Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads

Carolin A Müller[1+*], Michael A Boemo[1+], Paolo Spingardi[2], Benedikt M Kessler[3], Skirmantas Kriaucionis[2], Jared T Simpson[4,5], Conrad A Nieduszynski[1*]

Supplemental Information

Supplemental tables:

Supplemental figures:

Supplemental tables

Supplemental Table S1: Information relating to D-NAscent, mass spectrometry and BrdU-seq samples. 'sample_info' tab: summary information for each experiment and analysed sample; 'Nanopore' tab: summary information for each MinION DNA sequencing library and sequencing run; 'Mass_spec' tab: summary information for each sample analysed by Mass spectrometry; 'file download' tab: summary of raw and processed files deposited at NCBI GEO.

Supplemental Table S2: Oligonucleotide information.

Oligonucleotide	*Sequence*
CA1218	5'-GAATTGGGCCCGCTCAGC-3'
CA1219	5'-GCTCCCGGGTCCGCGGCGATGCTCAGGCTC TGTGTCTGCTGAGCGGGCCCAATTC-3'

Supplemental Table S3: Strain information.

Strain	*Genotype*	*Source*
YVL11	*MATa ADE2 cdc21::kanMX leu2::LEU2-GAL-hENT1 LYS2 RAD5 trp1::TRP1-GAL-dNK ura3-1*	Vernis *et al*. (2003)
E3087	*MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, URA3::GPD-TK(5x), AUR1c::ADH-hENT1, RAD5+*	Magiera *et al*. (2014)

Supplemental figure legends

Supplemental Figure S1: Nanopore sequencing can distinguish thymidine from analogues. Event distributions for six adjacent 6-mers containing either thymidine (blue) or BrdU (green), FdU (red), (C) IdU (black) or (D) EdU (orange). The data were generated using the ONT MinION R9 and R9.5 pore with sequencing speeds of 250 bp/s (**a**) or 450 bp/s (**b**), respectively.

Supplemental Figure S2: Mass spectrometry quantification of BrdU in genomic DNA. Genomic DNA was hydrolysed and the nucleosides deoxyadenosine (dA), thymidine (dT) and 5-Bromo-2'-deoxyuridine (BrdU) were quantified by mass spectrometry. Bar graphs show the percentage of thymidine to deoxyadenosine (blue), BrdU to deoxyadenosine (orange) or BrdU to the sum of BrdU and thymidine (red) in samples taken from (**a**) yeast cultures grown constitutively in different concentrations of BrdU (relating to Fig. 2), (**b**) the cell cycle experiment to generate hemi-labelled DNA using 400 μg/ml (1.3 mM) BrdU (relating to Fig. 3 a, b), (**c**) the pulse-chase experiment in the presence of hydroxyurea (relating to Fig. 3 c-g) and (**d**) the cell cycle experiment with a limiting concentration of BrdU (relating to Fig. 4). Bargraphs in (a) show the average of two independent measurements for each sample, with grey circles representing the individual values.

Supplemental Figure S3: BrdU-seq data from cells grown constitutively in BrdU. Thymidine-auxotroph *S. cerevisiae* cells were grown constitutively in media containing different ratios of thymidine to BrdU, yielding BrdU incorporation rates between 15% and 79%. Barcoded genomic DNA from BrdU-containing samples was pooled, immuno-precipitated using an anti-BrdU antibody and prepared for Illumina sequencing. Sequencing reads were demultiplexed, mapped to the *S. cerevisiae* genome and read tag counts were determined for uniquely mapping reads in 100 bp bins. The ratio between immuno-precipitated material and the corresponding input sample was calculated for each bin as described in the methods. The ratios were summed over 5 kb windows across the yeast genome and plotted as distributions.

Supplemental Figure S4: Comparison of pore models trained using different incorporation rates of BrdU. Scatter plots comparing the mean of the distribution for BrdU-containing 6-mers when these distributions were found using genomic DNA with different BrdU incorporation rates. The mean of each distribution trained on (**a**) 15% incorporation, (**b**) 26% incorporation, and (**c**) 79% incorporation is compared to the mean of the distribution trained on 49% incorporation. The diagonal is indicated with a red line. (**d**-**f**) Corresponding histograms of the differences means for each 6-mer that has KL-divergence >2 compared with the ONT pore model. Correlations are Pearson's R.

Supplemental Figure S5: Parameters for detection of BrdU in genomic DNA. (**a**) Histogram of distances (mean = 21 bp) on the *S. cerevisiae* genome between 6-mers in the BrdU pore model that had a KL-divergence >2 against the ONT model. (**b**) Receiver operating characteristic (ROC) curves specifying the true positive and false positive rates for various log-likelihood thresholds of BrdU when using a BrdU pore model that only contained 6-mers with KL-divergence >0 (blue), >1 (green) or >2 (red) against the ONT pore model. Numbers near points specify the log-likelihood threshold above which a position in a read is classified as BrdU.

Supplemental Figure S6: Hidden Markov model for BrdU detection. Design of the hidden Markov model (HMM) used to evaluate the log-likelihood that a 6-mer of interest (green box) from our BrdU pore model contains at least one BrdU. The HMM has states corresponding to

the sequence 20 bp on either side of the 6-mer of interest. Each of the 41 bases corresponds to a deletion state, an insertion state, and a match state. Deletion states are silent and insertion states emit a uniform distribution in the range of event signals commonly observed. Match states emit a normal distribution of the expected signal for the 6-mer at that position. Two HMMs are built for the 41 base pairs: In the first (top), all of the Gaussian match state (blue) emission distributions are taken from the pore model provided by ONT, which specifies the mean and standard deviation of the expected current produced by each thymidine-only 6-mer. The second HMM (bottom) is similar, but the Gaussian match state emission distribution corresponding to the 6-mer of interest is taken from our new pore model, which specifies the mean and standard deviation of the expected current produced by 6-mers that contain at least one BrdU (red). For each HMM, running the forward algorithm on the events corresponding to the 41 bases gives the probability of observing those events given the states and structure of the HMM. Computing the log ratio between these two probabilities gives a log-likelihood that the 6-mer of interest contains at least one BrdU.

Supplemental Figure S7: Genome-wide BrdU-seq profile from hemi-labelled cells. Yeast cells were arrested in G1 and released into S phase in the presence of 400 μg/ml BrdU (1.3 mM). Plots show the number of Illumina sequencing reads (IP normalised to input material) with data points from non-unique sequences and those greater than six standard deviations above the mean removed. Replication origins are shown as yellow circles.

