

# ADVANCED MATERIALS

## Supporting Information

for *Adv. Mater.*, DOI: 10.1002/adma.202204957

### Massively Parallel Selection of NanoCluster Beacons

*Yu-An Kuo, Cheulhee Jung, Yu-An Chen, Hung-Che Kuo, Oliver S. Zhao, Trung D. Nguyen, James R. Rybarski, Soonwoo Hong, Yuan-I Chen, Dennis C. Wylie, John A. Hawkins, Jada N. Walker, Samuel W. J. Shields, Jennifer S. Brodbelt, Jeffrey T. Petty, Ilya J. Finkelstein, and Hsin-Chih Yeh\**

Supporting Information

**Massively Parallel Selection of NanoCluster Beacons**

*Yu-An Kuo, Cheulhee Jung, Yu-An Chen, Hung-Che Kuo, Oliver S. Zhao, Trung D. Nguyen, James R. Rybarski, Soonwoo Hong, Yuan-I Chen, Dennis C. Wylie, John A. Hawkins, Jada N. Walker, Samuel W. Shields, Jennifer S. Brodbelt, Jeffrey T. Petty, Ilya J. Finkelstein, and Hsin-Chih Yeh\**

**Experimental methods:***NC probe preparation:*

The sequences of all DNA used in this report can be found in **Table S1-S12**. All oligonucleotides were purchased from Integrated DNA Technologies (IDT) and most of them were purified by desalting. Only the dye-labeled oligos were purified by HPLC. Sodium phosphate dibasic anhydrous ( $\text{Na}_2\text{HPO}_4$ ; F.W. 141.96), sodium phosphate monobasic monohydrate  $\text{Na}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$ ; F.W. 137.99) and sodium borohydride ( $\text{NaBH}_4$ ) were purchased from Fisher Scientific, whereas silver nitrate ( $\text{AgNO}_3$ ) was acquired from Sigma-Aldrich. Deionized (DI) water (18  $\text{M}\Omega \cdot \text{cm}$ ) was used for all solution preparations.

In a typical preparation, a 15  $\mu\text{M}$  (final concentration) NC probe solution was prepared by adding 12.5  $\mu\text{l}$  of 1.2 mM NC probe (C55, **Table S1**) to 940  $\mu\text{l}$  of 20 mM sodium phosphate buffer (pH 6.6). The solution was vortexed for 2 s, and 45  $\mu\text{l}$  of 4 mM silver nitrate solution was added to it. Again, the mixture was vortexed for 2 s. The solution was allowed to sit in the dark for 10 min at room temperature. For silver cluster formation, 7  $\mu\text{l}$  of freshly prepared 13.2 mM  $\text{NaBH}_4$  solution was added to the reaction, resulting in a pale-yellow-orange mixture, which was then stored in the dark overnight. The resulting NC probe solution had the [NC probe]: [ $\text{Ag}^+$ ]: [ $\text{NaBH}_4$ ] molar ratio of 1:12:6.

Before *MiSeq* chip experiment, a 0.5 ml centrifugal filter (cat. no. UFC503024, MilliporeSigma) was employed to remove excess silver ions. Purification protocol followed the manufacturer guidelines. The filtered solution was then diluted to 500 nM (DNA concentration, which was verified using the NanoDrop 2000 UV-Vis spectrophotometer, Thermo Scientific) before being injected into *MiSeq* chip.

*NCB preparation and in-solution validation:*

To activate NCB, 1.5  $\mu\text{l}$  of 1.2 mM activator solution was added to a 120  $\mu\text{l}$  aliquot of the previously prepared 15  $\mu\text{M}$  C55 probe solution. The mixture was vortexed, centrifuged, and immersed in a hot water bath (90-95°C) for 1 min, followed by gradually cooling down to room temperature for 1 hr. The activated NCB had the [NC probe]: [activator] molar ratio of 1:1. The fluorescence measurements started exactly at 1 hr after the addition of activator.

We quantified NCB fluorescence using a fluorometer (FluoroMax-4, Horiba) and a 100  $\mu\text{l}$  quartz cuvette (16.100F-Q-10/Z15, Starna Cells). Both the excitation and emission wavelength scan ranges were set to be from 400 nm to 800 nm using 5 nm slit size, 5 nm increment step, 0.1 s integration time. Two control samples, a 20 mM sodium phosphate pH 6.6 buffer only

sample and an NC probe only sample (with AgNCs but no activators) were also measured (**Figure S3**). The acquired spectra were saved as csv files and processed using a Python script.

*NCB fluorescence enhancement ratio calculation:*

We followed a similar definition described in ref. 7b to calculate the ensemble enhancement ratio of NCB after activation. However, in ref. 7b, 1D spectra based on 580 nm excitation were acquired and area integrated intensities were calculated over 595 to 740 nm emission range. Here we collected 2D spectra of samples and calculated volumetric integrated intensities over the red (Ex/Em: 620/60, 700/75 nm) and the yellow-orange (Ex/Em: 535/50, 605/70 nm) excitation/emission “windows” (**Figure S2**). From there we calculated the enhancement ratio:

$$\text{Enhancement ratio} = \frac{(I_{\text{NCB}} - I_{\text{background}}) - (I_{\text{NC probe}} - I_{\text{background}})}{I_{\text{NC probe}} - I_{\text{background}}} = \frac{I_{\text{NCB}} - I_{\text{NC probe}}}{I_{\text{NC probe}} - I_{\text{background}}}$$

, where  $I_{\text{NCB}}$  stands for the volumetric integrated intensity of NCB in red or yellow-orange window,  $I_{\text{NC probe}}$  represents the volumetric integrated intensity of dark AgNC on the C55 probe, and  $I_{\text{background}}$  is the volumetric integrated intensity of the sodium phosphate buffer. The improvement ratio is simply the ratio of the enhancement ratio of an activator to that of the standard activator (G12 or G15). Similarly, the POT difference ratio is simply the ratio of the enhancement ratios of the twins:

POT difference ratio =

(enhancement ratio of bright twin NCB) / (enhancement ratio of dark twin NCB)

*NCB fluorescence visualization:*

Color photos of inactivated (NC probe only) and activated (the duplex) NCBs were acquired using a digital camera (PowerShot SX 500 IS, Canon) on a Syngene gel imager (with 365 nm excitation) (**Figure 1B**). NCB fluorescence were also visualized using a gel imaging scanner (Typhoon 9500, GE Healthcare Life Sciences). For the Typhoon 9500 experiments, 240  $\mu\text{l}$  of NCB sample was placed in a single well on the 96 multi-well plate. The fluorescence was acquired using the built-in Cy3 channel (EX: 532 nm, EM: 575 nm long pass) or the Cy5 channel (EX: 635 nm, EM: 665 nm long pass) while the PMT gain was set to 400 and the pixel size was 10  $\mu\text{m}$ . The imaging results were saved as tiff files and changed to 16-bit false colors (yellow-orange for Cy3-channel imaging and red for Cy5-channel imaging) using ImageJ.

*Activator library design:*

All activator strands contained a universal, 30-nt-long TA-rich hybridization segment followed by an 18-nt-long variable region (the activator) and an 8-nt-long restriction site (**Table S1**). Additional adapters at 5'-end (P5 and SP1) and 3'-end (P7 and SP2) were designed by Illumina for sequencing purpose. To identify our library sequences, a 6-nt-long barcode was also needed and that was added to the 3'-end. The barcode for the canonical activator G15 was different from that of any other activators in order to monitor the sequencing yield. Three different libraries were established based on shifted frames, giving totally 40,068 unique activator sequences (**Table S1**).

#### *NGS library preparation:*

A standard PCR process was performed using Q5 high-fidelity DNA polymerase from NEB (cat. no. M0491S). All PCR primers were purchased from IDT (**Table S1**). The PCR procedure and the thermal cycler (Eppendorf, Mastercycler<sup>®</sup> nexus) settings followed the protocol provided by NEB. The library sequences and the canonical activator were amplified separately, reaching a final concentration greater than 5 ng/μl for each tube. After PCR amplification, the concentration of DNA library was verified using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific). Together with the fiducial markers (PhiX), the library sequences were immobilized and bridge amplified on an Illumina *MiSeq* chip, followed by sequencing using a PE300 reagent kit (v3, Illumina). For the 3-segment interrogation (**Table 1**), we targeted to have 1.2 million reads for the mutations in each segment. 10,000 additional reads were spiked in for the canonical activator G15. The fiducial markers, PhiX, were counted for 15%~20% of the overall coverage. The actual numbers of reads varied from batch to batch, but were within 5-13 million reads.

#### *MiSeq chip preparation:*

After sequencing, *MiSeq* chips were kept at 4°C in storage buffer (1X TBE buffer, cat. no. AM9865, Invitrogen). Before hybridizing with NC probes, all DNA strands covalently affixed to the *MiSeq* chip surface were denatured with 20 μl 0.1N NaOH solution for 5 min and then rinsed with 20 μl 1X TBE buffer 3 times to remove excess NaOH. This rinsing step removed untethered DNA strands containing residual fluorescent dyes from sequencing. Before the NCB screening experiment, chip was rinsed with working buffer (150 μl, 200 mM sodium phosphate buffer pH 6.6) three times. To cleave the unwanted sequence beyond the activator sequence, a 32-nt-long strand complementary to the restriction site (RE strand in **Table S1**) was introduced to the chip and the chip was annealed at 40°C for 40 min. After annealing, 1X TBE buffer was

used to rinse the chip, followed by MauBI (cat. no. ER2081, Thermo Fisher Scientific) restriction enzyme digestion. The reaction buffer was prepared following the manufacturer's protocol and 20  $\mu\text{l}$  enzyme solution was kept in the *MiSeq* chip at 40°C for 40 min. After digestion, the chip was washed with 20  $\mu\text{l}$  0.1N NaOH solution for 5 min and rinsed with 20  $\mu\text{l}$  1X TBE buffer for three times. The PhiX sequences were labeled using an Alexa488-tagged probe (500 nM and 20  $\mu\text{l}$ , **Table S1**) for fiducial marker imaging. To optimize the annealing conditions, we evaluated the intensities of NCBs that had gone through different temperature treatments (40°C for 40 min, room temperature for 40 min, and 90°C for 10 min, **Figure S6**). Holding the chip at 40°C for 40 min not only gave an excellent annealing result but also extended the chip life to up to 20 runs of NCB activation experiments. After testing different concentrations of C55 probes for the NCB screening experiment (**Figure S11**), we chose to use 20  $\mu\text{l}$  of 500 nM C55 probe solution for all our chip experiments. The chip was imaged at room temperature with microscope settings stated below. After each experiment, the chip was washed with 20  $\mu\text{l}$  0.1N NaOH solution for 5 min, followed by rinsing with 20  $\mu\text{l}$  1X TBE buffer for three times and storing at 4°C.

*Fluorescence microscopy and image acquisition:*

An open-source software, Micro-Manager,<sup>[1]</sup> was used to control an sCMOS camera (ORCA-Flash 4.0, Hamamatsu), an xyz translation stage (ProScan III, Prior Scientific), and an auto-shutter (Lambda SC, Shutter Instrument) on an Olympus IX71 fluorescence microscope for all our *MiSeq* screening experiments. A metal-halide illuminator (Lumen 200, Prior) and a 60 $\times$  water-immersion objective (UPLSAPO60XW, Olympus) were used in the IX71 system. We developed a MATLAB script to generate the position list of each field of view for automatic acquisition. On each *MiSeq* chip, we acquired 60 images (FOV: 220 x 220  $\mu\text{m}^2$ ) per row for 3 rows on both floor and ceiling (totally 360 images), covering a total surface area of 5.81  $\text{mm}^2$ . It is worthwhile to note that colonies in some regions of the chip were not registered in the Illumina sequencing files. To bypass most of these unregistered regions, we shifted the imaging starting position by 380  $\mu\text{m}$  in the vertical direction and 1,611  $\mu\text{m}$  in the horizontal direction with respect to the reference point at the bottom left corner (**Figure S7C**). We first recorded fiducial marker images (Alexa488, FOV: 220 x 220  $\mu\text{m}^2$ , 1 second exposure time, green channel, Ex/Em: 480/40, 535/50 nm, cat. no. 51006, Chroma), and then recorded NCB images in both red (Ex/Em: 620/60 nm, 700/75 nm, cat. no. 49006, Chroma) and yellow-orange channels (Ex/Em: 535/50 nm, 605/70 nm, cat. no. 49004, Chroma) under the same imaging settings. Photobleaching on the *MiSeq* chip (2-7% decrement after 200 ms exposure, **Figure S10**) did

not prevent us from correctly ranking the brightness of NCBs, as the excitation power ( $\sim 10$  W/cm<sup>2</sup>), the exposure time (200 ms), and the scanning step size (220  $\mu$ m per step) were all precisely controlled by our auto-scan algorithm and shutter control. As shown in the **Figure S12-S13**, our NGS screening results were consistent and reproducible.

*Flat-field correction:*

We implemented flat-field correction to eliminate the variation of fluorescence background across the field of view (FOV, **Figure S7A-B**). A Gaussian-blur filter was applied to generate the flat-field reference image for each FOV. We found that a Gaussian-blur filter with sigma equal to 50 best fit our purpose. The corrected images used for the following analysis were computed as follows:

$$I_{\text{corrected}} = (I_{\text{measure}} - I_{\text{dark}}) / (I_{\text{flat-field image}} - I_{\text{dark}})$$

where  $I_{\text{measure}}$  is the recorded fluorescence images,  $I_{\text{flat-field image}}$  is the flat-field reference image generated by the Gaussian-blur filter, and  $I_{\text{dark}}$  is the dark image recorded using 1 second exposure time while illuminator is turned off.

*Fluorescence correlation spectroscopy (FCS) measurements:*

FCS measurements were carried out using a confocal system (Alba v5, ISS) built around a Nikon microscope body (Eclipse TE2000-U, Nikon). A super-continuum laser (SuperK EVO EU-4, NTK Photonics) and a 60 $\times$  water-immersion objective (CFI Plan Apochromat VC, Nikon) were used in the FCS experiments. To validate the NCB chemical yield measurement results, an Atto647N-labeled ssDNA probe and an Atto532N-labeled ssDNA probe were used to generate the concentration calibration curve shown in **Figure S27-S28**. All FCS measurements were carried out using 200  $\mu$ l samples in 8-well chamber slides (Nunc Lab-Tek, Thermo Fisher Scientific). The laser beam was focused 25  $\mu$ m into the sample for all FCS measurements in this work.

*FCS analysis:*

Autocorrelation curves were fitted using the software package provided by ISS, giving estimates on the average number of emitters in the detection volume ( $N$ ) and the average translational diffusion time constant ( $\tau$ ). Single-emitter brightness (SEB) was computed based on:

$$\text{SEB} = \text{Average photon count rate (kHz)} / N$$

The “activated” NC probe concentration was derived from the calibration curve established by the Atto647N probe (**Figure S27**) and Atto532N probe (**Figure S28**). NCB chemical yield was then computed as:

NCB chemical yield = (Activated NC probe concentration in nM) × (dilution constant) / 15,000 where 15,000 nM represents the DNA concentration in the original reaction (i.e., 15 μM).

*Absorption measurements:*

Absorption spectra of NCBs were measured using Cary 5000 UV-Vis-NIR spectrometer from Agilent. 500 μl of 15 μM NCB solution was prepared following aforementioned protocol and was injected into a 700 μl Micro Fluorescence Cuvette from Thorlabs. The dual-beam mode was used with baseline/zero correction. All absorption spectra were measured from 300 nm to 800 nm with slit size of 2 nm (**Figure S29**). The acquired data were processed and analyzed using Python scripts.

*Image alignment algorithm:*

A custom bioinformatics and imaging processing pipeline named CHAMP (Chip-Hybridized Associated Mapping Platform) was developed by Finkelstein’s group and the detailed algorithm description can be found in ref. 8. CHAMP helped decipher the activator sequence behind each activated NCB spot (termed the NCB-CHAMP selection method, **Figure 1C** and **Figure S39**). In brief, mapping the alignment markers was done at four stages. First, a rough alignment was carried out using Fourier-based cross correlation, followed by a precision alignment using least-squares constellation mapping between FASTQ and *de novo* extracted NCB spots. We built up the consensus sequences and their corresponding information (e.g., lane number, tile number, and x-y coordinates) at all reported positions in the FASTQ file using the *map* command. Second, the *init* command was executed to record the metadata of imaging settings (e.g., rotation and scaling). Third, the *h5* command was applied to generate a single hdf5 file containing all 512×512 PhiX fiducial marker images. Fourth, the *align* command transformed the processed sequence information into pseudo-images and performed precise alignment. The output files were saved individually by image positions. The content included x, y coordinates of each sequence and the corresponding sequence ID. To analyze our NCB images, we developed an additional function named *ncb*, which corrected the uneven illumination using flat-field correction. A bootstrap method was then performed to derive the median intensity of each activator (baseline corrected) in order to rank the NCB brightness (**Figure S39**).

*Feature extraction/selection and machine learning model establishment/validation:*

The feature extraction was performed using MERCI<sup>[2]</sup> and was targeted library\_1 sequences only. For all extraction processes, both positive and negative thresholds were set to 5%, the maximal motif length was set to 6 bases, and the maximal number of wildcard nucleotide (A, T, G, C, or nothing) was set to 1 base. For example, in 5-fold cross-validation, the threshold was set to be 144 (5% of 2,880 sequences for each class). Separately, to extract “bright” features, the entire bright and dark classes were used (3,600 sequences for each class) and 180 was set as the threshold. The dark feature extraction was performed by simply swapping the bright and dark classes with the same parameter settings. The extracted motifs were then processed with Python scripts to include the position information. 339 bright yellow-orange, 567 dark yellow-orange features, 402 bright red and 1,164 dark red features were separately identified.

To decrease the chance of overfitting, we further narrowed down to a set of the most discriminative features with 61 bright yellow-orange, 121 dark yellow-orange, 103 bright red and 112 dark red features using Weka<sup>[3]</sup> – a process we termed feature selection. The attribute evaluator was set to “CfsSubsetEval”<sup>[4]</sup> and the search method was set to “GreedyStepwise”.<sup>[3]</sup> CFS scored a feature subset based on high correlation of features with predictive classes and low inter-correlation of features.<sup>[4]</sup> The “greedy” algorithm started with an empty set and iteratively added the feature that maximizes the gain in the CFS score. Feature selection process stopped when any additional feature decreased the CFS score. All other parameters were set to default values.

After feature extraction and selection, classification models were established based on various ML algorithms including logistic regression (LR), linear discriminant analysis (LDA), decision tree (DT), AdaBoost (ADA), and support vector machines (SVM), using the scikit-learn package in Python. 5-fold cross-validation was performed to evaluate the model performance. The best model (i.e., LR model) was employed to design bright and multicolor NCBs *in silico* (**Figure S32**).

*In silico design of bright NCBs:*

To design red and yellow-orange NCBs *in silico*, we again divided the 18-nt-long activator into 3 segments. Based on the most discriminative features identified by Weka, we sampled the distribution of these features in each segment and generated a list of common motifs with their corresponding positions. Please note that the features selected from each segment could slightly go beyond the range of that segment. We discovered that less than 3 features were favored for

the bright yellow activators, while 3 or more features were favored for the bright red activators (**Figure S40**). Consequently, to construct a red NCB candidate, we assigned 2 or 3 features to the blank 18-nt template for yellow and red candidates, respectively. The candidate design process started with feature\_1 insertion into segment\_1. As feature\_1 could go beyond segment\_1, feature\_2 might have an overlap with feature\_1 when being inserted into segment\_2. In that situation, the design algorithm would replace feature\_2 with another feature to ensure no overlap. However, if any two features shared identical bases at their overlapping site, they were considered as “compatible” and could be inserted into the same template. For example, as shown in **Figure S33**, feature C\_CTG (positions 1-5) and feature GGG\_GC (positions 5-10) shared a guanine base at the overlapping site (position 5). Consequently, they were compatible and were used in constructing a bright NCB candidate. The same procedure was repeated until a compatible feature for segment\_3 was found. Once all three features were inserted into the template, the remaining blank positions were filled up based on the composition popularity (at the same positions) from the bright class sequences. The edit distance<sup>[5]</sup> of the new candidate was then assessed. We only selected new candidates with edit distance between 3 to 5 from the top 200 bright activators screened on chip for test-tube investigation (**Table S10-S11**).

To mitigate the overfitting issue in our machine learning models, we varied the number of features in our models and evaluated the resulting test error and training error (**Figure S32**). After evaluating various feature numbers in the models, we found that the optimal number of features was close to the number of discriminative features selected by Weka (**Figure S32C**).

#### *NCB mobility evaluation in native PAGE gels:*

30  $\mu\text{M}$  of dark C55 probes were prepared following the previously stated protocol. The same molar ratio of [NC probe]: [ $\text{Ag}^+$ ]: [ $\text{NaBH}_4$ ] = 1:12:6 was used while doubling the amount of each chemical. 30  $\mu\text{M}$  NCB sample was prepared by adding 3  $\mu\text{l}$  of 1.2 mM activator to the 120  $\mu\text{l}$ , 30  $\mu\text{M}$  C55 probe solution, followed by the aforementioned hybridization and buffer exchange protocol.

Hand-cast native polyacrylamide gels were prepared by mixing 3.75 ml 40% acrylamide solution (cat. no. HC2040, Invitrogen), 0.75 ml TBE 1X buffer (Invitrogen), 3 ml DI water, 75  $\mu\text{l}$  10% w/w ammonium persulfate solution (cat. no. HC2005, Invitrogen), and 7.5  $\mu\text{l}$  TEMED (cat. no. 45-000-226, Fisher Scientific), targeting gel percentage at 20%. Gel solution was then poured onto the 1.0 mm empty gel cassette (cat. no. NC2010, Life Technologies). The cured gel was pre-run at 60V, 5 mA for 30 min in 1X TBE buffer before loading any samples. The DNA ladder was prepared by mixing 4  $\mu\text{l}$  of 100  $\mu\text{M}$  20/100 ladder (cat. no. 51-05-15-02,

Integrated DNA Technology) with 11  $\mu\text{l}$  DI water, 4  $\mu\text{l}$  5X loading dye (cat. no. LC6678, Invitrogen), and 1  $\mu\text{l}$  SYBR Gold nucleic acid gel stain (cat. no. S11494, Invitrogen), while NCB samples were prepared by mixing the 16  $\mu\text{l}$ , 30  $\mu\text{M}$  NCB solution with 4  $\mu\text{l}$  5X loading dye. After the gel was pre-run for 30 min, each lane was loaded with 20  $\mu\text{l}$  NCB samples or ladder and the gel was run at 50V, 5 mA for 800 min.

To evaluate the NCB purity after gel purification, the PAGE gel was stained by soaking the gel in SYBR Gold solution (2.5  $\mu\text{l}$  10,000 $\times$  SYBR Gold in 50 ml DI water, cat. no. S11494, Invitrogen). For better visualization of the post-staining results, NCBs were diluted to 5  $\mu\text{M}$  before mixing with the loading dye in the gel experiments (**Figure S30**). All gels were imaged on the Syngene gel imager (with 365 nm excitation). A 495 nm long-pass filter (cat. no. FGL495S, Thorlabs) was used to block the UV excitation from the imager.

#### *NCB elution from native PAGE gels:*

NCB gel bands were extracted using a gel band cutter and collected in a 1.5 ml Eppendorf tube, followed by smashing the gel into pieces with a plastic stick (**Figure S30**). For elution, 450  $\mu\text{l}$  of 20 mM SPB pH6.6 solution was added to each tube. The tube was shaken for 1 hr and stored at room temperature overnight. The suspension was then filtered using a micro-centrifugal filter (cat. no. F2517-5, Thermo Scientific) at 8,000 g for 20 min, followed by buffer exchange using the 0.5 ml centrifugal filter described above (cat. no. UFC503024, MilliporeSigma) and 10 mM ammonium acetate buffer (cat. no. AM9070G, Invitrogen).

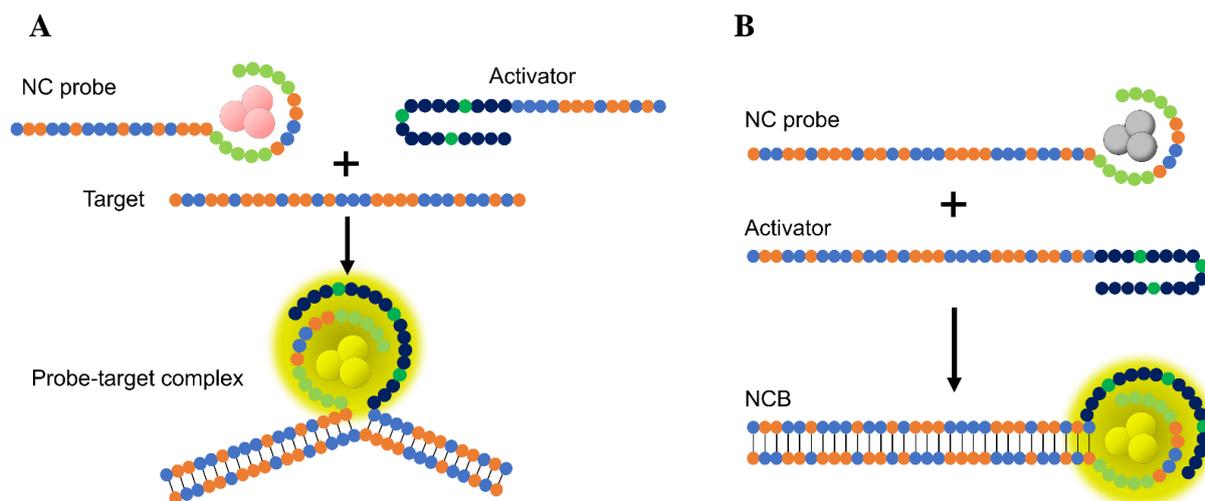
#### *Sample preparation for native mass spectrometry:*

Four gel-purified NCB samples (yPOT5, yPOT6, rPOT5 and rPOT6) were desalted and buffer exchanged into 10 mM ammonium acetate using a spin column (Micro Bio-Spin™ P-6 Gel Columns, Bio-Rad). Octylamine was added to aliquots of NCB solution at a concentration of 0.1% (v/v), to reduce the extensive metal cationic adduction that is commonly seen for ESI-MS analysis of oligonucleotides > 20 nt<sup>[6]</sup> (**Figure S31**).

