Supplemental material

RNA-DNA hybrid (R-loop) immunoprecipitation mapping:

an analytical workflow to evaluate inherent biases

László Halász, Zsolt Karányi, Beáta Boros-Oláh, Tímea Kuik-Rózsa, Éva Sipos, Éva Nagy, Ágnes Mosolygó-L, Anett Mázló, Éva Rajnavölgyi, Gábor Halmos, Lóránt Székvölgyi *

* Correspondence:
Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Hungary; 4032 Debrecen Hungary, Egyetem tér 1.
E-mail: lorantsz@med.unideb.hu

Table of contents

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
S9.6	S9.6 ATCC ® HB-8730™ Mus musculus (B cell)	HB-8730
Goat anti-mouse IgG marked with HRP	Santa Cruz Biotechnology	sc-2005
Rabbit anti-mouse IgG Alexa647	Thermo Fisher Scientific	A16168
Chemicals, Peptides, and Recombinant Proteins		
UltraPure Paraformaldehyde	Sigma-Aldrich	P6148-500G
RNase A solution, 10 mg/ml	UD-GenoMed Ltd.	UDV0322
RNase H	New England Biolabs	M0297L
Proteinase K	Thermo Fisher Scientific	EO0492
HindIII, EcoRI, BsrGI, XbaI and SspI	New England Biolabs	*https://www.neb.com/*
Dynabeads Protein A, 5 ml	Life Technologies	10002D
Penicillin-Streptomycin Solution, 100 ml	Sigma-Aldrich	P4333-100ML
ROX solution 50 uM 1ml	Thermo Scientific	34094
Nitrocellulose Blotting Membrane	GE Healthcare	10600020
Phenol-chloroform-isoamyl alcohol mixture	Sigma-Aldrich	77617-100ml
RPMI-1640	Sigma-Aldrich	R8758-500ML

Critical Commercial Assays		
NucleoSpin Tissue Kit	Macherey-Nagel	740952.50
LightCycler 480 SYBR Green I Master, 2x qPCR master mix	Roche	4887352001
Naive CD4+ T-cell Isolation Kit	Miltenyi Biotec	130-094-131
NucleoSpin Gel and PCR Clean-up (250 preps)	Macherey-Nagel	740609.250
SuperSignal West Femto Trial Kit	Thermo Scientific	34094
Deposited Data		
Raw and analyzed data	This paper	SRP095885
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	*http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/*
Human GRCh37 blacklisted regions	ENCODE Consortium	ENCFF419RSJ
NTERA2 Chromatin State Data	ENCODE Consortium	ENCSR403MYH
MCF7 Chromatin State Data	Taberlay et al., 2014	GEO: GSE57498
K562, IMR90, HEK and Primary Fibroblast Core 15-State Models	NIH Roadmap Epigenomics Mapping Consortium	E123, E017, E086 and E055
NTERA2 DRIP-sequencing data	(Ginno et al., 2012)	GEO: GSE45530
NTERA2 DRIP-sequencing data	(Sanz et al., 2016)	GEO: GSE70189

K562 DRIP-sequencing data	(Sanz et al., 2016)	GEO: GSE70189
IMR90 DRIP-sequencing data	(Nadel et al., 2015)	GEO: GSE68953
HEK DRIP-sequencing data	(Nadel et al., 2015)	GEO: GSE68953
Primary Fibroblast DRIP-sequencing data	(Lim et al., 2015)	GEO: GSE57353
MCF7 DRIP-sequencing data	(Stork et al., 2016)	GEO: GSE81851
Human Protein coding genes, exons and introns	Ensembl	http://www.ensembl.org/
List of Repetitive Elements	UCSC; RepeatMasker	https://genome.ucsc.edu/
Experimental Models: Cell Lines		
Human Jurkat T-lymphoblastoid cell line	Sigma-Aldrich	https://www.sigmaaldrich.com/
Human Naive CD4+ T-cells	This study	N/A
Human GM12878 B-lymphoblastoid cell line	Coriell Institute	https://www.coriell.org/
Sequence-Based Reagents		
TruSeq ChIP Sample Preparation Kit	Illumina	www.illumina.com
Primers for DRIP-qPCR, see Table S3	This study	N/A
Software and Algorithms		
R programming language	R core team	https://www.r-project.org/
ggplot2	(Wickham, 2009)	https://cran.r-project.org/
ggbio	(Yin, Cook, & Lawrence, 2012)	https://bioconductor.org/

GenomicRanges	(Lawrence et al., 2013)	https://bioconductor.org/
DECIPHER	(Wright, 2016)	https://bioconductor.org/
Circos	(Krzywinski et al., 2009)	http://www.circos.ca/
BWA	(Li & Durbin, 2009)	http://bio-bwa.sourceforge.net/
SAMtools	(Li et al., 2009)	http://samtools.sourceforge.net/
Picard tools		https://broadinstitute.github.io/picard/
BEDTools	(Quinlan & Hall, 2010)	http://bedtools.readthedocs.io/en/latest/
MACS2	(Zhang et al., 2008)	https://github.com/taoliu/MACS
deepTools	(Ramírez, Dündar, Diehl, Grüning, & Manke, 2014)	https://github.com/fidelram/deepTools
pROC	(Robin et al., 2011)	https://cran.r-project.org/

Supplemental Methods

Cell cultures

The Jurkat human T lymphoblastoid cell line was cultured in RPMI-1640 medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum albumin (BSA), L-glutamine and gentamycin at 37 °C, in a humidified / 5% CO2 chamber. 100 million exponentially growing cells were washed twice with 1 x PBS and divided into equal aliquots for the twenty-four DRIP experiments. In DRIP experiments #5 and #13, the GM12878 B lymphoblastoid cell line was used for comparison with the Jurkat cell line. The GM12878 cells obtained from the Coriell Institute for Medical Research. Cells were grown at 37 °C in a humidified / 5% CO2 chamber in vented 25 cm^2 cell culture flasks, containing 10-20 mls of RPMI-1640, L-glutamine and gentamycin.

Naive CD4+ T-cell Isolation

Leukocyte enriched buffy coats were obtained from healthy blood donors (individual donations) drawn at the Regional Blood Center of the Hungarian National Blood Transfusion Service (Debrecen, Hungary) in accordance with the written approval of the Director of the National Blood Transfusion Service and the Regional and Institutional Ethics Committee of the University of Debrecen, Medical and Health Science Center (Hungary). PBMCs were separated by a standard density gradient centrifugation with Ficoll-Paque Plus (Amersham Biosciences, Uppsala, Sweden). Naive T-cells were separated from human blood mononuclear cells using the naive $CD4^+$ T-cell isolation kit based on negative selection according to the manufacturer's instruction (Miltenyi Biotec). Using the $CD4^+$ T Cell Isolation Kit, human $CD4^+$ T helper cells are isolated by negative selection. Non-target cells are labeled with a cocktail of biotin-conjugated monoclonal antibodies and the $CD4^+$ T Cell MicroBead Cocktail. The magnetically labeled non-target T cells are depleted by retaining them on a MACS® Column in the magnetic field of a MACS Separator, while the unlabeled T helper cells pass through the column.