Supplemental Figure S8: Flow cytometry analysis of cellular DNA content. Cell cycle synchrony from an alpha factor arrest (0 min) into (**a**) 400 μg/ml (1.3 mM) BrdU to generate hemi-labelled DNA (relating to Fig. 3 a, b), (**b**) HU and (40 μg/ml, 130 μM) BrdU followed by a thymidine chase (relating to Fig. 3 c - g) or (**c**) 40 μg/ml (130 μM) BrdU concentrations (relating to Fig. 4). Time points shown in red where taken when BrdU was present in the media. Time points taken during the thymidine chase are shown in blue.

Supplemental Figure S9: Sister fork progression in HU. Histogram of the length of BrdU-positive regions on either side of the origin for each read in **Fig 3f**, interpreted as the distance that each fork travelled from the origin before BrdU was depleted. Bars away from the central diagonal correspond to reads that had asymmetric fork progression.

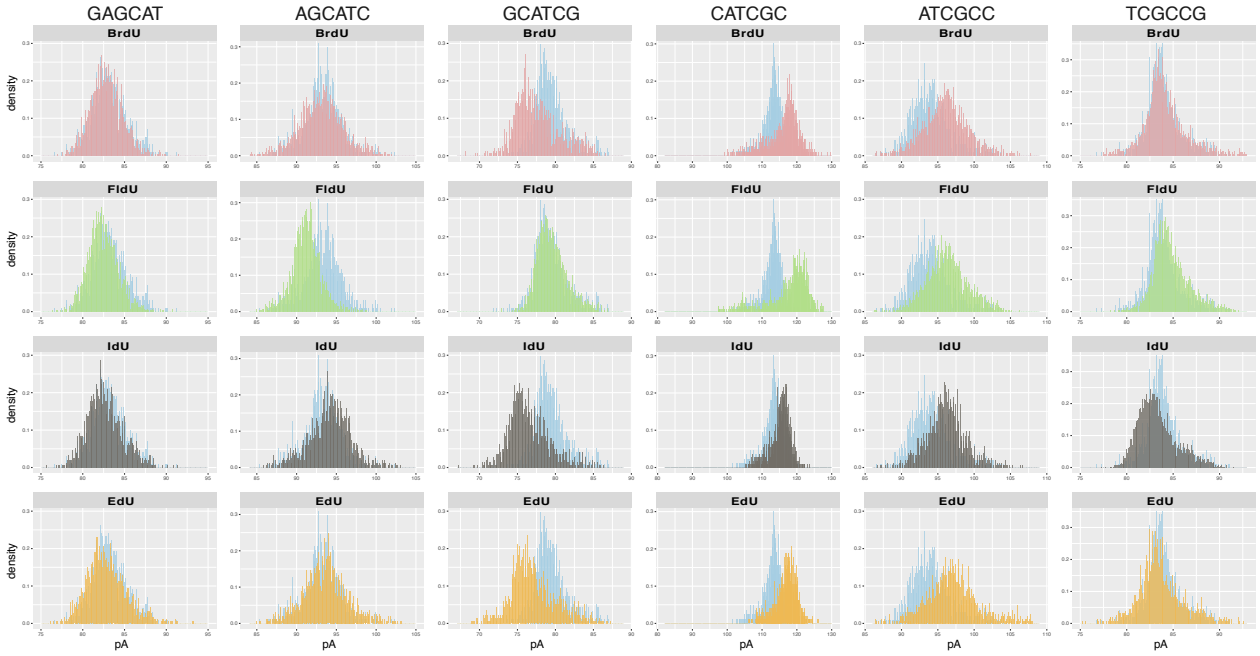
Supplemental Figure S10: Correlation between BrdU-seq data and median replication time. A scatter plot showing the relationship between published median replication time (Trep; Müller et al. 2014) and BrdU-seq data for timepoint t2 (90 min post release).

Supplemental Figure S11: D-NAscent identification of clustered and dispersed initiation sites. (**a**) Histogram showing the distance between origins called by D-NAscent and the closest confirmed or likely origin from OriDB (grey; N=6070; mean=3.9 kb; s.d.=3 kb; median=1.7 kb) and the distance between the same number of randomly generated positions on the genome and the nearest confirmed or likely origin from OriDB (red; N=6070; mean=7.8 kb; s.d.=5.9 kb; median = 5.6 kb). (**b**) Called origins were considered to be close to a confirmed or likely origin from OriDB if the distance between them was less than 3.9 kb (and far if the distance was greater than 3.9kb); fraction of close (blue) and far (green) origins that do not have another origin called by D-NAscent within the indicated (x-axis) window size of their location. (**c**) Violin-plot representations of the distribution of BrdU z-scores for 2 kb windows spanning origins that replicate in early (left), mid (center) or late (right) S phase. The top tick denotes the maximum and the middle tick denotes the mean.

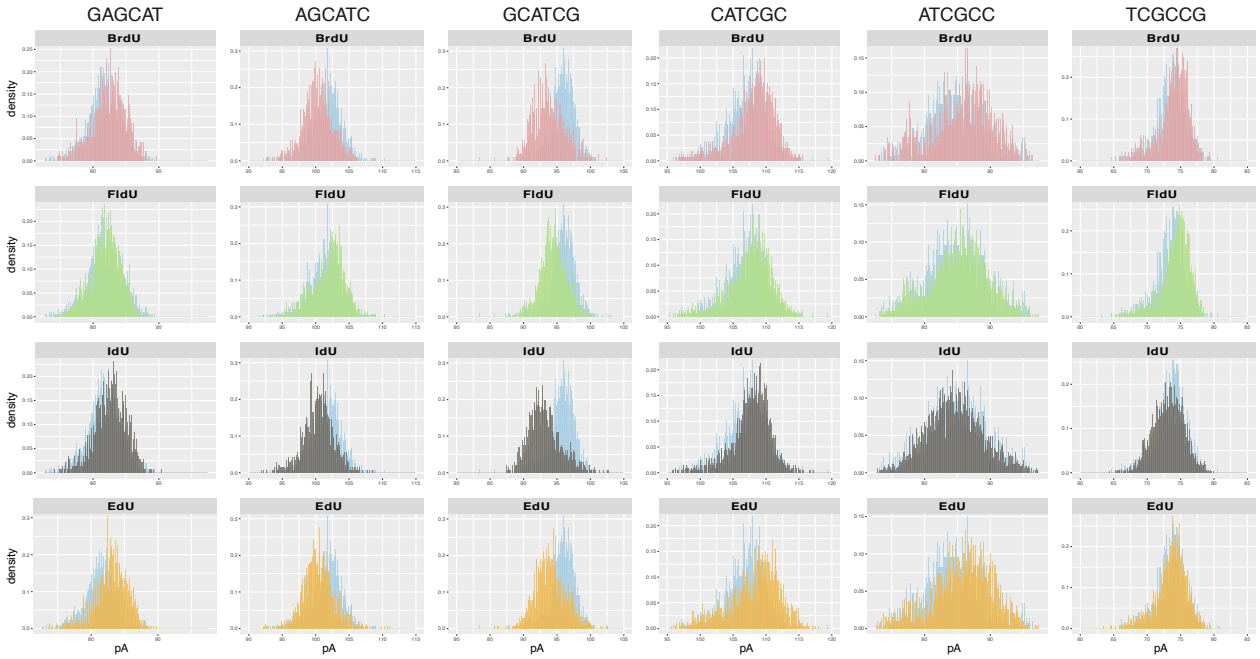
Supplemental Figure S12: Distribution of nanopore read lengths for the results shown in Fig. 3 and Fig. 4. (**a**) Nanopore read lengths corresponding to the results shown in Fig. 3 c-g. (**b**) Nanopore read lengths corresponding to the results shown in Fig. 4.