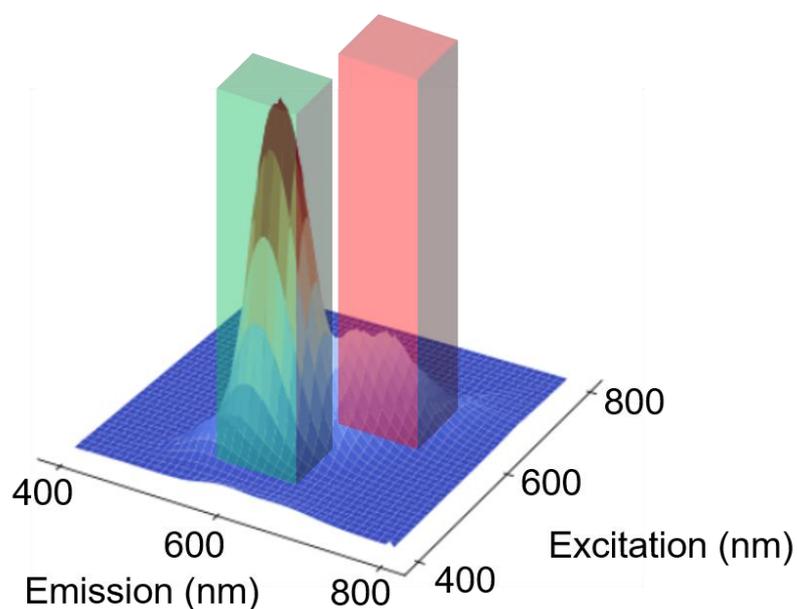
#### *Native mass spectrometry by direct infusion:*

3-5  $\mu\text{l}$  of 5  $\mu\text{M}$  purified yPOT5, yPOT6, rPOT5 and rPOT6 NCB solutions were loaded into Au/Pd-coated nanospray borosilicate static tips (prepared in-house) for nano electrospray ionization (nESI). All direct infusion experiments were performed on a Thermo Scientific Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer. A spray voltage of 0.65-0.8 kV and heated capillary temperature of 150°C were used to ionize and desolvate the NCB

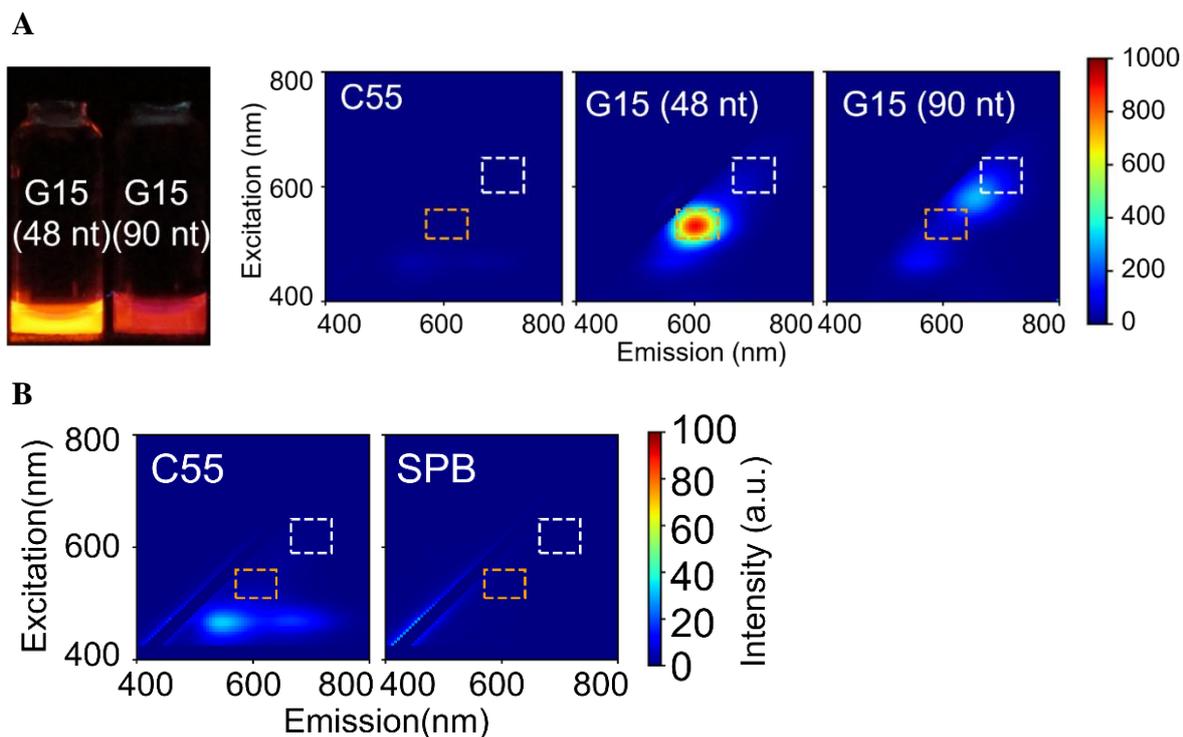
duplexes, facilitating their transmission into the gas-phase. In-source CID (100-130 eV) was also utilized to enhance transmission and reduce cationic adduction of the NCB complexes. MS1 spectra were collected at a resolution of 240K (@  $m/z$  200) and using 100 averages (**Figure S31**).



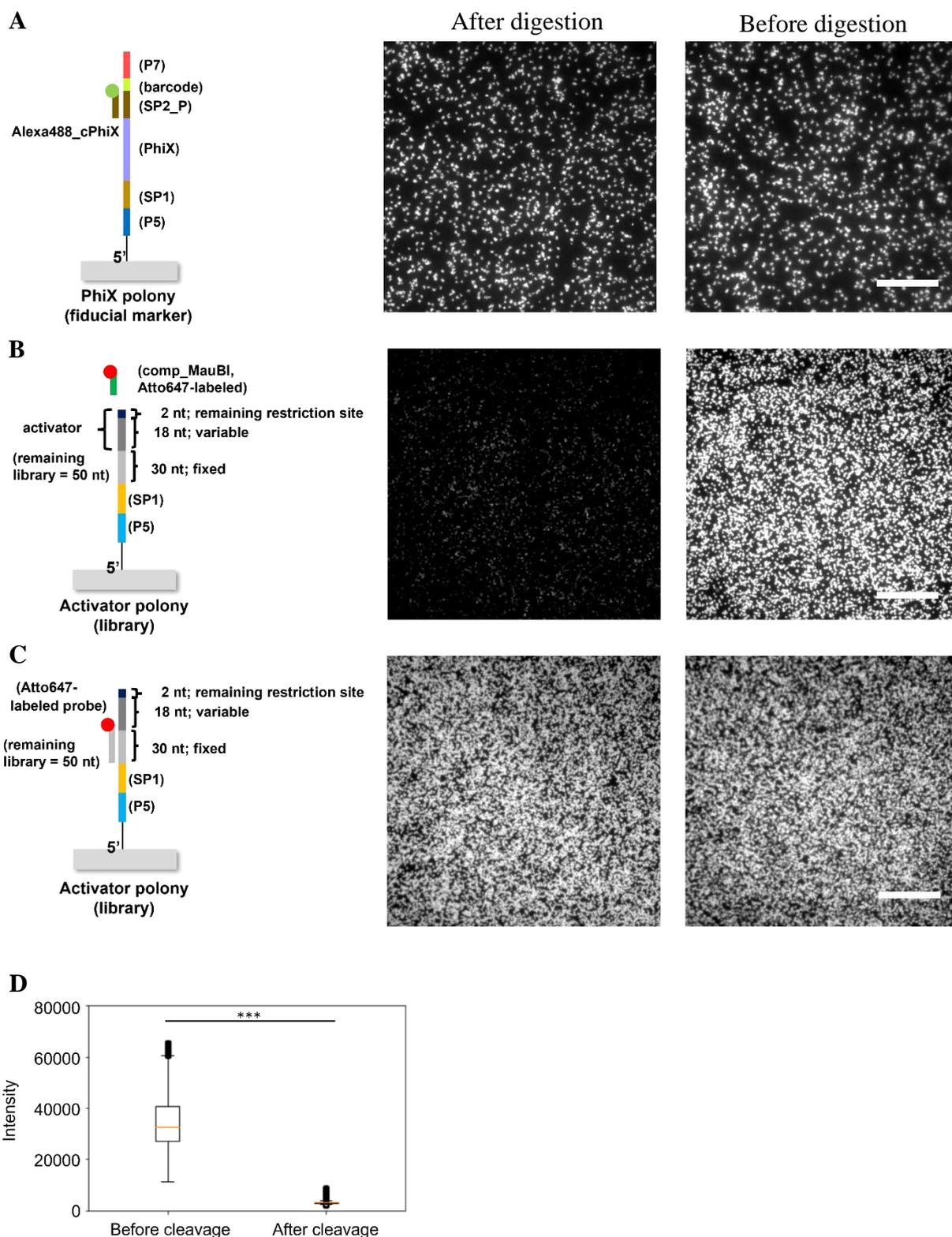
**Figure S1.** Schematic of NanoCluster Beacon (NCB) working principle. A) An NCB consists of an NC probe and an activator probe<sup>[7]</sup>. Upon binding to a target, the dark silver nanocluster (AgNC) interacts with the activator and lights up. NCBs remain dark when there is no target in the solution. B) In the study, we eliminated the target and only focused on the interactions between the NC probe and the activator.



**Figure S2.** Schematic of volumetric integrated intensity. From each 2D fluorescence spectrum, we can calculate the volumetric integrated intensities in the yellow-orange channel (Ex/Em: 535/50, 605/70 nm) and the red channel (Ex/Em: 620/60, 700/75 nm), respectively. The volumetric integrated intensity refers to the volumetric integral under the 2D spectrum surface.

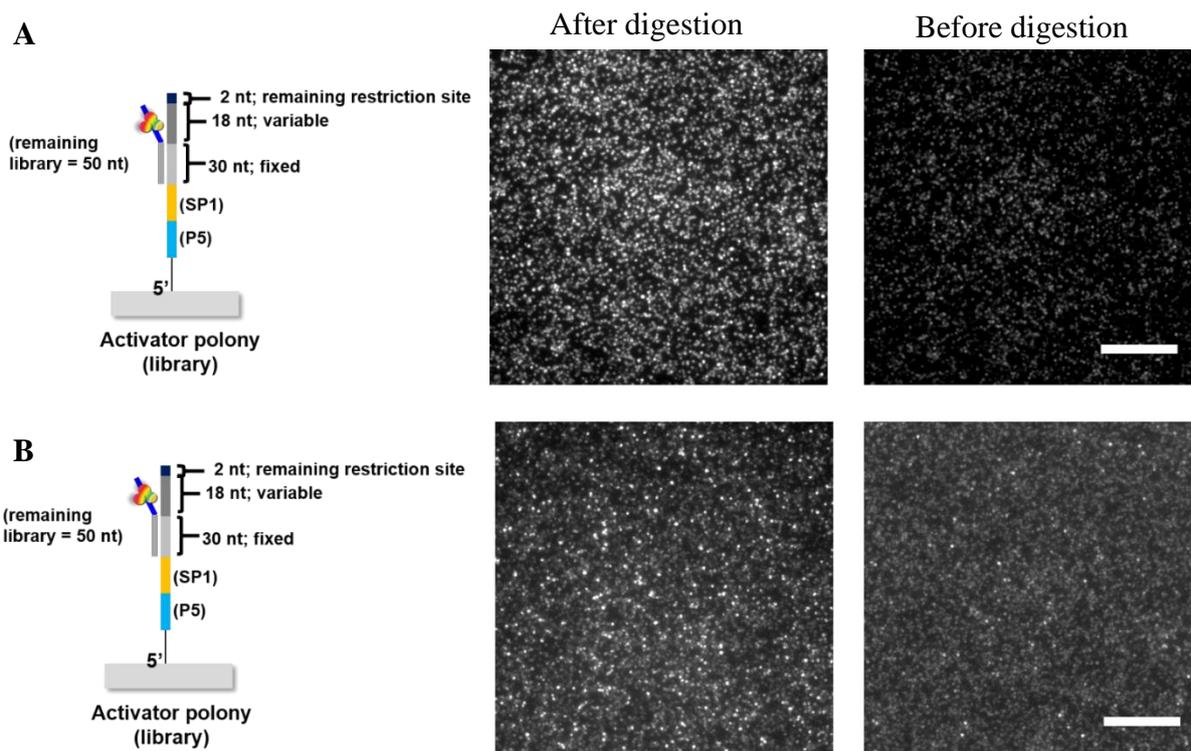


**Figure S3.** Illumination comparison of canonical NCBs with different activator length. A) We observed red-shifted fluorescence for NCBs having longer G15 activator (90-nt long) compared to canonical G15 activator (48-nt long). B) By adjusting the intensity range of the plot, the very low level of green-yellow fluorescence of C55 probes became visible. However, there was almost no fluorescence for both C55 probes and SPB in the yellow-orange and red windows (dashed boxes).

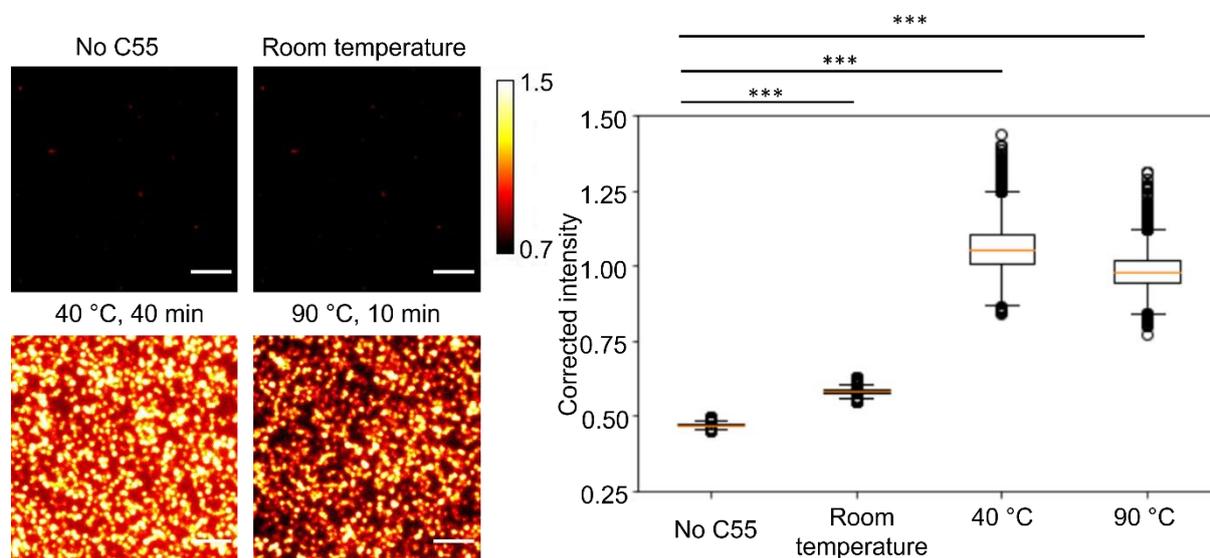


**Figure S4.** MiSeq chip images before and after restriction enzyme digestion. A) PhiX image is unaffected after cleavage. B) Overhang is successfully cleaved as minimal Atto647N-labeled comp\_SP2 probes can still bind with the activator polonies. C) The library sequences are unaffected after digestion. D) We averaged one row of Atto647N-tagged fluorescence images.

The median intensity dropped ~90% after cleavage. Box plot visualized five summary statistics: the median, two hinges (the lower and upper hinges corresponded to the first and third quartiles) and two whiskers. The upper whisker extended from the hinge to the largest value no further than  $1.5 \times \text{IQR}$  from the hinge, and the lower whisker extended from the hinge to the smallest value at most  $1.5 \times \text{IQR}$  of the hinge. The statistical significance relating to (D) panel was performed by Mann-Whitney U-test. Differences were ranked significant when \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001. Empty circles represented the outliers. (B) and (C) panels had the same contrast setting, while the contrast setting of (A) panel was different. Scale bar: 25  $\mu\text{m}$ .



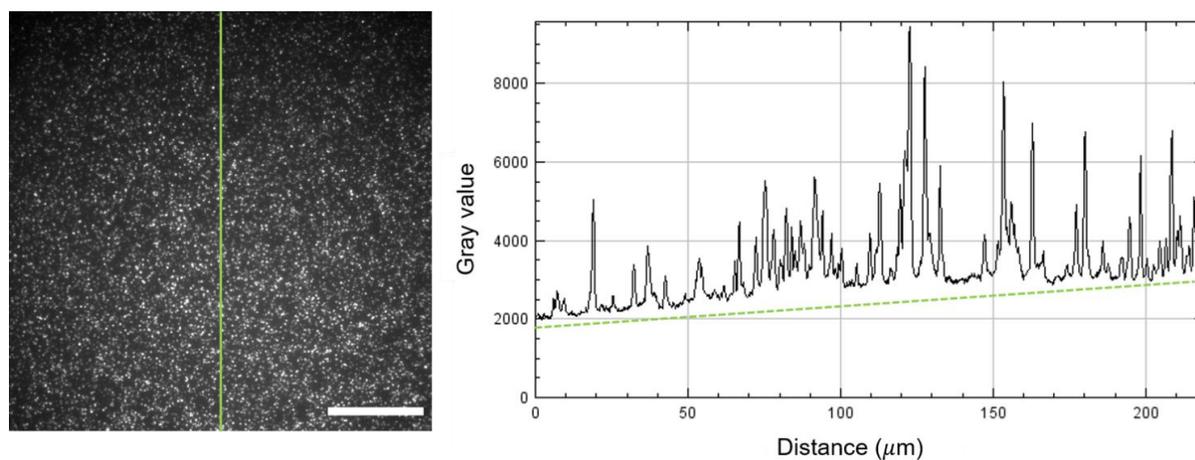
**Figure S5.** NCBs images before and after digestion. A) After restriction enzyme digestion, stronger NCBs signals are observed in the red channel. B) Stronger NCBs signals are observed in the yellow-orange channel. (A) and (B) panels had the same contrast setting. Scale bar: 25  $\mu\text{m}$ .



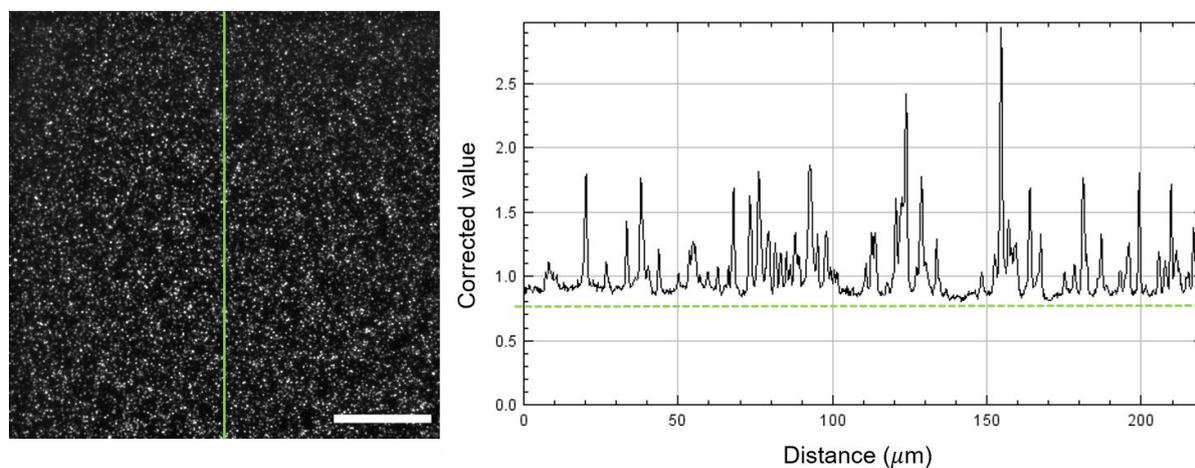
**Figure S6.** Intensity comparison of NCBs on the chip under different hybridizing temperature. We tested the NCBs intensity on the chip with different hybridizing temperature. We found that 40°C condition gave the brightest mean intensity compared to 90°C condition, which was the condition similar as test-tube validation. Furthermore, as 40°C gave a more moderate condition to the delicate *MiSeq* chip, we applied 40°C condition to the chip

experiments throughout this report. Box plot visualized five summary statistics: the median, two hinges (the lower and upper hinges corresponded to the first and third quartiles) and two whiskers. The upper whisker extended from the hinge to the largest value no further than  $1.5 \times$  IQR from the hinge, and the lower whisker extended from the hinge to the smallest value at most  $1.5 \times$  IQR of the hinge. Empty circles represented the outliers. The statistical significance here was performed by Mann-Whitney U-test. Differences were ranked significant when \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001.

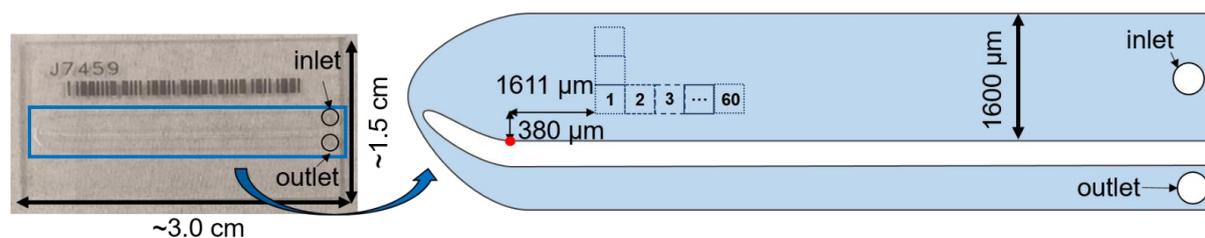
## A. Before pseudo-flat field correction



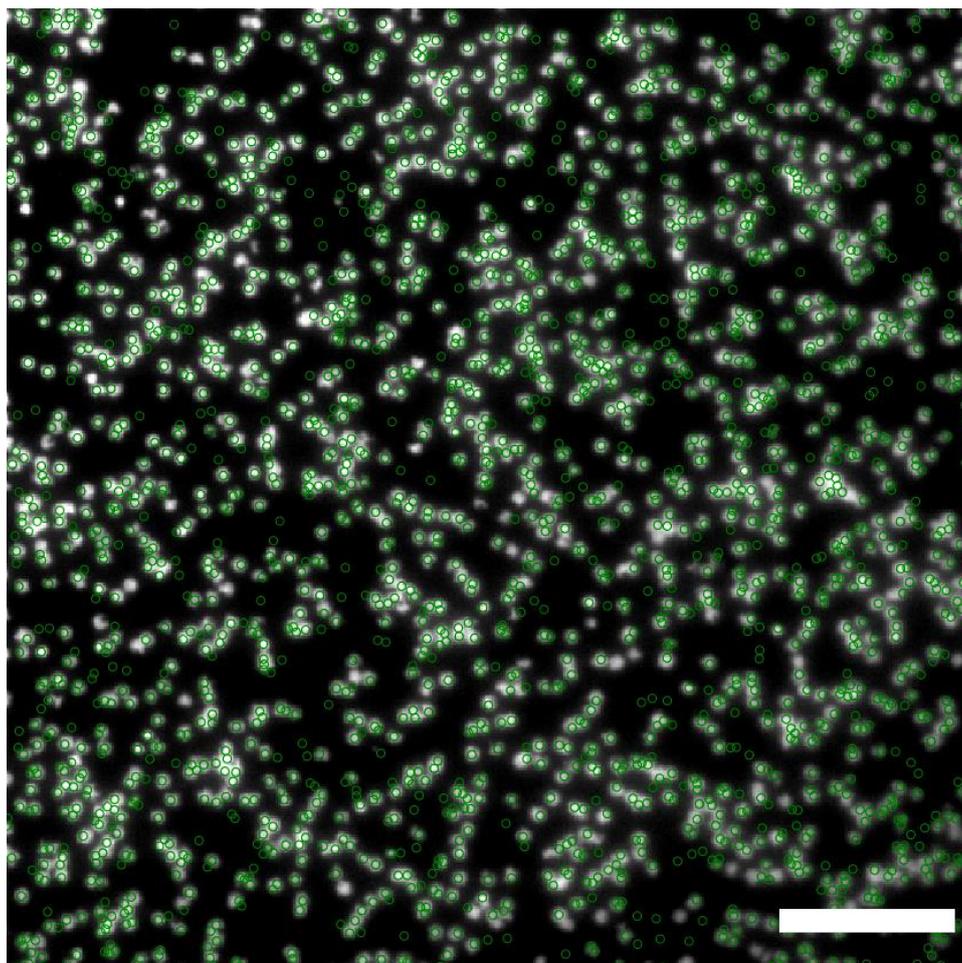
## B. After pseudo-flat field correction



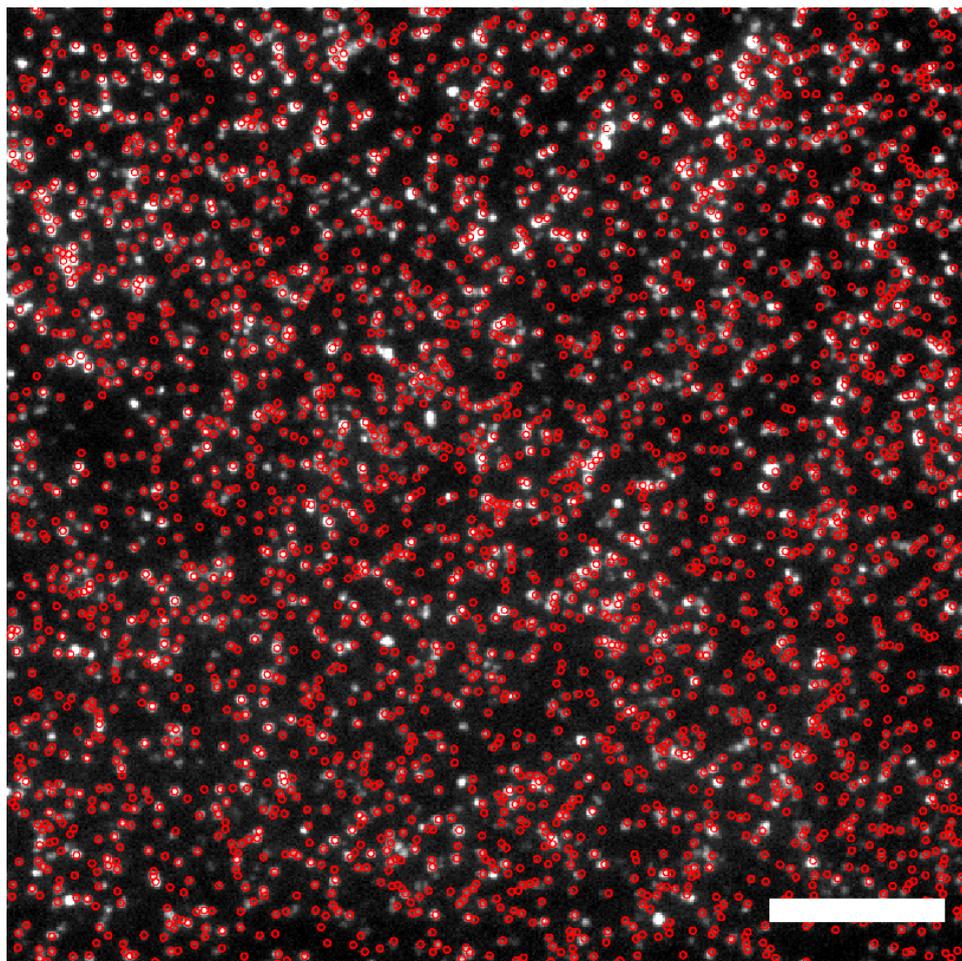
## C. Schematic of a MiSeq chip



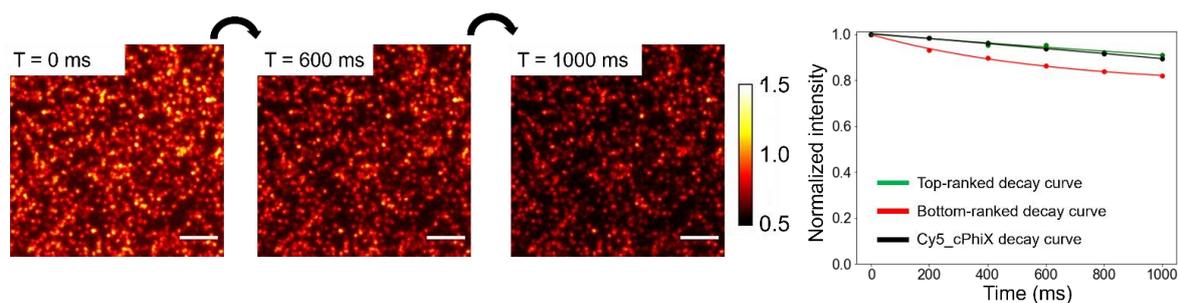
**Figure S7.** *MiSeq* chip images before and after flat-field correction. A) A representative field-of-view (FOV, 1,024×1,024 pixel) shows uneven illumination that leads to inconsistent intensity baseline across the FOV. B) The same FOV is corrected using a flat-field correction method. The intensity baseline is uniform throughout the FOV after correction. Distance: from top to bottom. Scale bar: 50 μm. C) A *MiSeq* chip is 1.5 cm long and 3.0 cm wide (left). To bypass most of these unregistered regions, we shifted the imaging starting position by 380 μm vertically and 1,611 μm horizontally with respect to the reference point (red dot) at the bottom left corner (right).



