Supplementary Fig. 1

a

b

of CSAs with >100 predicted binding sites

Design: Full CSA

Design: CSA consensus monomer

chr2 — Full CSA: *n* = 4,392 / Consensus: *n* = 14
chr8 — Full CSA: *n* = 2,599 / Consensus: *n* = 32
chr21 — Full CSA: *n* = 600 / Consensus: *n* = 38
chrY — Full CSA: *n* = 306 / Consensus: *n* = 3
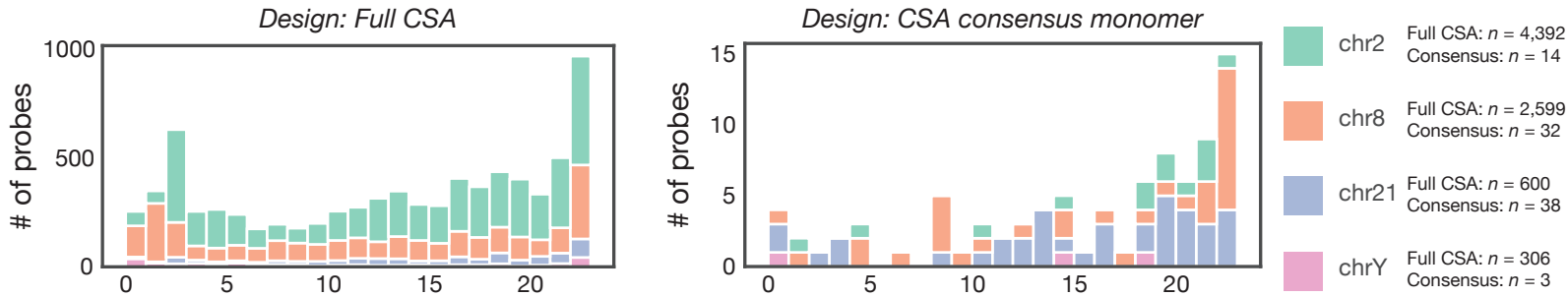
c

Classes of predicted binding profiles

Design: Full CSA

Design: CSA consensus monomer

Only binds target CSA
Binds 1 additonal CSA
Binds >1 additonal CSA

chr2 — Full CSA: *n* = 4,392 / Consensus: *n* = 14
chr8 — Full CSA: *n* = 2,599 / Consensus: *n* = 32
chr21 — Full CSA: *n* = 600 / Consensus: *n* = 38
chrY — Full CSA: *n* = 306 / Consensus: *n* = 3
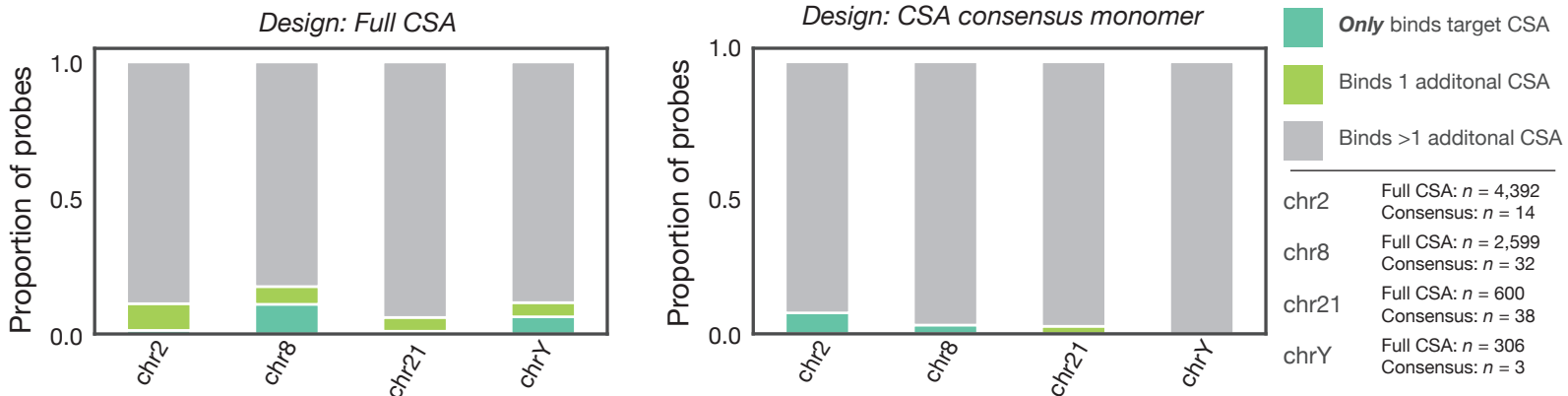
d

