studies lie within coding sequences. While this preference for noncoding variation is likely driven by functional and evolutionary constraints, it is intriguing to consider that a DNA repair bias in favor of exons might reduce the input of coding mutations on which evolution can operate. The diminished exonic mutation burden also suggests a possible connection between evolution of gene architecture and mutation. How entangled are the exon-defining and exon-protective functions of H3K36me3? Has preferential recruitment of repair proteins to protein-coding sequences facilitated the evolution of intronic sequences, or has the repair machinery co-opted an exon-defining chromatin state to drive down mutation rates in coding sequences?

Continuing efforts to discover cancer driver genes[6] should take into account this new form of local mutation rate heterogeneity. In particular, the possibility that certain genes deviate from this general trend should be kept in mind. Future studies, harnessing ever-larger catalogs of cancer mutations, could address how the within-gene mutational biases uncovered here interact with other forms of mutational heterogeneity at smaller and larger scales. In addition, studies of somatic mutations, *de novo* mutations in family pedigrees[8], mutational patterns inferred from population-scale sequencing[2] and studies in other species could investigate whether preferential DNA repair in exons has left a mark on non-cancer genomes. Cancer genomics continues to highlight new forms of mutational complexity and unexpected mechanisms of mutagenesis[14,15]. The study by Frigola *et al.*[9] adds an important piece to the puzzle of how gene regulation, chromatin structure and DNA repair work together to shape the mutational landscape of the genome.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Youssoufian, H. *et al. Nature* **324**, 380–382 (1986).
2. Carlson, J. *et al. bioRxiv,* preprint at https://doi.org/10.1101/108290 (2017).
3. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. *Nature* **532**, 264–267 (2016).
4. Perera, D. *et al. Nature* **532**, 259–263 (2016).
5. Schuster-Böckler, B. & Lehner, B. *Nature* **488**, 504–507 (2012).
6. Lawrence, M.S. *et al. Nature* **499**, 214–218 (2013).
7. Haradhvala, N.J. *et al. Cell* **164**, 538–549 (2016).
8. Jónsson, H. *et al. Nature* **549**, 519–522 (2017).
9. Frigola, J. *et al. Nat. Genet.* **49**, 1684–1692 (2017).
10. Li, J. *et al. PLoS Comput. Biol.* **11**, e1004583 (2015).
11. Li, F. *et al. Cell* **153**, 590–600 (2013).
12. Schwartz, S., Meshorer, E. & Ast, G. *Nat. Struct. Mol. Biol.* **16**, 990–995 (2009).
13. King, M.C. & Wilson, A.C. *Science* **188**, 107–116 (1975).
14. Supek, F. & Lehner, B. *Cell* **170**, 534–547 (2017).
15. Imielinski, M., Guo, G. & Meyerson, M. *Cell* **168**, 460–472 (2017).

Correcting CRISPR for copy number

John Paul Shen & Trey Ideker

The CRISPR–Cas9 system enables global screens of gene function with high sensitivity and specificity, but off-target effects have been reported for CRISPR guide RNAs targeting genes that are amplified at high copy number. A new study describes a computational approach to correct for this copy number effect, increasing the specificity of CRIPSR screens to identify essential genes.

Clustered regularly interspaced short palindromic repeat (CRISPR)–Cas9 technology has revolutionized the field of functional genomic screening owing to its high sensitivity and specificity in disrupting genes and the fact that CRISPR can be easily scaled to test tens of thousands of gene disruptions in a single pooled experiment[1]. Performing these genomic screens over large numbers of cancer cell lines has allowed for the identification of core essential genes and the differentiation of these from tumor-specific dependencies[2]. Already, CRISPR screens have been used successfully to identify potential drug targets in many cancer types[3]. Additionally, CRISPR technology can be multiplexed to knock out multiple genes in the same cell to facilitate identification of gene–gene and gene–drug interactions[4,5].

Recently, multiple groups have reported that single guide RNAs (sgRNAs) targeting copy number–amplified (CNA) regions of the genome can decrease cell proliferation in a manner that is independent of the specific gene being targeted[6–8]. sgRNAs targeting amplified regions create a greater number of DNA double-strand breaks than those targeting non-amplified regions, and these double-strand breaks lead to induction of the DNA damage response and, ultimately, cell cycle arrest[6]. Because this hypoproliferative effect is indistinguishable from that of an sgRNA targeting an essential gene in a pooled CRISPR screen, copy number effects prevent the accurate assessment of functional importance for genes in CNA regions (**Fig. 1**).

John Paul Shen and Trey Ideker are in the Department of Medicine, University of California, San Diego, La Jolla, California, USA.
e-mail: tideker@ucsd.edu

Pervasive copy number effects

On page 1779 of this issue, Aviad Tsherniak, William Hahn and colleagues[9] present a computational approach to correct for CRISPR copy number effects. Using new data from genome-wide CRISPR screens in 342 cancer cell lines, Meyers *et al.*[9] first showed that they could recapitulate the strong association between the number of copies of the locus targeted by an sgRNA and the degree of depletion of the sgRNA from the pool of cells. Remarkably, this effect was seen in every cell line examined, and the degree of depletion was of a magnitude similar to that for sgRNAs targeting essential genes. Similar results were seen in a reanalysis of two previously published genome-wide CRISPR–Cas9 loss-of-function screens[6,10]. Interestingly, the effect on sgRNA depletion of increasing numbers of genomic cut sites was lessened in cell lines with mutant *TP53*, presumably because these cells have defects in the activation of cell cycle checkpoints in response to DNA damage.

