In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription

Amjad Askary1, Luis Sanchez-Guardado1, James M. Linton1, Duncan M. Chadly1, Mark W. Budde1, Long Cai1, Carlos Lois1 and Michael B. Elowitz1,2*

1Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. 2Howard Hughes Medical Institute and Department of Applied Physics, California Institute of Technology, Pasadena, CA, USA. *e-mail: melowitz@caltech.edu
Supplementary Figure 1

Zombie active sites are only found in the cells were they are made, whereas the individual transcripts can diffuse away and be detected in cells other than their cell of origin.

(a) Co-culture of mES-Z1 cells with non-transgenic parental cells. Active sites are only found in CFP+ cells. However, small diffraction limited dots are found in all cells including the non-transgenic ones. (b) Same image as a, but with blue and cyan channels turned off for better visibility of the barcode signal. A few examples of individual barcode transcripts are marked by circles. The inset shows a magnified and contrast adjusted view of a CFP negative cell, marked by the square, which contains some small barcode dots, indicating diffusion of individual RNA molecules from the cells in which they are produced. (c) In the absence of any transgenic cells, background non-specific signal is low. Indicating that the signal observed in the presence of transgenic cells is not non-specific HCR amplification. (d) Same image as c, but with blue and cyan channels turned off to make the lack of small dots more evident. The experiments were independently repeated three times with similar results. Scale bar is 25 μm.
Supplementary Figure 2

Histogram of intensity for the brightest dot in each cell.

Bimodal distribution is consistent with presence or absence of active sites in the cells.
Supplementary Figure 3

Mutual information analysis of pairwise correlations between Z3 barcodes.

Diagonal elements are set to 1 by definition. Off-diagonal elements represent normalized mutual information (i.e. uncertainty coefficient) between detection of indicated barcode pairs. Low values are consistent with independent detection. For each pair of barcodes, detecting one at a given site does not significantly alter the probability of detection of the other (chi-square test, p>0.1). Total of 564 cells were analyzed.
Supplementary Figure 4

Reliable detection of 20bp targets with individual HCR probes.

Images show same treatment as in Figure 2B, except with HCR amplification. Images are scaled to different intensity ranges compared to Figure 2B. The experiment was independently repeated three times with similar results. Scale bar is 25 μm.
Supplementary Figure 5

In the absence of competition, probes with a single mismatch can bind to their targets in the active site and generate significant fluorescent signal.

The signal from probes with a single mismatch (A, G, or C) is minimal when they are hybridized together with a match probe (T) in an equimolar mixture (Figure 3C, reproduced here in the box with the panels not relevant to the current experiment shaded). However, when hybridized individually, without competition, they generate considerable signal in the active sites. Representative images are shown outside the box, next to their corresponding condition. All images were acquired and processed under the same conditions for each channel. Histograms show the distribution of signal intensity (natural log of the intensity of the brightest dot in each cell) for the mismatch probes in the presence and absence of competition. Total of 2374 cells were analyzed for the no competition conditions, with at least 295 cells for each condition. The bimodal distributions, in the absence of competition, reflect a subset of cells with bright active sites. This background signal is largely reduced in the presence of competition. These results suggest that probe competition is necessary for discrimination of single nucleotide variants.
Supplementary Figure 6

Representative images showing discrimination based on a single nucleotide mismatch.

All images are acquired under the same conditions, and brightness for each channel is adjusted identically across all the images. Each column represents one experiment in which four probes with a single nucleotide variation and orthogonal HCR initiators (B1, B2, B3, and B4) were mixed and hybridized to the sample. The identity of the variable nucleotide is shown by the letter on the panels. HCR initiators and the fluorescent channels used for each probe are shown next to the rows. See Figure 3d for quantification of the results. Scale bar is 10 μm.
Supplementary Figure 7

SNV detection is robust to the position of the variant base pair in the barcode sequence.

Matrices represent SNV analysis, as in Figure 3D, with four distinct color permutations, with the indicated target nucleotide at positions 1 through 7 (starting from the 5' end of the probe). Accurate discrimination can be achieved for positions 2 to 7. Even position 1 provides discrimination ability. Total of 9364 cells were analyzed, with at least 234 cells for each permutation.
Supplementary Figure 8

Bootstrap analysis of active site classification.

For each barcode, boxplots show the fraction of active sites that were classified differently when the data were resampled with replacement. The central red line in each box indicates the median, across 5000 bootstrap rounds, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Red lines at zero, in design 2, indicate that no dots changed their classification.
Design 1 barcodes can be classified based on signal intensity in edited and unedited channels.

