File S1: Description of supplemental material. This document contains detailed descriptions for all the supplemental files and tables for:

Effects of sheared chromatin length on ChIP-seq quality and sensitivity

Cheryl A. Keller, Alexander Q. Wixom, Elisabeth F. Heuston, Belinda Giardine, Chris C.-S. Hsiung, Maria R. Long, Amber Miller, Stacie M. Anderson, April Cockburn, Gerd A. Blobel, David M. Bodine, Ross C. Hardison

File S2: Code. This file contains a combination of custom bash scripts and commands to run FIMO, bedtools, and deepTools used to extract, quantitate, and visualize peak data with and without CTCF or TAL1 motifs.

Table S1: Datasets. This table contains a complete listing of datasets and metadata for all ChIP-seq samples, and includes the following information:

Column A: Dataset category – classification of the dataset (main, retrospective, low input, or hematopoietic progenitor)
Column B: Figures – list of figures in which the dataset appears
Column C: Target – ChIP-seq target (CTCF, TAL1, or POL2)
Column D: Cell type – name of cell type used for ChIP-seq experiment
Column E: Library ID – unique dataset identifier
Column F: Cell number – number of cells used for ChIP-seq experiment
Column G: Treatment – cell treatment prior to ChIP
Column H: Fixative – fixative used to cross-link the cells before sonication
Column I: Cycles – number of sonication cycles
Column J: Average size (bp) - mean size of unenriched chromatin as measured between 100-500bp using the Agilent Bioanalyzer 7500 DNA chip
Column K: Platform – Illumina sequencing platform
Column L: Assembly – genome assembly to which the dataset was mapped
Column M: GEO – relevant GEO accession numbers
Column N: Mapped Reads – numbers of mapped reads
Column O: Peaks – number of peaks called by MACS
Column P: Percent GC – percentage of CG content
Column Q: Duplication rate – percentage of duplicated reads
Column R: Complexity – fraction of non-redundant reads
Column S: Percent mapped – percentage of reads mapped to genome
Column T: NSC - normalized strand coefficient quality metric score
Column U: RSC - reverse strand coefficient quality metric score
Column V: QTag - Quality tag score based on thresholded RSC
Column W: FRiP – Fraction of Reads in Peaks score
Column X: FRiP – Fraction of Reads in Peaks (as percentage)
Column Y: Subjective quality assessment (pass, low pass, or fail)

Table S2: Peaks and motifs statistics. This table contains statistics related to numbers of peaks and motifs.

Column A: Dataset category – classification of the dataset (main)
Column B: Figures – list of figures in which the dataset appears
Column C: Target – ChIP-seq target (CTCF or TAL1)
Column D: Cell number – number of cells used for ChIP-seq experiment
Column E: Library ID – unique dataset identifier
Column F: Cycles – number of sonication cycles
Column G: Peaks – number of peaks called by MACS
Column H: Peaks with motif ($p<0.0001$) – number of peaks with a significant motif ($p<0.0001$)
Column I: Peaks without motif – number of peaks without a motif
Column J: Percentage of peaks with motif ($p<0.0001$) – percentage of peaks with a significant motif ($p<0.0001$)
Column K: Overlapped-hc peaks – number of peaks that overlapped with high confidence peaks
Column L: Percentage of peaks that overlapped with hc peaks – percentage of peaks that overlapped with high confidence peaks
Column M: Overlapped-hc peaks motif ($p<0.0001$) – number of overlapped-hc peaks with a significant motif ($p<0.0001$)
Column N: Overlapped-hc peaks without motif – number of overlapped-hc peaks without a motif ($p<0.0001$)
Column O: Percentage of overlapped-hc peaks with motif ($p<0.0001$) – percentage of overlapped-hc peaks with a significant motif ($p<0.0001$)

Table S3: Input datasets. This table contains a complete listing of datasets and metadata for all input samples, and includes the following information:

Column A: Dataset category – classification of the dataset (input)
Column B: Cell type – name of cell type used for ChIP-seq experiment
Column C: Library ID – unique dataset identifier
Column D: Treatment – cell treatment
Column E: Fixative – fixative used to cross-link the cells before sonication
Column C: Unique library ID
Column D: Treatment
Column E: Fixative used to cross-link the cells before sonication
Column F: Platform – Illumina sequencing platform
Column G: Assembly – genome assembly to which the dataset was mapped
Column H: GEO – relevant GEO accession numbers
Column I: Mapped Reads – numbers of mapped reads
Column J: Percent GC – percentage of CG content
Column K: Duplication rate – percentage of duplicated reads
Column L: Percent mapped – percentage of reads mapped to genome

Column M: Used for Library IDs - list of unique library IDs for which the input data was used.
Column N: Comments not covered by above categories

Table S4: Hematopoietic progenitors. This table contains information related to the isolation of mouse eight primary hematopoietic progenitor cells from mouse bone marrow.
Column A: Cell type – name of cell type used for ChIP-seq experiment
Column B: Library ID – unique dataset identifier
Column C: Isolation/sort markers – list of markers used for isolation and sorting of cell types via FACS
Columns D-AG: Isolation dates and cell numbers (in millions of cells) for each cell type, as indicated by row
Column AH: Totals – total cell numbers (in millions) isolated for each cell type
Column AI: Comments not covered by above categories