JBC Papers in Press. Published on June 1, 2012 as Manuscript M112.361790
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M112.361790

Retinoic acid receptors recognise the mouse genome through binding elements with diverse spacing and topology

Emmanuel Moutier*, Tao Ye*, Mohamed-Amin Choukrallah[1], Sylvia Urban, Judit Osz, Amandine Chatagnon[2], Laurence Delacroix[3], Diana Langer, Natacha Rochel, Dino Moras, Gerard Benoit[2], and Irwin Davidson#.

From the Institut de Génétique et de Biologie Moléculaire et Cellulaire.

CNRS/INSERM/UDS., 1 Rue Laurent Fries, 67404 Illkirch Cédex. France.

1.Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland. 2. Centre de Génétique et de Physiologie Moléculaire et Cellulaire, UMR5534 CNRS, Université de Lyon 1, 43 Boulevard du 11 novembre 1918 F-69622 Villeurbanne Cedex, France. 3. GIGA-Neurosciences, CHU, BAT B36 +1, 1 ave de l'Hôpital, 4000 Liège, Belgium.

Running Title : *Characterisation of novel RAREs.*

# To whom correspondence should be addressed, E mail : irwin@titus.u-strasbg.fr. FAX: 33 3 88 65 32 01. TEL: 33 3 88 65 34 40 (45) * These authors contributed equally to this study

Key words : F9 cells, neuronal differentiation, vitamin D3 receptor, ChIP-seq, RNA-seq, Isothermal titration calorimetry.

Background. Retinoic Acid Receptors (RARs) heterodimerise with Retinoid X receptors (RXRs) to regulate gene expression.

Results. This heterodimer recognises the genome via a large and diverse repertoire of direct and inverted repeat DNA elements.

Conclusion. The observed diversity of binding elements changes the paradigm of how RAR-RXR recognises the genome.
Significance. Half site spacing in the DNA binding element allosterically regulates RAR function.

Summary.
Retinoic Acid Receptors (RARs) heterodimerise with Retinoid X Receptors (RXRs) and bind to RA-response elements (RAREs) in the regulatory regions of their target genes. While previous studies on limited sets of RA-regulated genes have defined canonical RAREs as direct repeats of the consensus RGKTCA separated by 1, 2 or 5 nucleotides (DR1, DR2, DR5), we show that in mouse embryoid bodies or F9 embryonal carcinoma cells, RARs occupy a large repertoire of sites with DR0, DR8 and IR0 (inverted repeat 0) elements. Recombinant RAR-RXR binds these non-canonical spacings *in vitro* with comparable affinities to DR2 and DR5. Most DR8 elements comprise three half sites with DR2 and DR0 spacings. This specific half site organisation constitutes a previously unrecognised, but frequent signature of RAR binding elements. In functional assays, DR8 and IR0 elements act as independent RAREs, while DR0 does not. Our results reveal an unexpected diversity in the spacing and topology of binding elements for the RAR-RXR heterodimer. The differential ability of RAR-RXR bound to DR0 compared to DR2, DR5 and DR8 to mediate RA-dependent transcriptional activation indicates that half site spacing allosterically regulates RAR function.

Introduction.

All trans-retinoic acid (RA), the naturally active vitamin A metabolite exerts a wide range of effects on vertebrate development and adult tissue homeostasis by regulating cell proliferation, differentiation and apoptosis (1-3). RA acts through binding to three members of the nuclear receptor (NR) superfamily, RARα RARβ and RARγ that function as ligand-dependent transcriptional regulators. RARs form heterodimers with rexinoid receptors (RXRs α β and γ) and bind to RA response elements (RAREs) located in the regulatory regions of target genes (4,5). RAREs are classically described as direct repeats of the consensus sequence 5'-RGKTCA-3' separated by 1, 2 or 5 nucleotides [for review see (6-8)].

We have previously used chromatin immunoprecipitation (ChIP) coupled with array hybridisation (ChIP-chip) to identify RAR occupied sites in mouse embryonic fibroblasts (MEFs) and in undifferentiated embryonic stem (ES) cells showing that RAR occupancy of target loci is cell-type specific (9). Analysis of these RAR occupied sites revealed that a majority did not comprise canonical DR1, 2 or 5 elements. This paucity can in part be accounted for by the fact that RAR-RXR bind DR1, DR2 or DR5 elements with non-canonical half sequences. An additional possibility that we did not address was the existence of DRs with non-canonical spacings. Evidence for the use of alternate spacing has previously been described (10), but their prevalence at RAR-occupied loci has not been described. We have analysed ChIP-seq data sets from RAR occupancy in ES cells grown as embryoid body precursors to neuronal differentiation and from F9 embryonal carcinoma cells. We identify DR0 as the most abundant RAR-binding element and we identify a novel composite DR8 as a common, but previously unrecognised RARE. Our results reveal an unexpected diversity in the spacing and topology of the DNA binding elements that is unique for the RAR-RXR heterodimer.

Experimental procedures.

ChIP and ChIP-seq.