Next, Meyers *et al.*[9] developed a computational tool they call CERES, which models sgRNA depletion mathematically as sum of (specific) gene-knockout and (non-specific) copy number effects, scaled by guide activity to account for differing sgRNA cutting efficiencies. This model was then used to 'normalize out' the effects of copy number, leaving only the specific, gene-knockout effects. Using CERES, they reanalyzed their CRISPR data set to show that the correlation between copy number and sgRNA depletion was greatly decreased. Recall of common core

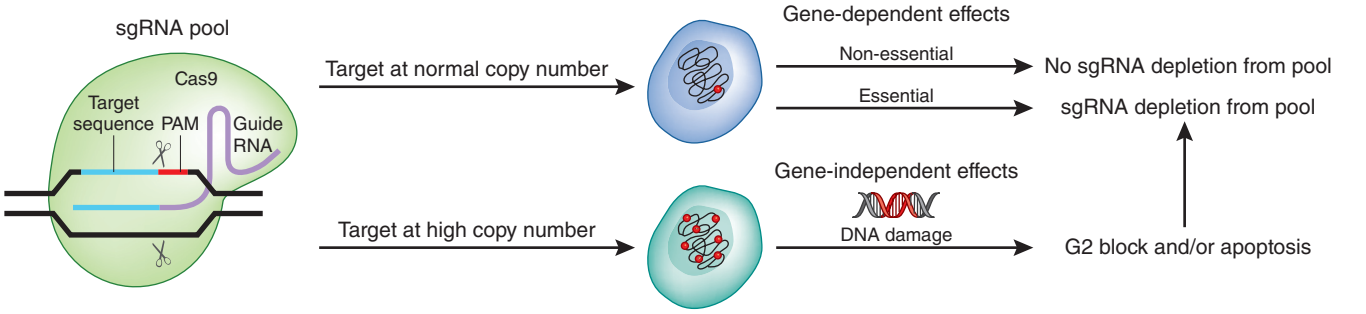

Figure 1 Gene-independent copy number effects in CRISPR–Cas9 essentiality screens. sgRNAs targeting amplified regions create more DNA double-strand breaks than those targeting non-amplified regions, leading to induction of DNA damage response and cell cycle arrest. This effect is indistinguishable from that of an sgRNA targeting an essential gene in a pooled CRISPR screen, preventing accurate assessment of the functional importance of genes in CNA regions. PAM, prospacer-adjacent motif.

essential genes[1] was improved by an average of 14% per cell line. Importantly, CERES does not require removal of genes at high copy number; thus, an average of 134 additional genes per cell line, which would have otherwise been removed by simple filtering methods[10], could be analyzed.

Finding drivers in amplicons

Induction of double-strand breaks, leading to DNA damage response and apoptosis, is not the only reason why genes at high copy number might be associated with sgRNA depletion. There is another quite plausible explanation: high–copy number regions serve to amplify oncogenes essential for tumor proliferation, with sgRNA targeting of these genes leading—as expected—to sgRNA depletion. To address this point, Meyers *et al.*[9] confirmed that CERES is able to identify known oncogenic driver genes, such as *KRAS* and *BRAF*, in high-CNA regions, suggesting that CERES does not entirely normalize away these gene-specific, on-target effects. As a related point, they showed convincingly that the model is able to identify the specific oncogenes driving proliferation from broad CNA regions. This ability is of particular interest given that CNA regions in tumors can be large, encompassing many genes.

Another feature of note from this resource is the estimated activity score for each sgRNA. It is well known that sgRNAs targeting the same gene can differ greatly in activity according to location of the specific sgRNA target as well as differences in the efficiency of Cas9 recruitment[11–13]. However, it is only practical to validate the activity of a small fraction of the tens of thousands of sgRNAs in a pooled genome-wide library. To account for variability in sgRNA activity, pooled CRISPR–Cas9 screens are performed with as many as 12 unique sgRNAs per gene[1]. More accurate estimates of sgRNA activity, such as those derived from applying CERES to over 300 cell lines, could allow future screens to be performed with far fewer sgRNAs, reducing library size and, in turn, decreasing costs for both tissue culture and sequencing.

One of the primary values of any large CRISPR knockout data set is the identification of cell-specific genetic vulnerabilities. Above all, Meyers *et al.*[9] have shown how heavily confounded this analysis can be: before correction, as much as 30% of CRISPR hits are due to copy number effects, a percentage that drops to less than 5% after analysis by CERES. Given the magnitude of these confounding effects, it may be time to revisit many previous CRISPR screens and certainly to add CERES to the bioinformatics toolbox.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

1. Hart, T. *et al. Cell* **163**, 1515–1526 (2015).
2. Tsherniak, A. *et al. Cell* **170**, 564–576 (2017).
3. Fellmann, C., Gowen, B.G., Lin, P.-C., Doudna, J.A. & Corn, J.E. *Nat. Rev. Drug Discov.* **16**, 89–100 (2017).
4. Han, K. *et al. Nat. Biotechnol.* **35**, 463–474 (2017).
5. Shen, J.P. *et al. Nat. Methods* **14**, 573–576 (2017).
6. Aguirre, A.J. *et al. Cancer Discov.* **6**, 914–929 (2016).
7. Munoz, D.M. *et al. Cancer Discov.* **6**, 900–913 (2016).
8. Wang, T. *et al. Science* **350**, 1096–1101 (2015).
9. Meyers, R.M. *et al. Nat. Genet.* **49**, 1779–1784 (2017).
10. Wang, T. *et al. Cell* **168**, 890–903 (2017).
11. Chari, R., Mali, P., Moosburner, M. & Church, G.M. *Nat. Methods* **12**, 823–826 (2015).
12. Doench, J.G. *et al. Nat. Biotechnol.* **32**, 1262–1267 (2014).
13. Shi, J. *et al. Nat. Biotechnol.* **33**, 661–667 (2015).