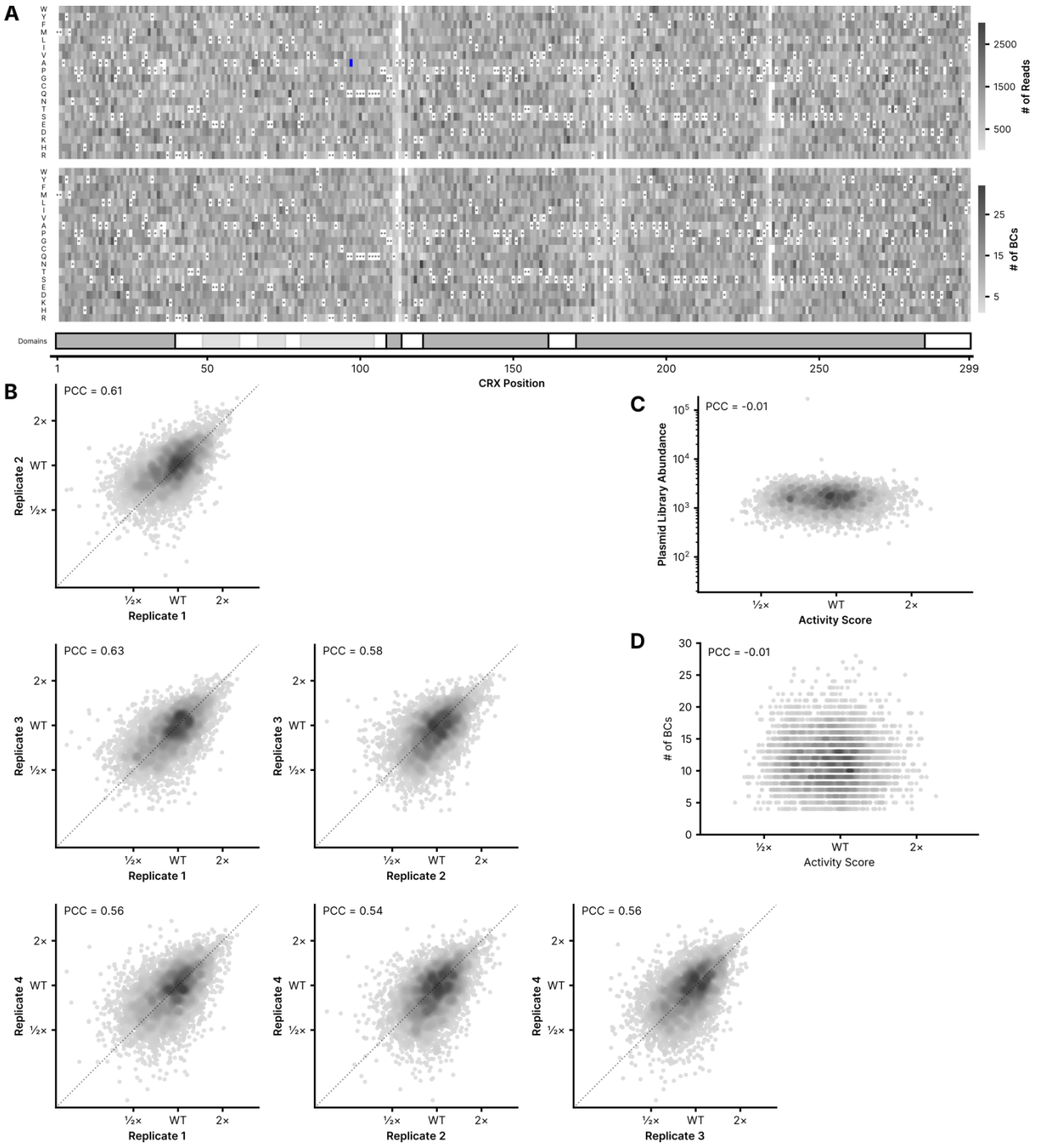
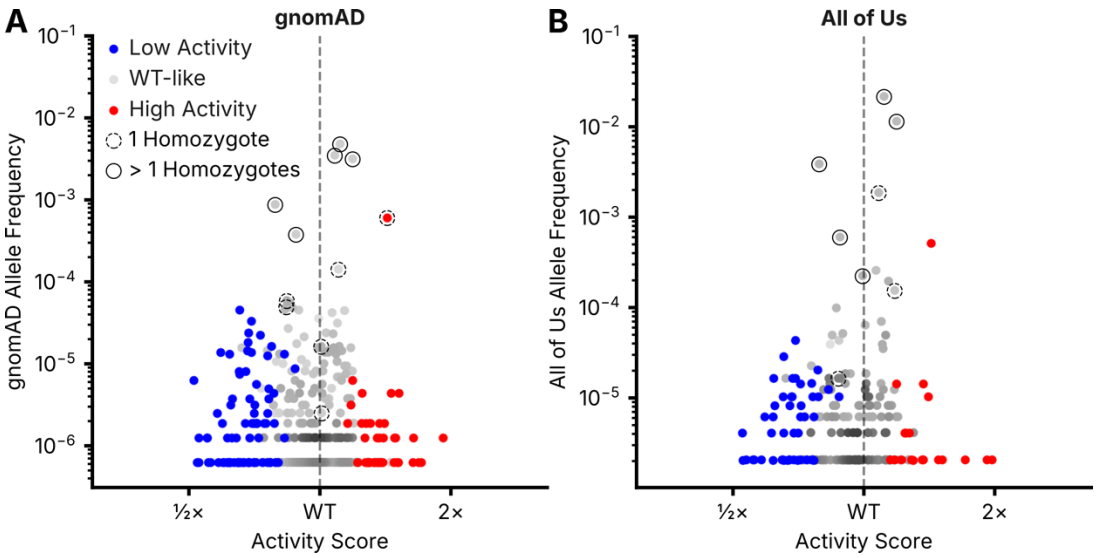
SUPPLEMENTAL FIGURES