ChIP and ChIP-seq experiments were performed on chromatin from murine embryonic stem cells grown as embryoid bodies for 4 days in absence of RA and then treated for 2 hours with all-trans retinoic acid as previously described (9,11) 2010, (12). Further details are provided in the Supplemental material. ChIP-seq was performed using the panRAR antibody (Sc-773, Santa Cruz USA), sequenced on an Illumina GAIIx sequencer and the raw data analysed by the Illumina Eland pipeline V1.7 or V1.8 and aligned to the genome with Bowtie. Peak detection was performed using the MACS software (http://liulab.dfci.harvard.edu/MACS/) (13) under settings where an anti-GFP ChIP from the embryoid bodies was used as a negative control. Peaks were then annotated using GPAT ((14), http://bips.u-strasbg.fr/GPAT/Gpat_home.html) with a window of +/- 20kb with respect to the coordinates of the beginning and end of Ensembl genes (release 64). Cluster comparison of ChIP-seq data sets was performed with seqMINER (15). F9 embryonal carcinoma cells were grown under standard conditions and treated with 10^{-6}M RA for 2 hours prior to ChIP-seq with the panRAR antibody (Chatagnon et al., manuscript in preparation). ChIP-seq analysis and annotation was performed as described above.

Electrophoretic Mobility Shift Assay

EMSA assays were performed essentially as previously described (9) using purified bacterial recombinant RARαΔAB-RXRαΔAB as described (16). The RAR binding elements in each oligonucleotide were centered and surrounded by their native flanking sequences. The sequences of the oligonucleotides are available on request. After electrophoresis the gels were dried and exposed to autoradiographic film or a PhosphorImager plate.

Isothermal titration calorimetry.

ITC measurements were performed at 25°C on a MicroCal ITC_{200} (MicroCal). Double stranded DNA and purified proteins bound to 9cis RA were dialyzed extensively against the same buffer used in the ITC experiments. The buffer contained 50 mM Hepes pH 8.0, 100 mM sodium chloride, 2% glycerol and 1 mM TCEP. In a typical experiment 2 µl aliquots of DNA at 80 to 150 µM were injected into a 10 µM RAR–RXR solution (200 µl sample cell). The delay between injections was 120 to 180 s to permit the signal to return to baseline before the next injection. ITC titration curves were analyzed using the software Origin 7.0 (OriginLab). Standard free energies of binding and entropic contributions were obtained, respectively, as $\Delta G = -RT \ln(K_a)$ and $T\Delta S = \Delta H - \Delta G$, from the K_a and ΔH values derived from ITC curve fitting.

Bioinformatics analysis.

The 150 nucleotides surrounding the ChIP-seq peaks were analysed using a custom JAVA API application to detect perfect consensus 5'-RGKTCA-3' half sites with the different spacings. Analysis of the 150 bp regions from ChIP-seq peaks where no canonical DR elements were found using the MEME programme (http://meme.nbcr.net/meme4_1/cgi-bin/meme.cgi) (17)) identified the pseudo DR0 consensus. MEME analysis was also used to derive the consensus DR0 and DR8 sequences by analysis of 200 ChIP-seq peaks of each class. FIMO (http://meme.sdsc.edu/meme/fimo-intro.html) was used to search for the FOXA1 consensus motif (http://jaspar.genereg.net/cgi-bin/jaspar_db.pl) and was run in default parameters with a p-value cutoff of 1e-4. For SPAMO analysis (http://meme.sdsc.edu/meme/doc/spamo.html) we used the motifs identified in the MEME analysis as the principle and spaced motifs. DRs and their constituent half sites were mapped in a 150 bp window

using the peak summit from MACS analysis as the central position. An in house JAVA application was used to align all DRs on the same strand to insure the sense and antisense matches gave homogeneous positions. The motif position profiles are based on a .wig like representation with a step size of 1bp to calculate the number of overlapping motifs present at each position in the 150 bp window.

Cell culture and transfection assays.

JI ES cells (129SV/Jae) were grown on iactivated fibroblast feeder cells in the presence of LIF under standard conditions. Cells were passaged three times in absence of feeders before transfer to bacterial dishes for embryoid body formation. For reporter assays, transfections comprised 1ug of the TATA-chloramphenicol acetyl transferase (CAT) reporters, 1ug pCH110 expressing β-galactosidase as internal standard, 1 ug of pCMV hRARa and RXRa expression vectors. Transfections were performed with Fugene (Roche, Basel Switzerland) and 24 hours after transfection 10^{-6}M RA was added for an additional 24 hours. Extract preparation, β-galactosidase assays and CAT assays were performed using a Roche CAT-Elisa kit as previously described (18). To generate the TATA-CAT reporter plasmids, the various wild-type and repeat elements were generated by DNA synthesis (GeneArt, Germany) with flanking BglII and NotI restriction sites. The plasmids containing these regions were amplified and the inserts purified and recloned between the equivalent sites in the previously described TATA-CAT plasmid (11,18).

mRNA-seq.

Total RNA was extracted from duplicate cultures of embryoid bodies grown for 4 days in absence of RA and after 24 hours of RA-treatment. The mRNA-seq libraries were prepared following the Illumina protocol (see Supporting text). Sequence reads mapped to reference genome mm9/NCBI37 using Tophat (19). Quantification of gene expression was done using Cufflinks (20) and annotations from Ensembl release 62. For each transcript the resulting RPKM were converted into raw read counts and these counts were added for each gene locus. Data normalization was performed as described by Anders et al. (21) and implemented in the DESeq Bioconductor package. Only regulated transcripts with an RPKM>2, an adjusted pvalue of <0.1 and a log2 fold change of >1 and <-1 were considered.

Results.

Representation of DR spacings at RAR-occupied loci.

