



Reply to: Fitness effects of CRISPR/Cas9-targeting of long noncoding RNA genes

Ying Liu¹, Zhiheng Liu^{1,2}, Zhongzheng Cao^{1,2} and Wensheng Wei¹✉

REPLYING TO Horlbeck, M.A. et al. *Nature Biotechnology* <https://doi.org/10.1038/s41587-020-0428-0> (2020)

In this issue, Weissman and colleagues¹ argue that the results in our recent paper² on CRISPR/Cas9-based screening for functional long noncoding RNAs (lncRNAs) could have been affected by (1) copy-number effects on single-guide RNA (sgRNA) dropout and (2) lncRNA genes that overlap with protein-coding genes. We have carefully reanalyzed our screening results to evaluate the validity of their concerns.

Using the ENCODE Consortium copy-number data, we reanalyzed our previous screening results in K562 and HeLa cells², two cell lines with known copy-number variations (CNVs). In contrast, the third cell line used in our study², GM12878, has a relatively normal karyotype. In K562 cells, among the 230 essential lncRNAs identified from the splicing-targeting screening, 28 show copy-number amplifications (Tables 1 and 2), and 19 of these 28 are located within amplified regions in chromosome 22. Many of these lncRNAs are indeed top-ranked from our screening. It is therefore possible that CNVs influenced the screening scores, thus generating false-positive hits.

To further assess the effect of CNVs, we analyzed one lncRNA, *MIR17HG*, with copy-number amplification ($\log_2 R = 0.8155$). Its essentiality has been confirmed by a CRISPR interference (CRISPRi) strategy using sgRNAs from a previously published report³ (Supplementary Fig. 14d in our original paper²), indicating that its growth phenotype is not due to copy-number effect. Nevertheless, cleavage of amplified regions could lead to higher rates of DNA damage, thereby reducing cell viability, which in turn could make lncRNAs located inside these regions appear essential for cell viability. This requires validation using orthogonal methods, such as CRISPRi or antisense oligonucleotide targeting. Unfortunately, current copy-number data for nearly half of the detected loci have a resolution of >1,000 kilobases, making it difficult to accurately account for shorter amplifications.

In HeLa cells, among 115 essential lncRNAs, 7 hits could potentially be affected by copy-number amplification. For the two strongest hits in the HeLa screen that fall within the human papillomavirus 18 (HPV-18) integration locus (chr8 128,228,000–128,244,000, hg19)⁴, cancer susceptibility 19 (*CASC19*) and colon cancer-associated transcript 1 (*CCAT1*) are located at chr8 128,197,880–128,215,467 and chr8 128,220,111–128,231,333, respectively, according to the GENCODE dataset V20 that we referred to². *CASC19* is actually located in a copy-number-normal region. For *CCAT1*, there are seven sgRNAs targeting its splice sites in our library. Target sites of five sgRNAs are within the HPV-18 locus, but one other sgRNA targeting a normal region (cut site chr8 128,221,962) also showed significant dropout effect (Supplementary Table 4 in ref. ²). *CCAT1*

was also shown to affect cell growth in HeLa cells through CRISPRi screening³. Furthermore, *CCAT1* has been previously reported to promote HeLa cell proliferation, an effect that was validated by both transcript overexpression and siRNA knockdown⁵.

Table 1 | Essential lncRNA hits inside copy-number-amplified regions in the K562 cell line

lncRNA	$\log_2 R$	Chromosome location
<i>RP11-544A12.4</i>	0.9197	chr.9
<i>GPC5-IT1</i>	0.8393	chr.13
<i>LINC00379</i>	0.8155	chr.13
<i>LINC00380</i>	0.8155	chr.13
<i>LINC00410</i>	0.8155	chr.13
<i>LINC01049</i>	0.8155	chr.13
<i>MIR17HG</i>	0.8155	chr.13
<i>LL22NC03-102D1.18</i>	0.8012	chr.22
<i>RP11-83J21.3</i>	0.7043	chr.9
<i>ANKRD62P1-PARP4P3</i>	0.6691	chr.22
<i>KB-7G2.8</i>	0.6691	chr.22
<i>TPTEP1</i>	0.6691	chr.22
<i>LL22NC03-23C6.13</i>	0.6683	chr.22
<i>LL22NC03-86G7.1</i>	0.6683	chr.22
<i>TMEM191C</i>	0.6683	chr.22
<i>RP11-57J16.1</i>	0.6363	chr.18
<i>AC002472.11</i>	0.5761	chr.22
<i>AC007050.17</i>	0.5761	chr.22
<i>AC007308.7</i>	0.5761	chr.22
<i>THAP7-AS1</i>	0.5761	chr.22
<i>TMEM191A</i>	0.5761	chr.22
<i>XXbac-B135H6.15</i>	0.5761	chr.22
<i>AC004471.10</i>	0.5474	chr.22
<i>AC004471.9</i>	0.5474	chr.22
<i>AC006547.13</i>	0.5474	chr.22
<i>DGCR5</i>	0.5474	chr.22
<i>XXbac-B444P24.8</i>	0.5474	chr.22
<i>BMS1P20</i>	0.3971	chr.22

¹Biodynamic Optical Imaging Center, Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. ²Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ✉e-mail: wswwei@pku.edu.cn

Table 2 | Essential lncRNA hits in each cell line within copy-number-amplified regions and overlapping with protein-coding genes^a

	K562			HeLa			GM12878			All lines	
	Y	N	N	Y	N	N	Y	N	N	Y	N
Amplified ($\log_2 R > 0.3$)											
Overlapping with coding exon		Y	N	Y		N	Y		N	Y	N
Screen score ≥ 2	28	70	152	7	26	82	0	85	135	5	12
Overlapping with essential gene		1		2				3		0	
Overlapping with nonessential gene			69			24			82		
Including non-exon-targeting sgRNAs showing dropout effects			39			13			41		

^aNon-exon-targeting sgRNAs represent those cutting at the intronic regions (at least 10 bp from the splice sites) of coding genes or intergenic regions. The sgRNA dropout was defined using a threshold of sgRNA fold change < 0.7 . N, no; Y, yes.