Supplemental Figure S1: CRX expression in Landing Pad + Reporter cells. Western blot for CRX in HEK 293-derived LP + Reporter cells, either with wild-type CRX integrated in the landing pad (“LP + CRX”) or no CRX integrated, demonstrating a lack of expression of the endogenous copy of CRX.

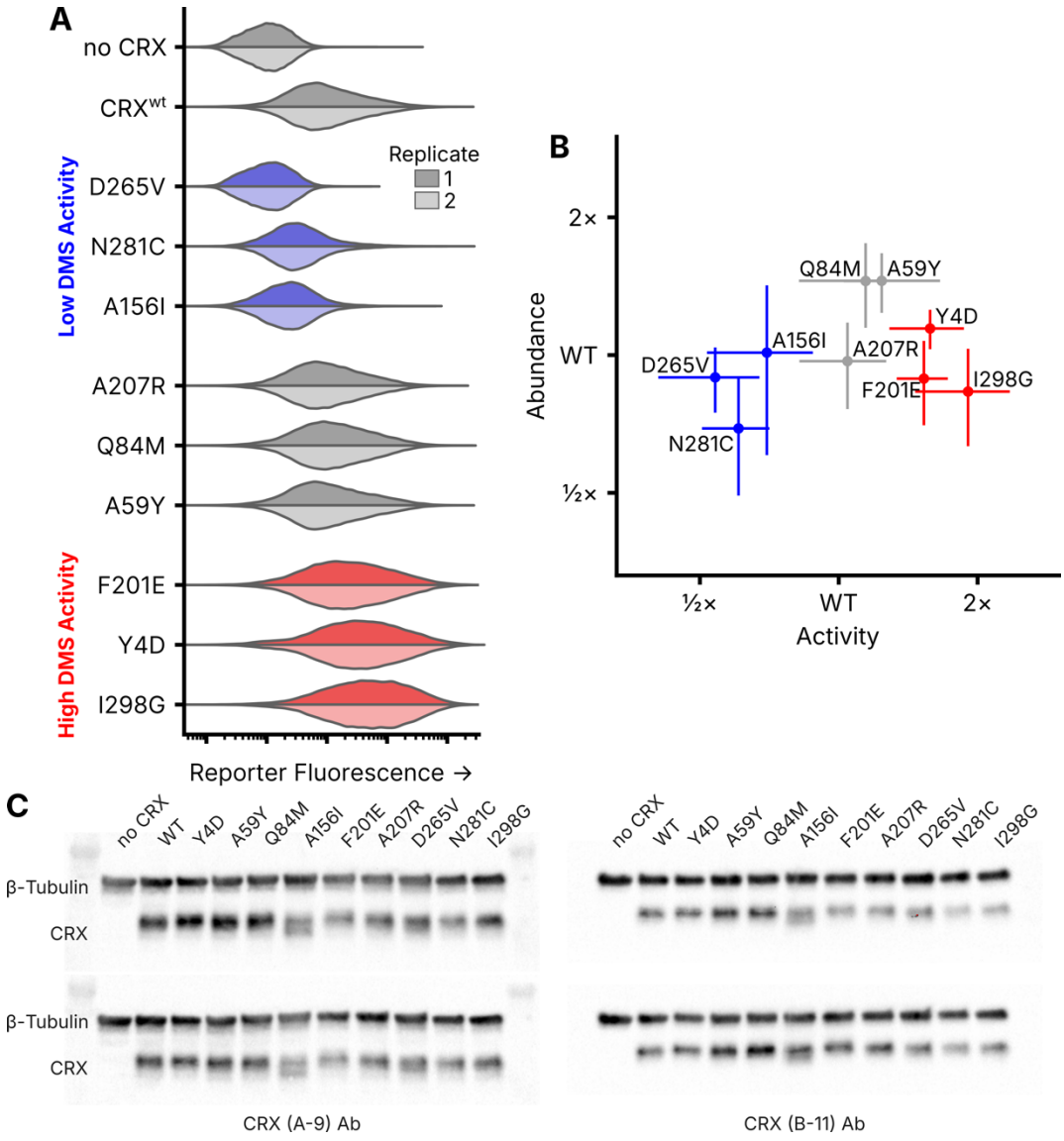


Supplemental Figure S2: Library construction and quality control. (**A**) Number of reads (top) and barcodes (bottom) per variant as determined by PacBio long read sequencing of the cloned *CRX* variant library. The blue box indicates a single variant (Q97A; 10,305 reads) which was overrepresented