Computational resource usage

Design: Full CSA

Design: CSA consensus monomer
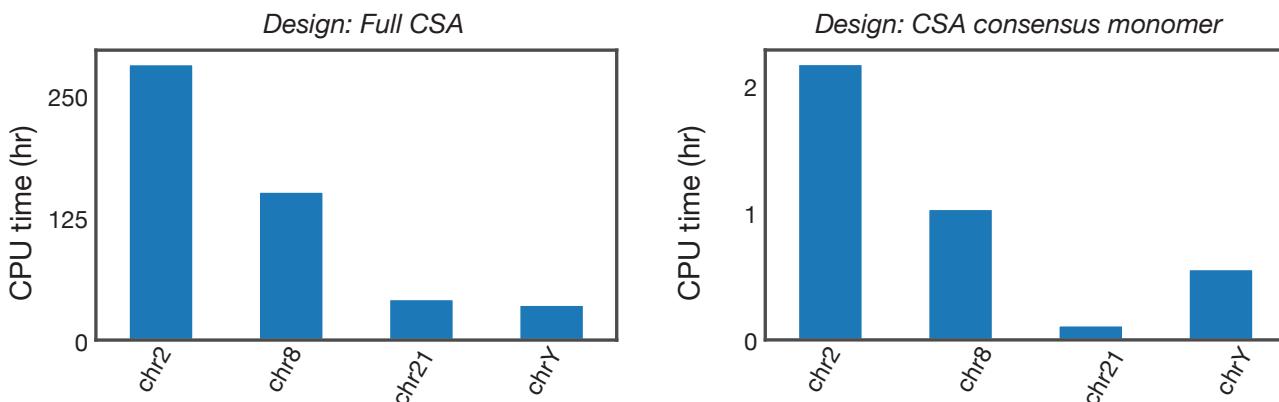
**Supplementary Fig. 1 | Probe design against four chromosome-specific alpha satellite repeats. a**, Schematic of the probe design process against four Chromosome-Specific Alpha satellite arrays (CSAs) in the human genome. **b**, Histograms of the # of CSAs with >100 predicted binding sites for probes designed from the full CSAs (left) and the CSA consensus monomers (right). **c**, Predicted binding classes for probes designed from the full CSAs (left) and the CSA consensus monomers (right). **d**, Aggregate CPU time needed to run the probe discovery and analysis.
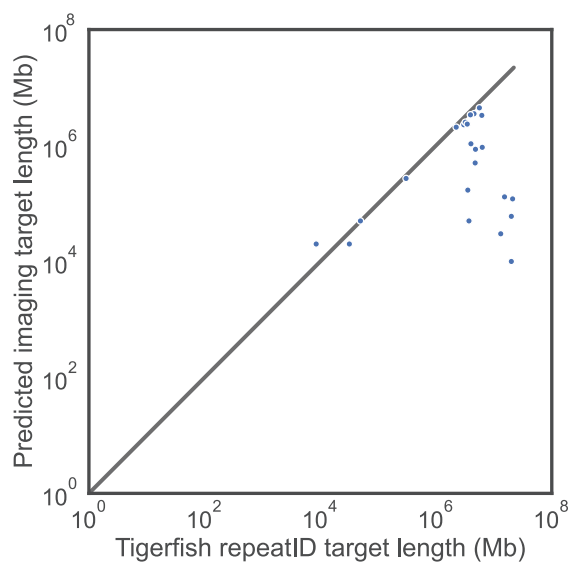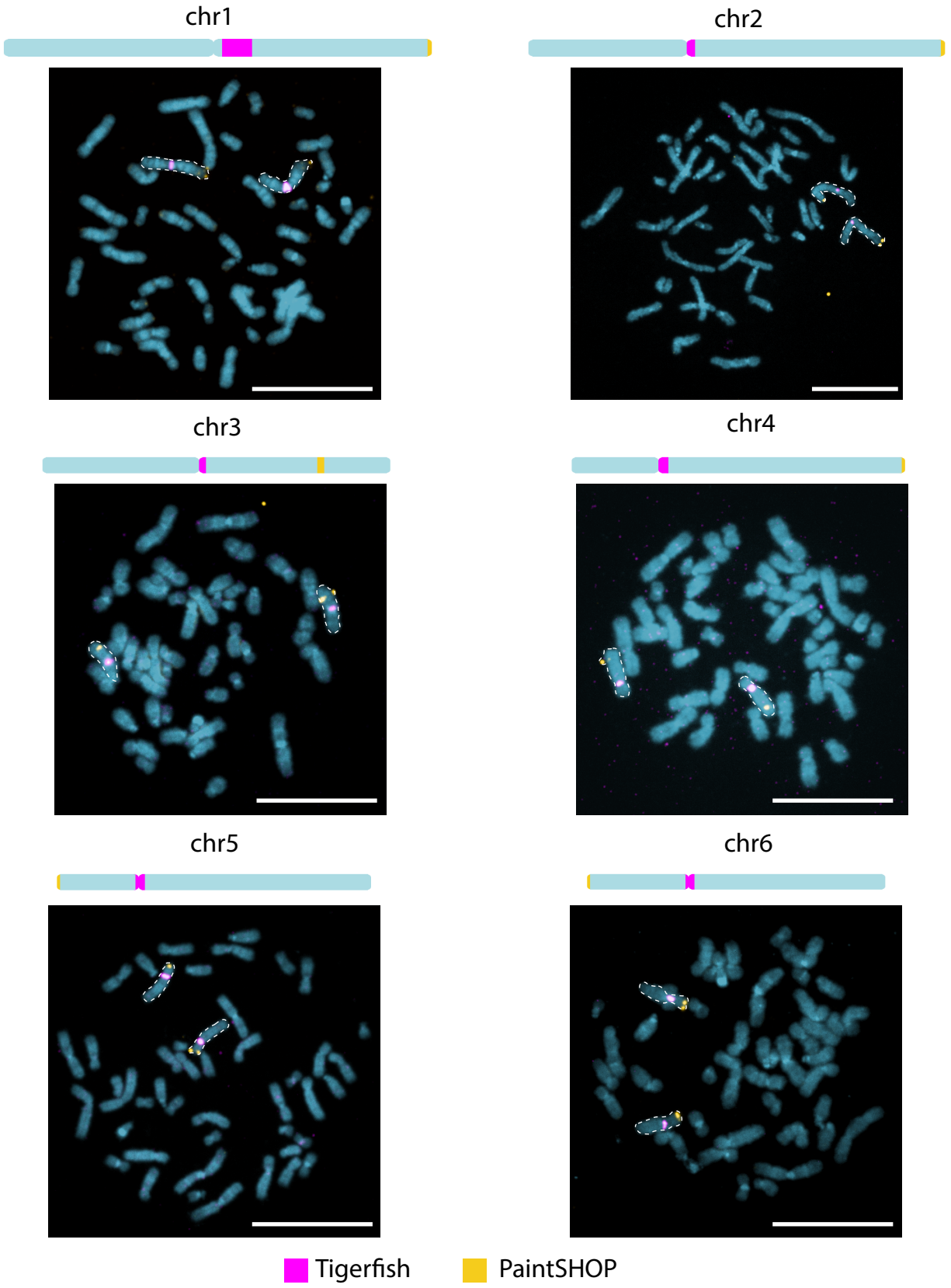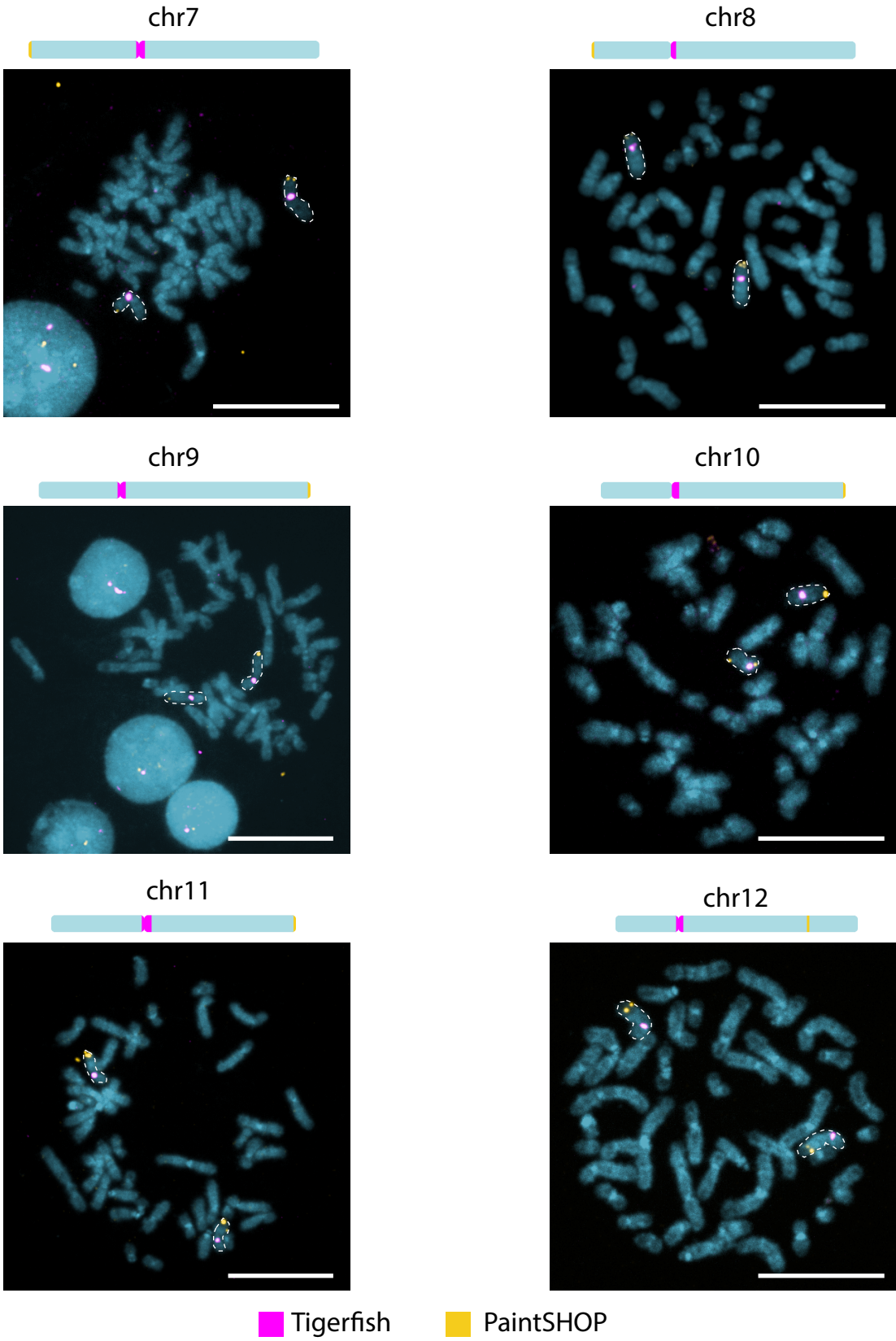
Supplementary Fig. 2

