y axis represents the number of genes. The x axis indicates the specific number of sgRNAs. **E-F**. Boxplot indicates the distribution of read coverage per sgRNA in all CRISPRko (E) and CRISPRa (F) screening samples. The x axis represents the different screening replicates. The y axis represents the read depth of each sgRNA.

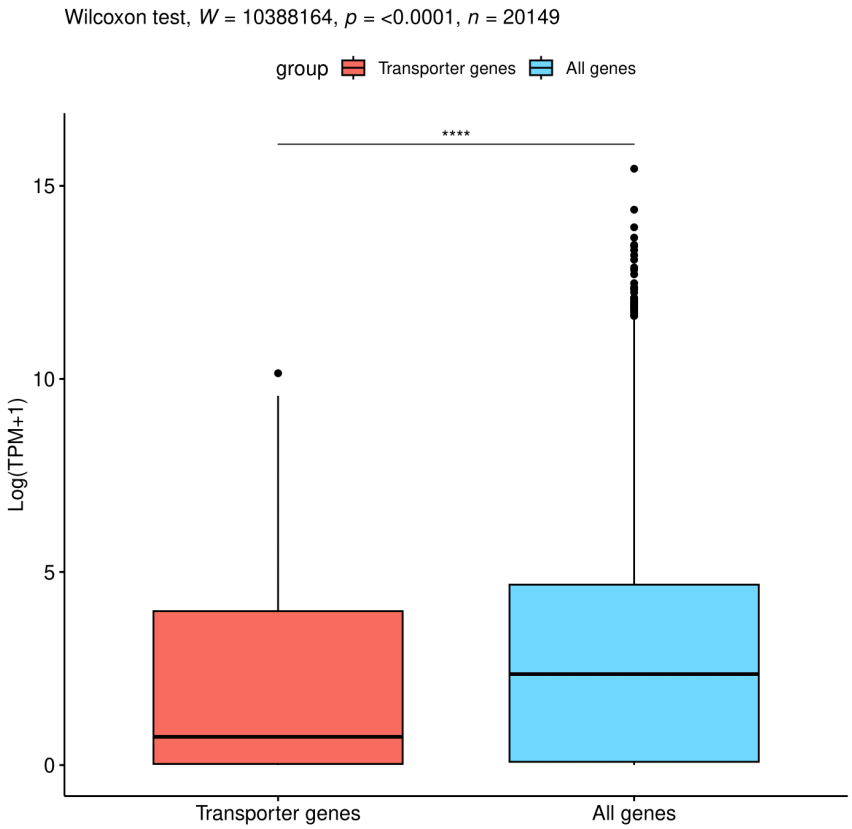
Fig 3: Identification of potential genes involved in doxorubicin transport from CRISPRko and CRISPRa screenings. **A**. Potential hits from L10 population from CRISPRko screening. **B**. Potential hits from H10 population from CRISPRko screening. **C**. Potential hits from L10 population from CRISPRa screening. **D**. Potential hits from H10 population from CRISPRa screening. The y axis represents MAGeCK RRA score. The x axis represents the ranking of the genes. The positively selected genes were indicated in red and negatively selected genes were indicated in blue

using the cutoff of MAGeCK RRA score < 0.001. The top 5 genes in both selection directions were highlighted by their gene symbols. **E**. Venn diagram shows the overlapping hits potentially involved in doxorubicin drug export from CRISPRko and CRISPRa screens. **F**. Venn diagram shows the overlapping hits potentially involved in doxorubicin drug import from CRISPRko and CRISPRa screens.

Fig. 4: Drug accumulation regulated by the ABCB1 gene during transient drug exposure.
A. qPCR was used to quantify the upregulation of ABCB1 gene in K562 ABCB1-CRISPRa clones. **B.** Western blotting was used to confirm the expression of ABCB1 in K562 ABCB1-CRISPRa clones. Vinculin was used as the loading control. **C.** FACS was used to quantify the uptake of drugs in K562 ABCB1-CRISPRa clones. For each group, cells were treated with doxorubicin, aclarubicin, daunorubicin, epirubicin, and idarubicin respectively at the final concentration of 2 μM for 2 h. Then fluorescence intensity of the drugs was quantified by FACS. **D.** The specific ABCB1 inhibitor tariquidar blocked the reduced uptake of doxorubicin in K562 ABCB1-CRISPRa clones. Cells were pre-treated with 1 μM tariquidar for 2 h, and then treated with 2 μM doxorubicin for 2 h. The fluorescence intensity of doxorubicin was quantified by FACS. **E.** CellTiter-Blue assay was used to quantify the cell viability. Cells were exposed to a serial dilution of doxorubicin for 72 h, then the live cells were measured. Bars show mean value ± s.e.m. (n = 2 or 3). *p < 0.05, **p < 0.005, and ***p < 0.0001 (versus the control), calculated using Student's t-test.

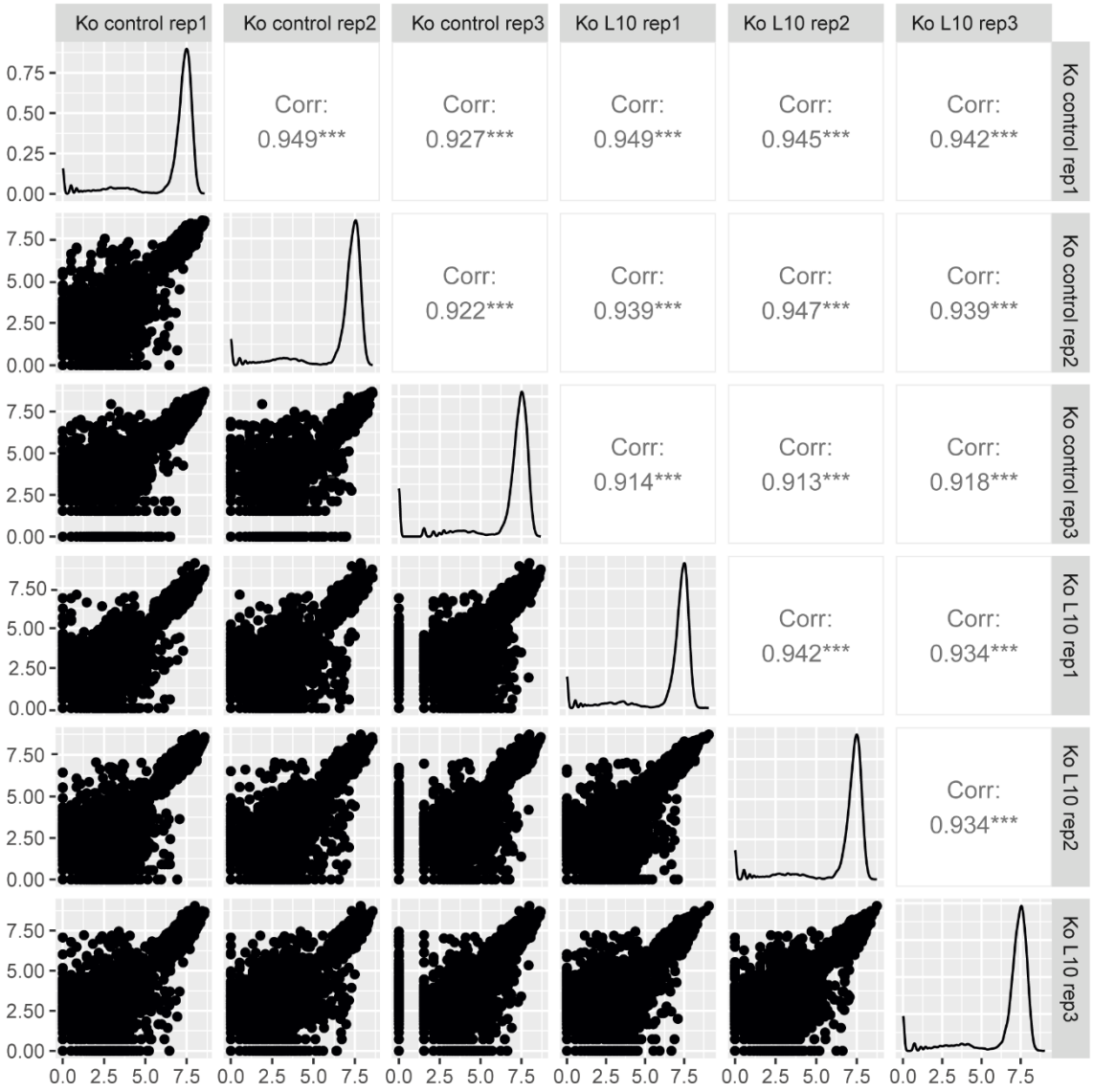
Fig. 5: SLC2A3 gene serving as a novel doxorubicin importer and response marker. **A.** qPCR was used to quantify the upregulation of SLC2A3 gene in K562 SLC2A3-CRISPRa clones. **B.** Western blotting was used to confirm the expression of SLC2A3 in SLC2A3-CRISPRa clones. Vinculin was used as the loading control. **C.** Quantification of doxorubicin uptake in K562 SLC2A3-CRISPRa clones. Cells were treated with doxorubicin at the final concentration of 2 μM for 2 h. Then the fluorescence intensity of the drugs was quantified by FACS. **D.** CellTiter-Blue assay was used to measure the cell viability. Cells were exposed to serial dilutions of doxorubicin for 72 h, then the live cells were measured. Bars show mean value ± s.e.m. (n = 2 or 3). *p < 0.05, **p < 0.005, and ***p < 0.0001 (versus the control), calculated using Student's t-test. **E.** Kaplan-Meier analysis of overall survival

of AML patients from TARGET study based on the SCL2A3 expression level. The red line indicates survival probability from the AML patients with higher expression of SLC2A3 (n=73). The blue line indicates survival probability from the AML patients with lower expression of SLC2A3 (n=72). The p value was calculated using the log rank test. **F.** Dot plot showing SLC2A3 expression across 33 different cancer types and their paired normal tissues. The red dots represent normalized SLC2A3 expression of samples from specific cancer type. The green dots represent normalized SLC2A3 expression of samples from specific normal tissue type. The cancer type with higher SLC2A3 expression than its matched normal tissue is indicated by red. The cancer type with lower SLC2A3 expression than its matched normal tissue is indicated by green. TPM, transcripts per million; ACC, adrenocortical carcinoma; BLCA, Bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangio carcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocelular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma.

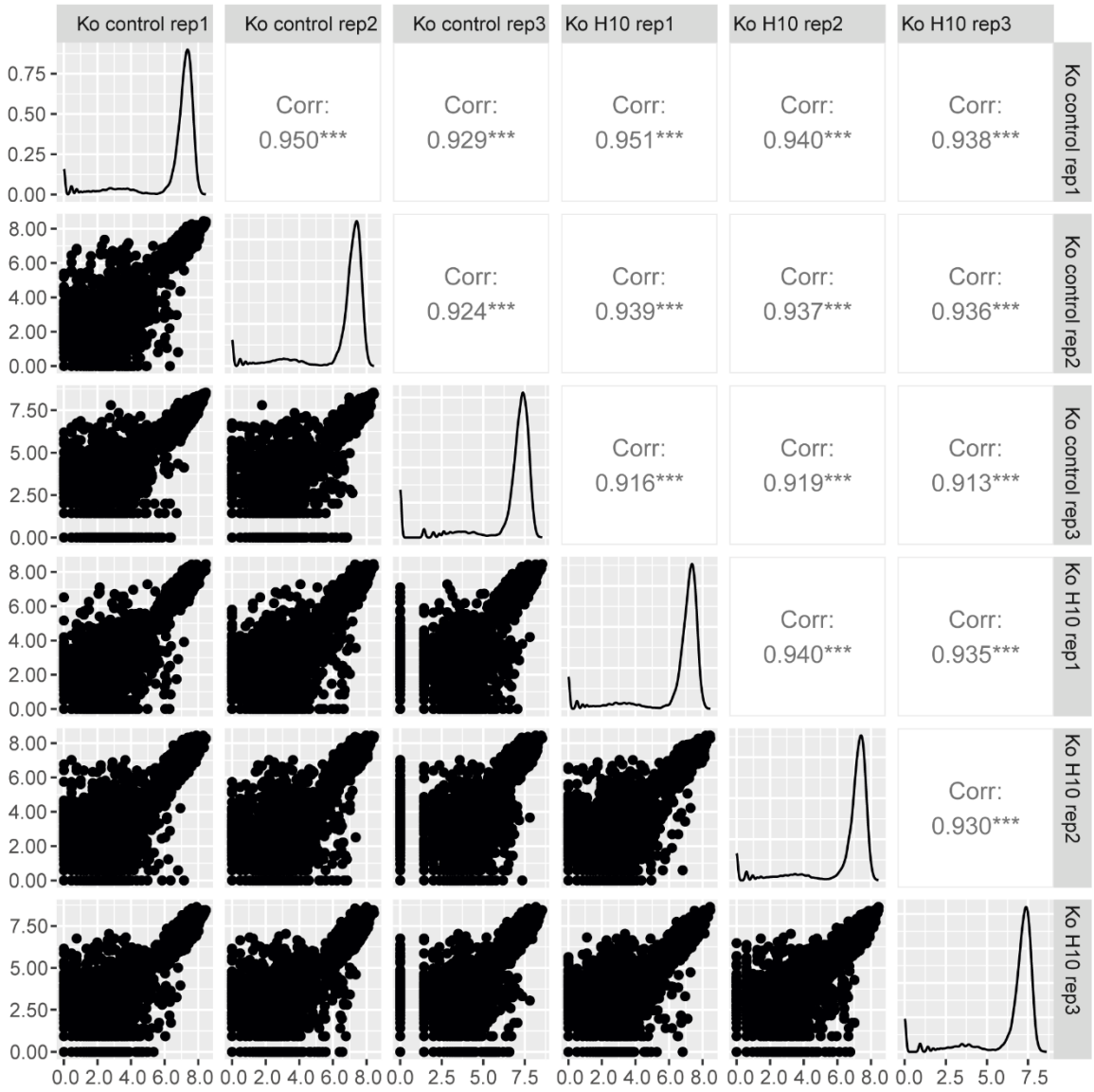


Supplementary Fig. 1: Gene expression of transporter genes compared to all genes. The *y* axis represents the logarithm of TPM normalized expression level of genes plus 1. The p value was calculated using unpaired two-sided Wilcoxon Rank Sum test. n = 979 (only transporter) or 19,193 (all) genes. K562 RNA-seq data were downloaded from Dependency Map Portal. ****p < 0.0001.

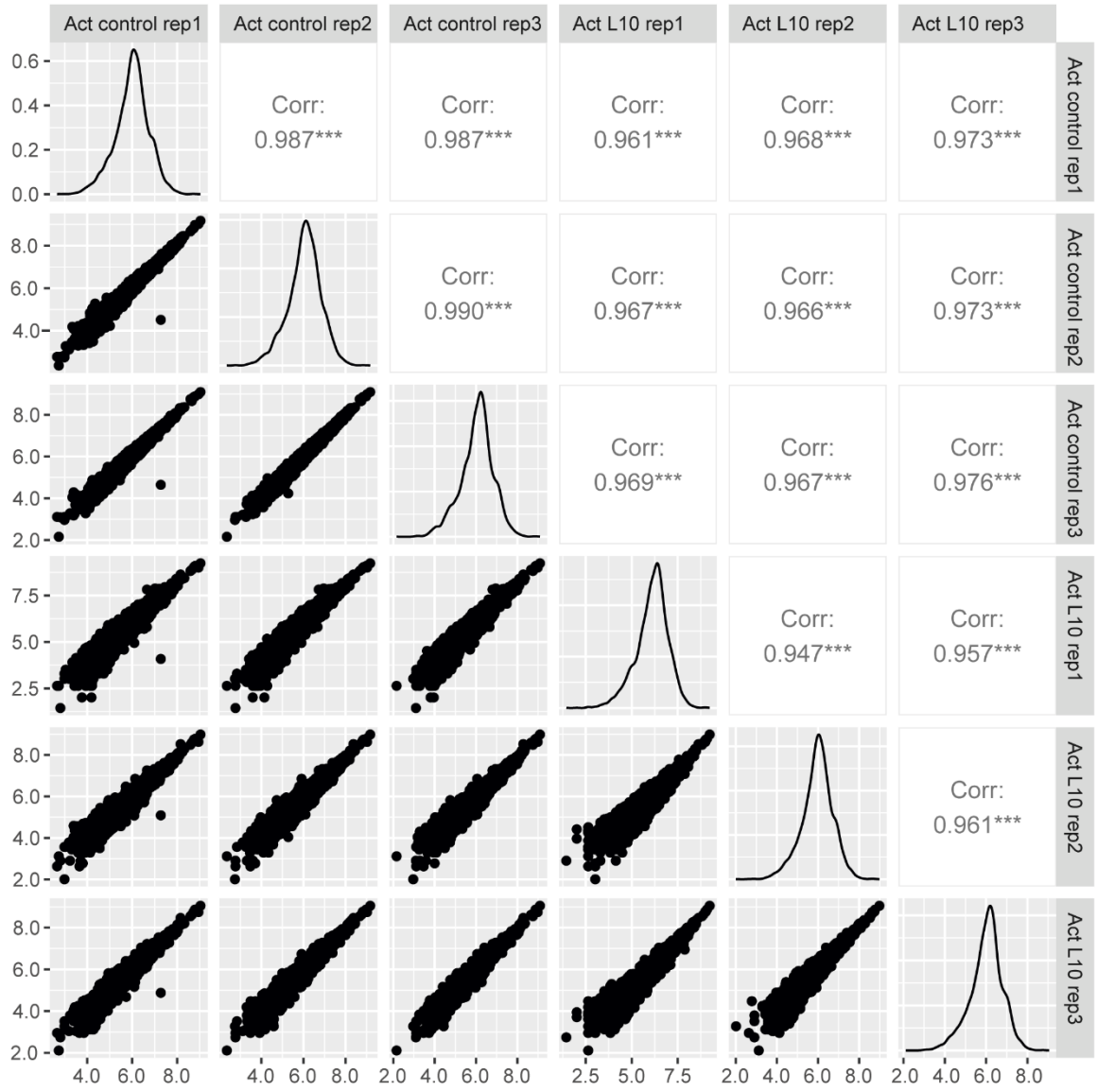
A


Ko control rep1
Ko control rep2
Ko control rep3
Ko L10 rep1
Ko L10 rep2
Ko L10 rep3
Corr: 0.949***
Corr: 0.927***
Corr: 0.949***
Corr: 0.945***
Corr: 0.942***
Corr: 0.922***
Corr: 0.939***
Corr: 0.947***
Corr: 0.939***
Corr: 0.914***
Corr: 0.913***
Corr: 0.918***
Corr: 0.942***
Corr: 0.934***
Corr: 0.934***
0.00
2.50
5.00
7.50
0.25
0.50
0.75
0.0 2.5 5.0 7.5

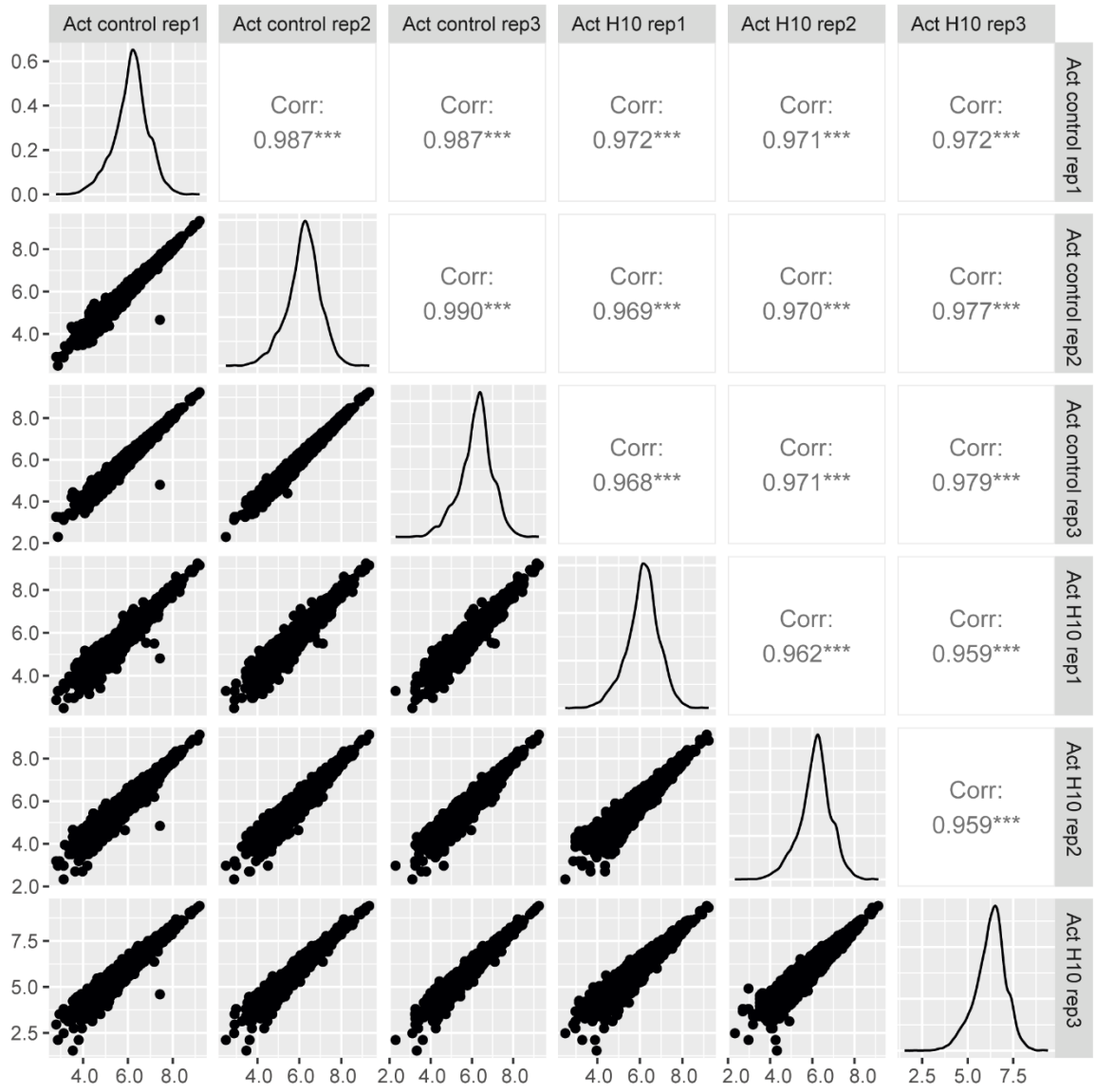
B


Ko control rep1
Ko control rep2
Ko control rep3
Ko H10 rep1
Ko H10 rep2
Ko H10 rep3
Corr: 0.950***
Corr: 0.929***
Corr: 0.951***
Corr: 0.940***
Corr: 0.938***
Corr: 0.924***
Corr: 0.939***
Corr: 0.937***
Corr: 0.936***
Corr: 0.916***
Corr: 0.919***
Corr: 0.913***
Corr: 0.940***
Corr: 0.935***
Corr: 0.930***
0.0 2.0 4.0 6.0 8.0
0.0 2.5 5.0 7.5

C

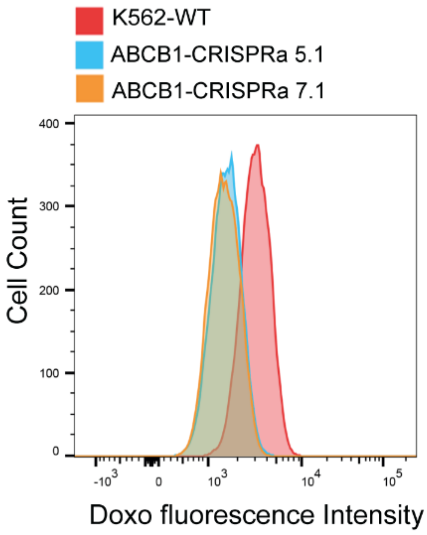

Act control rep1
Act control rep2
Act control rep3
Act L10 rep1
Act L10 rep2
Act L10 rep3
Corr: 0.987***
Corr: 0.987***
Corr: 0.961***
Corr: 0.968***
Corr: 0.973***
Corr: 0.990***
Corr: 0.967***
Corr: 0.966***
Corr: 0.973***
Corr: 0.969***
Corr: 0.967***
Corr: 0.976***
Corr: 0.947***
Corr: 0.957***
Corr: 0.961***

D

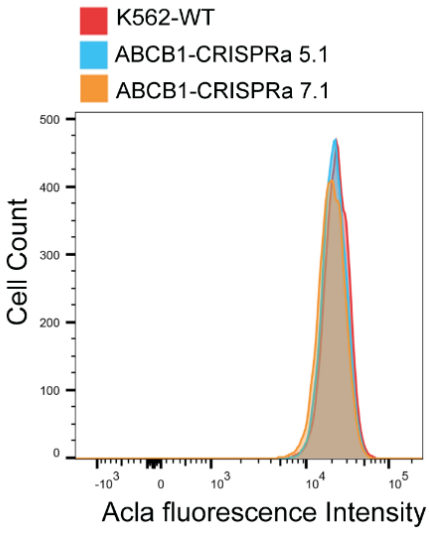


Supplementary Fig. 2: Pairwise correlation of CRISPRko and CRISPRa screen replicates. **A**. Pairwise correlation of L10 population from the CRISPRko screenings with the control group. **B**. Pairwise correlation of H10 population from the CRISPRko screenings with the control group. **C**. Pairwise correlation of L10 population from the CRISPRa screenings with the control group. **D**. Pairwise correlation of H10 population from the CRISPRa screenings with the control group. The logarithm of read count plus 1 of each replicate was plotted. Pearson correlation coefficient was calculated to measure the correlation between each replicate pair.

