Supplement: Data exploration, quality control and statistical analysis of ChIP-exo experiments

Rene Welch[1], Dongjun Chung[6], Irene Ong[3], Jeffrey Grass[3,4], Robert Landick[3,4,5], and Sündüz Keleş[1,2]

[1]Department of Statistics, University of Wisconsin Madison
[2]Department of Biostatistics and Medical Informatics, University of Wisconsin Madison
[3]Great Lakes Bioenergy Research Center, University of Wisconsin Madison
[4]Department of Biochemistry, University of Wisconsin Madison
[5]Department of Bacteriology, University of Wisconsin Madison
[6]Department of Public Health Sciences, Medical University of South Carolina

Contents

1 Comparison of ChIP-exo with ChIP-Seq data.

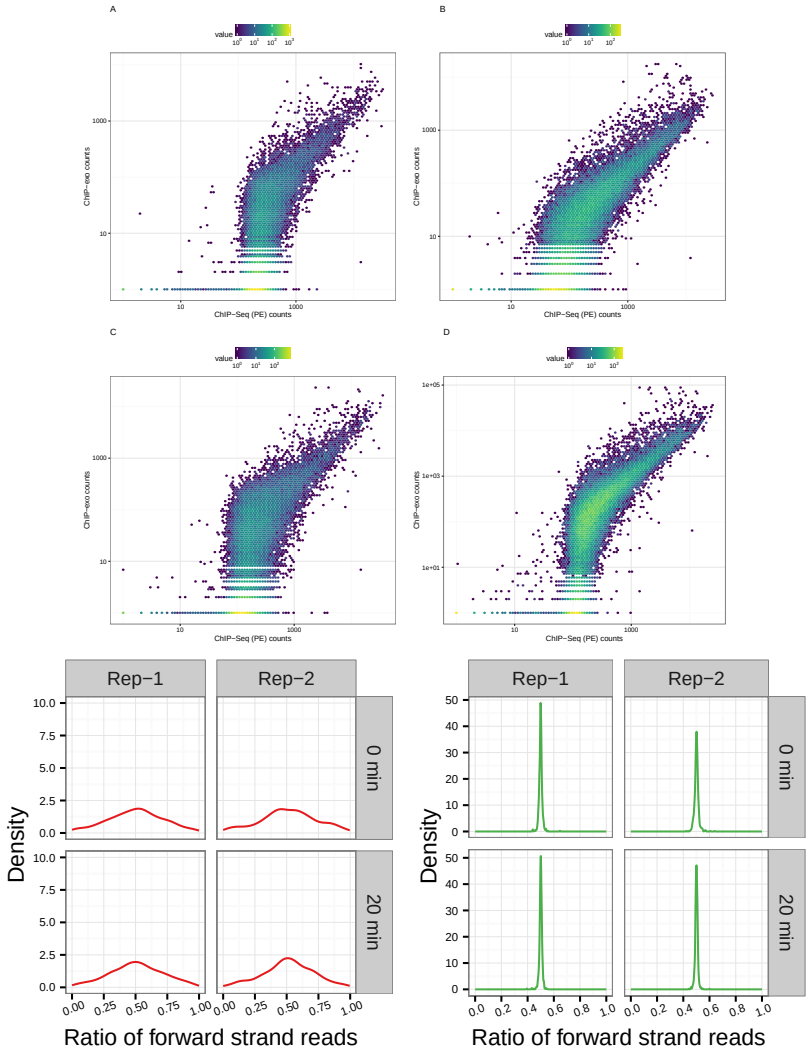


Figure S1: Hexbin plots of PE ChIP-Seq bin counts vs. ChIP-exo bin counts for σ^{70} second biosample: A) E2-1, B) E2-2 C) E2-3 and D) E2-4 samples. E) Forward Strand Ratio densities for SE ChIP-Seq and ChIP-exo peaks for σ^{70}, S2 and E2 groups respectively.

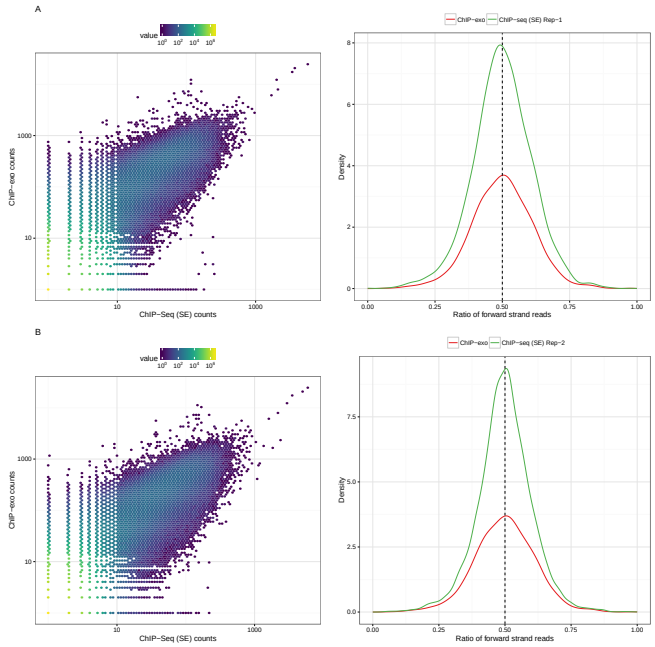


Figure S2: Hexbin plot comparing A) Rep-1 and B) Rep-2 ENCODE SE ChIP-seq bin counts vs. Pugh ChIP-exo bin counts for CTCF in HeLa cell. C) Rep-1 and D) Rep-2 ChIP-seq compared against ChIP-exo Forward Strand Ratio densities.

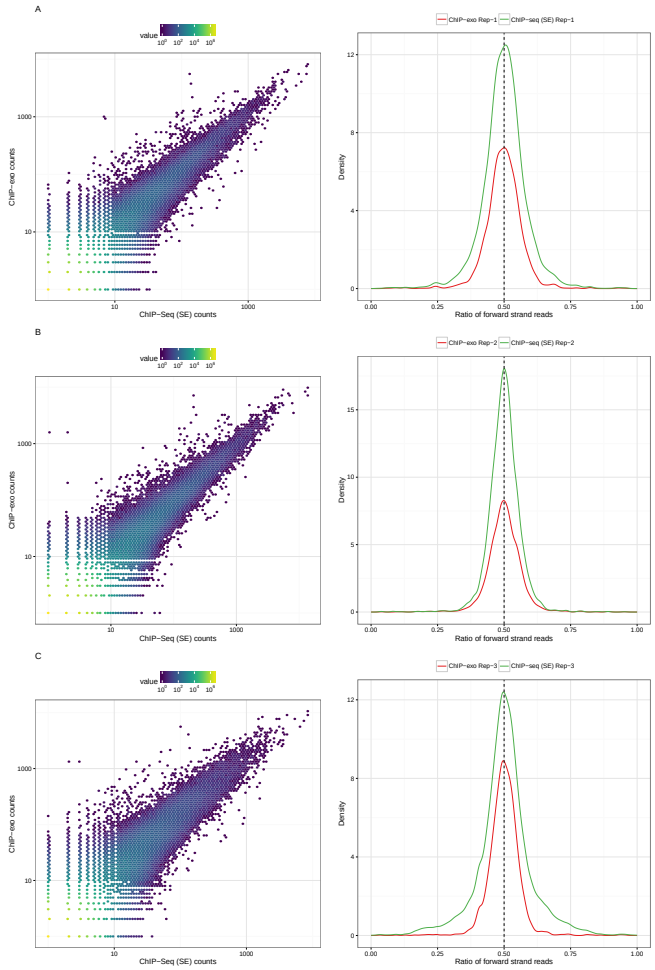


Figure S3: Hexbin plots comparing SE ChIP-Seq bin counts vs. ChIP-exo bin counts for ER in human MCF-7 cell lines.