a
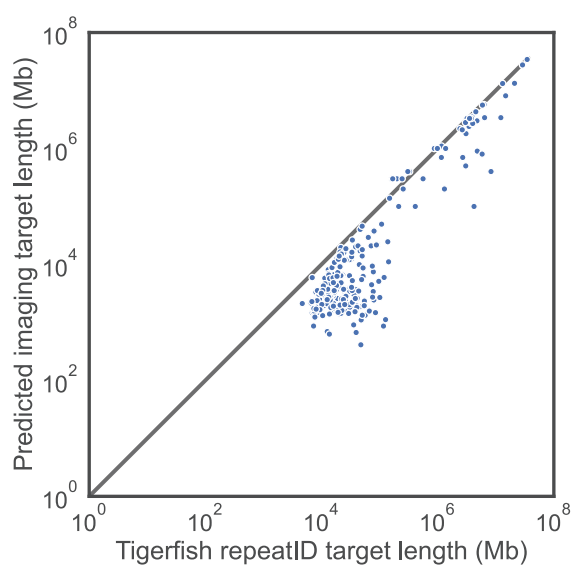
b
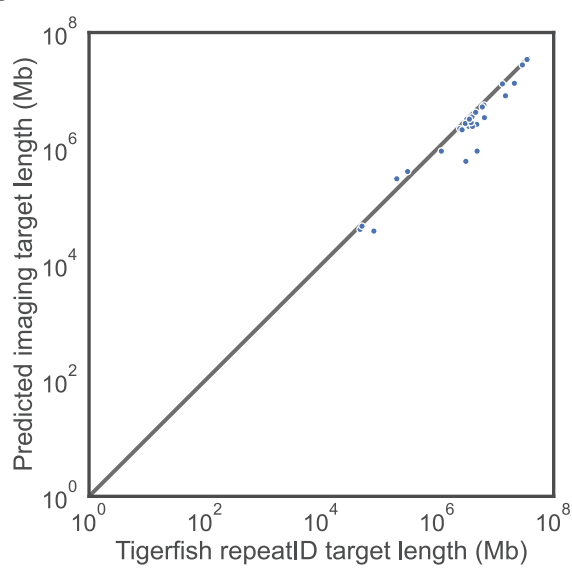
c

Supplementary Fig. 2 | Effective lengths of Tigerfish target intervals. a, Scatter plot depicting the effective lengths of intervals identified and processed successfully for probe design using permissive parameters (Y-axis) and the length of these intervals when first identified at the "Repeat discovery" step. **b**, Scatter plot depicting the effective lengths of intervals identified and processed successfully for probe design using conservative parameters (Y-axis) and the length of these intervals when first identified at the "Repeat discovery" step. **c**, Scatter plot depicting the effective lengths of intervals inputted for probe design (Y-axis) and the length of these intervals when first identified at the "Repeat discovery" step for the 24-target panel used for in situ validation experiments. *n* = 263 intervals.

Supplementary Fig. 3

Supplementary Fig. 3 | Full-field metaphase spreads for chromosomes 1–6. Full-field images of the metaphase spreads from which the crops depicted in Figure 4b originated showing the staining pattern of the indicated Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 20 µm.

Supplementary Fig. 4

chr7

chr8

chr9

chr10

chr11

chr12

■ Tigerfish   ■ PaintSHOP

Supplementary Fig. 4 | Full-field metaphase spreads for chromosomes 7–12. Full-field images of the metaphase spreads from which the crops depicted in Figure 4b originated showing the staining pattern of the indicated Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 20 μm.
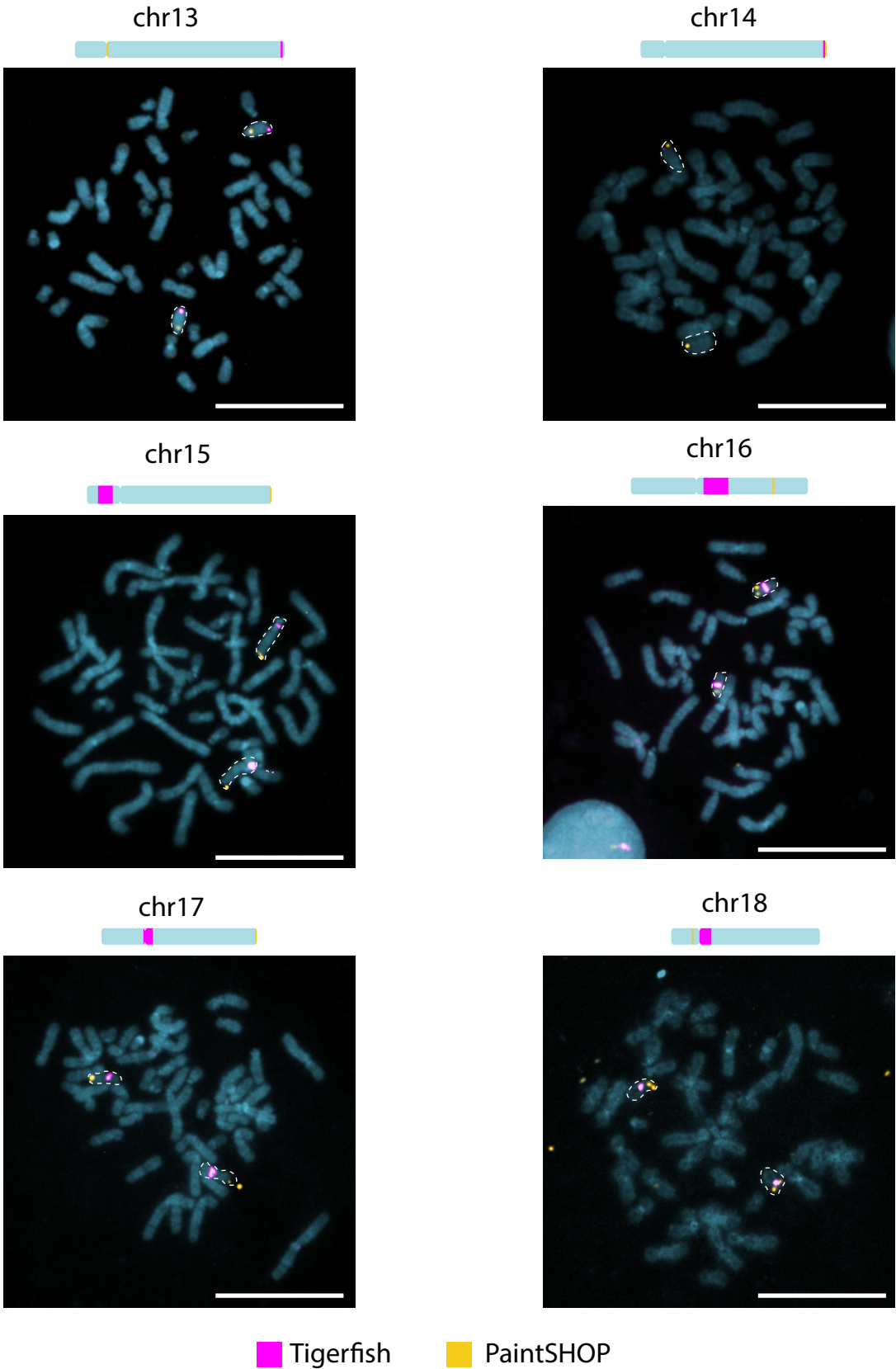
Supplementary Fig. 5