Supplemental Fig. S1

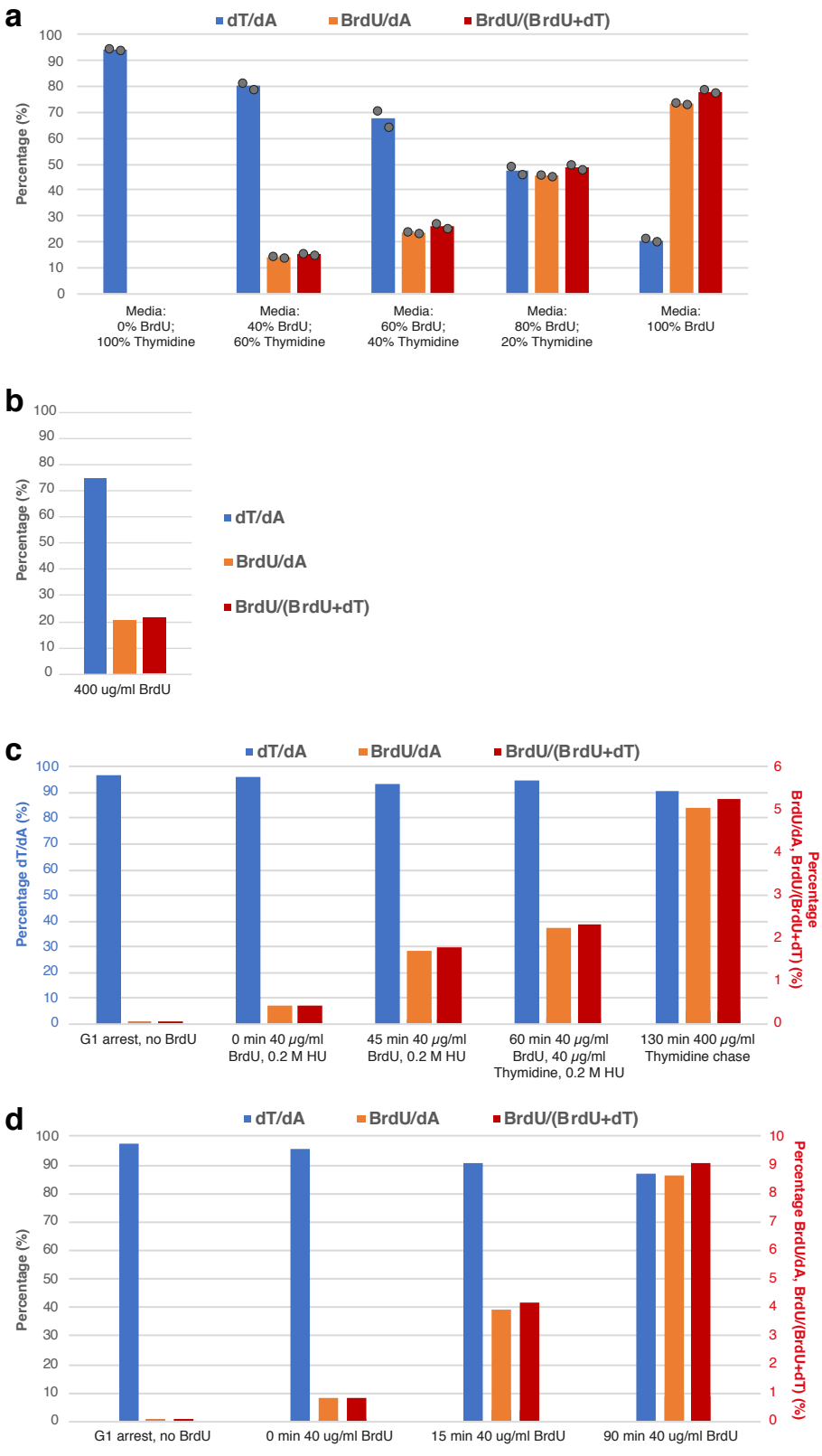
a



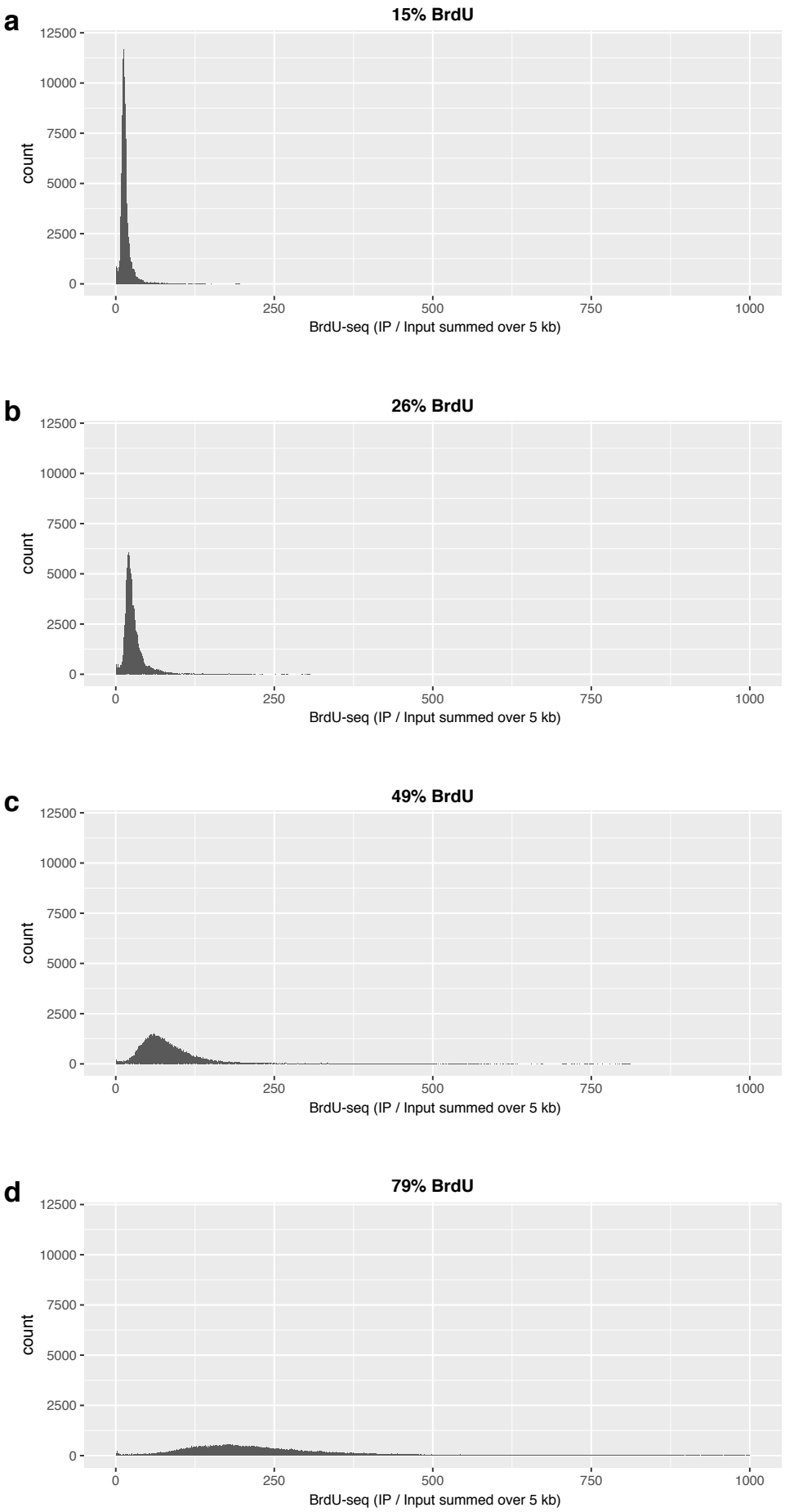
b



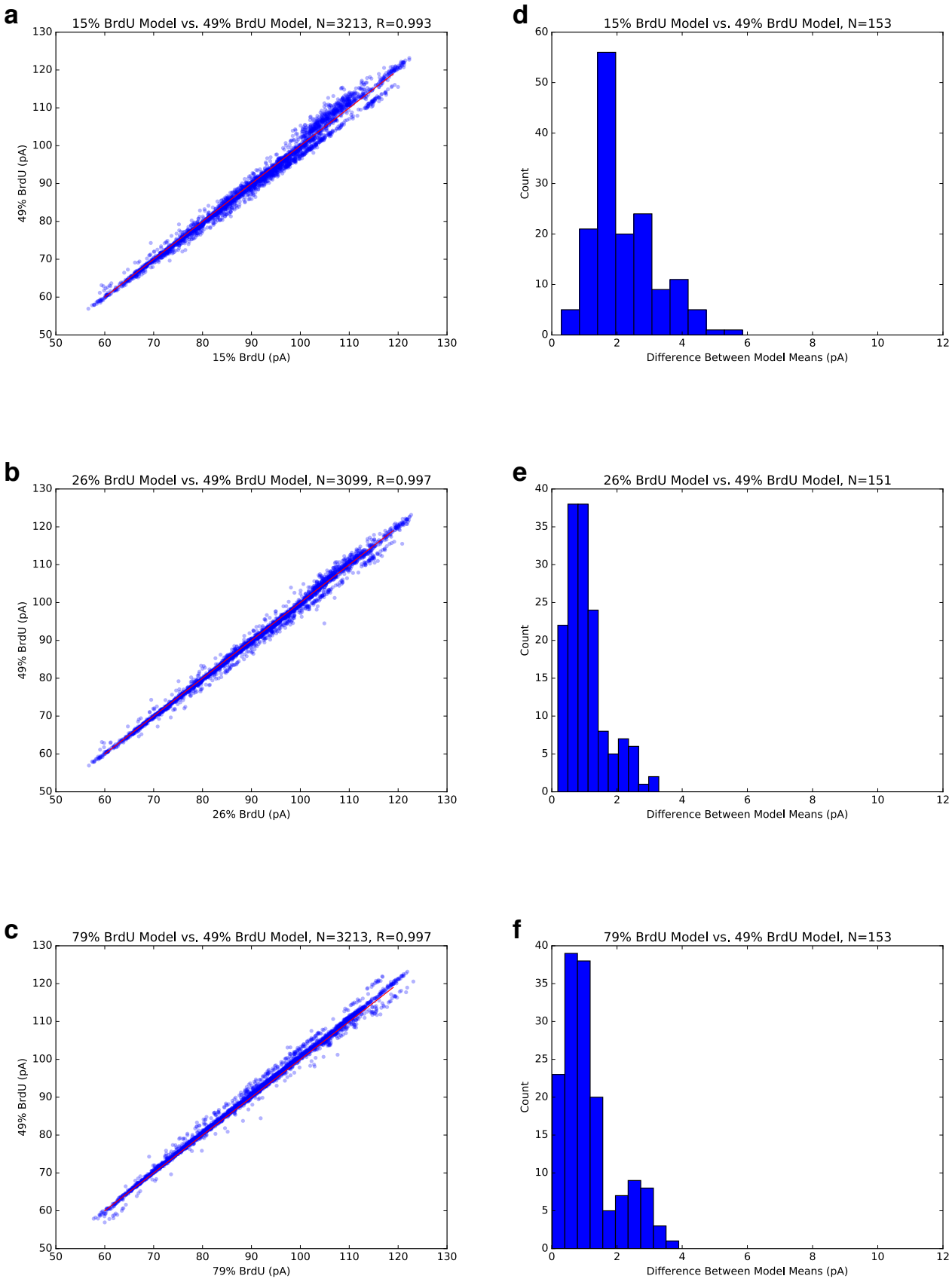