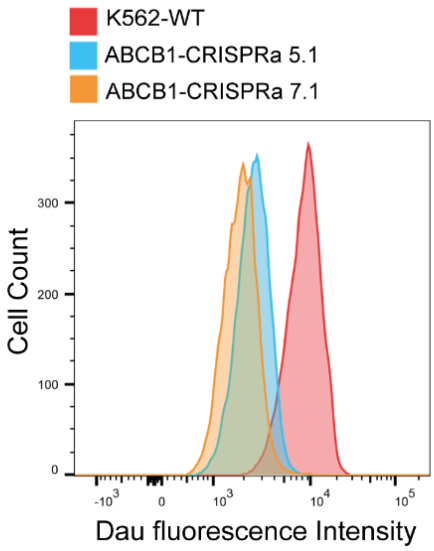
A


K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
Cell Count
Doxo fluorescence Intensity

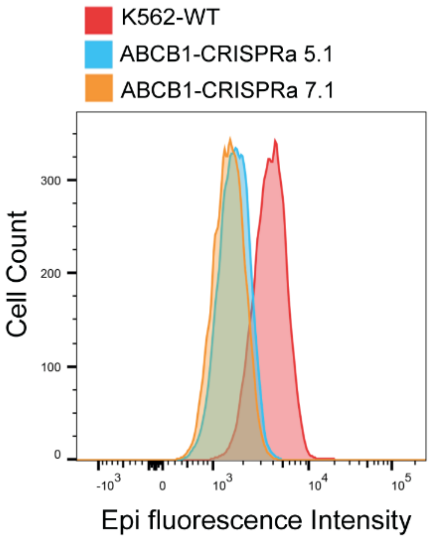
B


K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
Cell Count
Acla fluorescence Intensity

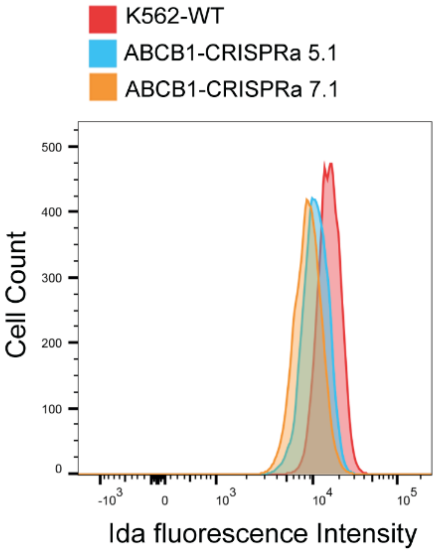
C


K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
Cell Count
Dau fluorescence Intensity

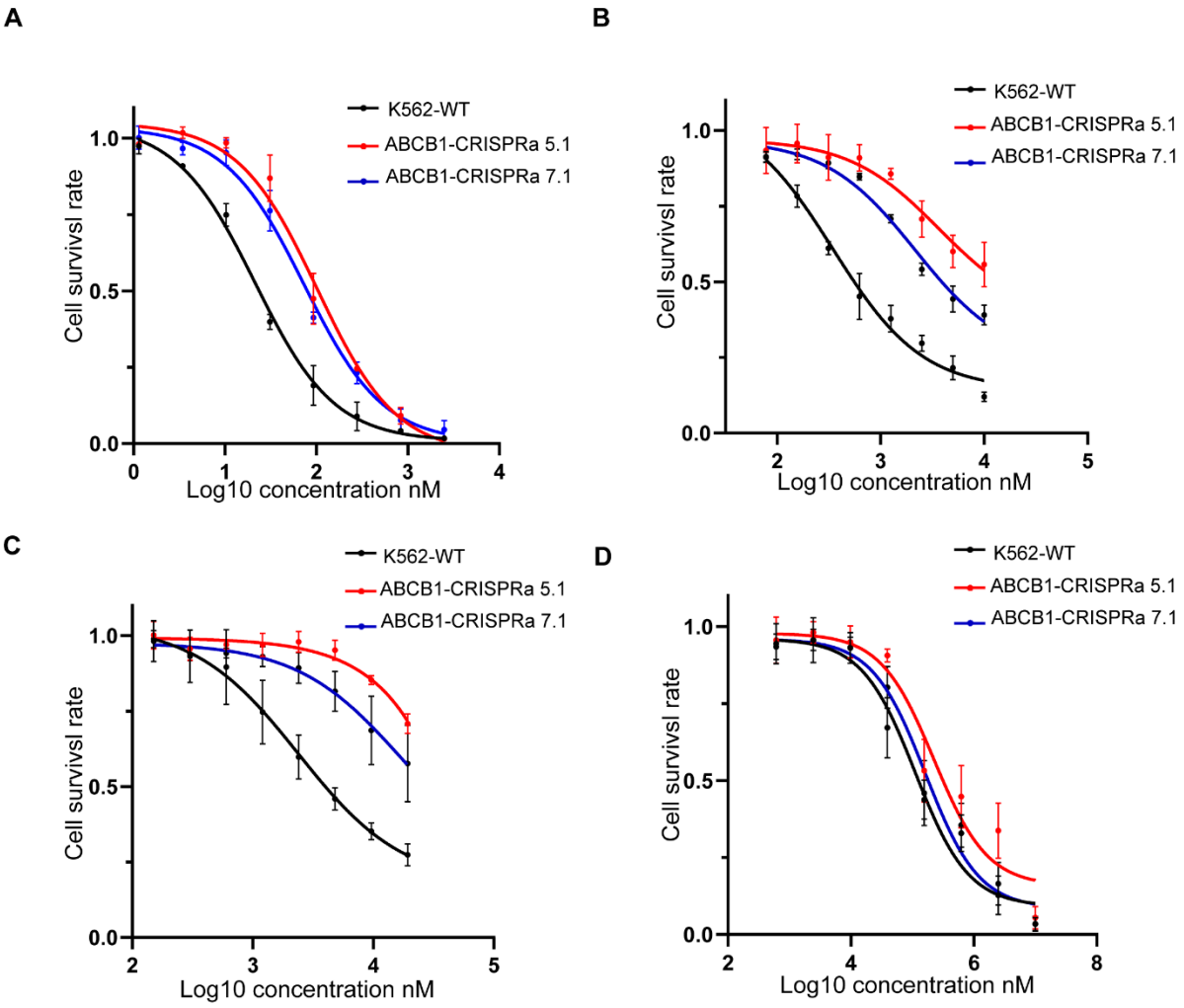
D


K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
Cell Count
Epi fluorescence Intensity

E

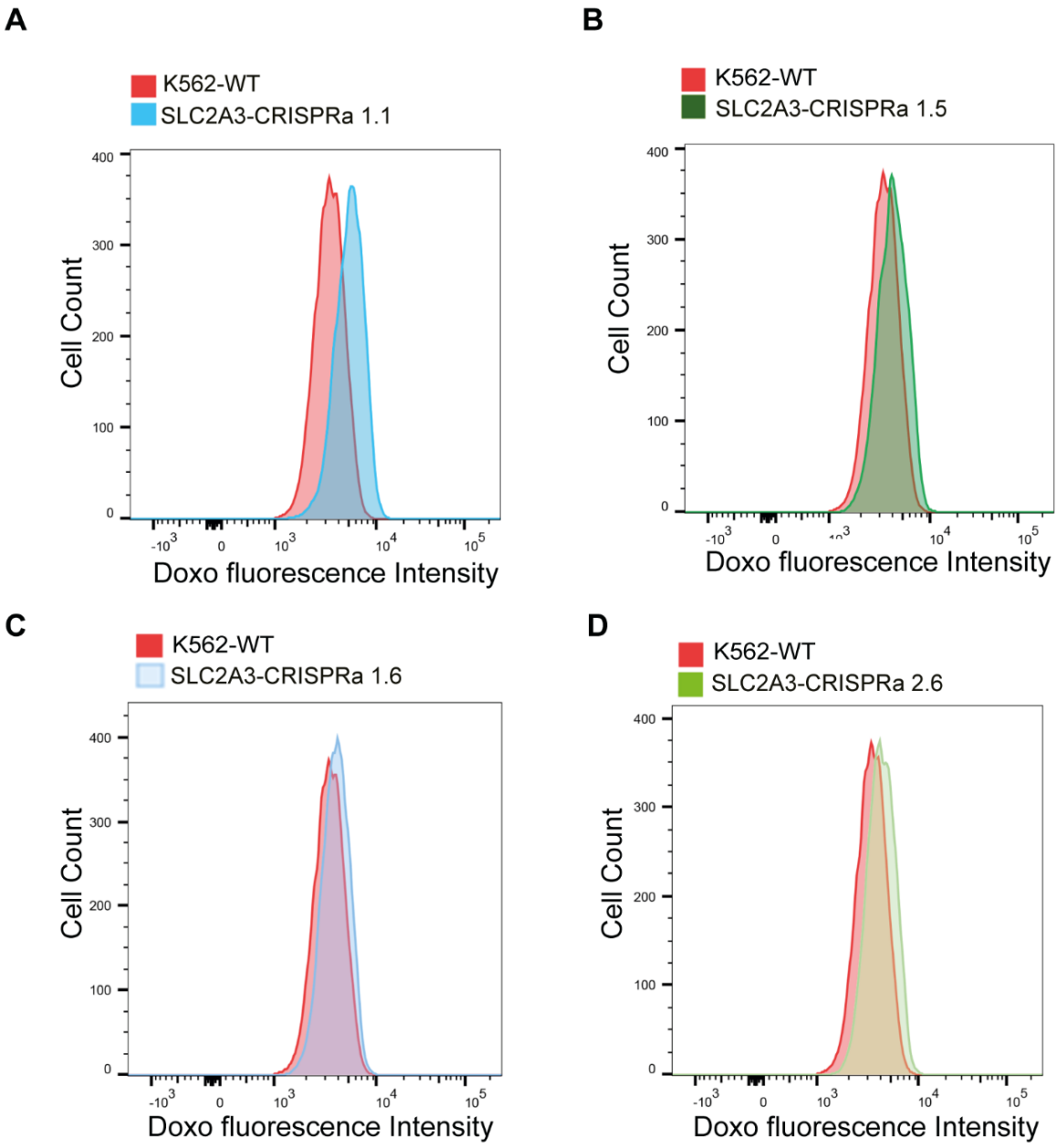

K562-WT
ABCB1-CRISPRa 5.1
ABCB1-CRISPRa 7.1
Cell Count
Ida fluorescence Intensity

Supplementary Fig. 3: Drug accumulation regulated by the ABCB1 gene during transient drug exposure. **A-E**. FACS histograms from the quantification of the drug uptake in K562 ABCB1-CRISPRa clones. For each group, cells were treated with doxorubicin (A), aclarubicin (B), daunorubicin (C), epirubicin (D), and idarubicin (E) respectively at the final concentration of 2 μM for 2 h. Then fluorescence intensity of the drugs was quantified by FACS.

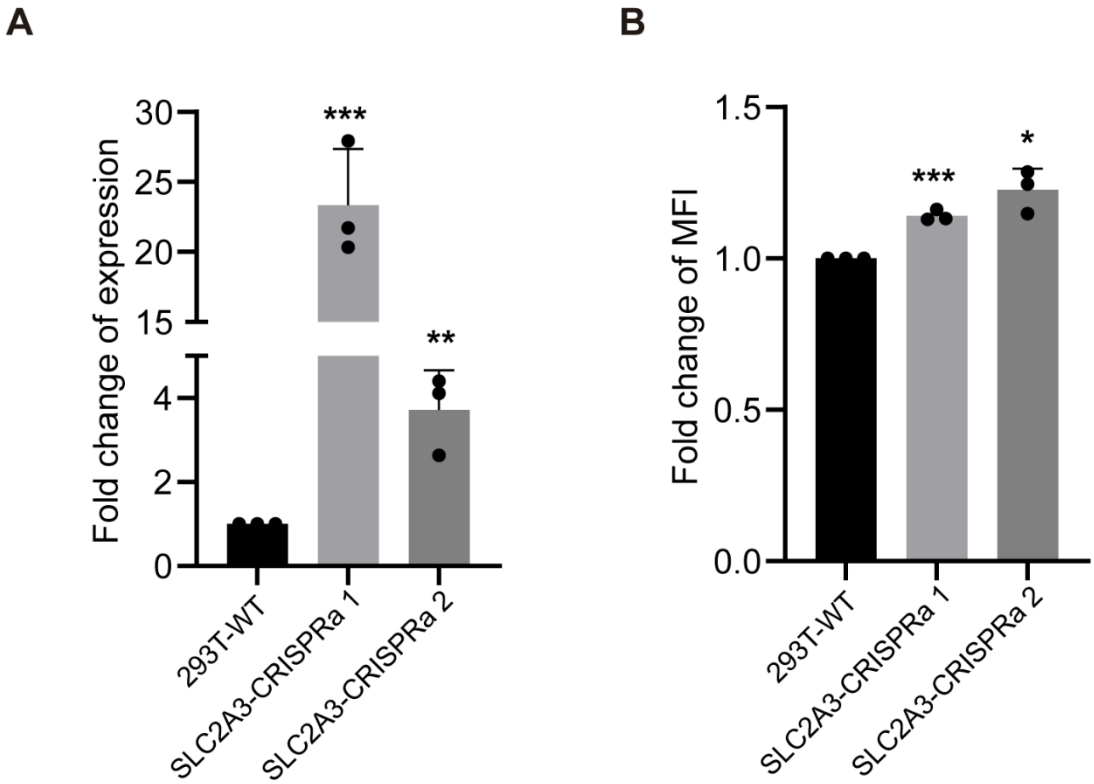


Supplementary Fig. 4: Cell survival regulated by the ABCB1 gene. For each group, cells were treated with aclarubicin, daunorubicin, epirubicin, and idarubicin respectively at a serial dilution of drugs for 72 h. CellTiter-Blue assay was used to quantify the cell viability. **A**. Cells were exposed to a serial dilution of aclarubicin for 72 h, then the live cells were measured. **B**. Cells were exposed to a serial dilution of daunorubicin for 72 h, then the live cells were measured. **C**. Cells were exposed to a

serial dilution of epirubicin for 72 h, then the live cells were measured. **D**. Cells were exposed to a serial dilution of idarubicin for 72 h, then the live cells were measured.



Supplementary Fig. 5: Drug accumulation regulated by the SLC2A3 gene during transient drug exposure. **A-D**. FACS histograms from the quantification of the drug uptake in K562 SLC2A3-CRISPRa clones. For each clone, cells were treated with doxorubicin at the final concentration of 2 μM for 2 h. Then fluorescence intensity of the drugs was quantified by FACS.



Supplementary Fig. 6: Doxorubicin uptake was enhanced in 293T cells with elevated SLC2A3 expression. **A**. qPCR was used to quantify the expression of SLC2A3 in 293T SLC2A3-CRISPRa bulk cells. **B.** FACS was used to quantify the uptake of doxorubicin in 293T SLC2A3-CRISPRa bulk cells. For each cell line, cells were treated with doxorubicin at the final concentration of 2 μM for 2 h. The fluorescence intensity of the drugs was quantified by FACS. Bars show mean value ± s.e.m. and significance was calculated using Student's t-test (n = 2 or 3). *p < 0.05, **p < 0.005, and ***p < 0.0001 (versus the 293T-WT cells).

Chapter 3

The genome-wide dual-CRISPR screening identifies essential non-coding regulatory elements

Reviewed by ***Nature Methods***

The genome-wide dual-CRISPR screening identifies essential non-coding regulatory elements

Yufeng Li[1,7], Minkang Tan[1,7], Almira Akkari-Henic[1,7], Ningning Xu[1], Shengnan Sun[1], Yavuz Ariyurek[2], Susan L. Kloet[2], Richard Davis[3], Harald Mikkers[1], Joshua J. Gruber[4], Michael P. Snyder[5]*, Xiao Li[6]*, Baoxu Pang[1]*

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands

[2]Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, The Netherlands

[3]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, The Netherlands

[4]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA

[5]Department of Genetics, Stanford University, Stanford, California, USA

[6]Department of Biochemistry, The Center for RNA Science and Therapeutics, Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH, USA.

[7]These authors contributed equally

*Correspondence: b.pang@lumc.nl (B.P.), mpsnyder@stanford.edu (M.P.S), xiao.li9@case.edu (X.L.),

Abstract

Non-coding regulatory elements (NCRE) represent a major fraction of the human genome. We developed a straightforward dual-CRISPR screening system capable of deleting thousands of NCREs genome-wide to study their functions in distinct biological contexts in K562 cells and 293T cells. We show that many NCREs, including ultraconserved elements, have silencer activity and play essential roles in cell growth and drug response. NCREs with redundant functions could also be identified from the screening data. This dual CRISPR system is also compatible with single-cell sequencing. Removing the hit PAX6_Tarzan from human embryonic stem cells led to defects in cardiomyocyte differentiation, indicating this UCE might be critical in heart development. Our study provides further evidence that many NCREs have important biological functions contributing to human biology and diseases.

Introduction

Protein-coding genes only represent less than 2% of the human genome, and the rest is non-coding, many of which contain regulatory elements that guide the transcription of genes at the right time and within the right tissue[1,2]. Based on the known functions, the non-coding regulatory elements (NCREs) are categorized into various segments, such as non-coding RNAs, promoters, enhancers, silencers, insulators, etc[2-7]. Assigning and understanding the function of the non-coding regulatory genome have been one main focus in genetic research during the past decades[2]. With the efforts from individual research groups and large consortia such as ENCODE and Roadmap Epigenomics, putative biological roles have been assigned to many NCREs[8-11]. In general, different types of NCREs have unique combinations of epigenetic modifications. Therefore, using next-generation sequencing methods like ChIP-seq against various histone modifications or transcription factors, or methods to profile accessible chromatins such as DNase I hypersensitivity sites combined with sequencing, FAIRE-seq, and ATAC-seq, many NCREs have been mapped in the human genome[12,13]. For instance, insulator regions are often enriched for CTCF binding sites[14,15], while enhancer regions are usually decorated with a combination of H3K27ac and H3K4me1[16]. Simultaneously, the recent development of massively parallel reporter systems also facilitates the direct biological activity measurement and identifications of enhancers and silencers[4,17,18]. However, all these methods do not measure the biological functions of these NCREs in their endogenous environment.

Recent advances in CRISPR genome editing have paved a new venue to study both the coding and non-coding parts of the genome[19-22]. CRISPR–Cas9 recognizes 20-bp genomic regions followed by PAM (5'-NGG-3') and typically introduces genetic changes of a few nucleotide deletions or insertions around the targeting sites in the genome[23]. This is especially useful in studying gene functions as mutation-induced frameshifts in the coding regions would render the proteins not functional anymore. On the other hand, NCREs often range from 50-200 bp in length, with multiple transcription factor (TF) binding sites[24]. Therefore, a single

CRISPR–Cas9-mediated genome editing must completely destroy the TF binding sites[25,26], which is also limited by the possible CRISPR–Cas9 targeting sites throughout the human genome. Due to such limitations, single guide RNAs tiling an entire testing region were often used to study NCREs regulating a few important genes. The modified CRISPR system that uses catalytic inactive Cas9 proteins (dCas9), either linked to a transcription activation or repression system, can also be used to study the functions of enhancers and insulators[26,27]. However, pre-knowledge of NCRE functions is needed to select the proper CRISPR–dCas9-mediated activation or repression systems[27,28]. Such selection could be further complicated by the recent evidence showing that certain NCREs could be bifunctional, e.g., functioning as enhancers or silencers depending on the cellular context[7,18]. Similarly, the dCas9 system is also limited by the availability of single guide RNAs at the targeting regions. Several dual CRISPR systems have been used to study the non-coding RNAs or enhancers, however, their throughputs are still limited due to either the barcoding system or the design of the screening system; therefore only limited regions were targeted[29,30]. Thus far, no systematic study of the NCREs in a genome-wide fashion, especially focusing on enhancers and silencers, has been performed.

We have developed a new dual-CRISPR screening system that could delete thousands of NCREs in a systematic and genome-wide fashion. This dual CRISPR screening system is easy to construct and sequence, without the need for additional barcoding. As target regions of more than 200 bp in size are removed from the genome, NCREs could be studied irrespective of their specific biological functions. We designed dual-CRISPR libraries targeting 4,047 Ultraconserved elements (UCEs) in the human genome from UCNEbase[31], 1,527 in vivo-validated conserved enhancers from VISTA Enhancer Browser[32], and all 13,539 predicted enhancers in K562 cells from ENCODE[16]. Using this system, we studied the biological functions of the UCEs in the human genome and identified regions that would affect cell survival and drug response in K562 and 293T cells. We found many UCEs have silencer activities, and many enhancers would show dual

functions. Furthermore, clusters of NCREs that play important functions could also be identified. Identified UCE region PAX6_Tarzan showed key function in cardiomyocyte differentiation from hESCs, underscoring the feasibility of our approach to gain insight into the function of UCEs. Here in this study, we provide a versatile tool and pipeline to study the function of NCREs and other non-coding parts of the genome.

Results

Development of the dual-CRISPR system

NCREs could be positive transcriptional regulators, i.e., enhancers; negative transcriptional regulators, i.e., silencers; or genome structure regulators, i.e. insulators. To systematically study NCREs irrespective of their different biological functions in their endogenous chromosome context, a versatile high-throughput dual-CRISPR screen system was developed. In this dual-CRISPR system, two different RNA polymerase III promoters, U6 and H1, are positioned in a convergent orientation to drive the transcription of two CRISPR guide RNAs (Supplementary Fig. 1a). To test this plasmid system, two guide RNAs targeting the 5 and 3 prime ends of one DNA fragment in the genome were inserted. After transfecting the cells, two functional guide RNAs were expressed and able to successfully delete the targeted region from the human genome (Supplementary Fig. 1b and 1c).

To target thousands of potential NCREs and test their functions in different contexts, a lentiviral system and cloning strategies analogous to the method described above were developed. In short, after NCREs were selected, all potential single guide RNAs that target both ends of each NCRE were designed. Guide RNAs were then selected based on their targeting efficiency and off-target potential. After that, guide RNAs were paired to be able to remove the targeting regions. Paired 20-nucleotide (nt) crRNA sequences are then properly orientated to follow the direction of the convergent promoters (Supplementary Fig. 1d). A restriction enzyme site was placed between the paired protospacer sequences in order to open the plasmids to insert the guide RNA scaffolds in the subsequent

steps, and oligo pools were ordered. A two-step cloning strategy was then used to assemble the full functional dual-CRISPR library (Methods). In brief, first, the oligo pool that contains only the paired 20-nt crRNA sequences were cloned into the lentiviral vector. After propagation, the plasmids containing the paired 20-nt crRNA sequences were opened up by restriction enzyme digestion, followed by the second round of cloning to insert the two tracrRNA scaffold sequences. The final plasmid libraries then contain paired functional guide RNAs that would remove the respective individual NCREs. There are several advantages of the new dual-CRISPR system. Two distinct promoter sequences should reduce the chance of potential recombination events in plasmids or infected cells. Furthermore, arranging guide RNAs in a convergent orientation allows direct PCR steps to amplify the fragments containing the paired guide RNA sequences from the infected cells, which are compatible with direct high-throughput pair-end sequencing (Supplementary Fig. 1e).

Identification of essential ultraconserved non-coding elements

UCEs are non-coding genetic sequences that are identical among different species [33-35]. Given such stringent conservation during evolution, it is expected that many UCEs should have biological functions and be pivotal in different species[36,37]. However, recent research showed that ultraconserved enhancers did not require perfect sequence matches to maintain their functions, when 23 of such enhancer UCEs were studied in vivo in mouse models[38]. To study the function of UCEs in a high-throughput fashion, a dual CRISPR library was assembled based on the published computation pipeline[39], which targets 4,047 UCEs in the human genome from UCNEbase[31], and 1,527 in vivo-validated conserved enhancers from VISTA Enhancer Browser[32]. In total 63,879 dual-CRISPRs were designed, including 1,070 control guides from a previous study[29]. The dual-CRISPR library was packaged into lentivirus, then used to infect K562 cells stably expressing Cas9 proteins (Supplementary Fig. 2a). Infected cells were kept in culture for 15 days to identify NCREs that would affect cell growth. Genomic DNA was isolated, and PCR was performed to extract the dual-CRISPRs. The abundance of different dual-

CRISPRs was compared to that from the initial population (day 0). Two biological replicate experiments were performed to extract reliable hits. The replicates correlated well with each other, indicating the screen system is stable and reliable (Supplementary Fig. 2b). After filtering low coverage, more than 90 percent of the target NCREs were matched with the paired guide RNAs across all replicates (Supplementary Fig. 2c). The robust ranking algorithm MAGeCK was used to identify potential NCREs that affect cell growth based on the screen data[40]. There were 346 UCEs and other potential NCREs depleted in the cell population that further grew for 15 days, compared to the initial population, suggesting these UCEs are potentially essential NCREs in K562 cells (Fig. 1a, Supplementary Table 1). Unexpectedly, we also identified a previously unannotated intergenic region that affected the growth of K562 cells. Thus, the dual-CRISPR system was capable of interrogating both previously annotated and unannotated genomic regions in an unbiased fashion.

To validate the identified UCEs and the potential new NCRE, dual-CRISPR pairs targeting the potential essential NCREs were used to generate knockout clones in K562 cells (Fig. 1, b, c, d and e; upper panel). We first determined cell growth rates in these clones. When UCEs PBX3_Claudia (referred as PBX3_Cl), FOXP1_Flora (referred as FOXP1_Fl), PAX6_Tarzan (referred as PAX6_Ta) and one potential NCRE (referred as de_novo_1) were removed from K562 cells, respective clones grew significantly slower compared to the wild-type K562 cells, indicating these UCEs regulate important genes or pathways in K562 cells (Fig. 1b, c, d and e; lower panel, and Supplementary Fig. 2d). We then surveyed the epigenetic modifications surrounding these UCEs. No clear combinations of epigenetic signatures were found to predict the function of these regions (Supplementary Fig. 2e), except that the de_novo_1 region sits next to a CTCF binding site (within 1 kb to the center of the CTCF binding site) which does not directly involve in chromosome looping based on ChIA-PET data (Supplementary Fig. 2f). As these UCEs are potential regulatory elements, we then tested their transcriptional regulatory activities using luciferase assays. Genomic fragments containing the

UCEs were cloned into two different luciferase reporter systems for the detection of enhancer and silencer activities respectively. Interestingly, no significant enhancer activity was observed from these regions using a commonly used minimal-promoter-driven luciferase reporter (Fig. 1f). However, two UCEs showed silencer activity monitored by a PGK-promoter-driven luciferase reporter (Fig. 1g).

NCREs usually function in a tissue-specific manner. To test the functions of UCEs in a different cell type of origin, another NCRE essentiality screen was performed in 293T cells using the same dual-CRISPR library (Supplementary Fig. 3a, Supplementary Table 2). As expected, only less than 10% of the essential UCEs were shared between 293T cells and K562 cells (Supplementary Fig. 3b). However, when genes that fall within 1Mb surrounding the potential essential UCEs were compared, more than 30% of the genes were shared between the two cell lines (Supplementary Fig. 3c), suggesting that although a different set of UCEs that may function in these two cell lines, there might still be some overlapping activities shared among these different UCEs that regulate a common set of genes that affect cell growth.