The annotation of lncRNAs has been updated continuously, and recent changes include a redefinition of the *CASC19* transcript and the discovery of gene fusions between the viral oncogenes *E6/E7* and *CCAT1/CASC19*. In this case, both splice-site targeting and CRISPRi could disrupt the fused viral oncogenes, thus requiring orthogonal methods to distinguish the function of these lncRNA loci.

We also reanalyzed our data to understand the effect on our results of overlaps between lncRNA and protein-coding genes. In our original design, we allowed sgRNAs targeting regions to overlap with protein-coding genes to a certain extent to cover more lncRNAs. Nevertheless, we ensured that none of them targets any essential gene conserved in multiple cell lines, including K562 (ref. ⁶).

In K562 cells, among 230 essential lncRNAs, 70 hits have at least one sgRNA targeting an exon of a protein-coding gene. However, only one lncRNA (*AC137932.4*) overlaps with an essential gene, ankyrin repeat domain 11 (*ANKRD11*)⁶ (Table 2). There are four sgRNAs targeting this lncRNA (*AC137932.4*), including one dropout sgRNA targeting the intronic region (> 800 bp from the splice site) of *ANKRD11*. This result suggests that the disruption of this lncRNA, instead of its overlapping essential gene, led to cell death or growth inhibition, a result that needs to be further confirmed by orthogonal methods. Furthermore, among the other 69 lncRNAs overlapping with nonessential coding genes in K562, 39 of them had sgRNAs targeting non-exonic regions that showed significant dropout effects (Table 2).

In GM12878 cells, three lncRNAs (*LINC00969*, *RP11-2C24.4* and *TNK2-AS1*) are questionable according to previously published data on the essentiality of protein-coding genes in GM12878 cells⁷ (Table 2). However, the essentiality of all three of these three lncRNAs was supported by sgRNAs targeting intronic regions (at least 10 bp from the splice sites) of coding genes or intergenic regions. In HeLa cells, only two lncRNAs (*CTC-250I14.6* and *RP11-227G15.3*) are questionable as they overlap with essential genes found in a previously published screening⁸. For lncRNA *RP11-227G15.3*, one targeting sgRNA hits the 5' untranslated region of the overlapping coding gene *ZNF207*, which also showed dropout effect. Moreover, similarly to the results in K562, among those lncRNAs overlapping with nonessential coding genes in GM12878 and HeLa cell lines, about half of them have dropout sgRNAs cutting at non-exonic regions (Table 2).

Among 17 essential lncRNAs common to all three cell lines (removing three lncRNAs with CNV issues in HeLa cells, $\log_2 R > 0.3$), only five overlap with exons of nonessential genes (Table 2). To sum up, for those lncRNAs that overlap with essential protein-coding

genes, which are similar to those neighbor hits (lncRNAs neighboring essential coding genes) reported in the CRISPRi strategy³, more rigorous validations are required to determine their essential roles.

In addition to the two main concerns above, Horlbeck et al.¹ state that the validation strategy of paired-guide RNA (pgRNA) deletion is not sufficiently orthogonal to the primary screening with sgRNAs that target the splice sites. We think that, except for targeting copy-number amplification regions, pgRNA-mediated deletion could still be a useful validation method to acquire independent evidence if one adheres to the following strict criteria: first, the deletion regions should not overlap with any exons or promoters of protein-coding or noncoding genes; second, the pgRNAs should be sufficiently specific to minimize potential multiple cleavages in the genome (see Methods section in ref. ²); and third, the assay should include proper controls targeting safe regions (for example, the *AAVS1* site).

Different types of perturbation have their own limitations. It would therefore be helpful to employ other orthogonal strategies without introducing DNA double-strand breaks to validate candidate lncRNAs whenever possible. We believe this discussion will help improve the current methodologies for identifying functional lncRNAs in future studies.

Received: 4 July 2019; Accepted: 17 December 2019;

Published online: 24 February 2020

References

- Horlbeck, M. A. et al. Fitness effects induced by CRISPR/Cas9-targeting of long noncoding RNA gene loci. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0428-0> (2020).
- Liu, Y. et al. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nature Biotechnol.* **36**, 1203–1210 (2018).
- Liu, S. J. et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, aah7111 (2017).
- Adey, A. et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* **500**, 207–211 (2013).
- Jia, L., Zhang, Y., Tian, F., Chu, Z. & Xin, H. Long noncoding RNA colon cancer associated transcript1 promotes the proliferation, migration and invasion of cervical cancer. *Mol. Med. Rep.* **16**, 5587–5591 (2017).
- Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
- Ma, Y. et al. CRISPR/Cas9 Screens Reveal Epstein–Barr Virus-Transformed B Cell Host Dependency Factors. *Cell Host Microbe* **21**, 580–591 e587 (2017).
- Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Data availability

All data used are from our original paper².

Code availability

Source code for the computational analysis described in this paper is available in Supplementary Code.

Author contributions

Z.L. performed the bioinformatics analysis. Y.L. and Z.C. integrated all the results. Y.L., Z.C. and W.W. wrote the manuscript with the help of Z.L.

Competing interests

W.W. is a founder of and scientific advisor to EdiGene.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-020-0431-5>.

Correspondence and requests for materials should be addressed to W.W.

Reprints and permissions information is available at www.nature.com/reprints.