in the library cloning process and is clipped in the colormap. Wild-type CRX was measured with 426,393 reads and 5,514 barcodes. (**B**) Variant activity score reproducibility across four replicates; DMS activity scores in the main text are a per-variant average of these four replicates. The dotted line

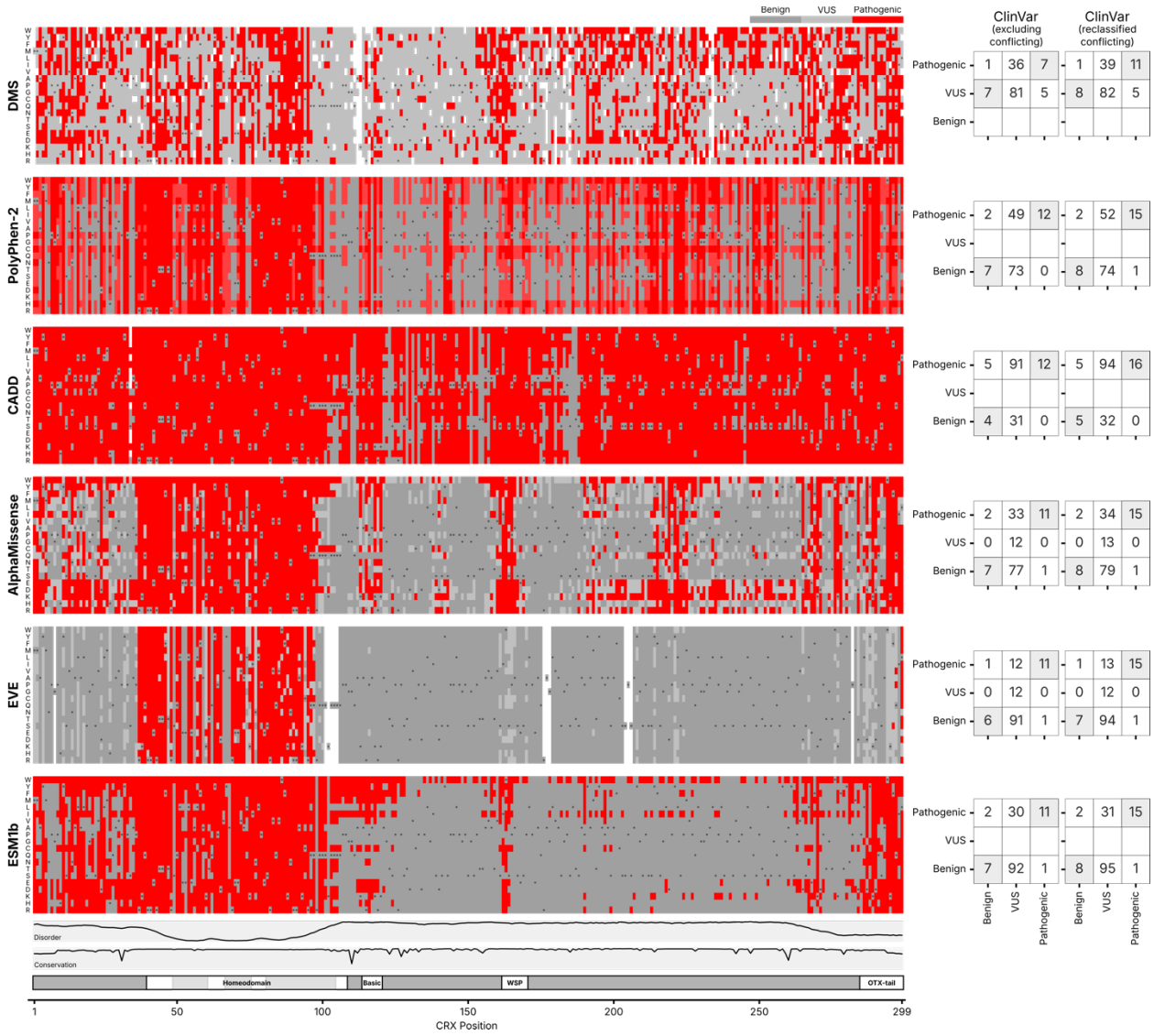
in each plot is the line y=x, for visual reference. (**C**) Correlation between per-variant DMS activity scores and variant abundance in the input plasmid library. (**D**) Correlation between per-variant DMS activity scores and number of variant barcodes in the input plasmid library.