Supplementary Fig. 5 | Full-field metaphase spreads for chromosomes 13–18. Full-field images of the metaphase spreads from which the crops depicted in Figure 4b originated showing the staining pattern of the indicated Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 20 μm.

Supplementary Fig. 6

chr19

chr20

chr21

chr22

chrX

chrY

■ Tigerfish   ■ PaintSHOP

Supplementary Fig. 6 | Full-field metaphase spreads for chromosomes 19–Y. Full-field images of the metaphase spreads from which the crops depicted in Figure 4b originated showing the staining pattern of the indicated Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 20 μm.
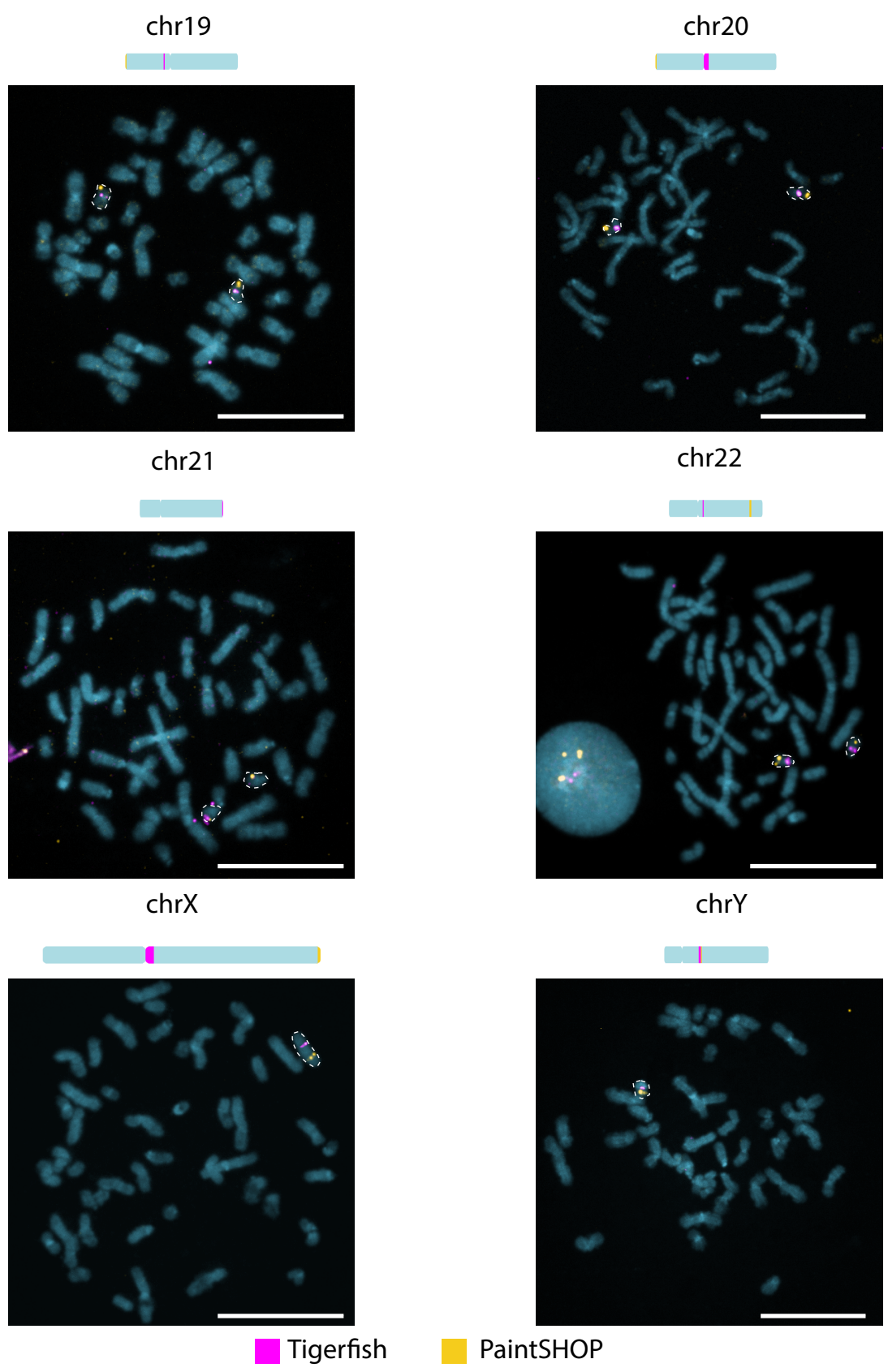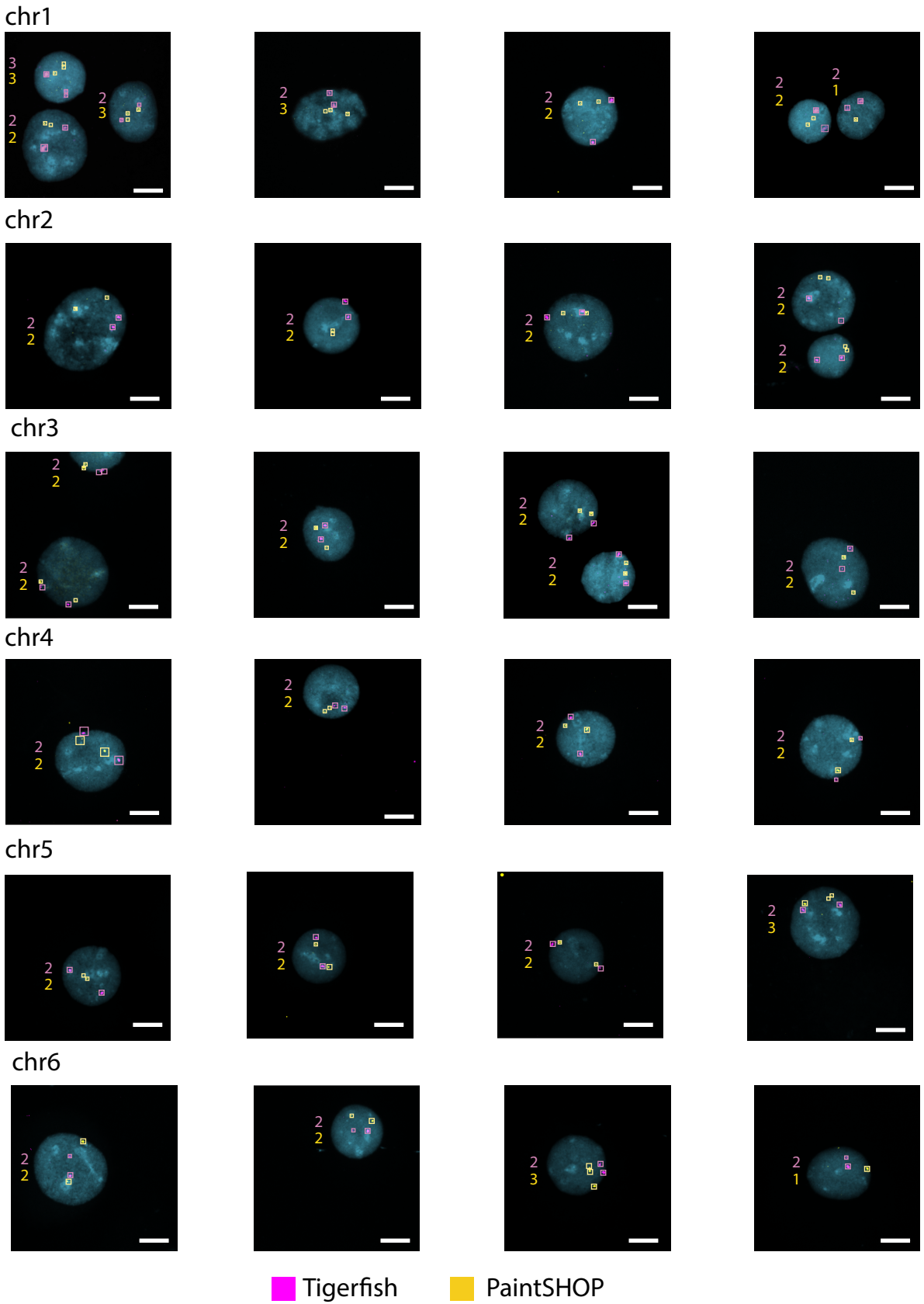
Supplementary Fig. 7