We performed RAR ChIP-seq in mouse ES cells grown for 4 days as embryoid body (EBs) precursor to neuronal differentiation (22) treated for two hours with RA. Analysis of this data set revealed 13385 RAR occupied loci that could be annotated to 12250 Ensembl and predicted transcripts (equivalent of 6628 RefSeq genes) (Supplemental Table 1). As seen for other nuclear receptors (23,24), more than half of the RAR-occupied sites were located greater than 20 kb upstream or downstream of the transcription start site (TSS) with only 10% residing in the proximal promoter region (-5kb to +2kb with respect to the TSS, Supplemental Fig. 1).

We analysed the frequency of all potential DR spacings from DR0 to DR10 (5'-RGKTCA-$_{0\text{-}10}$-RGKTCA-3') in the 150 bps surrounding the peak summit of the 1000 highest occupied sites based on the number of sequence tags forming the peak. Similarly, as it has been reported that RAR occupied elements are closely associated with estrogen receptor (ER) binding elements (IR3) in MCF7 cells (25), we analysed the RAR occupied loci for IR0-IR10. This analysis showed that DR0 was by far the most frequently represented spacing followed by DR2, DR8, DR5, and

DR1 (Fig. 1A). Similar results were obtained when analysing 1000 medium and 1000 low occupied sites. In each case the DR0 and DR2 spacings remained the most frequent followed by DR5 and DR8. However, the total number of loci comprising canonical half sites with these spacings decreases with occupancy. The higher occupied sites comprise a large proportion of elements with canonical half site repeats, whereas the lower occupied must contain either more elements with degenerate half sites and/or other spacings (compare the total number of DRs in each class in Fig. 1A).

Analysis of the IR frequencies in each class showed that IR0 is the most frequently found in the highly occupied class, while IR9 is also represented in the medium occupied class (Fig. 1A). Again, the frequency of IR elements is higher in the most occupied class than in the lower occupied classes.

The analysis of the total 13385 sites confirmed that DR0 is the most represented spacing followed by DR2, DR5, and DR8 together with IR0 and IR9 (Fig. 1B). Extending the analysis to DR/IR11-DR/IR20 showed that none of these larger spacings was strongly enriched in the data set (data not shown). DR0, not the canonical DR5, is thus the most highly represented DR element at RAR-occupied loci in EBs irrespective of the degree of occupancy in the cells.

We further analysed the frequency of the DR and IR elements in the RAR occupied sites in the proximal promoter region (-5kb to +2kb) to ask if any of these elements are selectively enriched. DR0 remains the most represented element, but the ratios of DR0/DR5 and DR5/DR8 indicate a modest relative enrichment of DR5 at promoters compared to the highest occupied class, but not when compared to the total data set. (Fig. 1C).

It has been suggested that RAR-occupied motifs in MCF7 cells are closely associated with ER binding sites (IR3) and with sites for the pioneer factor FOXA1 (25,26). No significant representation of IR3 was seen at RAR-occupied sites in EBs (and F9 cells, Fig. 1A and D). Analysis of the 150 bp around the top 1000 RAR-occupied peaks in EBs revealed only 47 FOXA1 consensus sites, while the same analysis in a 500 bp window revealed 259 potential sites (data not shown). Thus, while no close association of RAR and ER binding elements are seen in EBs, around 25% of RAR-occupied loci show association with FOXA1 when a wider window is used.

To ask if the above DR frequencies are representative of what can be seen in other cell types, we analysed the RAR ChIP-seq data set obtained from RA-treated F9 embryonal carcinoma cells (Chatagnon et al., manuscript in preparation). Although this data set comprises around 32000 RAR-occupied loci when analysed by MACS using the same parameters as for the EB data set, we restricted our analysis to the top 13385 peaks. Comparison with the EB data set indicated a large fraction of shared sites and potential target genes, and a set of sites specific for each cell type (Supplemental Fig. 2 and data not shown). Analysis of the F9 data set indicated that DR0 is again the most frequent at the high, medium or low occupied peaks followed by DR2, DR5, DR8, DR1, and IR0 (Fig. 1D). Overall, the relative DR and IR frequencies are therefore similar, but not identical to those in the EB data set (Figs. 1B and E) in accordance with the existence of a set of sites specific to F9 cells.

To determine if this spectrum of DR frequencies is found at sites occupied by other nuclear receptors, we analysed DR and IR usage at sites occupied by the vitamin D3 receptor (VDR) in the data set of Ramagopalan et al., (27). In striking contrast to what is oberved for RAR, for VDR the DR3 element is strongly represented followed by the DR4, but there is no high frequency of DR0 or IR0 elements at these sites (Fig. 1F). Our analysis concords with that reported by the

authors of this study who defined DR3 as strongly enriched at these sites. Thus, the RAR-RXR heterodimer binds to DR elements with a much larger variety of spacings than the VDR-RXR heterodimer whose specificity is more restricted.

***In vitro* binding of RAR-RXR to non-canonical spaced DR motifs.**

To determine if DRs with non-canonical spacings can directly bind RAR-RXR, we performed EMSA analysis with oligonucleotide probes derived from sites occupied by RAR and purified bacterially produced recombinant RARαΔAB-RXRαΔAB [(28), hereafter designated simply RAR-RXR].

We first investigated the ability of DR0 elements with fully canonical half sites, including that present in the suppressor of cytokine signalling 3 (*Socs3*) gene locus (Fig. 2A) or half sites with a single base mismatch to compete for RAR-RXR bound to the DR5 from the *Rarb* gene. Several consensus DR0 elements were able to compete formation of the RAR-RXR-DR5 complex (see lanes 6-8, in Fig 2B), while DR0s with one or several mismatches were poor competitors or did not compete (lanes 2-4). On the other hand, a DR0 from the AE binding protein 2 (*Aebp2*) locus with a single mismatch was a good competitor. Thus, as we previously reported (9) in some elements mismatches are tolerated, whereas at others they are detrimental to *in vitro* RAR-RXR binding.