Detection of RNA-DNA hybrids by Dot Blot Assay

For dot blot immunoassays 6 µg of phenol-chloroform extracted and sonicated Jurkat nucleic acid were treated with 1 µl RNase A (10 mg/ml; UD-GenoMed Ltd.) in TE buffer with different salt concentration (from 25 mM to 500 mM) at 37 °C for 1 hour. The RNase H digestion was performed using 1-8 µl of RNase H (5000 U/ml; NEB) at 37 °C, overnight. For control samples, we used sonicated nucleic acid without any further treatment, alkali-treated sonicated nucleic acid (incubated with 1 µl of 50 mM NaOH at 65 °C for 10 min) and sonicated nucleic acid resuspended in 1x RNase

H Reaction Buffer (NEB) at incubated at 37 °C without RNase H enzyme. Both the RNase A and RNase H digested samples were spotted on a nitrocellulose membrane (GE Healthcare) at two different concentrations (600 ng and 125 ng) in triplicates. After drying the spots at room temperature, the membrane was fixed with UV for 5 minutes and blocked by 5% bovine serum albumin (BSA) in PBST buffer (PBS containing 0.25 % Tween) at room temperature for 20 minutes. The blocked membrane was then incubated with the S9.6 antibody in PBST buffer containing 5% bovine serum albumin at room temperature for 2 hours. The membrane was washed five times with PBST and incubated with goat anti-mouse IgG marked with HRP (Santa Cruz Biotechnology) at room temperature for 1 hour. After five washes with PBST, the signal was detected by the SuperSignal West Pico Chemiluminescent substrate (Thermo Fisher Scientific) and imaged using FlourChemQ (ProteinSimple).

Immunofluorescent Labeling of RNA-DNA hybrids

Jurkat cells were resuspended in pre-warmed (37 °C) 1 x PBS to a density of $6x10^6$ / ml and diluted 4-fold in 1% molten low melting point agarose dissolved in PBS. 22 µl of the cell suspension (~ 33.000 cells) was spread in each well of an 8 chamber Ibidi slide. After the gel set, agarose embedded cells were washed three times in PBS (500 µl / well) on ice. The permeabilization, lysis and nuclei preparation were performed in one step using a Lysis Buffer consisting of 1 % (v/v) TritonX-100, and 2 M NaCl in PBS/EDTA (500 µl/well, 10 minutes on ice). The samples were blocked by 5 mg/ml BSA dissolved in PBS / 5mM EDTA, on ice for 30 minutes. RNA-DNA hybrids were labeled by the S9.6 monoclonal mouse antibody and a rabbit anti-mouse Alexa647 secondary antibody. Imaging was carried out using a Zeiss Axiovert 200M confocal laser scanning microscope. Signal intensities were quantified by ImageJ.

Step-by-step protocol of the best-performing DRIP experiment (exp. 5)

Crosslinking of cells was done with 1% paraformaldehyde (UP) for 10 minutes, then quenched with 2.5 M glycine (pH 6, final concentration: 500 mM) for 5 minutes at room temperature. Cells were lysed in the lysis buffer provided by the NucleoSpin Tissue kit (Macherey-Nagel) at 65°C for 7 h (according to the kit protocol), or at 37°C overnight (where indicated in the main text).Total nucleic acid was isolated by a NucleoSpin Tissue Kit (Macherey-Nagel) and eluted in 100 µl of elution buffer (5 mM Tris-HCl pH 8.5). The purified nucleic acid prep was fragmented by sonication in 300 µl of Tris-HCl pH 8.5 (high salt concentration: 300 mM NaCl) for 2 x 5 min (30 sec ON, 30 sec OFF, LOW, Bioruptor) to yield an average DNA fragment size of ~500 bp. Fragment analysis was done by using 1 % agarose gelelectophoresis. If it was necessary, further sonication was applied. The

sonicated DNA sample was purified by a NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) and eluted in 100 µl of elution buffer (5 mM Tris-HCl pH 8.5). Twelve micrograms of DNA was diluted with 5 mM Tris-HCl pH 8.5 to a total volume of 100 µl. Two percent of the sample was kept as input DNA. Half of the sample was treated with 8 µl of RNase H (5000 U/ml; NEB) in a total volume of 80 µl at 37 °C, overnight. Dynabeads Protein A magnetic beads (ThermoFisher Scientific) were pre-blocked with PBS/5 mM EDTA containing 0.5% BSA. To immobilize the S9.6 antibody, 50 µl pre-blocked Dynabeads Protein A was incubated with 10 µg of S9.6 antibody in IP buffer (50 mM Hepes/KOH at pH 7,5; 0,14 M NaCl; 5 mM EDTA; 1% Triton X-100; 0,1 % Na-Deoxycholate, ddH2O) at 4°C for 4 hours with rotation. Six micrograms of genomic DNA was added to the mixture and gently rotated at 4°C, overnight. Beads were recovered and washed successively with 1ml lysis buffer 1 (low salt, 50 mM Hepes/KOH pH 7.5, 0.14 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1 % Na-Deoxycholate), 1ml lysis buffer 2 (high salt, 50 mM Hepes/KOH pH 7.5, 0.5 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1 % Na-Deoxycholate), 1ml wash buffer (10 mM Tris-HCl pH 8, 0.25M LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1 mM EDTA pH 8) and 1ml TE (100 mM Tris-HCl pH 8, 10 mM EDTA pH 8) at 4°C, two times. Elution was performed in 100 µl of elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1 % SDS) for 15 min at 65 ˚C. After purification by NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel), nucleic acids were eluted in 55 µl of elution buffer (5 mM Tris-HCl pH 8.5). The recovered DNA was then analyzed by quantitative real-time PCR (qPCR). qPCR was performed with LightCycler 480 SYBR Green I Master (Roche) and analyzed by a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific). The qPCR data were evaluated by the comparative C_T method. RNA-DNA hybrid enrichment was calculated based on the IP/Input ratios.

DNA-RNA immunoprecipitation (DRIP) sequencing

DRIP-seq libraries were prepared according to the Illumina's TruSeq ChIP Sample Preparation protocol. Briefly, the enriched DRIP DNA was end-repaired and indexed adapters were ligated to the inserts. Purified ligation products were then amplified by PCR. Amplified libraries were prepared and sequenced in the Genomic Medicine and Bioinformatics Core Facility of the University of Debrecen (1x50 bp read length, single-end, Illumina HiScan SQ) and at the EMBL Genomics Core Facility, Heidelberg (2x150 bp read length, paired-end, Illumina HiSeq 2500).

Sequenced reads were aligned to the Human reference genome (hg19) using default parameters of BWA MEM (Li & Durbin, 2009) algorithm. Low mapping quality, supplementary alignments, reads mapped to blacklisted regions and redundant reads were omitted (Li et al., 2009; Quinlan & Hall,

2010) from downstream analysis. Replicate experiments (rep1, rep2) were merged and then MACS2 (Zhang et al., 2008) was used to identify enriched regions (at FDR 1%) of the genome normalized to input datasets.

Processed and merged alignments were subjected to bamCoverage (Ramírez et al., 2014) to generate signal files. RPKM (Reads Per Kilobase Million) values were calculated for 20 bp bins for each sample and smoothed using a 60 bp sliding window (--binSize 20 --smoothLength 60 –normalizeUsingRPKM --extendReads 300). The generated signal files were visualized in R 3.2.2, using the ggplot2 (Wickham, 2009) and ggbio (Yin et al., 2012) packages.

Genomic Annotation of RNA-DNA hybrids

We used GenomicRanges (Lawrence et al., 2013) to determine the genomic distribution of DRIP peaks, allowing us to calculate the intersecting area between binding sites and the corresponding annotation categories. Areas occupied by the intersected regions were compared to a randomized peak coverage. Random peak sets were generated for each chromosome by permutation, considering the chromosomal distribution of chromatin states and omitting blacklisted regions.

***In silico* restriction enzyme digestion**

To calculate the expected (theoretical) fragment length distribution generated by a combination of restriction enzymes (HindIII, EcoRI, BsrGI, XbaI and SspI), we cut the human (hg19) and yeast (sacCer3) genomes *in silico* with the DECIPHER R package (Wright, 2016). From the cutting site positions, we calculated the length of restriction fragments. Statistical comparison of the resulting fragment length distributions was performed by the Wilcoxon Rank Sum test by randomly sampling 300 values 100 times. P-values were adjusted with Benjamini & Hochberg correction.