Supplemental Figure S3: Population allele frequency of CRX variants. (**A, B**) DMS activity score plotted versus variant allele frequency of CRX variants reported in gnomAD (v4.1, **A**) and All of Us (accessed 2024-06, **B**). Variants with a single homozygous individual reported are circled with a dotted line; variants with multiple homozygous individuals reported are circled with a solid line. Variants are colored as in Figure 3D.



Supplemental Figure S4: Per-variant activity and abundance relative to wild-type CRX. (**A**) Fluorescent reporter measurements (arbitrary units) of individual CRX variants integrated into the LP+Reporter cell line, separated by their classification in the DMS assay. Two independent biological replicate experiments per sample (distributions plotted from 40,000 cells). (**B**) Protein abundance of the variants in (A) measured by western blot quantification relative to β-tubulin (arbitrary units). DMS activity scores plotted versus mean abundance from four replicate blots with abundance normalized to wild-type CRX across replicate blots; standard deviations shown as cross bars. (**C**) Source western blot images for the variant abundance measurements shown in (B).



Supplemental Figure S5: Comparison of CRX DMS results with computational variant effect predictors. For each method, classifications of all scored variants are shown in the heatmap on the left and classifications compared to ClinVar (for the subset of variants in ClinVar scored by the method) are shown in the tables on the right (either excluding conflicting variants or with manual reclassification). In the ClinVar tables, rows are left empty if the method does not produce the corresponding classification. For the DMS assay, any variant with an activity score significantly altered from wild-type CRX is called as “Pathogenic”, and other variants are called “VUS”. For PolyPhen-2, model-produced “HumVar” classifications are shown, with variants in the “possibly damaging” class in light red. For CADD, the recommended Phred score pathogenicity cutoff of 20 was used. For AlphaMissense, model-produced classifications are shown. For EVE, the “75_pct_retained_ASM” classifications were used. For ESM1b, the recommended pathogenicity cutoff of -7.5 was used. Disorder, domains, and conservation are shown as in Figure 1.

SUPPLEMENTAL FILES

Supplemental Table S1: Oligo sequences used in this study.

Supplemental Table S2: Plasmids used in this study.

Supplemental Table S3: Per-variant DMS activity scores.

Supplemental Movie S1: Average DMS activity scores superimposed on a predicted structure of the CRX homeodomain in complex with DNA. An animated 360° rotation of the structure shown in Figure 4. A cartoon ribbon model is shown on the left and a space-filling atomic model is on the right.

Supplemental Interactive S1: DMS activity scores for all measured single amino acid CRX substitutions. Activity scores were normalized to wild-type CRX; the wild-type amino acid at each position is indicated by the gray circle in each column. The average row shows the mean activity score for all substitutions at each position. Disorder, domains, and conservation are shown as in Figure 1. Empty boxes indicate variants not measured in the DMS assay, due to drop-out during the library cloning or variant measurement steps. Scores for specific variants or positions can be visualized by hovering the mouse over the position of that variant.