Competition experiments with increasing quantities of competitor olgonucleotide indicated that the *Rarb* DR5 motif efficiently competes with between 25 and 50 fold of cold competitor (Fig. 2C lanes 4 and 5). In comparison, the DR0 motif from the *Socs3* gene is less efficient (lanes 7-10), whereas that from the Musashi homolog 2 (*Msi2*) gene is comparable to DR5 (lanes 11-14). Binding of recombinant RAR-RXR to the *Rarb* DR5 and the *Socs3* and *Msi2* DR0 was also assessed by isothermal titration calorimetry (ITC). A dissociation constant (Kd) of 73 nM for the *Rarb* DR5, and 79 nM for a canonical DR2 from the *Hoxa10* gene was determined, compared with 76 nM for the *Msi2* DR0 and 110 nM for the *Socs3* DR0 (Fig. 3A and B). Thus, consensus DR0s are *bone fide* RAR-RXR binding elements with *in vitro* affinities comparable to the canonical DR5. Close analysis of thermodynamic parameters observed for the binding of RAR-RXR to DNA suggests that enthalpy solely drives this protein-DNA interaction and that increase in the binding affinity to DNA correlates with overall favorable enthalpy.

ChIP-seq identified loci with IR0 elements that efficiently competed complex formation (Supplemental Fig. 3A and B, lanes 4, 6 and 8), while mutated versions conserving only one half site showed no competition (lanes 5, 7, and 9). Titration experiments showed that the IR0s from the methyl-CpG binding domain protein 6 (*Mbd6*) and tripartite motif containing 16 (*Trim16*) genes competed with efficiences comparable to that of the DR5 element (Fig. 2D). This was confirmed by ITC showing a Kd of 58 nM for the *Trim16* IR0 element, and 97 nM for the consensus IR0 from the *Vat1* gene locus (Fig. 3A and B).

A novel composite DR8 RAR binding element.

DR8 is comparable in frequency to DR5 in the EB and F9 data sets. These elements fall into two classes comprising either a simple DR8 with no other potential half sites in the spacer sequence or composite half sites formed by juxtaposition of three half sites with DR2 and DR0 spacings (Fig. 4A). Of the 431 DR0 elements in the top 1000 EB sites (Fig. 1A), 183 form DR8 with an additional 5'-RGKTCA-3' half site. MEME analysis of the remaining DR0 motifs revealed the presence of a variant (5'-RGATCA-3') half site at a further 61 loci (Fig. 4A). This unique and specific topological arrangement of half sites therefore constitutes a highly represented

element at RAR occupied loci and leaves only 187 DR0s with no recognisable 5' half site with a DR2 spacing. Similarly in F9 cells, 63 of the 74 DR8s in highly occupied class have an additional DR2 spaced half site (Supplemental Fig. 4A).

We next asked whether RAR bound the DR2, DR0 or DR8 spacings of these elements. We chose two composite DR8 elements identified by ChIP-seq at the v-maf musculoaponeurotic fibrosarcoma oncogene family, protein A (*Mafa*) and CD97 antigen (*Cd97*) gene loci (Supplemental Fig. 4A). The wild-type *Mafa* and *Cd97* elements both efficiently competed complex formation in EMSA assays (Supplemental Fig. 4B, lanes 4 and 9, and Fig. 4B lanes 7-10). Oligonucleotides mutated in the central half site leaving the DR8 spacing also competed (Supplemental Fig. 4B, lanes 5 and 12 and Fig. 4B lanes 19-22). Mutation of the 5' half site leaving only the DR0 resulted in much less efficient competition (Supplemental Fig. 4B, lanes 6 and 10 and Fig. 4B, lanes 11-14). In contrast, mutation of the 3' site leaving the DR2 led to efficient competition (Supplemental Fig. 4B, lanes 7 and 11 and Fig. 4B, lanes 15-18). Mutation of the 5' and 3' sites leaving only a single half site essentially abolished competition, (Supplemental Fig. 4B, lanes 8 and 13 and Fig. 4B, lanes 23-26). Further evidence that DR8 can bind RAR-RXR comes from the *Dedd* gene locus whose simple DR8 is an efficient competitor in EMSA assays (Supplemental Fig. 5A, compare lanes 1-4 with 9-12 and the mutated element in lanes 5-8). It is noteworthy that at least *in vitro*, DR8 elements with the variant 5'-RGATCA-3' half site are much less efficient competitors than those with the 5'-RGKTCA-3' half sites (Supplemental Fig. 5B).

The binding of the RAR-RXR to the *Mafa* element was also analysed by ITC, showing that the *Mafa* DR2 had a high affinity with a Kd of 25 nM, while this DR0 had a lower affinity than the other DR0s with a Kd of 180 nM, (Fig. 3A and B). ITC performed with the WT *Mafa* element containing all three half sites showed bimodal binding. The first phase indicated occupancy of a high affinity site with a Kd of 13 nM, followed by a second phase with much lower affinity of 180 nM. These results are consistent with binding of high affinity DR2 or DR8 and lower affinity to the DR0.