Supplemental References

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009). Circos: an information esthetic for comparative genomics. *Genome Res*, *19*(604), 1639–1645. https://doi.org/10.1101/gr.092759.109

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., … Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, *9*(8), 1–10. https://doi.org/10.1371/journal.pcbi.1003118

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(W1), 1–5. https://doi.org/10.1093/nar/gku365

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

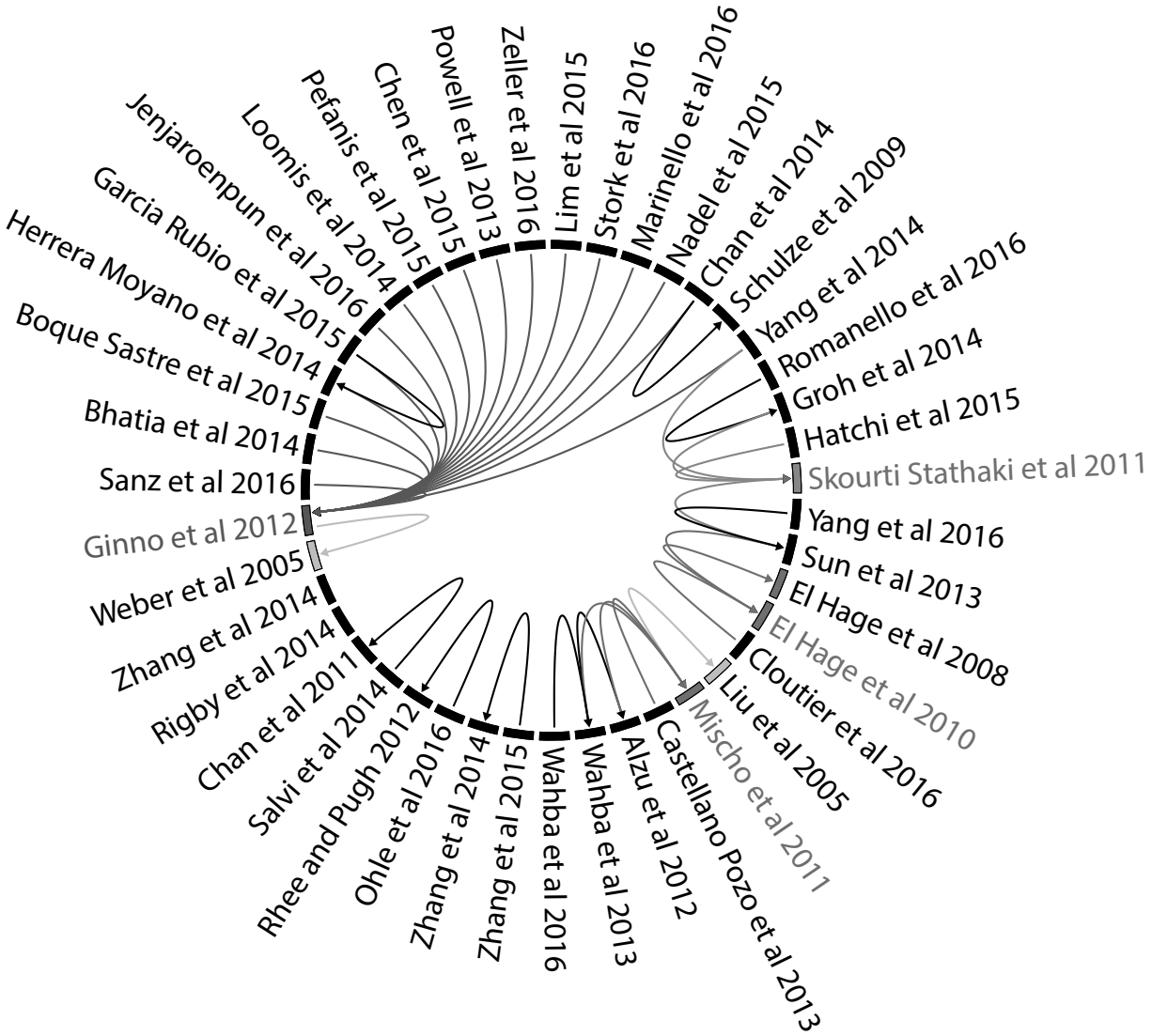
Wickham, H. (2009). Ggplot. *Media*, *35*(July), 211. https://doi.org/10.1007/978-0-387-98141-3

Wright, E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, *8*(1), 352–359.

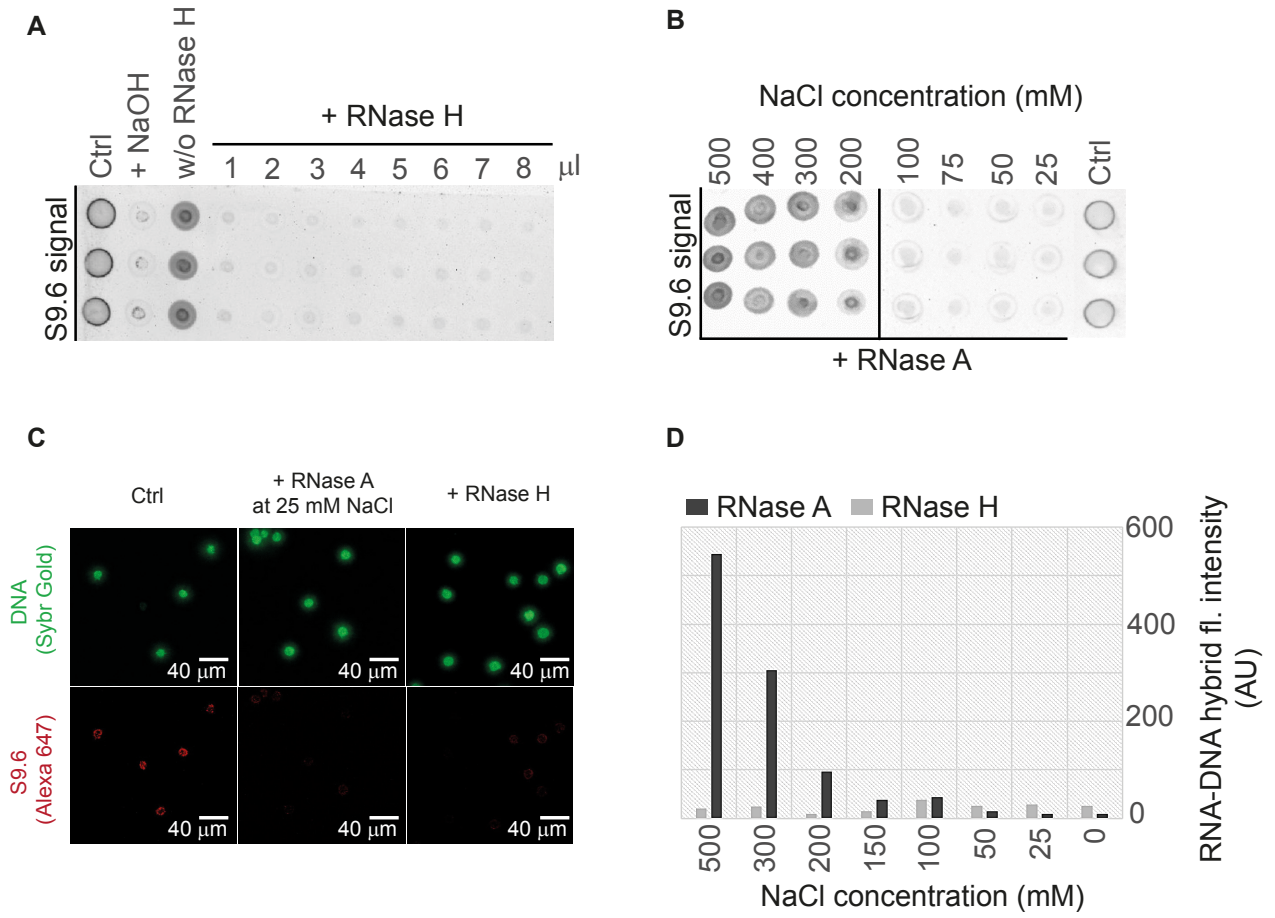
Yin, T., Cook, D., & Lawrence, M. (2012). ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biology*, *13*(8), R77. https://doi.org/10.1186/gb-2012-13-8-r77

Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

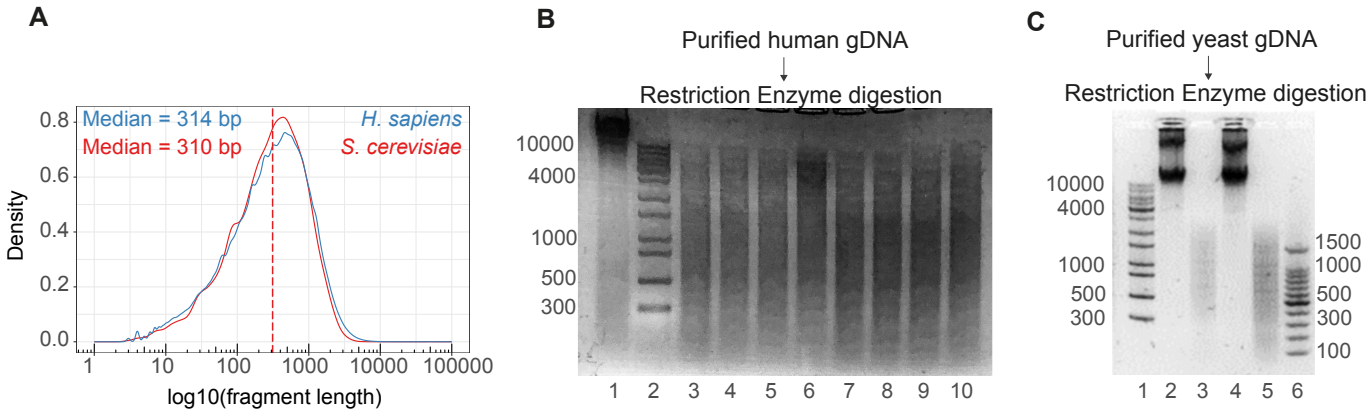
Supplemental Figures and Figure Legends



Supplemental Figure S1. Reference Network of the DRIP Experiments Based on Scientific Literature. The Circos plot shows the DRIP studies (nodes) and the DRIP methods referenced by each study (arrows). Base-nodes (in bold) point to studies that receive many citations; most DRIP experiments originate from 2-3 chief studies. Light-gray edge highlights the MeDIP approach (methylated DNA immunoprecipitation, Weber et al 2005) forming the basis of the original DRIP protocol. A detailed summary of the studies is given in Supplemental Table S1.

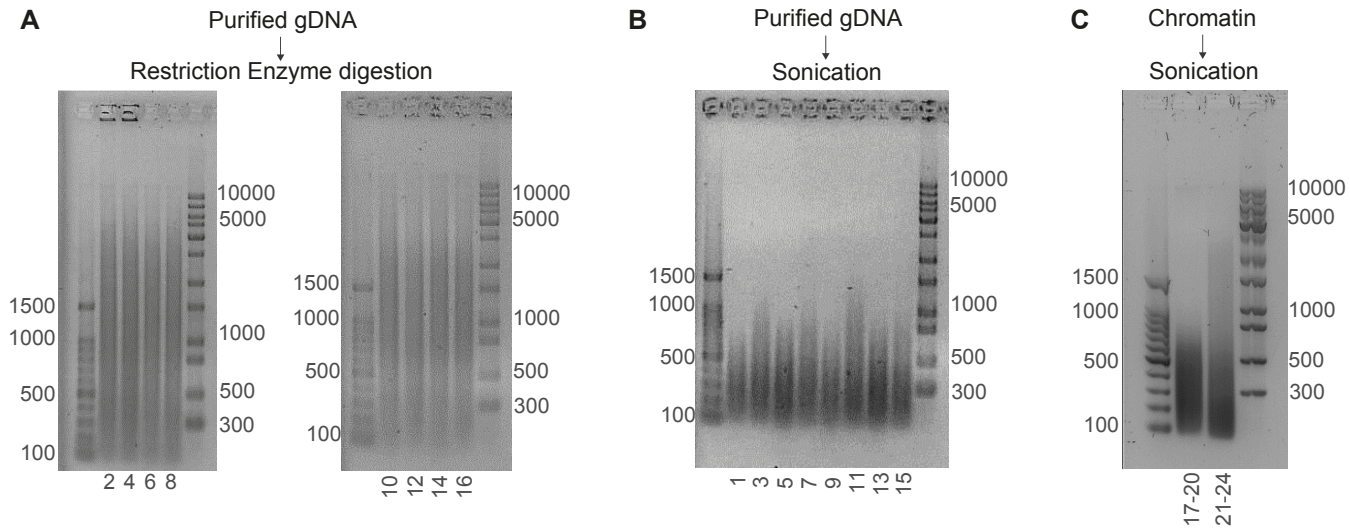


Supplemental Figure S2. The Effect of Ribonucleolytic Treatment on the Level of R-loops.
(*A*) S9.6 dot blot hybridization showing the decrease of RNA-DNA hybrid level as a result of RNase H digestion. Each spot contains 5 µg of sonicated nucleic acid pipetted in triplicates onto the membrane. The first three columns represent control assays: (1. no treatment; 2: alkaline hydrolysis of free RNA and RNA-DNA hybrids by 50 mM sodium hydroxide (NaOH); 3: buffer control (w/o RNase H). The remaining columns show the effect of RNase H added in increasing amount.
(*B*) S9.6 dot blot hybridization showing the decrease of RNA-DNA hybrid level as a result of RNase A digestion. RNA-DNA hybrids become sensitive to RNase A as a function of decreasing monovalent (NaCl) concentration. Last column: negative (buffer only) control.
(*C*) Confirmation of the salt-dependent RNase H-like hybridase activity of RNase A by fluorescent microscopy. Permeabilized Jurkat cells were treated with RNase A (at low ionic strength) or RNase H, and RNA-DNA hybrid were immunofluorescently labeled by the S9.6 antibody. Green channel: DNA stained by SybrGold. Red channel: rabbit anti-mouse Alexa647 secondary antibody.
(*D*) Microscopic quantification of S9.6 signal intensities upon RNase A digestion performed at decreasing NaCl concentrations. The majority of RNA-DNA hybrids were not destroyed above 300 mM NaCl, but became efficiently digested below 100 mM NaCl, in line with the dot blot hybridization results. In parallel with each RNase A digestion reactions, nuclear preps were treated with RNase H (as a negative control).