(a) Scatter plots show natural log of signal intensity in the edited (y-axis) versus unedited (x-axis) channels for each barcode. Negative controls lacking ABE and gRNA (blue) are superimposed on samples transfected with all components (orange). Edited (upper orange cloud) and unedited (lower blue cloud) active sites are broadly distinguishable. (b) Natural log of signal intensity in edited versus unedited channels color coded based on the frequency (%) by which classification of each dot, as edited or unedited, is changed across 5000 rounds of bootstrap resampling. All points here are from samples transfected with GFP, ABE, and the gRNA of the specified barcode.
Design 2: negative control versus cells transfected with ABE and gRNA

Supplementary Figure 10
Design 2 barcodes can be classified based on signal intensity in edited and unedited channels.

(a) Scatter plots show natural log of signal intensity in the edited (y-axis) versus unedited (x-axis) channels for each barcode. Negative controls lacking ABE and gRNA (blue) are superimposed on samples transfected with all components (orange). Edited (upper orange cloud) and unedited (lower blue cloud) active sites are broadly distinguishable. (b) Natural log of signal intensity in edited versus unedited channels color coded based on the frequency (%) by which classification of each dot, as edited or unedited, is changed across 5000 rounds of bootstrap resampling. All points here are from samples transfected with CFP, ABE, and the gRNA of the specified barcode.
Supplementary Figure 11

Edit frequencies measured by Zombie are similar to those measured by next generation sequencing.

HEK293T cells containing multiple integrations of lentivirally delivered design 1 memory array were transiently transfected with plasmids for ABE7.10, a barcode specific gRNA, and CFP (orange points). As negative control, we also transfected a separate group with CFP but not ABE7.10 and gRNA (blue points). 5 days after transfection, some cells from each group were analyzed by Zombie, similar to Fig. 4, and the rest were analyzed by next generation sequencing (see methods). We detected no edits by Zombie in four negative control samples (not plotted). Error bars are 95% binomial confidence intervals, calculated using Clopper-Pearson method. Number of active sites analyzed by Zombie were 4251, 1237, 2910, 3466, 4883, 3095, 2465, 1501, and 1991 for barcodes 1 through 10, respectively, in ABE and gRNA positive condition (orange) and 3650, 3293, 4496, 5508, 5347, 3986, 5605, 5020, 2790, and 2142 for barcodes 1 through 10, respectively, in the control condition (blue).
Supplementary Figure 12

Zombie accurately discriminates barcodes with single nucleotide variations in mouse brain tissue.

Scatter plots showing natural log of signal intensity for two variants (A and G) of two barcodes (1 and 2), as in Fig. 5G, for lentivirus pairs 2 (left) and 3 (right). Each point represents one active site. The experiment was performed on brain sections from two mice. Biological duplicates showed similar results.
Supplementary Figure 13

Overlapping barcode integration sites can result in underestimation of Zombie SNV detection accuracy in mouse brain sections.

(a) Correlation between two SNVs engineered in the same virus can be used to estimate SNV detection accuracy in tissue samples transduced by the viral mix (Schematic reproduced from Fig. 5F). The lentivirus pairs are designed so that each active site incorporates either an A in both barcodes 1 and 2, or a G in both barcodes. (b) Maximum intensity projection of a confocal stack shows transduced cells in a section of mouse olfactory bulb. Scale bar is 50 μm. (c) Injection of lentivirus mix into the olfactory bulb can result in the integration of multiple viral genomes, containing different barcodes, in the same cell. Imaging reveals multiple “GG” (arrows) and “AA” (arrowheads) integration sites in the same cell, which permit accurate classification. (d) In some cases, integration sites for two virus pairs overlap in the nucleus (dashed circle), leading to an erroneous SNV call. Upper and lower images are identical overlays of the four images in (c), but the lower image also includes CFP fluorescence in gray. The experiment was repeated on two biologically independent samples with similar results. Scale bar is 5 μm.
Supplementary Figure 14

In situ readout of a viral library with 81 combinations in three rounds of hybridization and imaging.

In each round, tissue sections were analyzed using 4 probes, in distinct fluorescence channels, corresponding to three variants in one of the barcode positions 1, 2, or 3 and one variant in position 4 (see Fig. 6A for the design of the library). As a result, in each round, some active sites were visualized in two channels (shown as semi-circles in this illustration). Information from images of all three rounds was then combined to decode the identity of each active site.
Supplementary Figure 15

Zombie barcode detection is compatible with in situ detection of endogenous gene expression in tissue sections.