We also labelled oligonucleotides comprising the WT and mutated *Mafa* DR8, the *Msi2* DR0 and the *Dedd* DR8 and tested their ability to form a complex with recombinant RAR-RXR. The WT *Mafa* element formed a complex with mobility identical to *Rarb* DR5 or *Hoxa*10 DR2 elements (Supplemental Fig. 5C, lanes 4-5, 1-2, 14-15). Complex formation was also seen with the DR8 and DR2 combinations, while binding to the DR0 was less efficient (lanes 6-11). The *Msi2* DR0 and the *Dedd* DR8 also form complexes with RAR-RXR (lanes 12-13, 16-17). These results confirm those of the EMSA competition and ITC showing that DR0 and DR8 form complexes with RAR-RXR. Furthermore, a single RAR-RXR heterodimer appears to bind to the WT *Mafa* DR8, as no slower migrating complex corresponding to an additional RAR and/or RXR bound to the third half site was observed.

We analysed the position of the constitutive half sites in the 183 DR8s from the top 1000 sites of EB data set with respect to the ChIP-seq peak summit. For each half site two populations are seen. The first is consistent with occupancy of the DR2 motif localising the two half sites of the DR2 close to the summit and the 3' half site of the DR0 downstream (Fig. 4C). The second population is more consistent with occupancy of the DR0 and localises the 5' half-site of the DR2 to a more upstream position and the 3' half site of the DR0 towards the peak center. As a consequence of the fact that the first of these two populations is more abundant, DR8 elements (in this case all of the DR8s

from the data set) show a skewed localisation biased towards the 3' side of the peak center (Fig. 4D). Consistent with this, analysis of the total DR2 population shows a bias towards the 5' side as a subpopulation of DR2s are present in DR8 elements where the DR0 is occupied, but not the DR2 (Fig. 4D). In addition, analysis of the total DR0 population shows elements localised 3' to the center corresponding to DR8 motifs where the DR2 is occupied and a second population located around the summit corresponding to occupancy of the DR0.

This analysis is consistent with the idea that the DR2 element of the composite DR8 is preferentially occupied with a subpopulation showing occupancy of the DR0 spacing. Our analysis does not reveal whether this is a stochastic effect on the general DR8 population, or whether differences in half site sequence and context favour occupancy of a specific spacing at different subpopulations of sites.

DR8 elements act as RA responsive elements.

We next asked if the DR0, IR0 and DR8 elements could act as RAREs. Three copies of the wild-type or mutated elements were inserted upstream of a TATA element in a chloramphenicol acetyl transferase (CAT) reporter vector (Fig. 5A). These vectors were transfected into COS1 cells along with vectors expressing RAR and RXR and CAT activity measured in presence and absence of RA.

A positive control vector with 3 copies of the Wt *Rarb* DR5 element strongly responded to RA, while mutation of one half site in each copy abolished the response and no effect of RA was seen using the empty vector (Fig. 5B). Vectors comprising IR0 and simple DR8 all showed a robust response to RA that was lost when one the the half sites was mutated (Fig. 5B). In contrast, no significant response was seen with vectors carrying DR0 elements. These results show that IR0 and the DR8 elements act as independent RAREs, but the DR0 does not have this activity.

To dissect the activity of the composite DR8 element, we transfected vectors in which mutations leave the DR8, DR2 or DR0 elements intact. As in the first set of experiments both the simple and composite DR8s show a robust RA response (Fig. 5B and C). Mutation leaving only the DR8 shows a strong RA response (Fig, 5C) in agreement with the observation that a simple DR8 acts as an RARE. The DR2 is also RA responsive, whereas when only the DR0 or a single half site are present, no RA response is seen. These results indicate that in the composite DR8, the DR2 and DR8, but not DR0, combinations are RA-responsive.

Characterisation of a degenerate pseudo-DR0 element.

The combination of all the above half-site spacings account for only a subset of the RAR-occupied loci. For example, in the EB data set, DR0, DR1, DR2, DR5, DR8 and IR0 with canonical half-site sequence are present at 6061 of the 13385 sites, while in the F9 data set, they account for 4574 of the 13385 selected sites. It should be noted however that in EBs there are 1010 elements in the 1000 highest occupied sites and almost all loci have at least one element with the above spacings.

To determine if there are other sequences that may be recognised by the RAR-RXR, we selected a series of sites with no consensus DR element and performed *de novo* motif detection using the MEME programme. One of the motifs generated by this analysis (pseudo DR0) resembles a degenerate DR0 type motif (Fig. 6A, and Supplemental Fig. 6). This motif was present at RAR-occupied sites such as those at the *Hoxb13*, and the WD repeat and SOCS box-containing 2 (*Wsb2*) gene loci (Fig. 6B and C and Supplemental Fig. 6). A more degenerate version of this sequence with a non-consensus 3'-half site was found at the and Rho family GTPase 3

(*Rnd3*) gene locus (Fig. 6C and Supplemental Fig. 6).

We performed EMSA competition with wild-type and mutated versions of the *Hoxb13* and *Rnd3* elements. The wild-type elements efficiently competed (Fig. 6C, lanes 5 and 9), while versions in which the central 5'-TCAA-3' core is mutated did not compete (lanes 6 and 10). Mutation of single nucleotides in the second pesudo half site did not however affect competition (lanes 7-8, 11-12). ITC measurement of binding to these elements indicated a Kd for *Hoxb13* of 116 nM and of 93 nM for *Rnd3* (Fig. 3B). Thus, despite its non-consensus 3'-half site sequence, the *Rnd3* element binds RAR-RXR with an affinity comparable to the canonical DR5.