2 SCC curves for ChIP-exo and ChIP-nexus data.

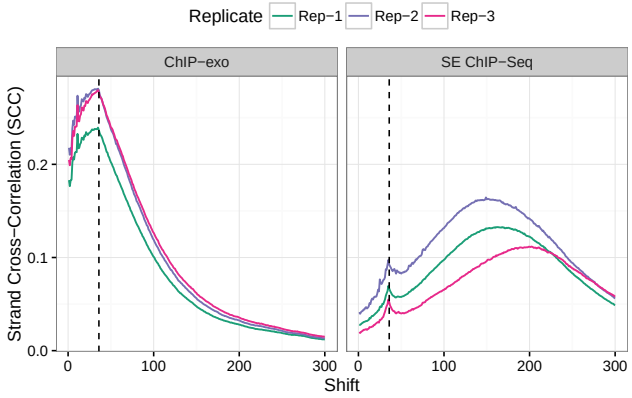


Figure S4: Comparison of ChIP-exo vs. ChIP-Seq SCC curves for ER factor in human MCF-7 cell lines.

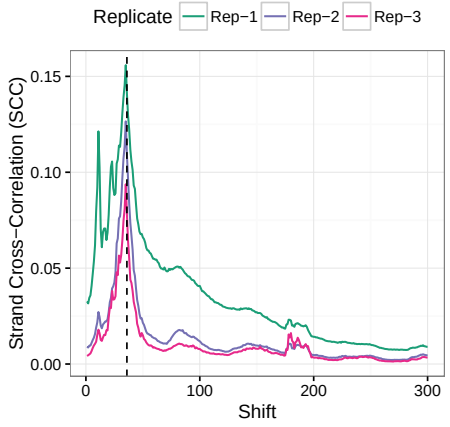


Figure S5: SCC curves for FoxA1 factor in mouse liver cell lines.

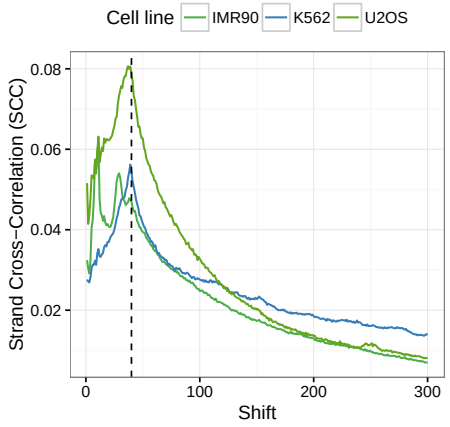


Figure S6: SCC curves for GR factor in IMR90, K562 and U2OS human cell lines.

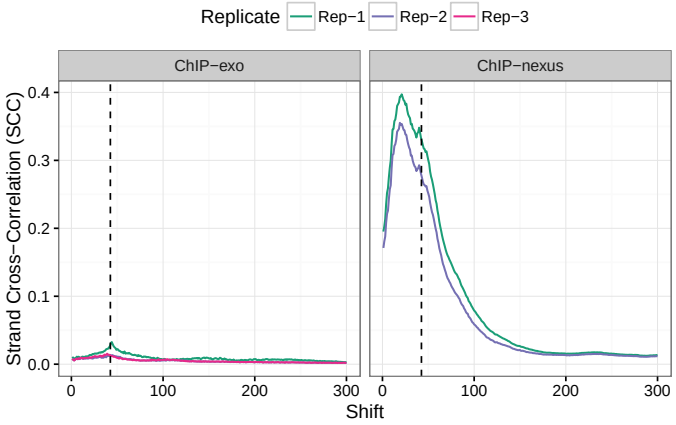


Figure S7: Comparison of ChIP-exo vs. ChIP-nexus SCC curves for TBP factor in human K562 cell lines.

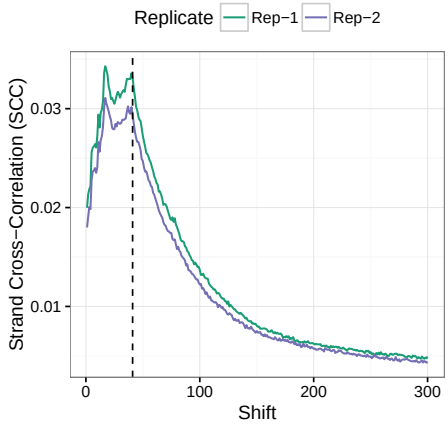


Figure S8: SCC curves for dorsal factor in embryo *D. Melanogaster* cell lines.

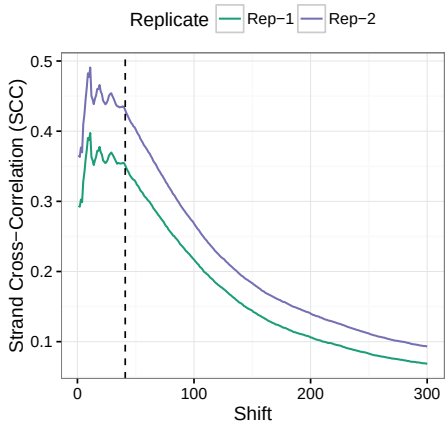


Figure S9: SCC curves for twist factor in embryo *D. Melanogaster* cell lines.

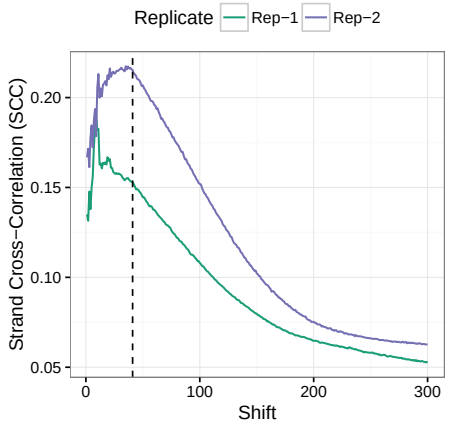


Figure S10: SCC curves for Max factor in S2 *D. Melanogaster* cell lines.

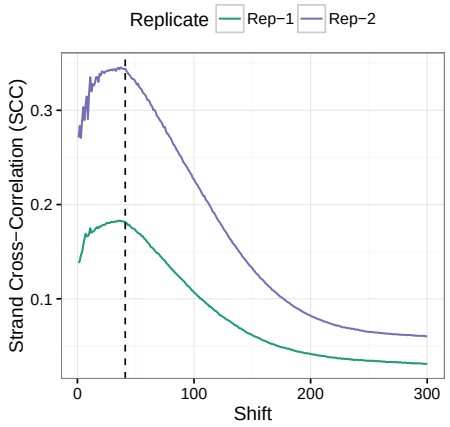


Figure S11: SCC curves for MyC factor in S2 *D. Melanogaster* cell lines.

