SUPPLEMENTARY MATERIALS AND METHODS

Chromatin immunoprecipitations.

1 ml of chromatin was incubated with 5 µl of anti-POLR3A, 5 µl of anti-POLR3D, 50 µl of anti-POLR2B, 10 µl of anti-H3K36me3, or 1.5 µL of anti-H3K4me3 in buffer A (20 mM Tris/HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA) overnight at 4°C on a rotating wheel. Ten µl of protein A bead suspension (25% slurry in buffer A), preblocked with 10 µg/ml of salmon sperm DNA and BSA at 4˚C overnight, was then added and incubation was continued for 1 hour at room temperature on a rotating wheel. The beads were then washed with dialysis buffer and ChIP wash buffer as described (O'Geen et al. 2006). Protein-DNA complexes were eluted from the beads, de-crosslinked, treated with RNase A and proteinase K as described (O'Geen et al. 2006). DNA concentration was determined by fluorometry on the Qubit system (Invitrogen). 10-12 ng DNA were used for library preparation. Libraries for ultra high throughput sequencing were prepared with the ChIP-Seq DNA sample kit (Illumina) as recommended by the manufacturer.

Analysis of ChIP-Seq data

Tag alignment. RNAP-III transcription units are rich in repeated sequences resulting from retrotransposition events that gave rise to numerous pseudogenes as well as from gene duplications. To maximize our ability to attribute sequence tags to a unique genomic position, we obtained 76 nucleotide-long reads rather than the standard 35 nucleotide reads, and we only aligned tags with perfect genomic matches. To avoid losing the information contained in tags with several possible matches in the genome for identification of RNAP-III-occupied loci, we aligned tags onto the mouse genome assembly (mcb137/mm9) with the "fetchGWI" software (www.isrec.isb-sib.ch/tagger/) (Iseli et al. 2007), which allowed us to align tags with multiple matches in the genome. We included repeated tags mapping up to 500 genomic locations. As in our previous work, we did not allow any mismatch for tag alignment (Canella et al. 2010). Table S8 shows the number of tags aligned with single matches on the mouse genome, 2-10 matches, 11 to 100, and 101 to 500. About 94.7% of the tags mapped onto a single genomic location.

Peak detection. In the sissrs peak detection, we obtained peaks on 241 tRNA genes, all *Rn5s* loci in Table S2, all genes in Table S3 except the two *Rn7sl* genes, which are not annotated (and were found in the list of loci not annotated as RNAP-III loci, see below). In addition to the *Rn7sk* sequence in Table S3, we found three additional loci annotated *Rn7sk*: they had 0%, 1%, and 15.4% unique tags, and visual inspection revealed all the tags to be within the *Rn7sk* coding sequence, with no tags in the immediate 5' and 3' flanking sequence as would be expected of true RNAP-III-transcribed loci (see (Canella et al. 2010); these three loci were, therefore, eliminated. In addition to loci annotated as RNAP-III genes, we obtained a total of 835 reproducible loci not annotated as RNAP-III genes. Of these 835 loci, we eliminated peaks containing less than 50% unique tags, which left 598 (72%) loci. The 598 loci were each submitted to visual inspection: peaks with shapes never found in known RNAP-III genes (for example peaks with

rectangular shapes resulting from artefactual accumulation of tags), or with identical shapes in all samples (rep1 and rep2 for POLR3A and POLR3D ChIPs), or in which no offset was observed between the peaks obtained with anti-POLR3A and anti-POLR3D antibodies, were eliminated. The loci left, shown in Tables S4 and S5, were the most convincing cases of loci occupied by RNAP-III. It is possible, however, that some of the peaks eliminated by visual inspection represent in fact loci occupied by RNAP-III. Further, our method may have eliminated loci occupied by very low levels of RNAP-III, as those rarely present convincing peaks. Finally, the peaks constituted by more than 50% repeated tags (and not corresponding to known RNAP-III genes) were not retained.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Pearson correlations of log2 scores obtained with the anti-POLR3A and anti-POLR3D antibodies for the rep 1 and rep2 biological replicates, as indicated. All loci listed in Tables S1-S5 are included. The regression line is indicated in red, the y=x line in blue. The Pearson correlation coefficients are indicated in the squares in the upper right part of the figure.