UCEs regulate a cascade of gene pathways affecting cell growth

NCREs may regulate proximal and distal genes, especially genes that are within the same topologically associating domains (TAD)[4,41]. To find out which genes might be affected by the deletion of these identified NCREs and potentially lead to the growth disadvantage, we integrated Hi-C data to identify genes that share the same TAD with the respective NCREs[42]. Transcriptional changes of the genes in the same TAD with the NCRE and in a close-by TAD were monitored between WT and UCE-knockout K562 cells by qPCR. Transcription of PTPRD and RCN1 genes were significantly upregulated in the knockout clones of de_novo_1 and PAX6_Ta respectively (Fig. 1h, and Supplementary Fig. 4a). In addition, these results also indicate that some of these UCEs are potential silencers, as corroborated by the luciferase assay from de_novo_1 (Fig. 1g). To survey the regulatory effects of NCRE on global transcription, RNA-seq was performed on K562_PAX6_Ta_KO

clones. Transcriptional changes of genes that are 5 Mb surrounding the PAX6 regions were compared. RCN1 gene was significantly upregulated (Supplementary Fig. 4b), confirming the qPCR results (Supplementary Fig. 4a). Furthermore, CD59, EHF, ABTB2, CD44 and PRR5L genes were also significantly upregulated. Among these genes, EHF is a transcription factor, which may amplify the effects of PAX6_Ta regulation. PRR5L interacts with the mTORC2 complex, and its upregulation would promote apoptosis[43], which may also contribute to the growth delay phenotype. It should be noted that some of these de-regulated genes may be indirectly regulated by PAX6_Ta.

To link the cell growth phenotype to the genes regulated by the identified regions, we further studied the PTPRD gene that was upregulated by the de_nono_1 NCRE removal. PTPRD gene codes a transmembrane receptor protein tyrosine phosphatase with tumor suppressor functions[44]. CRISPR activation system was used to upregulate the PTPRD gene directly in K562 cells (Fig. 1i), which mimics the effects of de_nono_1 NCRE removal. A similar growth disadvantage was observed in cells with the PTPRD gene directly upregulated (Fig. 1j), suggesting that the NCRE de_novo_1 may impair cell growth by regulating the PTPRD gene.

UCEs that function in drug-resistance

Mutations or genetic changes in the NCREs could contribute to different diseases [45-48]. However, it is unknown if NCREs are directly involved in drug responses. To test this, K562 cells infected with the dual-CRISPR library targeting UCEs were exposed to the tyrosine kinase inhibitor imatinib that targets the BCR-ABL fusion kinase in K562 cells and other related chronic myeloid leukemias. After 15 days, surviving cells were collected, and changes in the abundance of dual-CRISPRs were analyzed.

After comparing the initial screening cell population, cells growing for 15 days without any drug treatment and cells growing under imatinib for 15 days, 81 NCREs possibly involved in resistance to imatinib treatment were enriched (Fig.

2a, Supplementary Table 3)[49]. First, luciferase assays were performed to identify whether these regions serve as enhancers or silencers. Two UCEs ZNF503_Ophelia (ZNF503_Op) and QKI_Jonathan (QKI_Jo) were studied and showed significant silencer activities when using a PGK-promoter luciferase reporter (Fig. 2b and c). Individual K562 cell clones with UCEs ZNF503_Op and QKI_Jo deleted were made respectively (Fig. 2d and e). These cells became more resistant to imatinib treatment than the control cells (Fig. 2f), indicating that the dual-CRISPR screenings identified potential NCREs that may play a role in imatinib resistance. To identify potential genes that these regions may regulate, transcription changes of the nearby genes were studied. In the ZNF503_Op KO clone, genes SAMD8, VDAC2 and ZNF503 were up-regulated (Fig. 2g), while in the QKI_Jo KO clone, genes PACRG and QKI were upregulated (Fig. 2h). These data also corroborate that these two NCREs may function as silencers (Fig. 2c), which contribute survival advantages of these cells during imatinib treatment, directly or indirectly. For instance, gene ZNF503 is a transcriptional repressor that may regulate genes that drive tumor cell proliferation[50,51].

Identification of essential enhancers using optimized dual-CRISPR systems

Enhancer regions have been extensively studied and well-defined during the past few decades. However, most annotations are based on the combinations of key epigenetic modifications computationally or using ectopic enhancer reporter assays. Using the single-guide RNA CRISPR system, enhancers could be studied in their endogenous loci[25,27]. However, only limited regions could be studied due to technical issues or chosen biological readouts. It has been challenging to dissect the roles of enhancers in their endogenous loci in a comprehensive manner[52], and therefore no related systematic study has been done yet.

Using the dual CRISPR system, we designed 254,203 guide RNAs targeting all 13,539 potential enhancers in K562 cells predicted by the ENCODE project[16]. Because for many enhancers, it is not possible to identify paired guide RNAs to remove the enhancer regions completely, a complementary strategy was used to

306 design guide RNA pairs targeting the inside of the enhancer regions to aim to
307 remove the core sequences in the center (Fig. 3a). While constructing these
308 libraries, we further optimized the cloning procedures and final structure of the dual
309 CRISPR library system (Fig. 3b), which we named dual-CRISPR-2.0. The main
310 improvement is that the distance between the two scaffolds was increased to 200
311 bp for optimal NGS sequencing efficiency (Fig. 3b). Using these genome-wide
312 dual-CRISPR screening libraries, all potential enhancers in K562 cells were
313 assayed and 1,005 enhancers were found to affect cell growth (Fig. 3c,
314 Supplementary Table 4). We focused first on the top hits by deleting these potential
315 essential enhancer regions, which caused decreased cell growth (Fig. 3d and
316 Supplementary Fig. 5a, b, c, d and e), indicating these potential enhancer regions
317 may regulate cell growth in K562 cells. Luciferase assays were then performed to
318 test the enhancer activity of these regions, where 4 out 6 regions led to a strong
319 and significant gene upregulation (Fig. 3e). These regions shared a typical
320 enhancer signature in K562 cells (Supplementary Fig. 5f, g, h, i , j and k). We then
321 tested the potential silencer activity of these regions. Interestingly 2 out of the 6
322 regions showed a significant silencer activity (Fig. 3f), but these 2 regions also
323 exerted significant enhancer activity (Fig. 3e). These seemingly contradictory
324 results in fact corroborate other recent studies where many NCREs were shown
325 to have both enhancer and silencer activities in different biological contexts and
326 when regulating different promoters[7,18,53]. In these studies, the promoters used for
327 the enhancer and silencer reporter systems were different. Furthermore, another
328 two NCRE regions did not show either enhancer or silencer activity, despite that
329 clear growth disadvantage was observed when these two NCREs were removed
330 (Fig. 3d and Supplementary Fig. 5d). One possibility is that these NCREs do not
331 regulate the specific promoters used in the luciferase assays. Therefore, studying
332 the functions of NCREs in their endogenous loci also provides complementary data
333 compared to other widely used experimental and computational methods to define
334 NCREs.
335

To gain more insights into the mechanisms of the growth defects, we further monitored the potential genes that may be regulated by some of these NCREs. For instance, enhancer E22:23590 is located within the BCR region, which forms chromosomal translocation with the Abl gene present in many patients with chronic myelogenous leukemia (CML)[54,55] and the K562 cell line. Although no clear regulatory functions were observed of this potential enhancer in K562, two nearby genes Rab36 and BCR-ABL were differentially deregulated (Fig. 3g). Gene Rab36 was downregulated, while BCR-ABL fusion gene was upregulated, suggesting a complex function of this NCRE. Downregulation of Rab36 was shown to inhibit cell growth[56]. Also, it has been shown that upregulation of BCR-ABL stimulated the TGF-β pathway causing cell growth arrest[57]. Therefore, the deregulation of these genes may collectively result in the growth disadvantage induced by removing enhancer E22:23590.

Inferring functional enhancer clusters containing redundant units

Multiple NCREs, especially enhancers, are often present in close vicinity surrounding their target genes. Often these enhancers play redundant roles in regulating the same gene or genes[58,59], which complicates the assignment of the biological functions of such enhancers (Fig. 4a). So far, most other groups and we have focused on identifying single NCRE or enhancer with strong phenotypes, such as on cell growth. It has been challenging to study clusters of enhancers in a high-throughput and genome-wide fashion. We hypothesized that for a cluster of enhancers with redundant activity, individually each of them would show some but not significant screening enrichment scores. However, when these enhancers are considered as a cluster, their combined biological effects would stand out. Based on this, a new analysis was performed on the enhancer essentiality screening (Fig. 3), where the targeted enhancers were further grouped according to previously defined clusters based on a distance metric[60]. To reliably capture the essential enhancer clusters, two computational models GSEA and MAGeCK RRA were applied[40,61], and P-values obtained by the two methods were corrected by BH procedure for FDR control (Methods). The shared top clusters computed by both

methods were then used for the downstream study (Venn diagram in Fig. 4b, Supplementary Table 5). The individual enhancers from the top 3 enhancer clusters were further studied (Fig. 4b). Luciferase assays were performed to determine the enhancer activity of these individual regions, and only enhancers from chromosome 6 showed strong enhancer activity using the pGL.4.23 reporter (Fig. 4c).

To test the functions of these individual enhancers and the clusters they reside in their endogenous loci, first, only one enhancer was removed from the clusters containing three enhancers using dual-CRISPR targeting. Only moderate growth disadvantage was observed (Fig. 4d, and Supplementary Fig. 6a and d). Then two enhancers were removed, which resulted in three possible combinations of two-enhancer deletion clones. In all cases, the two-enhancer deletion clones had stronger growth suppression compared to the deletion of any single enhancer (Fig. 4d and e, and Supplementary Fig. 6a, b, d and e), suggesting that some redundant functions are shared between these enhancer archipelagos as described before[49]. When all three enhancers were removed, proliferation was strongly impaired (Fig. 4f, and Supplementary Fig. 6c and f). These data indicate that the dual-CRISPR screening system is also capable of capturing clusters of NCREs with redundant biological functions in a genome-wide and systematic study.

Studying the functions of NCREs using the dual-CRISPR system at the single-cell level

Coupling the NCREs with their respective regulated genes has been challenging. Assays such as ChIA-PET or Hi-C that probe the 3D chromatin interactions could provide indications for genome-wide gene-NCRE physical interactions, though no direct measurement of the transcriptional regulation could be made. Single-cell (sc) RNA-seq combined with CRISPR perturbation (Perturb-seq) has been used to couple genetic perturbation, either targeting genes or enhancers, with broader transcriptome changes in single cells[62,63]. However, these studies still relied on single-guide-RNA-mediated perturbation, and prior knowledge of the potential

function of the NCREs is required to choose the right dCas9-repressor system. Removing a larger fragment of the genome while monitoring the transcriptome changes could provide a useful tool to study both the coding and non-coding part (irrespective of the prior knowledge of the NCRE) of the genome at the single-cell level. To explore this option, the dual-CRISPR screening system was modified to capture the transcriptome changes of single cells upon removal of NCREs. Two distinct capture sequences were added to the dual-scaffold of guide RNAs[64], allowing direct capture of the two guide RNAs and mRNAs within a single cell (Fig. 5a). As a pilot test, 82 pairs of guide RNAs targeting 42 different NCREs from the top lists of our genome-wide screening were selected and cloned into the sc-dual-CRISPR system that contains GFP as a marker. K562/Cas9 cells were infected with lentivirus containing the sc-dual-CRISPR library at a low MOI 0.3 to make sure each cell only contained one pair of guide RNAs. Then GFP positive cells were sorted and processed for scRNA-seq (Supplementary Fig. 7a, b, c, d, e and f).

Two single-cell sequencing libraries, either aiming to target up to 10,000 cells per chip or 30,000 cells per chip were made. After perturbation index assignment and filtering multiplet cell data, 1,199 and 3,271 usable single cells were harvested from two batches respectively. Therefore, it is possible to retrieve more usable single cells per perturbation with a higher number of input cells (Supplementary Fig. 7i and j), when pairs of guide RNAs were used as additional cell barcodes. The two experiments were combined with a mean of 110 cells per NCRE target achieved for downstream analysis (Supplementary Table 6), and a median of 4,454 genes per cell and a median of 21,446 UMIs of mRNA molecules per cell were observed (Supplementary Fig. 7j). We observed a similar trend of deregulation of genes between bulk RNA-seq data (Fig. 5b) and scRNA-seq data where PAX6_Ta region was removed, around the TADs of PAX6_Ta (Fig. 5c; 127 cells). To identify significantly deregulated genes targeted by the tested NCREs around their vicinity, differential expression analyses were performed by MAST[65], a method tailored to fit a two-part generalized linear model for zero-inflated and bimodal distributed single-cell gene expression data. Due to the detection limit on lowly expressed

transcripts[66], only genes with acceptable mean normalized expression levels and detected in a sufficient percentage of cells were used for DE test by MAST (Methods). Positive controls with guides targeting promoters of 4 ribosomal genes (RPL18A, RPL13, RPL21 and RPL8) confirmed the feasibility of the sc-dual-CRISPR system to capture target genes of NCREs (Supplementary Fig. 7k). There were 22 significant NCRE-gene pairs identified using this method (Fig. 5d, and Supplementary Fig. 7l, m and n, and Supplementary Table 7). For example, a potential enhancer E11:125334 that was identified to play a role in imatinib resistance was found to down-regulate gene El24 upon dual-CRISPR editing (Fig. 5d). Gene El24 is indeed associated with resistance to many chemotherapeutic drugs including imatinib[67,68], suggesting the enhancer E11:125334 may regulate El24 gene to exert imatinib resistance. In addition, a potential enhancer E16:30551 identified to affect cell growth was found to downregulate genes PPP4C and BOLA2B and simultaneously upregulate gene ZNF689 (Supplementary Fig. 7l). For instance, gene PPP4C encodes protein phosphatase 4 catalytic subunit and PPP4C-deficient thymocytes showed decreased proliferation and enhanced apoptosis in vivo[69]. The gene BOLA2B representing BolA family member 2 was reported to associate with human hepatocellular carcinoma progression and the BOLA2B-knockout mice model shows a slow tumorigenicity[70]. Together, deregulation of these genes together upon E16:30551 deletion may lead to cell growth disadvantage. These data show that it is possible to combine dual-CRISPR-mediated NCRE deletion with scRNA-seq to identify potential genes regulated by NCREs.

UCE in PAX6 region affects hESC cardiomyocyte differentiation

Assigning specific biological functions to NCREs is still challenging. UCEs are especially interesting as these regions are conserved among different species, and for a long time these regions have been speculated to play fundamental functions in evolution. However, recent research indicated that mice with a few UCEs removed showed no abnormalities, and sequence conservation did not play an important role in the enhancer function within UCEs in mice[71,72]. A caveat though

is that only a small percentage of the UCEs have been studied, and only the ones with enhancer function were considered. It is possible that other molecular functions may be associated with these UCEs in different developmental stages. It is therefore important to first define the potential functions of UCEs, which could be facilitated by systematically studying the functions of UCEs in different biological contexts using the dual-CRISPR system.

Global transcriptome analyses showed that in K562_PAX6_Ta_KO clones, several heart-related pathways were enriched (Fig. 6a). It suggests that this UCE may play a role in cardiomyocyte function. Although no enhancer or silencer activity of PAX6_Ta was shown in canonical luciferase assays in K562 cells (Fig. 1g and h), both qPCR and RNA-seq data measuring the surrounding gene expression of this UCE suggest that this UCE might be a silencer (Supplementary Fig. 4). As NCREs often function in a tissue- and promoter-specific manner, we then constructed a silencer reporter driven by a super-core promoter SCP1 for expressing GFP, which measures silencer activity on a different promoter from the PGK promoter used in luciferase assays. A reduction of GFP fluorescence was observed, indicating UCE_PAX6 may be a potential silencer in this context in K562 cells (Fig. 6b and Supplementary Fig. 8a).

To further investigate the potential function of PAX6_Ta in cardiac development, NKX2-5(eGFP/w) human embryonic stem cells (hESCs) were transfected with dual-CRISPR targeting the PAX6_Ta region[73]. Efficient genome editing and complete knockout were observed in the PAX6_Ta_KO_bulk hESCs (Fig. 6c and Supplementary Fig. 8b). Then both the parental NKX2-5(eGFP/w) hESCs and the PAX6_Ta_KO_bulk hESCs were subjected to cardiomyocyte differentiation. The parental NKX2-5(eGFP/w) hESCs became GFP^+ upon expression of the cardiac marker gene NKX2-5, once they committed into cardiomyocytes successfully [73]. In contrast, a significant reduction of GFP^+ cells was observed in PAX6_Ta_KO_bulk hESCs that underwent cardiac differentiation (Fig. 6d and e). Furthermore, the small percentage of the GFP^+ cardiomyocytes formed from the

PAX6_Ta_KO_bulk hESCs showed irregular beat patterns compared to WT NKX2-5(eGFP/w) hESCs (Movie 1, 2, 3 and 4). When the transcription changes of the genes surrounding UCE PAX6_Ta were measured, gene RCN1 was significantly upregulated in PAX6_Ta_KO_bulk hESCs compared to the parental cells after cardiomyocyte differentiation. In contrast, no transcriptional changes were observed before differentiation (Fig. 6f). Furthermore, several genes contributing to cardiomyocyte functionality such as CACNG8, RYR2 and MYH6 genes were also similarly de-regulated (Supplementary Fig. 8c and d). These data suggest that PAX6_Ta may be a key NCRE in cardiomyocyte differentiation via regulation of surrounding genes and possibly other genes that are key to heart development and physiology.

Discussion

NCREs are essential in regulating the transcription of genes and coordinating genomic information to form complex organisms. NCREs may function in activating, repressing, or insulating the transcription activities. In this study, we developed an adaptable dual-CRISPR system that could be used to study NCREs irrespective of their biological functions in a genome-wide fashion. The integrated dual-CRISPR libraries could be amplified and sequenced using routine methods, without the need for custom sequencing primers or barcodes to infer pairs of guide RNAs. We constructed several dual-CRISPR libraries targeting 4,047 UCEs, 1,527 *in vivo*-validated conserved enhancers, and all potential 13,539 enhancers predicted in K562 cells. Using these libraries, the functions of NCREs in cell survival and drug responses were studied in K562 cells and 293T cells.

Results showed that many UCEs might play important roles in cell survival or resistance to imatinib treatment. In the previous in vivo studies using mouse models, genetic editing of a few selected UCEs would not show strong phenotypes, although UCEs are expected to be functionally essential due to their high evolutionary conservation. It is possible that some UCEs may only function in defined tissues or developmental stages. Results from our research and future

investigations may help to narrow down the potential tissues for in vivo studies to better understand these evolutionary puzzles. Intriguingly, many of the UCEs showed silencer activities based on luciferase assays or CRISPR editing (Fig. 1, 2 and 6). Compared to enhancers or insulators, silencers are less well-studied and their roles in different biological pathways are still to be identified[7]. Although silencers are over-represented in the top hits, it could be biased from the screening readouts which looked for growth disadvantages. The real representation of silencers and enhancers in UCEs may benefit from future similar studies focusing on distinct biological pathways.

Many NCREs are expected to regulate multiple genes, which together may contribute to the phenotypes when NCREs were manipulated in their endogenous loci. Our results from studying all potential enhancers predicted from K562 cells showed that many enhancers may show silencer activities or no activities in luciferase assays (Fig. 3e and f), although biological functions were observed and nearby genes were deregulated when these NCREs were removed from their endogenous loci (Fig. 3d and g). This indicates that studying these NCREs in their endogenous loci is needed to define their potential functions and regulated pathways, where the dual-CRISPR screening system and its single-cell application extension are useful. Furthermore, results showed that in their endogenous loci NCREs may have multiple regulatory functions, e.g., acting as both enhancers and silencers (Fig. 3g), which has been previously observed and proposed based on reporter assays[7,18]. Therefore, this dual-CRISPR system could be used to study NCREs with complex regulatory functions or even without prior defined functions (Fig. 1e), which would not be possible using the CRISPR–dCas9-activation/repression systems. It has been shown that many NCREs function in clusters and may have redundant roles. Results showed that the dual-CRISPR system was able to identify not only individual functional NCREs but also NCRE clusters in biological contexts (Fig. 4).

Unlike targeting genes, where genome-wide one-fit-for-all CRISPR libraries could be used for different cell types of the same genome, NCREs often function in a tissue-specific manner, which requires a versatile CRISPR system that could be tailored to target different NCREs with potential distinct functions. This dual CRISPR screening system has multiple advantages over existing similar systems, and we expect that this system will have broad applications in studying the functions of NCREs and other non-coding parts of the genome. Our results also showed that NCREs might play important roles in drug resistance, and we identified a critical UCE that regulates cardiomyocyte differentiation, which may have future translational perspectives.

Methods

Cell culture

K562 cells were cultured in RPMI 1640 + L-Glutamine (Gibco), 10% fetal bovine serum (Biowest), and 1% Penicillin-Streptomycin (Gibco). 293T cells were cultured in DMEM (Gibco), 10% FBS and 1% Penicillin-Streptomycin. Cell density and culture conditions were maintained according to the ENCODE Cell Culture Guidelines. NKX2-5eGFP/w human embryonic stem cells (hESCs) were cultured in StemFlex™ Medium (Thermofisher A3349401) on Biolaminin (LN521-02) coated 6-well plate, and passaged using TrypLE Select (Thermofisher 12563011).

Dual-CRISPR plasmid construction

The lentiviral dual-CRISPR plasmid used for the screening was made based on the lentiGuide-Puro (Addgene #52963). The human U6 and H1 promoters were cloned to replace the U6 promoter of the lentiGuide-Puro plasmid. The plasmid is referred to as Lenti-dual-CRISPR-U6-H1 (pBP43).

dual-CRISPR plasmid containing Cas9 used to generate knockout clones for screening validations was made based on the pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene #62988). The human U6 and H1 promoters were cloned to replace

the U6 promoter in the PX459V2. And the 3XFLAG tag on the Cas9 protein was replaced with the HA tag. The plasmid is referred to as Dual-CRISPR-Cas9-U6-H1 (pBP48).

The dual-scaffold was cloned to the pUC19 backbone. The plasmid is referred to as pUC19-dual-scaffold-1.0 (pBP44) or pUC19-dual-scafffold-2.0 (pBP49).
The gRNAs were designed and cloned as the following:
U6 side: GTGGAAAGGACGAAACACCGN_{20} (gRNA target sequence) GTTTTAGAGCTAGAAATAGC
H1 side: TATGAGACCACTCTTTCCCG N_{20} (gRNA target sequence) GTTTTAGAGCTAGAAATAGC

Dual-CRISPR-Cas9-U6-H1 was digested with BbsI, and the dual-scaffold was isolated from pUC19_dual_scaffold digested with BbsI. Then these four fragments were assembled using NEBuilder HiFi according to the manufacturer's protocol.

Dual-CRISPR library design

The paired guide RNA sequences of the dual-CRISPR library targeting the ultraconserved elements (UCE) and some validated enhancers were from the published computation pipeline CRISPETa (referred as UCE library)[39].

The predicted K562 enhancers based on ENCODE ChIP-seq data using machine learning model were selected as targets for dual CRISPR screening[16]. The software package CRISPRseek was used to search for potential protospacers sequences with PAM NGG pattern as the potential CRISPR targeting regions[74]. Guide RNAs with high predicted cleavage efficiency and specificity were chosen with the following parameters cutoff: *gRNAefficacy* > 0.15, *top5OfftargetTotalScore* < 47 and *top10OfftargetTotalScore* < 50. Guide RNA were further filtered to ensure no overlapping with exons and repetitive regions.

First, the immediate upstream and downstream flanking regions (200 bp in size) of the enhancers were selected for guide RNA design. Single gRNAs from the two flanking regions were paired, resulting in up to 25 pairs of gRNAs per targeting enhancer region (referred to as enhancer library 1 in Fig. 5a targeting 3,995 enhancers). Due to genomic repeats and other constraints, not all enhancers could be targeted by the previous design strategy. To target the rest of the predicted enhancers in K562 cells, pairs of guide RNAs were selected within the enhancer regions (5' proximal and 3' proximal of the enhancers), with around 14 pairs of guide RNAs per enhancer (referred as enhancer library 2 in Fig. 5b targeting 13,020 enhancers). In this way, all predicted enhancers in K562 cells by ENCODE were targeted.

The oligo pools were then designed and ordered from CustomArray/GenScript. For UCE library and enhancer library 1, pairs of gRNAs were designed to be compatible with Lenti-dual-CRISPR-1.0 as following:
ATCTTGTGGAAAGGACGAAACACCG-[gRNA1, 20nt]-gttttgagacgggatccCGTCTCAAAAC
[reverse complement of gRNA2, 20nt]-CGGGAAAGAGTGGTCTCATACAGAACTTAT

For enhancer library 2, pairs of gRNAs were designed to be compatible with Lenti-dual-CRISPR-2.0 as following:
ATCTTGTGGAAAGGACGAAACACCG-[gRNA1,20nt]-gttttagagctaGAAAtagcaagttGAGACG-[barcode,10nt]-CGTCTCAACTTGCTATTTCTAGCTCTAAAAC-[reverse complement of gRNA2, 20nt]-CGGGAAAGAGTGGTCTCATACAGAACTTAT.

Each dual-CRISPR library also contains control paired guide RNAs from a previous study, which target promoters, exons, and introns of 17 ribosomal genes and 3 cancer-related genes (FOXA1, HOXB13 and EZH2), non-targeting guide

RNAs, guide RNAs targeting the adeno-associated virus integration site 1 (AAVS1) loci[29].

Single-cell dual-CRISPR library design

For single-cell dual-CRISPR library design, 82 paired guide RNA sequences were selected from the pooled dual CRISPR screen experiments to target 42 significant NCREs affecting cell growth and imatinib resistance, with 1-3 unique pairs of guide RNAs chosen for each region. 10 extra pairs of guide RNAs were selected to target the AAVS1 region as negative controls. The oligo pool was then ordered and assembled into the single-cell dual-CRISPR library.