chr1

chr2

chr3

chr4

chr5

chr6

■ Tigerfish   ■ PaintSHOP

Supplementary Fig. 7 | Representative enumeration images for chr1–6. Four representative images of interphase nuclei and corresponding puncta counts for the specified Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 10 µm.

Supplementary Fig. 8

chr7

chr8

chr9

chr10

chr11

chr12

■ Tigerfish   ■ PaintSHOP

Supplementary Fig. 8 | Representative enumeration images for chr7–12. Four representative images of interphase nuclei and corresponding puncta counts for the specified Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 10 μm.
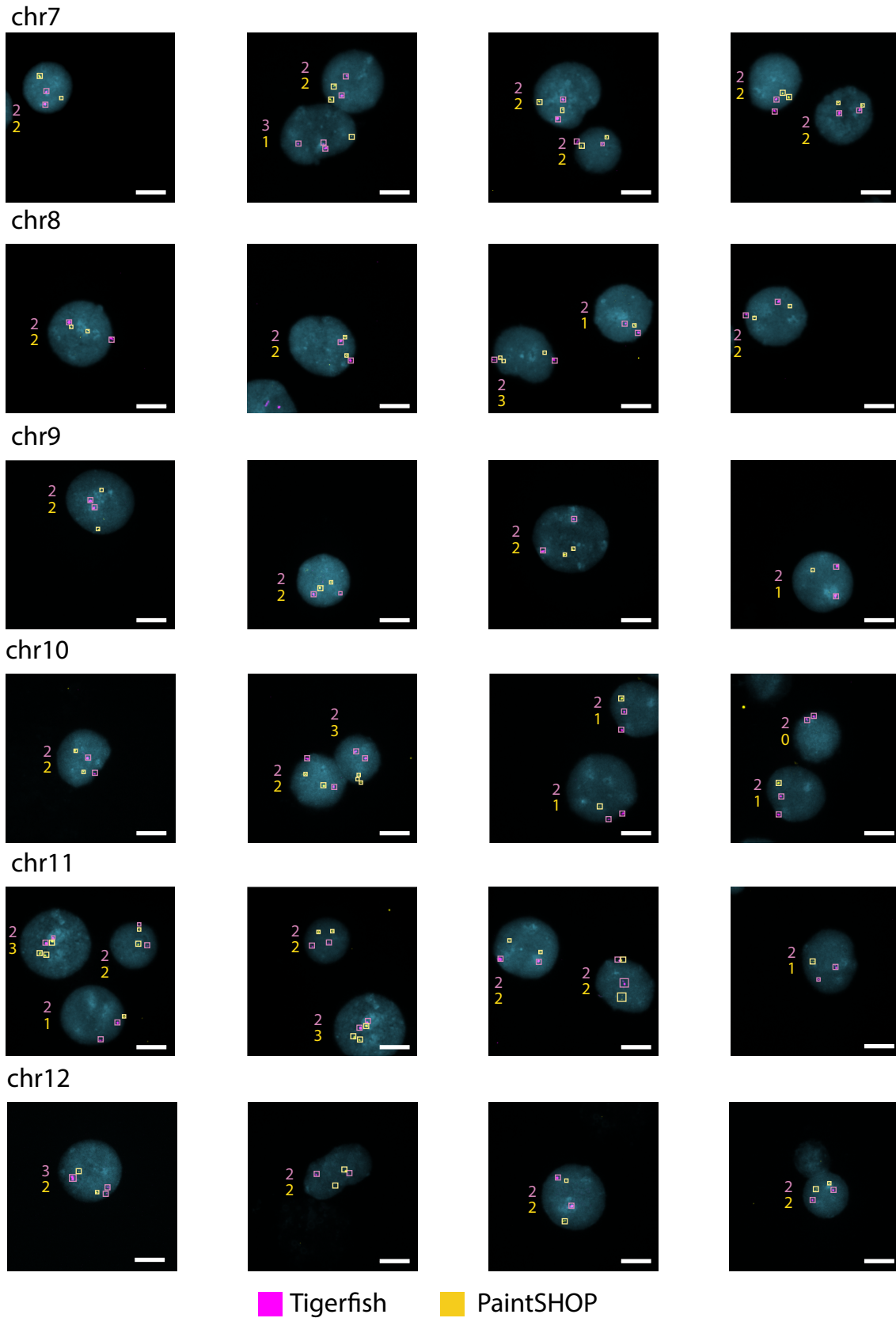
Supplementary Fig. 9