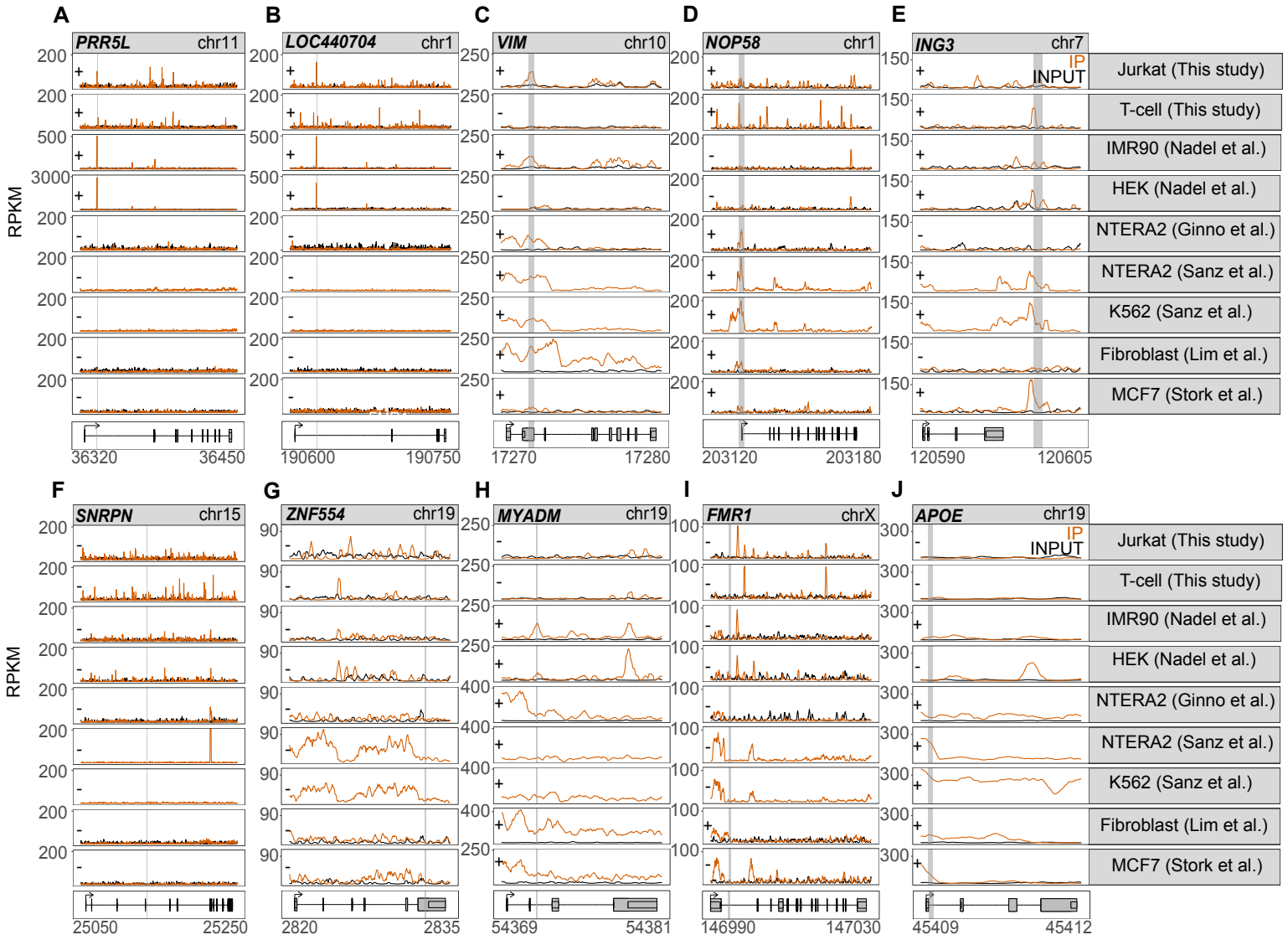


Supplemental Figure S3. Distribution of DNA Fragment Length of *Homo Sapiens* and *Saccharomyces cerevisiae* After Restriction Enzyme Cleavage.
(*A*) Theoretical DNA fragment size distribution as a result of an *in silico* restriction enzyme cocktail fragmentation. The obtained fragment length distributions are similar in both species (Wilcoxon rank sum test: p=0.944, not significant) with a median size of 310-314 bp.
(*B*-*C*) Restriction fragment length distribution obtained as a result of restriction enzyme cocktail fragmentation in a real digestion reaction. Digestions were performed by a mix of HindIII, EcoRI, BsrGI, XbaI and SspI on genomic DNA purified from human and budding yeast cells, respectively. The observed DNA fragment length distribution of yeast DNA matches with the theoretical distribution, which is not the case for human DNA samples.
Lanes of the agarose gel in panel *B*:
1. Undigested gDNA
2. 1Kb Plus DNA ladder
3. 5µg gDNA + 20U for each restriction enzyme in NEB2.1 buffer
4. 5µg gDNA + 20U for each restriction enzyme in NEB2.1 buffer + 0.1mg/ml BSA
5. 5µg gDNA + 20U for each restriction enzyme in Tango buffer
6. 5µg gDNA + 20U for each restriction enzyme in CS buffer
7. 5µg gDNA + 40U for each restriction enzyme in NEB2.1 buffer
8. 5µg gDNA + 40U for each restriction enzyme in NEB2.1 buffer + 0.1mg/ml BSA
9. 5µg gDNA + 40U for each restriction enzyme in Tango buffer
10. 5µg gDNA + 40U for each restriction enzyme in CS buffer

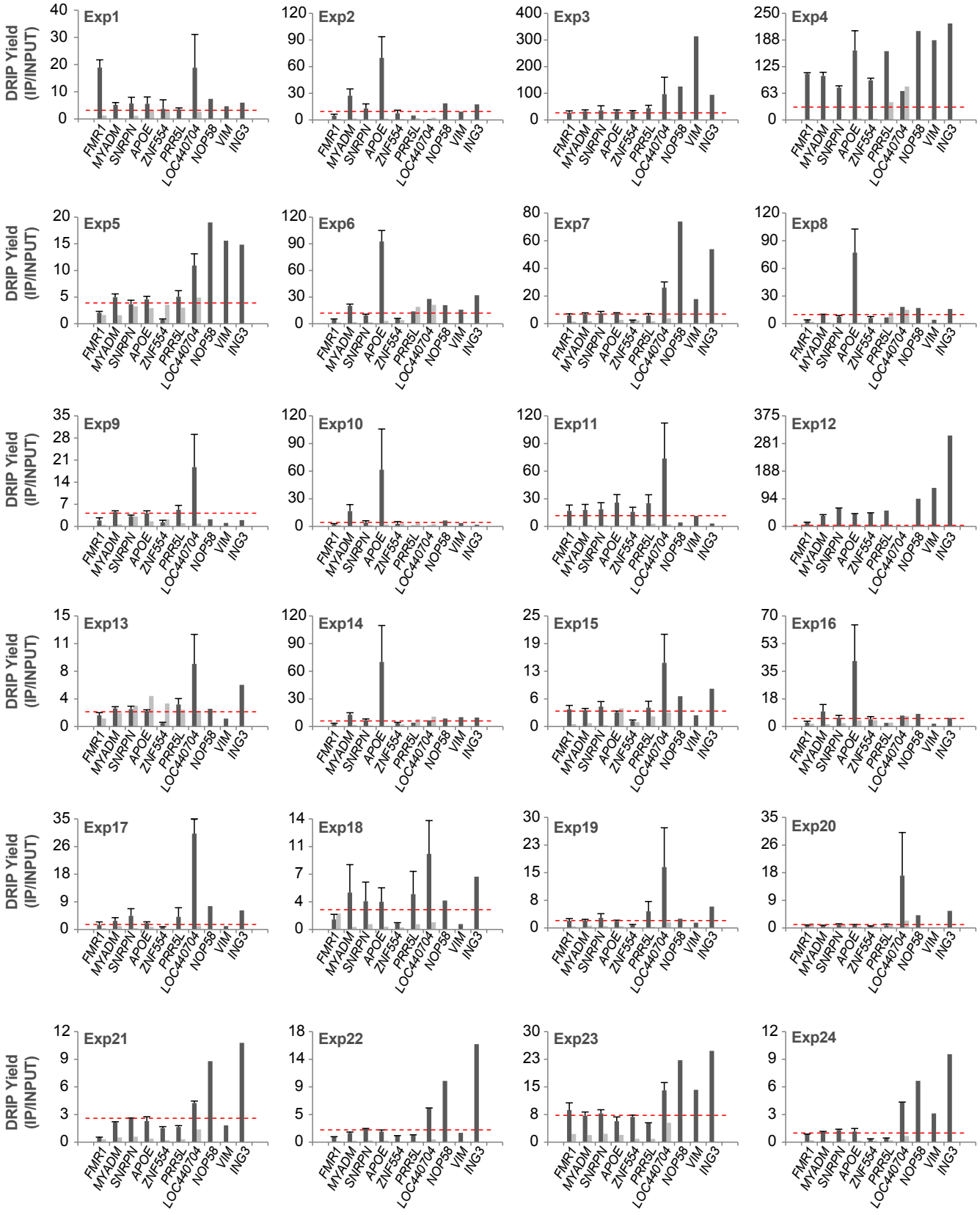
Lanes of the agarose gel in panel *C*:
1. 1Kb Plus DNA ladder
2. Undigested DNA (BY4741)
3. 2µg BY4741 gDNA + 20U for each restriction enzyme in NEB2.1 buffer
4. Undigested gDNA (BY4742)
5. 2µg BY4742 gDNA + 20U for each restriction enzyme in NEB2.1 buffer
6. 100 bp DNA ladder