Construction of the pooled and single-cell dual-CRISPR libraries

The synthesized oligo pools were first amplified by PCR using the following primer:
Forward primer: ATCTTGTGGAAAGGACGAAA;
Reverse primer: ATAAGTTCTGTATGAGACCA;

For Lenti-dual-CRISPR-1.0 libraries, PCR procedures using NEBNext High-Fidelity 2X PCR Master Mix were: 98 °C for 30 s, 18 cycles of 98 °C for 10 s, 68 °C for 30 s and 72 °C for 30 s, and 72 °C for 5 min. For each reaction, 80 ng of the oligo pool was used for a 100 µl PCR reaction, and 20 reactions per library were pooled. The pooled PCR products were further size-selected and gel-purified using QIAGEN MinElute column. In the first step of library construction, the amplified oligo libraries were assembled into the digested Lenti-dual-CRISPR-U6-H1 plasmids using Gibson assembly. The assembly mix was made using 200 ng of digested dual-CRISPR-U6-H1 plasmids, 30 ng insert DNA (at molar ratio 1:10) and 10 µl of 2× Gibson Assembly Master Mix to produce a final volume of 20 µl. The assembly mix was incubated at 50 °C for 60 min, and in total 10 reactions were pooled for each library. The pooled reaction products were purified by ethanol precipitation and resuspended in 10 µl water, from which 2 µl of the products was electroporated into 25 µl of Endura electrocompetent cells (Endura 60242-2). In total 5 electroporation reactions were pooled and grown in 5 ml SOC recovery

673 medium for 2 h. Then 5 µl from the 5 ml SOC recovery medium was used to
674 perform serial dilution plating to determine the transformation efficiency, and the
675 rest was further cultured in 1000 ml LB medium with 100 µg ml^{-1} carbenicillin
676 overnight. The plasmids containing the oligos were extracted using the Qiagen
677 Maxiprep kit, and were further digested with BsmBI to open the plasmids at the
678 restriction enzyme sites placed on oligo inserts. To assemble the final Lenti-dual-
679 CRISPR-1.0 libraries, the BsmBI-digested plasmids containing the oligos were
680 ligated with the dual-scaffold fragments (digested and isolated from pUC19-dual-
681 scaffold-1.0) using T7 DNA ligase (NEB M0318). The ligation mix was made using
682 20 µg of digested plasmids, 1.2 µg of the dual-scaffold fragment, 200 µl of 2× T7
683 buffer and 40 µl T7 DNA ligase to produce a final volume of 400 µl. The ligation
684 mix was incubated at 16°C overnight. The ligation reaction products were purified
685 by ethanol precipitation, resuspended in 210 µl of water, and treated with Plasmid-
686 Safe (Epicentre, E8101K). The products were then purified using QIAGEN
687 MinElute column and eluted in 10 µl of water, which was electroporated into 50 µl
688 of Endura electrocompetent cells and grown in 2 ml SOC recovery medium for 2
689 h. Then 2 µl from the 2 ml SOC recovery medium was used to perform serial
690 dilution plating to determine the transformation efficiency, and the rest was further
691 cultured in 500 ml LB medium with 100 µg ml^{-1} carbenicillin. The final lenti-dual-
692 CRISPR-1.0 libraries were extracted using the Qiagen Maxiprep Kit.
693
694 For Lenti-dual-CRISPR-2.0 libraries, the procedures to clone the oligo pools into
695 the lenti-dual-CRISPR-U6-H1 plasmids were similar. Gibson assembly was used
696 to clone the dual-scaffold-2.0 (with 200 bp random sequences inserted between
697 the two scaffold sequences) to make the final lenti-dual-CRISPR-2.0 libraries.
698
699 ***Lentivirus production***
700 For each library, the 293T cells were grown in five T175 flasks at 50% confluency
701 before transfection. For each flask of 293T cells grown in 25 ml of fresh medium,
702 15 µg of library plasmids, 7 µg of psPAX2, 3.5 µg of pCMV-VSV-G, and 76.5 µl of
703 X-tremeGENE 9 DNA Transfection Reagent were mixed in 1 ml of serum-free

medium and used for transfection. Fresh medium was added the day following transfection. Media supernatant containing virus particles was collected on the second and third days after transfection, pooled and further concentrated using Lenti-X according to the manufacturer's protocol. Virus titer was then determined by making serial (10^{-3} to 10^{-10}) dilutions of 4 µl of frozen virus supernatant in media containing 8 µg ml^{-1} of polybrene to infect 293T cells. Two days after infection, cells were selected with 2 µg ml^{-1} puromycin for an additional 7 d. Virus titer was then calculated based on the survival colonies and the related dilution.

Pooled dual-CRISPR screen

K562/Cas9 cells were infected with the respective virus libraries at a multiplicity of infection = 0.2 by spin-infection. For spin-infection, 3×10^6 cells in each well of a 12-well plate were infected in 1 ml of medium containing 8 µg ml^{-1} of polybrene. In total, four plates were used for each infection to infect a total of 1.5×10^8 cells, which would result in ~300× to ~500× coverage of the dual-CRISPR libraries. Two days after infection, cells were selected by 2 mg ml^{-1} of puromycin for a further 6 d. Dead cells were then removed with Histopaque-1077 (Sigma) by centrifuging cells at 400g for 30 minutes at room temperature. For each biological replicate experiment, the lentivirus was produced again and infection was repeated.

After puromycin selection, an aliquot of 10^8 K562/Cas9 cells infected with the respective dual-CRISPR libraries were frozen as the control population (day 0). From the same cell population, 10^8 K562 cells were further cultured for another 15 days to identify essential NCREs that affect cell growth; and another 10^8 cells were cultured under 0.1 µM imatinib treatment for 15 days. Dead cells were removed, and drugs were refreshed during the subculture of imatinib-treated cells. On day 15, 10^8 cells for each culture were collected and frozen for the next step. For 293T cells, procedures were similar, except that the dead cells were removed by refreshing the medium.

Single-cell RNA-seq combined with the dual-CRISPR assay

The single-cell dual-CRISPR library was introduced into K562-Cas9 expressing cells via spin infection at a final MOI of <0.3. After one week, cells that successfully incorporated the single-cell dual-CRISPR system were sorted based on GFP expression using FACS sort. To prepare for single-cell processing, 0.5×10^6 cells at a concentration of 1,500 cells ml^{-1} in 0.04% BSA-PBS were used for single-cell RNA and guide RNA capture according to the 10x Genomics protocol. The 10x Genomics Chromium platform was used to generate the single-cell libraries according to the manufacturer's protocol, which were sequenced on the Illumina NovaSeq 6000 platform.

Library sequencing

Genomic DNA was isolated using the QIAamp DNA Blood Maxi Kit (QIAGEN). Dual-CRISPR regions integrated into the chromosomes were then PCR amplified using LongAmp Taq DNA Polymerase (NEB, #M0323) using primers annealing to U6 and H1 promoters (Supplementary Table 8). Stagger sequences were introduced to the PCR primers to increase the diversity of the next-generation sequencing libraries, when the flowcell only contains the dual-CRISPR sequencing libraries[75]. The stagger primers are not necessary if libraries are pooled with other sequencing samples with diverse sequences, and a one-step PCR using primers containing Illumina sequencing adaptors will be sufficient. For each 100 µl PCR reaction, 10 µg of genomic DNA, 50 µl of 2× LongAmp master mix, 2 µl of 25 µM U6_stagger primer, 2 µl of 25 µM H1_stagger primer were used. In total, 7 PCR reactions with different pairs of stagger primer were used and pooled, assaying 70 µg of genomic DNA. PCR procedures were: 94 °C for 30 s, 25 cycles of 94 °C for 15 s, 60 °C for 15 s and 65 °C for 60 s, and 65 °C for 10 min. These fragments were then cleaned up and gel-purified using MinElute PCR Purification Kit. Then the Illumina TruSeq adapters were ligated, and sequencing libraries were prepared according to the ENCODE protocol and sequenced on the Illumina HiSeq4000 platform.

Pooled dual-CRISPR screen analysis

765 Cutadapt 3.4 was used to extract the unique 20nt protospacer sequences from
766 each pair of guide RNA sequences by locating the U6/H1 promoter sequences
767 from the 5' end and scaffold sequence from the 3' end of the 20nt protospacer
768 sequence (U6 promoter, ATATATCTTGTGGAAAGGACGAAA; H1 promoter,
769 ATAAGTTCTGTATGAGACCACTCTT)[76]. The read pairs that did not contain the
770 correct promoter and scaffold sequences were not considered. To ensure
771 functional CRISPR guide RNA sequences were counted, the protospacer
772 sequence and the additional 20 bp (for both Lenti-dual-CRISRP-1.0/ Lenti-dual-
773 CRISPR-2.0) into the CRISPR scaffold sequences were kept during the trimming
774 of the sequence reads (tracrRNA sequence, AAGTTAAAAT). Trimmed reads were
775 then mapped to the indexed paired protospacers references generated by Bowtie2
776 based on our initial library designs, and only aligned reads with mapping quality
777 (MAPQ) score over 23 were used for downstream analyses[77].
778
779 MAGeCK RRA was used to identify the significant hits depleted after 15-day
780 culture compared with the day 0 initial cell population, with the cutoff of RRA score
781 < 0.01[40]. MAGeCK MLE was performed to identify NCREs that confer imatinib
782 resistance, by comparing the 15-day imatinib-treated cell population, 15-day
783 culture cell population and day 0 initial cell population[78]. The NCREs whose loss
784 confers Imatinib resistance were identified as regions that were positively selected
785 (i.e., NCREs with beta scores > mean + 2×s.d.) in the 15-day imatinib-treated
786 populations but are weakly selected in the 15-day culture populations.
787 Gene pathway over-representation analyses were based on GO term and KEGG
788 pathway definitions with the cutoff of FDR < 0.25. Genes potentially affected by
789 NCREs were identified as the ones located in the range of +/- 1Mb of the NCRE
790 regions.
791
792 ***Single-cell dual-CRISPR screen analysis***
793 Cell Ranger 6.0.1 pipeline was used to process Chromium single-cell data to align
794 reads, generate feature-barcode matrices for the mRNAs, and capture guide RNAs.
795 Seurat 4.0.2 was used to process the single-cell RNA-seq data[79]. Gene expression

matrix was normalized using the “LogNormalize" method with scale factor of 10,000 and log-transformed. Cell-cycle-related scores were regressed during data scaling. Differential expression tests and p-value calculations were performed by the MAST-fitted model based on whole transcriptome data in all single cells except for genes with mean normalized expression < 0.01[65]. To control the confounding factors in the differential expression testing between perturbed and control cell groups, we included logarithms of the total number of expressed genes per cell and the total number of gRNAs detected per cell as covariates in the MAST regression model to overcome test miscalibration.

TADs around the candidate NCREs were used to narrow down the potential genes deregulated by CRISPR editing. The K562 TAD dataset was downloaded from 3D Genome Browser, and the median window size is around 1 Mbp. The neighboring gene coordinates were extracted using biomaRt[80]. The gene expression changes were quantified as the log2 fold change of mean of normalized gene expression from perturbed cell population divided by the mean of normalized gene expression from the negative control cell population.

RNA-seq sample preparation and analysis

RNA was isolated using ISOLATE II RNA Mini Kit (BIO-52073), and sequencing libraries were made with the Invitrogen Collibri Stranded RNA Library Preparation kit (A39003024), according to the manufacturer’s protocol. The whole-transcriptome sequencing was done on the Illumina NovaSeq 6000 platform.

Snakemake pipeline (https://github.com/snakemake-workflows/rna-seq-star-deseq2) was used to process the bulk RNA-seq sample. Briefly, Cutadapt v3.4 was used to trim adapters from reads, STAR was used to align the spliced transcripts to the reference genome (GRCh37 assembly release 75) and quantify the read counts per gene[81]. DESeq2 was used to perform differential expression analysis[82]. P-values were corrected by the BH procedure for FDR control.

The KEGG pathway enrichment analysis was based on the GSEA method, which integrates the expression level of individual genes and aggregates gene

827 expression in the pathway analysis to manifest the phenotypic differences, i.e., to
828 show if the pathway is activated or suppressed[61].
829
830 ***Enhancer cluster analysis***
831 Individual enhancers from the dual-CRISPR libraries (enhancer libraries 1 and 2)
832 were first clustered based on the super-enhancer annotations in K562 cells[60]. Two
833 computational methods GSEA and MAGeCK RRA were used to compute the
834 depletion scores of enhancer clusters associated with cell growth[40,61]. P-values
835 obtained by the two methods are corrected by the BH procedure for FDR control.
836 Among the top 20 clusters identified from both methods, 10 clusters were shared,
837 which were considered for further validation.
838
839 ***Luciferase assay***
840 Candidate NCREs were amplified with primers containing a homologous arm from
841 the genomic DNA of K562 cells. These fragments were then inserted in front of the
842 promoters of the luciferase plasmids pGL4.23 (Promega, with some modification
843 on the cloning sites, detecting the enhancer activity) and pGL4.53 (Promega,
844 detecting the silencer activity) by using NEBuilder HiFi. Cells were then co-
845 transfected with the pRL-CMV Renilla reporter vector and the pGL4.53/modified
846 pGL4.23 vectors with the NCRE sequence inserted. The luciferase assay was
847 performed using the Dual-Luciferase Reporter Assay System from Promega
848 according to the manufacturer's protocol. Original luciferase plasmid without any
849 insertion was used as the control. All luciferase assays were from three
850 independent transfections performed on different days. All tested regions and
851 associated primers are listed in Supplementary Table 8.
852
853 ***Dual-CRISPR–Cas9-guided NCRE knockout***
854 Paired guide RNAs targeting the 5' and 3' ends of the NCREs were selected from
855 the screening libraries. All selected guide design falling in intron regions were
856 checked to be at least 10bp away from adjacent splicing sites. The gRNA
857 sequences were cloned into the dual-CRISPR-Cas9-U6-H1 (pBP48) plasmid

containing the gRNA scaffold and Cas9 sequence. K562 cells were transfected with the plasmids containing the respective pairs of guide RNAs and then selected for successful transfection using puromycin. K562 cells transfected with the dual-CRISPR plasmid containing a pair of guide RNAs targeting the GFP sequences were used as the CRISPR-editing control (control_KO). Single clones of cells were picked and verified using PCR and Sanger sequencing.

Quantitative PCR

RNA was extracted using ISOLATE II RNA Mini Kit (Bioline, BIO-52073), and the cDNA was synthesized using SuperScript IV VILO Master Mix (Invitrogen, 11756050). Real-time PCR was performed with SensiFAST SYBR No-ROX Kit (Bioline, BIO-98020) on Biorad CFX Opus 384 Real-Time PCR Systems system. The expression of the housekeeping gene *GAPDH* was used as the control.

Cell proliferation assay

Cell proliferation assay was performed by mixing the NCRE_KO cell lines with cells expressing GFP at a 1:1 ratio. The changes in GFP percentage were monitored at indicated time points by FACS. Ctrl_KO cells were used as the negative control.
To test the imatinib response, NCRE_KO clones and K562 cells were seeded into a 96-well plate and treated with 0,4 µM Imatinib (STI571, Selleckchem) for 3 days. The cell viability was then measured using the CellTiter-Blue viability assay (Promega, G8082). Relative survival was normalized to their respective untreated NCRE_KO clones and K562 cells and corrected for background signal.

CRISPR–Cas9-guided NCRE knockout in stem cells

The crRNAs targeting the 5' and 3' ends of UCE PAX6_Tarzan were ordered from Integrated DNA Technologies (IDT). To form the crRNA: tracrRNA duplex, 2.2 µl of crRNA and 2.2 µl of tracrRNA (IDT) were mixed in 0.6 µl Nuclease-free duplex buffer. The mix was heated at 95°C for 5 min, then cooled to room temperature for 10 min. Then 0.5 µl of crRNA: tracrRNA duplex, 0.24 µl of Cas9 Nuclease V3 (IDT), and 0.76 µl of buffer R (Neon™ Transfection System) were mixed and incubated

for 20 min at room temperature to form the CRISPR–Cas9 complex. Electroporation was performed by mixing 0.25 million NKX2-5$^{eGFP/w}$ hESCs in 22 µl buffer R with 2 µl of CRISPR–Cas9 complex, according to the protocol of Neon nucleofector. Cells were then transferred to a laminin-coated 12-well plate for culturing. Genomic DNA was isolated 4 days later, and the deletion of PAX6_Tarzan was verified using PCR and Sanger sequencing. Primer and guide RNA sequences are listed in Table S8. The maintenance of NKX2-5eGFP/w hESCs and cardiomyocyte differentiation were performed as described previously[83].

Acknowledgments

This work was supported by a Gisela Thier Fellowship from Leiden University Medical Center, a KWF Young Investigator Grant from the Dutch Cancer Society (11707), and an ERC Starting Grant from the European Research Council (950655-Silencer) awarded to B.P.

Author contributions

Conceptualization, B.P.; Methodology, B.P., and Y.L.; Investigation, Y.L., B.P., A.H., and S.S.; Software and Formal Analysis, X.L., M.T., Y.L., and N.X.; Writing – Original Draft, B.P., and Y.L.; Writing – Review & Editing, B.P., M.T., N.X., A.H., S.S., X.L., R.D., H.M., J.J.G., and M.P.S.; Resources, Y.A., S.L.K., R.D., H.M., J.J.G., and M.P.S.; Supervision, B.P., X.L., and M.P.S.; Funding Acquisition, B.P.

Competing interests

M.P.S. is a founder and member of the science advisory boards of Personalis, SensOmics, Qbio, January, Mirvie and Filtricine, and a member of the science advisory boards of Genapsys and Epinomics.

References

1 Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53, doi:10.1126/science.abj6987 (2022).

2 Abascal, F. *et al.* Perspectives on ENCODE. *Nature* **583**, 693-698, doi:10.1038/s41586-020-2449-8 (2020).

3 Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews Genetics* **21**, 292-310, doi:10.1038/s41576-019-0209-0 (2020).

4 Pang, B. & Snyder, M. P. Systematic identification of silencers in human cells. *Nature Genetics* **52**, 254-263, doi:10.1038/s41588-020-0578-5 (2020).

5 Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics* **53**, 1064-1074, doi:10.1038/s41588-021-00863-6 (2021).

6 van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol* **35**, 145-153, doi:10.1038/nbt.3754 (2017).

7 Pang, B., van Weerd, J. H., Hamoen, F. L. & Snyder, M. P. Identification of non-coding silencer elements and their regulation of gene expression. *Nature Reviews Molecular Cell Biology*, doi:10.1038/s41580-022-00549-9 (2022).

8 Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710, doi:10.1038/s41586-020-2493-4 (2020).

9 Zhou, X. *et al.* Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. *Nat Biotechnol* **33**, 345-346, doi:10.1038/nbt.3158 (2015).

10 Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

11 Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology* **28**, 1045-1048, doi:10.1038/nbt1010-1045 (2010).

12 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).

13 Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212-224, doi:10.1016/j.cell.2010.09.009 (2010).

14 Kim, T. H. *et al.* Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell* **128**, 1231-1245, doi:10.1016/j.cell.2006.12.048 (2007).

15 Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396, doi:10.1016/s0092-8674(00)81967-4 (1999).

16 Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol* **13**, R48, doi:10.1186/gb-2012-13-9-r48 (2012).

17 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).

18 Gisselbrecht, S. S. *et al.* Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Mol Cell* **77**, 324-337.e328, doi:10.1016/j.molcel.2019.10.004 (2020).

19 Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:doi:10.1126/science.1225829 (2012).

20 Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84-87, doi:doi:10.1126/science.1247005 (2014).

21 Klann, T. S., Black, J. B. & Gersbach, C. A. CRISPR-based methods for high-throughput annotation of regulatory DNA. *Current Opinion in Biotechnology* **52**, 32-41, doi:https://doi.org/10.1016/j.copbio.2018.02.004 (2018).

22 Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature Reviews Molecular Cell Biology* **17**, 5-15, doi:10.1038/nrm.2015.2 (2016).

23 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology* **32**, 347-355, doi:10.1038/nbt.2842 (2014).

24 Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M. B. Annotating non-coding regions of the genome. *Nat Rev Genet* **11**, 559-571, doi:10.1038/nrg2814 (2010).

25 Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* **34**, 192-198, doi:10.1038/nbt.3450 (2016).

26 Tycko, J. *et al.* Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat Commun* **10**, 4063, doi:10.1038/s41467-019-11955-7 (2019).

27 Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology* **33**, 510-517, doi:10.1038/nbt.3199 (2015).

28 Klann, T. S. *et al.* CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nature Biotechnology* **35**, 561-568, doi:10.1038/nbt.3853 (2017).

29 Zhu, S. *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nature Biotechnology* **34**, 1279-1286, doi:10.1038/nbt.3715 (2016).

30 Chen, P. B. *et al.* Systematic discovery and functional dissection of enhancers needed for cancer cell fitness and proliferation. *Cell Rep* **41**, 111630, doi:10.1016/j.celrep.2022.111630 (2022).

31 Dimitrieva, S. & Bucher, P. UCNEbase--a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res* **41**, D101-109, doi:10.1093/nar/gks1092 (2013).

32 Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92, doi:10.1093/nar/gkl822 (2007).

33 Bejerano, G. *et al.* Ultraconserved Elements in the Human Genome. *Science* **304**, 1321-1325, doi:doi:10.1126/science.1098119 (2004).

34 Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research* **15**, 1034-1050, doi:10.1101/gr.3715005 (2005).

35 Miller, W. *et al.* 28-Way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Research* **17**, 1797-1808, doi:10.1101/gr.6761107 (2007).

36 Wong, E. S. *et al.* Deep conservation of the enhancer regulatory code in animals. *Science* **370**, doi:10.1126/science.aax8137 (2020).

37 Dermitzakis, E. T., Reymond, A. & Antonarakis, S. E. Conserved non-genic sequences — an unexpected feature of mammalian genomes. *Nature Reviews Genetics* **6**, 151-157, doi:10.1038/nrg1527 (2005).

38 Snetkova, V. *et al.* Ultraconserved enhancer function does not require perfect sequence conservation. *Nature Genetics* **53**, 521-528, doi:10.1038/s41588-021-00812-3 (2021).

39 Pulido-Quetglas, C. *et al.* Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion. *PLoS Comput Biol* **13**, e1005341, doi:10.1371/journal.pcbi.1005341 (2017).

40 Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology* **15**, 554, doi:10.1186/s13059-014-0554-4 (2014).

41 Krijger, P. H. L. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nature Reviews Molecular Cell Biology* **17**, 771-782, doi:10.1038/nrm.2016.138 (2016).

42 Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology* **19**, 151, doi:10.1186/s13059-018-1519-9 (2018).

43 Thedieck, K. *et al.* PRAS40 and PRR5-like protein are new mTOR interactors that regulate apoptosis. *PLoS One* **2**, e1217, doi:10.1371/journal.pone.0001217 (2007).

44 Veeriah, S. *et al.* The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proceedings of the National Academy of Sciences* **106**, 9435-9440, doi:doi:10.1073/pnas.0900571106 (2009).

45 Bal, E. *et al.* Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. *Nature* **607**, 808-815, doi:10.1038/s41586-022-04906-8 (2022).

46 Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458, doi:10.1126/science.aad9024 (2016).

47 Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nature Genetics* **47**, 818-821, doi:10.1038/ng.3335 (2015).

48 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).

49 Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* **14**, 756-780, doi:10.1038/s41596-018-0113-7 (2019).

50 Shahi, P. *et al.* The Transcriptional Repressor ZNF503/Zeppo2 Promotes Mammary Epithelial Cell Proliferation and Enhances Cell Invasion. *J Biol Chem* **290**, 3803-3813, doi:10.1074/jbc.M114.611202 (2015).

51 Shahi, P. *et al.* ZNF503/Zpo2 drives aggressive breast cancer progression by down-regulation of GATA3 expression. *Proc Natl Acad Sci U S A* **114**, 3169-3174, doi:10.1073/pnas.1701690114 (2017).

52 Blobel, G. A., Higgs, D. R., Mitchell, J. A., Notani, D. & Young, R. A. Testing the super-enhancer concept. *Nature Reviews Genetics* **22**, 749-755, doi:10.1038/s41576-021-00398-w (2021).

53 Martinez-Ara, M., Comoglio, F., van Arensbergen, J. & van Steensel, B. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol Cell* **82**, 2519-2531.e2516, doi:10.1016/j.molcel.2022.04.009 (2022).

54 Grosveld, G. *et al.* The chronic myelocytic cell line K562 contains a breakpoint in bcr and produces a chimeric bcr/c-abl transcript. *Mol Cell Biol* **6**, 607-616, doi:10.1128/mcb.6.2.607-616.1986 (1986).

55 Groffen, J. *et al.* Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell* **36**, 93-99, doi:10.1016/0092-8674(84)90077-1 (1984).

56 Zhu, Y., Liang, S., Pan, H., Cheng, Z. & Rui, X. Inhibition of miR-1247 on cell proliferation and invasion in bladder cancer through its downstream target of RAB36. *J Biosci* **43**, 365-373 (2018).

57 Møller, G. M., Frost, V., Melo, J. V. & Chantry, A. Upregulation of the TGFbeta signalling pathway by Bcr-Abl: implications for haemopoietic cell growth and chronic myeloid leukaemia. *FEBS Lett* **581**, 1329-1334, doi:10.1016/j.febslet.2007.02.048 (2007).

58 Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nature Reviews Genetics* **22**, 324-336, doi:10.1038/s41576-020-00311-x (2021).

59 Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239-243, doi:10.1038/nature25461 (2018).

60 Jiang, Y. *et al.* SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res* **47**, D235-d243, doi:10.1093/nar/gky1025 (2019).

61 Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550, doi:doi:10.1073/pnas.0506580102 (2005).

62 Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat Methods* **17**, 629-635, doi:10.1038/s41592-020-0837-5 (2020).

63 Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377-390.e319, doi:10.1016/j.cell.2018.11.029 (2019).

64 Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol* **38**, 954-961, doi:10.1038/s41587-020-0470-y (2020).

65 Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278, doi:10.1186/s13059-015-0844-5 (2015).

66 Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nature Methods* **17**, 629-635, doi:10.1038/s41592-020-0837-5 (2020).

67 Duan, L. *et al.* EI24 Inhibits Cell Proliferation and Drug Resistance of Esophageal Squamous Cell Carcinoma. *Front Oncol* **10**, 1570, doi:10.3389/fonc.2020.01570 (2020).

68 Choi, J. M. *et al.* Reduced expression of EI24 confers resistance to gefitinib through IGF-1R signaling in PC9 NSCLC cells. *Lung Cancer* **90**, 175-181, doi:10.1016/j.lungcan.2015.08.019 (2015).

