Supplemental Material

Material and Methods.

ChIP.

Cells were crosslinked with formaldehyde 0.4% for 10 min at room temperature and the reaction was stopped by adding glycine to final concentration 0.2M for 10 minutes at room temperature. The ES cells were passed each day for 3 days in absence of feeders before treatment with RA and then formaldehyde. Fixed cells were rinsed twice with PBS and resuspended in 100 ul of lysis buffer (10 mM EDTA pH 8, 50 mM Tris-HCl pH 8, SDS 1 %). Lysate was sonicated 5 min (30 sec on / 30 sec off) in Diagenode water bath-sonicator and centrifuged at 14000 rpm for 10 min. The cleared supernatant was used immediately in ChIP experiments or stored at -80°C.

50-150 ug of sonicated chromatin was diluted 10 times in ChIP Dilution Buffer (SDS 0.01%, Triton X-100 1.1%, 1.2 mM EDTA pH 8, 16.7 mM Tris-HCl pH 8 and 167 mM NaCl) and pre-cleared for 1 hour, rotating at 4°C, with 50 ul blocked beads (Protein G Sepharose (PGS) 50% slurry incubated for 4 hours minimum with BSA 0.5 mg/mL and yeast tRNA 0.2 mg/mL) before the overnight incubation with 2-5ug of specific antibody. The bound material was recovered after a 2 hours incubation, rotating at 4°C, with 30 ul blocked PGS. The beads were washed, for 5 minutes, once in Low Salt Buffer (SDS 0.1%, Triton X-100 1%, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8 and 150 mM NaCl), twice in High Salt Buffer (SDS 0.1%, Triton X-100 1%, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8 and 500 mM NaCl), twice in LiCl Buffer (0.25 M LiCl, NP-40 1%, Na Deoxycholate 1%, 1 mM EDTA pH 8 and 10 mM Tris-HCl pH 8) and twice in TE. ChIPed material was eluted by two 15 minute incubations at room temperature with 250 ul Elution Buffer (SDS 1%, 0.1 M NaHCO3). Chromatin was reverse-crosslinked by adding 20 ul of NaCl 5M and incubated at 65°C for 4

hours minimum and DNA was submitted to RNase and proteinase K digestion and extracted by phenol-chloroform.

Tandem ChIP was performed by eluting the first ChIP in reduced volume and diluting 10 times in ChIP Dilution Buffer before the second round of ChIP.

ChIP-chip

For ChIP-chip, input chromatin and ChIPed material were amplified using the GenomePlex Complete Whole Genome Amplification Kit (WGA2 087K044, Sigma) and labelled. Samples were hybridised to the extended promoter array from Agilent covering -5kb to +2kb regions of around 17000 mouse cellular promoters.

Data were extracted with Agilent *Feature extraction* and analysed by *ChIP Analytics*. Data were subjected to Blank substraction, Inter-array median, Intra-array (dye-bias) median and Intra-array Lowness (intensity-dependent) normalisations. The Whitehead error model was used and the peak detection was performed with Pre-defined Peak Shape detection v2.0 with a p-value < 0.005 for non-parametric test and a peak-score > 3 for EVD-based score (based on T. Kaplan & N. Friedman "Model-Based Analysis of High resolution Chromatin Immunoprecipitation" Technical Report 2006-11, School of Computer Science & Engineering, Hebrew University 2006).

The output .tsv tables were further analysed using a script to detect individual oligonuleotides whose IP/Input ratio in the ChIP from the tagged cells was more than two-fold higher than in the ChIP from control cells. The selected probes were then screened to determine whether their upstream and downstream neighbours also had similar characteristics. The probes of this list were then sorted for those with a normalised log ratio greater than 1.5. Peaks were determined where a minimum of 3 adjacent probes were found to reproducibly fit the above criteria in the tandem Flag ChIPs and in the HA ChIP.

ChIP-seq.