Supplemental Data S1: Fluorescence distributions for the CRX DMS library integrated in LP+Reporter cells. Fluorescence distributions and gating strategies for the four replicates of the DMS library as measured reported by the Sony SY3200 fluorescence-activated cell sorter.

Supplemental Code: Scripts and software used for the analyses presented in this manuscript.

SUPPLEMENTAL METHODS

For a list of primers and plasmids used in this study, please see Supplemental Table S1 and S2, respectively.

Generation of the landing pad cell line

All cell lines were cultured in 90% DMEM (Gibco #11965092), 10% heat-inactivated fetal bovine serum (Gibco #16140089) supplemented with Penicillin-Streptomycin (Gibco #15140122). Unless otherwise noted, all transfections were carried out with cells plated at a density of 6×10^5 cells/well in a six-well plate 24 hours prior to transfection. Each well of cells was transfected with the indicated plasmids and 4 µL each of Lipfectamine 3000 and P3000 reagent (Invitrogen #L3000015), suspended in 250 µL Opti-MEM (Gibco #31985062). For all transfections, the Lipofectamine 3000 was mixed with half the volume of Opti-MEM, while the P3000 reagent and DNA were combined in the remaining Opti-MEM. The Opti-MEM was mixed, and incubated at room temperature for 15 min, before being added dropwise to the prepared cells. All flow cytometry was performed on a Cyoflex S flow cytometer (Beckman-Coulter, V4-B2-Y4-R3 model).

In experiments where constructs were integrated into the landing pad, cells were transfected with a mixture of the indicated constructs and Addgene #11916, a plasmid expressing Cre recombinase. Three days post-transfection, media was replaced with fresh media supplemented with 5 nM AP1903/Rimiducid (MedChemExpress #HY-16046), which selects against the iCasp9 inducible caspase present in the naïve landing pad (Straathof et al. 2005). After 24 hours, media was replaced again, with fresh media supplemented with 5 nM AP1903 and 1 µg/mL puromycin (Sigma-Aldrich #P8833), which selects for the presence of puromycin N-acetyltransferase carried on the integrated

construct. Surviving cells were allowed to recover and expand, with replating as needed to normalize cell density. Following integration, cells were continually grown in media supplemented with 5 nM AP1903 and 1 µg/mL puromycin.

To generate the landing pad cell line, HEK 293 cells were co-transfected with with 400 ng pJLS83, a plasmid carrying the landing pad sequence, and 400 ng Addgene #105927, a plasmid expressing Cas9 and an sgRNA targeting the human Rosa26 safe harbor locus. Eight days post-transfection, individual cells were sorted into single wells of multiple 96-well plates pre-filled with 150 µL of a 1:1 mixture of fresh media and 0.22 µm-filtered conditioned media exposed to wild-type HEK 293 cells in culture for 72 hours. The sort was performed on a Cytoflex SRT (Beckman-Coulter, V5-B2-Y5-R3 model); sorted cells were gated for positive blue channel (B450) fluorescence to select for expression of the mTagBFP2 fluorescent protein carried on the landing pad construct.

Over the course of three weeks, individual clones were allowed to expand and screened to identify lines with low-variance mTagBFP2 expression and normal growth characteristics and morphology. To screen out lines with integration of the landing pad construct on multiple alleles, clones were co-transfected with 600 ng Addgene #11916, a plasmid expressing Cre recombinase, and a 1:1 mixture of 200 ng pJLS119 and 200 ng pJLS120, plasmids carrying the mEmerald and mCherry2 fluorescent proteins, respectively, along with the necessary recombinase sites for landing pad recombination. Cells were measured by flow cytometry after AP1903+puromycin purification of LP-integrated cells, and any clones yielding cells exhibiting both green and red fluorescence were discarded. The final validated clonal landing pad cell line was frozen at low passage numbers, in CryoStor CS10 freezing medium (Biolife Solutions #210102).

Generation of the CRX reporter cell line

A synthetic CRX reporter was synthesized, carrying eight repeats of a strong consensus CRX binding motif, CTAATCCC, each padded with a two base pair spacer motif (AG) (White et al. 2016). This reporter sequence was cloned upstream of the CMV-T6 minimal promoter (Loew et al. 2010) and the mEmerald fluorescent protein to produce pJLS96.