69 Shui, J. W., Hu, M. C. & Tan, T. H. Conditional knockout mice reveal an essential role of protein phosphatase 4 in thymocyte development and pre-T-cell receptor signaling. *Mol Cell Biol* **27**, 79-91, doi:10.1128/mcb.00799-06 (2007).

70 Luo, J. *et al.* BolA family member 2 enhances cell proliferation and predicts a poor prognosis in hepatocellular carcinoma with tumor hemorrhage. *J Cancer* **10**, 4293-4304, doi:10.7150/jca.31829 (2019).

71 Snetkova, V. *et al.* Ultraconserved enhancer function does not require perfect sequence conservation. *Nat Genet* **53**, 521-528, doi:10.1038/s41588-021-00812-3 (2021).

72 Ahituv, N. *et al.* Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5**, e234, doi:10.1371/journal.pbio.0050234 (2007).

73 Elliott, D. A. *et al.* NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat Methods* **8**, 1037-1040, doi:10.1038/nmeth.1740 (2011).

74 Zhu, L. J., Holmes, B. R., Aronin, N. & Brodsky, M. H. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One* **9**, e108424, doi:10.1371/journal.pone.0108424 (2014).

75 Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863, doi:10.1038/nprot.2017.016 (2017).

76 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

77 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

78 Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* **16**, 281, doi:10.1186/s13059-015-0843-6 (2015).

79 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

80 Smedley, D. *et al.* BioMart--biological queries made easy. *BMC Genomics* **10**, 22, doi:10.1186/1471-2164-10-22 (2009).

81 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

82 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

83 Campostrini, G. *et al.* Generation, functional analysis and applications of isogenic three-dimensional self-aggregating cardiac microtissues from human pluripotent stem cells. *Nat Protoc* **16**, 2213-2256, doi:10.1038/s41596-021-00497-2 (2021).

Figure legends

Fig. 1: Identification of essential UCEs and validated enhancers

(a) MAGeCK algorithm was used to identify significant hits depleted from the cells cultured for an additional 15 days compared to the initial population. The Manhattan plot shows the distribution of all the target regions. Significant hits were above the dash line, indicating the cutoff of MAGeCK RRA score 0.001. Different colors represent different chromosomes. The UCEs selected for downstream analyses were indicated. The NCRE de_novo_1 is an intergenic fragment included in the library design.

(b, c, d, e) Knockout (KO) essential NCREs from K562 cells using the dual-CRISPR system. Upper panel: K562 cells were transfected with the respective guide RNA pairs that target the indicated NCREs. Single-cell clones with the respective NCRE deletion were selected. The blue arrow indicates the intact genomic regions. The red arrow indicates the NCRE deletions. Lower panel: cell proliferation assay was performed by mixing the KO cell lines with cells expressing GFP at a 1:1 ratio. The changes in GFP percentage were monitored at indicated time points by FACS. Cells with dual-CRISPR guide RNAs targeting GFP sequences served as negative controls (Ctrl_KO). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The ratio of cells in the initial mixture was set as 100% (n=3 biological independent samples; values are shown as the mean ± s.d.; PBX3_CI_KO#1 ***P = 0.0003, PBX3_CI_KO#2 ****P < 0.0001, FOXP1_FI_KO#1 ***P = 0.0004, FOXP1_FI_KO#2 ****P < 0.0001, PAX6_Ta_KO#1 ***P = 0.0026, PAX6_Ta _KO#2 ****P < 0.0001, de_novo_1_KO#1 ***P = 0.0016, de_novo_1_KO#2 ***P = 0.0003 calculated using two-way ANOVA).

(f) Enhancer activities were measured by luciferase assay. The respective NCREs were cloned by PCR into the enhancer reporter plasmid pGL4.23 with a minimal promoter. The empty pGL4.23 plasmid was used as the control for the baseline

luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.23 empty plasmids (n = 3 biological independent samples; bars show mean value ± s.d.; de_novo_1 ^{ns}P = 0.1867, FOXP1_FI ^{ns}P = 0.7006, PBX3_CI $^{**}P$ = 0.0065, PAX6_Ta ^{ns}P = 0.7339 calculated using two-tailed unpaired t-test).

(g) Silencer activities were measured by luciferase assay. The respective NCREs were cloned by PCR into the silencer reporter plasmid pGL4.53 with a PGK promoter. The empty pGL4.53 plasmid was used as the control for the baseline luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.53 empty plasmids (n = 3 biological independent samples; bars show mean value ± s.d.; de_novo_1 $^{***}P$ = 0.0001, FOXP1_FI ^{ns}P = 0.1718, PBX3_CI $^{***}P$ = 0.0007, PAX6_Ta ^{ns}P = 0.4754 calculated using two-tailed unpaired t-test).

(h) TADs identified by Hi-C surrounding the NCRE de_novo_1 are shown, and the location of de_novo_1 is indicated by the vertical blue bar. Horizontal yellow and blue bars indicate distinct TADs. Transcription of nearby genes was quantified by qPCR (*n*=3 biological independent samples; bars show mean value ± s.d.; UHRF2:Ctrl_KO ^{ns}P = 0.7381, UHRF2:de_novo_KO#1 ^{ns}P = 0.0786, UHRF2:de_novo_KO#2 ^{ns}P = 0.3125, KDM4C:Ctrl_KO ^{ns}P = 0.0777, KDM4C:de_novo_KO#1 ^{ns}P = 0.0786, KDM4C:de_novo_KO#2 ^{ns}P = 0.3125, PTPRD:Ctrl_KO ^{ns}P = 0.3794, PTPRD:de_novo_KO#1, $^{**}P$ = 0.0063, PTPRD:de_novo_KO#2, $^{**}P$ = 0.0091 calculated using two-tailed unpaired t-test).

(i) The expression level of gene PTPRD was measured by qPCR for K562 cells and K562 cells with CRISPR/dCas9-SAM activation systems targeting the PTPRD gene (PTPRD SAM) (*n*=3 biological independent samples; dots represent mean value ± s.d.; PTPRD SAM1 $^{**}P$ = 0.0057, PTPRD SAM2 $^{*}P$ = 0.0135 calculated using two-tailed unpaired t-test).

(j) Cell proliferation assay for K562 WT and K562 cells with CRISPR/dCas9-SAM activation systems targeting the PTPRD gene (PTPRD SAM). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The

ratio of cells in the initial mixture was set as 100% (*n*=3 biological independent samples; values are shown as the mean ± s.d.; PTPRD SAM1 ***P = 0.0002, PTPRD SAM2 ****P < 0.0001 calculated using two-way ANOVA).

Fig. 2: Drug resistance regulated by UCEs

(a). UCE–drug interactions study in K562 cells. MAGeCK MLE algorithm was used to identify UCEs involved in imatinib resistance based on dual-CRISPR screenings. Three different cell populations were used: the day 0 population, the 15-day imatinib treatment population and the 15-day non-treatment control population. Beta score indicates the degree of selection upon UCE removal relative to the day 0 initial population. The y axis represents beta scores of the15-day imatinib treatment. The x axis shows beta scores of the 15-day non-treatment condition. The horizontal and vertical dashed lines indicate the mean plus or minus one standard deviation of the 15-day imatinib treatment and the 15-day non-treatment control beta score, respectively. The diagonal dashed line indicates the mean plus or minus one standard deviation of the differential beta scores, which can be calculated by subtracting the 15-day non-treatment control from the 15-day imatinib treatment beta score. The orange group shows UCEs conferring imatinib resistance upon removal; the purple group shows UCEs sensitizing cells to imatinib treatment upon removal. Selected UCEs for downstream analyses were marked.

(b) Enhancer activities were measured by luciferase assay. The respective UCEs were cloned by PCR into the enhancer reporter plasmid pGL4.23 with a minimal promoter. The empty pGL4.23 plasmid was used as the control for the baseline luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.23 empty plasmids ($n = 3$ biological independent samples; bars show mean value ± s.d.; ZNF503_Op ^{ns}P = 0.9705 and QKI_Jo **P = 0.0029 calculated using two-tailed unpaired t-test).

(c) Silencer activities were measured by luciferase assay. The respective UCEs were cloned by PCR into the silencer reporter plasmid pGL4.53 with a PGK promoter. The empty pGL4.53 plasmid was used as the control for the baseline luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.53 empty plasmids ($n = 3$ biological independent samples; bars show mean value ± s.d.; ZNF503_Op ***P = 0.0002 and QKI_Jo ****P < 0.0001 calculated using two-tailed unpaired t-test).

(d, e) Knockout (KO) ZNF503_Op and QKI_Jo from K562 cells using the dual-CRISPR system. K562 cells were transfected with the respective guide RNA pairs that target the indicated UCEs. Single-cell clones with the respective UCE deletion were selected. The blue arrow indicates the intact genomic regions. The red arrow indicates the NCRE deletions.

(f) Drug resistance conferred by knockouts of ZNF503_Op and QKI_Jo. The ZNF503_Op and QKI_Jo KO cells were treated with 0.4 μM imatinib for three days. Cell viability was measured by CellTiter-Blue (n=3 biological independent samples; Bars show mean value ± s.d.; ZNF503_Op *P = 0.0366 and QKI_Jo *P = 0.0108 calculated using two-tailed unpaired t-test).

(g, h) Potential genes regulated by the UCEs. TADs identified by Hi-C surrounding ZNF503_Op and QKI_Jo are shown, and the locations are indicated by vertical blue bars. Transcriptions of nearby genes were quantified by qPCR (n = 3 biological independent samples; bars show mean value ± s.d; SAMD8 **P = 0.004941, VDAC8 **P = 0.007008, ZNF503 ***P = 0.000786, PACRG *P = 0.011125, QKI **P = 0.001101 calculated using two-tailed unpaired t-test).

Fig. 3: Identification of essential enhancers in K562 cells

(a) Outline of the design of dual-CRISPR libraries targeting all potential enhancers in K562 cells. Enhancers were predicted by ENCODE based mainly on the combination of H3K4me1 and H3K27ac and other markers. Two strategies were

used to design paired guide RNAs to target all the potential enhancers in K562 cells.

(b) The outline of the optimization of the dual CRISPR system. To have an optimal NGS sequencing efficiency, the distance between the two scaffolds was increased from 50bp to 200bp by nonsense sequences.

(c) Essential enhancers in K562 cells were identified using dual-CRISPR screenings. MAGeCK algorithm was used to identify significant hits depleted from cells cultured for an additional 15 days compared to the initial population. The Manhattan plot shows the distribution of all the target regions. Significant hits were above the dash line, indicating the cutoff of MAGeCK RRA score 0.001. Different colors represent different chromosomes. The essential enhancers selected for downstream analyses were indicated.

(d) Cell proliferation assay was performed by mixing the enhancer E22:23590 KO cell lines with cells expressing GFP. The changes in GFP percentage were monitored at indicated time points by FACS. Cells with dual-CRISPR guide RNAs targeting GFP sequences served as negative controls (Ctrl_KO). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The ratio of cells in the initial mixture was set as 100% (n=3 biological independent samples; values are shown as the mean ± s.d; ****P < 0.0001 calculated using two-way ANOVA).

(e) Enhancer activities were measured by luciferase assay. The essential K562 putative enhancers were cloned by PCR into the enhancer reporter plasmid pGL4.23 with a minimal promoter. The empty pGL4.23 plasmid was used as the control for the baseline luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.23 empty plasmids (n = 3 biological independent samples; bars show mean value ± s.d.; E6:52372 ****P < 0.0001, E7:135735 ^{ns}P = 0.4104, E8:124178 ***P = 0.0002, E12:123591 **P = 0.0010, E14:71791 ****P < 0.0001, E22:23590 ^{ns}P = 0.2432 calculated using two-tailed unpaired t-test).

(f) Silencer activities were measured by luciferase assay. The essential K562 putative enhancers were cloned by PCR into the silencer reporter plasmid pGL4.53 with a PGK promoter. The empty pGL4.53 plasmid was used as the control for the baseline luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.53 empty plasmids (n = 3 biological independent samples; bars show mean value ± s.d.; E6:52372 $*P$ = 0.4655, E7:135735 ^{ns}P = 0.4104, E8:124178 $***P$ = 0.0002, E12:123591 ^{ns}P = 0.2297, E14:71791 $****P$ < 0.0001, E22:23590 ^{ns}P = 0.2432 calculated using two-tailed unpaired t-test).

(g) TADs identified by Hi-C surrounding the E22:23590 are shown and the location of E22:23590 is indicated by the vertical blue bar. Horizontal yellow and blue bars indicate distinct TADs. Transcription of nearby genes was quantified by qPCR (n=3 biological independent samples; bars show mean value ± s.d.; RAB36:KO#1 and KO#2 $****P$ < 0.0001, BCL-ABL:KO#1 $***P$ = 0.0009 and BCL-ABL:KO#2 $**P$ = 0.0021 calculated using two-tailed unpaired t-test).

Fig. 4: Identifying essential cluster of enhancers with redundant functions

(a) Illustration of a cluster of enhancers regulating gene transcription.

(b) Venn diagram indicates the overlapping clusters identified by GSEA and RRA. The plot shows the ranking of the individual enhancers from the top 3 clusters in the original screening analysis. The y axis represents the MAGeCK RRA score. The x axis represents the ranking of the individual enhancers based on the RRA score.

(c) Enhancer activities were measured by luciferase assay. Individual putative enhancer elements from the top 3 enhancer clusters were cloned by PCR into the enhancer reporter plasmid pGL4.23 with a minimal promoter. The empty pGL4.23 plasmid was used as the control for the baseline luciferase activities. The y axis represents the relative unit of luciferase activity compared to that of pGL4.23 empty plasmids (n = 3 biological independent samples; bars show mean value ± s.d.; chr6R1 ^{ns}P = 0.6800, chr6R2 ^{ns}P = 0.1147, chr6R3 ^{ns}P = 0.4200, chr10R1 ^{ns}P =

0.3176, chr10R2 ^{ns}P = 0.1312, chr10R3 ^{ns}P = 0.0721, chr22R1 ***P = 0.0006, chr22R2 **P = 0.0017, chr22R3 *P = 0.0122 calculated using two-tailed unpaired t-test).

(d, e, f) Growth effect of one enhancer (d), two-enhancer (e), three-enhancer (f) removal from the cluster. Cell proliferation assay was performed by mixing the KO cell lines with cells expressing GFP at a 1:1 ratio. The changes in GFP percentage were monitored at indicated time points by FACS. Cells with dual-CRISPR guide RNAs targeting GFP sequences served as negative controls (Ctrl_KO). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The ratio of cells in the initial mixture was set as 100%. R1, R2 and R3 represent the three individual potential enhancers targeted by the dual-CRISPR libraries (n=3 biological independent samples; values are shown as the mean ± s.d.; R2_KO *P =0.0348 and R1+R3_KO, R1+R2_KO, R2+R3_KO, All_KO ****P < 0.0001, calculated using two-way ANOVA).

(g) Upper panel, epigenetic signatures surrounding chr22 enhancer cluster. The RNA Pol2 ChIA-PET loops are shown. The red arrow indicates the location of gene SLC25A1. Lower panel, transcription of *SLC25A1* in different enhancer KO clones were quantified by qPCR (n=3 biological independent samples; values are shown as the mean ± s.d. for each bar; R1_KO **P = 0.0066, R2_KO **P = 0.0075, R3_KO ***P = 0.0001, R1+R2_KO *P = 0.0195, R2+R3_KO **P = 0.0015, R1+R3_KO **P = 0.0062 and All_KO **P = 0.0025 calculated using two-tailed unpaired t-test).

Fig. 5: Single-cell RNA-seq coupled with the dual-CRISPR system

(a) Illustration of adapting the dual-CRISPR system with scRNA-seq. Two distinct capture sequences were inserted into the stem loops of the guide RNA scaffolds. The guide RNAs could then be captured together with the mRNA within individual single cells.

(b) Transcriptional changes of genes surrounding PAX6_Ta measured by bulk RNA-seq. The y axis shows the logarithm fold change of genes between the PAX6_Ta KO clone and control K562 cells. The x axis indicates the genomic coordinate. The dashed line indicates the location of PAX6_Ta. Distinct colors represent different TADs.
(c) Transcriptional changes of genes surrounding PAX6_Ta measured by scRNA-seq. The y axis shows the logarithm fold change of the captured genes between single cells with the guide RNA pair targeting PAX_Ta and control K562 single cells. The x axis indicates the genomic coordinate. Dashed line indicates the location of PAX6_Ta. Distinct colors represent different TADs.
(d) The upper panel shows epigenetic signatures surrounding enhancer E11:125334 as indicated by the red line. The lower panel shows the differential gene expression testing results. Violin plots show the normalized expression levels of candidate genes in perturbed (188 cells) and control (4282 cells) groups. The gene *EI24* was significantly down-regulated (*P* value < 0.05, calculated by MAST fitted model; n.s., not significant).

Fig. 6: UCE PAX6_Ta regulating the cardiomyocyte differentiation in hESCs.
(a) KEGG pathway enrichment analysis of genes differentially regulated in PAX6_Ta KO cells (+/-5Mbp around PAX6_Ta). The Y-axis represents the significantly enriched KEGG pathways, and the x-axis represents their corresponding gene ratios, i.e., the ratio between the number of unique genes in a specific pathway and the number of unique genes mapped to all these pathways. The left panel presents all significantly up-regulated pathways, and the right panel presents all down-regulated ones. The dot size denotes the number of genes mapped in a pathway, and the color indicates the significance level of the enrichment (FDR < 0.25).

(b) Silencer activities were measured by the repressive ability on the super-core (SC) promoter driving the expression of GFP. A previously published silencer served as the positive control (validated silencer). The empty vector (SC promoter) was used as the control to normalize the respective repressive activities (n=3

biological independent samples; values are shown as the mean ± s.d. for each bar; Validated silencer ***P = 0.0002 and PAX6_Ta ****P < 0.0001 calculated by using two-tailed unpaired t-test).

(c) Knockout (KO) PAX6_Ta in NKX2-5$^{eGFP/w}$ hESC. PCR results show the removal of PAX6_Ta in NKX2-5$^{eGFP/w}$ hESC bulk cells.

(d) The removal of PAX6_Ta caused the defect of cardiomyocyte differentiation, as measured by NKX2-5/GFP expression. Upper panel, NKX2-5$^{eGFP/w}$ hESC cells (hESC_NKX) after differentiation; lower panel, NKX2-5$^{eGFP/w}$ hESC PAX6_Ta KO bulk cells (PAX6_Ta_KO) after differentiation (Scale bar: 275 μm).

(e) Quantification of the cardiomyocyte differentiation as shown in (d) by FACS. The cardiomyocyte differentiation efficiency measured by GFP in NKX2-5$^{eGFP/w}$ hESC was set to 100%. (n=2 biological independent samples; values are shown as the mean ± s.d.; PAX6_Ta_KO **P = 0.0081, calculated using two-tailed unpaired t-test).

(f) The transcription of gene RCN1 was quantified by qPCR before and after differentiation in both NKX2-5eGFP/w hESC cells (hESC_NKX) and NKX2-5eGFP/w hESC PAX6_Ta KO bulk cells (PAX6_Ta_KO) (n=2 biological independent samples; values are shown as the mean ± s.d.; hECS-CM: PAX6_Ta_KO *P = 0.0252, calculated using two-tailed unpaired t-test).

Figure 1

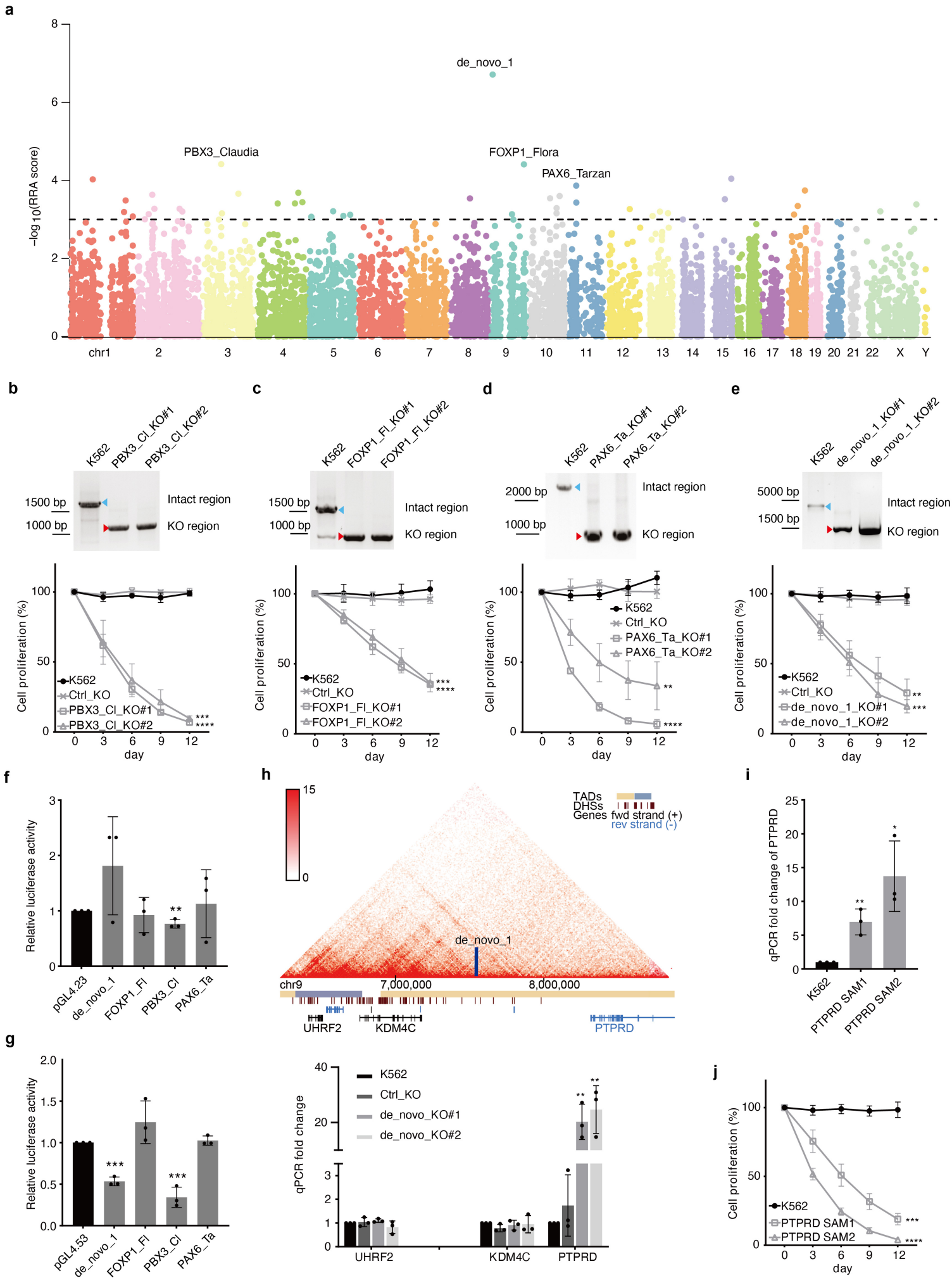

a
de_novo_1
PBX3_Claudia
FOXP1_Flora
PAX6_Tarzan
−log10(RRA score)
chr1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
b
K562
PBX3_Cl_KO#1
PBX3_Cl_KO#2
1500 bp
1000 bp
Intact region
KO region
Cell proliferation (%)
day
K562
Ctrl_KO
PBX3_Cl_KO#1
PBX3_Cl_KO#2
c
FOXP1_Fl_KO#1
FOXP1_Fl_KO#2
FOXP1_Fl_KO#1
FOXP1_Fl_KO#2
d
PAX6_Ta_KO#1
PAX6_Ta_KO#2
2000 bp
PAX6_Ta_KO#1
PAX6_Ta_KO#2
e
de_novo_1_KO#1
de_novo_1_KO#2
5000 bp
de_novo_1_KO#1
de_novo_1_KO#2
f
Relative luciferase activity
pGL4.23
de_novo_1
FOXP1_Fl
PBX3_Cl
PAX6_Ta
h
TADs
DHSs
Genes fwd strand (+)
rev strand (-)
de_novo_1
chr9
7,000,000
8,000,000
UHRF2
KDM4C
PTPRD
i
qPCR fold change of PTPRD
K562
PTPRD SAM1
PTPRD SAM2
g
Relative luciferase activity
pGL4.53
de_novo_1
FOXP1_Fl
PBX3_Cl
PAX6_Ta
qPCR fold change
K562
Ctrl_KO
de_novo_KO#1
de_novo_KO#2
UHRF2
KDM4C
PTPRD
j
Cell proliferation (%)
day
K562
PTPRD SAM1
PTPRD SAM2