Figure S2. Alignment of the fifty *Rn5*s genes on chromosome 8. The sequences were aligned with ClustalW2 set on the default parameters. The gene number on the left indicates the ordinal position of the gene within the cluster. Boxes D, A, and C as well as the intermediate element (IE), as defined in (Nielsen et al. 1993) and (Pieler et al. 1987), are highlighted on top in purple, the *Rn5*s coding sequence (which contains the A, IE, and C boxes) is highlighted in yellow. The eight genes highlighted in blue on the left have the lowest scores (see Table S2): the correlating mutations at positions +1 and +2, or +2, or +3 relative to the TSS are highlighted in blue in the sequence, those in box C are highlighted in green. All other mutations are highlighted in grey.

Figure S3. Alignment of type 3 promoter elements for the known type 3 RNAP-III genes. The numbers on the very left refer to row numbers in Table S3 (1 to 11) and Table S1 (153). The coordinates refer to the first and last nucleotide of the sequence shown relative to the TSS (+1).

Figure S4. Alignment of *Rn4.5s* sequences. The sequences were aligned with Clustal W2 set on the default parameters. The first number on the left in bold indicates the row number of the corresponding entry in Table S4. This number is followed by the chromosome 6 coordinates of the sequence used in the alignment. Only part of the alignment is shown, and the numbers on the right indicate the position relative to the fragments defined by the coordinates on the left. Sequences numbered 1, 2, and 3 on the left correspond to 4.5S genes 1', 1, and 2 in (Gogolevskaya and Kramerov 2010). The other sequences correspond to sequences annotated as 4.5S in the UCSC RNA table. The RNA coding sequence as defined by (Gogolevskaya and Kramerov 2010) for sequences 1, 2, and 3 is indicated in yellow, sequences with similarities to A and B boxes in

purple. Putative corresponding RNA coding sequence and A and B boxes are similarly indicated for sequences 4-9. The underlined sequences in sequences 1, 2, and 3 correspond to the conserved GC$^{C}/_{A}$ACGCCT and AGAAT elements.

Figure S5. Results of univariate regressions. The best pattern, highest score pattern, and most common nucleotides (nt) are indicated for the A and B boxes. The significance threshold was set at 1%. Positions for which no base was significant in the regression (in general because the bases at these positions did not vary between low and high scoring tRNA genes) are labeled with an asterisk. Nucleotides that correlated with increased score are indicated in upper case, and nucleotides that correlated with decreased score are indicated in lower case in parenthesis.

Figure S6. The X and Y axis show RNAP-II and RNAP-III scores (log2 of value+1). The numbers along the axis indicate the number of points counted in vertical respectively horizontal slices of 1 unit starting at 0. The colors of the dots indicate amounts of H3K3me3 tags at the gene, as indicate in the lower right corner.

Figure S7. Distribution of scores on 50.000 random regions of the mappable mm9 mouse genome. The X axis indicate scores, the Y axis indicate the number of random regions for a given score (frequency).

Figure S8. Distribution of sequence tags obtained from input (panel A) or material immunoprecipitated with the anti-POLR3D antibody (panel B) in the loci listed in Tables S1 to S5.

Figure S9. Panels D and G of Figure 3 were repeated with scores calculated as log ((y+c)/i+c)), where y=ChIP score, i=input score, and c=pseudocounts set at 5. The results are similar to those in Figure 3.

SUPPLEMENTARY REFERENCES

Canella D, Praz V, Reina JH, Cousin P, Hernandez N. 2010. Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Res* **20**(6): 710-721.

Gogolevskaya IK, Kramerov DA. 2010. 4.5SI RNA genes and the role of their 5'-flanking sequences in the gene transcription. *Gene* **451**(1-2): 32-37.

Iseli C, Ambrosini G, Bucher P, Jongeneel CV. 2007. Indexing strategies for rapid searches of short words in genome sequences. *PLoS One* **2**(6): e579.

Nielsen JN, Hallenberg C, Frederiksen S, Sorensen PD, Lomholt B. 1993. Transcription of human 5S rRNA genes is influenced by an upstream DNA sequence. *Nucleic Acids Res* **21**(16): 3631-3636.

O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ. 2006. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques* **41**(5): 577-580.

Pieler T, Hamm J, Roeder RG. 1987. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. *Cell* **48**(1): 91-100.