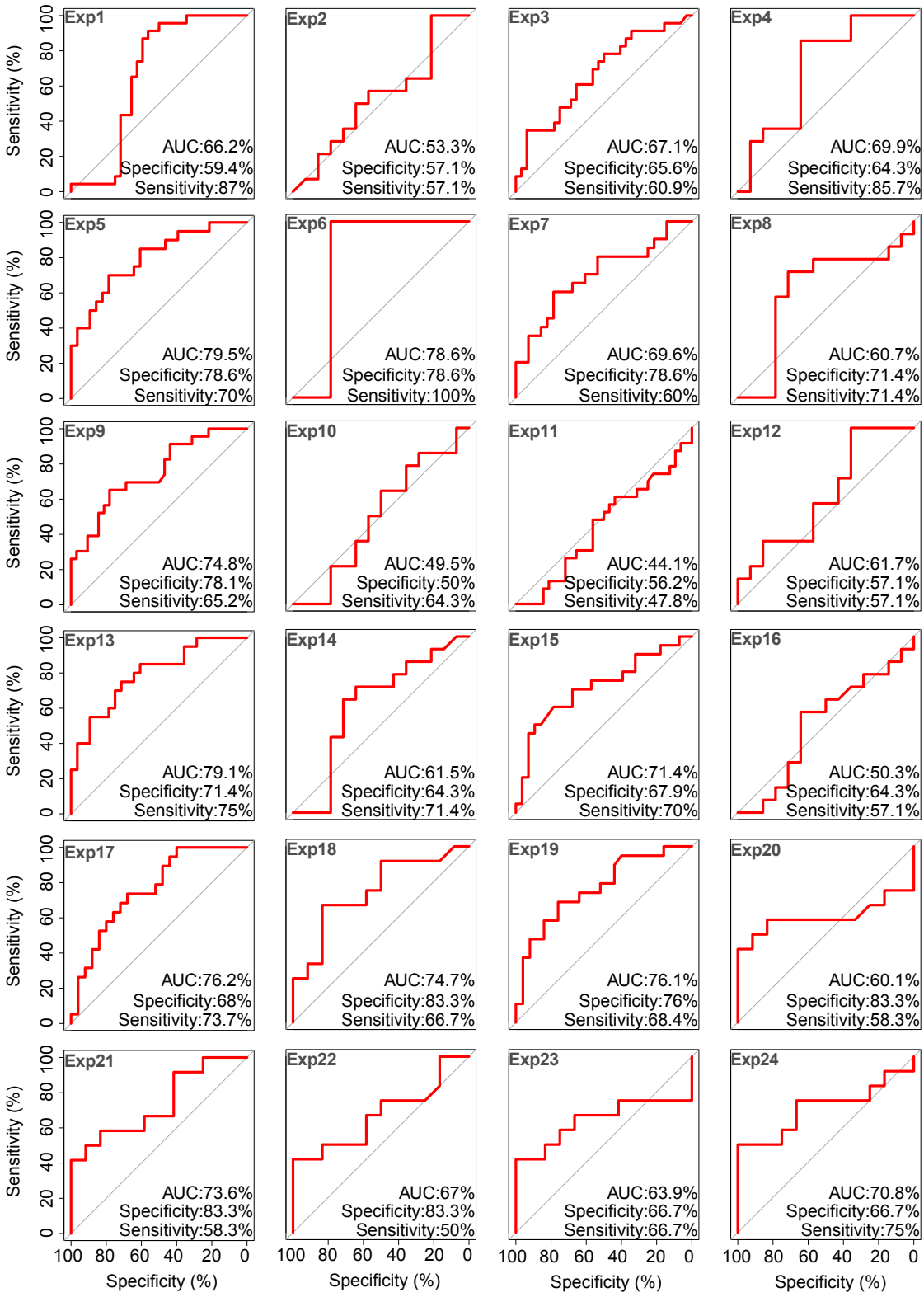
Supplemental Figure S4. Comparison of the Effect of Restriction Enzyme Cocktail Fragmentation and Sonication to *Homo sapiens* Genomic DNA.
(*A*-*B*) DNA fragment length distribution as a result of restriction enzyme cocktail fragmentation (*A*) and sonication (*B*). Both restriction enzyme digestions and sonication were performed on naked genomic DNA purified from Jurkat cells.
(*C*) Fragmentation of chromatin by sonication rather than naked DNA. The numbers below the agarose gels indicate the relevant DRIP scheme IDs.



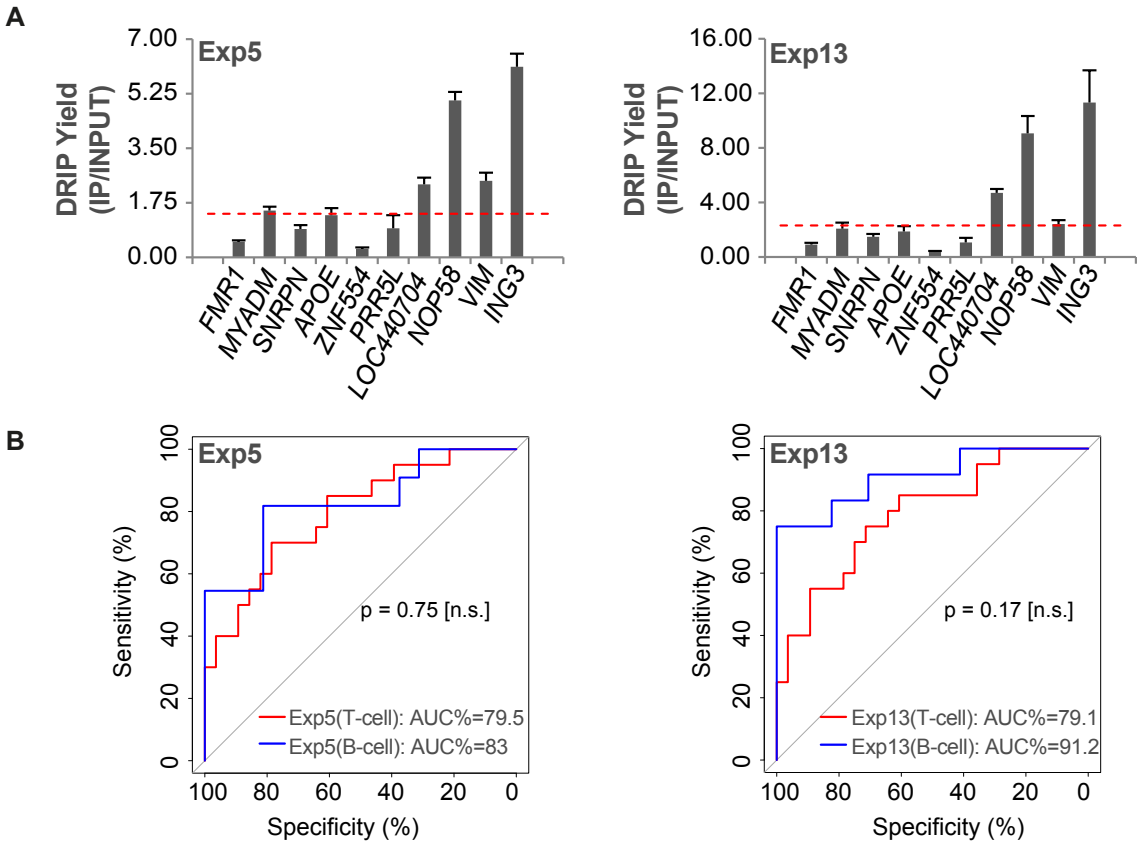
Supplemental Figure S5. Making a Reference R-loop Set by DRIP-Seq Mapping.
DNA-RNA hybrid mapping (DRIP-seq) was performed in two closely related human cell types (Jurkat T-cell leukemia cell line and T CD4+ lymphocytes). Genome browser tracks show the IP (orange) and input (gray) signals over the selected test regions. (*A*-*J*) Test loci used for benchmarking the DRIP classifiers. DRIP profiles over the same test regions, obtained in other cell types, are also displayed. Test regions that are positive or negative for the presence of an R-loop (gray shading) are indicated by + and -, respectively. RPKM: reads per kilobase per million reads. Locus names and chromosome numbers are indicated on the top of each panel. Vertical light-gray boxes highlight the regions tested by DRIP-qPCR.