The pseudo DR0 from the SET nuclear oncogene (*Set*) has a sequence identical to that of *Hoxb13* and efficiently competes (Fig. 6C, lane 13). In contrast, the tetraspanin 9 (*Tspan9*) gene element does not compete (lanes 14). Similarly, the *Wsb2* gene element also shows almost no ability to compete (lane 15). Consequently at these loci, it is probable that RAR-binding is mediated by other as yet unidentified elements. The above elements contain the 5'-TCAA-3' core, but the conserved G at position 2 in canonical half sites and in the pseudo half sites of the *Hoxb13* and *Rnd3* elements is not conserved. Although these motifs do not contain 2 canonical half sites, their ability to bind RAR-RXR requires conservation of the G residue at position 2 in the first and second half sites.

We tested the ability of the *Hoxb13* pesudo DR0 to act as an independent RARE by inserting 3 copies in the above described CAT reporter vector. After transfection, no RA response is observed consistent with the fact that consensus DR0 elements do not show RA responsiveness in this assay (Fig. 5B).

Association of DR0 and DR8 elements with other motifs.

While performing the MEME analysis on the 150 bp at the DR0 and DR8 containing sites, we noted that these were not the only motifs represented at these sites. Both the MEME analysis and a subsequent SPAMO [Spaced motif analysis tool, (29)] analysis revealed the presence of several other motifs specifically located both 5' and 3' to the DR0 elements often with a rather precise spacing (Supplemental Fig. 7A). TOMTOM analysis (30) of these motifs revealed that one of them located 5' to the DR0/8 was a potential PITX2 binding site, the other motifs showing no significant similarity to known transcription factor binding sites (Supplemental Fig. 7B). This analysis indicates that DR0/8 lie within a longer and more complex element made up of several highly represented motifs with specific locations with respect to the DR0/8 (Supplemental Fig. 7C). Note however that most regions contain a subset of these motifs, only a few comprise all of them. This characteristic is specific for the DR0/8 as no such conservation was seen around the DR5 elements (data not shown).

DR5 is enriched at RAR-occupied sites associated with RA-regulated genes.

We next examined the relationship between RAR occupancy and RA-regulated gene expression. EBs were grown for 4 days in absence of RA and then treated with RA for 24 hours. The transcripts whose expression is induced or repressed by RA compared to the untreated EBs were then assessed by RNA-seq. Transcripts showing a greater than $Log_2$1-fold change in expression were determined (Supplemental Table 2) identifying 824 induced and 379 repressed transcripts. The induced transcripts are strongly enriched in Hox-familly and other homeobox-containing transcription factors responsible for subsequent EB patterning. No specific

class was represented in the repressed genes.

This data was then compared to those having an RAR-occupied element in the 20 kb upstream/downstream or anywhere within the gene body in the 2 hour ChIP-seq data set. This analysis does not take into account RAR-occupied sites in far intergenic regions that cannot be readily be assigned to potential target genes. Keeping in mind this caveat, 495 induced and 200 repressed transcripts were identified as potential direct RAR targets (Fig. 7A). A majority of the transcripts whose expression is induced and about half those repressed by RA at 24 hours are potentially directly regulated by the RARs.

We next asked whether the RAR-occupied sites associated with RA-regulated transcripts were enriched in a particular class of DR/IR element. The 495 up-regulated transcripts were associated with 943 RAR-occupied sites showing a profile where although DR0 remained the most common element, the DR5 was relatively enriched (Fig. 7B). At the highly occupied EB sites, the ratio of DR5 to DR0 was 0.18 compared to 0.3 in the total data set, while at the up-regulated genes it is enriched to 0.46 (Fig. 7D). Similarly the DR8 element remains prominent in this category. The IR0 element is well represented at the up-regulated transcripts, with 20 elements at the 943 sites compared to the 12 at the 1000 most occupied sites (Figs. 7B and 1A). At RA-repressed transcripts, DR0 and DR2 elements remain the most represented, with a low number of DR5 and DR8 elements, but IR0 elements are strongly depleted (Figs. 7C and D). Thus, transcripts whose expression is activated by RA in the EBs are relatively enriched in DR5 and IR0, and DR2 and DR8 elements remain highly represented.

Discussion.

Flexibility in half site spacing and topology of binding elements bound by the RAR-RXR heterodimer.

Here we show that DR0 not DR5 is the most frequent half-site spacing seen at RAR occupied sites in EBs or F9 cells. DR0 elements can be subdivided into three classes, with either an additional canonical half site or a variant 5'-RGATCA-3' half site with DR2 spacing forming composite DR8s, or DR0s with no additional 5' half-sites. DR0 of the latter class bind RAR-RXR *in vitro* with affinities comparable to those of DR2 and DR5. The location of a subset of DR0 elements at the center of the ChIP-seq peak is also consistent with their occupancy in cells. In addition, we also identified pseudo DR0 elements with 3' half sites differing significantly from the consensus sequence. Remarkably, some of these elements have a high affinity *in vitro* for RAR-RXR.

Despite the prevelance of the DR0 element at RAR-occupied loci, it does not act as an independent RARE when placed upstream of a minimal TATA element. It remains to be determined whether the DR0 can confer RA responsiveness in the context of natural promoters acting in combination with other transcription factors. DR0s are components of a larger element composed of several sequence motifs amongst which are potential PITX2 binding sites. DR0s may function in the context of this larger element in cell types where the appropriate additional factors are present.