RW: need to generate this curves for sig70 samples, compared against PE and SE ChIP-seq

3 QC pipeline applied to ChIP-exo and ChIP-nexus data.

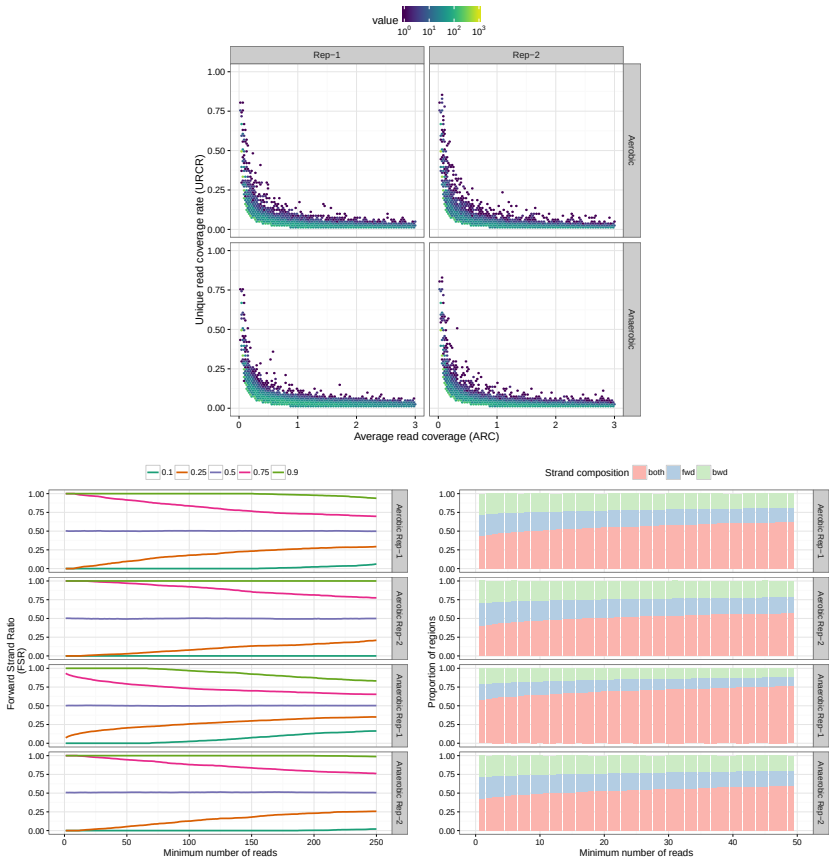


Figure S12: ChIP-exo QC pipeline diagnostics for σ^{70} E1 biosample in *E. Coli*.

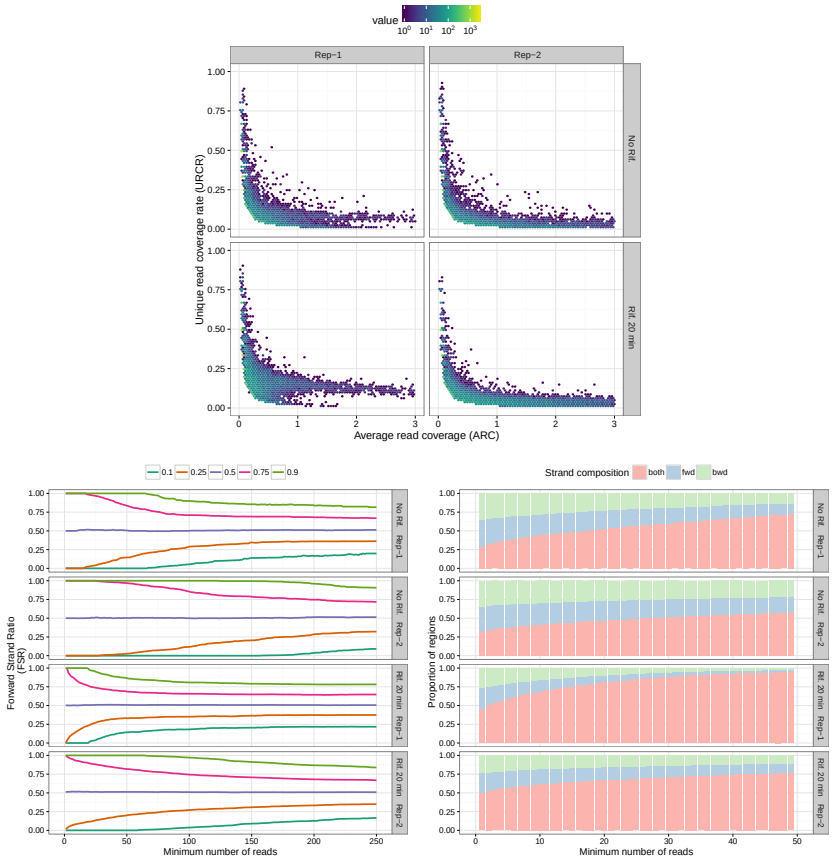


Figure S13: ChIP-exo QC pipeline diagnostics for σ^{70} E2 biosample in *E. Coli*.

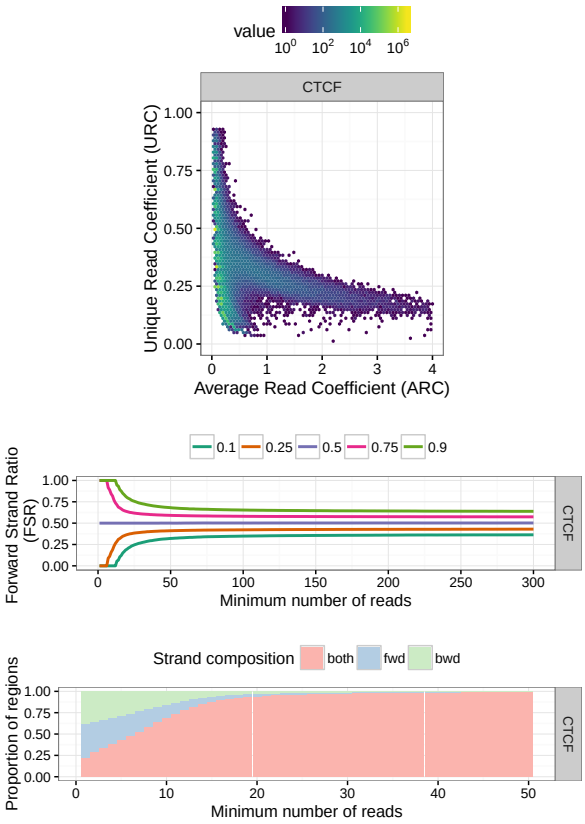


Figure S14: ChIP-exo QC pipeline diagnostics for CTCF factor in human HeLa cell lines.

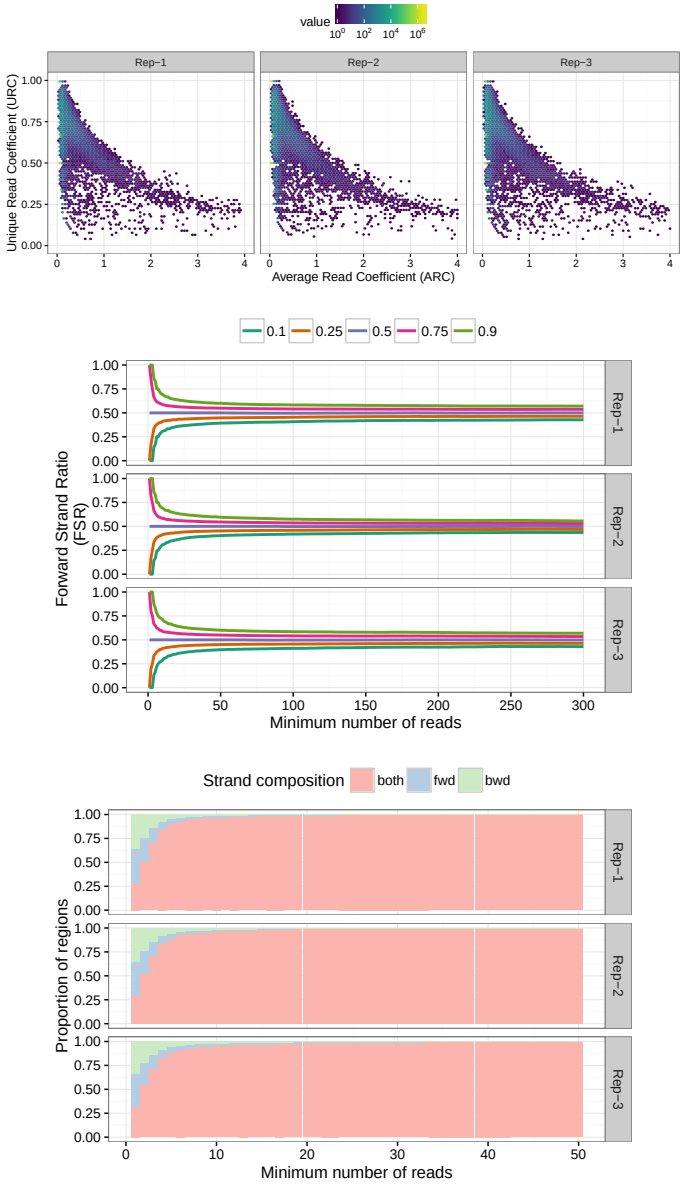


Figure S15: ChIP-exo QC pipeline diagnostics for ER factor in human MCF-7 cell lines.