(a) Maximum projected confocal images of an olfactory bulb section are tiled to show a larger field of view. The barcode was delivered by injection of a lentivirus that also expresses H2B-Cerulean under human UbiC promoter. Expression of Tbx21 and Tyrosine hydroxylase (Th) was visualized by HCR FISH. CFP is detected based on its native fluorescence, without any further staining. (b) Although there is a correlation between expression of CFP and detection of Zombie active sites, there are instances of cells with low or no CFP that have an active site (arrow), as well as those that show CFP expression but no active site (arrowhead). We speculate that the former is caused by lack of expression of CFP from the integrated viral genome (e.g. due to silencing) and the latter is indicative of imperfect barcode detection efficiency. (c, d) Magnified views showing Tbx21 (green) and Th (red) endogenous mRNA detected by HCR in two orthogonal channels. Four biological replicates showed similar results. Scale bar is 200μm.
Supplementary Figure 16

Formaldehyde (PFA) fixation prior to in situ transcription results in a drastic decrease in detection efficiency.

Histogram of intensity of the brightest dot in each cell is shown for different fixation and permeabilization conditions. Fraction of cells with active sites decreases significantly when cells are fixed by 1% PFA and permeabilized by 3:1 mixture of methanol and acetic acid (MAA). Fixation by 2 and 4% PFA leads to almost complete lack of Zombie active site in cells. For this reason, PFA fixation is not used prior to in situ transcription.
Supplementary Figure 17

Effect of transcription time and fixation on detection efficiency.

Increasing transcription time from 15 min to 3 hours has a modest effect on transcription efficiency. However, fixing with MAA (3:1 mix of methanol and acetic acid) increases efficiency considerably compared to fixing with 100% methanol.
Supplementary Figure 18

The ratio of acetic acid to methanol in the fixation step prior to in situ transcription affects detection efficiency.

Histogram of intensity of the brightest dot in each cell is shown for different acetic acid to methanol ratios. In this paper, we use 25% acetic acid in methanol for fixation. However a modest gain in efficiency can be obtained by increasing acetic acid to 35 or 50 percent.
**Supplementary Table 2. List of experimental conditions and their effect on barcode detection efficiency.**

| Condition                        | Description                                                                 | Result                                                                                                                                 |
|----------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| PFA fixation                     | Fixed with 1, 2, and 4% formaldehyde solution in PBS, followed by permeabilization by either 3:1 (v/v) mix of methanol and acetic acid or 70% ethanol. | Fixation with PFA prior to transcription step drastically reduced the detection efficiency (see Fig. S4)                              |
| Methanol and acetic acid fixation| Fixed with 100% methanol as well as 5, 15, 25, 35, and 50% acetic acid in methanol solutions. | Mix of acetic acid in methanol provides the best results (see Figures S5 and S6).                                                    |
| Clarke’s fluid fixation          | Fixed with 3:1 (v/v) mix of Ethanol:Acetic Acid for 15 minutes at room temperature. | Observed a decrease in detection efficiency compared to 3:1 MAA fixation.                                                              |
| Methanol and acetone fixation    | Fixed with 1:1 (v/v) mix of methanol and acetone for 15 minutes at room temperature. | Observed a drastic decrease in the detection efficiency.                                                                                 |
| Proteinase K treatment           | Permeabilized the cells initially using 1, 5, and 10ug/ul Proteinase K for 11min at room temperature and in a subsequent experiment using 1 ug/ul of Proteinase K for 1, 2, 5, and 10 min at room temperature. | All of these treatments led to loss of most cells. We don’t rule out the possibility that it may improve permeability in some tissues, but in our experience so far it has not been necessary. |
| Triton X-100                     | Washed the cells with 0.5% Triton X-100 for 10 minutes at room temperature after fixation by 3:1 MAA mix. | Observed no advantage over not washing the cells with this solution.                                                                   |
| SDS                              | Washed with 0.1% SDS for 10 minutes at room temperature after fixation by 3:1 MAA mix. | It severely affected the cell morphology, therefore we decided not to use this treatment.                                             |
| Histone wash                     | Washed with 2 mg/ml Dextran sulfate (MW 500,000), 0.2 mg/ml Heparin sodium salt, 0.1% IGEPAL CA-630, 10mM EDTA, 10mM Tris pH = 8.0 in nuclease free water for 10 minutes at room temperature after fixation by 3:1 MAA mix. | Observed a slight decrease in detection efficiency.                                                                                   |
| RNA polymerase concentration     | Performed transcription with 2, 5, 10, 15, and 20U/ul T7 RNA polymerase at 37°C for 3 hours. | Observed no gain in efficiency for concentrations above 5U/ul.                                                                       |
| Duration of transcription reaction| Performed transcription with T7 RNA polymerase for 15, 30, 60, and 180 minutes at 37°C. | Duration of the transcription reaction only has a modest effect on detection efficiency (see Fig. S5)                                |

PFA, paraformaldehyde; MAA, mix of methanol and acetic acid (v/v).