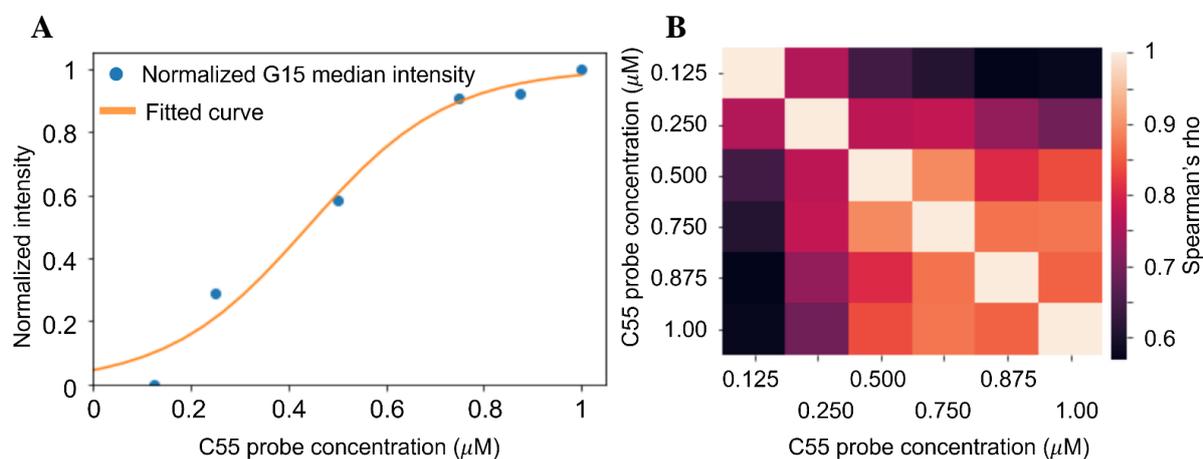
**Figure S8.** Alignment of PhiX fiducial markers. A representative FOV (512×512 pixel; 110×110  $\mu\text{m}^2$ ) of aligned PhiX fiducial markers acquired under the green channel (EX/EM: 480/40, 535/50 nm). The PhiX fiducial markers are labeled with Alexa488 and the registered PhiX positions from the fastq data file are circled in green. Scale bar: 20  $\mu\text{m}$ .



**Figure S9.** Identification of activator sequences from the fastq data. A representative FOV ( $512 \times 512$  pixel;  $110 \times 110 \mu\text{m}^2$ ) of aligned red NCBs polonies acquired under the red channel (EX/EM: 620/60, 700/75). The library sequences (i.e., activators) are hybridized with the common NC probe (i.e., C55) and form activated NCBs on the chip. The registered activator positions from the fastq data file are circled in red. These two examples (**Figure S8-S9**) demonstrate the accuracy of the NCB-CHAMP<sup>[8]</sup> mapping algorithm. Scale bar:  $20 \mu\text{m}$ .

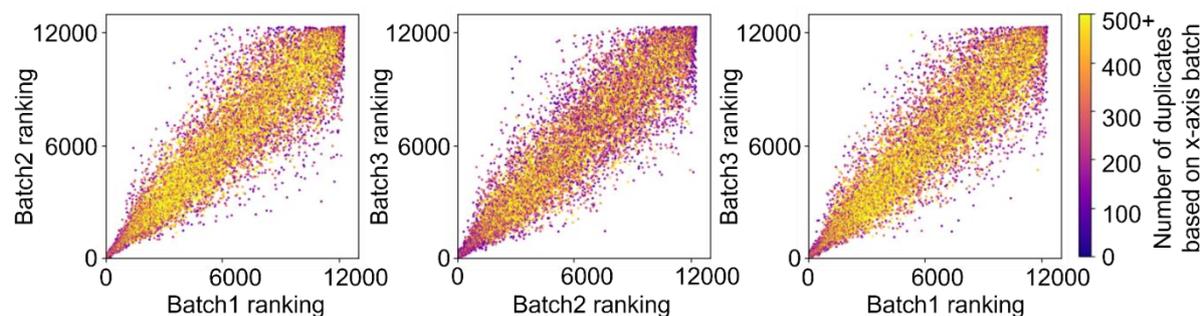


**Figure S10.** NCB intensity decay on the chip. Photobleaching on the MiSeq chip (2-7% decrement after 200 ms exposure) did not prevent us from correctly ranking the brightness of NCBs, as the excitation power ( $\sim 10 \text{ W/cm}^2$ ), the exposure time (200 ms), and the scanning step size ( $220 \mu\text{m}$  per step) were all precisely controlled by our auto-scan algorithm and shutter control. Scale bar:  $10 \mu\text{m}$ .

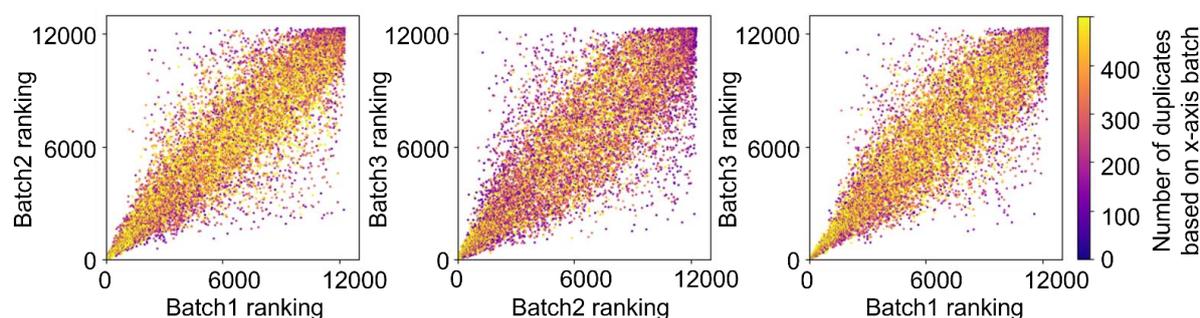
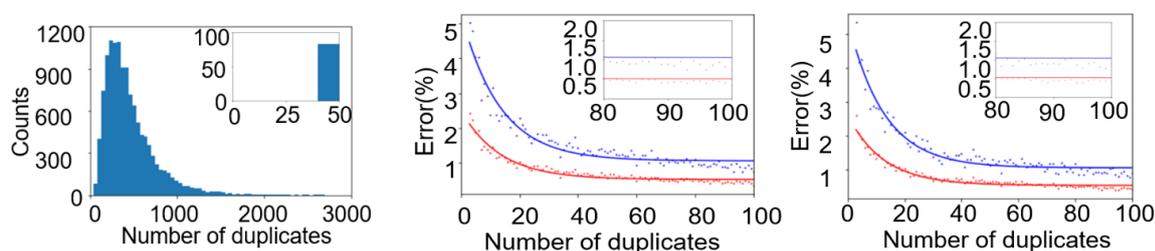


**Figure S11.** Titration curve of G15 NCB intensity on the chip. A) To find out the optimal C55 concentration for NCB screening on *MiSeq* chip, we used the G15 NCB intensity as the calibration standard in a titration experiment. C55 probes at 6 different concentrations were delivered to the chip. The normalized G15 NCB median intensity reached a plateau when the C55 probe concentration was about  $0.8 \mu\text{M}$ . B) In fact, we observed highly consistent ranking results when the C55 probe concentration was higher than  $0.5 \mu\text{M}$ .

## A. Library\_1 (red):



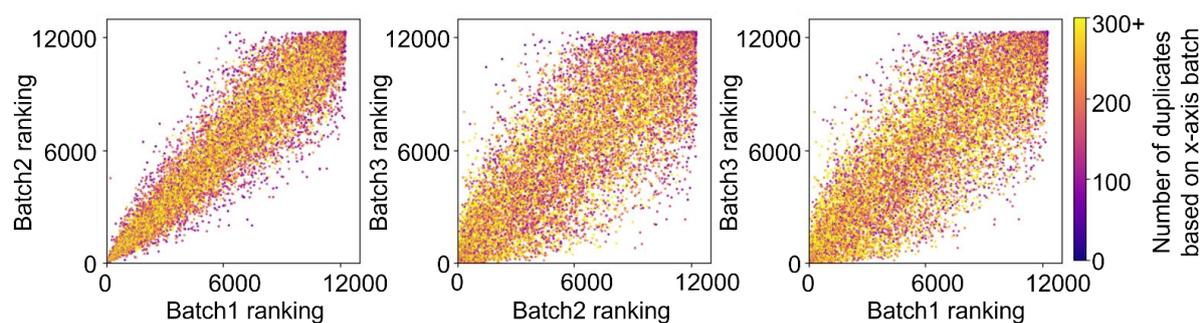
## B. Library\_1 (yellow-orange):

C. Statistics of *MiSeq* chip results for Library\_1

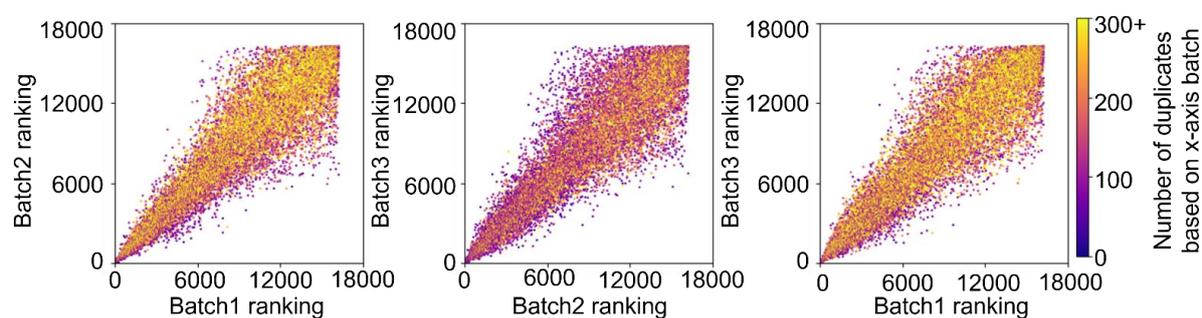
**Figure S12.** Batch to batch variations in *MiSeq* chip selection for library\_1. Nonparametric measure is applied to evaluate the ranking correlation among repeated experiments. A) For library\_1, all 12,286 distinct sequences are found on chip. The Spearman's rho for the 3 comparisons in red channel are 0.92, 0.93, and 0.93, with the  $R^2$  of 0.85, 0.86 and 0.86 (left to right). B) The Spearman's rho for the 3 comparisons in yellow-orange channel are 0.86, 0.91, and 0.86, with the  $R^2$  of 0.75, 0.83 and 0.75 (left to right). C) (left) Distribution of the number of activator duplicates in library\_1. All activators had at least 20 duplicates observed. On average, each activator had  $457 \pm 308$  polonies on a *MiSeq* chip. (middle and right) Estimation of error in the NCB brightness characterization after bootstrapping. To improve the accuracy of our high-throughput screening, we performed 100 rounds of bootstrapping processes by random sampling 75% of observed duplicates intensity and assigned median intensity (baseline corrected; *Methods*) as the NCB on-chip intensity. Bootstrap intensity values were calculated

for the standard sequence (i.e., G15) with all numbers of clusters between 3 and 100. Shown are the average errors (red points) and 90% confidence intervals of error (blue points), using the median intensity with either 200 (middle) or 20,000 clusters (right) for 10,000 rounds as reference. Solid lines indicate a fit to the data.

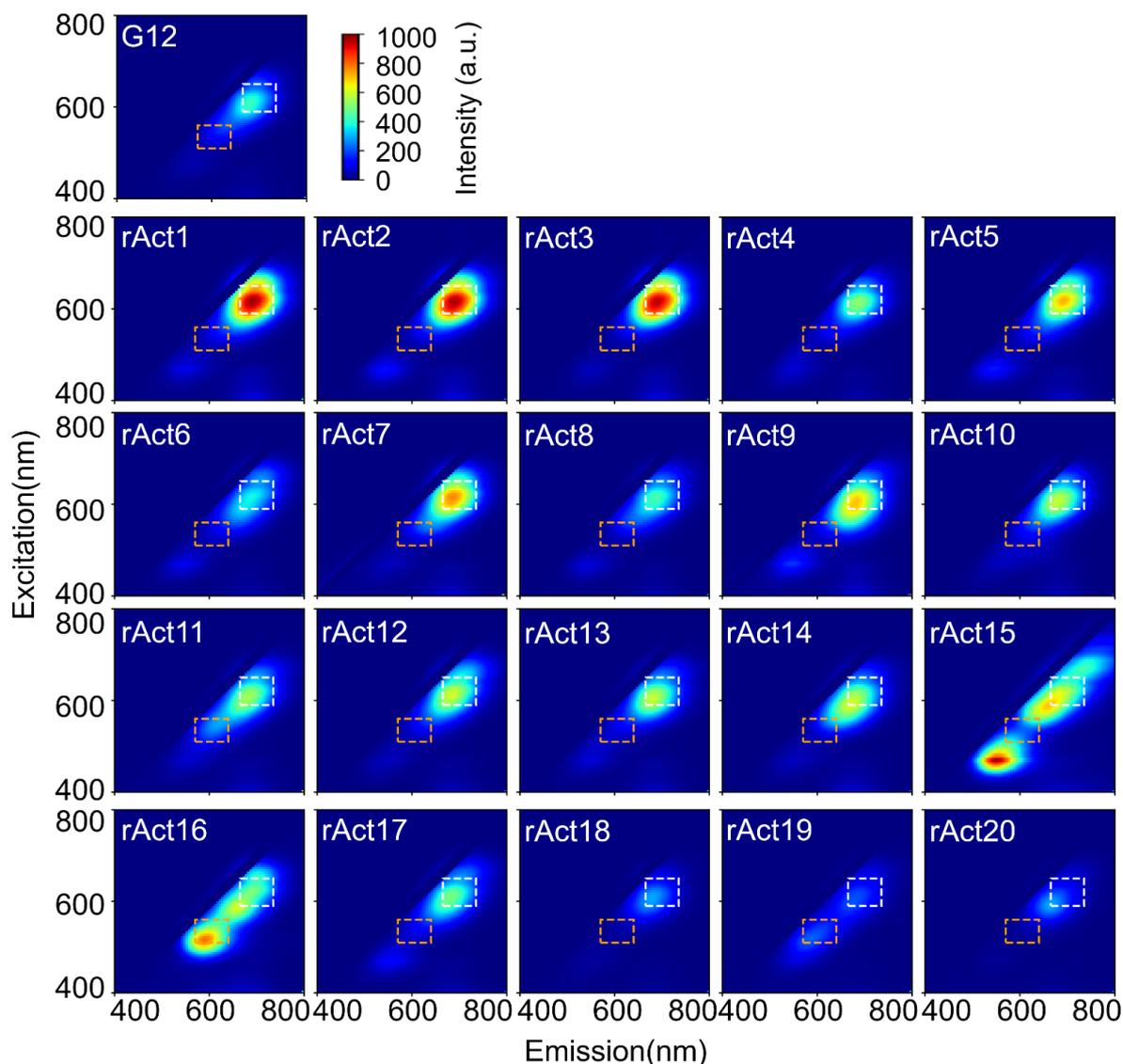
## A. Library\_2 (red):



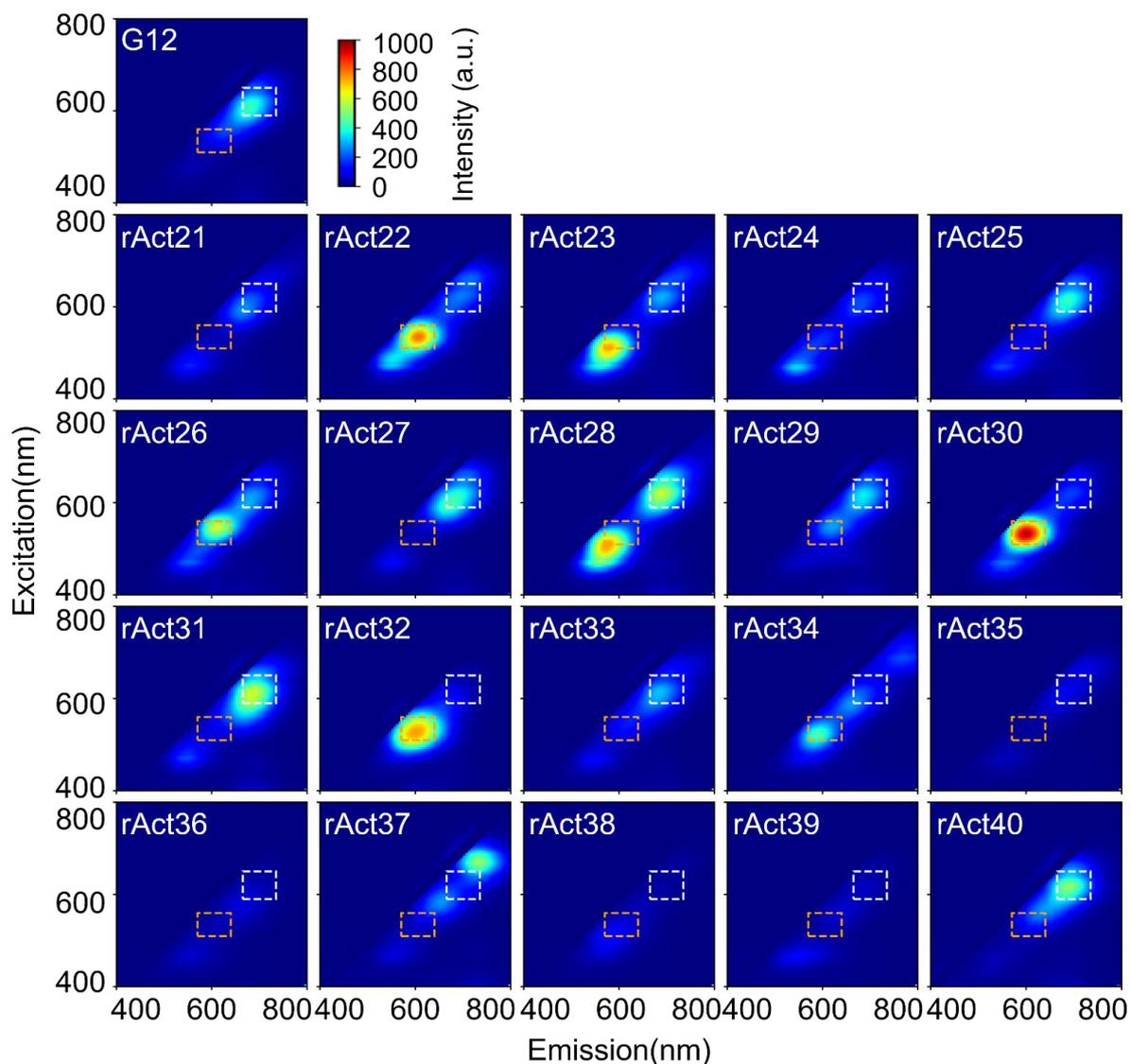
## B. Library\_3 (red):



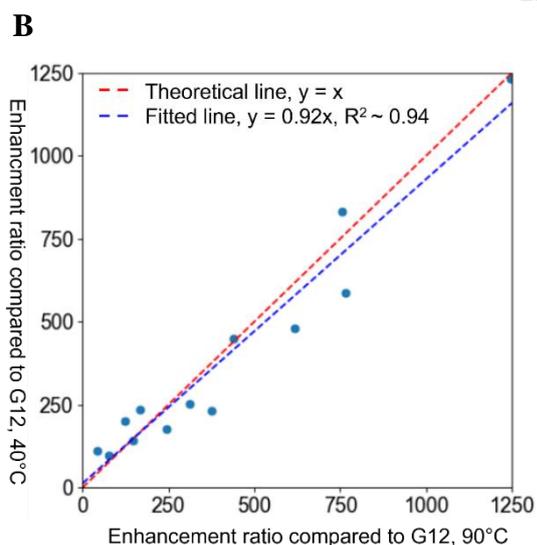
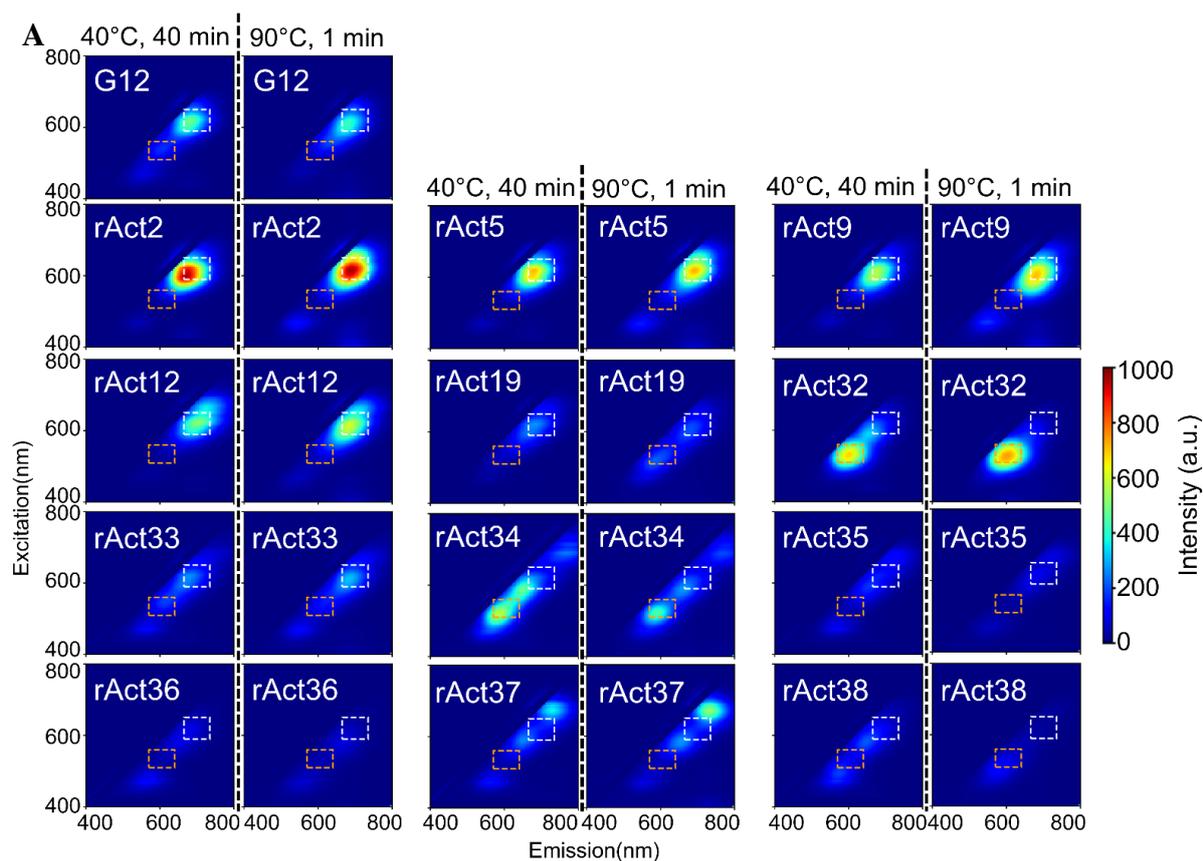
**Figure S13.** Batch to batch variations in *MiSeq* chip selection for library\_2 and library\_3. A) For library\_2, all 12,286 distinct sequences are found on chip. The Spearman's rho for the 3 comparisons in red channel are 0.93, 0.83 and 0.80 (left to right), with the  $R^2$  of 0.87, 0.68, and 0.65 (left to right). B) For library\_3, all 16,255 distinct sequences are found on chip. The Spearman's rho for the 3 comparisons in red channel are 0.91, 0.86 and 0.89 (left to right), with the  $R^2$  of 0.82, 0.73 and 0.79 (left to right).



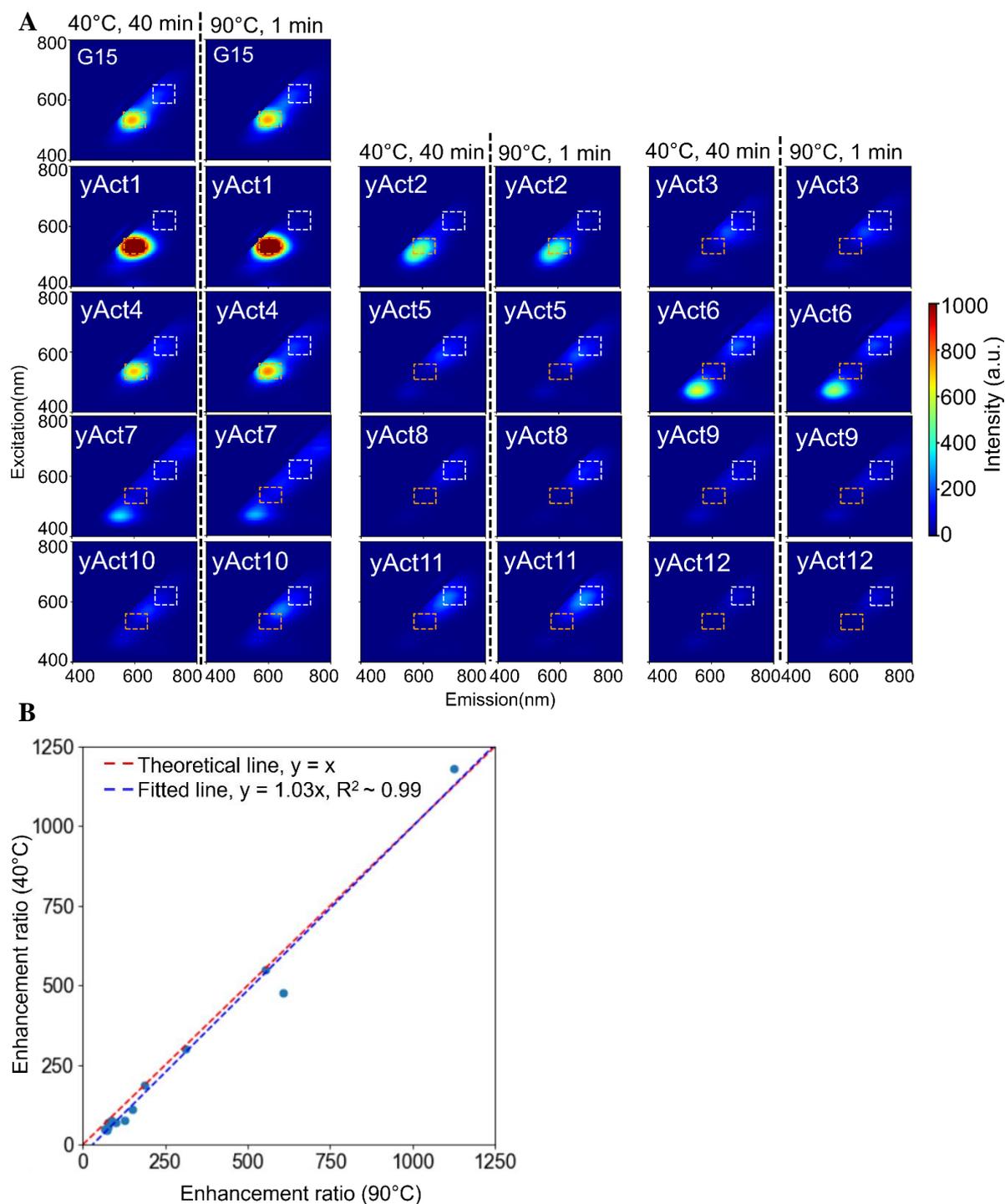
**Figure S14.** 2D spectra of red activator candidates were randomly selected from the 827 sequences brighter than G12 (including the 3 false positive selections: rACT18, rACT19 and rACT20). Compared to G12 NCB (ATCCGGGGTGGGGTGGGG), 17 out of 20 bright red activator candidates (selected by the chip screening method) have the improvement ratio greater than one (85% accuracy). In particular, rAct1 NCB (TCCATTGGTGGGGTGGGG) has the improvement ratio of 2.94. The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm). See **Table S2A** for details.