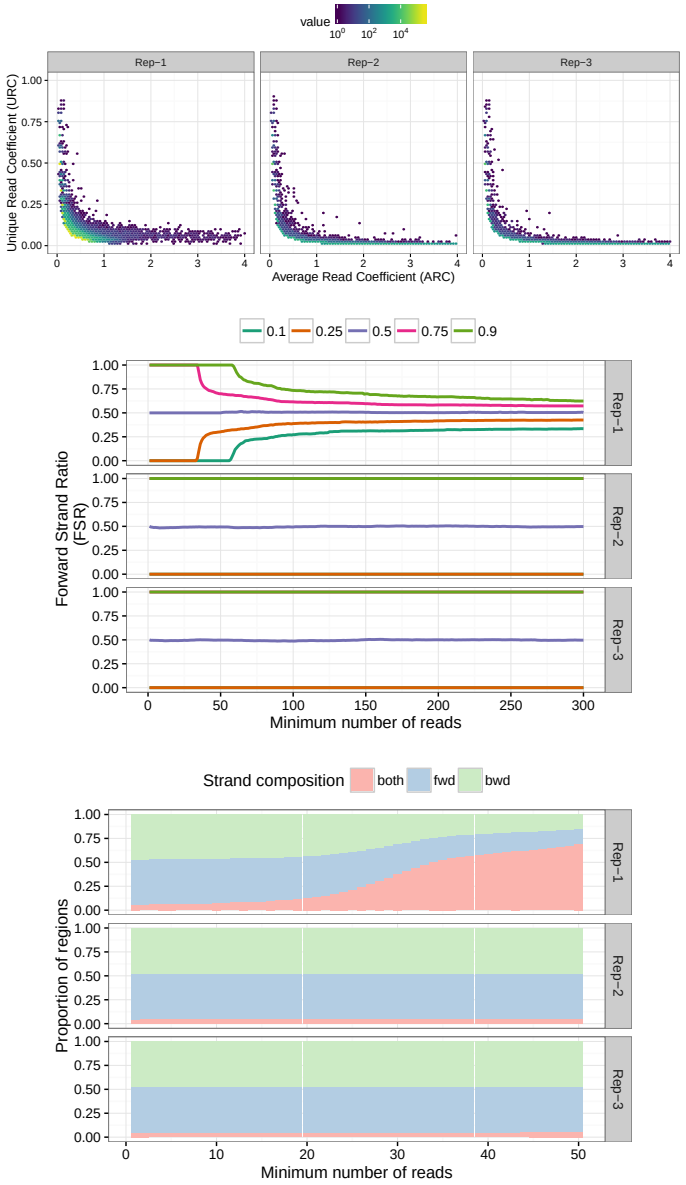


Figure S16: ChIP-exo QC pipeline diagnostics for TBP factor in human K562 cell lines.

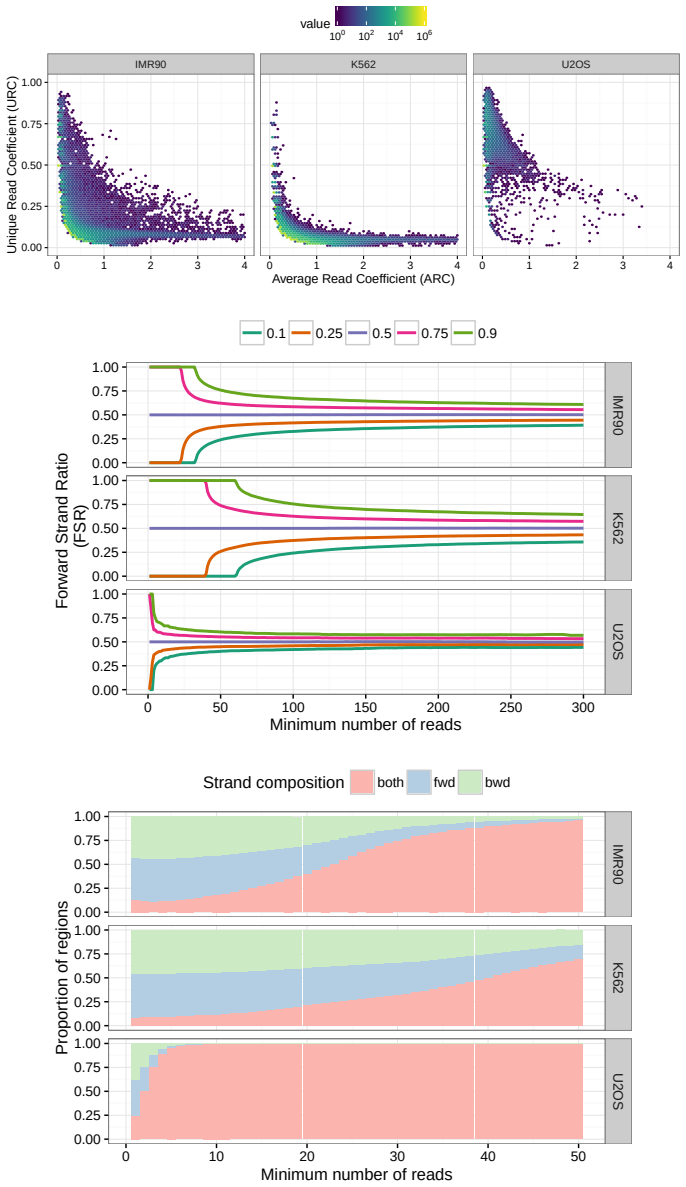


Figure S17: ChIP-exo QC pipeline diagnostics for GR factor in human IMR90, K562 and U2OS cell lines respectively.

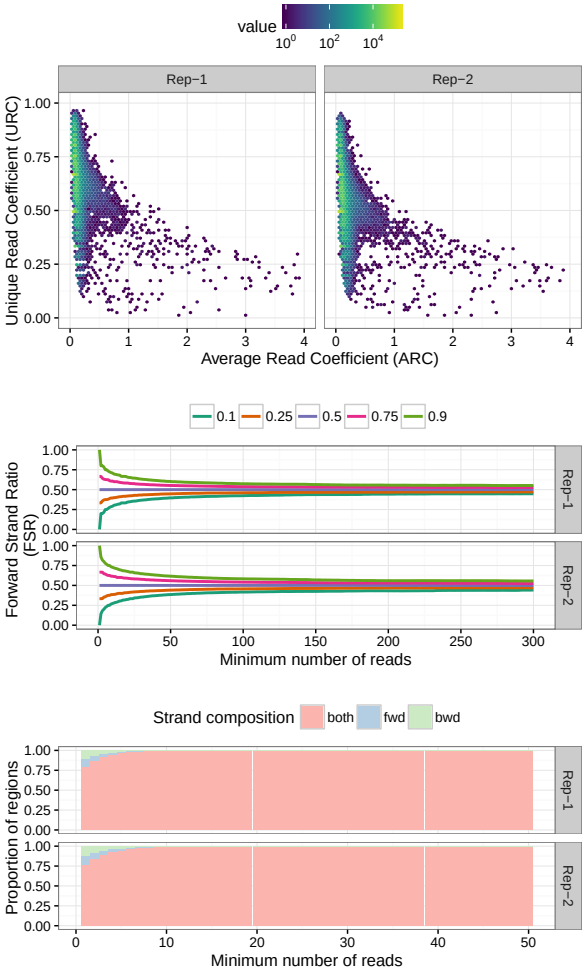


Figure S18: ChIP-exo QC pipeline diagnostics for ChIP-nexus experiment of dorsal factor in embryo *D. Melanogaster* cell lines.