The ChIPed DNA prepared as described above was prepared and sequenced on an Illumina GAII sequencer according to the Illumina protocols. The raw sequence data was aligned using Eland V1.3.2 and peak detection performed using the MACS software (http://liulab.dfci.harvard.edu/MACS/). The statistics of the peak detection programme are indicated in Supporting Table 4. The peaks were then annotated using the GPAT software (http://bips.u-strasbg.fr/GPAT/Gpat_home.html (2)).

Multiple alignments and motif search at Chiped Loci.

In a first search for DR elements, the 1200 nt surrounding the highest scoring probe in each peak was extracted and the presence of DRs with consensus RGKTCA{1,2,5}RGKTCA detected by UNRREAL, an algorithm available at http://idoubt.fr. UNRREAL, Universal Nuclear Receptor Response Element Analytic Localizer, enables the detection of degenerate consensus half-sites separated by a specified number of base pairs.

A second search of the 100 most sharply defined MEF loci and 211 ES loci was performed on the 300nt surrounding the highest scoring probe. In this case, DRs with consensus RGKTCA/G{1,2,5}RGKTCA/G and half sites with consensus RGKTCA/G were determined, by employing UNRREAL and the *fuzznuc* algorithm from the EMBOSS package (6), respectively. Multiple alignments of the regions containing the DRs or the half-sites was performed using MAFFT (http://www.ebi.ac.uk/Tools/mafft/index.html) and CLUSTALW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The presence of other potential common motifs within these sequences was also analysed by employing two motif discovery algorithms: MDSCAN (3), available at

http://ai.stanford.edu/~xsliu/MDscanhttp://ai.stanford.edu/~xsliu/MDscan/ ;

and WEEDER (5), available at http://159.149.109.9/modtools .

EREs and FOXA1 sites were searched using UNRREAL and *fuzznuc*.

Legends to supplemental figures and tables.

Supplemental Figure 1.

A. Examples of RAR binding sites in MEFs. Graphic representation of tandem-Flag ChIP-chip results in UCSC web browser at the indicated loci. The values on the Y axis show the normalised IP/Input ratio. Results are shown for the tandem-Flag ChIP-chip on RARα and RARγ as well as the HA ChIP on RARγ and the tandem-Flag ChIP-chip on control cells. **B**. Examples of RAR binding sites in ES cells. Results are shown for the duplicate tandem-Flag ChIP-chip on cells expressing tagged RARγ and control cells. **C.** Differential occupation of HoxA and B loci in MEFs and ES cells. RAR occupies the *Hoxa10* locus in MEFs, but the *Hoxa5* lcous in ES cells, while it occupies the *Hoxb4* locus in ES cells, but not in MEFs. **D.** Examples of RAR binding to repressed and active promoters. UCSC view of RAR ChIP-chip and H3K4me3 ChIP-seq in MEFs where RAR is bound to the repressed *Faim2* promoter devoid of H3K4me3 and the active *Bsg* promoter with high levels of H3K4me3.

Supplemental Fig. 2.

Multiple alignments of half sites within RAR bound loci. **A.** The 60 nucleotides surrounding each half site were aligned using MAFFT. The locus of origin and coordiantes of the half site are shown to the left of the figure. In several instances, overlapping 60 nucleotides fragments have been used as they contained closely spaced half sites. In these cases the algorithm aligns the larger overlapping region rather than centering on the half site. **B**. MAFFT alignment of the regions containing identfied DR elements. Here the algorithm successfully identified most of the repeated half sites, but did not reveal any additional sites present in a majority of loci.

Supplemental Table1.

RAR bound genes in MEFs. Excel table of annotated loci bound by RARs. Page 1 shows the loci bound by RARγ with gene name, identity of the peak oligonucleotide, normalised ratio, peak score (both from a representative tandem Flag-ChIP-chip), location of binding sites with respect to TSS, accession number and secondary annotation. Page 2 shows genes bound in common by RARγ and RARα. NB indicates no significant binding. Page 3 compares ChIPed loci and the previously described transcriptome data (1). The gene symbol and fold change (Log_2) for all of the ChIPed genes showing RA regulation are listed in decreasing order.