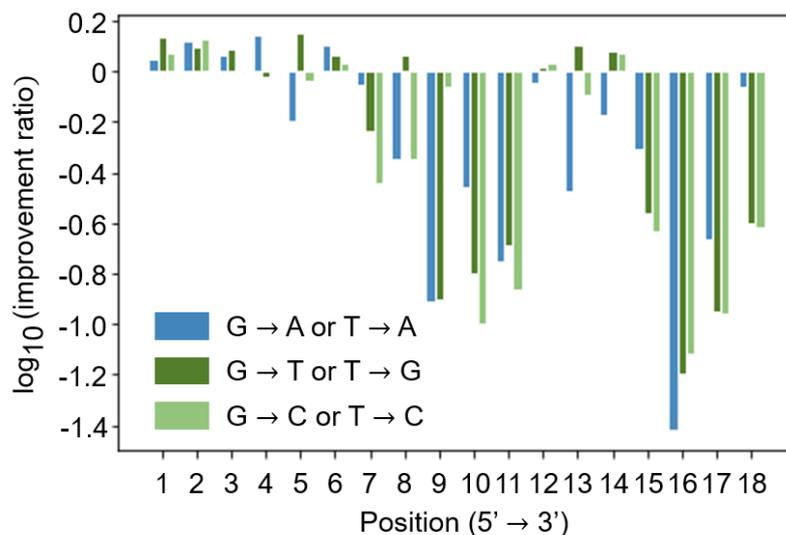
**Figure S15.** 2D spectra of dark red activator candidates were randomly selected from the 11,458 sequences darker than G12 (including 3 false negative selections: rACT28, rACT31 and rACT40). Compared to G12 NCB (ATCCGGGGTGGGGTGGGG), 17 out of 20 dark red candidates (selected by the chip screening method) have the improvement ratio less than one (85% accuracy). In particular, rAct38 NCB (GGGTGGGTTTATGTGGGG) has the improvement ratio of 0.10. The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm). See **Table S2B** for details.



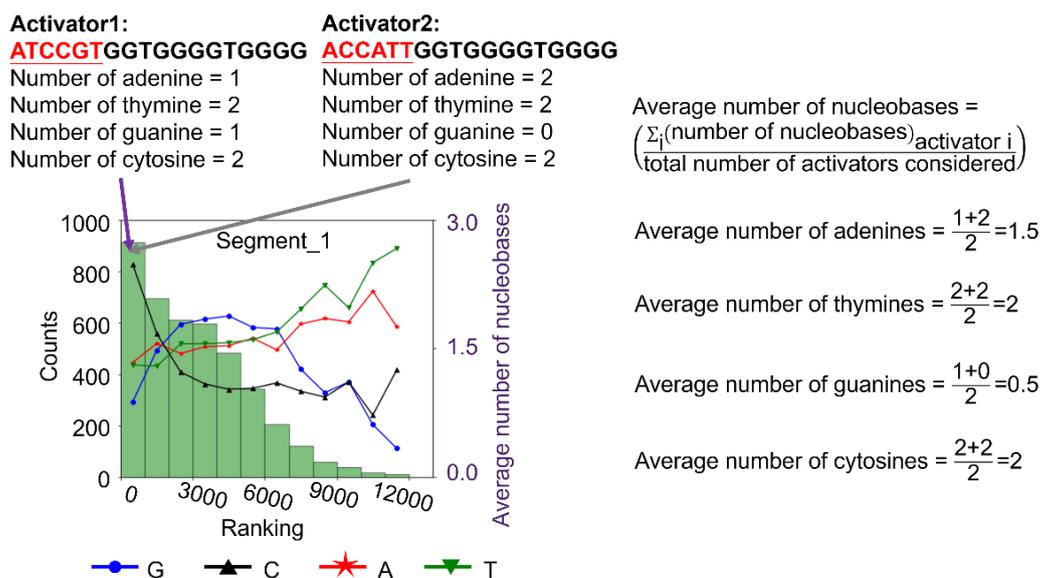
**Figure S16.** Comparison between two hybridization conditions of red activator candidates. Twelve red activator candidates from **Table S2** were selected and performed in-solution validation by heating at 40°C/40 minutes. A) When comparing the two hybridization conditions in test tubes (40°C/40 minutes vs. 90°C/1 minute then gradually cools down to room temperature), we found their resulting fluorescence activation on NCBs indistinguishable. B) We computed the enhancement ratio of the samples under two conditions. The result was consistent and well correlated.



**Figure S17.** Comparison between two hybridization conditions of yellow-orange activator candidates. Twelve yellow-orange activator candidates from **Table S4** were selected and performed in-solution validation by heating at 40°C/40 minutes. A) When comparing the two hybridization conditions in test tubes (40°C/40 minutes vs. 90°C/1 minute then gradually cools down to room temperature), we found their resulting fluorescence activation on NCBs indistinguishable. B) We computed the enhancement ratio of the samples under two conditions. The result was consistent and well correlated.

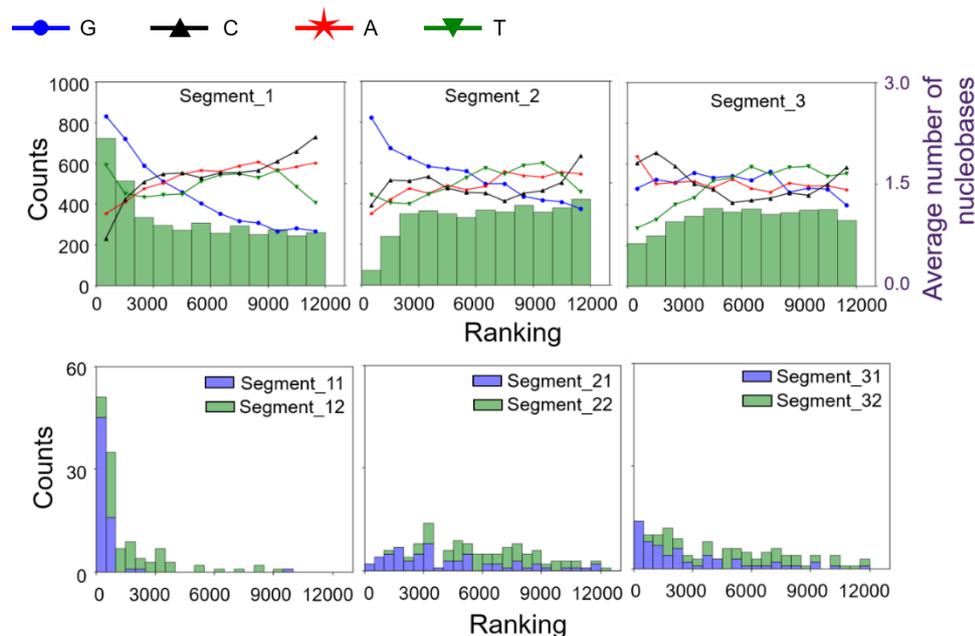


**Figure S18.** Small-scale test-tube investigation of single nucleotide variants of G15 in the yellow-orange channel. The improvement ratios of 54 single nucleotide variants of G15 NCBs are put into this base-10 logarithm chart. By substituting G to T at position 5 (i.e., GGGT**T**GGGTGGGGTGGGG), the largest improvement in the enhancement ratio is observed, which is only 1.41-fold higher than the enhancement ratio of G15 NCB in the yellow-orange channel. This result demonstrates that a small-scale investigation cannot improve the brightness of an existing NCB by more than 2-fold. Interestingly, by substituting G to A at position 16 (i.e., GGGTGGGGTGGGGT**A**GG), we observed a pair of POTs with POT difference ratio ~25.

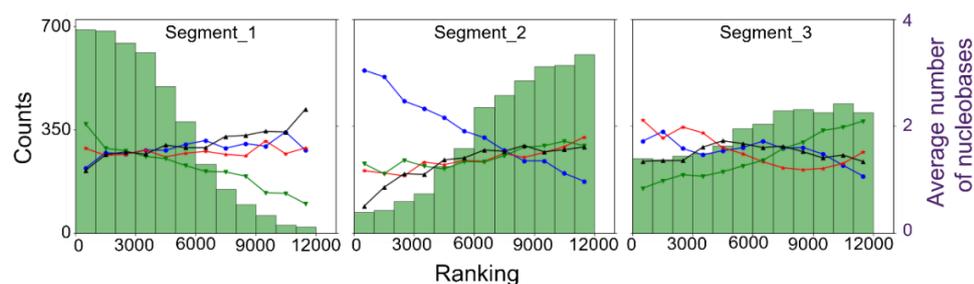


**Figure S19.** Detailed explanation of Figure 2b histogram. Here we demonstrated calculation of the average number of nucleobases from two top-ranked sequences. For instance, Activator1 (Left, ATCCGT GGTGGG GTGGGG, Rank12) has 1 adenine, 2 thymine, 1 guanine, and 2 cytosine, while the Activator2 (Right, ATCCGT GGTGGG GTGGGG, Rank14) has 2 adenine, 2 thymine, 0 guanine, and 2 cytosine. Both activators were highly ranked (i.e., they create bright NCBs) and located in the first bin (ranking 1-1,000) of the histogram. Each histogram contained 4,096 sequences.

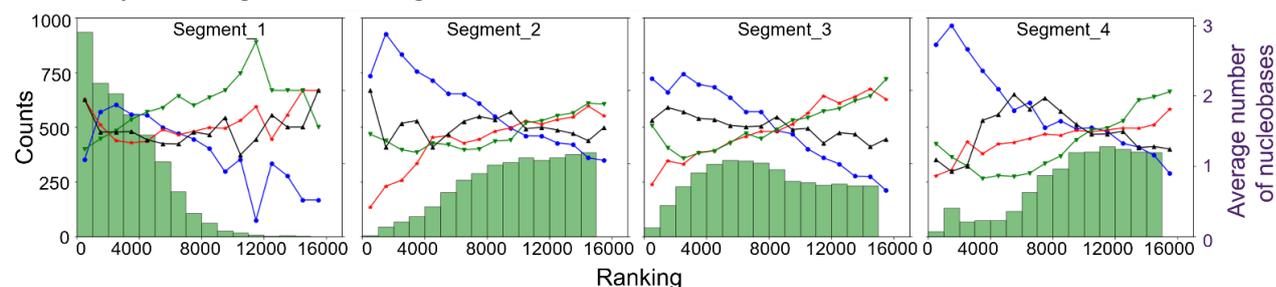
## A. Library\_1, 3- and 6-segment interrogation on yellow-orange NCBs



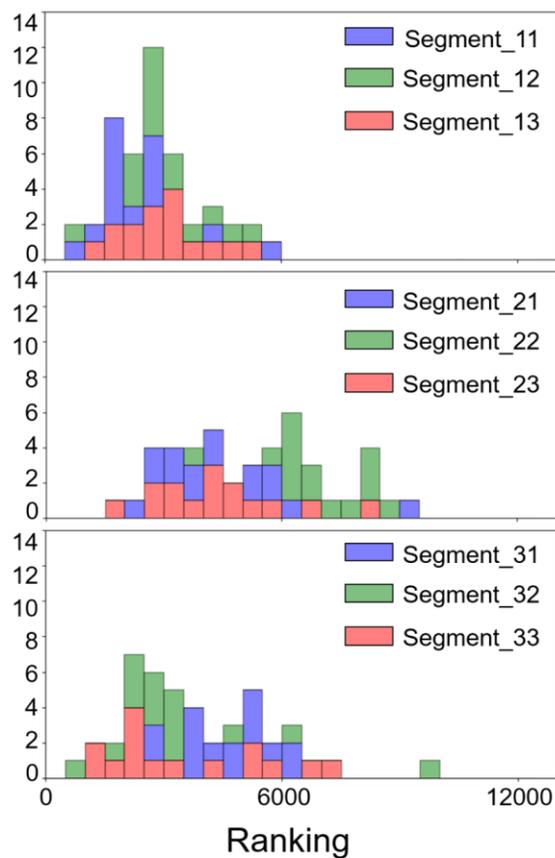
## B. Library\_2, 3-segment interrogation on red NCBs



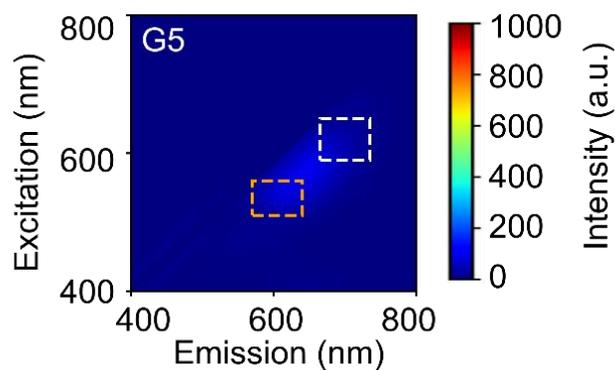
## C. Library\_3, 4-segment interrogation on red NCBs



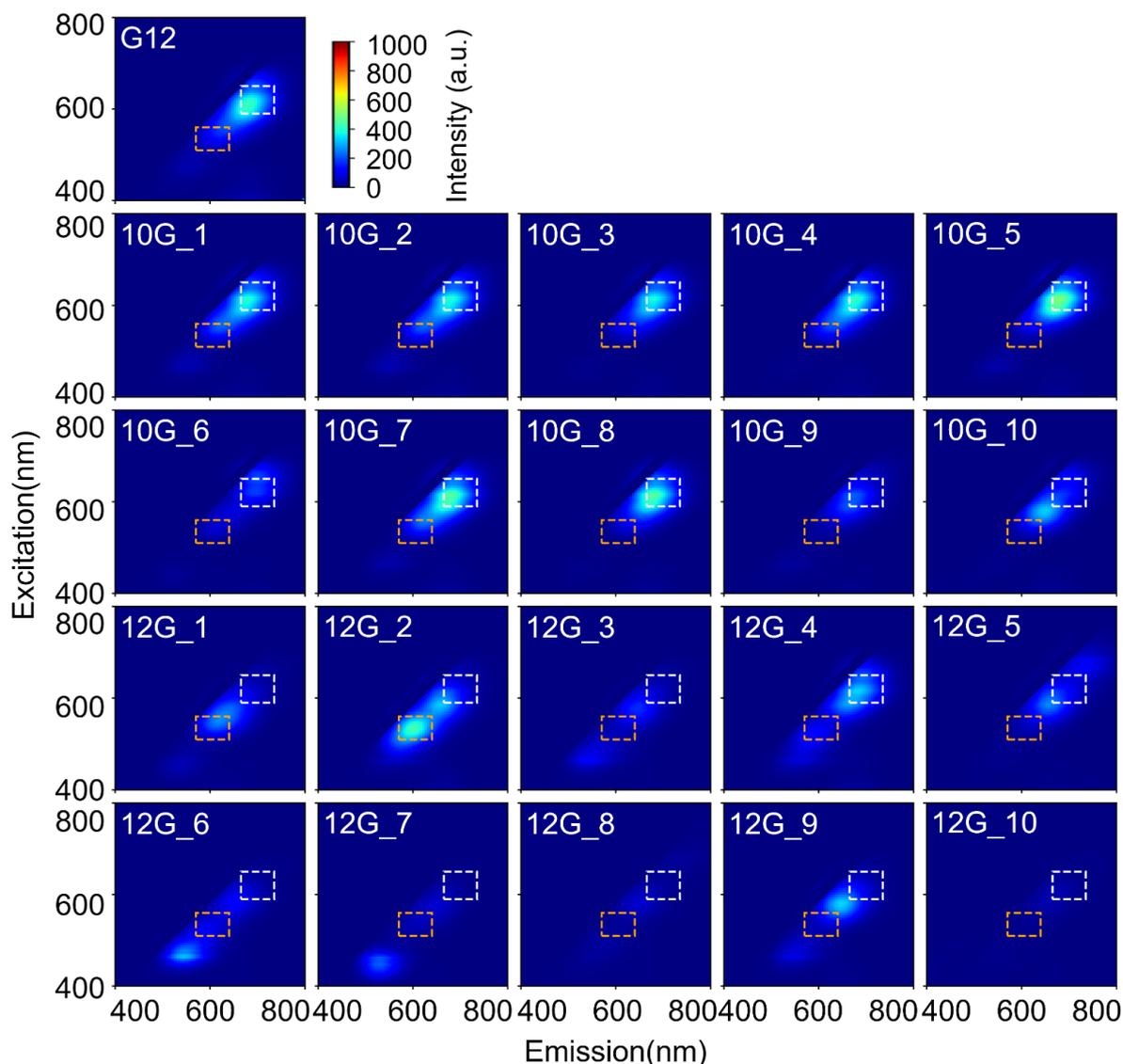
**Figure S20.** Influence of activator mutations on NCB brightness. A) Here we showed the influence of activator mutations on yellow-orange NCB brightness. Bottom row is stacked histograms. B) When randomizing the 3 segments in library\_2, positions 10 to 15 were found to be the interaction hot zone. The segment definition for all three libraries could be found in **Table S1C**. C) When randomizing the 4 segments in library\_3, positions 10 to 12 were found to be the interaction hot zone. The results from (B) and (C) were consistent with the library\_1 result (**Figure 2**), which showed positions 10-12 were the interaction hot zone for creating bright NCBs.



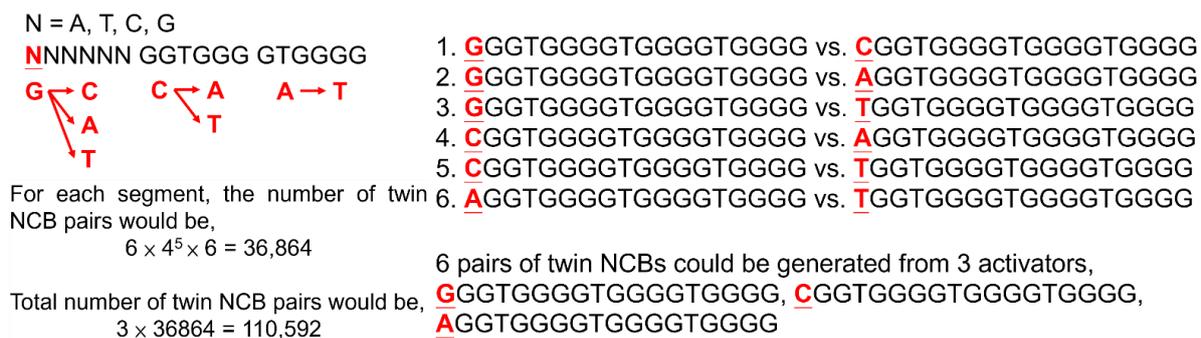
**Figure S21.** Nine-segment interrogation on the library\_1 for red NCB brightness. Here we further divide the library\_1 activator into 9 segments (**Table S1B**) and investigate each segment's influence on red NCBs brightness demonstrated by stacked histograms. In segment\_2, segment\_22 (positions 9-10) and segment\_23 (positions 11-12) are the critical zones as the ranking shifts toward the dark side when these segments are randomized.



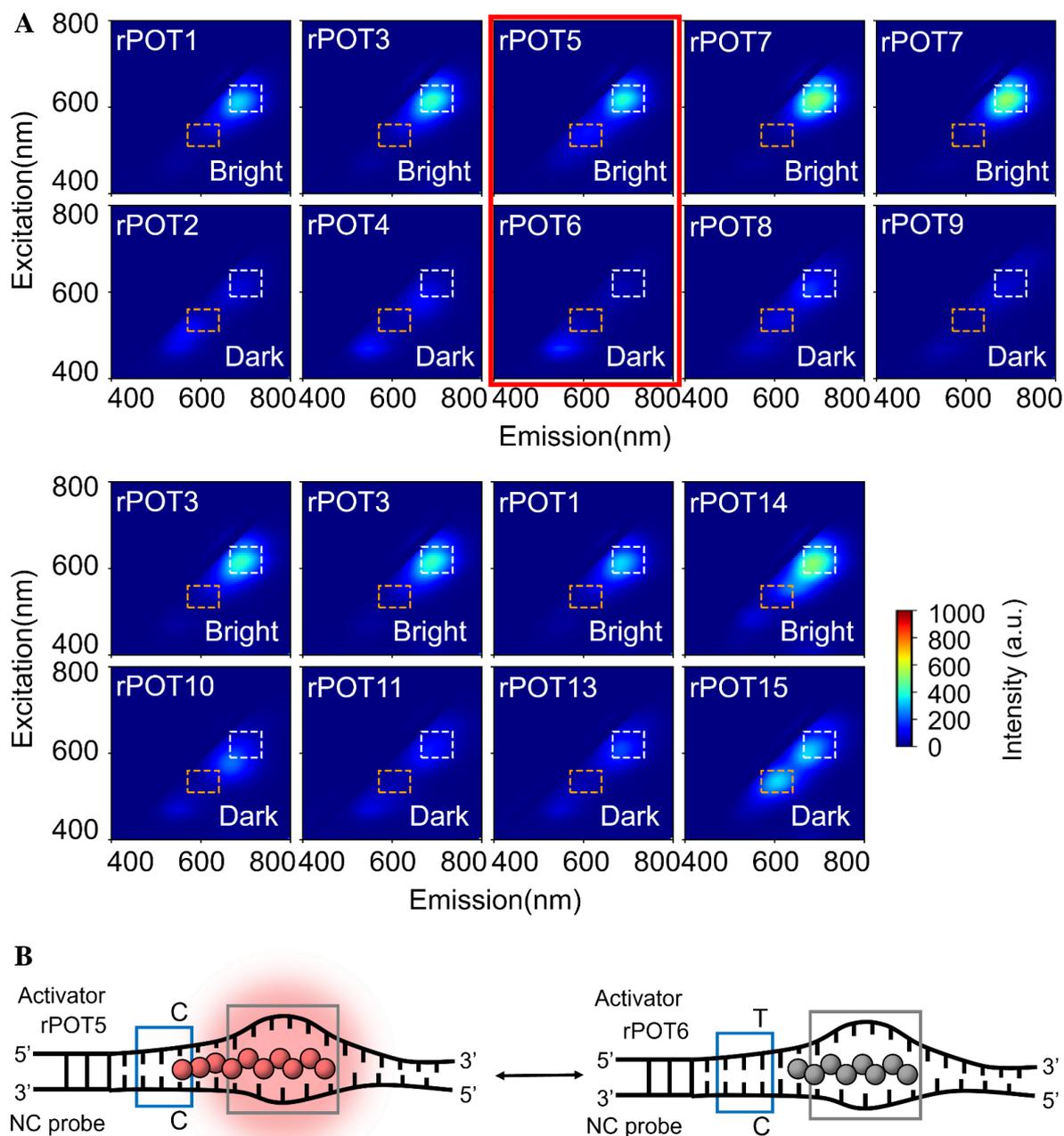
**Figure S22.** 2D spectra of G5 NCB. Based on the design rules discussed in **Figure 2**, we speculated that this G5 activator (CCCCCGCGGGGTTTCCC) would lead to a bright NCB. However, the result was actually a low red enhancement ratio (39, as compared to 439 for G12; **Table S9**). This result clearly indicated that segments do not work alone – cooperativities among the segments determine the activation color and intensity of an NCB.



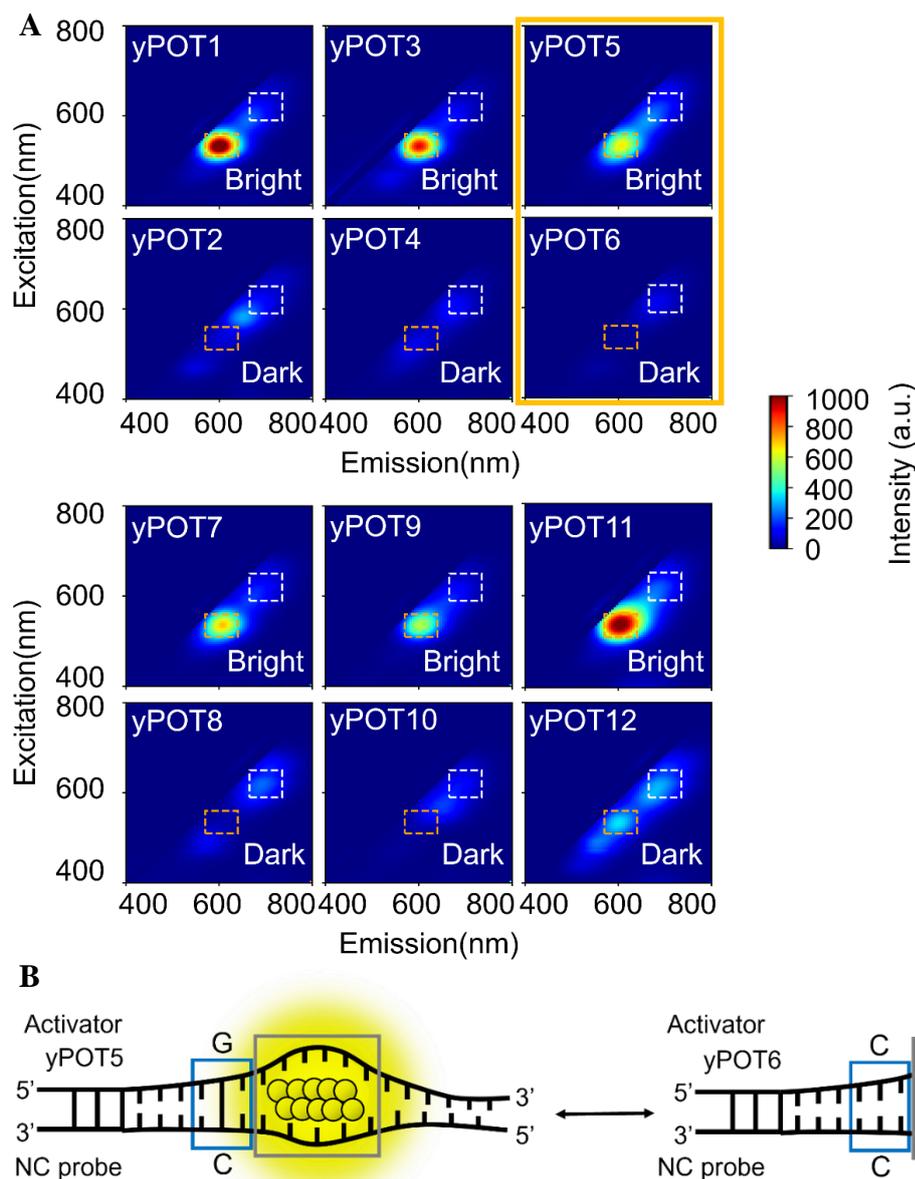
**Figure S23.** 2D spectra of activators with various numbers of guanine bases. Based on chip selection results, ten 10G activators can potentially be brighter than G12 NCB (**Table S5**) and ten 12G activators can potentially be darker than G12 NCB (**Table S6**). Test-tube investigation proves that 7 of the selected 10G activators have their enhancement ratios comparable to that of G12 in the red channel (improvement ratio  $\geq 0.9$ ), and all selected 12G activators are darker than G12 in the red channel (improvement ratio  $< 0.6$ ). This result indicates that it is possible to create bright red NCBs with fewer numbers of guanine. The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm).