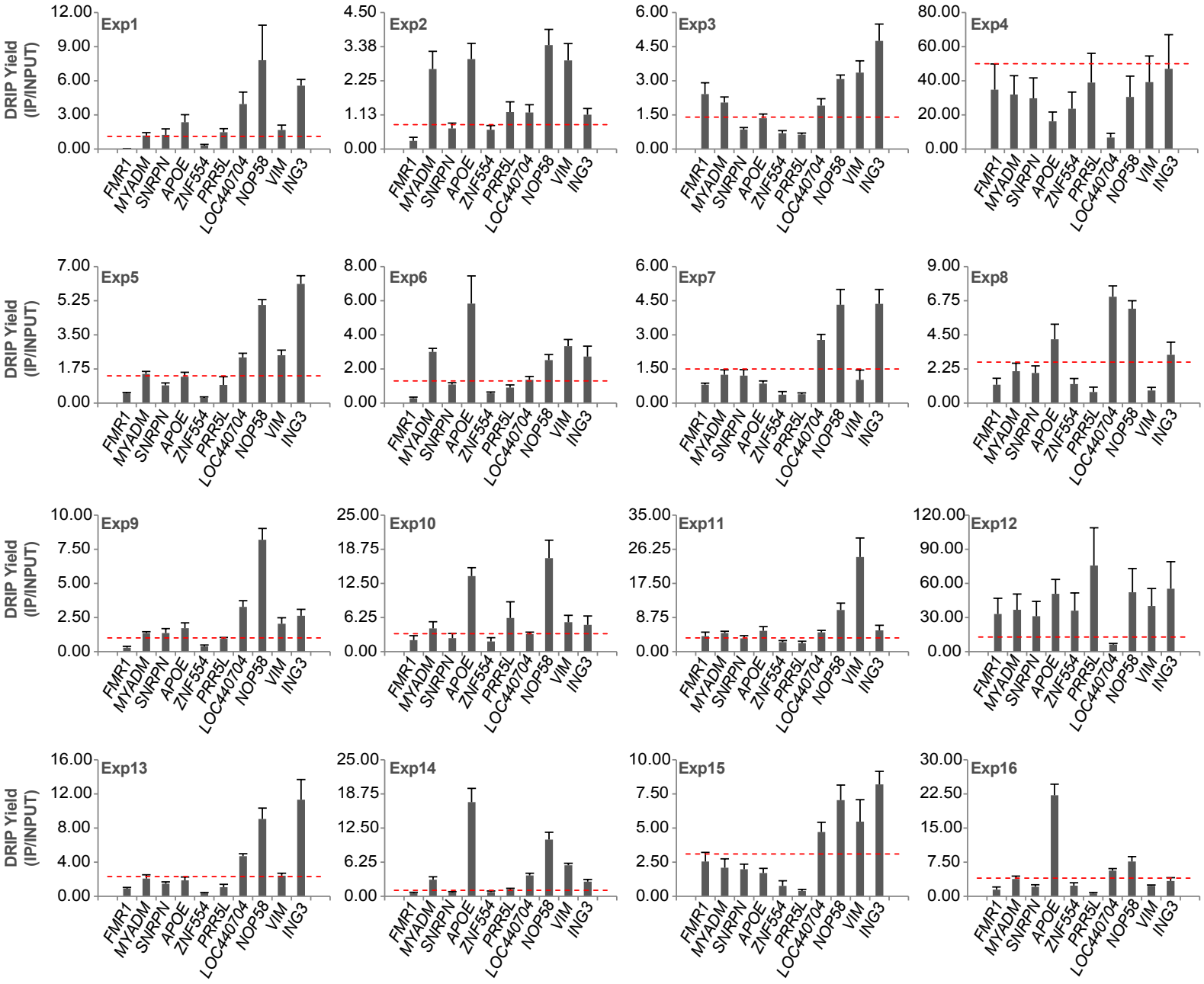
Supplemental Figure S6. DRIP Enrichment Scores Determined by qPCR Over the Test Loci. DRIP-qPCR yield is shown for the twenty-four DRIP experiments over the selected reference loci. Black and grey bars represent DRIP yields from control and RNase H treated nucleic acid samples, respectively. The first five loci are negative controls (based on the lack of DRIP-seq enrichment), while the remaining five loci are positive controls (showing significant R-loop enrichment by DRIP-Seq). Horizontal dotted line represents the cutoff separating the real R-loop signal from background (extracted from the ROC curves, see Supplemental Material, Figure S7). Optimal separation of negative and positive test loci is obtained in exp 5, exp 13, exp17 and exp18. We highlight these methods as “preferred”. On the contrary, positive and negative test loci are not properly distinguished by exp 2, exp 10, exp11 and exp16. We highlight these methods as “not preferred”.