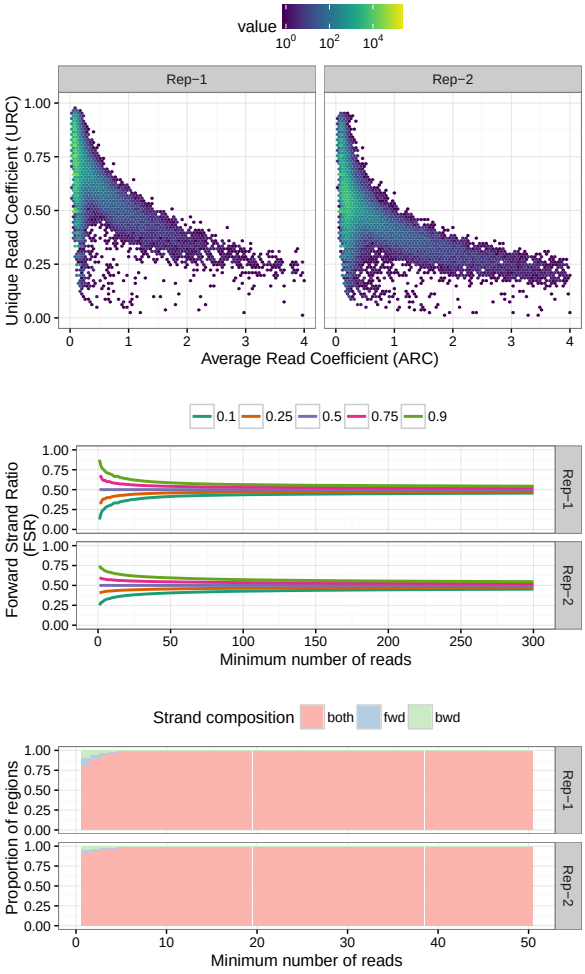


Figure S19: ChIP-exo QC pipeline diagnostics for ChIP-nexus experiment of twist factor in embryo *D. Melanogaster* cell lines.

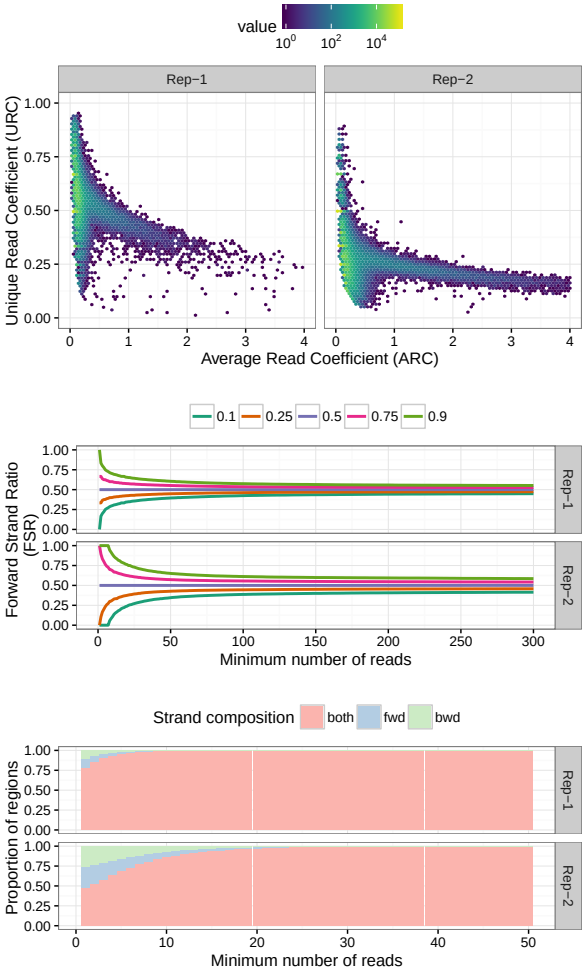


Figure S20: ChIP-exo QC pipeline diagnostics for ChIP-nexus experiment of Max factor in S2 *D. Melanogaster* cell lines.

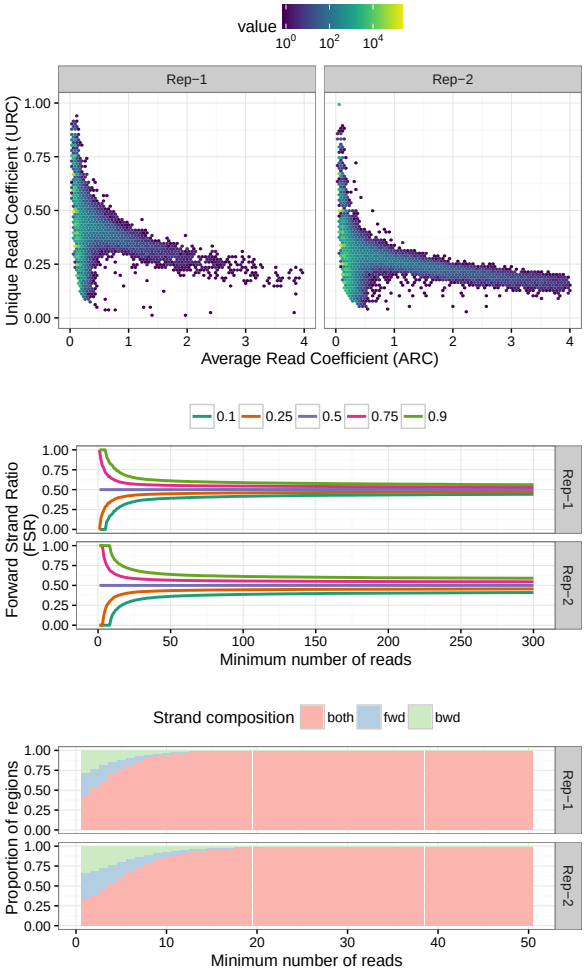


Figure S21: ChIP-exo QC pipeline diagnostics for ChIP-nexus experiment of MyC factor in S2 *D. Melanogaster* cell lines.