Figure 2

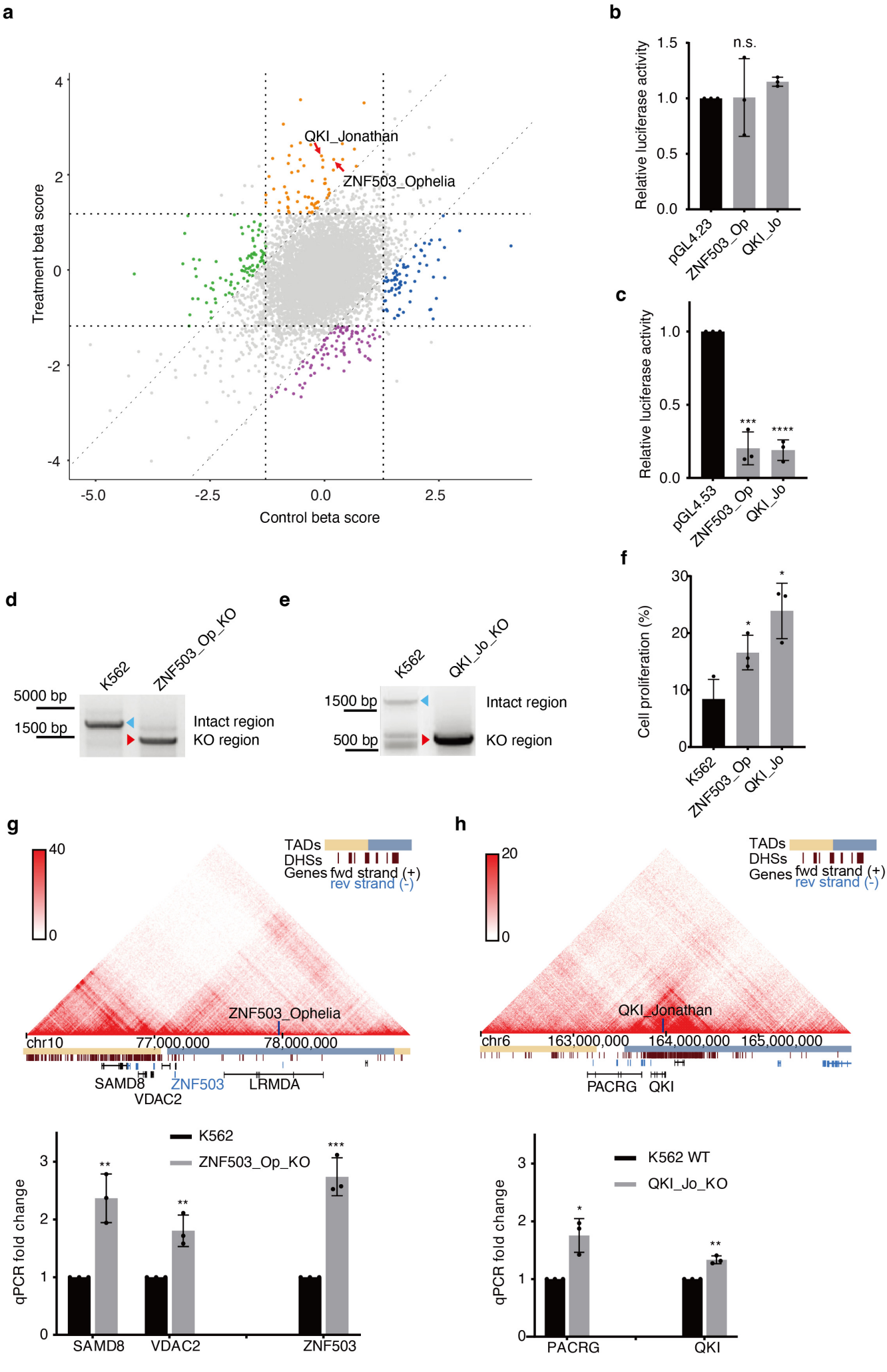

a
QKI_Jonathan
ZNF503_Ophelia
Treatment beta score
Control beta score
b
n.s.
Relative luciferase activity
pGL4.23
ZNF503_Op
QKI_Jo
c
Relative luciferase activity
pGL4.53
ZNF503_Op
QKI_Jo
d
K562
ZNF503_Op_KO
5000 bp
1500 bp
Intact region
KO region
e
K562
QKI_Jo_KO
1500 bp
500 bp
Intact region
KO region
f
Cell proliferation (%)
K562
ZNF503_Op
QKI_Jo
g
TADs
DHSs
Genes fwd strand (+)
rev strand (-)
ZNF503_Ophelia
chr10
77,000,000
78,000,000
SAMD8
ZNF503
LRMDA
VDAC2
h
TADs
DHSs
Genes fwd strand (+)
rev strand (-)
QKI_Jonathan
chr6
163,000,000
164,000,000
165,000,000
PACRG
QKI
qPCR fold change
K562
ZNF503_Op_KO
SAMD8
VDAC2
ZNF503
K562 WT
QKI_Jo_KO
PACRG
QKI

Figure 3

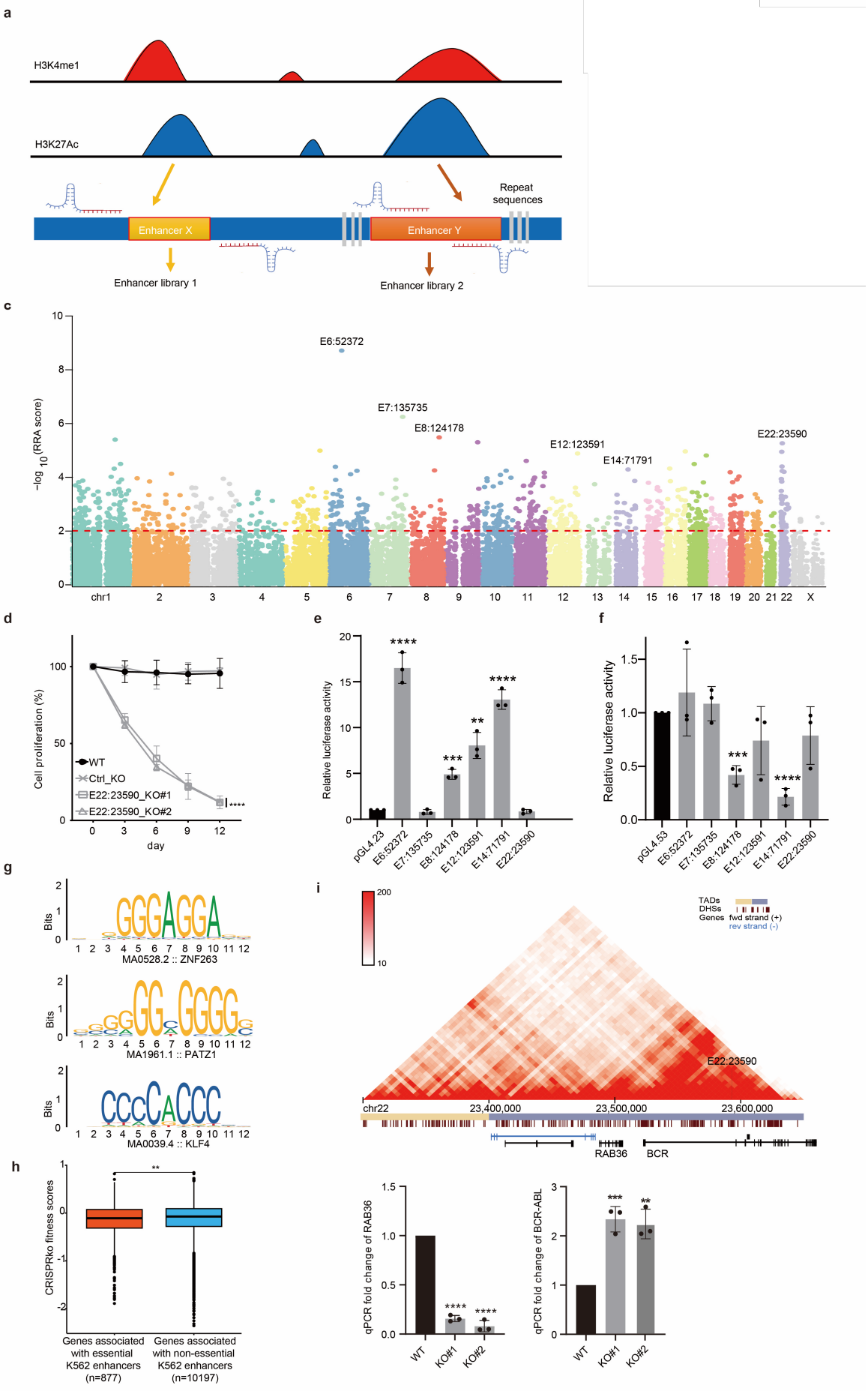

a
H3K4me1
H3K27Ac
Repeat sequences
Enhancer X
Enhancer Y
Enhancer library 1
Enhancer library 2
c
E6:52372
E7:135735
E8:124178
E12:123591
E14:71791
E22:23590
−log 10 (RRA score)
chr1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
d
Cell proliferation (%)
WT
Ctrl_KO
E22:23590_KO#1
E22:23590_KO#2
****
day
e
Relative luciferase activity
pGL4.23
E6:52372
E7:135735
E8:124178
E12:123591
E14:71791
E22:23590
f
Relative luciferase activity
pGL4.53
E6:52372
E7:135735
E8:124178
E12:123591
E14:71791
E22:23590
g
Bits
MA0528.2 :: ZNF263
MA1961.1 :: PATZ1
MA0039.4 :: KLF4
h
CRISPRko fitness scores
Genes associated with essential K562 enhancers (n=877)
Genes associated with non-essential K562 enhancers (n=10197)
i
TADs
DHSs
Genes
fwd strand (+)
rev strand (-)
E22:23590
chr22
23,400,000
23,500,000
23,600,000
RAB36
BCR
qPCR fold change of RAB36
WT
KO#1
KO#2
qPCR fold change of BCR-ABL
WT
KO#1
KO#2

Figure 4

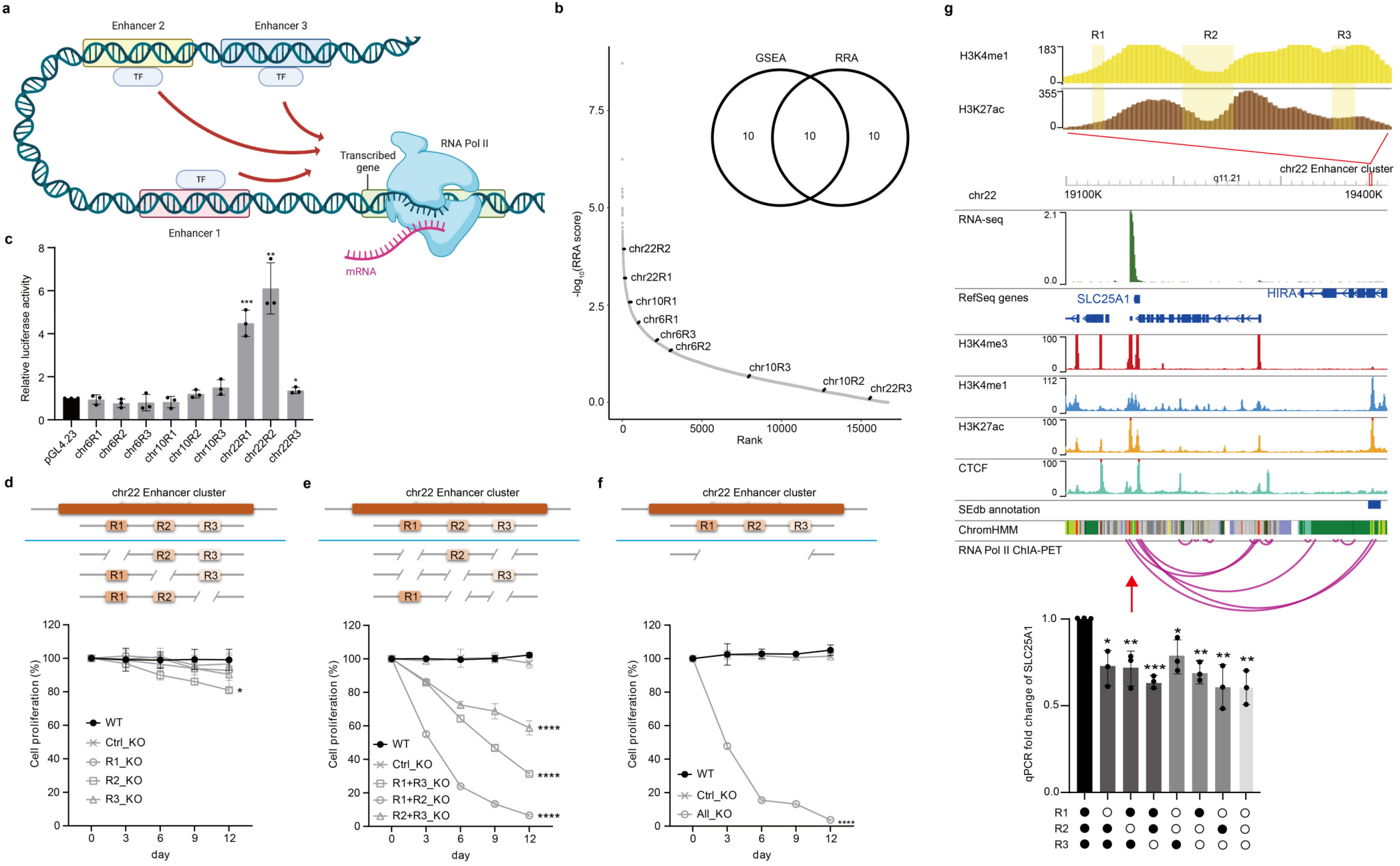

a
Enhancer 2
Enhancer 3
TF
TF
TF
Transcribed gene
RNA Pol II
Enhancer 1
mRNA
b
GSEA
RRA
10
10
10
-log10(RRA score)
Rank
chr22R2
chr22R1
chr10R1
chr6R1
chr6R3
chr6R2
chr10R3
chr10R2
chr22R3
c
Relative luciferase activity
pGL4.23
chr6R1
chr6R2
chr6R3
chr10R1
chr10R2
chr10R3
chr22R1
chr22R2
chr22R3
d
chr22 Enhancer cluster
R1
R2
R3
Cell proliferation (%)
day
WT
Ctrl_KO
R1_KO
R2_KO
R3_KO
e
chr22 Enhancer cluster
Cell proliferation (%)
day
WT
Ctrl_KO
R1+R3_KO
R1+R2_KO
R2+R3_KO
f
chr22 Enhancer cluster
Cell proliferation (%)
day
WT
Ctrl_KO
All_KO
g
R1
R2
R3
H3K4me1
H3K27ac
chr22 Enhancer cluster
chr22
19100K
19400K
RNA-seq
RefSeq genes
SLC25A1
HIRA
H3K4me3
H3K4me1
H3K27ac
CTCF
SEdb annotation
ChromHMM
RNA Pol II ChIA-PET
qPCR fold change of SLC25A1
R1
R2
R3

Figure 5

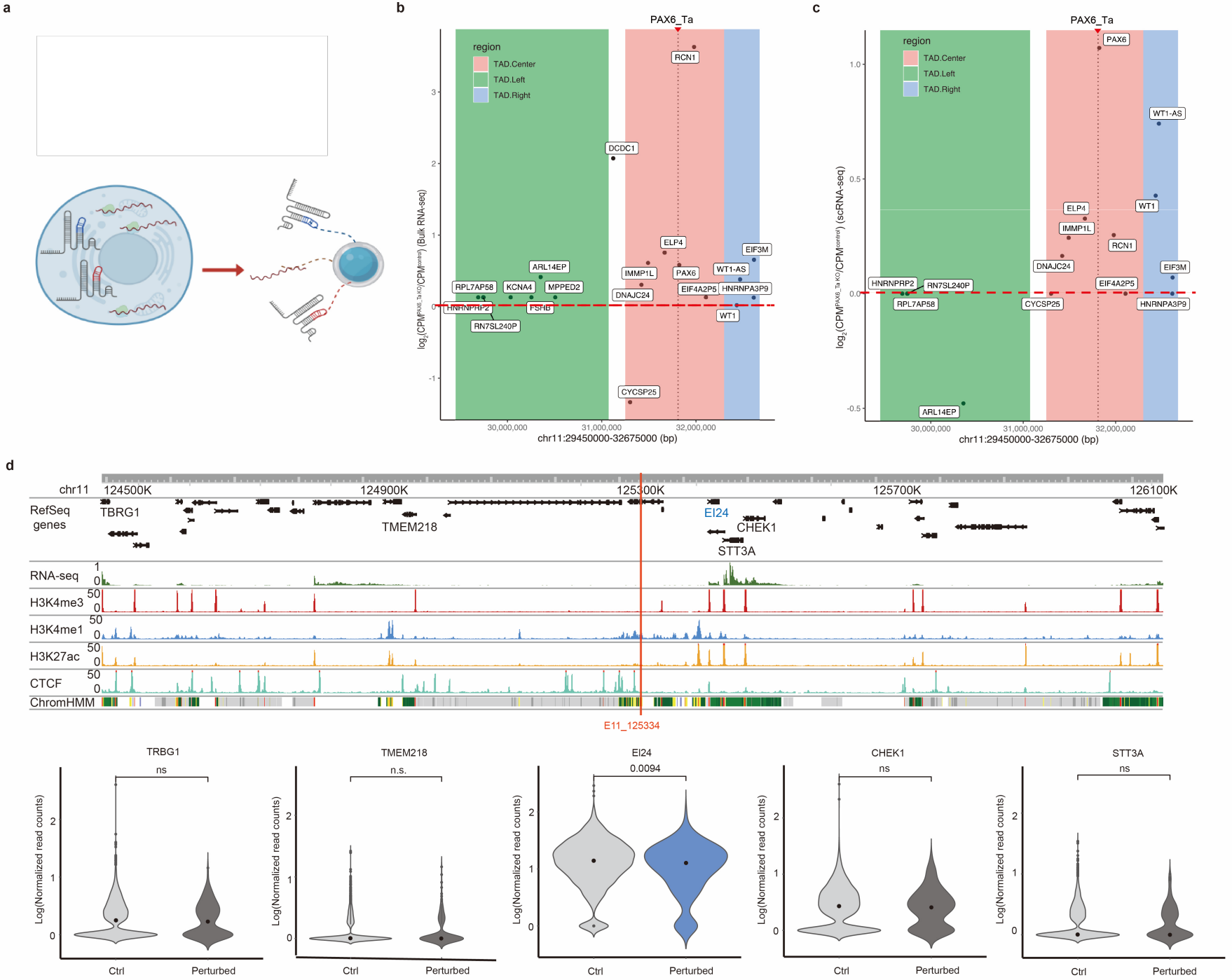
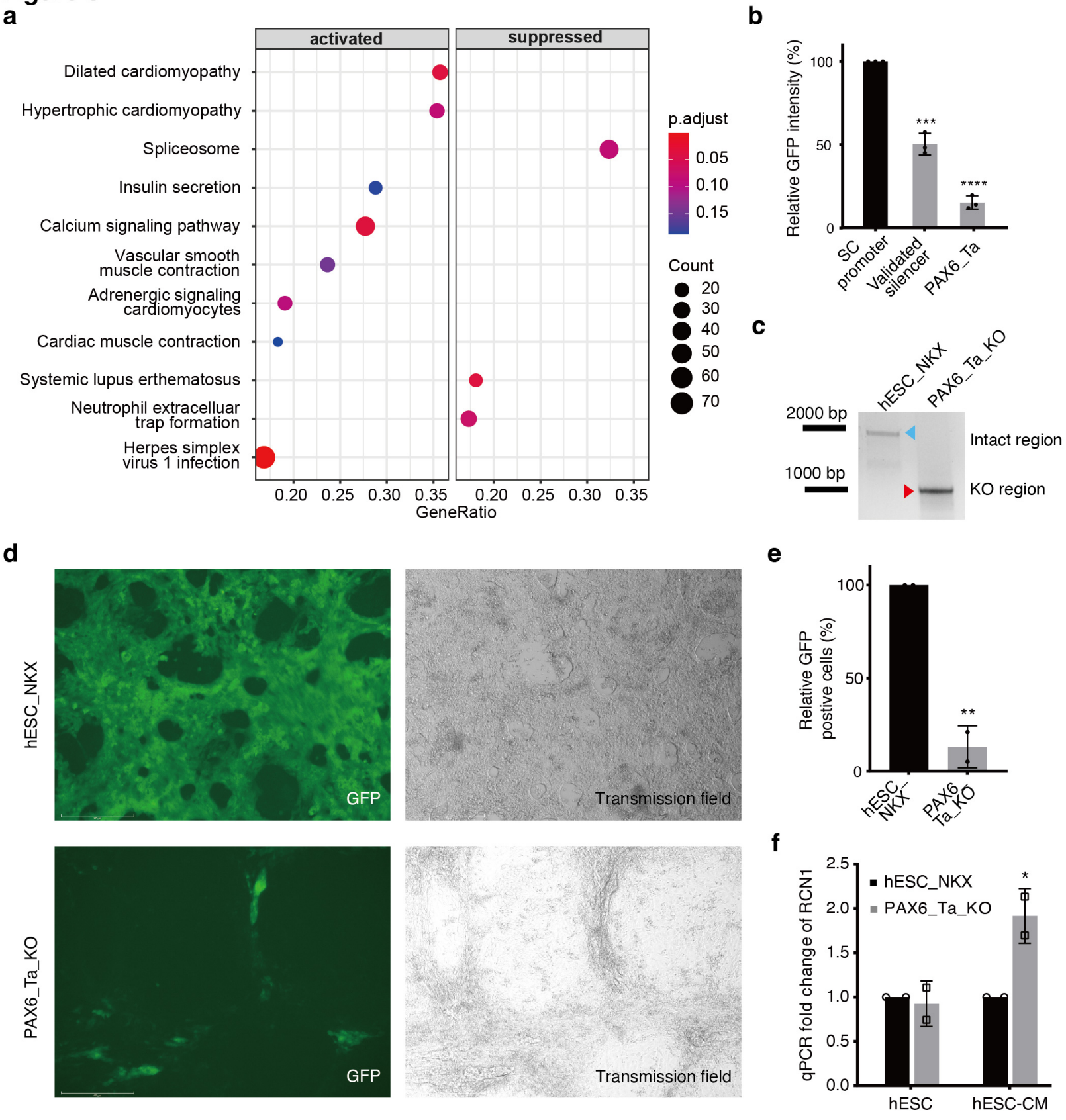

a
b
PAX6_Ta
region
TAD.Center
TAD.Left
TAD.Right
RCN1
DCDC1
ELP4
EIF3M
ARL14EP
IMMP1L
PAX6
WT1-AS
RPL7AP58
KCNA4
MPPED2
DNAJC24
EIF4A2P5
HNRNPA3P9
HNRNPRP2
FSHB
WT1
RN7SL240P
CYCSP25
chr11:29450000-32675000 (bp)
c
PAX6_Ta
PAX6
WT1-AS
ELP4
WT1
IMMP1L
RCN1
DNAJC24
EIF3M
HNRNPRP2
RN7SL240P
EIF4A2P5
RPL7AP58
CYCSP25
HNRNPA3P9
ARL14EP
chr11:29450000-32675000 (bp)
d
chr11
124500K
124900K
125300K
125700K
126100K
RefSeq genes
TBRG1
TMEM218
EI24
CHEK1
STT3A
RNA-seq
H3K4me3
H3K4me1
H3K27ac
CTCF
ChromHMM
E11_125334
TRBG1
ns
TMEM218
n.s.
EI24
0.0094
CHEK1
ns
STT3A
ns
Log(Normalized read counts)
Ctrl
Perturbed

Figure 6


a
activated
suppressed
Dilated cardiomyopathy
Hypertrophic cardiomyopathy
Spliceosome
Insulin secretion
Calcium signaling pathway
Vascular smooth muscle contraction
Adrenergic signaling cardiomyocytes
Cardiac muscle contraction
Systemic lupus erthematosus
Neutrophil extracelluar trap formation
Herpes simplex virus 1 infection
0.20 0.25 0.30 0.35
0.20 0.25 0.30 0.35
GeneRatio
p.adjust
0.05
0.10
0.15
Count
20
30
40
50
60
70
b
Relative GFP intensity (%)
100
50
0
***
****
SC promoter
Validated silencer
PAX6_Ta
c
hESC_NKX
PAX6_Ta_KO
2000 bp
1000 bp
Intact region
KO region
d
hESC_NKX
PAX6_Ta_KO
GFP
Transmission field
GFP
Transmission field
e
Relative GFP postive cells (%)
100
50
0
**
hESC_NKX
PAX6_Ta_KO
f
qPCR fold change of RCN1
2.5
2.0
1.5
1.0
0.5
0.0
hESC_NKX
PAX6_Ta_KO
*
hESC
hESC-CM

Supplementary information related to

The genome-wide dual-CRISPR screening identifies essential non-coding regulatory elements

Yufeng Li[1,7], Minkang Tan[1,7], Almira Akkari-Henic[1,7], Ningning Xu[1], Shengnan Sun[1], Yavuz Ariyurek[2], Susan L. Kloet[2], Richard Davis[3], Harald Mikkers[1], Joshua J. Gruber[4], Michael P. Snyder[5]*, Xiao Li[6]*, Baoxu Pang[1]*

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands

[2]Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, The Netherlands

[3]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, The Netherlands

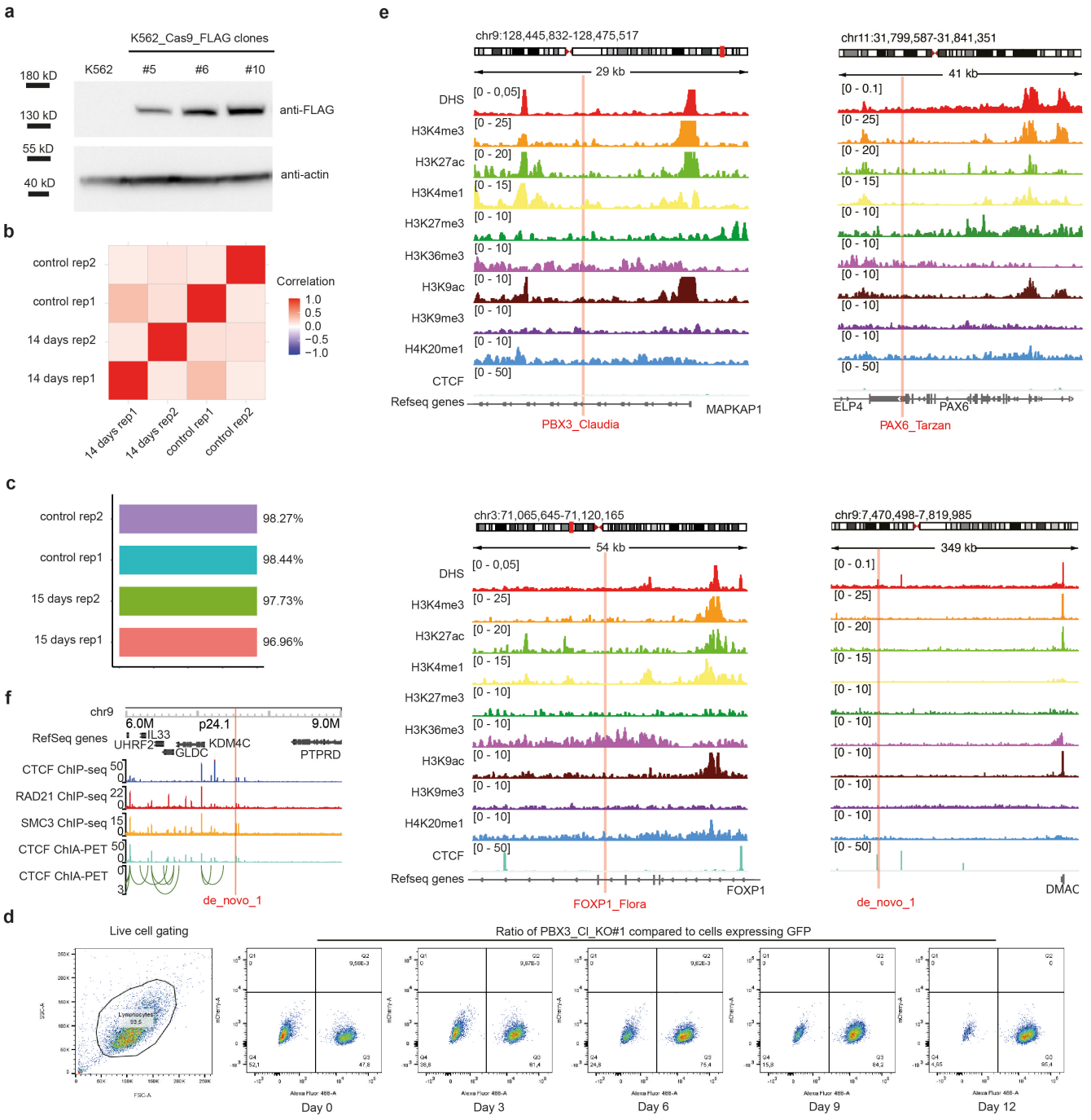
[4]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA

[5]Department of Genetics, Stanford University, Stanford, California, USA

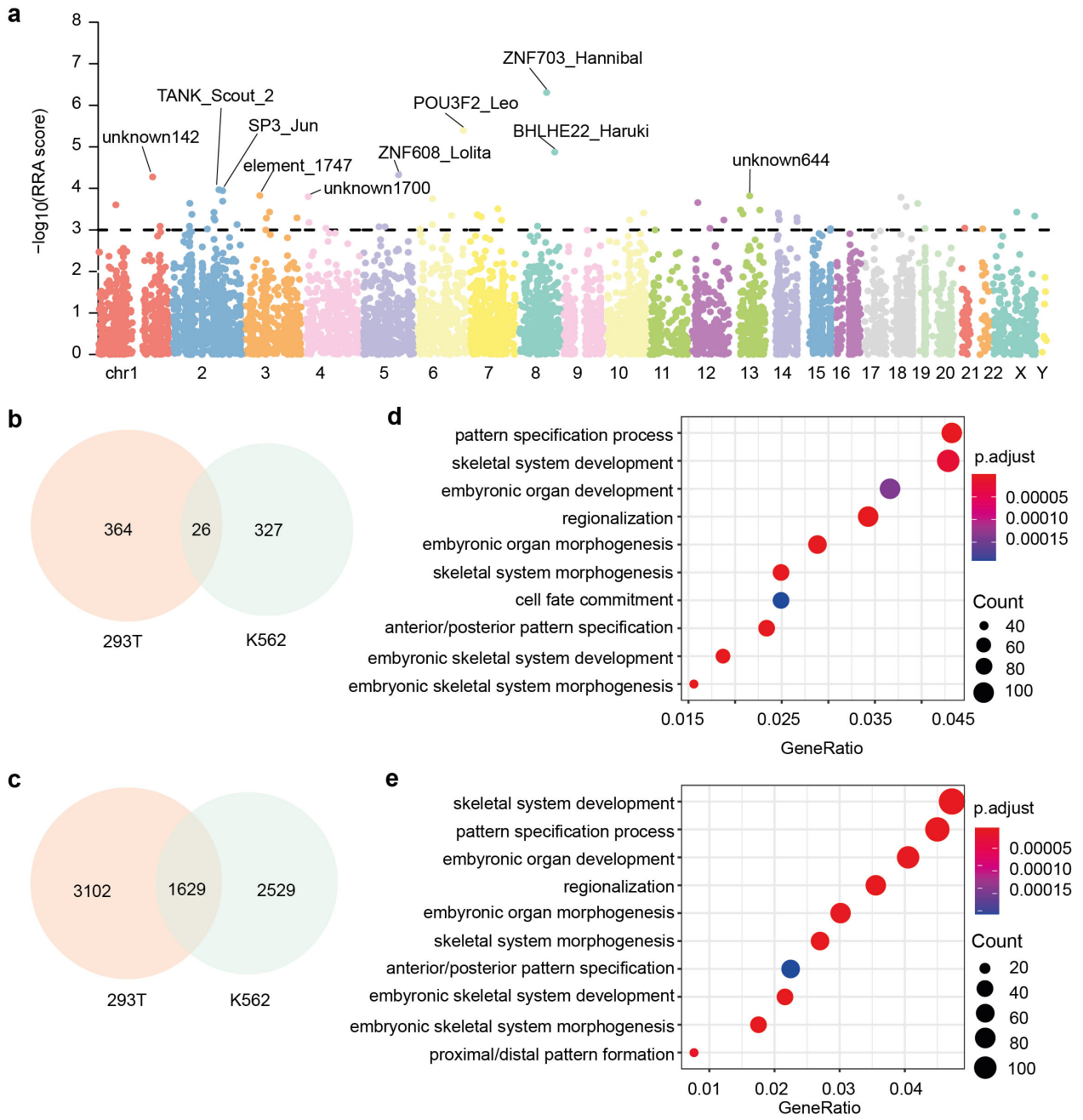
[6]Department of Biochemistry, The Center for RNA Science and Therapeutics, Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH, USA.

[7]These authors contributed equally

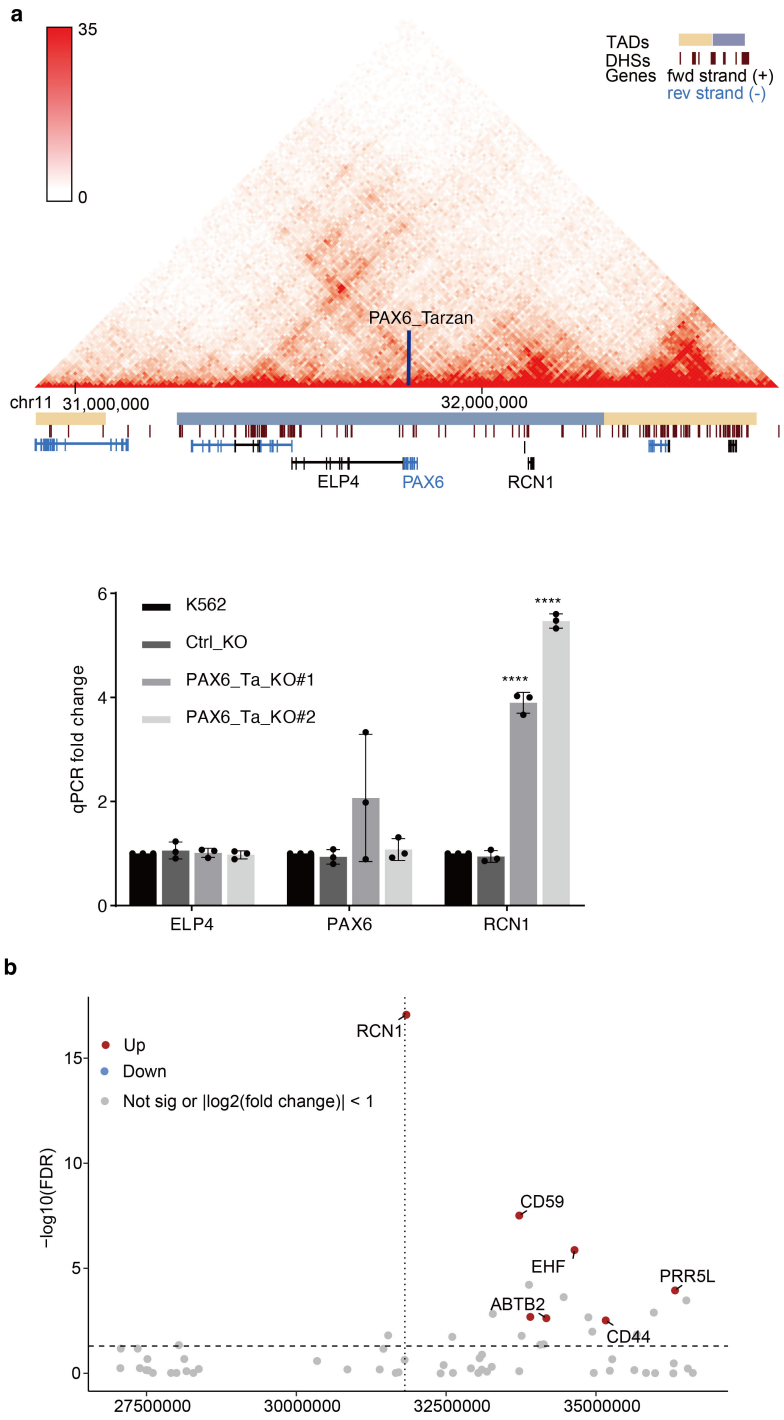
*Correspondence: b.pang@lumc.nl (B.P.), mpsnyder@stanford.edu (M.P.S), xiao.li9@case.edu (X.L.),



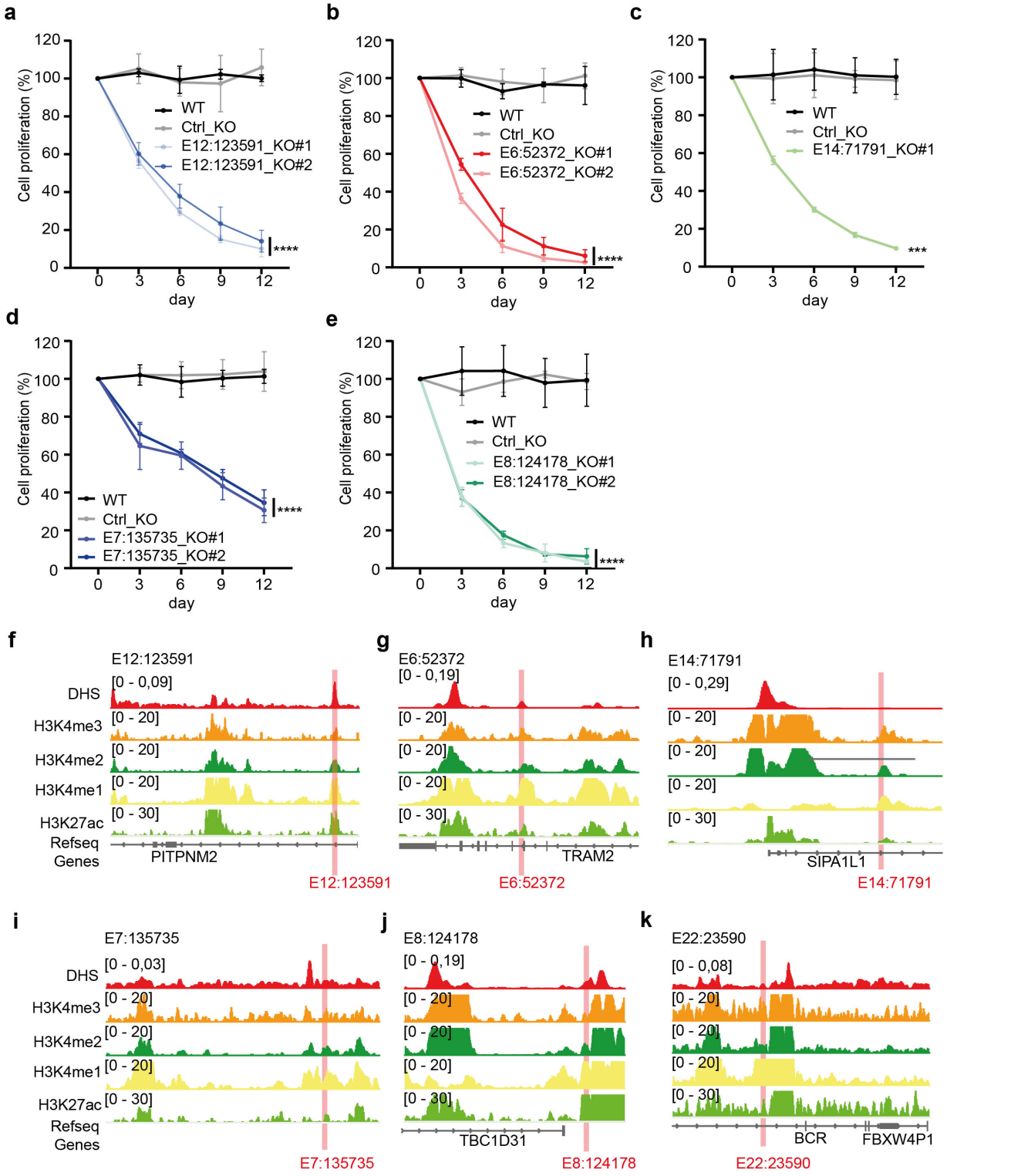
Supplementary Figure 2: UCE dual-CRISPR screening in K562 cells.
(a) Western blot to validate the Cas9-expressing K562 cells. The expression of Cas9-FLAG proteins in different clones was probed with the anti-FLAG antibody. Actin was used as the loading control. K562 _Cas9_FLAG clone #10 was used for subsequent screenings. **(b)** Pearson correlation of the dual-CRISPR library biological replicates. (control, day 0 of the screening samples; 15 days, cells grew for another 15 days). **(c)** Percentage of the recovered target UCEs and validated enhancers from different samples during the dual-CRISPR screenings in K562 cells. **(d)** FACS gating examples of cell proliferation assays. **(e)** Epigenetic signatures surrounding the UCEs PBX3_CI, FOXP1_FI, PAX6_Ta and the NCRE De_novo_1. **(f)** Chromatin looping around NCRE de_novo_1, as shown by CTCF ChIA-PET loops and in situ Hi-C matrix.



Supplementary Figure 3: Essential UCEs shared between K562 and 293T cells. **(a)** Essential UCEs in 293T are identified using the dual-CRISPR screening. MAGeCK algorithm was used to identify significant hits depleted from the cells cultured for an additional 14 days compared to the initial population. The Manhattan plot was used to show the distribution of all the target regions. Significant hits were above the dashed line, indicating the cutoff MAGeCK RRA score of 0.001. Different colors represent different chromosomes. The top essential hits were shown. **(b)** Venn diagram shows the shared essential UCEs between K562 (353 UCEs with RRA score < 0.01) and 293T (410 UCEs with RRA score < 0.01) cell lines, respectively. Only 26 (<10%) of the essential UCEs identified from the respective screenings were shared between K562 and 293T cells. **(c)** Venn diagram shows the genes around the essential UCEs between K562 (4158 genes that are +/- 1Mbp around the essential UCEs) and 293T (4731 genes that are +/- 1Mbp around the essential UCEs) cell lines. 1629 (>30%) of the genes were shared between K562 and 293T cells. **(d-e)** Over Representation Analysis of Gene Ontology terms in Biological Process sub-ontology based on the potential genes that might be regulated by the essential UCEs in 293T cells (d) and K562 cells (e), respectively.

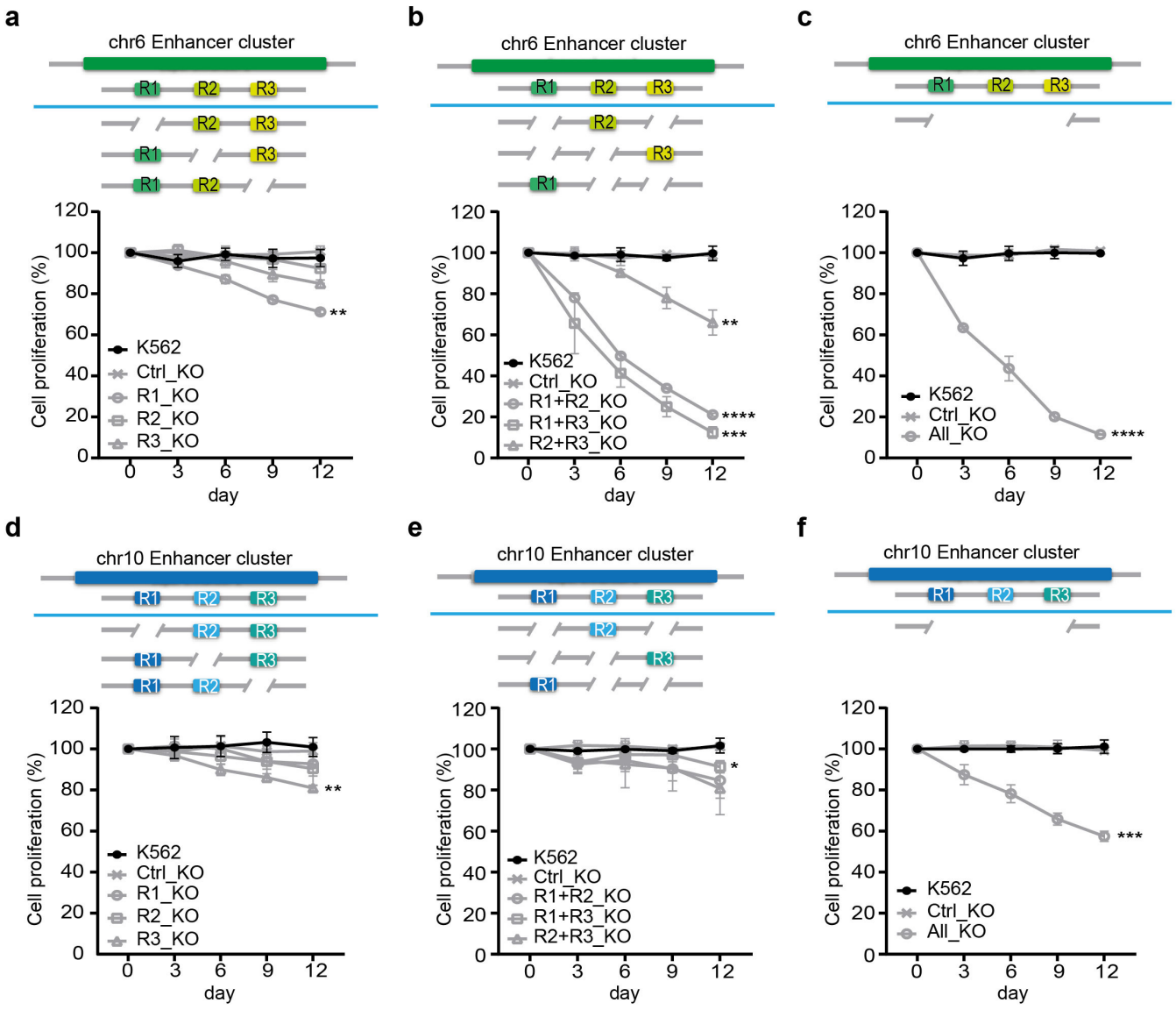


Supplementary Figure 4: NCREs regulate their surrounding genes to exert their functions
(a) TADs identified by Hi-C surrounding PAX6_Ta are shown, and the location of PAX6_Ta is indicated by the vertical blue bar. Horizontal yellow and blue bars indicate distinct TADs. Transcription of three nearby genes within the same TAD as PAX6_Ta was quantified by qPCR (n = 3 biological independent samples; bars show mean value ± s.d.; ****P < 0.0001, calculated using unpaired t-test). **(b)** RNA-seq was performed to identify differentially expressed genes around NCRE PAX6_Ta from PAX6_Ta-KO cells versus K562 WT cells. The x axis represents the coordinates of genes surrounding PAX6_Ta ranging from - 5Mb to + 5 Mb. The y axis shows the -log10(FDR) of nearby genes by DESeq2. The dashed horizontal line indicates the FDR cutoff of 0.05, and the dotted vertical line indicates the location of PAX6_Ta. Each dot represents one gene, the red dots represent significantly up-regulated genes, blue dots represent significantly down-regulated genes, and the gray ones are either non-significant genes or genes with |log2FC|<1.



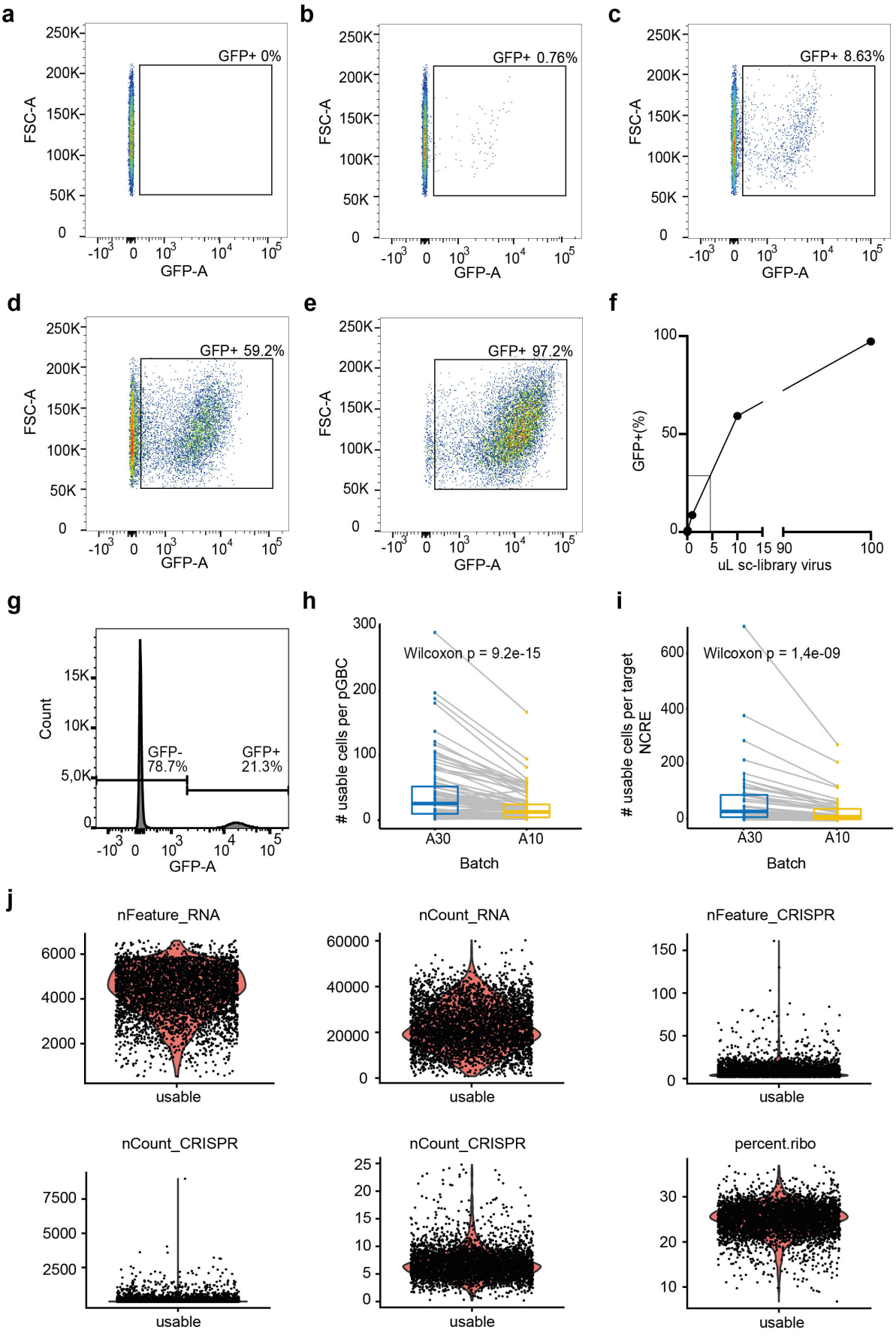
Supplementary Figure 5: Essential enhancers in K562 cells.

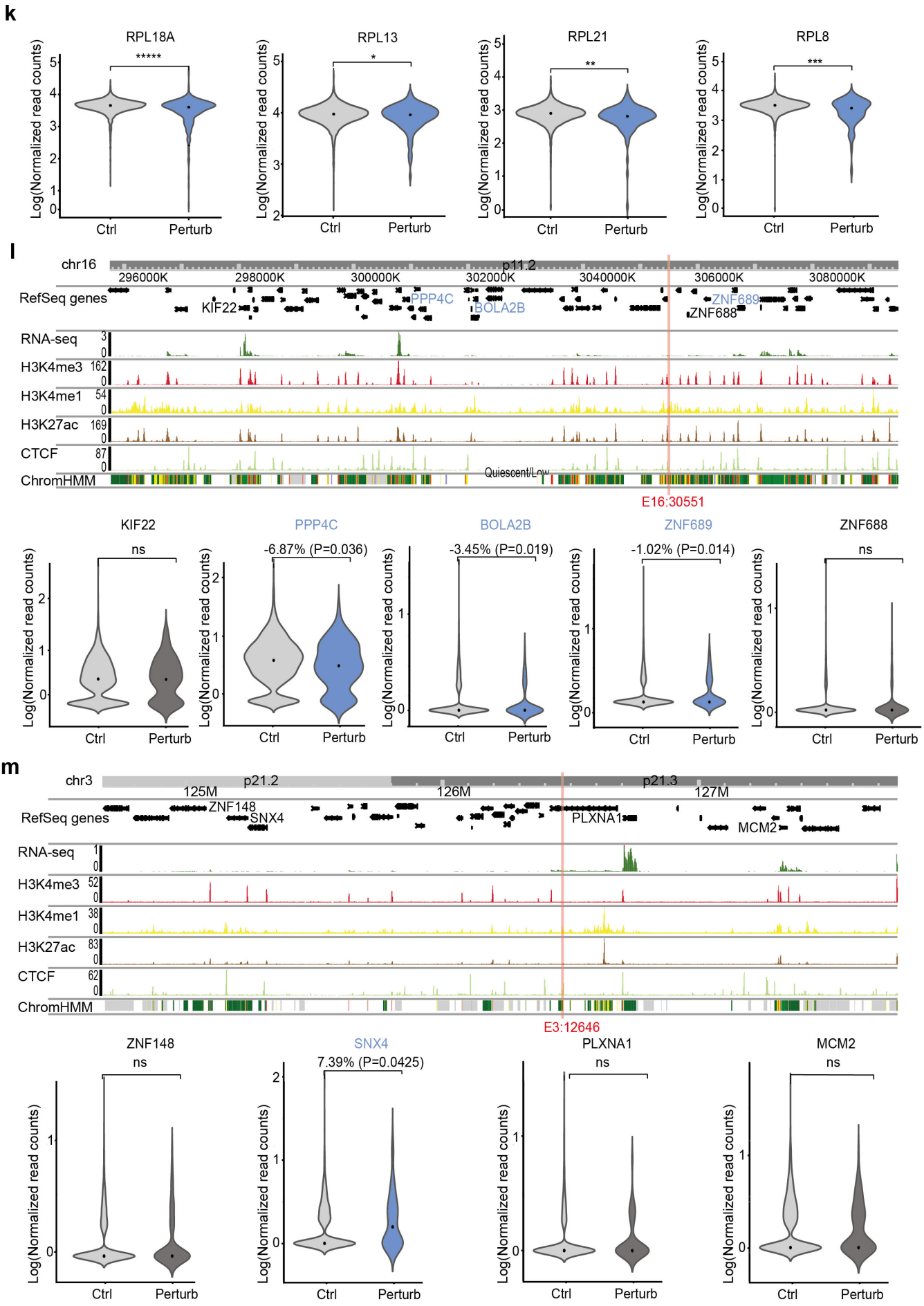
(a-e) The removal of essential enhancers E12:123591 (a), E6:52372 (b), E14:71791 (c), E7:135735 (d), E8:124178 (e) caused cell growth disadvantage in K562 cells. Cell proliferation assay was performed by mixing the KO cell lines with cells expressing GFP at a 1:1 ratio. The changes in GFP percentage were monitored at indicated time points by FACS. Cells with dual-CRISPR guide RNAs targeting GFP sequences served as negative controls (Ctrl-KO). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The ratio of cells in the initial mixture was set as 100% (n=3 biological independent samples; values are shown as the mean ± s.d.; E14:71791_KO#1 ***P < 0.001 and E12:123591, E6:52372, E7:135735, E8:124178 ****P < 0.0001, calculated using two-way ANOVA).**(f-k)** Epigenetic signatures surrounding the identified essential enhancers in K562 cells. E12:123591 (f), E6:52372 (g), E14:71791 (h), E7:135735 (i), E8:124178 (j) and E22:23590 (k).