The inability of the RAR to activate transcription when bound to DR0 compared to DR2, DR5 and DR8 indicates that half site spacing exerts allosteric control on RAR activity. We have modelled the structure of the RAR-RXR DBDs bound to DR0 and IR0 spacings based on the known structures of these DBDs bound to DR5 or of the ecdysone receptor heterodimer bound to a IR1 [(31) and Supplemental Fig. 8A]. 3-dimensional models indicate that a conformational change of the DBD may be required to

circumvent steric hindrance upon binding to IR0 or to favor dimer interactions on DR0 (Supplemental Figs. 8B and C). These conformational changes likely affect the overall conformation of the heterodimer and co-regulator positioning. Future biophysical and structural experiments will precisely define these changes and why they are compatible with function in the case of IR0, but not in the case of DR0. Nevertheless, these observations reinforce previous examples of allostric control of RAR-RXR function (32) and are reminiscient of those showing that the DNA binding element can act as an allosteric regulator of glucocorticoid receptor (33) .

In the context of embryoid bodies it is interesting to note that RAR-RXR binding to DR0s may antagonise the activity of GCNF (germ cell nuclear factor or NR6A1) a distantly related member of the NR superfamily that binds DR0 and appears important for down-regulation of pluripotency genes in RA-induced differentiation (34,35). RAR-RXR and GCNF may therefore compete for binding to DR0 elements and hence antagonize each other. It remains to be determined whether GCNF and RAR-RXR compete for binding to DR0s or bind different sets of DR0s.

EMSA and ITC showed that IR0 elements are high affinity RAR-RXR binding sites and transfection assays showed that they can act as independent RAREs. Previously, an IR0 type RARE has been reported in the mouse gene nuclear receptor subfamily 2, group C, member 1 (*Nr2c1*) gene (36) showing that this element acts as an RARE in natural promoters and it acts as an RARE when the RAR was expressed in yeast (37). This is consistant with their enrichment at RA-regulated transcripts suggesting many of these elements act as RAREs. Nevertheless, although IR0 is the most frequent IR element, it is much less abundant than DRs at RAR-occupied sites. We also observed representation of IR9 at RAR occupied sites, but no binding of RAR-RXR to these sites was seen in EMSA (our unpublished data).

Together our data indicate that the RAR-RXR heterodimer can bind to a variety of half sites spacings. This property is not seen for VDR-RXR that exhibits a strong specificity for DR3 (38), nor for PPAR-RXR that recognises DR1 (39). The ability to recognise such diverse spacings can be explained by the lack of defined secondary structure of the hinge connecting the DNA and ligand binding domains of the RAR and RXR (28). This contrasts with the α-helical structure of the VDR and TR hinges that imposes a constraint on half-site spacing (28,40). For PPAR-RXR, constraint is imposed by the additional interactions of the PPAR hinge with the 5'-AACT-3' motif positioned 5' of the half-site.

Composite DR8, a novel signature of RAR-occupied loci.

An important observation of our study is the high frequency of composite DR8 elements at RAR-occupied sites shortly after RA treatment. Composite DR8 elements efficiently bind RAR-RXR *in vitro* and dissection of the DR2, DR8 and DR0 components shows that the DR2 and DR8 combinations both bind efficiently, whereas the tested DR0s are less efficient. It remains to be determined why the DR0 in this context appears to be less favoured. RAR-RXR bind the DR8 spacing as indicated by the fact that DR8 elements with no intervening half site efficiently compete RAR-RXR in EMSA *in vitro* and they act as RAREs in cells. Similarly mutations of the composite DR8 leaving only the DR8 spacing also act as RAREs in cells. Moreover, a simple DR8 element has been previously reported to mediate the response to 9-cis RA in the *IGFBP3* promoter (41), or the human H1(0) histone gene (42).

Despite these observations, our data are more compatible with preferential occupation of the DR2 and the DR0 than

of the DR8 in cells. Moreover, as both the DR2 or DR8 spacings act as RAREs, what is the significance of the DR2-DR0 half site arrangement in the composite DR8? One possibility is that an RAR-RXR heterodimer may bind to the DR2 or DR0 spacings along with an orphan receptor that could bind to the additional 5' or 3' half site. Although further experiments will be required to determine whether this occurs *in vivo*, our data show that this previously unrecognised half site organisation is a signature of a large number of RAR-occupied sites.

Our results are not the first to report overlapping or composite RAR binding elements. For example, competitive binding to a composite DR3-DR9 arrangement of three half sites has been shown to mediate RA inhibition of VDR activity at the *Itgb3* promoter (43). Similarly, composite elements mediating both RAR and ER responses have been identified in the lactoferrin and placental lactogen promoters (44,45), but such composite elements also mediate antagonism between these pathways in breast cancer cells (25). Also a composite DR/IR sequence forms part of a complex regulatory element involved in regulation of γF-crystallin expression by RAR, PAX6 and large MAF proteins (46). While the above illustrate specific examples of overlapping elements, the composite DR8 described here represents a more specific and frequent half site organisation.

Comparison of RAR genomic occupancy and RA-regulated transcription shows that a majority of transcripts regulated by RA at 24 hours are potential direct targets showing occupancy by RAR at 2 hours. This is consistent with the observation that the major changes in expression of many of the transcripts takes place over the first 12 hours [(47) and our unpublished data]. Although DR5 elements are enriched at sites associated with RA-regulated transcription, consistent with their known role as RAREs, DR2, DR8 and IR0 are also strongly represented at sites associated with RA-regulated transcripts. This together with their ability to act as independent RAREs suggest that each type of element may contribute to the RA response in EBs.