To produce lentivirus carrying the reporter, HEK 293T cells were seeded in a 10 cm dish, with 10 mL of media containing cells at a concentration of 3×10^5 cells/mL. Cells were transfected with a mixture of 8 µg Addgene #12260, 1 µg Addgene #12259, 1 µg pJLS96, and 40 µL PEI in 500 µL Opti-MEM. 72 hours post-transfection, media was aspirated and concentrated using Lenti-X Concentrator (TaKaRa #631231) following the manufacturer's recommended protocol. To create a "dead" reporter control, a second batch of lentivirus was produced using pJLS97, a variant of pJLS96 in which the CRX motifs were replaced with a variant known to abrogate CRX binding (CTACTCCC) (White et al. 2016).

Landing pad cells were separately transduced with the pJLS96 and pJLS97 lentiviruses. After 72 hours, the bulk populations of transduced cells were transfected with 400 ng pJLS38, a plasmid expressing human CRX. Three days post-transfection, individual pJLS96-transduced cells were sorted into single wells of multiple 96-well plates pre-filled with 150 µL of a 1:1 mixture of fresh medium and 0.22 µm-filtered conditioned media exposed to wild-type HEK 293 cells in culture for 72 hours. The sort was performed on a Cytoflex SRT (Beckman-Coulter, V5-B2-Y5-R3 model); sorted cells were gated for positive blue channel fluorescence to select for presence of the landing pad construct and

for positive green fluorescence, relative to pJLS38-transfected cells transduced with the pJLS97 “dead” reporter construct.

Over the course of three weeks, individual clones were allowed to expand and screened to identify lines with low-variance mTagBFP2 expression and normal growth characteristics and morphology. Selected lines were co-transfected with 400 ng Addgene #11916 and 400 ng of either pJLS99, pJLS100, or pJLS101, plasmids carrying wild-type CRX, the p.R90W hypomorphic CRX variant, or miRFP670 (a transcriptionally inactive control protein), respectively. Following successful isolation of integrated cells, cells were screened by flow cytometry to maximize the dynamic range of mEmerald fluorescence between the pJLS99- and pJLS100-transfected cells. The final validated clonal landing pad + reporter cell line was frozen at low passage numbers, in CryoStor CS10 freezing medium.

Cloning the CRX variant library

A DNA library comprising all possible human CRX single residue substitution variants was ordered as a combinatorial variant library from Twist Bioscience, with synthesis of a designed codon for each variant. The library construct was ligated between the AflII (NEB #R0520) and NheI (NEB #R3131) sites in pJLS84v2, a plasmid carrying recombinase sites matching the landing pad. The library fragment and plasmid backbone were mixed at a 3:1 molar ratio, targeting a total DNA amount of approx. 100 ng. The ligation was performed with T4 Ligase (NEB #M0202) at 16°C for 16 hours.

The ligation product was purified using a Monarch PCR & DNA Cleanup Kit (NEB #T1030) following the manufacturer’s recommended protocol, with elution in 6 µL H2O. The purified ligation product was transformed into 10-beta Electrocompetent *E. coli* (NEB

#C3020). Briefly, 50 µL of electrocompetent cells were mixed with 5 µL purified ligation product on ice, and then split into two pre-chilled electroporation cuvettes (0.1 cm gap, Bio-Rad #1652089). Transformations were performed using a Bio-Rad GenePulser Xcell Electroporation System with PC Module (Bio-Rad #1652662), with the following conditions: 2000 V, 200 Ω, and 25 µF.

Following electroporation, 975 µL of pre-warmed 10-beta/Stable Outgrowth Medium was added (NEB #B9035) to each cuvette, and the outgrowths were pooled and transferred to a 14 mL round-bottom culture tube (Falcon #352059) for incubation with agitation at 37°C for 1 hour. 1:1000 and 1:10,000 dilutions of the outgrowth were plated on LB plates carrying 50 µg/mL kanamycin (LB+Kan50, Sigma-Aldrich #K1377) and incubated at 37°C for 16 hours. 1700 µL of the outgrowth was inoculated into 200 mL LB+Kan50 and incubated, with agitation, at 37°C for approx. 10 hours, until an OD600 of 3.0 was reached. Based on colony counts from the dilution plates, the 200 mL culture was inoculated with approx. 5.4 million colony-equivalents.