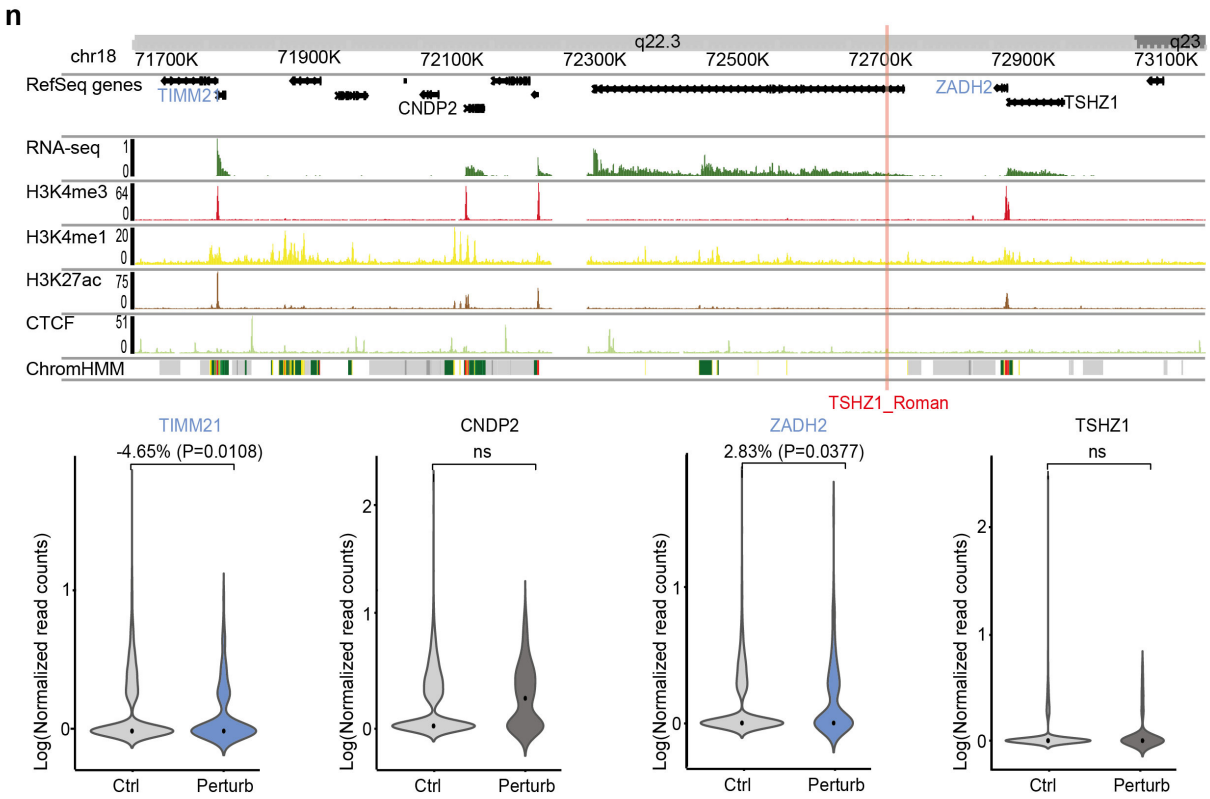


Supplementary Figure 6: Identifying essential enhancer clusters.

(a, b, c) Growth effect of one enhancer (a), two-enhancer (b), and three-enhancer (c) removal from the cluster in chromosome 6. Cell proliferation assay was performed by mixing the KO cell lines with cells expressing GFP at a 1:1 ratio. The changes in GFP percentage were monitored at indicated time points by FACS. Cells with dual-CRISPR guide RNAs targeting GFP sequences served as negative controls (Ctrl_KO). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The ratio of cells in the initial mixture was set as 100%. R1, R2, and R3 represent the three individual potential enhancers targeted by the dual-CRISPR libraries (n=3 biological independent samples; values are shown as the mean ± s.d.; R1_KO **P = 0,0017, R2+R3_KO **P = 0,0025, R1+R3_KO ***P = 0,0001, R1+R2_KO and All_KO ****P < 0.0001 calculated using two-way ANOVA). **(d, e, f)** Growth effect of one enhancer (d), two-enhancer (e), and three-enhancer (f) removal from the cluster in chromosome 10. Cell proliferation assay was performed by mixing the KO cell lines with cells expressing GFP at a 1:1 ratio. The changes in GFP percentage were monitored at indicated time points by FACS. Cells with dual-CRISPR guide RNAs targeting GFP sequences served as negative controls (Ctrl_KO). The y axis represents the relative ratio of the GFP negative cells to the positive cells. The ratio of cells in the initial mixture was set as 100%. R1, R2, and R3 represent the three individual potential enhancers targeted by the dual-CRISPR libraries (n=3 biological independent samples; values are shown as the mean ± s.d.; R1+R3_KO *P = 0,0346, R3_KO **P = 0,01, All_KO ***P = 0.0003 calculated using two-way ANOVA).

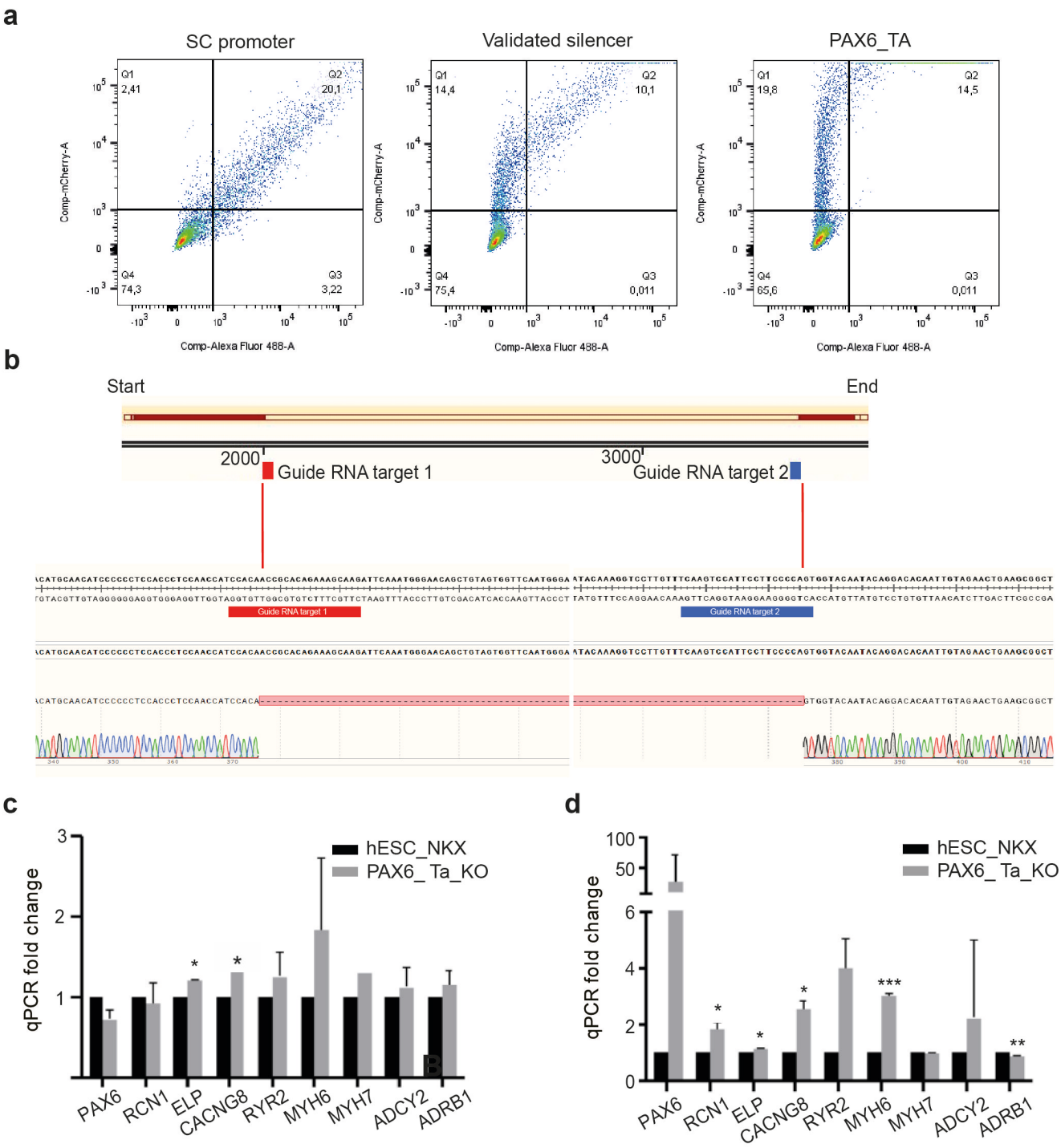

a
GFP+ 0%
b
GFP+ 0.76%
c
GFP+ 8.63%
d
GFP+ 59.2%
e
GFP+ 97.2%
f
GFP+(%)
uL sc-library virus
g
Count
GFP-
78.7%
GFP+
21.3%
h
Wilcoxon p = 9.2e-15
usable cells per pGBC
Batch
i
Wilcoxon p = 1,4e-09
usable cells per target NCRE
Batch
j
nFeature_RNA
nCount_RNA
nFeature_CRISPR
nCount_CRISPR
nCount_CRISPR
percent.ribo
usable


k
RPL18A
*****
RPL13
*
RPL21
**
RPL8
***
Log(Normalized read counts)
Ctrl
Perturb
l
chr16
p11.2
296000K
298000K
300000K
302000K
3040000K
306000K
3080000K
RefSeq genes
KIF22
PPP4C
BOLA2B
ZNF689
ZNF688
RNA-seq
H3K4me3
H3K4me1
H3K27ac
CTCF
ChromHMM
Quiescent/Low
E16:30551
KIF22
ns
PPP4C
-6.87% (P=0.036)
BOLA2B
-3.45% (P=0.019)
ZNF689
-1.02% (P=0.014)
ZNF688
ns
m
chr3
p21.2
p21.3
125M
126M
127M
RefSeq genes
ZNF148
SNX4
PLXNA1
MCM2
RNA-seq
H3K4me3
H3K4me1
H3K27ac
CTCF
ChromHMM
E3:12646
ZNF148
ns
SNX4
7.39% (P=0.0425)
PLXNA1
ns
MCM2
ns



Supplementary Figure 7: Single-cell dual-CRISPR screening.

(a-f) Determining the lentiviral titer of the single-cell dual-CRISPR library (sc-library) in K562 cells, with a range of 0μL (a), 0.1μL (b), 1μL (c), 10μL (d) and 100μL (e) was tested. GFP expression from the sc-library was used as the marker to determine the virus infection rate. Individual plots show 0% GFP-expressing cells (a), 0,76% GFP-expressing cells (b), 8,63% GFP-expressing cells (c), 59,2% GFP-expressing cells (d), and 97,2% GFP-expressing cells (e), which was plotted in (f). A predicted MOI <0.3 using 5uL of sc-library lentivirus was then determined. (**g)** Cells containing the sc-library for the downstream analysis. FACS showed that 20% of total cells had the sc-library, as monitored by GFP expression, representing an MOI of 0.2 in the final experiment. This ensured that every single cell contained only one pair of guide RNAs from the sc-library. (**h)** Comparison of usable cells per paired guide RNA barcodes (pGBCs) between two batches of scRNA-seq data: A10 and A30 show 10,000 and 30,000 cells were loaded per 10X Genomics chip, respectively. 1,199 and 3,271 usable cells were retrieved from A10 and A30 batches, respectively. Wilcoxon signed-rank test was used to compare two datasets. (**i)** Comparison of usable cells per target NCRE (multiple pairs of guide RNA were included in the sc-library) between two batches of scRNA-seq data. Wilcoxon signed-rank test was used to compare two batch data. (**j)** Basic quality control of 4,470 single cells containing the sc-library after cell filtering and multiplets filtering. Violin plots show the total number of genes, total UMIs of transcripts, the total number of unique sgRNA, total UMIs of sgRNA detected, and the percentage of transcripts encoding mitochondrial and ribosomal genes detected in every single cell. (**k)** The dual-CRISPRs targeting the promoters of ribosomal genes: RPL18A, RPL13, RPL21, and RPL8 served as the positive controls. The gene expression matrix was normalized using the LogNormalize method with a scale factor of 10,000. Identities of cell perturbation were assigned based on whether pGBCs were captured. Differential gene expression testing between perturbed cells (Perturb) and negative control cells (NCT) were done by the MAST algorithm with the number of gene UMIs observed per cell and the number of gRNA UMIs per cell as covariates (*****P<0.00001, ***P<0.001, **P<0.01, *P<0.05). (**l-n)** The upper panel shows epigenetic signatures surrounding NCREs E16:30551 (l), E3:126465 (m), and TSHZ1_Roman (n), indicated by the red line. The lower panel shows the differential gene expression testing results using MAST. Violin plots show the normalized expression levels of candidate genes in perturbed and control cells (P < 0.05, calculated by MAST fitted model; ns: not significant).



Supplementary Figure 8: Removing UCE PAX6 affecting cardiac development.
(**a**) FACS gating examples of the silencer reporter assay. (**b**) Sanger sequencing of the PAX6_Ta region showed that the two guide RNAs made the deletion in NKX2-5eGFP/w hESC. (**c-d**) The transcription of heart-related genes PAX6, RCN1, ELP, CACNG8, RYR2, MYH6, MYH7, ADRY2, and ADRB1 was quantified by qPCR before (b) and after (c) differentiation in both NKX2-5eGFP/w hESC cells (hESC_NKX) and NKX2-5eGFP/w hESC PAX6_Ta KO bulk cells (PAX6_Ta_KO) (n=2 biological independent samples; values are shown as the mean ± s.d. for each bar; before_PAX6 ^{ns}P = 0.083652, before_RCN1 ^{ns}P = 0.719522, before_ELP *P = 0.001540, before_CACNGB *P = 0.020124, before_RYR2 ^{ns}P = 0.369053, before_MYH6 ^{ns}P = 0.319187, before_ADCY2 ^{ns}P = 0.545715, before_ADRB1 ^{ns}P = 0.390106, after_PAX6 ^{ns}P = 0.480798, after_RCN1 *P = 0.024307, after_ELP *P = 0.015280, after_CACNGB *P = 0.018328, after_RYR2 ^{ns}P = 0.057717, after_MYH6 ***P = 0.000667, after_MYH7 ^{ns}P = 0.209353 after_ADCY2 ^{ns}P = 0.593984, after_ADRB1 **P = 0.007835 calculated using two-tailed unpaired t-test).

Supplementary Tables

Supplementary Table 1: Essential ultra-conserved elements identified from K562 cells using an RRA score cutoff of 0.01. MAGeCK RRA algorithm was used to identify significantly depleted pgRNA targeting ultra-conserved elements. The RRA score, P value, FDR, and genomic location are listed.

Supplementary Table 2: Essential ultra-conserved elements identified from 293T cells using an RRA score cutoff of 0.01. MAGeCK RRA algorithm was used to identify significantly depleted pgRNA targeting ultra-conserved elements. The RRA score, P-value, FDR, and genomic location are listed.

Supplementary Table 3: Ultraconserved elements exerting imatinib resistance identified from K562 cells using MAGeCK MLE beta score as cutoff. The absolute value of control beta score smaller than one standard deviation and Imatinib treatment beta score bigger than one standard deviation. Control beta score, imatinib treatment beta score, and genomic location are listed.

Supplementary Table 4: Essential putative enhancers identified from K562 cells using MAGeCK RRA score cutoff of 0.01. Two batches of pooled dual CRISPR screening were performed with enhancer library 1, which was designed to delete target enhancers from flanking regions; and enhancer library 2, which was designed to delete the core motifs of enhancers from inside. MAGeCK RRA algorithm was used to identify significantly depleted pgRNA targeting putative enhancers in K562 cells, respectively, and merged results from two batches. Batch information, RRA score, P-value, FDR, and genomic location are listed.

Supplementary Table 5: Top 20 enhancer clusters identified by GSEA and MAGeCK RRA in the K562 cell line. The SEdb_ID column is the names of super-enhancers in Sedb; the Size column indicates the number of regions from our designed libraries, and the Elements column indicates the detailed coordinates of the regions. The P-values, FDR, enrichment score and RRA score were obtained by GSEA and MAGeCK algorithms. FDR was obtained by BH procedure.

Supplementary Table 6: Statistics of usable single cells per pGBC captured by sc-dualCRISPR screening assay. Cell ranger was used to process scRNA-seq data. Seurat was used to filter multiplets and assign cell perturbation identity. The numbers of usable cells per pGBC from two batches were listed.

Supplementary Table 7: Significant NCRE-gene pairs identified from sc-dual CRISPR screening using a P-value cutoff of 0.05. The LogNormalize method was used to normalize scRNA-seq data using the scale factor of 10,000. MAST was used for differential gene expression testing with the logarithm of total numbers UMIs representation gene and gRNA as covariates of the MAST model to regress out. P-value was calculated by MAST testing. NCRE, target gene, mean expression levels in Perturb and NCT cell groups, logarithm fold change, and genomic location are listed.

Supplementary Table 8: Related information on tested regions. All tested regions and associated primers are listed.

Supplementary Movies

Removing PAX6_Ta causes defects in cardiomyocyte differentiation.

Movie 1: NKX2-5eGFP/w hESC WT GFP channel.

Movie 2: NKX2-5eGFP/w hESC WT brightfield.

Movie 3: NKX2-5eGFP/w hESC PAX6_Ta KO GFP channel.

Movie 4: NKX2-5eGFP/w hESC PAX6_Ta KO brightfield. (Scale bars: 275 μm).

Chapter 4

Summary

Summary

Understanding the meaning of the human genome codes is one of the keys to unlock the secrets of life. Despite having a fairly good grasp of the sequences of the human genome, we are still far from understanding the functions of most parts of the genome and their involvement in diseases. The application of the CRISPR-Cas9 genome-editing systems revolutionized the way to study the function of the genome, not only the coding genes but also the non-coding genome. In this thesis, multiple CRISPR screening systems were designed and used to study the transport of chemotherapeutic drugs and the functions of non-coding regulatory elements in distinct biological pathways.

In **Chapter 2**, in order to focus on drug transport and at the same time reduce the background interference from a genome-wide screening, custom CRISPR-knockout and CRISPR-activation libraries were designed and assembled to target all potential membrane-associated transporters. Taking advantage of the autofluorescence of doxorubicin, FACS sorting was used to enrich CRISPR-edited cells with either high or low drug accumulation during the screenings, in which drug uptake phenotype was directly measured. Such a strategy should avoid confounding factors or screening hits from other screening studies with similar aims that are based on cell survival as an indirect readout for drug transport. We showed that such screenings with focused libraries are very robust, as the same top hits were usually present in distinct populations indicating a similar potential drug transport function within the same screening, which in a way re-confirming their potential roles. Using this method, we identified previously known drug exporters such as ABCB1 and ABCG2. In addition, we identified a new doxorubicin importer gene SLC2A3 (GLUT3). We also realized that the CRISPR-activation screening complements the CRISPR-knockout system, which most of the research uses. For future screening efforts, it is better to combine the two screening systems to get a comprehensive identification of the potential factors involved in defined biological readouts.

In **Chapter 3**, we used an innovative dual-CRISPR system to study the functions of non-coding regulatory elements (NCREs) in their endogenous genomic environment. Using this method, we were able to study 4,047 UCEs in the human genome from UCNEbase, 1,527 in vivo-validated conserved enhancers from VISTA Enhancer Browser, and 13,539 potential K562-cell enhancers predicted by the ENCODE project, by deleting these NCREs one by one. We were able to identify many NCREs many affect cell fitness and drug responses. In addition, many of the NCREs actually have silencer (transcriptionally repressive) activity, in contrast to most commonly known enhancer elements. This is interesting and important because the roles of silencers in gene regulation in the whole genome were not well studied before. Our data also support the idea that silencers may exist broadly in the human genome and play important roles, similar to other well-studied NCREs such as enhancers and insulators. We were able to show that many of these NCREs regulate nearby genes, therefore affecting different biological pathways. One of the ultraconserved NCRE PAX6_Tarzan would affect the cardiomyocyte differentiation in an hESC model, suggesting many ultraconserved NCRE may exert their functions in a tissue-dependent context. We provide the broad research community with a new tool to study the functions of NCREs in different biological aspects, and it is expected many more important NCREs will be discovered in the future, which would not only be relevant to fundamental biology but also relevant to human diseases.

Nederlandse samenvatting

Het begrijpen van de betekenis van de codes van het menselijk genoom is een van de sleutels om de geheimen van het leven te ontrafelen. Ondanks dat we de sequentie van het menselijk genoom weten, begrijpen we nog lang niet de functies van de meeste delen van het genoom en hun betrokkenheid bij ziekten. De toepassing van de CRISPR-Cas9 techniek bracht een revolutie teweeg in de manier waarop de functie van het genoom werd bestudeerd, niet alleen voor de coderende genen maar ook het niet-coderende gedeelte van het genoom. In dit proefschrift zijn meerdere CRISPR-screeningsystemen ontworpen en gebruikt om het transport van chemotherapeutica en de functies van niet-coderende regulerende elementen in verschillende biologische routes te bestuderen.

Om zich te concentreren op medicijntransport en tegelijkertijd de achtergrond interferentie van een genoom brede screening te verminderen, werden in Hoofdstuk 2 aangepaste CRISPR-knock-out- en CRISPR-activeringsbibliotheken ontworpen en samengesteld om alle potentiële membraan-geassocieerde transporters te beïnvloeden. Door gebruik te maken van de autofluorescerende eigenschappen van doxorubicine, werd FACS-sortering gebruikt om CRISPR-bewerkte cellen te verrijken met hoge of lage geneesmiddelaccumulatie tijdens de screenings, waarbij het fenotype van de geneesmiddelopname direct werd gemeten. Een dergelijke strategie moet verwarrende factoren of screeninghits van andere screeningstudies met vergelijkbare doelstellingen die zijn gebaseerd op celoverleving als indirecte uitlezing voor geneesmiddeltransport, vermijden. We hebben aangetoond dat dergelijke screenings met gefocuste bibliotheken zeer robuust zijn, aangezien dezelfde tophits meestal aanwezig waren in verschillende populaties, wat wijst op een vergelijkbare potentiële medicijntransportfunctie binnen dezelfde screening, wat in zekere zin hun potentiële rol opnieuw bevestigt. Met behulp van deze methode hebben we eerder bekende drugsexporteurs geïdentificeerd, zoals ABCB1 en ABCG2. Daarnaast hebben we een nieuw doxorubicine-importergen SLC2A3 (GLUT3) geïdentificeerd. We realiseerden ons ook dat de CRISPR-activeringsscreening

een aanvulling vormt op het CRISPR-knock-outsysteem, dat in de meeste onderzoeken wordt gebruikt. Voor toekomstige screeningsinspanningen is het beter om de twee screeningsystemen te combineren om een uitgebreide identificatie te krijgen van de mogelijke factoren die betrokken zijn bij gedefinieerde biologische uitlezingen.

In hoofdstuk 3 hebben we een innovatief dual-CRISPR-systeem gebruikt om de functies van niet-coderende regulerende elementen (NCRE's) in hun endogene genomische omgeving te bestuderen. Met behulp van deze methode konden we 4.047 UCE's in het menselijk genoom van UCNEbase, 1.527 in vivo gevalideerde geconserveerde versterkers van VISTA Enhancer Browser en 13.539 potentiële K562-celversterkers, voorspeld door het ENCODE-project, bestuderen door deze NCRE's één voor één te verwijderen. We waren in staat om veel NCRE's te identificeren die van invloed zijn op celfitness en medicijnreacties. Bovendien hebben veel van de NCRE's eigenlijk silencer (transcriptioneel repressieve) activiteit, in tegenstelling tot de meest algemeen bekende enhancer-elementen. Dit is interessant en belangrijk omdat de rol van silencers in genregulatie in het hele genoom niet eerder goed bestudeerd was. Onze gegevens ondersteunen ook het idee dat silencers breed in het menselijk genoom kunnen voorkomen en een belangrijke rol spelen, vergelijkbaar met andere goed bestudeerde NCRE's zoals versterkers en lange afstand regulatoren. We konden aantonen dat veel van deze NCRE's nabijgelegen genen reguleren en daardoor verschillende biologische routes beïnvloeden. Een van de ultra geconserveerde NCRE PAX6_Tarzan zou de differentiatie van cardiomyocyten in een hESC-model beïnvloeden, wat suggereert dat veel ultraconserved NCRE hun functies in een weefselafhankelijke context kunnen uitoefenen. We bieden de brede onderzoeksgemeenschap een nieuwe tool om de functies van NCRE's in verschillende biologische aspecten te bestuderen, en er wordt verwacht dat er in de toekomst nog veel meer belangrijke NCRE's zullen worden ontdekt, die niet alleen relevant zijn voor de fundamentele biologie, maar ook voor de humane ziekten.