chr13

chr14

chr15

chr16

chr17

chr18

■ Tigerfish   ■ PaintSHOP

Supplementary Fig. 9 | Representative enumeration images for chr13–18. Four representative images of interphase nuclei and corresponding puncta counts for the specified Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 10 μm.
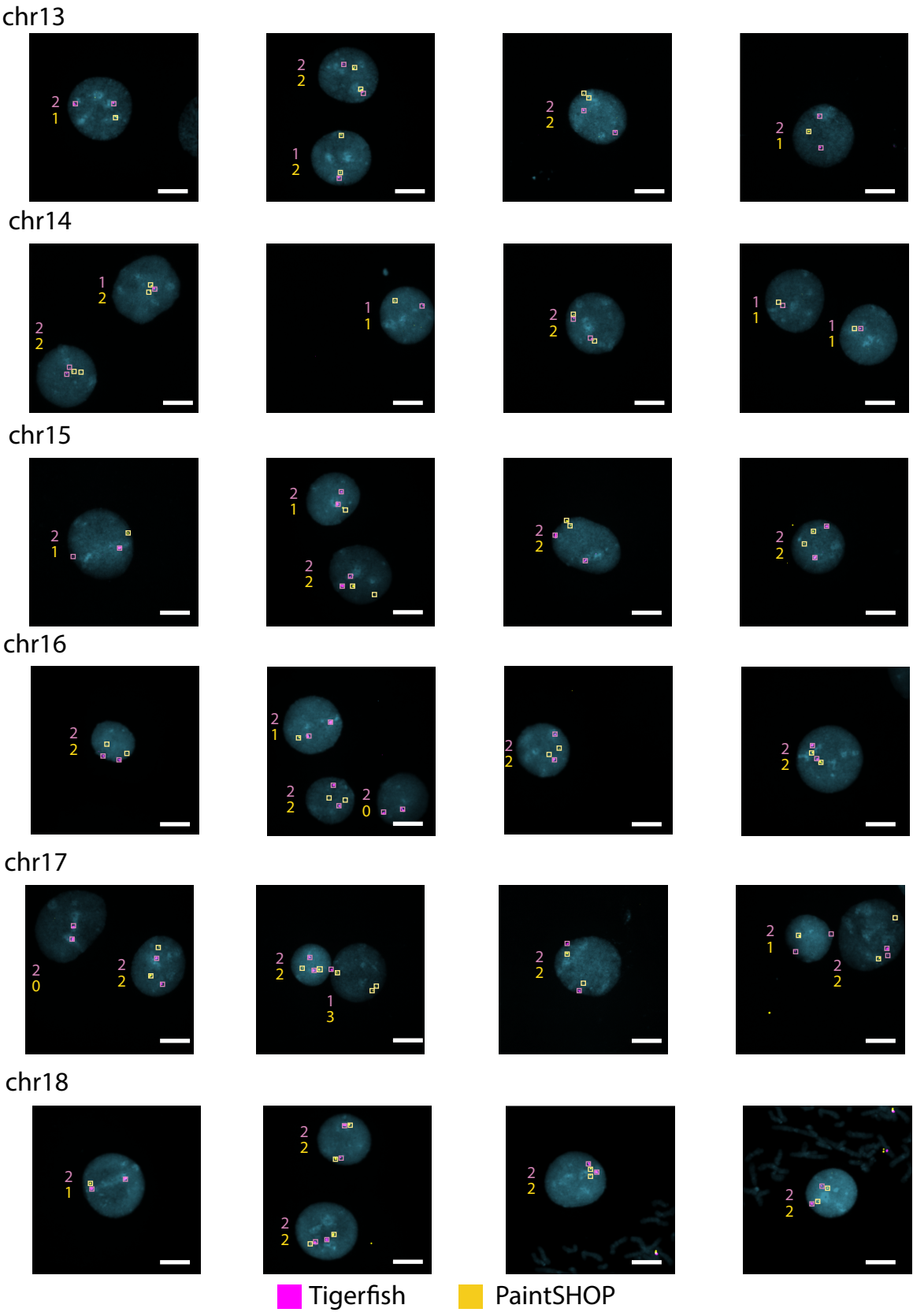
Supplementary Fig. 10