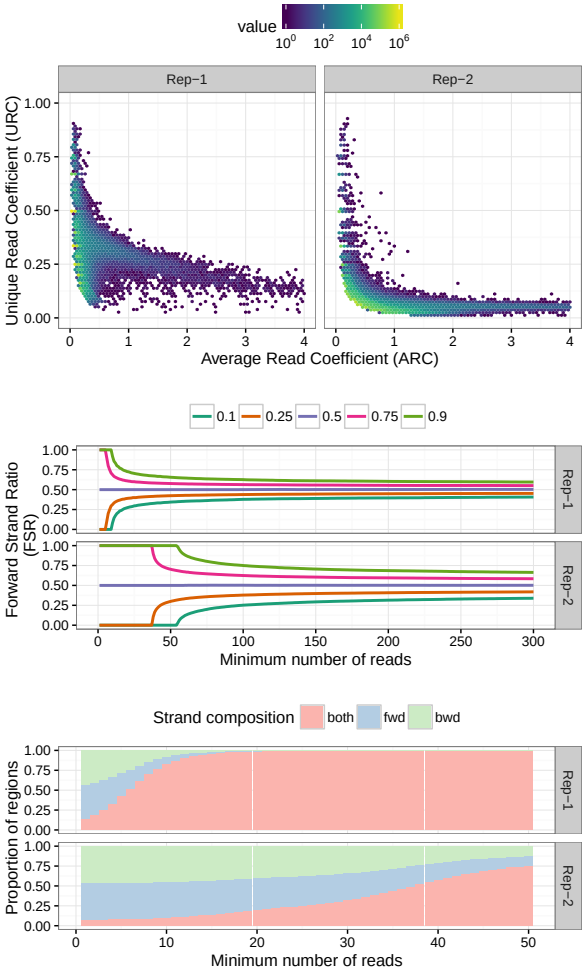
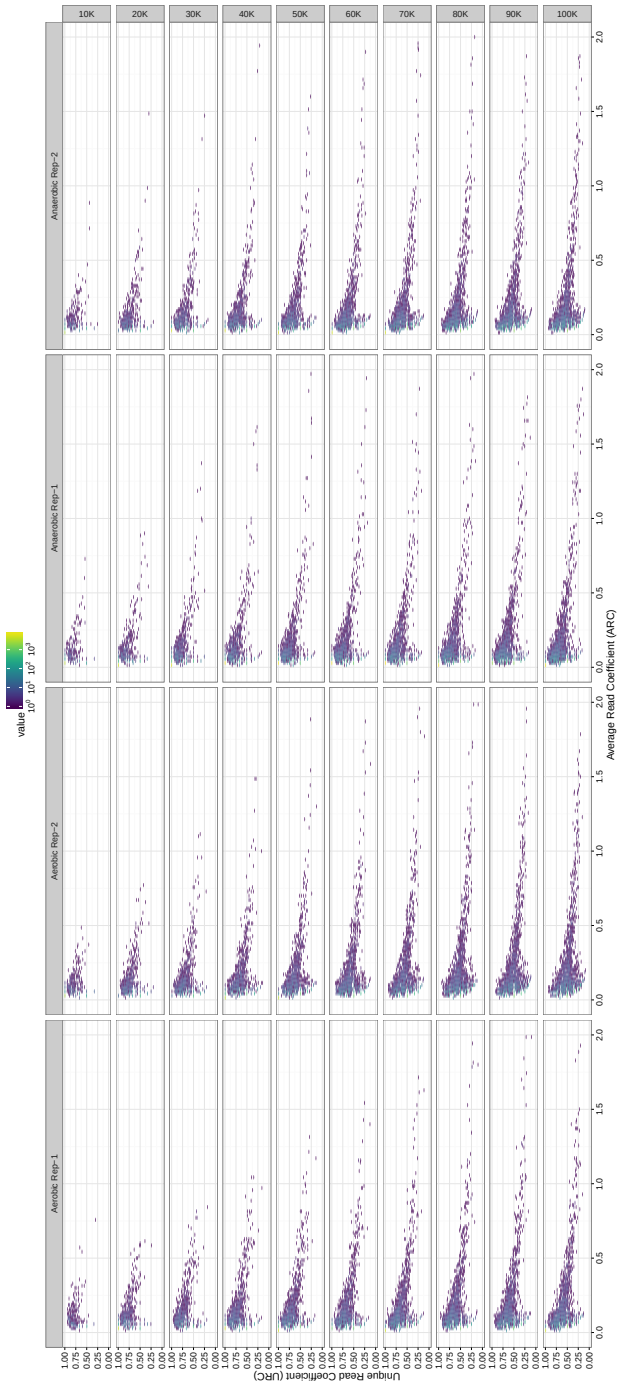


Figure S22: ChIP-exo QC pipeline diagnostics for ChIP-nexus experiment of TBP factor in K562 human cell lines.

Comparison of QC pipeline for different sub-sampled sequencing depths.