Supplemental Fig. S2



Supplemental Fig. S3

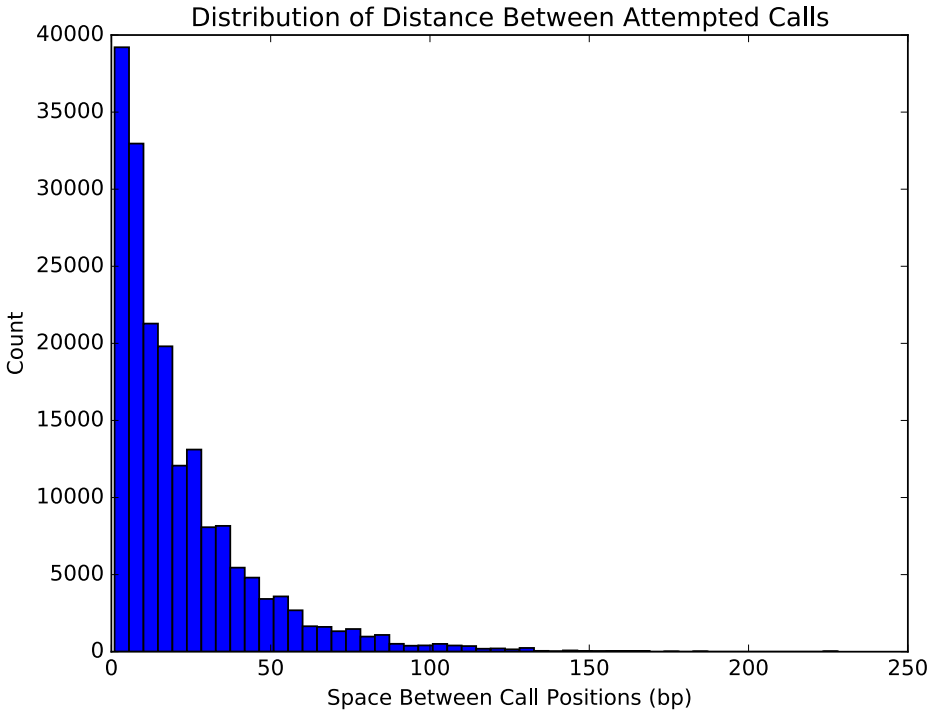


Supplemental Fig. S4

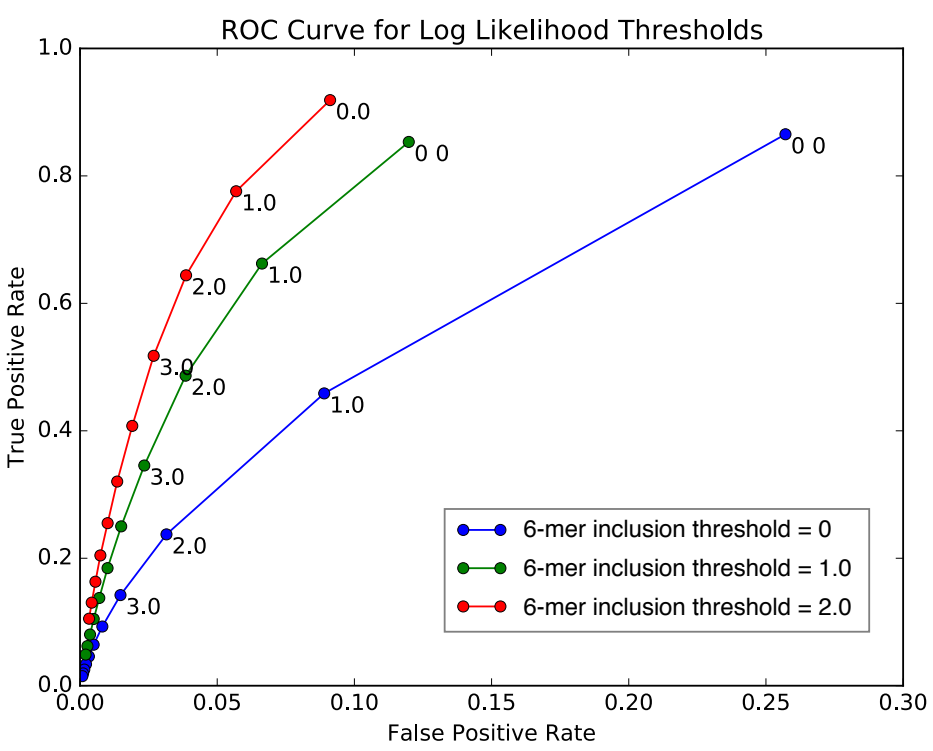


Supplemental Fig. S5