The 200 mL culture was centrifuged at 3100 ×g for 10 min to pellet the E. coli, and plasmid DNA was purified from the resulting cell pellet using a GenElute HP Plasmid Maxiprep Kit (Sigma-Aldrich #NA0310) following the manufacturer's recommended protocol with elution in 3 mL H2O. The purified plasmid library was concentrated using an Eppendorf Vacufuge plus, yielding the "step one" CRX DMS library.

To add barcodes to this library, a short random barcoding oligo ("variant barcode", "vBC") was ordered as an Ultramer DNA Oligo from Integrated DNA Technologies (pJLS84v2+BC1). 1 µg of the step one plasmid library was linearized with XhoI (NEB #R0146) at 37°C for 16 hours. The digest product was run on a 1% agarose gel and the

band was excised and purified using a Monarch DNA Gel Extraction Kit (NEB #T1020) following the manufacturer's recommended protocol. The linearized plasmid library and barcoding oligo were assembled with NEBuilder HiFi DNA Assembly Master Mix (NEB #E2621) at a 1:5 molar ratio, targeting approx. 200 fmol total DNA in the assembly reaction. The assembly reaction was incubated at 50°C for 1 hour, and then purified using a Monarch PCR & DNA Cleanup Kit.

The purified, assembled, barcoded library was transformed following the same protocol as the step one DMS library. The 200 mL culture was inoculated with 40 µL of outgrowth and grown to an OD600 of 2.0 (approx. 13 hours). Based on colony counts from the dilution plates, the 200 mL culture was inoculated with approx. 85,000 colony-equivalents. Plasmid DNA was purified and concentrated using the same procedure as for the step one DMS library, yielding the "step two" CRX DMS library.

To add the puromycin N-acetyltransferase (PAC) cassette to the step two library, a fragment carrying an internal ribosome entry site (IRES) and PAC coding sequence was amplified from pJLS98 using primers JClPr88 and JClPr89. 1 µg of the step two plasmid library was linearized with SpeI-HF (NEB #R3133) at 37°C for 16 hours, and purified following the same protocol as the step two library. The IRES-PAC cassette and linearized step two library were assembled and purified following the same protocol used to generate the step two library.

The purified, assembled, PAC-containing library was transformed following the same protocol used to generate the step one DMS library. The 200 mL culture was inoculated with 1850 µL of outgrowth and grown to an OD600 of 3.5 (approx. 11 hours).

Plasmid DNA was purified and concentrated using the same procedure as for the step one DMS library, yielding the “step three” CRX DMS library.

A second short random barcoding oligo (“random barcode”, “rBC”) was ordered as an Ultramer DNA Oligo from Integrated DNA Technologies (pJLS84v2+BC2). 1 µg of the step three plasmid library was linearized with XhoI at 37°C for 16 hours, and purified following the same protocol as the step two library. The barcoding oligo and linearized step three library were assembled and purified following the same protocol used to generate the step two library.

The purified, assembled, barcoded library was transformed following the same protocol used to generate the step one DMS library. The 200 mL culture was inoculated with 1850 µL of outgrowth and grown to an OD600 of 3.0 (approx. 10 hours). Plasmid DNA was purified and concentrated using the same procedure as for the step one DMS library, yielding the final CRX DMS library.

Associating variants with barcodes

To associate CRX variants with barcodes, the step three CRX DMS library was sequenced using a PacBio Revio long-read sequencer. Briefly, 4 µg of the barcoded plasmid library was linearized by digestion with NruI-HF (NEB #R3192) at 37°C for 16 hours. The linearized library was purified using a Monarch PCR & DNA Cleanup Kit following the manufacturer’s recommended protocol with elution in 20 µL H2O. The purified linearized library was used to prepare a PacBio sequencing library using a SMRTbell prep kit 3.0 (PacBio #102-141-700), following the manufacturer’s recommended protocol (PacBio Protocol #102-166-600 REV02) with the following modifications: DNA shearing was skipped and the 20 µL of purified linearized library was

used directly as input for Repair and A-tailing after the addition of 27 µL of Low TE buffer; the final cleanup with SMRTbell cleanup beads was not performed (no size selection). The prepared library was sequenced on a single Revio SMRT cell.