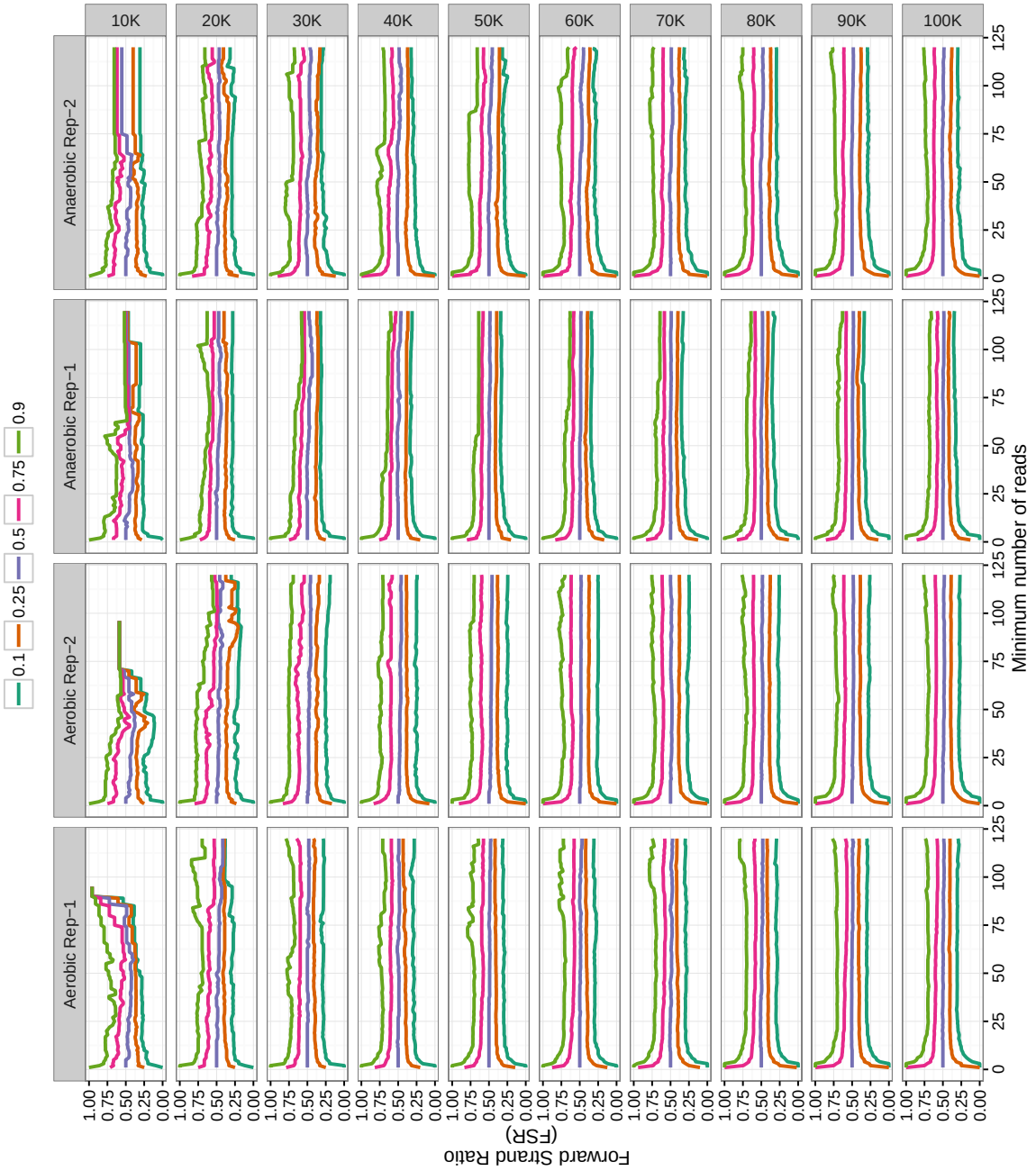
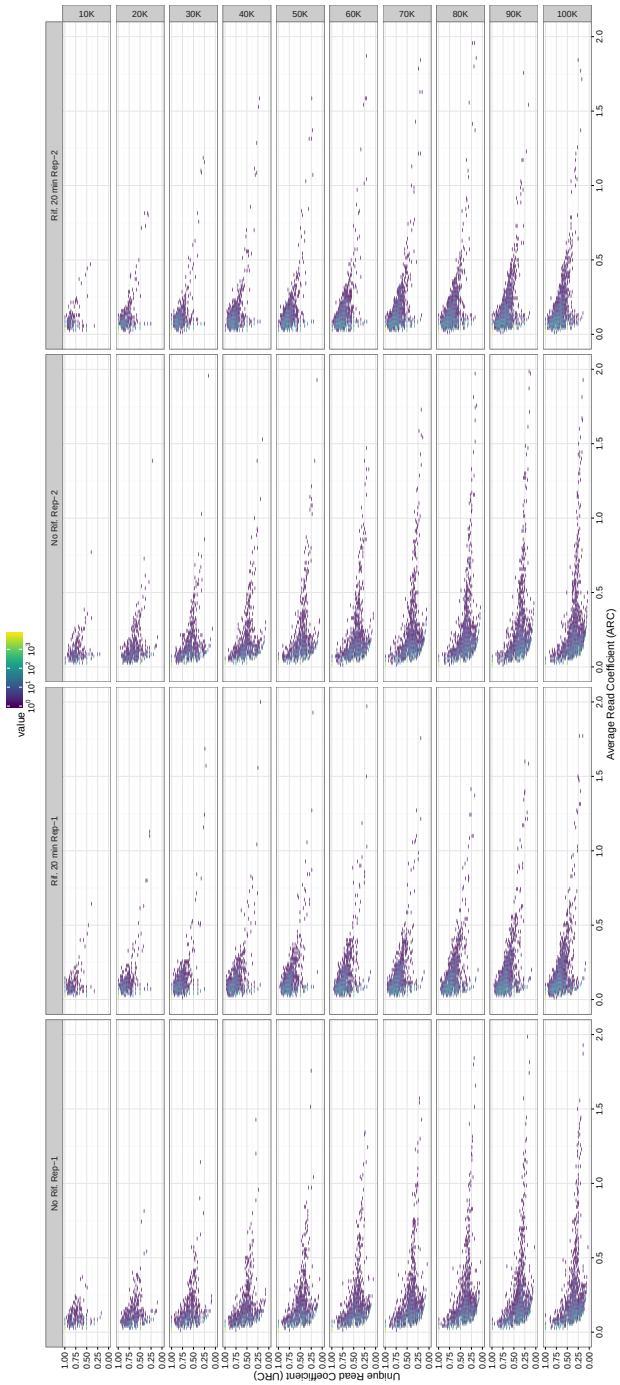
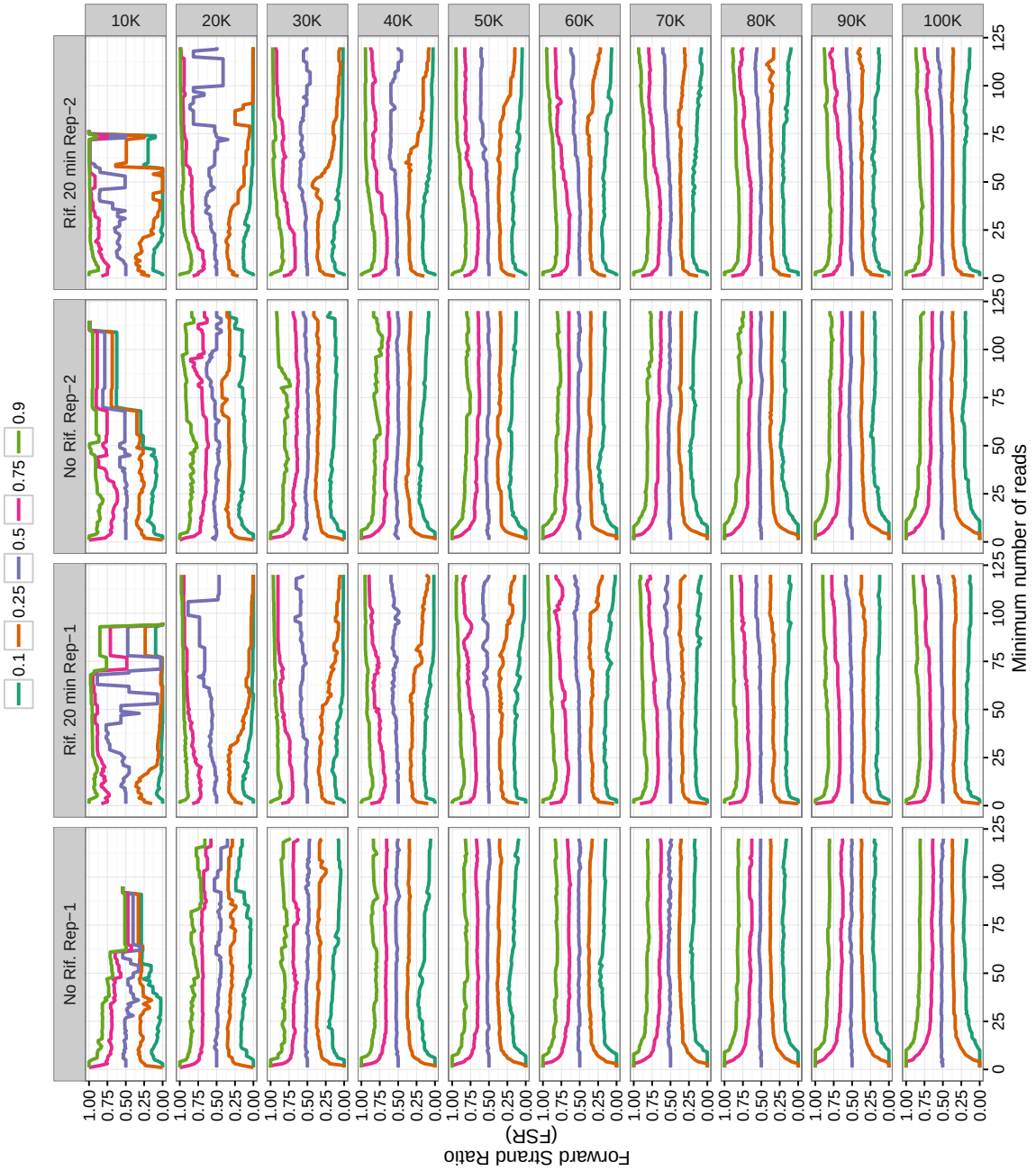

0.1
0.25
0.5
0.75
0.9
Aerobic Rep−1
Aerobic Rep−2
Anaerobic Rep−1
Anaerobic Rep−2
10K
20K
30K
40K
50K
60K
70K
80K
90K
100K
Forward Strand Ratio (FSR)
Minimum number of reads

Figure S23: ChIP-exo QC pipeline diagnostics for σ^{70} E1 biosample on *E. Coli* when sampling 10K to 100K reads from each experiment


10K
20K
30K
40K
50K
60K
70K
80K
90K
100K
No Rif. Rep−1
Rif. 20 mn Rep−1
No Rif. Rep−2
Rif. 20 mn Rep−2
Unique Read Coefficient (URC)
Average Read Coefficient (ARC)
value
10^0 10^1 10^2 10^3


0.1
0.25
0.5
0.75
0.9
No Rif. Rep−1
Rif. 20 min Rep−1
No Rif. Rep−2
Rif. 20 min Rep−2
10K
20K
30K
40K
50K
60K
70K
80K
90K
100K
Forward Strand Ratio (FSR)
Minimum number of reads