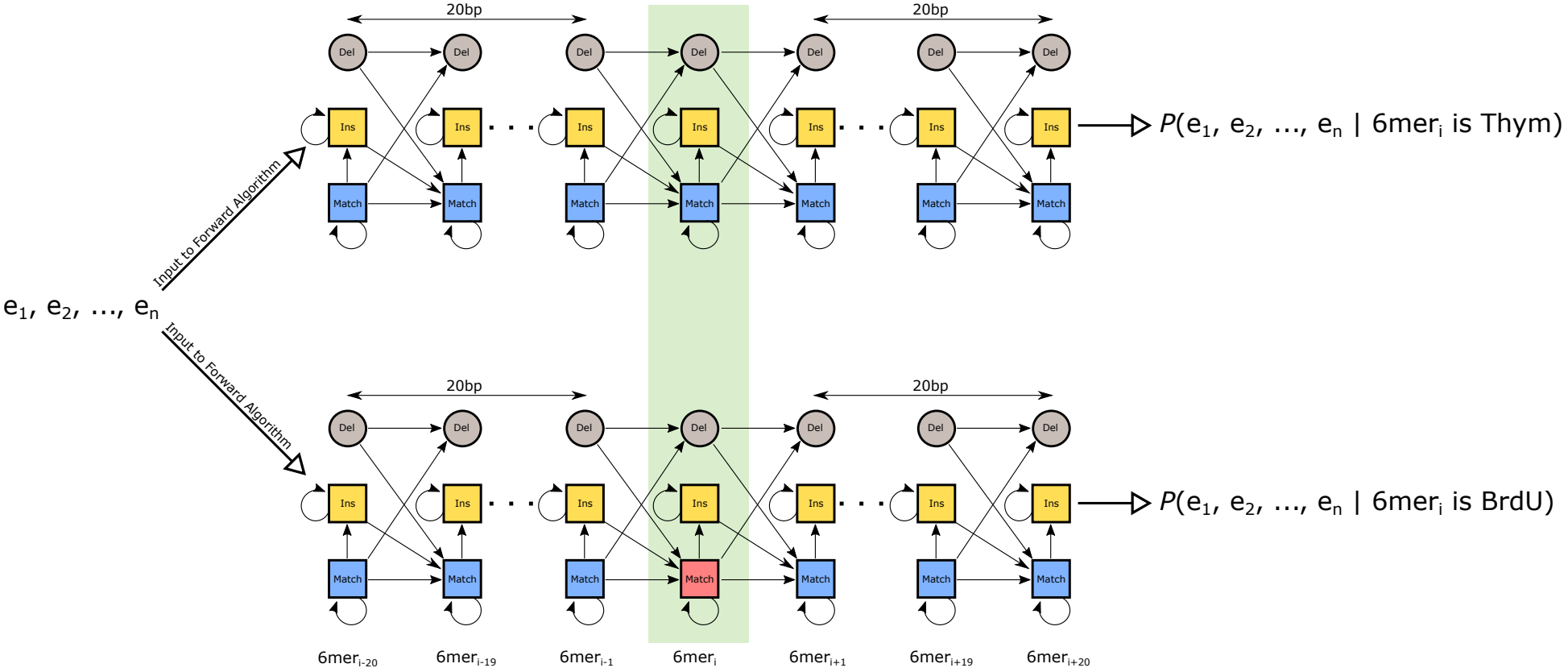
a



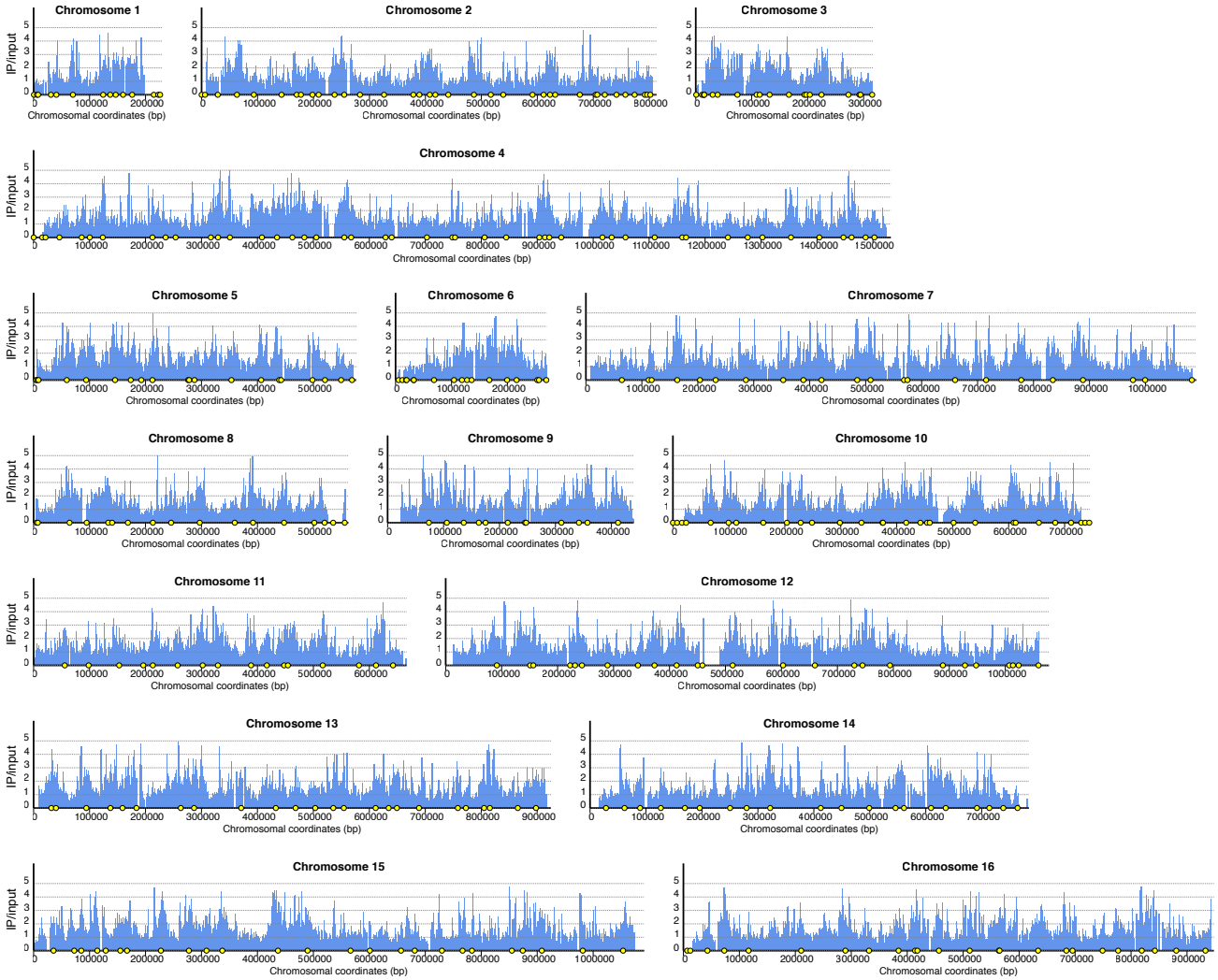
b



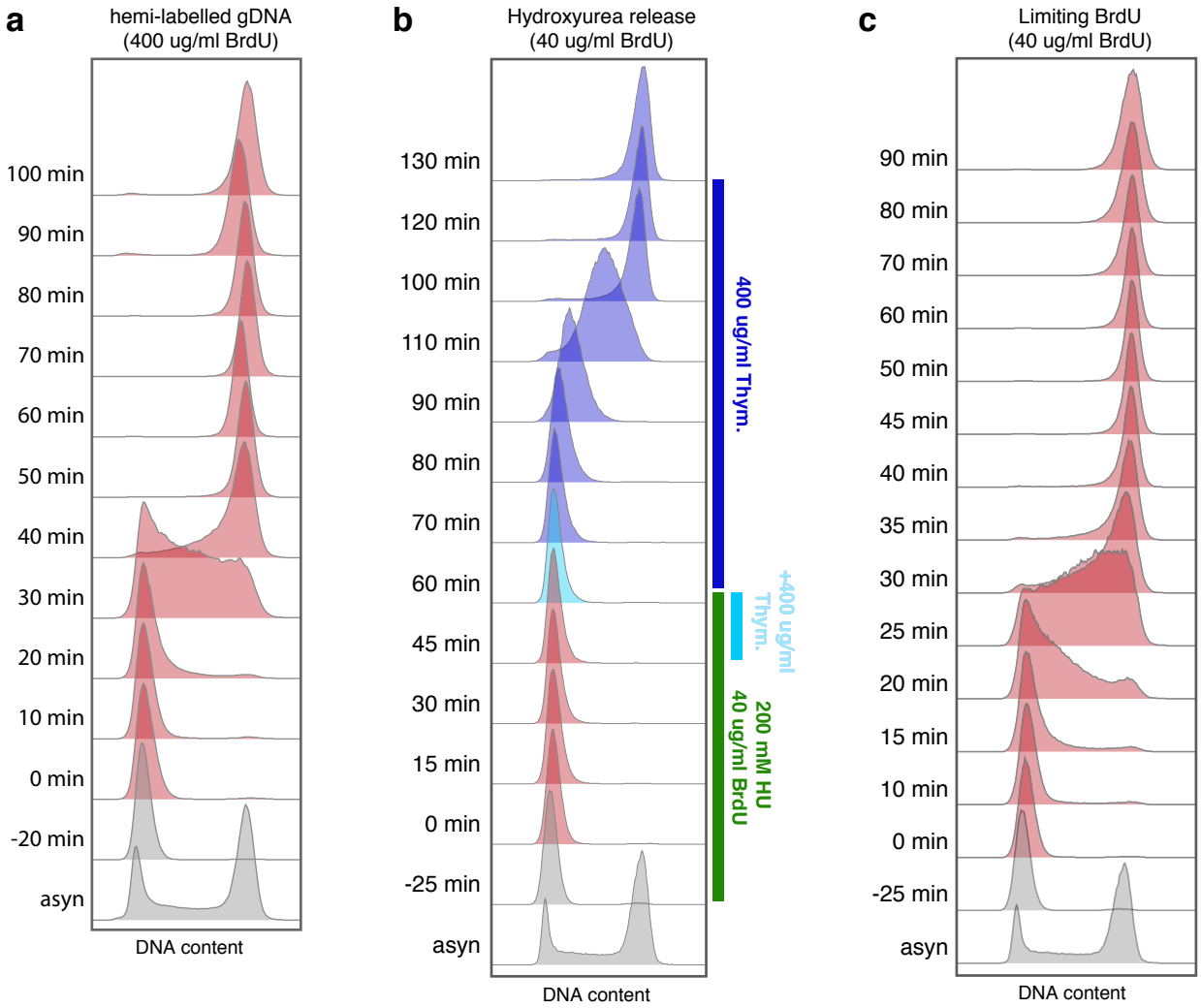
Supplemental Fig. S6