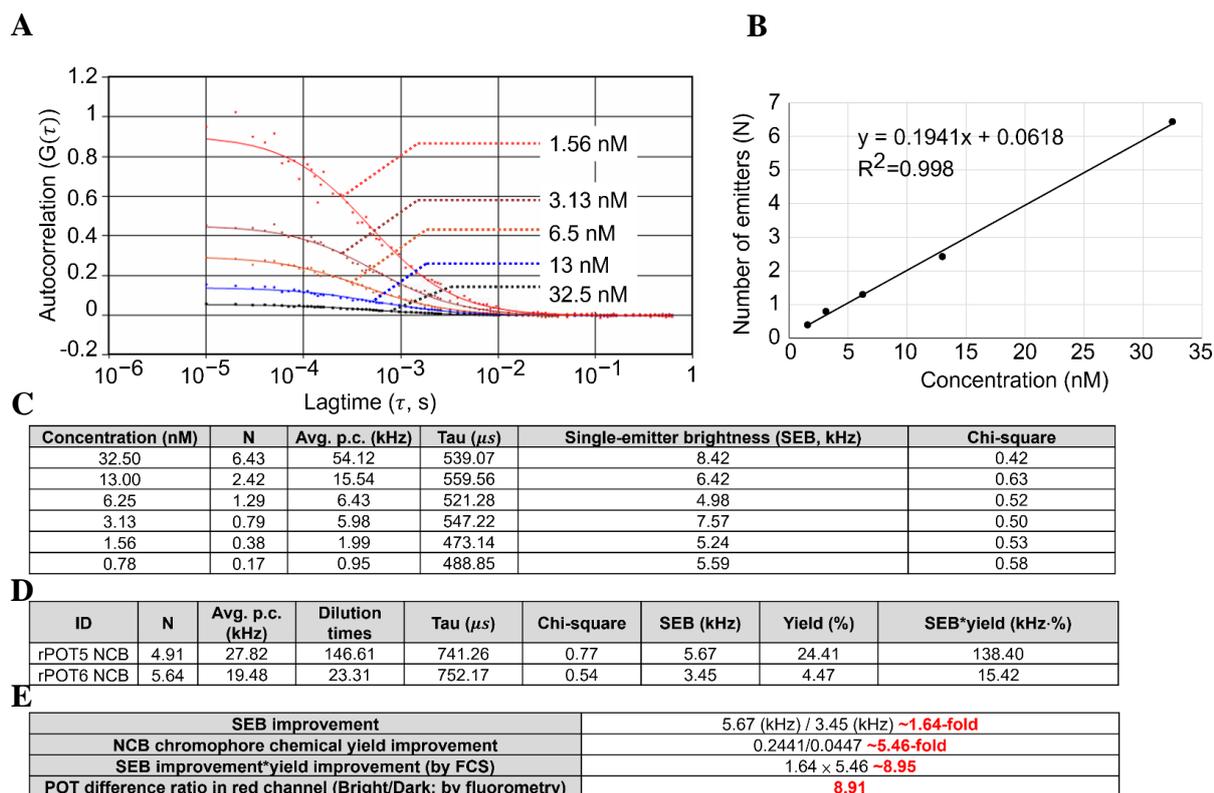
**Figure S24.** Calculation for the number of twin NCB pairs in library\_1. Although library\_1 only has 12,286 activators, they give us totally 110,592 twin NCB pairs. For each position in segment\_1, there are 6 scenarios for creating twin NCBs at that position. Considering we have 6 positions in segment\_1 and we fill up the rest of the 5 positions using the  $4^5$  combinations, we have  $6 \times 4^5 \times 6 = 36,864$  distinct twin NCB pairs just for segment\_1. For 3 segments, there are totally  $3 \times 36,864 = 110,592$  distinct twin NCB pairs in library\_1.



**Figure S25.** 2D spectra of red POT candidates. A) Based on chip selection results, 9 sets of red POT candidates are evaluated in test tubes. All these candidates have their POT difference ratios greater than 1.63, with the largest difference ratio of 8.91 (rPOT5/rPOT6 NCBs, highlighted in solid red box, **Table S7**). B) Based on the hotspots from **Figure 3**, we hypothesized that the disruption of silver-mediated pair C-Ag<sup>+</sup>-C would darken red NCB samples and form red POT pairs. The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm). The blue box represents the hotspots of red POTs. The blue and gray boxes represent the bag position of bright member of red POTs and its counterpart, respectively.



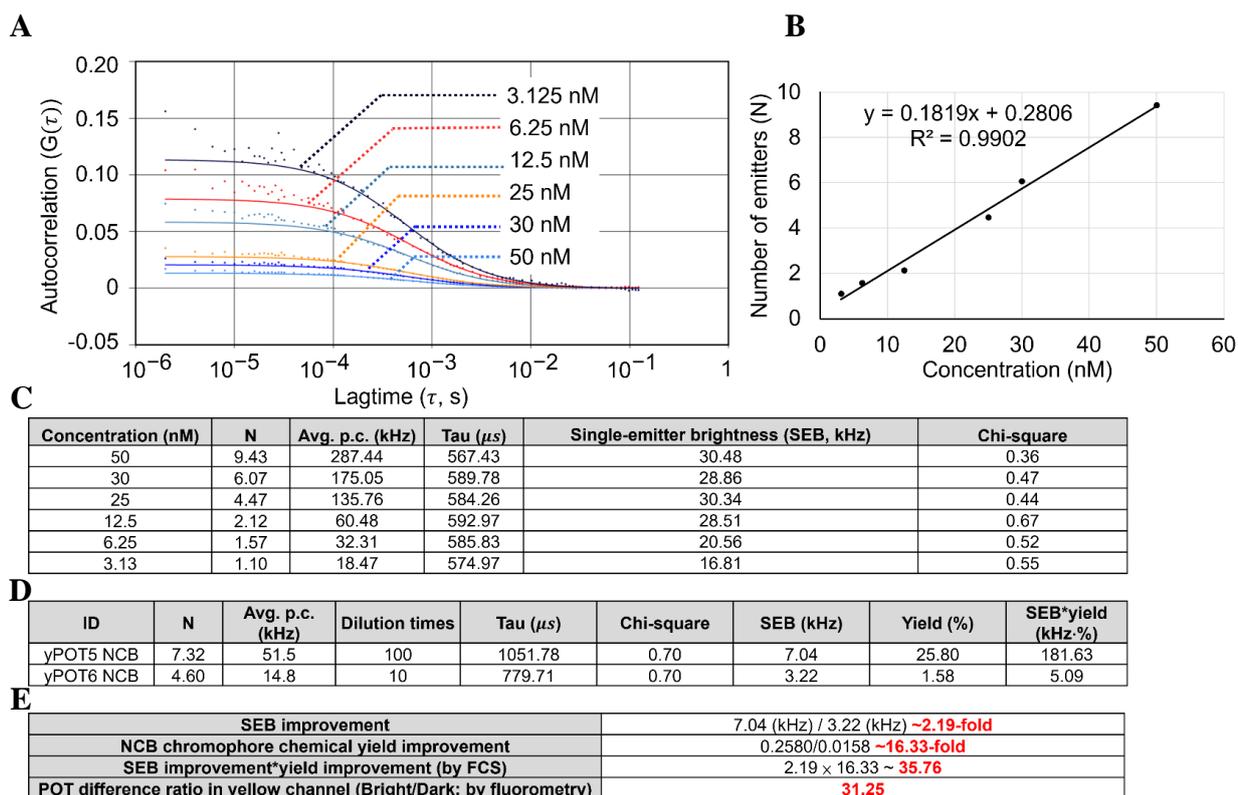
**Figure S26.** 2D spectra of yellow-orange POT candidates. A) Based on chip selection results, 9 sets of yellow-orange POT candidates are evaluated in test tubes. All these candidates have their POT difference ratios greater than 3.29, with the largest difference ratio of 31.25 (yPOT5/yPOT6 NCBs, highlighted in solid orange box, **Table S8**). B) Based on the hotspots from **Figure 3**, we hypothesized that the disruption of WC pair, GC pairing, would darken yellow-orange NCB samples and form yellow-orange POT pairs. The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm). The blue box represents the hotspots of yellow-orange POTs. The blue and gray boxes represent the bag position of bright member of yellow-orange POTs and its counterpart, respectively.



**Avg. p. c.** (average photon count); **N** (number of emitters in detection volume);

**Tau** (dwell time of emitters in detection volume); **SEB** (single-emitter brightness)

**Figure S27.** Fluorescence correlation spectroscopy (FCS) results on the red NCBs. A) The amplitude of autocorrelation function,  $G(0)$ , is inversely proportional to the fluorophore concentration (Atto647N-labeled ssDNA, for calibration purpose), demonstrating no optical saturation in our FCS experiments. B) Number of emitters in the detection volume ( $1/G(0)$ ) shows a linear relationship with the emitter concentration, generating a calibration curve. C) The fitting parameters of the calibration FCS experiment on Atto647N-labeled ssDNA. D) The fitting parameters of the FCS experiment on rPOT5 and rPOT6 NCBs, which are an extreme POT (**Figure 3**). E) From the FCS experiment, it is clearly to see that a single rPOT5 emitter is 1.64-fold brighter than a single rPOT6 emitter (which we term “single-emitter brightness”, SEB), and the concentration of rPOT5 emitter is 5.46 higher than that of rPOT6 emitter (which we call “chromophore chemical yield” or “yield” – not all NC probes carry a AgNC that can be activated). The product of SEB improvement and chromophore chemical yield improvement (8.95) is about the same as the improvement in ensemble enhancement ratio identified by the fluorometer (8.91), indicating that the intensity difference seen in rPOT NCBs is a result of different yield and different SEB. For FCS setup and analysis, please refer to **Methods**.

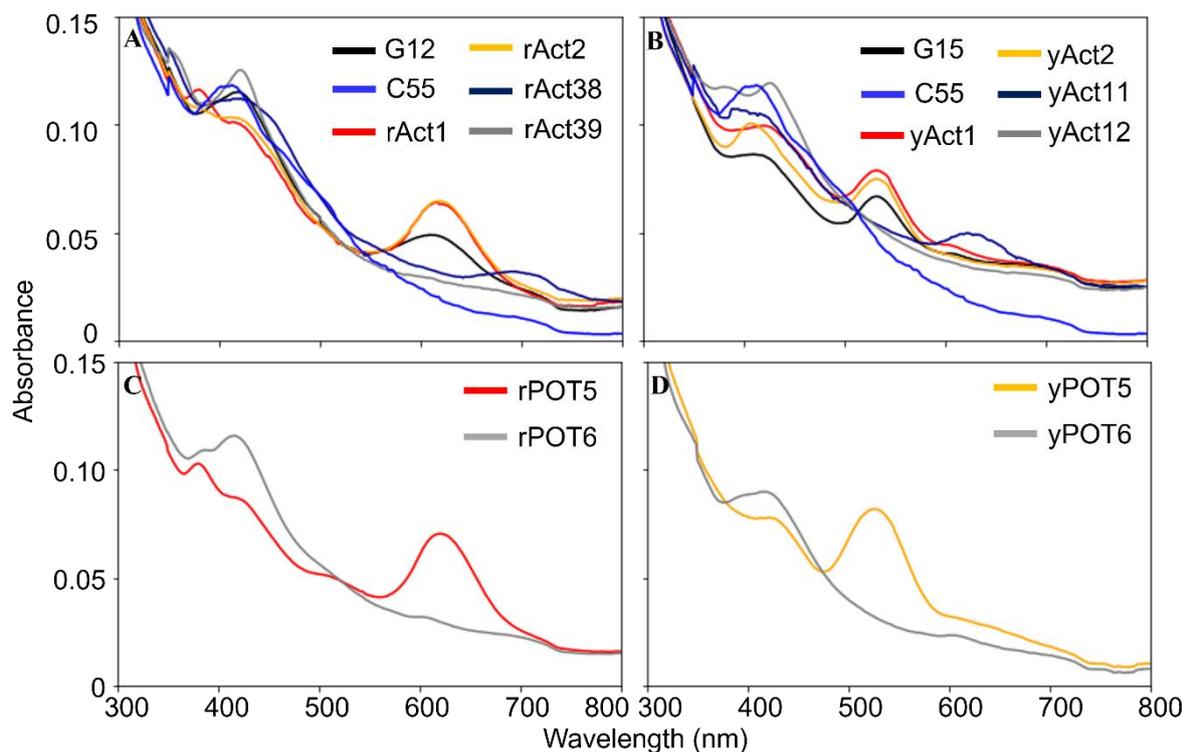


**Avg. p. c.** (average photon count); **N** (number of emitters in detection volume);

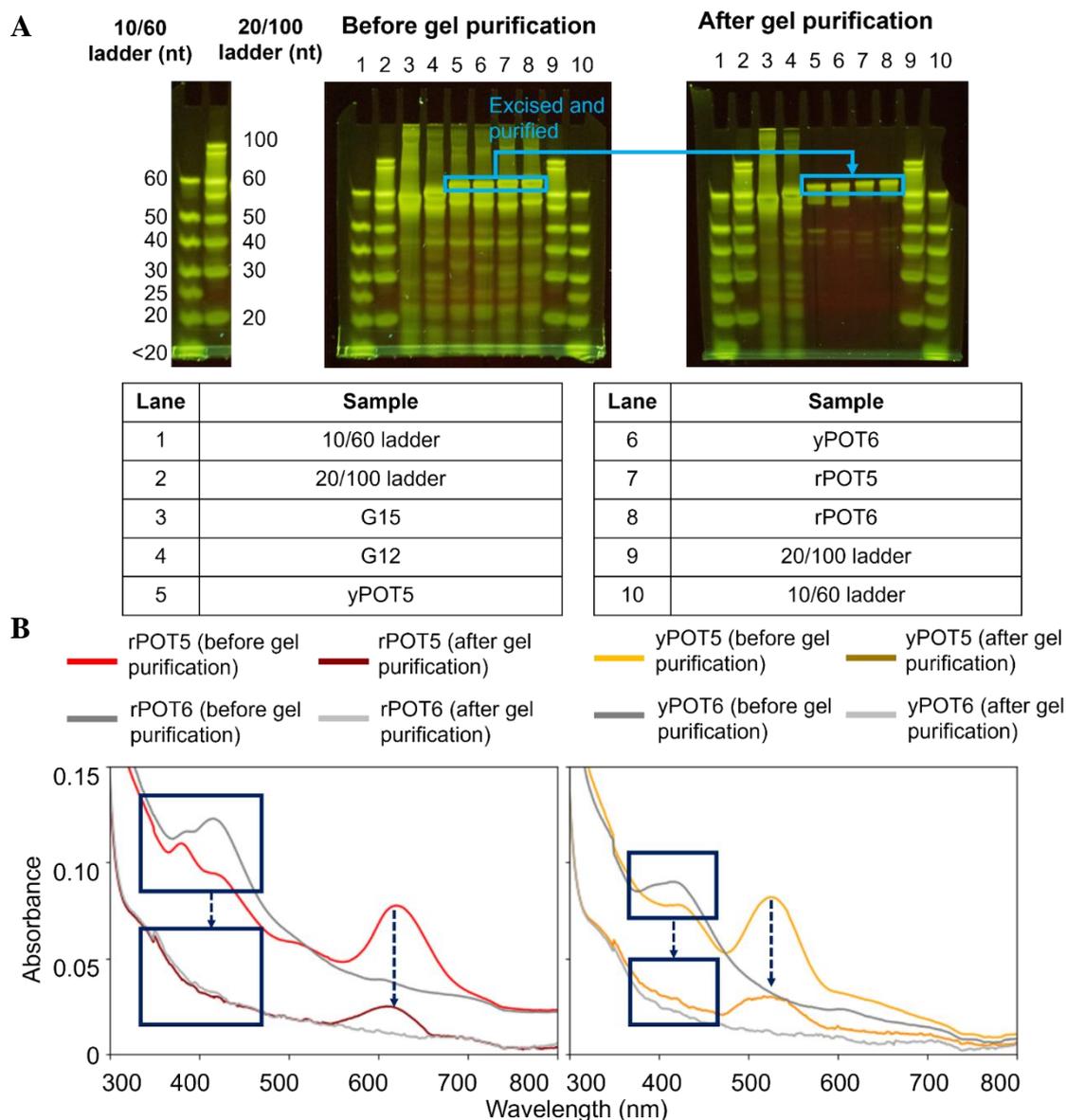
**Tau** (dwell time of emitters in detection volume); **SEB** (single-emitter brightness)

**Figure S28.** Fluorescence correlation spectroscopy (FCS) results on the yellow-orange NCBs.

A) The amplitude of autocorrelation function,  $G(0)$ , is inversely proportional to the fluorophore concentration (Atto532N-labeled ssDNA, for calibration purpose), demonstrating no optical saturation in our FCS experiments. B) Number of emitters in the detection volume ( $1/G(0)$ ) shows a linear relationship with the emitter concentration, generating a calibration curve. C) The fitting parameters of the calibration FCS experiment on Atto532N-labeled ssDNA. D) The fitting parameters of the FCS experiment on yPOT5 and yPOT6 NCBs, which are an extreme POT (**Figure 3**). E) From the FCS experiment, it is clearly to see that the SEB of yPOT5 emitter is 2.19-fold brighter than that of yPOT6 emitter, and the chromophore chemical yield of yPOT5 emitter is 16.33 higher than that of yPOT6 emitter. The product of SEB improvement and chromophore chemical yield improvement (35.76) is about the same as the improvement in ensemble enhancement ratio identified by the fluorometer (31.25), indicating that the intensity difference seen in yPOT NCBs is a result of different yield and different SEB. For FCS setup and analysis, please refer to *Methods*.



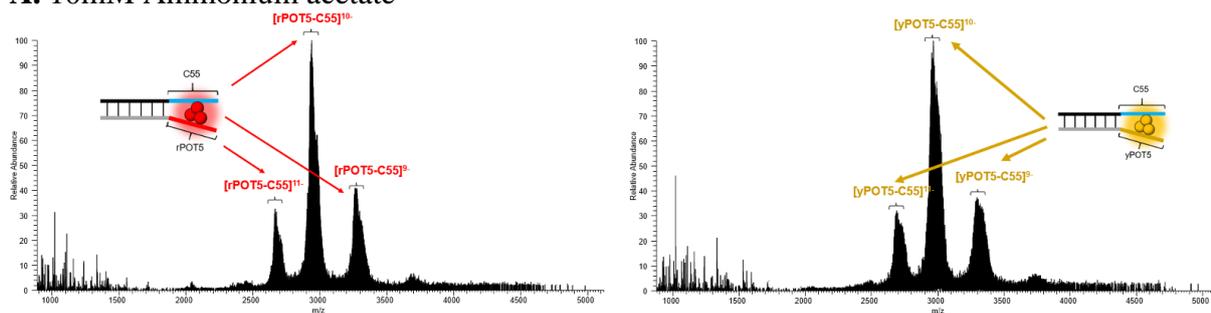
**Figure S29.** Absorption spectra of selected NCBs and POTs. A) For the 4 selected red NCBs, we observed the highest absorbance (0.064) for rAct1 NCB at 610 nm, while that of the 4 selected yellow-orange NCBs reached 0.079 at 525 nm for yAct1 as shown in (B). C) For rPOT5 and rPOT6 NCBs, differences in their absorption spectra around 610 nm were observed (0.070 for rPOT5 and 0.030 for rPOT6). D) For yPOT5 and yPOT6 NCBs, differences in their absorption spectra around 525 nm were observed (0.082 for yPOT5 and 0.032 for yPOT6). The buffer background has been subtracted from these absorption spectra (baseline corrected). The C55 concentrations in these samples are at 15  $\mu\text{M}$ . The spectra are taken on a Cary 5000 UV-Vis-NIR Spectrophotometer.



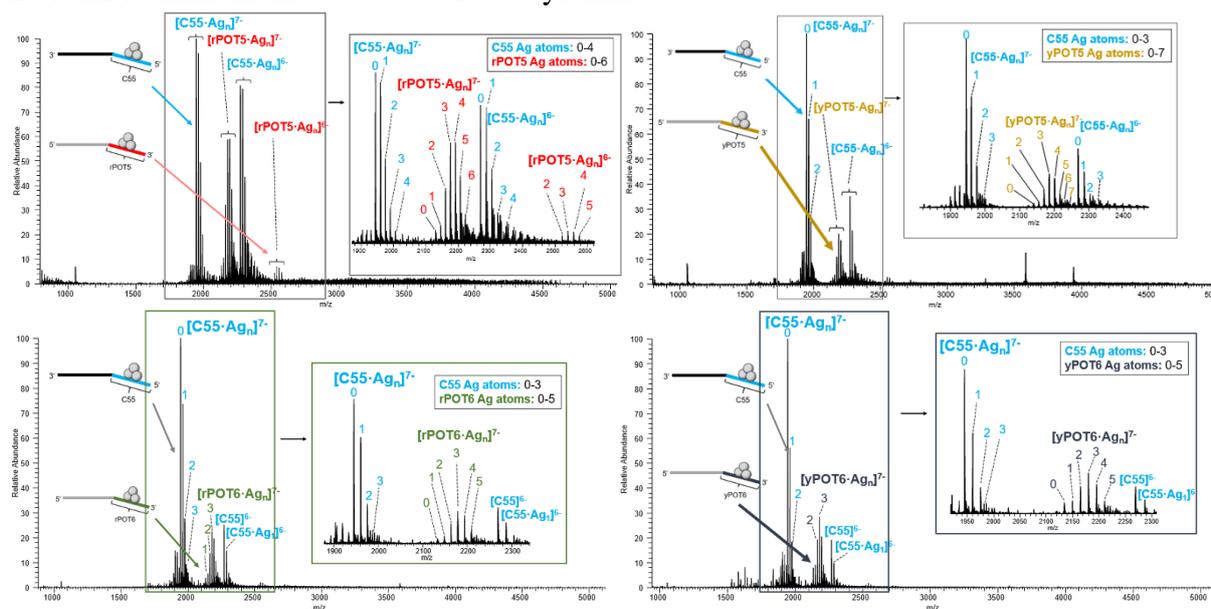
**Figure S30.** Native PAGE gel purification of NCBs. A) Gel of the six NCB samples before and after purification. SYBR Gold DNA dye was used here to stain all species for visualization. In a separate gel, the fluorescent species from the blue box region were excised and eluted. When running the purified species in the native PAGE gel, we could clearly see the removal of subspecies. Please note that G15 and G12 NCBs (Lane 3 and 4) were not gel purified and they served as a control for comparison. The gel photos were taken with a 495 nm long-pass filter in front of the camera to remove the UV excitation light. B) The absorption spectra of gel-purified POT pairs. We could clearly see the differences in absorption spectra before and after gel purification – the absorption peaks around 400 nm (the blue boxes and the dashed arrows) disappeared after purification in both POT pairs. We believe the absorption peaks around 610 nm represented the red NCB species observed in the rPOT5-rPOT6 pair, while the absorption

peaks around 525 nm represented the yellow-orange NCB species in the yPOT5-yPOT6 pair. The buffer background has been subtracted from these absorption spectra (baseline corrected). The spectra are taken on a Cary 5000 UV-Vis-NIR Spectrophotometer.