Supplementary Fig. 10 | Representative enumeration images for chr19–Y. Four representative images of interphase nuclei and corresponding puncta counts for the specified Tigerfish (magenta) and PaintSHOP (yellow) probe sets. Images are maximum intensity projections in Z. Scale bars, 10 µm.
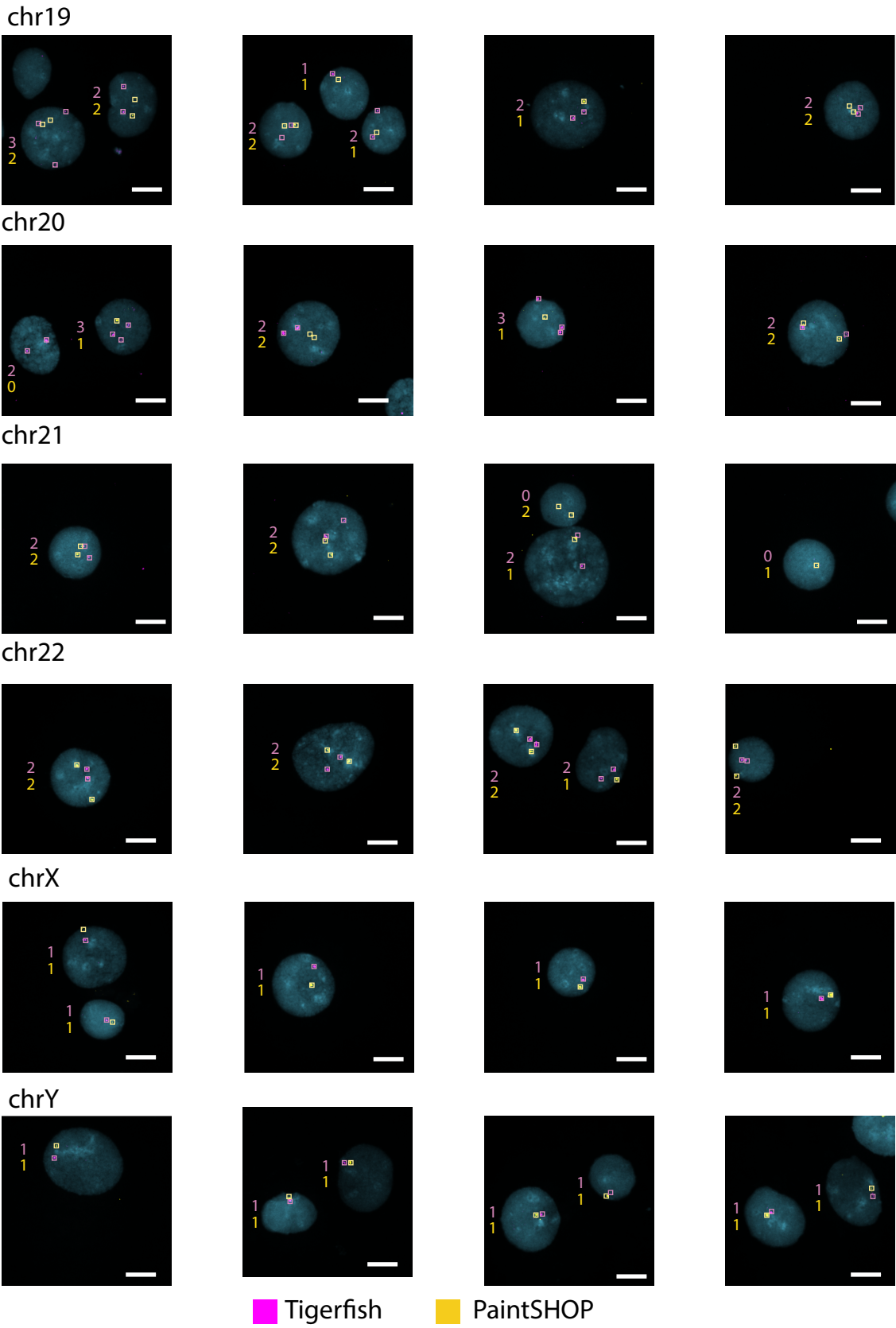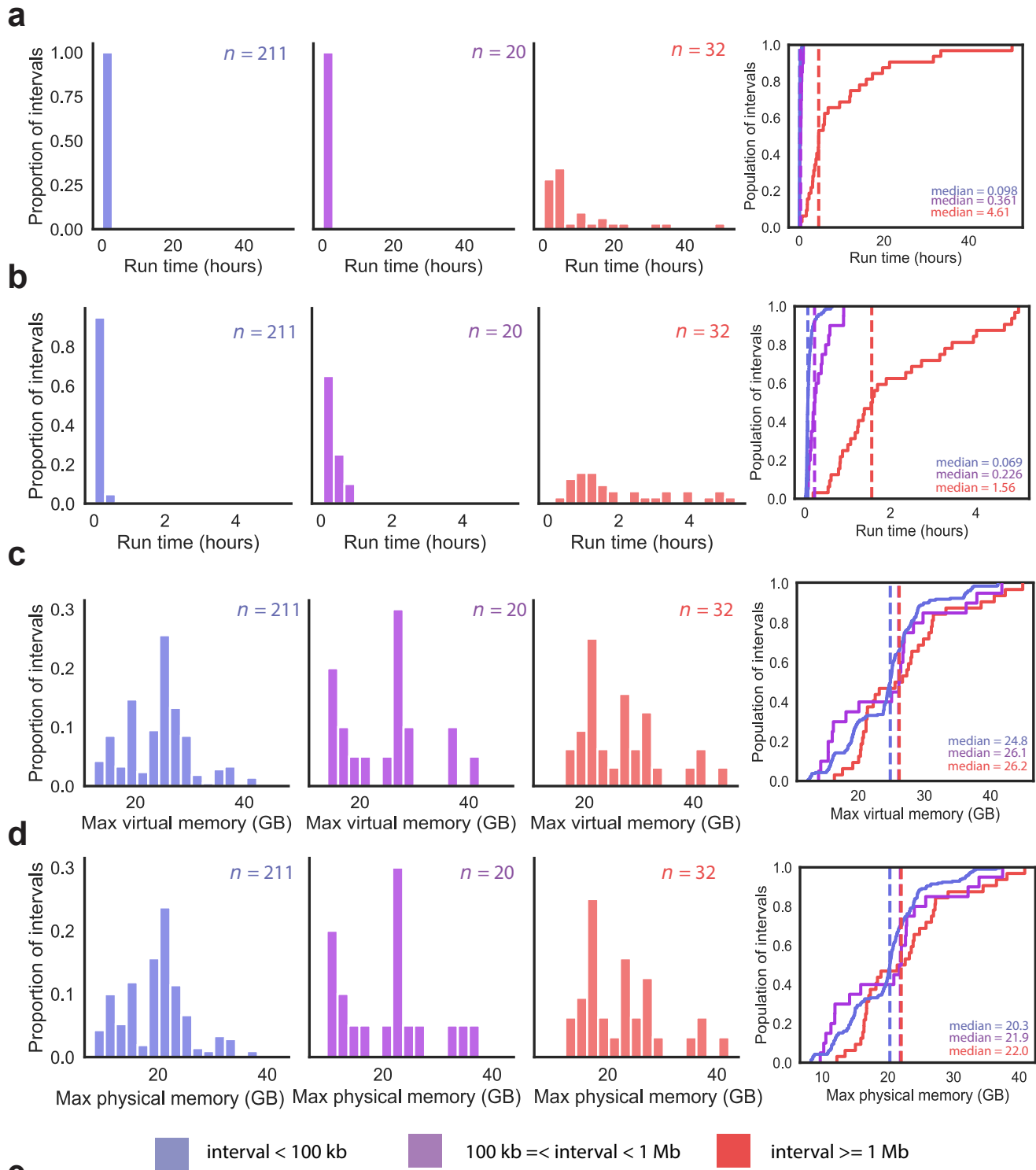
Supplementary Fig. 11

Supplementary Fig. 11 | Tigerfish computational requirements by repeat region length. a, Histograms (left) and empirical cumulative distributions (right) of the wall-clock runtime recorded for running the 263 conservative and permissive intervals stratified by the target interval length. **b**, Histograms (left) and empirical cumulative distributions (right) of the CPU runtime recorded for running the 263 conservative and permissive intervals stratified by the target interval length. **c**, Histograms (left) and empirical cumulative distributions (right) of the maximum recorded virtual memory allocation for running the 263 conservative and permissive intervals. **d**, Histograms (left) and empirical cumulative distributions (right) of the maximum recorded physical memory allocation for running the 263 conservative and permissive intervals. Vertical dashed lines in the cumulative distribution plots correspond to the median values. **e**, Summary statistics for the values presented in panels a–d.