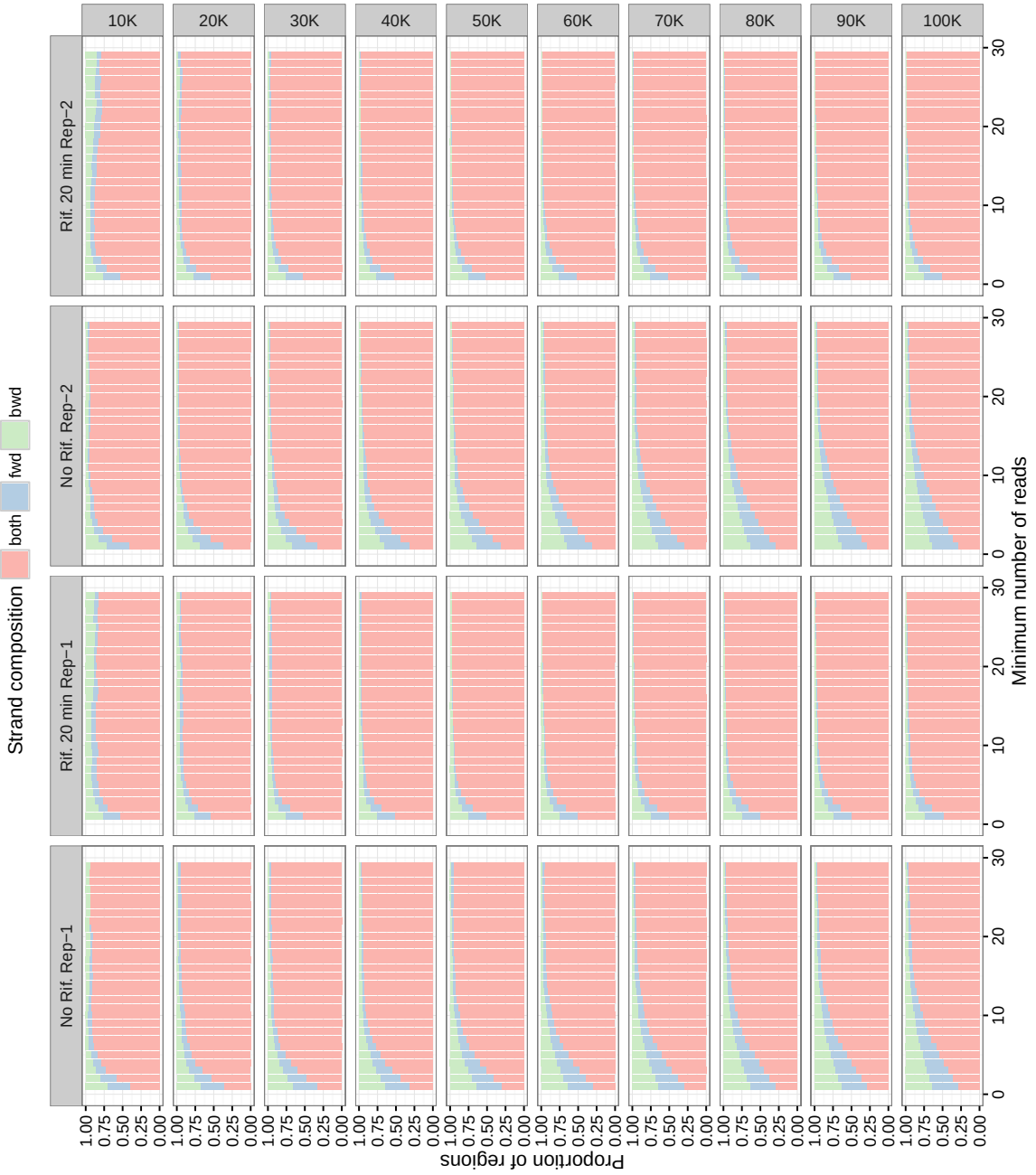


Figure S24: ChIP-exo QC pipeline diagnostics for σ^{70} E2 biosample on *E. Coli* when sampling 10K to 100K reads from each experiment

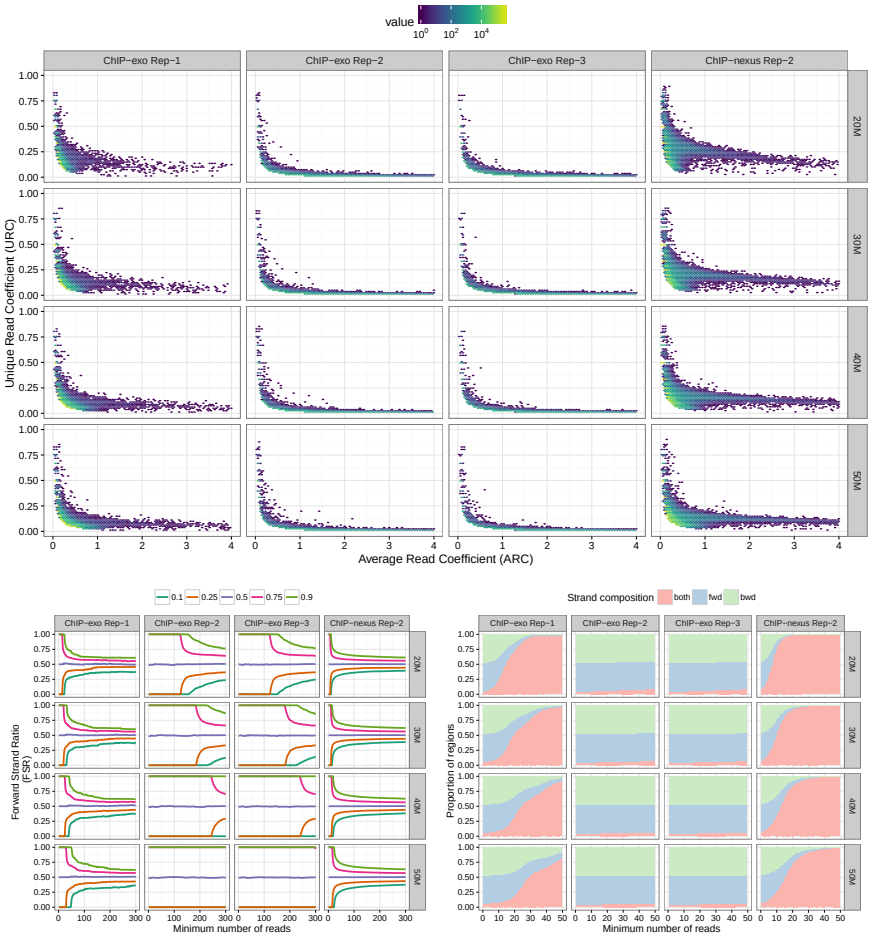


Figure S25: ChIP-exo QC pipeline diagnostics applied to ChIP-nexus and ChIP-exo experiments of TBP factor in K562 human cell lines when sampling 20M to 50M reads from each experiment.

3.1 FoxA1 peak overlaps with high quality regions.

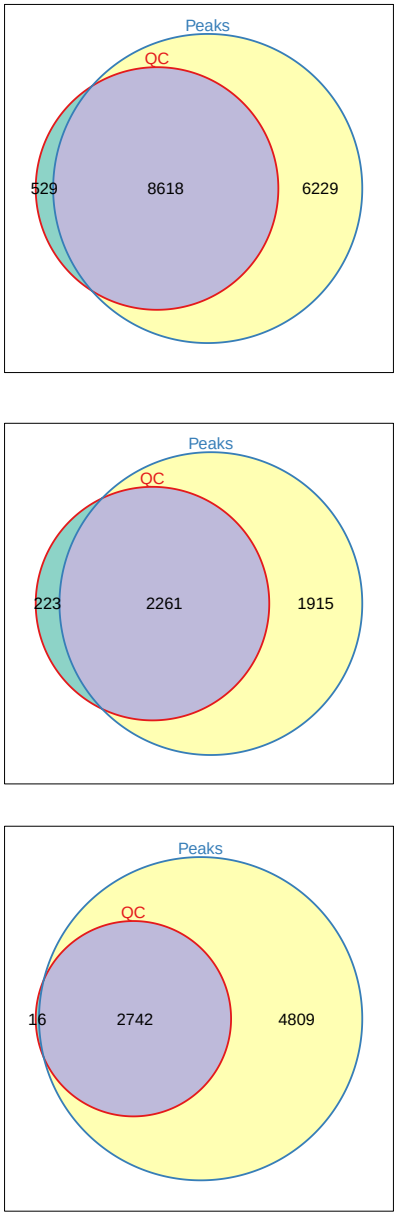


Figure S26: Venn diagrams of high quality regions that overlap peaks for FoxA1 factor in mouse liver cell lines. Top) Rep-1, Middle) Rep-2 and Bottom) Rep-3.