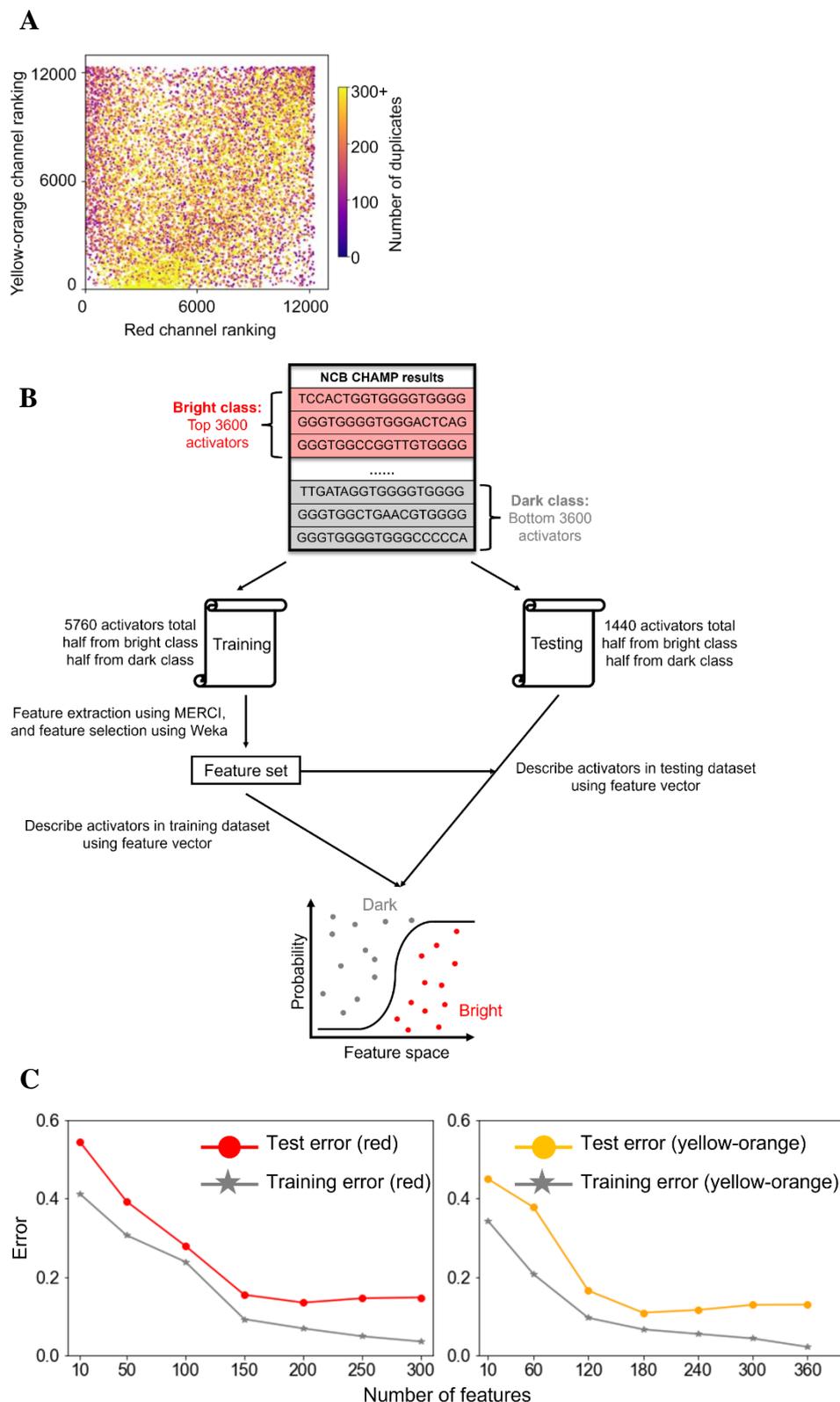
## A. 10mM Ammonium acetate



## B. 10mM Ammonium acetate + 0.1% octylamine

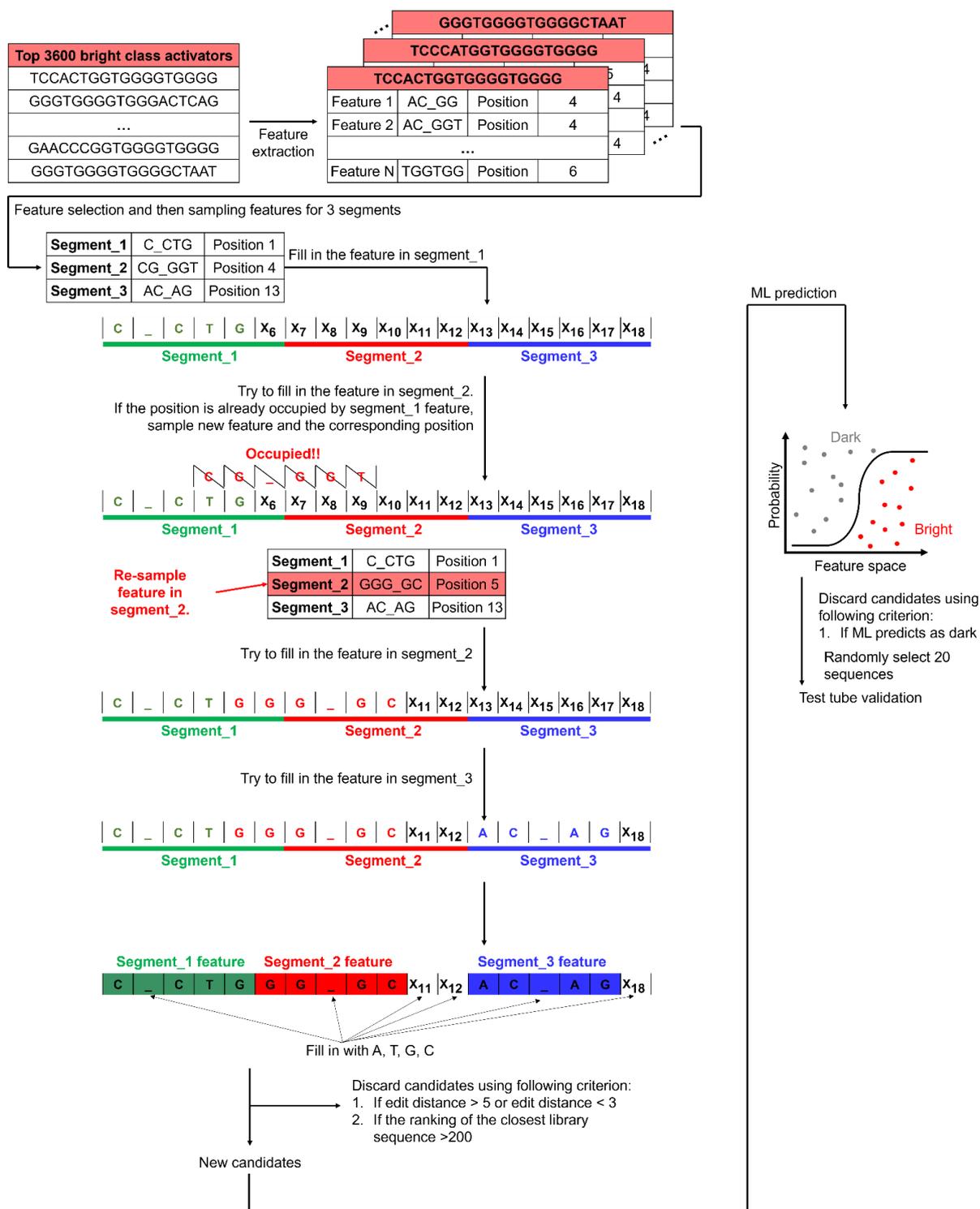


**Figure S31.** ESI-MS analysis of selective NCBs. A) Following the purification using gel electrophoresis, several NCB samples (yPOT5-C55, rPOT5-C55, yPOT6-C55 and rPOT6-C55) were desalted and buffer exchanged into 10 mM ammonium acetate. The resultant mixture was then analyzed by ESI-MS to evaluate the silver stoichiometry of the NCB duplexes. B) Octylamine was added to aliquots of yPOT5-C55 and rPOT5-C55 NCBs (upper row), yPOT6-C55 and rPOT6-C55 (bottom row) at a concentration of 0.1% (v/v), in attempt to reduce the extensive cationic metal adduction that is commonly seen for ESI-MS analysis of oligonucleotides > 20 nt<sup>[6]</sup>. Although the addition of octylamine disassembles the NCB duplexes into their respective NC probe and an activator sequence, we found the C55 NC probes from the rPOT5 NCB carry 0-4 and the others carry 0-3 silver atoms, while the activators from yPOT5, rPOT5, yPOT6 and rPOT6 NCBs carry 0-7, 0-6, 0-5 and 0-5 silver atoms, respectively. These results indicate that the original silver stoichiometry for the intact yPOT5 NCB may be larger than that of the intact rPOT5 NCB. Moreover, the bright member of POTs may be larger than its counterpart.



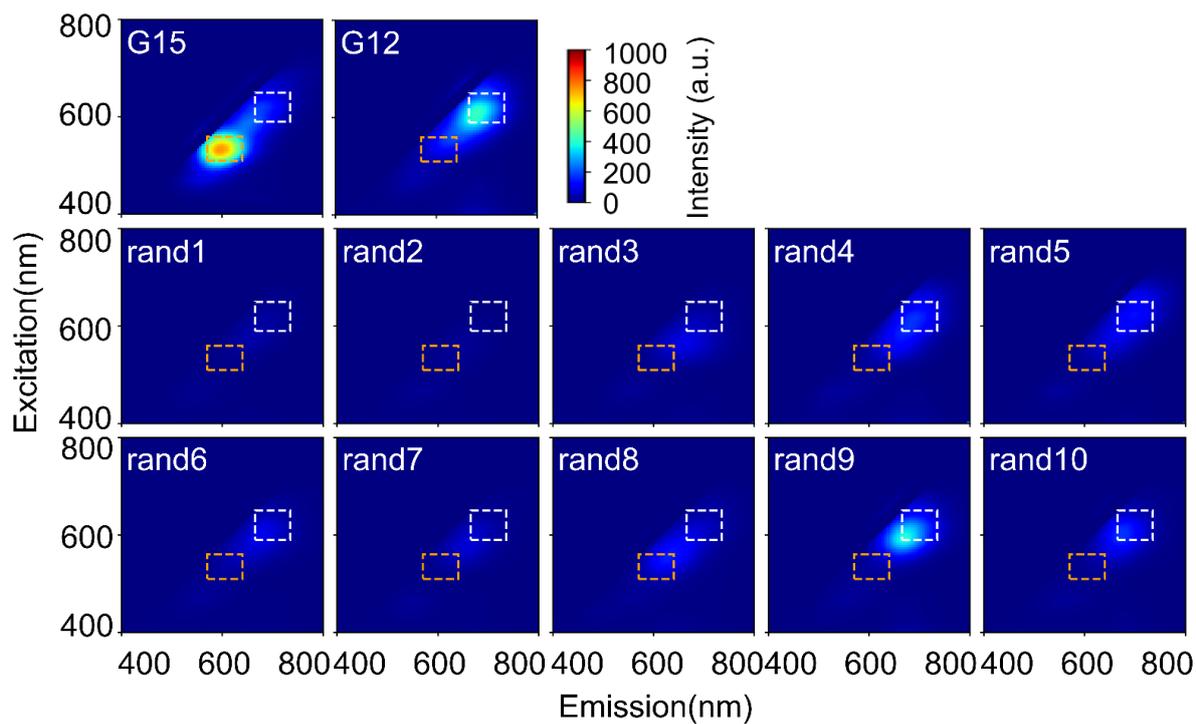
**Figure S32.** Workflow for establishing machine learning models to classify screened NCBs or NCB candidates. A) As the ranking of the activators were quite different between red and yellow-orange color channels, we built ML models based on “bright yellow-orange vs. dark yellow-orange classification” and “bright red vs. dark red classification”, separately. B) In this

report, we performed 5-fold CV to classify our library sequences. Following the approaches proposed by Copp and Gwinn,<sup>[5b, 9]</sup> we labeled the top 30% NCBs as “bright” class and the bottom 30% as “dark” class. The feature extraction process was proceeded using MERCI.<sup>[2]</sup> The extracted motifs were processed with Python scripts to include the position information. We then identified the most discriminative set of features using Weka.<sup>[3]</sup> Based on the selected features, a number of models were established for classifying the chip screening results and we found the model built on LR has the best performance. C) By using different number of features to evaluate LR modeling error, we found that the optimal number of features was close to the number of discriminative features selected by Weka. In our ML modeling process, only sequences from library\_1 (12,286 variants) were implemented. We emphasize that our 3 libraries only covered  $5.8 \times 10^{-5}$  % of the overall sequence space in 18-mers, which is substantially smaller than the 0.1% coverage of 10-mer space in the previous studies.<sup>[5b, 9]</sup>

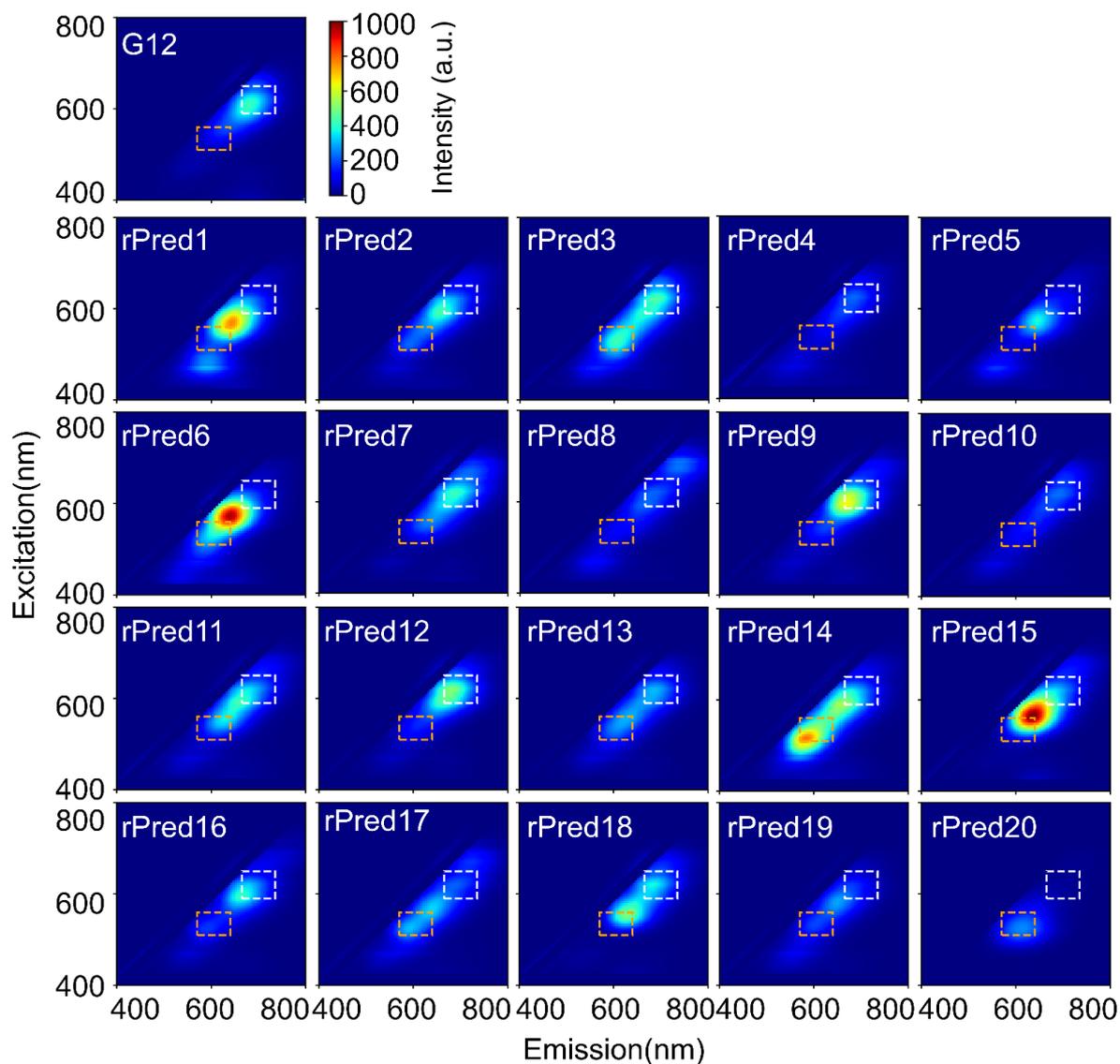


**Figure S33.** Workflow to design bright NCBs *in silico*. Based on the most discriminative features identified by Weka, we sampled the distribution of these features in each segment and generated a list of common motifs with their corresponding positions. To construct a red NCB candidate, we assigned 3 features to the blank 18-nt template, starting with feature\_1 insertion into segment\_1. As feature\_2 might have an overlap with feature\_1 when being inserted into segment\_2. In that situation, the design algorithm would replace feature\_2 with another feature

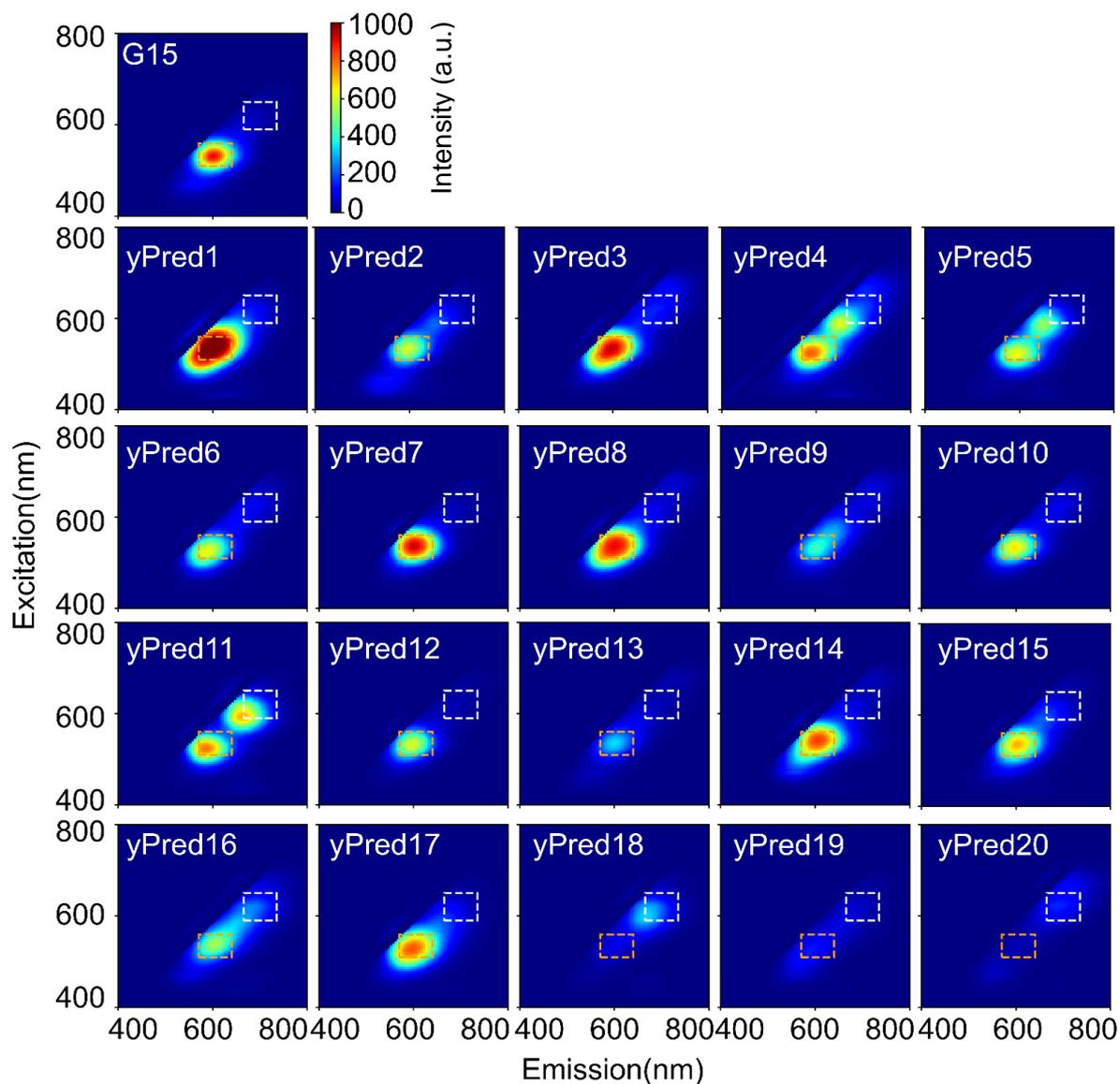
to ensure no overlap. However, if any two features shared identical bases at their overlapping site, they were considered as “compatible” and could be inserted into the same template. For example, as shown above, feature C\_CTG (positions 1-5) and feature GGG\_GC (positions 5-10) shared a guanine base at the overlapping site (position 5). Consequently, they were compatible and were used in constructing a bright NCB candidate. The same procedure was repeated until a compatible feature for segment\_3 was found. Once all three features were inserted into the template, the remaining blank positions were filled up based on the composition popularity (at the same positions) from the bright class sequences. The edit distance<sup>[5]</sup> of the new candidate was then assessed. We only selected new candidates with edit distance between 3 to 5 from the top 200 bright activators screened on chip for test-tube investigation (**Table S10-S11**).



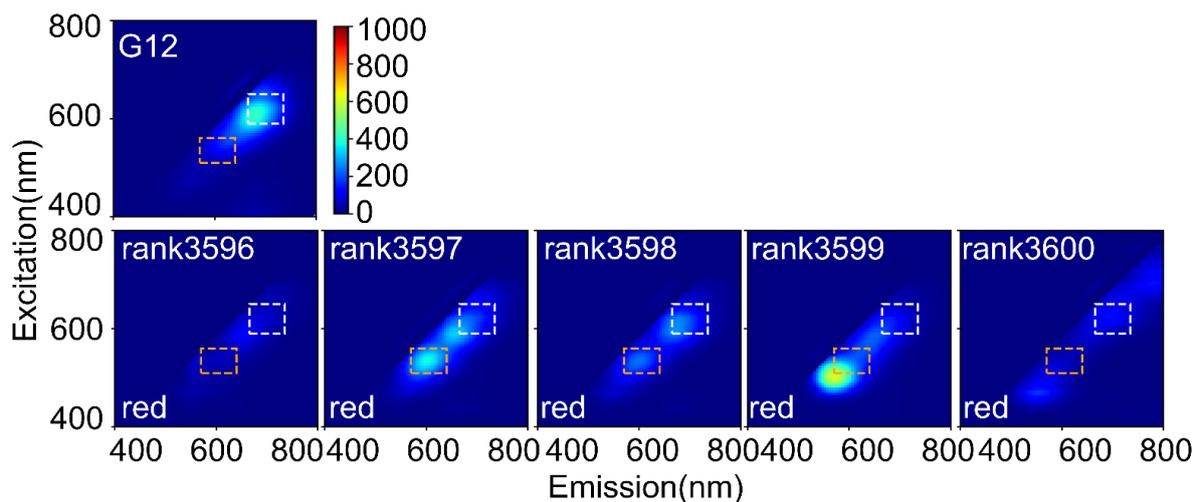
**Figure S34.** 2D spectra of randomly designed NCBs. Ten activators are randomly designed and evaluated in test tubes. The enhancement ratio of rand8 passed the threshold in the yellow-orange channel, while rand4, rand5, rand9 and rand10 passed the threshold in the red channel (**Table S9**). The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm).



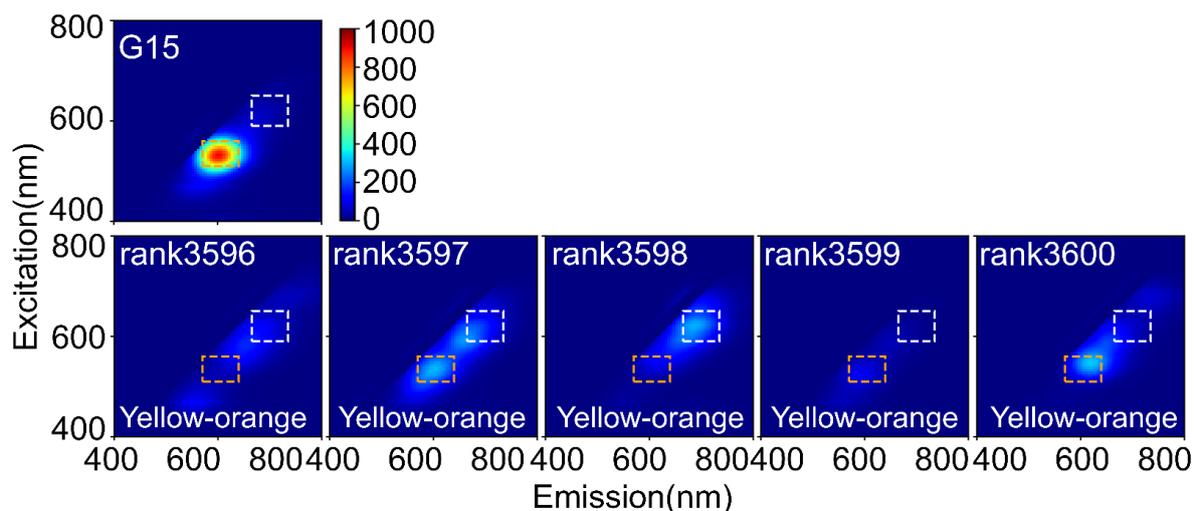
**Figure S35.** 2D spectra of *in silico* designed red NCBs. Twenty activators are designed based on the machine learning results and evaluated in test tubes. On average, the enhancement ratio was 291 for these twenty designs (**Table S10**). The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm).



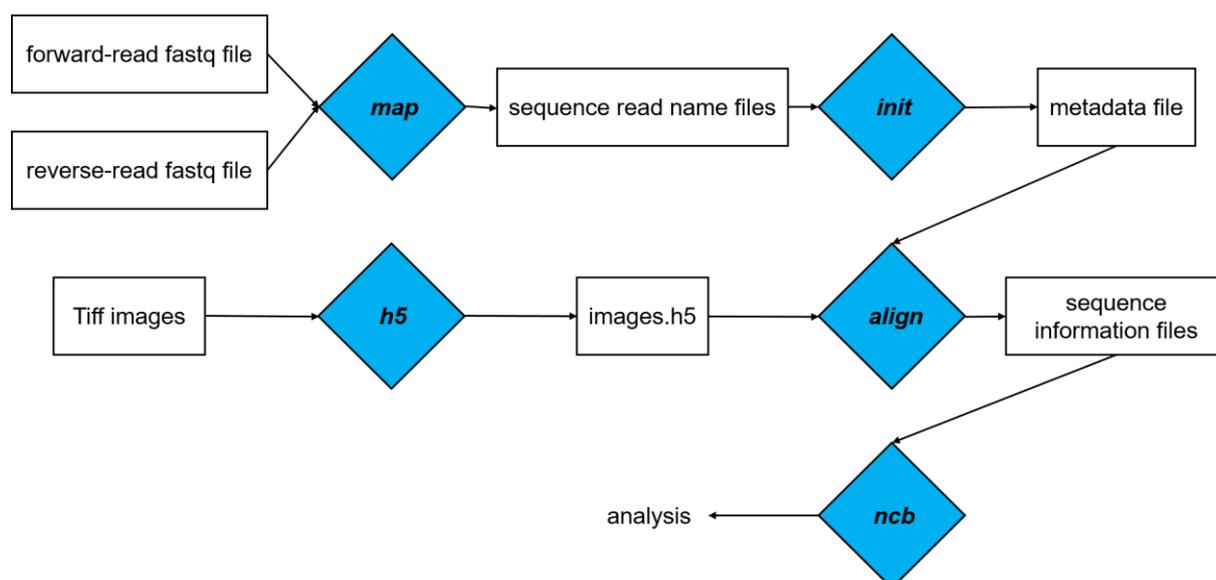
**Figure S36.** 2D spectra of *in silico* designed yellow-orange NCBs. Twenty activators are designed based on the machine learning results and evaluated in test tubes. On average, the enhancement ratio was 532 for these twenty designs (**Table S11**). The white dashed box represents the integrated region of red channel (Ex/Em: 620/60, 700/75 nm), and the orange dashed box represents the integrated region of yellow-orange channel (Ex/Em: 535/50, 605/70 nm).



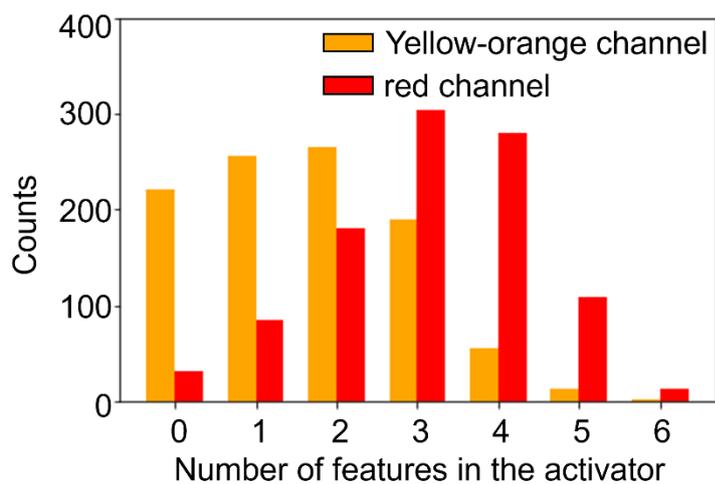
**Figure S37.** 2D spectra of red NCBs near Rank3600. As we apply top 3,600 sequences as our bright class to perform ML modeling. We evaluate the fluorescence intensity of the NCBs ranking number near 3,600. The median enhancement ratio 145 was set as our threshold to evaluate the *in silico* designed red NCBs (**Table S12A**).



**Figure S38.** 2D spectra of yellow-orange NCBs near Rank3600. As we apply top 3,600 sequences as our bright class to perform ML modeling. We evaluate the fluorescence intensity of the NCBs ranking number near 3,600. The median enhancement ratio 66 was set as our threshold to evaluate the *in silico* designed yellow-orange NCBs (**Table S12B**).



**Figure S39.** CHAMP workflow. A custom bioinformatics and imaging processing pipeline named CHAMP (Chip-Hybridized Associated Mapping Platform) was developed by Finkelstein's group and the detailed algorithm description can be found in ref. 8. CHAMP helped decipher the activator sequence behind each activated NCB spot (termed the NCB-CHAMP selection method, **Figure 1C** and **Figure S9**). In brief, mapping the alignment markers was done at four stages. First, a rough alignment was carried out using Fourier-based cross correlation, followed by a precision alignment using least-squares constellation mapping between fastq and *de novo* extracted NCB spots. We built up the consensus sequences and their corresponding information (e.g., lane number, tile number, and x-y coordinates) at all reported positions in the fastq file using the *map* command. Second, the *init* command was executed to record the metadata of imaging settings (e.g., rotation and scaling). Third, the *h5* command was applied to generate a single hdf5 file containing all 512×512 PhiX fiducial marker images. Fourth, the *align* command transformed the processed sequence information into pseudo-images and performed precise alignment. The output files were saved individually by image positions. The content included x, y coordinates of each sequence and the corresponding sequence ID. To analyze our NCB images, we developed an additional function named *ncb*, which corrected the uneven illumination using flat-field correction. A bootstrap method was then performed to derive the median intensity of each activator (baseline corrected; **Methods**) in order to rank the NCB brightness.



**Figure S40.** Feature distribution for the top 1,000 library sequences for red and yellow-orange channels. By evaluating the selected bright features within the top 1,000 library sequences for red and yellow-orange channels, we found the optimal number of features to create bright NCBs would be 2 and 3 for yellow-orange and red channels, respectively.