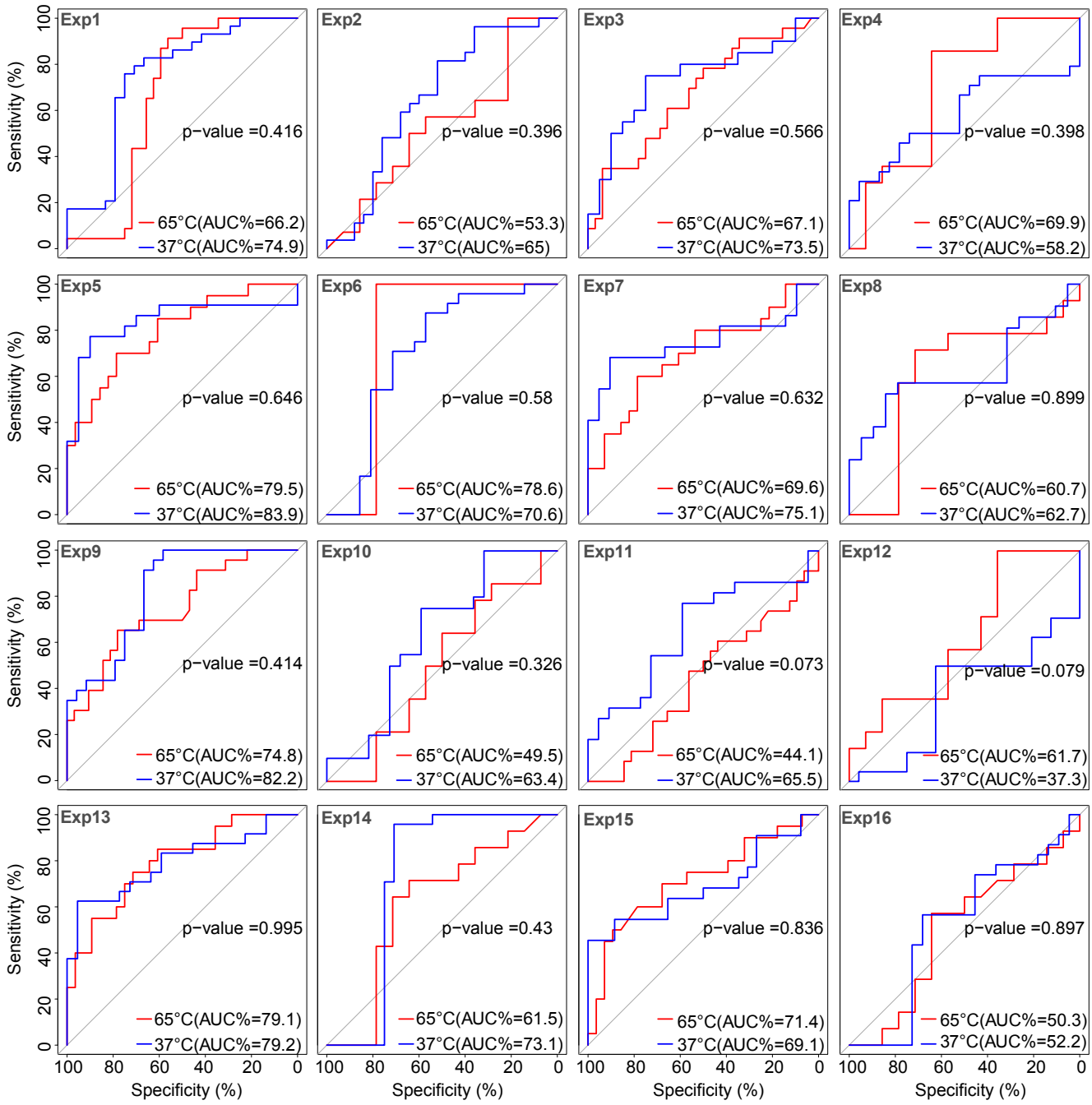
Supplemental Figure S7. ROC Analysis of the DRIP Classifiers.
ROC plots illustrating the efficacy of the twenty-four DRIP protocols. Area under the curve (AUC), specificity and sensitivity are labelled within each plot. The diagonal indicates an AUC of 0.5, corresponding to random answers obtained from the experiments.



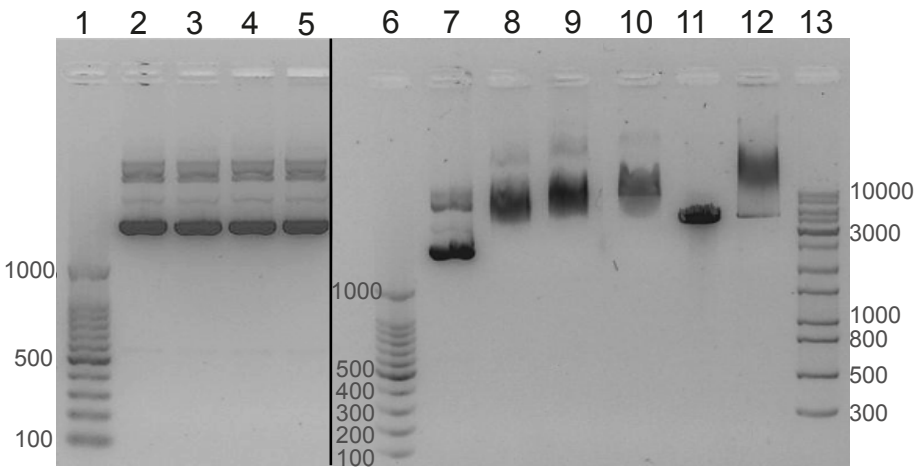
Supplemental Figure S8. The best-performing DRIP protocols work equally well in other cell types. T and B lymphoblastoid cell lines (Jurkat and GM12878) were compared in exp5 and exp13, respectively. (*A*) DRIP-qPCR enrichment scores of the GM12878 cell line displayed over the ten test regions. Horizontal dotted line represent the cutoff value (calculated from the ROC curves) separating the true R-loop signal from background. (*B*) Paired ROC plots comparing the efficacy of the DRIP protocols in Jurkat (red line) and GM12878 (blue line) cells. No significant difference was observed between the cell lines.



Supplemental Figure S9. The effect of cell lysis performed at 37 °C.
DRIP yields were measured by qPCR in sixteen DRIP experiments (with the cell lysis step performed at 37 °C) over the selected reference loci. Horizontal dotted line represents the cutoff separating the real RNA-DNA hybrid signal from background (extracted from the ROC curves, see Supplemental Material, Figure S10).



Supplemental Figure S10. The effect of cell lysis at 65 °C and 37 °C on the specificity and sensitivity of the DRIP experiment.
Paired ROC plots compare the efficacy of sixteen DRIP protocols performed at 65 °C and 37 °C, respectively. None of the tested conditions cause a significant difference between the two temperatures. Area under the curve (AUC), specificity and sensitivity are labelled on each plot.



Supplemental Figure S11. Evidence for the DNA binding of RNase A.
A plasmid DNA incubated with DRIP samples (lanes 2-5) do not show any change in its electrophoretic mobility. Incubation of a plasmid DNA with RNase A (lanes 7-12) significantly changes the electrophoretic mobility via the DNA binding activity of the ribonuclease. The band shift occurs on supercoiled, nicked circular and linearized plasmid templates.
Lanes of the gel:
1. 100 bp marker
2. pmCherry-N1 plasmid (supercoiled form) incubated with DRIP IP sample 5 (65 °C)
3. pmCherry-N1 plasmid (supercoiled form) incubated with DRIP IP sample 5 (37 °C)
4. pmCherry-N1 plasmid (supercoiled form) incubated with DRIP IP sample 7 (65 °C)
5. pmCherry-N1 plasmid (supercoiled form) incubated with DRIP IP sample 7 (37 °C)
6. 100 bp marker
7. pmCherry-N1 plasmid (supercoiled form) in 300 mM NaCl/10 mM Tris-Cl pH 7.5
8. pmCherry-N1 plasmid (supercoiled form) + 2 µl RNase A (10 mg / ml, UDG) in 300 mM NaCl/10 mM Tris-Cl pH 7.5
9. pmCherry-N1 plasmid (supercoiled form) + 2 µl RNase A (10 mg / ml, NEB) in 300 mM NaCl/10 mM Tris-Cl pH 7.5
10. pmCherry-N1 plasmid (nicked circular form) + 2 µl RNase A (10 mg / ml, UDG) in 300 mM NaCl/10 mM Tris-Cl pH 7.5. Nicking was achieved by 30 min UV treatment.
11. pmCherry-N1 plasmid (linear) in 300 mM NaCl/10 mM Tris-Cl pH 7.5. Linearization was achieved by BamHI digestion of the plasmid. The digested plasmid was PCR clean up purified.
12. pmCherry-N1 plasmid (linear) + 2 µl RNase A (10 mg / ml, UDG) in 300 mM NaCl/10 mM Tris-Cl pH 7.5.
13. 1 kb marker