Supplemental Table 2.

Potential DR elements at RAR bound loci in MEFs. The Excel file contains 6 Pages. **1-3** show the DR1, 2 and 5 respectively with the gene symbol, chromosome, sequence of the DR, chromosomal coordinates of the DR, whether it is a perfect consensus or has the allowed 5'-TCA/G-3' mismatch. All sequences are shown in the same orientation, and the strand on which they are located is indicated. The results are summarised on page **4**. **5** shows the half sites associated with each locus. The spacing between neighbouring half sites on the same strand is shown. Consensus DRs within this group are also indicated. Results are summarised on page **6**.

Supplemental Table 3.

Excel file contains 9 pages. **1**. RAR bound genes in ES cells. Excel table of annotated loci bound by RARs showing the loci bound by RARγ with gene symbol, identity of the peak oligonucleotide, normalised ratio, peak score, location of binding sites with respect to TSS, accession number and secondary annotation. **2** shows genes bound in common by in MEFs and ES cells. **3-5** Potential DR1, 2 and 5 elements at RAR bound loci in ES cells with gene symbol, chromosome, sequence of the DR, chromosomal coordinates of the DR and whether

it is a perfect consensus. All sequences are shown in the same orientation, and the strand on which they are located is indicated. The results are summarised on page **6**. **7-9**. Potential half sites at 211 RAR bound loci in ES cells. The spacing between neighbouring half sites on the same strand is shown.

Supplemental Table 4.

Summary of H3K4me3 ChIP-seq results and intersection with RAR ChIP-chip. The Excel file contains two pages. Page 1 shows the statistics from the sequencing and peak detection. Page 2 shows the combination of the H3K4me3 ChIP-seq results for ES cells taken from supplemental Fig. S4 of Mikkelsen et al (4), our ChIP-seq from MEFs with the RAR ChIP-chip data. Column A) gene symbol, B) Refseq transcript identity, C) promoter class with respect to high, intermediate and low GC content, D-E) H3K4me3 status of the corresponding genes in ES and MEFs, F). H3K4me3 in ES combined with RAR ChIP-chip in ES, ND indicates that the gene is not ChIPed, K4 is ChIPed and H3K4me3 positive, none is ChIPed and no H3K4me3, G) H3K4me3 in MEFs combined with RAR ChIP-chip in MEFs, H). H3K4me3 in ES combined with RAR ChIP-chip in MEFs, I) H3K4me3 in MEFs combined with RAR ChIP-chip in MEFs. Each column can be sorted to reveal genes that are ChIPed in either cell line to reveal their methylation status. It should be noted that when generating the lists of ChIPed genes for this analysis, we included both genes in the case of the divergently transcribed promoters. The total number of potentially regulated genes is therefore higher than the number of bound loci indicated in tables 1 and 4.

References.

1. **Fadloun, A., D. Kobi, L. Delacroix, D. Dembele, I. Michel, A. Lardenois, J. Tisserand, R. Losson, G. Mengus, and I. Davidson.** 2007. Retinoic acid induces TGFbeta-dependent autocrine fibroblast growth. Oncogene.
2. **Krebs, A., M. Frontini, and L. Tora.** 2008. GPAT: retrieval of genomic annotation from large genomic position datasets. BMC Bioinformatics **9:**533.
3. **Liu, X. S., D. L. Brutlag, and J. S. Liu.** 2002. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol **20:**835-9.

4. **Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein.** 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature **448:**553-60.
5. **Pavesi, G., P. Mereghetti, G. Mauri, and G. Pesole.** 2004. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res **32:**W199-203.
6. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16:**276-7.