**Table S1.** Sequences of probes and library designs used in this report.

A) RE strand is used for restriction enzyme digestion (MauBI). The three Atto probes are used for digestion evaluation and NCB-CHAMP alignment. B) The 6-segment and 9-segment interrogation of library\_1. C) Three different library designs. For each of our library sequences on *MiSeq* chip, it consists of six parts: P5 (light blue), SP1 (gold), library sequence (gray for hybridization segment, purple for activator, and dark blue for restriction site), SP2 (orange), barcode (red), and P7 (green). P5 and P7 are adapters for surface attachment. SP1 and SP2 are sequencing-by-synthesis primer binding sites. Barcodes are reserved and used by *Illumina*. The 30-nt-long hybridization segment is for C55 hybridization and the 18-nt-long activator part is where we call “the library”. As for the library size, library\_1 contains 12,286 sequences, library\_2 contains 12,286 sequences, and library\_3 contains 16,255 sequences. The ‘**CG**’ represents the remaining nucleotides after cleavage by a restriction enzyme, and the vertical line represents the cutting site.

**A.**

Acronym	Sequence (5' → 3')
C55	CCC CCT TAA TCC CCC TAT AAT AAA TTT TAA ATA TTA TTT ATT AAT
G15	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG GTG GGG
G12	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA ATC CGG GGT GGG GTG GGG
RE strand	CAG ACG TGT GCT CTT CCG ATC TCG CGC GCG NN
Atto647N-tagged comp_MauBI	/5ATTO647N/ CAG ACG TGT GCT CTT CCG ATC TCG CGC GCG NN
Atto647N-tagged AT	/5ATTO647N/ CCC CCT TAA TCC CCC TAT AAT AAA TTT TAA ATA TTA TTT ATT AAT
Atto532N-tagged AT	/5ATTO532N/ CCC CCT TAA TCC CCC TAT AAT AAA TTT TAA ATA TTA TTT ATT AAT
Alexa488_cPhiX	/5Alex488N/CG GTC TCG GCA TTC CTG CTG AAC CGC TCT TCC GAT C
Forward primer	AAT GAT ACG GCG ACC ACC GAG A
Reverse primer	CAA GCA GAA GAC GGC ATA CGA GA
G15 (90 nt)	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG GTG GGG AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA CTT GTT CAT

**B.** 6-segment interrogation and 9-segment interrogation in library\_1

Acronym	Sequence (5' → 3')
<b>The 6-segment interrogation</b>	

Segment_11	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA <u>NNN</u> TGG GGT GGG GTG GGG
Segment_12	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG <u>NNN</u> GGT GGG GTG GGG
Segment_21	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG <u>NNN</u> GGG GTG GGG
Segment_22	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT <u>NNN</u> GTG GGG
Segment_31	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG <u>NNN</u> GGG
Segment_32	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG GTG <u>NNN</u>
<b>The 9-segment interrogation</b>	
Segment_11	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA <u>NNG</u> TGG GGT GGG GTG GGG
Segment_12	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA <u>GGN</u> <u>NGG</u> GGT GGG GTG GGG
Segment_13	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG <u>TNN</u> GGT GGG GTG GGG
Segment_21	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG <u>NNT</u> GGG GTG GGG
Segment_22	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG <u>GGN</u> <u>NGG</u> GTG GGG
Segment_23	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT <u>GNN</u> GTG GGG
Segment_31	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG <u>NNG</u> GGG
Segment_32	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG <u>GTN</u> <u>NGG</u>
Segment_33	ATT AAT AAA TAA TAT TTA AAA TTT ATT ATA GGG TGG GGT GGG GTG <u>GNN</u>

C. Sequence information of library\_1, library\_2 and library\_3

Acronym	Sequence (5' → 3')
Canonical activator G15 in library_1	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGGGTGGGGTGGGG</u> (activator) <u>CG</u>   CGCGCG (restriction site)</p> <p>AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>GTAGAG</i> (barcode)</p> <p>ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Canonical activator G15 in library_2 and library_3	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGGGTGGGGTGGGG</u> (activator)</p> <p>AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>GTAGAG</i> (barcode)</p> <p>ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Segment_1 activators In library_1	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>NNNNNNGGTGGGGTGGGG</u> (activator) <u>CG</u>   CGCGCG (restriction site)</p> <p>AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode)</p> <p>ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>

Segment_2 activators in library_1	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGNNNNNGTGGGG</u> (activator) <u>CG</u>   CGCGCG (restriction site)</p> <p>AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode)</p> <p>ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Segment_3 activators In library_1	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGGGTGGGNNNNN</u> (activator) <u>CG</u>   CGCGCG (restriction site)</p> <p>AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode)</p> <p>ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Segment_1 activators In library_2	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>GGGNNNNNGGGGTGGGG</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i> (barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Segment_2 activators In library_2	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGGGTNNNNNNGGG</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Segment_3 activators In library_2	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>NNNTGGGGTGGGGTGNNN</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
Segment_1 activators In library_3	<p>AATGATACGGCGACCACCGAGA (P5)</p> <p>TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1)</p> <p>ATTAATAAATAATATTTAAAATTTATTATA<u>NNNTGGNNNGGGGTGGGG</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>

<p>Segment_2 activators in library_3</p>	<p>AATGATACGGCGACCACCGAGA (P5) TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1) ATTAATAAATAATATTTAAAATTTATTATA<u>GGGNNNGGTNNNGTGGGG</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
<p>Segment_3 activators library_3</p>	<p>AATGATACGGCGACCACCGAGA (P5) TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1) ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGNNNGGGNNNGGG</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>
<p>Segment_4 activators In library_3</p>	<p>AATGATACGGCGACCACCGAGA (P5) TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SP1) ATTAATAAATAATATTTAAAATTTATTATA<u>GGGTGGGGTNNNGTGNNN</u> (activator) AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SP2) <i>TTGTTC</i>(barcode) ATCTCGTATGCCGTCTTCTGCTTG (P7)</p>

**Table S2.** Test-tube investigation of selected red activator candidates.

To validate our NCB-CHAMP selection method, we randomly select 20 activators from the 827 activators that are brighter than G12 (ranked 828<sup>th</sup>) and 20 activators from the 11,458 activators that are darker than G12 for further investigation in test tubes. A) Using G12 NCB (ATCCGGGGTGGGGTGGGG) as the standard for red emitter comparison, 17 out of 20 bright red candidates are found brighter than G12 NCB in test tubes (also see **Figure S14**). B) 17 out of 20 dark candidates are found darker than G12 NCB in test tubes (**Figure S15**). The formulas to compute “enhancement ratio” and “improvement ratio” are described in *Methods*. In short, we first calculate the volumetric integrated intensity (**Figure S2**) from the 2D spectrum of the sample in the red channel (Ex: 620/60 nm, Em: 700/75 nm). From there we calculate the enhancement ratio:

$$\text{Enhancement ratio} = (I_{\text{NCB}} - I_{\text{NC probe}}) / (I_{\text{NC probe}} - I_{\text{background}})$$

where  $I_{\text{NCB}}$  stands for the volumetric integrated intensity of NCB in red of yellow-orange window,  $I_{\text{NC probe}}$  represents the volumetric integrated intensity of dark AgNC on C55 probe, and  $I_{\text{background}}$  is the volumetric integrated intensity of buffer (i.e., sodium phosphate buffer). The improvement ratio is simply the ratio of the enhancement ratio of an NCB to that of the standard red NCB (G12). False selections are highlighted in gray below.

We also calculated the integrated intensity ratio to assess the spectral purity. When using 3.5 as a cutoff for spectrum purity, 14 out of 20 bright red NCBs are spectrally pure (70%). NCBs that are below threshold marked in dark yellow.

#### A. Selected red bright NCB candidates

ID	Activator (5' → 3')	Rank	Enhancement ratio in red channel	Improvement ratio (compared to G12)	Integrated intensity ratio (red/yellow-orange)
G12	ATCCGGGGTGGGGTGGGG	828	439	1	3.15
rAct1	TCCATTGGTGGGGTGGGG	3	1292	2.94	9.77
rAct2	TCCAATGGTGGGGTGGGG	4	1247	2.84	8.89
rAct3	TCCATAGGTGGGGTGGGG	5	1229	2.80	9.22
rAct4	ATCCGTGGTGGGGTGGGG	12	601	1.37	8.69
rAct5	TCCTATGGTGGGGTGGGG	55	756	1.72	6.48
rAct6	TCTCATGGTGGGGTGGGG	100	498	1.13	5.71
rAct7	ATCCCAGGTGGGGTGGGG	155	973	2.22	7.91
rAct8	CCTTCTGGTGGGGTGGGG	214	532	1.21	3.61
rAct9	CACATTGGTGGGGTGGGG	293	764	1.74	4.47
rAct10	CCCCAAGGTGGGGTGGGG	356	669	1.52	4.34
rAct11	CCTTGAGGTGGGGTGGGG	395	575	1.31	1.23
rAct12	TCACTAGGTGGGGTGGGG	410	616	1.40	4.78

rAct13	ACTCGTGGTGGGGTGGGG	517	583	1.33	6.04
rAct14	CCTGCAGGTGGGGTGGGG	549	678	1.54	4.42
rAct15	GGGTGGGGTGGGGCTAGA	577	833	1.90	3.01
rAct16	TGGGACGGTGGGGTGGGG	590	694	1.58	1.40
rAct17	TGAACAGGTGGGGTGGGG	596	516	1.18	3.67
rAct18	GCTACAGGTGGGGTGGGG	621	269	0.61	1.40
rAct19	CGGTTTGGTGGGGTGGGG	719	166	0.38	0.99
rAct20	CGCTTCGGTGGGGTGGGG	800	201	0.46	2.00

## B. Selected dark candidates

ID	Activator (5' → 3')	Rank	Enhancement ratio in red channel	Improvement ratio (compared to G12)	Integrated intensity ratio (red/yellow-orange)
G12	ATCCGGGGTGGGGTGGGG	828	439	1	3.15
rAct21	GGGTGGGGTGGGACGCTA	967	221	0.51	3.65
rAct22	ATCTGAGGTGGGGTGGGG	1078	236	0.54	0.80
rAct23	GGGTGGGGTGGGGACATT	1145	256	0.58	2.77
rAct24	AAGTTTGGTGGGGTGGGG	1236	139	0.32	1.27
rAct25	AACGATGGTGGGGTGGGG	1413	404	0.92	3.97
rAct26	TGGCTTGGTGGGGTGGGG	1517	272	0.62	1.73
rAct27	CTGCTTGGTGGGGTGGGG	1676	423	0.96	3.13
rAct28	GGGTGGGGTGGGGAGATC	1717	587	1.34	3.54
rAct29	CTGGCCGGTGGGGTGGGG	1859	347	0.79	3.05
rAct30	AAAAGGGGTGGGGTGGGG	1992	192	0.44	2.04
rAct31	ACGTTTGGTGGGGTGGGG	2084	575	1.30	4.76
rAct32	GGTGCAGGTGGGGTGGGG	3220	124	0.28	0.89
rAct33	ACAATAGGTGGGGTGGGG	5327	375	0.86	3.66
rAct34	GGGTGGGGTGTGGTGGGG	6044	243	0.55	1.49
rAct35	GGGTGGCGATTAGTGGGG	7236	148	0.35	2.78
rAct36	GGGTGGTAATGTGTGGGG	8231	78	0.19	1.36
rAct37	GGGTGGGGTGGGTGTAGG	10835	310	0.71	3.00
rAct38	GGGTGGGTTTATGTGGGG	11357	44	0.10	1.22
rAct39	GGGTGGTCAAAAGTGGGG	12124	81	0.19	1.85
rAct40	GGGTGGCCTCCAGTGGGG	12286	571	1.29	3.38

**Table S3.** Test-tube investigation of twelve selected red activator candidates under different heating condition.

To test whether the in-solution validation using annealing condition for chip experiments (40°C for 40 minutes) would lead to similar results as traditional test-tube heating process (90°C for 1 minute then gradually cools down to room temperature), we compared the 2D spectra of twelve selected red activator candidates and computed their enhancement ratio and improvement ratio between these two conditions (**Figure S16**). The improvement ratio of each candidate was calculated by dividing enhancement ratio of the candidate by the G12 enhancement ratio under two different conditions separately.

ID	Rank	Enhancement ratio, 90°C	Improvement ratio compared to G12, 90°C	Enhancement ratio, 40°C	Improvement ratio compared to G12, 40°C
<b>G12</b>	828	439	1	445	1
<b>rAct2</b>	4	1247	2.84	1233	2.77
<b>rAct5</b>	55	756	1.72	832	1.87
<b>rAct9</b>	293	764	1.74	587	1.32
<b>rAct12</b>	410	616	1.40	481	1.08
<b>rAct19</b>	719	166	0.38	236	0.53
<b>rAct32</b>	3220	124	0.28	200	0.45
<b>rAct33</b>	5327	375	0.86	231	0.52
<b>rAct34</b>	6044	243	0.55	178	0.40
<b>rAct35</b>	7236	148	0.35	142	0.32
<b>rAct36</b>	8231	78	0.19	98	0.22
<b>rAct37</b>	10835	310	0.71	254	0.57
<b>rAct38</b>	11357	44	0.10	111	0.25

**Table S4.** Test-tube investigation of twelve selected yellow-orange activator candidates under different heating condition.

To test whether the in-solution validation using annealing condition for chip experiments (40°C for 40 minutes) would lead to similar results as traditional test-tube heating process (90°C for 1 minute then gradually cools down to room temperature) for yellow-orange NCBs, we compared the 2D spectra of twelve selected yellow-orange activator candidates and computed their enhancement ratio and improvement ratio between these two conditions (**Figure S17**).

The formulas to compute “enhancement ratio” and “improvement ratio” are described in **Methods**. In short, we first calculate the volumetric integrated intensity (**Figure S2**) from the 2D spectrum of the sample in the yellow-orange channel (Ex: 535/50 nm, Em: 605/70 nm). From there we calculate the enhancement ratio:

$$\text{Enhancement ratio} = (I_{\text{NCB}} - I_{\text{NC probe}}) / (I_{\text{NC probe}} - I_{\text{background}})$$

where  $I_{\text{NCB}}$  stands for the volumetric integrated intensity of NCB in yellow-orange window,  $I_{\text{NC probe}}$  represents the volumetric integrated intensity of dark AgNC on C55 probe, and  $I_{\text{background}}$  is the volumetric integrated intensity of buffer (i.e., sodium phosphate buffer). The improvement ratio is simply the ratio of the enhancement ratio of an NCB to that of the standard yellow-orange activator (G15).

**A.** 90°C for 1 minute then gradually cools down to room temperature

ID	Activator (5' → 3')	Rank	Enhancement ratio, 90°C	Improvement ratio compared to G15, 90°C
<b>G15</b>	<b>GGGTGGGGTGGGGTGGGG</b>	<b>673</b>	<b>553</b>	<b>1</b>
yAct1	TTGGTGGGGTGGGGTGGGG	2	1125	2.03
yAct2	GAGTTCGGTGGGGTGGGG	1519	310	0.56
yAct3	GGGTGGGGTGGGCCACAG	2535	77	0.14
yAct4	GCCGGGGTGGGGTGGGG	3672	608	1.10
yAct5	GGGTGGGGTGGGCACGGC	4655	88	0.16
yAct6	GGGTGGGGTGGGGTAATA	5404	188	0.34
yAct7	GGGTGGGGTGGGTTTTCT	6540	149	0.27
yAct8	CGCTACGGTGGGGTGGGG	7481	72	0.13
yAct9	GGGTGGGGTGGGCTACGG	8102	77	0.14
yAct10	GGGTGGGGTGGGCCCTTT	9863	127	0.23
yAct11	GGGTGGGGCGCCGTGGGG	10087	100	0.18
yAct12	TCATTTCGGTGGGGTGGGG	11247	66	0.12

**B.** 40°C for 40 minutes

<b>ID</b>	<b>Activator (5' → 3')</b>	<b>Rank</b>	<b>Enhancement ratio, 40°C</b>	<b>Improvement ratio compared to G15, 40°C</b>
<b>G15</b>	<b>GGGTGGGGTGGGGTGGGG</b>	<b>673</b>	<b>549</b>	<b>1</b>
<b>yAct1</b>	TTGGTGGGTGGGGTGGGG	2	1179	2.15
<b>yAct2</b>	GAGTTCGGTGGGGTGGGG	1519	302	0.55
<b>yAct3</b>	GGGTGGGGTGGGCCACAG	2535	71	0.13
<b>yAct4</b>	GCCGGGGGTGGGGTGGGG	3672	478	0.87
<b>yAct5</b>	GGGTGGGGTGGGCACGGC	4655	77	0.14
<b>yAct6</b>	GGGTGGGGTGGGGTAATA	5404	186	0.34
<b>yAct7</b>	GGGTGGGGTGGGTTTTCT	6540	110	0.20
<b>yAct8</b>	CGCTACGGTGGGGTGGGG	7481	45	0.08
<b>yAct9</b>	GGGTGGGGTGGGCTACGG	8102	55	0.10
<b>yAct10</b>	GGGTGGGGTGGGCCCTTT	9863	77	0.14
<b>yAct11</b>	GGGTGGGGCGCCGTGGGG	10087	71	0.13
<b>yAct12</b>	TCATTCGGTGGGGTGGGG	11247	49	0.09

**Table S5.** Test-tube investigation of 10-guanine activators.

Based on chip selection results, ten 10G activators can potentially be brighter than G12 NCB (**Figure S23**). Test-tube investigation proves that 7 of the selected 10G activators have their enhancement ratios comparable to that of G12 in the red channel (improvement ratio  $\geq 0.9$ ). This result indicates that it is possible to create bright red NCBs with fewer numbers of guanine. False selections are highlighted in gray below.

ID	Sequence (5' → 3')	Enhancement ratio	Improvement ratio (compared to G12)
G12	ATCCGGGGTGGGGTGGGG	439	1
10G 1	AACCTTGGTGGGGTGGGG	415	0.95
10G 2	TCCAATGGTGGGGTGGGG	409	0.93
10G 3	ATCCATGGTGGGGTGGGG	397	0.91
10G 4	ATCCAGGTGGGGTGGGG	413	0.95
10G 5	TACCATGGTGGGGTGGGG	527	1.21
10G 6	GGGTGGTCCCCCGTGGGG	240	0.54
10G 7	AACCATGGTGGGGTGGGG	480	1.09
10G 8	CTCCATGGTGGGGTGGGG	473	1.09
10G 9	ACATCAGGTGGGGTGGGG	204	0.47
10G 10	GGGTGGCCCCCGTGGGG	221	0.51

**Table S6.** Test-tube investigation for activators having at least 12 guanines.

Based on chip selection results, ten 12G activators can potentially be darker than G12 NCB (**Figure S23**). Test-tube investigation proves that all of the selected 12G activators are darker than G12 in the red channel (improvement ratio < 0.6).

ID	Sequence (5' → 3')	Enhancement ratio	Improvement ratio (compared to G12)
G12	ATCCGGGGTGGGGTGGGG	439	1
12G_1	GGGTGGTTCGGACGTGGGG	61	0.15
12G_2	GGGTGGTGTCAGGTGGGG	124	0.29
12G_3	GGGTGGAAGAGGGTGGGG	51	0.12
12G_4	GGGTGGTTGCTGGTGGGG	260	0.59
12G_5	GGGTGGGTCGCCGTGGGG	126	0.29
12G_6	GGGTGGAGTGATGTGGGG	45	0.11
12G_7	GGGTGGTGAGACGTGGGG	26	0.06
12G_8	GGGTGGGCTGACGTGGGG	31	0.08
12G_9	GGGTGGAAGAGTGTGGGG	57	0.14
12G_10	GGGTGGACGACGGTGGGG	17	0.05

**Table S7.** Test-tube investigation of red POT candidates.

Based on chip selection results, 9 sets of highly ranked red POT candidates identified in NGS screening are evaluated in test tubes. We emphasize that these sequences were highly ranked POT candidates identified in NGS screening, around the POT hotspots. But among these highly ranked candidates, we did not cherry-pick the ones to present here. All these candidates have their POT difference ratios greater than 1.63, with the largest difference ratio of 8.91 (rPOT5/rPOT6, **Figure S25A**). Single-nucleotide differences in these POTs are marked in red. POT difference ratio is simply the ratio of the enhancement ratios of the twins, which is,

$$\text{POT difference ratio} = \frac{\text{Enhancement ratio of bright candidate}}{\text{Enhancement ratio of dark candidate}}$$

ID (Bright)	Sequence (5' → 3')	ID (Dark)	Sequence (5' → 3')	POT difference ratio in red channel (Bright/Dark)
rPOT1	AT <b>C</b> CGTGGTGGGGTGGGG	rPOT2	AT <b>T</b> CGTGGTGGGGTGGGG	4.43±0.68
rPOT3	T <b>C</b> CATTGGTGGGGTGGGG	rPOT4	T <b>T</b> CATTGGTGGGGTGGGG	6.55±0.92
rPOT5	AAT <b>C</b> CTGGTGGGGTGGGG	rPOT6	AAT <b>T</b> CTGGTGGGGTGGGG	8.91±1.31
rPOT7	T <b>C</b> CATAGGTGGGGTGGGG	rPOT8	T <b>G</b> CATAGGTGGGGTGGGG	3.10±0.55
rPOT7	T <b>C</b> CATAGGTGGGGTGGGG	rPOT9	T <b>A</b> ATAGGTGGGGTGGGG	8.32±1.81
rPOT3	T <b>C</b> CATTGGTGGGGTGGGG	rPOT10	T <b>A</b> CATTGGTGGGGTGGGG	3.10±0.39
rPOT3	T <b>C</b> CATTGGTGGGGTGGGG	rPOT11	T <b>C</b> GATTGGTGGGGTGGGG	3.39±1.11
rPOT1	AT <b>C</b> CGTGGTGGGGTGGGG	rPOT12	AT <b>C</b> AGTGGTGGGGTGGGG	2.82±1.34
rPOT13	AT <b>C</b> CGAGGTGGGGTGGGG	rPOT14	AT <b>C</b> GAGGTGGGGTGGGG	1.63±0.20

ID (Bright)	Sequence (5' → 3')	ID (Dark)	Sequence (5' → 3')	POT difference ratio in yellow-orange channel (Bright/Dark)
rPOT1	AT <b>C</b> CGTGGTGGGGTGGGG	rPOT2	AT <b>T</b> CGTGGTGGGGTGGGG	0.73±0.07
rPOT3	T <b>C</b> CATTGGTGGGGTGGGG	rPOT4	T <b>T</b> CATTGGTGGGGTGGGG	1.46±0.24
rPOT5	AAT <b>C</b> CTGGTGGGGTGGGG	rPOT6	AAT <b>T</b> CTGGTGGGGTGGGG	2.78±0.24
rPOT7	T <b>C</b> CATAGGTGGGGTGGGG	rPOT8	T <b>G</b> CATAGGTGGGGTGGGG	0.77±0.30
rPOT7	T <b>C</b> CATAGGTGGGGTGGGG	rPOT9	T <b>A</b> ATAGGTGGGGTGGGG	1.65±0.20
rPOT3	T <b>C</b> CATTGGTGGGGTGGGG	rPOT10	T <b>A</b> CATTGGTGGGGTGGGG	1.19±0.11
rPOT3	T <b>C</b> CATTGGTGGGGTGGGG	rPOT11	T <b>C</b> GATTGGTGGGGTGGGG	1.52±0.40
rPOT1	AT <b>C</b> CGTGGTGGGGTGGGG	rPOT12	AT <b>C</b> AGTGGTGGGGTGGGG	1.59±0.57
rPOT13	AT <b>C</b> CGAGGTGGGGTGGGG	rPOT14	AT <b>C</b> GAGGTGGGGTGGGG	0.51±0.05

**Table S8.** Test-tube investigation of yellow-orange POT candidates.

Based on chip selection results, 6 sets of highly ranked yellow-orange POT candidates identified in NGS screening are evaluated in test tubes. All these candidates have their POT difference ratios greater than 3.29, with the largest difference ratio of 31.25 (yPOT5/yPOT6 NCBs, **Figure S26A**). Single-nucleotide differences in these POTs are marked in red.

ID (Bright)	Sequence (5' → 3')	ID (Dark)	Sequence (5' → 3')	POT difference ratio in yellow-orange channel (Bright/Dark)
yPOT1	TAA <b>G</b> TGGGTGGGGTGGGG	yPOT2	TAA <b>C</b> TGGGTGGGGTGGGG	9.16±1.65
yPOT3	TTAGT <b>G</b> GGTGGGGTGGGG	yPOT4	TTAGT <b>C</b> GGTGGGGTGGGG	9.41±0.69
yPOT5	CAGT <b>G</b> AGGTGGGGTGGGG	yPOT6	CAGT <b>C</b> AGGTGGGGTGGGG	31.25±5.37
yPOT7	AGCT <b>G</b> AGGTGGGGTGGGG	yPOT8	AGCT <b>A</b> AGGTGGGGTGGGG	14.17±2.95
yPOT9	ACAG <b>T</b> GGGTGGGGTGGGG	yPOT10	ACAG <b>C</b> GGGTGGGGTGGGG	6.15±1.20
yPOT11	ACAG <b>I</b> GGGTGGGGTGGGG	yPOT12	ACAG <b>A</b> GGGTGGGGTGGGG	3.29±0.26

ID (Bright)	Sequence (5' → 3')	ID (Dark)	Sequence (5' → 3')	POT difference ratio in red channel (Bright/Dark)
yPOT1	TAA <b>G</b> TGGGTGGGGTGGGG	yPOT2	TAA <b>C</b> TGGGTGGGGTGGGG	0.88±0.07
yPOT3	TTAGT <b>G</b> GGTGGGGTGGGG	yPOT4	TTAGT <b>C</b> GGTGGGGTGGGG	0.69±0.24
yPOT5	CAGT <b>G</b> AGGTGGGGTGGGG	yPOT6	CAGT <b>C</b> AGGTGGGGTGGGG	2.10±0.34
yPOT7	AGCT <b>G</b> AGGTGGGGTGGGG	yPOT8	AGCT <b>A</b> AGGTGGGGTGGGG	0.83±0.15
yPOT9	ACAG <b>T</b> GGGTGGGGTGGGG	yPOT10	ACAG <b>C</b> GGGTGGGGTGGGG	0.84±0.07
yPOT11	ACAG <b>I</b> GGGTGGGGTGGGG	yPOT12	ACAG <b>A</b> GGGTGGGGTGGGG	0.68±0.05

**Table S9.** Test-tube investigation of randomly designed NCBs and G5.

Ten randomly designed activators and a hypothetical bright candidate (G5) designed based on **Figure 2** conclusion were evaluated in test tubes. Note that we do not preset any threshold of predicted success before selection here. On average, the enhancement ratio of the ten designs in yellow-orange and red channels were 19 and 126, respectively. Since we selected top 30% (3,600) activator sequences as the bright class, we used the median enhancement ratio value from Rank3595 to Rank3600 sequences as our new criteria for bright/dark categorization, which corresponded to 145 and 66 for red and yellow-orange channels, respectively (see **Table S12**). As a result, 1 out the 10 random sequences was identified as a “bright yellow-orange” activator and 4 out of the 10 random sequences were identified as “bright red” activators.