Supplemental Fig. S7

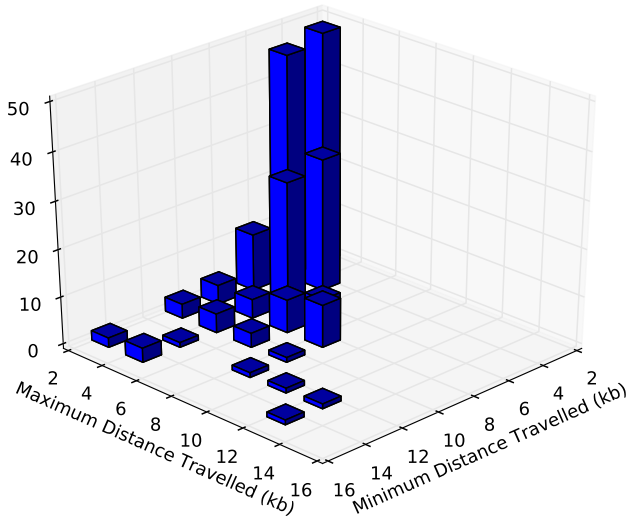


Supplemental Fig. S8

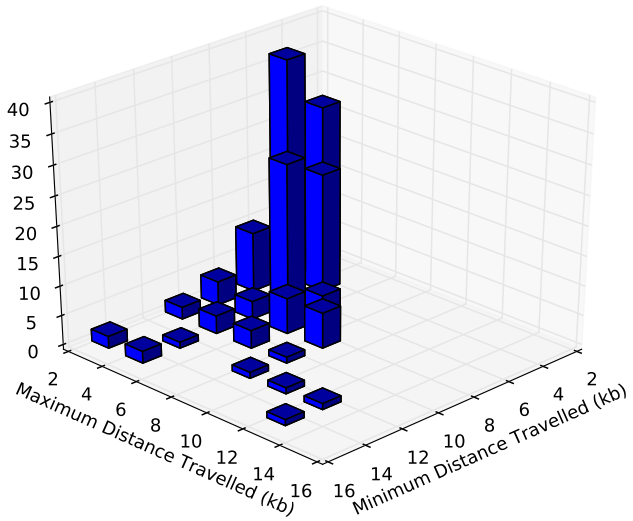


Supplemental Fig. S9

