Paired CRISPR screening libraries for studying the function of the non-coding genome at scale

Pang, B.X.

Paired CRISPR screening libraries for studying the function of the non-coding genome at scale

The majority of the human non-coding genome remains poorly studied. A user-friendly genome-wide screening system composed of thousands of paired single-guide RNAs for the deletion of non-coding regions revealed key functions of many non-coding elements in cell growth and cell differentiation and in cellular response to drugs.

This is a summary of:
Li, Y. et al. Genome-wide Cas9-mediated screening of essential non-coding regulatory elements via libraries of paired single-guide RNAs. *Nat. Biomed. Eng.* https://doi.org/10.1038/s41551-024-01204-8 (2024).

Published online: 31 May 2024

The question

More than 98% of the 3 billion nucleotide pairs in the human genome are non-coding and do not contain the information for making proteins[1]. Research during the past two decades has changed the understanding of the non-coding genome from 'junk DNA' to regions containing regulatory elements that direct gene transcription with spatial and temporal precision. These non-coding regulatory elements (NCREs) may have different functions and can be broadly categorized into distinct functional units, such as non-coding RNAs, promoters, enhancers, silencers and insulators, among others[2]. Although the research community has made great efforts to annotate NCREs, it is still challenging to study their biological functions in the endogenous environment, considering the large numbers and possible distinct functions of NCREs. Clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9)-mediated genome editing revolutionizes how the genome can be studied, opening a new venue to study the NCREs.

The solution

We developed a screening technique called the dual-CRISPR system to study the NCREs in a genome-wide fashion. Based on the CRISPR–Cas9 genome-editing technology, the dual-CRISPR system produces paired guide RNAs that target both ends of an individual NCRE for its removal from the genome. We can assemble thousands of paired guide RNAs into the dual-CRISPR screening library for the genome-wide deletion of thousands of NCREs to study their functions in diverse biological contexts. After genome-wide screening, the paired guide RNA sequences within the cells can be retrieved directly using one-step polymerase chain reaction (PCR), and are then ready for next-generation sequencing, without the need for additional barcoding to infer the content of the paired guide RNAs. Furthermore, the dual-CRISPR system is compatible with single-cell sequencing to infer the NCRE–gene interactions at the single-cell level.

We used the dual-CRISPR screening system to study 4,047 ultraconserved elements (UCEs) in the human genome from UCNEbase[3], 1,527 in vivo-validated conserved enhancers from VISTA Enhancer Browser, and all 13,539 predicted enhancers in K562 cells from ENCODE[4] (Fig. 1a). We identified essential UCEs and NCREs that may affect cell growth in human lymphoblast (K562) and embryonic kidney (293T) cell lines, and chemotherapeutic response in K562 cells. Interestingly, some validated UCEs and NCREs were found to be silencers that are still understudied[5]. Using another dual-CRISPR screening library, we also identified essential enhancers that may affect cell growth in K562 cells (Fig. 1b), and found that some enhancers have both enhancer and silencer activities[2]. Furthermore, we identified clusters of NCREs with functional redundancy. UCEs are non-coding sequences that are almost identical among different species and are assumed to have important biological roles because of the strong conservation during evolution[3]. We identified that UCE PAX6_Tarzan affected cardiomyocyte differentiation, once it was removed from human embryonic stem cells, indicating an important biological role of this UCE.

Future directions

The dual-CRISPR screening system could be used to investigate the roles of NCREs in a high-throughput capacity across distinct biology contexts. As the target regions are removed from the genome, NCREs can be studied irrespective of specific roles as silencers, enhancers or other regulatory elements with unknown functions. NCREs usually function in a tissue-specific way, and thus, unlike the CRISPR screening libraries targeting genes, no one-fit-for-all libraries are available for different cells and tissues. We designed the dual-CRISPR screening system in a user-friendly way so that it is possible for individual research groups to design and assemble custom screening libraries for their own model systems in a straightforward manner. Such broad applications could help the understanding of the functional roles of NCREs in health and disease conditions and may allow for identifying NCREs as future drug targets.

The dual-CRISPR system targets genomic regions for deletion, and thus it is not possible to pinpoint the functional units within the target regions. Like other CRISPR systems, the dual-CRISPR system also relies on highly efficient paired guide RNAs with low off-target effects that simultaneously delete the target regions.

With this versatile tool, we and other research groups are now better equipped to study the many questions that remain unanswered, such as how NCREs regulate gene pathways, function during tissue differentiation and play a role in the development of diseases.