Supplementary Note
Tigerfish inputs and outputs

Input(s)	Snakemake step	Output(s)	Option	Usage
FASTA	**generate_jf_count*** Generates a Jellyfish index file to count *k*-mers genome wide.	counts.jf	mer_val fasta_file	*Required.* *k*-mer size *Required.* Genomic reference FASTA
FASTA	**generate_bt2_indices*** Generates genome wide Bowtie2 (Bt2) index to align probes.	Bt2 Indices	fasta_file	*Required.* Genomic reference FASTA
counts.jf FASTA	**generate_jf_idx*** Generates *k*-mer count index files for each scaffold.	counts.txt index.txt scaffold FASTA	fasta_file mer_val sample	*Required.* Genomic reference FASTA *Required.* *k*-mer size *Required.* Scaffold name(s)
BED	**split_bed**** Takes BED file and splits coordinates into distinct files.	BED BED ...	bed_file sample	*Optional.* BED file with repeat target coordinates. *Required.* Scaffold name(s)
counts.txt index.txt	**repeat_ID**** Identifies repeat regions over queried chromosomes.	BED	sample file start window threshold composition mer_val	*Required.* Scaffold name(s) *Required.* Base position where repeat search begins. *Required.* Size of *k*-mer search window (W). *Required.* Min *k*-mer count value (T). *Required.* Proportion of elevated *k*-mers within W (C). *Required.* *k*-mer size
BED (repeat_ID) or BED (split_bed)	**design_probes**** Designs probes from identified repeat coords or user provided BED file.	probes.tsv region FASTA	fasta_file sample min_len max_len min_temp max_temp	*Required.* Genomic reference FASTA *Required.* Scaffold name(s) *Required.* Min size of candidate probe (bp) *Required.* Max size of candidate probe (bp) *Required.* Min melting temp of probe (C) *Required.* Max melting temp of probe (C)
probes.tsv counts.txt region FASTA	**kmer_filter** Computes each probe's on-target and off-target *k*-mer counts, then sorts probes by on-target count.	k_mer_sort.tsv	c1_val c2_val mer_val	*Required.* Constant to rank probes by copy_num. *Required.* Constant to rank probes by enrich_score *Required.* *k*-mer size
k_mer_sort.tsv	**probe_mer_filter** Filters probes based on probe *k*-mer similarity and target binding.	k_mer_filter.tsv	enrich_score copy_num mer_cutoff mer_val	*Required.* Min proportion a probe's *k*-mers must bind within a repeat region target. *Required.* Total sum of any probe's *k*-mers within a target repeat region. *Required.* Any probes within a target repeat exceeding this proportion of shared *k*-mers will be filtered *Required.* *k*-mer size
chrom.sizes	**generate_genome_bins:** Takes reference genome and creates bins.	alignment_bin.BED threshold_bin.BED	genome_windows chrom_sizes_file thresh_window	*Required.* Size of genome bins for alignment. *Required.* chrom.sizes file of queried genome. *Required.* Size of threshold bins to flag to determine imaging target coordinates.

* Denotes optional step if proper file paths/options are specified in config.yml
** Denotes step specific to either probe_design or repeat_ID run modes

Input(s)	Snakemake step	Output(s)	Option	Usage
k_mer_filter.tsv	**make_chrom_dir** Seperates probe files by repeat region.	repeat.txt		
repeat.txt Bt2 indices alignment_bin.BED	**alignment_filter** Filters candidate probes using Bt2 and NUPACK to predict in silico probe binding.	filtered_probes.tsv	target_sum bt2_alignments seed_length model_temp max_pdups min_on_target max_probe_return off_bin_thresh align_thresh ref_flag	*Required.* Total on-target sum of all probes desired. *Required.* Bt2will return alignments up to this val. *Required.* Controls Bt2 seed length. *Required.* Temperature of NUPACK predict model. *Required.* Probes in final dataset with predicted binding greater than this value will be filtered. *Required.* Min on target score for any given probe. *Required.* Max probes to be returned/repeat. *Required.* Off-target threshold over any non-target genomic bin. Required. Min binding sites required to flag a binned region as significant toward probe binding. Required. Provides intermediate output files.
filtered_probes.tsv	**merge_alignment_filter** Aggregates all candidate probes by scaffold.	chrom_probes.tsv		
chrom_probes.tsv	**split_rm_alignments** Creates a new directory containing all candidate probe files by repeat.	scaffold/region.tsv		
scaffold/region.tsv	**align_probes** Takes candidate probes corresponding to compute target specificity. SAM derived sequences are used to compute *in silico* binding.	region_align.txt	bt2_alignments seed_length model_temp mer_val	*Required.* Bt2 will return alignments up to this val. *Required.* Controls Bt2 seed length. *Required.* Temperature of NUPACK predict model. binding greater than this value will be filtered. *Required.* Any probes within a target repeat exceeding this proportion of shared *k*-mers will be filtered
region_align.txt	**derived_beds** Creates BED file from SAM derived sequence alignment for each candidate probe.	derived_align.BED		
region_align.txt	**get_region_bed** Creates BED file for the target repeat region.	repeat.BED		
derived_align.BED repeat.BED	**bedtools_intersect** Performs two bedtools intersects: 1. Derived alignments against threshold genome bins. 2. Repeat region against threshold genome bins.	derived_BEDtools.txt repeat_BEDtools.txt		

* Denotes optional step if proper file paths/options are specified in config.yml
** Denotes step specific to either probe_design or repeat_ID run modes

Input	Snakemake step	Output	Option	Usage
derived_BEDtools.txt repeat_BEDtools.txt region_align.txt chrom.sizes	**get_alignments** Computes genome wide binding summaries for all probes within a target repeat region.	binding_map.png thresh_summary.txt binding_quant.txt	align_thresh	*Required*. Binding sites required to flag a bin as significant toward probe signal.
filtered_probes.tsv binding_map.png thresh_summary.txt binding_quant.txt	**map_region_coords** Adds imaging target coordinates to the candidate probes file. Creates probe binding summary.	final_probes.tsv		
final_probes.tsv probe_summary.txt	**merge_mapping** Aggregates all repeat regions into a single file by scaffold.	probes_merged.tsv		
probes_merged.tsv	**summary** Summarizes probe binding and count by repeat region target.	probe_summary.txt		

* Denotes optional step if proper file paths/options are specified in config.yml
** Denotes step specific to either probe_design or repeat_ID run modes

Input	Snakemake step	Output	Option	Usage
filtered_probes.tsv	**gather_repeat_regions** Takes filtered candidate probes and splits them by scaffold. Input file should have one repeat per scaffold.	split_probes.txt	sample	*Required.* Scaffold name(s)
split_probes.txt	**align_cand_probes** Takes candidate probes corresponding to compute target specificity. SAM derived sequences are used to compute *in silico* binding.	region_align.txt	bt2_alignments seed_length model_temp mer_val	*Required.* Bt2 will return alignments up to this val. *Required.* Controls Bt2 seed length. *Required.* Temperature of NUPACK predict model. binding greater than this value will be filtered. *Required.* Any probes within a target repeat exceeding this proportion of shared *k*-mers will be filtered
region_align.txt	**derived_cand_beds** Creates BED file from SAM derived sequence alignment for each candidate probe.	derived_align.BED		
region_align.txt	**get_cand_region_bed** Creates BED file for the target repeat region.	repeat.BED		
derived_align.BED repeat.BED	**bedtools_cand_intersect** Performs two bedtools intersects: 1. Derived alignments against threshold genome bins. 2. Repeat region against threshold genome bins.	derived_BEDtools.txt repeat_BEDtools.txt		
derived_BEDtools.txt repeat_BEDtools.txt region_align.txt chrom.sizes	**get_cand_alignments** Computes genome wide binding summaries for all probes within a target repeat region.	binding_map.png thresh_summary.txt binding_quant.txt	align_thresh	*Required.* Binding sites required to flag a bin as significant toward probe signal.
chrom.sizes repeat.BED	**generate_cand_chromomap*** Creates a karyoplot using chromoMap of the repeat target.	chromomap.HTML		
filtered_probes.tsv binding_map.png thresh_summary.txt binding_quant.txt	**map_cand_region_coords** Adds imaging target coordinates to the candidate probes file. Creates probe binding summary.	final_probes.tsv		
final_probes.tsv probe_summary.txt	**merge_cand_mapping** Aggregates all repeat regions into a single file by scaffold.	probes_merged.tsv		
probes_merged.tsv	**summary** Summarizes probe binding and count by repeat region target.	probe_summary.txt		

* Denotes optional step if proper file paths/options are specified in config.yml
** Denotes step specific to either probe_design or repeat_ID run modes