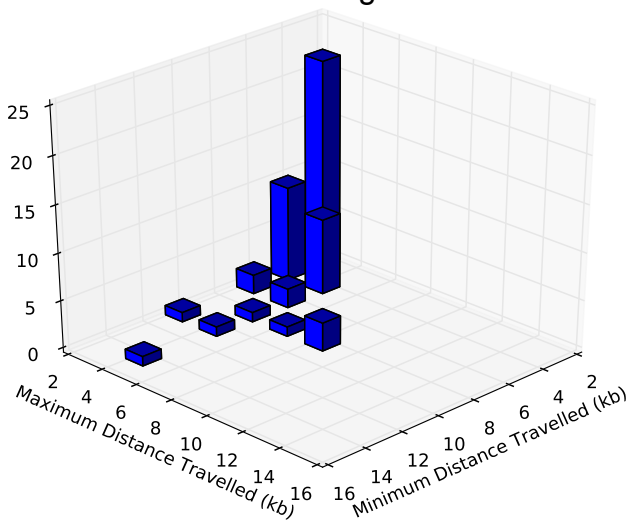
All origins



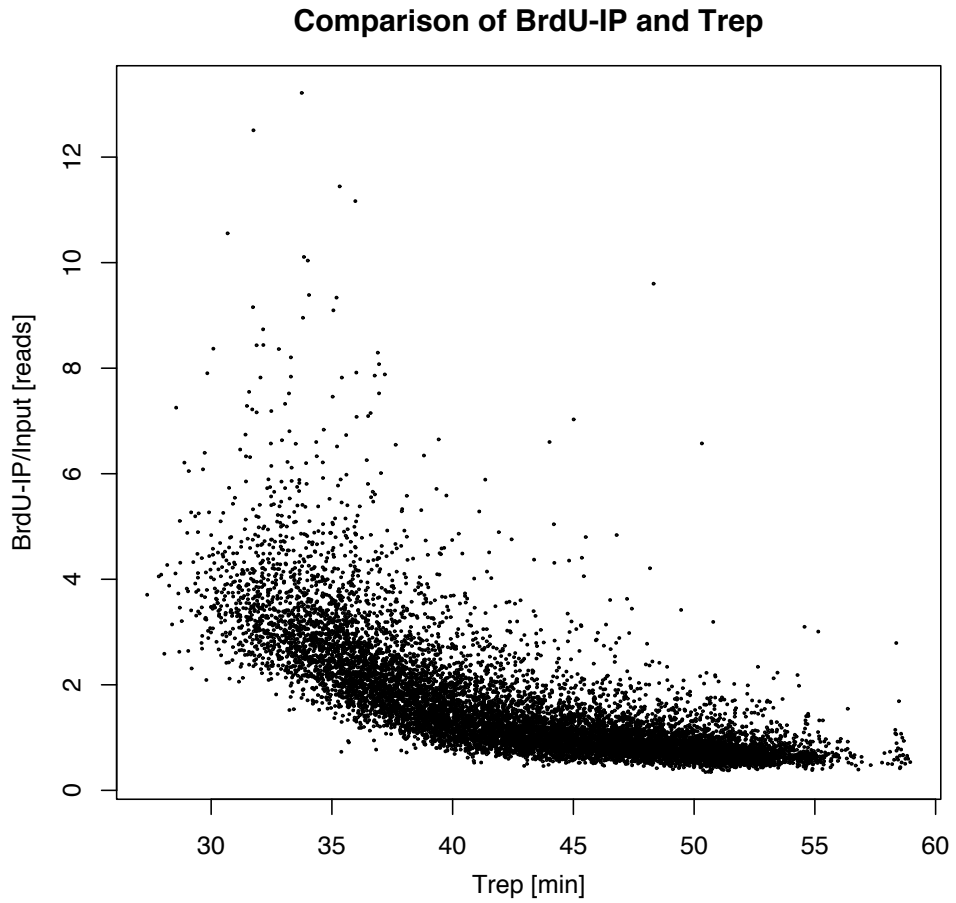
Unchecked origins



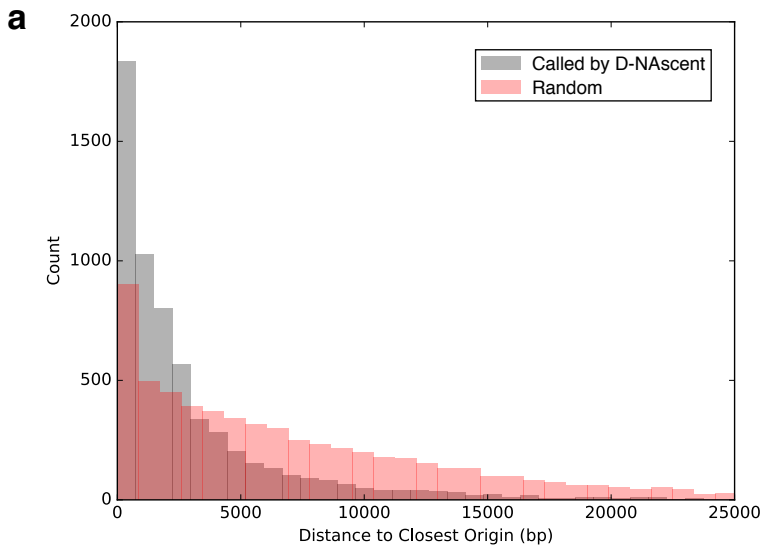
Checked origins