Reads were aligned to a synthetic reference sequence comprising the expected library plasmid structure with wild-type CRX and "N" nucleotides in place of the expected barcode location using minimap2 (https://github.com/lh3/minimap2, v2.24, with parameters -A2 -B4 -O12 -E2 –end-bonus=13 –secondary=no –cs=long). The resulting PAF file was parsed with a custom Python script (see Supplemental Code, "call_variants.py") to generate a barcode-to-variant map. From 9.5 million total HiFi reads, 63.8% contained a valid barcode and a single missense CRX variant or wild-type sequence. The remaining reads represent a mixture of sequencing errors, low-quality reads, or valid reads of constructs that failed barcoding, acquired indels during cloning, or contained more than a single CRX missense variant. Of 75,946 observed barcodes on full-length plasmid reads, 86.8% mapped to a single CRX variant, while 7.3% mapped to wild-type CRX. Of the remaining barcodes, 4.9% mapped to CRX constructs with more than a single missense variant, while only approx. 1% of barcodes could not be unambiguously assigned—i.e. were observed to co-occur with different variants in different reads. All barcodes not uniquely mapping to a single CRX missense variant or wild-type CRX were discarded.

Measuring variant activity

To conduct the DMS, landing pad + reporter cells were plated at a density of 3 × 10^5 cells/mL in 150 mm dishes, 30 mL per dish, three dishes per biological replicate. 24 hours later, each dish was transfected with 2 µg of the final CRX DMS library, as well as

8 µg Addgene #11916, 50 µL Lipofectamine 3000, and 50 µL P3000, in 3 mL Opti-MEM. Three days later, media was replaced with fresh media supplemented with 5 nM AP1903. Six hours later, media was replaced again with fresh media supplemented with 5 nM AP1903 and 1 µg/mL puromycin. Four days later, media was removed, Accutase (Innovative Cell Technologies #AT104) was added to dissociate cells, and cell suspensions from each of the three dishes for each biological replicate were pooled. Cell suspensions were centrifuged for 5 min at 150 ×g to pellet cells. Pelleted cells were resuspended in fresh media supplemented with 5 nM AP1903 and 1 µg/mL puromycin, and plated in a T-150 flask, one per replicate.

Cells were sorted into four bins based on reporter fluorescence on a Sony SY3200 fluorescence-activated cell sorter, recovering between 500,000 and 3,000,000 cells per bin. Sort distributions and gating strategies for each replicate are shown in Supplemental Data S1. Sorted fractions were pelleted by centrifugation for 5 min at 150 ×g, resuspended in 12 mL fresh media supplemented with 5 nM AP1903 and 1 µg/mL puromycin, and each plated in a fresh T-75 flask. 3–6 days post-sort, cell fractions were harvested by dissociation with 5 mL Accutase, centrifuged, and frozen in 1.5 mL microcentrifuge tubes.

Genomic DNA was extracted from each fraction using a Monarch Genomic DNA Purification Kit (NEB #T3010). Barcode sequences were amplified from gDNA using a two-step protocol: gDNA was first amplified with primers JLSPr141–144 and JLSPr165–168+171+172 using Q5 High-Fidelity DNA Polymerase (NEB #M0491) for 20 cycles with an annealing temperature of 65°C, and then with primers IDT10_i7_NN and IDT10_i5_NN, where NN is replaced with the unique indexing barcode ID, for sample

multiplexing (10 cycles, 65°C annealing temperature). Prepared fraction amplicon libraries were sequenced on an Illumina NovaSeq X Plus instrument in a series of shared 10B flow cells, targeting 100 million reads per replicate.

Reads were cleaned with fastp (https://github.com/OpenGene/fastp, v0.23, with default parameters). Barcodes were extracted, counted, and analyzed using a custom script (see Supplemental Code, “extract_bcs.sh”, “count_bcs.py”, and “analysis.ipynb”). vBCs were mapped to variants using the results of the PacBio long-read sequencing of the barcoded plasmid library. vBCs failing to match an expected barcode from the PacBio sequencing were error-corrected up to a Hamming distance of 1, if and only if they could be unambiguously mapped to an expected variant. All reads with non-mapping or ambiguous vBCs were discarded. vBC-rBC pairs with fewer than three reads were discarded. vBCs with fewer than 10 unique rBCs in any of the four bins were discarded. Raw activity scores were computed by summing the number of rBCs per vBC in each of the four bins divided by the sum of the number of rBCs across all four bins, weighted by the mean fluorescence intensity of each bin. Raw activity scores for each variant were divided by the wild-type CRX activity score, yielding the normalized DMS activity score used throughout this manuscript (Supplemental Table S3).