Baoxu Pang
Leiden University Medical Center, Leiden, the Netherlands.

FROM THE EDITOR

"This CRISPR screen for the deletion — systematically, genome-wide and at high throughput — of thousands of non-coding regulatory elements will facilitate the study of the functions of these elements." **Editorial Team, *Nature Biomedical Engineering*.**

FIGURE

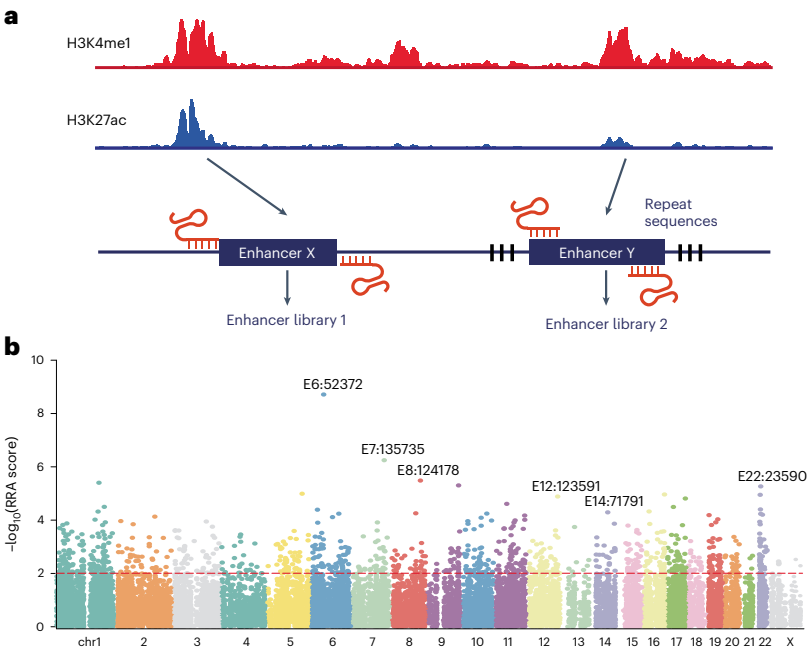

Fig. 1 | Identification of essential enhancers in K562 cells. a, Design of dual-CRISPR libraries targeting all potential enhancers in K562 cells, identified based on the combination of H3K4me1, H3K27ac and other markers. **b**, Essential enhancers in K562 cells were identified using dual-CRISPR screens. The model-based analysis of genome-wide CRISPR–Cas9 knockout (MAGeCK) algorithm was used to identify significant hits depleted from cells cultured for an additional 15 days compared with the initial population. The Manhattan plot shows the distribution of all the target regions. Significant hits were above the red dashed line, indicating the MAGeCK robust rank aggregation (RRA) cut-off score of 0.01. Different colours represent different chromosomes (chr). Panel **a** created with BioRender.com. Panel **b**: © 2024, Li, Y. et al., CC BY 4.0.

BEHIND THE PAPER

This project was envisioned at Stanford University in 2016 and became one of the first independent projects led by my research group at Leiden University Medical Center. Inspired by results from my other project using paired CRISPR–Cas9 to delete a single NCRE, I was excited about the possibilities a genome-wide screening system could bring, and scribbled the rudimentary dual-CRISPR screening system in my lab journal. With the support of the senior authors, the dual-CRISPR screening libraries were made possible, and my group members continued to carry out the biological experiments. Despite serious delays due to the COVID-19 pandemic, an unsuccessful grant application based on this research and the lead first author being critically ill, thanks to the team members' efforts, we could present this system to the research community. We hope our work will facilitate the discovery of more therapies targeting NCREs to treat diseases. **B.P.**

REFERENCES

1. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
   This paper provides the latest human genome sequence.
2. Pang, B., van Weerd, J. H., Hamoen, F. L. & Snyder, M. P. Identification of non-coding silencer elements and their regulation of gene expression. *Nat. Rev. Mol. Cell Biol.* **24**, 383–395 (2023).
   This review article highlights the various NCREs, including silencers.
3. Dimitrieva, S. & Bucher, P. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucl. Acids Res.* **41**, D101–D109 (2012).
   This paper catalogues the ultraconserved non-coding elements.
4. 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 (2012).
   This paper presents the ENCODE-predicted enhancers in K562 cells.
5. Pang, B. & Snyder, M. P. Systematic identification of silencers in human cells. *Nat. Genet.* **52**, 254–263 (2020).
   This paper reports the identification of NCREs—silencers.