ID	Sequence (5' → 3')	Emission peak (nm)	Enhancement ratio (yellow-orange)	Improvement ratio (compared to G15)	Enhancement ratio (red)	Improvement ratio (compared to G12)
G5	CCCCCGCGGGGTTTCCC	645	39	0.09	83	0.19
rand1	AGGGACTAGGTGGGCGCT	660	9	0.02	44	0.10
rand2	CGCGTGAGCGAGGTCGAG	630	10	0.02	9	0.02
rand3	GTACGGAGGTGAGCTTGG	660	23	0.04	66	0.15
rand4	TGTGCACAAGAGGGGAGG	685	30	0.05	250	0.57
rand5	GCTGATTGGGCGCTTGGG	695	24	0.04	206	0.47
rand6	GGCCGACTTGTGGGTAGG	675	24	0.04	92	0.21
rand7	TGAGGGCTGAGACGCCGG	660	19	0.03	53	0.12
rand8	GCTCGGGCCAGGTGGAAG	625	68	0.12	79	0.18
rand9	AGTGGGGATGAGTGTGCA	665	28	0.05	316	0.72
rand10	GCCGGGTTGTAGATGGGT	670	18	0.03	149	0.34

**Table S10.** Test-tube investigation of *in silico* designed bright red NCBs.

Twenty activators are designed based on the machine learning results and evaluated in test tubes. Following the observation shown in **Figure S40**, the new red candidates were generated if three bright red features were presented in the sequences as shown below. On average, the enhancement ratio was 291 for these twenty designs and the mean edit distance was 4.0. Here the enhancement ratio of 145 was set as the cutoff for bright red NCBs (**Table S12A**). Five out of the 20 *in silico* designed red NCBs below showed either low emission (rPred19 and rPred20, highlighted in gray) or yellow-orange emission (rPred14, rPred15 and rPred18, highlighted in yellow). Among the 20 NCBs below, rPred9 NCB was the brightest (1.30-fold brighter than G12 NCB, highlighted in red).

ID	Activator (5' → 3')	Emission peak (nm)	Minimal edit distance	Enhancement ratio	Improvement ratio compared to G12	Motif #1	Motif #2	Motif #3
rPred1	TCCCATGCGGGGCTCG GG	655	4	220	0.50	CCC	GGG_ C	TC_G G
rPred2	CCCGAAGGGGGATCG CG	640	4	250	0.57	CCC	GGGG A	TC_C G
rPred3	CCCGAAGGTGGGGCTC TG	655	4	487	1.11	CCC	GA_G GT	C_CT G
rPred4	TACCAAGGGGGGAACG GG	685	4	167	0.38	CA_G G	GGGG A	AA_G G
rPred5	ACCAGAGGGGTGGGCC CG	645	5	154	0.35	ACC_ G	AG_G GT	CCC
rPred6	TCCCAAGGTGGGGGGC AG	640	3	198	0.45	CCC	CA_G G	GGGG _C
rPred7	TCCCGAGGTTGGGTCT GG	685	3	408	0.93	CCC	CG_G GT	CTGG
rPred8	TCCAGCGGGGGAGGGG GC	735	4	184	0.42	TCC_ G	GGGG A	GGG_ C
rPred9	ATCCCTCGGGGAGGGG GC	670	5	571	1.30	CCC	GGGG A	GGGG _C
rPred10	CATCCGTTGGGGGACG GG	685	5	180	0.41	A_CC G	TTGG _G	GGGA C
rPred11	GCCCGAGGGGGGACG GG	655	3	373	0.85	CCC	C_AG G	GGGA C
rPred12	TCCAGTGGGGGGAGCG GG	680	4	505	1.15	TCC_ G	GGGG A	GCGG
rPred13	CCCGTAGGGTAGGTTG GG	685	4	316	0.72	CCC	TA_G GT	TT_G G
rPred14	CCCGAAGGGGGGGCA TG	580	5	531	1.21	CCC	AA_G G	GGG_ C
rPred15	TCCCGCGGGGGGACG GG	635	3	325	0.74	CCC	GGGA	GGGA C
rPred16	TCCGACGGGGGTGGG GG	660	3	277	0.63	TCC_ G	AC_G G	TGGG GG
rPred17	TCCCCAGGGGACTGG GG	640	3	170	0.39	CCC	GGGG A	CTGG
rPred18	ATCCTTCGGGGATCG GG	630	5	380	0.87	CTT_ G	GGGG A	A.CG G
rPred19	TCCAAGGGGTGGACTG GC	650	4	101	0.23	AA_G G	AG_G GT	CTGG
rPred20	TACCCAGGGGACTGG GC	650	4	31	0.07	CCC	CA_G G	ACT_ G

**Table S11.** Test-tube investigation of *in silico* designed yellow-orange NCBs.

Twenty activators are designed based on the machine learning results and evaluated in test tubes. Following the observation shown in **Figure S40**, the new yellow-orange candidates were generated if two bright yellow-orange features were presented in the sequences as shown below. On average, the enhancement ratio was 532 for these twenty designs and the mean edit distance was 3.5. Here the enhancement ratio of 66 was set as the cutoff for bright yellow-orange NCBs (**Table S12B**). Three out of the 20 *in silico* designed yellow-orange NCBs below showed either low emission (yPred18, yPred19 and rPred20, highlighted in gray) or red emission (yPred18 and rPred20). Thus, the overall test-tube validation accuracy was 85%. Among the 20 NCBs below, yPred1 NCB was the brightest (2.30-fold brighter than G15 NCB, highlighted in yellow-orange).

ID	Activator (5' → 3')	Emission peak (nm)	Minimal edit distance	Enhancement ratio	Improvement ratio compared to G15	Motif #1	Motif #2
yPred1	GTGTTGGGTGGTCGGGGG	585	3	1272	2.30	GTG TG	TGGGTG
yPred2	GGTGTGGGTGGGAAGGGC	595	3	371	0.67	GT TG	TGGGTG
yPred3	TGTGTGTGGGGATGGGG	595	3	968	1.75	GT TGG	GGGGG
yPred4	GCTGTGTGGGGTGTGGGG	585	3	724	1.31	GT TGG	GTGTGG
yPred5	GGAGTGGGTGGTGGTGGG	590	3	487	0.88	TGGGTG	GTG TG
yPred6	TCGGTGTGGTGTGTGGGG	585	4	299	0.54	GTGGTG	GTGTGG
yPred7	TGGTGTGGTTGGCGGGGG	600	3	946	1.71	GT TGG	T GCG
yPred8	AGTGTGGTGTGGGGGGG	595	5	619	1.12	GTGGTG	TTG GG
yPred9	GCTTGGGTGGGTGTGGGC	600	3	448	0.81	GT TGG	TGGGTG
yPred10	AGTGGGTGTGTGTGGGGG	595	4	680	1.23	GT TGG	TGGGTG
yPred11	GAGTTAGGGGTGTGGGGC	580	5	885	1.60	GT AG	GT TGG
yPred12	AGTGTGGGTGTGTGGGGG	595	3	481	0.87	GT TGG	TGGGTG
yPred13	GGTGTGGGTGTGTGGGGG	600	3	249	0.45	GT TGG	TGGGTG
yPred14	GTTGTGTTGGGAGGGGGG	600	4	559	1.01	GTGGTG	GA GGG
yPred15	GTATAGTGGGTGTGGGC	600	4	498	0.90	TGGGTG	GTGTGG
yPred16	GTCGTGGTGGTGGTGGGC	600	4	470	0.85	GTGGTG	GTG TG
yPred17	GAGGTGGTGGTGGTGGGG	595	3	514	0.93	GTGGTG	GTG TG
yPred18	TGTGGTGAGGGGAGGGG	665	3	53	0.10	TGA G	GGGG A
yPred19	GGTGTGGTGGTGGTGGGC	580	4	65	0.12	GTGGTG	GTG TG
yPred20	CGTGTGGTTTGGGGGGG	685	3	50	0.09	GT TGG	TTG GG

**Table S12.** Test-tube investigation of red and yellow-orange NCBs near Rank3600.

As we selected top 30% (3,600) activator sequences as the bright class, we used the median enhancement ratio value from Rank3595 to Rank3600 sequences as our new criteria for bright/dark categorization, which corresponded to 145 and 66 for red and yellow-orange channels, respectively.

**A. Red channel**

<b>ID</b>	<b>Sequence (5' → 3')</b>	<b>Emission peak (nm)</b>	<b>Enhancement ratio (red)</b>	<b>Improvement ratio (compared to G12)</b>
Rank3596	CTCGAAGGTGGGGTGGGG	650	83	0.19
Rank3597	TGGAAAGGTGGGGTGGGG	600	239	0.54
Rank3598	CGTAGTGGTGGGGTGGGG	670	233	0.53
Rank3599	GAACCCGGTGGGGTGGGG	570	143	0.33
Rank3600	GGGTGGGGTGGGGGTGGA	550	145	0.33

**B. Yellow-orange channel**

<b>ID</b>	<b>Sequence (5' → 3')</b>	<b>Emission peak (nm)</b>	<b>Enhancement ratio (yellow-orange)</b>	<b>Improvement ratio (compared to G15)</b>
Rank3596	GGGTGGGGTGGGTCAATC	655	55	0.10
Rank3597	CGAAGCGGTGGGGTGGGG	600	210	0.38
Rank3598	AAACCGGGTGGGGTGGGG	685	66	0.12
Rank3599	GGGTGGATGGCAGTGGGG	590	61	0.11
Rank3600	GGGTGGTGCAGCGTGGGG	615	182	0.33

**Table S13.** Machine learning model prediction results.

In this report, we evaluated the predictability across various machine learning models, including logistic regression<sup>[10]</sup> (LR), linear discriminant analysis<sup>[11]</sup> (LDA), decision tree<sup>[12]</sup> (DT), AdaBoost<sup>[13]</sup> (ADA), and support vector machines<sup>[14]</sup> (SVM). We observed that after feature selection using Weka, LR revealed the best predictability for both the red channel (accuracy: 0.87; marked in red) and yellow-orange channel (accuracy: 0.89; marked in yellow-orange) based on 5-fold cross validation. All the models were built based on “bright yellow-orange vs. dark yellow-orange classification” and “bright red vs. dark red classification” separately.

<b>Model (red)</b>	<b>Accuracy (Acc)</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>Positive prediction rate</b>	<b>Negative prediction rate</b>
<b>LR</b>	0.87	0.88	0.85	0.90	0.83
<b>LDA</b>	0.86	0.86	0.85	0.90	0.81
<b>DT</b>	0.84	0.86	0.80	0.86	0.80
<b>ADA</b>	0.86	0.87	0.85	0.90	0.83
<b>SVM</b>	0.86	0.86	0.85	0.90	0.81

<b>Model (yellow-orange)</b>	<b>Accuracy (Acc)</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>Positive prediction rate</b>	<b>Negative prediction rate</b>
<b>LR</b>	0.89	0.88	0.90	0.92	0.85
<b>LDA</b>	0.87	0.85	0.90	0.92	0.82
<b>DT</b>	0.83	0.84	0.82	0.86	0.79
<b>ADA</b>	0.87	0.86	0.88	0.90	0.83
<b>SVM</b>	0.88	0.87	0.90	0.92	0.84

**Table S14.** Selected bright and dark features for yellow-orange channel.

We defined the top 30% as the bright class and bottom 30% as the dark class. The feature extraction was processed using MERCI. We then used Weka to selected important features. The attribute evaluator was set to “CfsSubsetEval”<sup>[4]</sup> and the search method was set to “GreedyStepwise”.<sup>[3]</sup> All other parameters were set to default values.

**A.** Selected bright features (the number indicates the segment that the motif belongs to).

3-nt	4-nt	5-nt	6-nt
	AAGG_1	AG_AG_1	AAG_GT_1
	AGGG_1	CG_AG_1	AAG_GT_3
	AGGG_3	GAGGT_1	CGGG_T_1
	ATGG_1	G_AAG_1	GAGG_G_2
	CGGG_1	G_ATG_1	GA_GGG_2
	GAGG_1	GC_AG_1	GA_GGG_3
	GCGG_1	GCC_G_3	GA_TGG_1
	GTGT_1	GCT_G_3	G_AGTT_1
	TAGG_1	G_CTG_1	GCG_GG_2
	TTGG_1	G_GGGG_2	GC_TGG_1
		GT_AG_1	G_CGGT_1
		GTT_G_1	GGA_GG_1
		G_TAG_1	GGC_GG_1
		G_TGT_1	GGC_GG_2
		TGA_G_1	GGGA_A_1
		TGA_G_2	GGGA_G_1
		TG_AG_1	GGGG_A_1
		T_GCG_3	GGGG_A_3
		TT_GG_1	GG_GCG_1
			G_GGCA_1
			GTGGTG_2
			GTGTGG_3
			GTG_TG_1
			GTG_TG_3
			GT_TGG_1
			GT_TGG_2
			TGGGTG_1
			TGGGTG_2
			TG_GCG_1
			TTG_GG_2
			TTG_GG_3
			TT_GGT_3

**B. Selected dark features (the number indicates the segment that the motif belongs to).**

3-nt	4-nt	5-nt	6-nt				
AAT_2	ACGT_3	AA_CG_2	AA_TGG_1				GG_GTC_1
ACT_2	CCGG_1	AA_TG_1	AA_TGG_3				GGTGGA_1
ATC_1	GGAT_1	A_AGT_1	A_AGTG_1				GGTGGA_2
ATC_2	GGTT_1	AC_CG_2	A_AGTG_2				G_GGAA_1
ATT_2	TCGG_1	AC_TG_1	ACG_GG_1				G_GGAC_1
CAT_2		AT_CG_2	AC_GGT_1				G_GGAT_1
CGC_2		AT_CG_3	AC_GTG_3				G_GGCT_1
CTC_2		ATG_T_1	AC_TGG_1				G_GGTA_1
CTT_3		AT_TG_1	A_CGGT_1				G_GGTC_1
TCA_3		C_AGT_1	ATG_TG_1				G_GGTT_1
TCT_2		C_AGT_2	AT_GGT_1				T_AGTG_2
TTC_2		CCG_T_1	AT_GTG_1				TCGG_G_1
TTT_3		CC_TG_3	CAG_TG_1				TCG_GG_1
		CTC_G_2	CA_GGT_1				TC_TGG_1
		CT_TG_1	CA_TGG_1				T_CGGT_1
		C_TGT_1	C_AGGT_1				TGGA_A_1
		C_TGT_2	C_AGTG_2				TGGA_C_3
		GAT_G_1	CCG_TG_1				TGGA_G_1
		GT_CT_1	CC_GGT_1				TGG_AA_1
		GT_CT_3	CC_TGG_1				TGG_CT_1
		GT_TC_1	CGTGGG_1				TGGT_A_1
		G_TCG_1	CGTGGG_2				TGGT_A_3
		TA_TG_1	CT_GGT_1				TGGT_C_1
		TA_TG_3	CT_TGG_1				TGGT_T_3
		T_AGT_1	C_TGGT_1				TGG_TT_2
		TC_GT_3	C_TGTG_1				TG_GAA_1
		TC_TG_1	C_TGTG_2				TG_GAC_1
		TG_AT_3	G_CGTG_1				TG_GAT_1
		TG_TT_1	G_CGTG_2				TG_GCC_1
		TG_TT_3	GGG_CC_1				TG_GTA_1
		T_GTT_1	GGGT_A_1				TG_GTC_1
		TT_CG_3	GGGT_C_1				TG_GTT_1
		T_TGT_1	GGGT_T_1				T_GGAC_1
							T_GGCC_1
							T_GGCT_1
							T_GGTA_1
							TTG_TG_1

**Table S15.** Selected bright and dark features for red channel.

We defined the top 30% as the bright class and bottom 30% as the dark class. The feature extraction was processed using MERCI. We then used Weka to selected important features. The attribute evaluator was set to “CfsSubsetEval” and the search method was set to “GreedyStepwise”. All other parameters were set to default values.

**A.** Selected bright features (the number indicates the segment that the motif belongs to).

3-nt	4-nt	5-nt	6-nt				
CCC_1	AAGG_1	AA_GG_1	AA_GGT_2			CCT_G_3	TA_GGT_2
CCC_3	ATGG_1	AA_GG_3	ACG_GT_1			C_CAG_3	TC_GGT_1
	CTGG_3	A_AGG_3	ACG_GT_3			C_CGG_1	TC_GGT_3
	GAGG_1	ACA_G_1	AC_GGT_2			C_CGG_2	TGGGGG_1
	GCGG_1	AC_AG_3	AC_GGT_3			C_CTG_1	TGGGGG_3
	TAGG_1	ACC_G_1	AG_GGT_1			C_CTG_3	TGGGTG_1
	TTGG_1	AC_GG_2	AG_GGT_2			C_GGG_1	TGGGTG_3
		ACT_G_1	ATG_GT_3			CT_AG_1	TTGG_G_2
		ACT_G_3	A_TGGT_2			CT_AG_3	TTG_GT_3
		A_CCG_1	A_TGGT_3			CTC_G_1	T_TGGT_2
		A_CCG_3	CAG_GT_3			CTC_G_3	T_TGGT_3
		A_CGG_1	CCG_GT_3			CT_CG_1	
		A_CGG_3	CC_GGT_2			CT_CG_3	
		AT_GG_1	CG_GGT_2			CT_GG_2	
		CAC_G_1	CT_GGT_2			GCC_G_1	
		CAC_G_3	GA_GGT_2			GCC_G_3	
		CA_CG_1	GA_GGT_3			GC_GG_3	
		CA_CG_3	G_AGGT_1			GGGAC_3	
		CA_GG_1	GC_GGT_2			GGGGA_1	
		CA_GG_2	GC_GGT_3			GGGGA_2	
		C_AGG_2	GGGG_C_3			GGGGA_3	
		CCA_G_1	GGG_GC_1			GGGGG_1	
		CC_AG_3	GGG_GC_2			TA_GG_3	
		CC_CG_3	GTGGTG_1			T_AGG_3	
		CC_GG_1	GTGGTG_3			TCC_G_1	
		CCT_G_1	TAG_GT_3			TC_CG_1	
						TC_CG_3	
						TC_GG_3	
						T_CCG_3	
						TT_GG_1	
						TT_GG_3	

**B. Selected dark features (the number indicates the segment that the motif belongs to).**

3-nt	4-nt	5-nt	6-nt
AAA_1	CCGT_1	AA_TG_1	AAGT_G_1
AGA_1	CGTG_3	AGT_G_1	AA_GTG_2
AGA_3	GAGT_1	AGT_G_3	ACG_GG_2
ATA_1	GGAA_1	A_TGT_1	AC_TGG_1
TAA_1	GGAG_1	CGT_G_1	AC_TGG_2
TGA_1	GGAT_1	CGT_G_2	A_TGGG_2
	GGCA_1	C_TGT_1	CAG_GG_1
	GGCC_1	GAA_C_1	CAG_GG_2
	GGCT_1	GAA_G_3	CAGT_G_1
	GGTA_1	GA_CT_1	CA_TGG_2
	GGTT_1	G_AGC_1	CG_GGG_1
		GCA_G_3	CG_GGG_2
		GCT_G_3	GCG_TG_1
		GC_TC_1	GCG_TG_2
		G_GCC_1	GC_TGG_1
		G_GTC_1	GC_TGG_2
		GTA_G_1	GGGA_A_1
		GT_AT_1	GGG_CG_1
		GT_CA_1	GGGT_A_1
		GT_CC_1	GG_GTA_2
		GT_CT_1	GG_GTT_2
		GTG_C_2	GGTGGT_1
		GTGGC_1	GGTGGT_2
		GT_GA_2	GGTG_T_1
		G_TTA_1	GGT_TC_1
		G_TTG_1	GTG_AT_1
		TAGTG_1	GTGG_C_2
		TC_TG_1	GTGGGC_3

		TC_TG_2	GTG_TA_1
		TG_AA_3	GT_GTA_1
		TG_CA_1	GT_GTT_1
		TG_CG_1	TAG_GG_2
		TG_CT_1	TGG_AA_2
		TGTGG_2	TGG_AC_2
		TG_TA_3	TGG_AC_3
		T_GAA_1	TGG_AG_2
		T_GAC_3	TGG_AT_2
		T_GAG_3	TGGGTT_1
		T_GAT_3	TGGGTT_3
		T_GCT_1	TGGT_A_1
		T_GTA_1	TGG_TC_3
		T_GTC_1	TG_GAC_1
		T_GTC_3	TG_GAG_1
		TT_AG_1	TG_GCC_1
		TTC_G_3	TG_GCG_3
		TT_TG_1	T_GGAA_3
			T_GGTA_3
			T_TGGG_2
			T_TGGG_3

**Discussion:** a possible mechanism for zipper bag model.

In our model, the zipper may not necessarily be formed by the Watson-Crick (WC) basepairs – it can also be formed by the silver-mediated pairs (e.g., the C-Ag<sup>+</sup>-C pair).<sup>[15]</sup> Upon close examination, we believe the yellow-orange POT zippers at positions 4-6 are caused by a WC pair GC or a wobble pair GT (where G is on the activator and C or T is on the NC probe), as disrupting the GC or GT pair at these positions often creates dim yellow-orange NCB samples (**Figure S26B**). As aforementioned, these dim yellow-orange samples attribute to less emissive AgNC chromophores and lower chemical yield of the chromophores. In contrast, the red POT zippers at positions 2-4 may be formed by a silver-mediated pair C-Ag<sup>+</sup>-C. Disrupting the C-Ag<sup>+</sup>-C pair at these positions darkens red NCB samples (**Figure S25B**). One recent report showed evidence that silver-mediated heteroduplexes (e.g., C<sub>11</sub>-Ag<sup>+</sup><sub>N</sub>-T<sub>11</sub>) could be less stable than their homoduplex counterparts (e.g., C<sub>11</sub>-Ag<sup>+</sup><sub>N</sub>-C<sub>11</sub>),<sup>[15a]</sup> while another report showed a Ag<sup>+</sup>-mediated interaction at a place away from the AgNC core,<sup>[16]</sup> both supporting the hypothesis behind our model.

Whereas it was still difficult to conclude the exact silver stoichiometries in the original duplexes, introducing octylamine to the MS spray solution led to single-stranded species with reproducible numbers of silver atoms on them. The C55 NC probes from yPOT5 NCB and rPOT5 NCB carried 0-3 and 0-4 silver atoms, while the activators yPOT5 and rPOT5 carried 0-7 and 0-6 silver atoms, respectively. These results could possibly indicate that the original silver stoichiometry for the intact yPOT5 NCB may be larger than that of the intact rPOT5 NCB. Although our results are contradictory to previous investigations that suggest red AgNC chromophores have a larger core,<sup>[17]</sup> we predict that the red chromophores featured in this study have a larger footprint in the activator/C55 bag, owing to different AgNC shapes and DNA conformations.<sup>[18]</sup> In future studies, we aim to pinpoint the binding sites of AgNC chromophores within the activator/NC probe duplexes by employing activated-electron photodetachment mass spectrometry (a-EPD MS), a structural characterization technique and tandem MS<sup>n</sup> method that was previously shown to reveal the binding sites of AgNCs in shorter single-stranded DNA hosts (up to 28-nt long).<sup>[18c]</sup> Alternatively, X-ray crystallography of DNA-templated AgNCs can reveal not only the binding sites but also the binding geometries of surrounding bases to AgNCs,<sup>[16, 18b, 18d, 18e]</sup> provided that NCBs can be crystallized.

## References:

- [1] A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, N. Stuurman, *J Biol Methods* **2014**, 1.
- [2] C. Vens, M. N. Rosso, E. G. Danchin, *Bioinformatics* **2011**, 27, 1231.
- [3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, *ACM SIGKDD Explorations Newsletter* **2009**, 11, 10.
- [4] M. A. Hall, L. A. Smith, Vol. Volume 20 No 1, Springer, Conference held at Perth 1998.
- [5] a)E. S. Ristad, P. N. Yianilos, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1998**, 20, 522; b)S. M. Copp, S. M. Swasey, A. Gorovits, P. Bogdanov, E. G. Gwinn, *Chemistry of Materials* **2019**, 32, 430.
- [6] a)A. C. McGinnis, E. C. Grubb, M. G. Bartlett, *Rapid Commun Mass Spectrom* **2013**, 27, 2655; b)M. Scalabrin, M. Palumbo, S. N. Richter, *Anal Chem* **2017**, 89, 8632; c)J. M. Sutton, M. G. Bartlett, *Rapid Commun Mass Spectrom* **2020**, 34, e8696; d)J. M. Sutton, N. M. El Zahar, M. G. Bartlett, *J Am Soc Mass Spectrom* **2021**, 32, 497.
- [7] a)H. C. Yeh, J. Sharma, J. J. Han, J. S. Martinez, J. H. Werner, *Nano Letters* **2010**, 10, 3106; b)J. M. Obliosca, M. C. Babin, C. Liu, Y. L. Liu, Y. A. Chen, R. A. Batson, M. Ganguly, J. T. Petty, H. C. Yeh, *ACS Nano* **2014**, 8, 10150; c)H. C. Yeh, J. Sharma, M. Shih Ie, D. M. Vu, J. S. Martinez, J. H. Werner, *Journal of the American Chemical Society* **2012**, 134, 11550; d)Y. A. Chen, J. M. Obliosca, Y. L. Liu, C. Liu, M. L. Gwozdz, H. C. Yeh, *Journal of the American Chemical Society* **2015**, 137, 10476; e)J. Zhang, C. Li, X. Zhi, G. A. Ramon, Y. Liu, C. Zhang, F. Pan, D. Cui, *Analytical Chemistry* **2016**, 88, 1294.
- [8] C. Jung, J. A. Hawkins, S. K. Jones, Jr., Y. Xiao, J. R. Rybarski, K. E. Dillard, J. Hussmann, F. A. Saifuddin, C. A. Savran, A. D. Ellington, A. Ke, W. H. Press, I. J. Finkelstein, *Cell* **2017**, 170, 35.
- [9] a)S. M. Copp, P. Bogdanov, M. Debord, A. Singh, E. Gwinn, *Advanced Materials* **2014**, 26, 5839; b)S. M. Copp, A. Gorovits, S. M. Swasey, S. Gudibandi, P. Bogdanov, E. G. Gwinn, *ACS Nano* **2018**.
- [10] S. Menard, **2002**.
- [11] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K. R. Mullers, *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)* **1999**, 41.
- [12] S. R. Safavian, D. Landgrebe, *IEEE Transactions on Systems, Man, and Cybernetics* **1991**, 21, 660.
- [13] Y. Freund, R. E. Schapire, *Journal of Computer and System Sciences* **1997**, 55, 119.
- [14] C. Cortes, V. Vapnik, *Machine Learning* **1995**, 20, 273.
- [15] a)S. M. Swasey, L. E. Leal, O. Lopez-Acevedo, J. Pavlovich, E. G. Gwinn, *Sci Rep* **2015**, 5, 10163; b)J. Kondo, Y. Tada, T. Dairaku, Y. Hattori, H. Saneyoshi, A. Ono, Y. Tanaka, *Nat Chem* **2017**, 9, 956.
- [16] C. Cerretani, J. Kondo, T. Vosch, *RSC Advances* **2020**, 10, 23854.
- [17] S. M. Copp, D. Schultz, S. Swasey, J. Pavlovich, M. Debord, A. Chiu, K. Olsson, E. Gwinn, *J Phys Chem Lett* **2014**, 5, 959.
- [18] a)D. Schultz, K. Gardner, S. S. Oemrawsingh, N. Markesevic, K. Olsson, M. Debord, D. Bouwmeester, E. Gwinn, *Adv Mater* **2013**, 25, 2797; b)C. Cerretani, H. Kanazawa, T. Vosch, J. Kondo, *Angew Chem Int Ed Engl* **2019**; c)M. S. Blevins, D. Kim, C. M. Crittenden, S. Hong, H. C. Yeh, J. T. Petty, J. S. Brodbelt, *ACS Nano* **2019**, 13, 14070; d)D. J. E. Huard, A. Demissie, D. Kim, D. Lewis, R. M. Dickson, J. T. Petty, R. L. Lieberman, *J Am Chem Soc* **2019**, 141, 11465; e)C. Cerretani, J. Kondo, T. Vosch, *CrystEngComm* **2020**, 22, 8136.