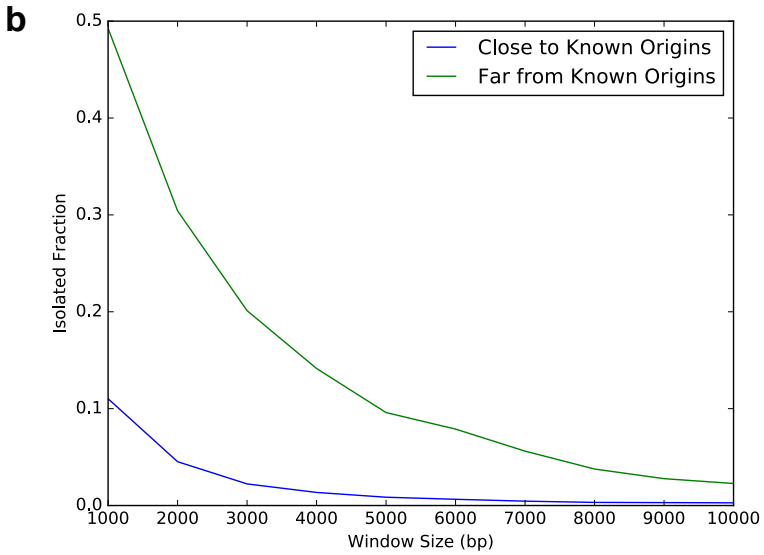


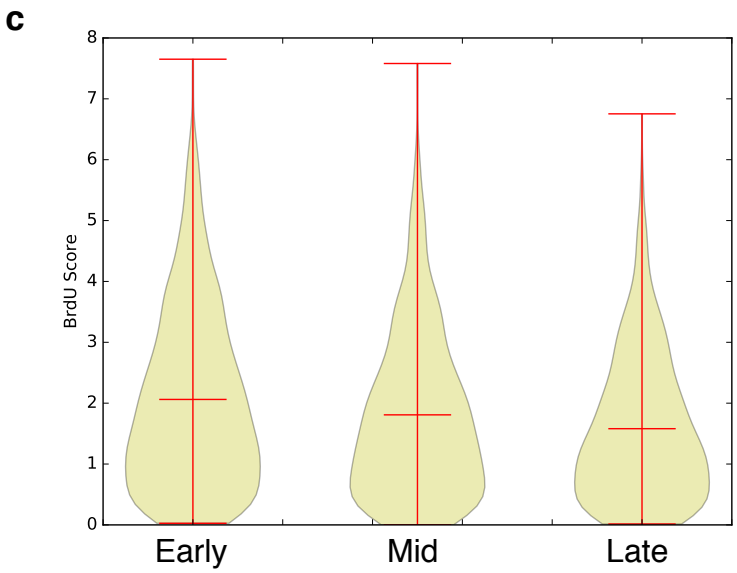
Supplemental Fig. S10



Supplemental Fig. S11







Supplemental Fig. S12