Comparison of our findings with those previously reported highlight similarities, but also major differences. Mahony *et al* (48), have previously reported ChIP-seq data from EBs treated with RA for 8 hours. These EBs were not grown under the same conditions as reported here and their data set comprised much less unique sequence reads resulting in 1924 RAR-occupied sites, far less than identified in our data sets. Mahony *et al*., analysed their data set for the presence of DR0-DR10 and IR0-IR10 and found DR5 and DR2 as the most highly represented elements with only a small number of DR0, and almost no IR0. In contrast, Mahoney et al., did not identify the composite DR8 motif in their data sets.

Similarly, the analysis of Hua *et al.,* of RAR-occupancy in MCF7 cells identified DR5 as the most frequent element, but also noted a low frequency of DR0 elements (25). Surprisingly however, they observed very few DR2 and DR8 elements at RAR occupied loci in this cell type. Unlike MCF7 cells, we did not observe a high frequency of IR3 ER binding elements colocalising with RAR in EBs or F9 cells. Differencies in species and/or cell type may explain these contrasting observations.

In summary, our results change the paradigm for how RAR-RXR recognises the genome from predominantly DR2 and DR5 elements to a more complex situation with a variety of half-site spacings and topologies and an allosteric regulation of RAR function by the half site spacing of the DNA binding element.

Acknowledgements
We thank Carole Peluso-Iltis for technical help in protein purification, and all members of the IGBMC high throughput sequencing platform in particular C Keime, for analysis of RNA-seq data and S. Le Gras for pipline analysis. This work was supported by grants from the CNRS, the INSERM, the Université de Strasbourg, the Association pour la Recherche contre le Cancer, the Ligue Nationale contre le Cancer, and the INCA grant N° 2008-037, the EuTRACC programme of the European union and the French ANR project Rapssodi. ID is an 'équipe labellisée' of the Ligue Nationale contre le Cancer. EM was supported by a fellowship from the ARC.

The data described in this paper are available at:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=tdinlicowokymvw&acc=GSE35599

Author contributions.
E.M. performed all the EMSA and functional analysis and contributed to the bioinformatics analysis. T.Y. performed most of the bioinformatics analysis. M-A. C. performed the 2 hour EB-ChIP-seq experiments. S.U. performed the 24 hour EB-ChIP-seq experiments and RNA-seq. J.O. and N.R. performed and analysed the ITC experiments. A.C. and G.B. performed the F9 ChIP-seq analysis. L.D initiated the analysis of non-canonical spacings. and D.L. performed EB culture and neuronal differentiation. E.M. M-A.C. D.M. G.B. and I.D. designed experiments analysed the data and wrote the paper.

References

1. Blomhoff, R., and Blomhoff, H. K. (2006) *J Neurobiol* **66**, 606-630
2. Mark, M., Ghyselinck, N. B., and Chambon, P. (2006) *Annu Rev Pharmacol Toxicol* **46**, 451-480
3. Niederreither, K., and Dolle, P. (2008) *Nat Rev Genet* **9**, 541-553
4. Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J., and Gronemeyer, H. (2006) *Pharmacol Rev* **58**, 760-772
5. Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J., and Gronemeyer, H. (2006) *Pharmacol Rev* **58**, 712-725
6. Bastien, J., and Rochette-Egly, C. (2004) *Gene* **328**, 1-16
7. Balmer, J. E., and Blomhoff, R. (2002) *J Lipid Res* **43**, 1773-1808
8. Balmer, J. E., and Blomhoff, R. (2005) *J Steroid Biochem Mol Biol* **96**, 347-354
9. Delacroix, L., Moutier, E., Altobelli, G., Legras, S., Poch, O., Choukrallah, M. A., Bertin, I., Jost, B., and Davidson, I. (2010) *Mol Cell Biol* **30**, 231-244
10. Mangelsdorf, D. J., Umesono, K., and Evans, R.M. (1994) *In The Retinoids: Biology, Chemistry and Medicine. M.B. Sporn, A.B. Roberts, and D.S. Goodman, eds. (New York: Raven Press)*, 319-349
11. Kobi, D., Steunou, A. L., Dembele, D., Legras, S., Larue, L., Nieto, L., and Davidson, I. (2010) *Pigment Cell Melanoma Res* **23**, 404-418
12. Martianov, I., Choukrallah, M. A., Krebs, A., Ye, T., Legras, S., Rijkers, E., Vanijcken, W., Jost, B., Sassone-Corsi, P., and Davidson, I. (2010) *BMC Genomics* **11**, 530

13. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008) *Genome Biol* **9**, R137
14. Krebs, A., Frontini, M., and Tora, L. (2008) *BMC Bioinformatics* **9**, 533
15. Ye, T., Krebs, A. R., Choukrallah, M. A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011) *Nucleic Acids Research* **39**, e35
16. Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mely, Y., Svergun, D. I., and Moras, D. (2011) *Nat Struct Mol Biol* **18**, 564-570
17. Bailey, T. L. a. E., C. (1994) *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* AAAI Press, Menlo Park, California, 28-36
18. Mengus, G., Fadloun, A., Kobi, D., Thibault, C., Perletti, L., Michel, I., and Davidson, I. (2005) *Embo J* **24**, 2753-2767
19. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) *Bioinformatics* **25**, 1105-1111
20. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) *Nat Biotechnol* **28**, 511-515
21. Anders, S., and Huber, W. (2010) *Genome Biol* **11**, R106
22. Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., and Barde, Y. A. (2004) *Nat Neurosci* **7**, 1003-1009
23. Welboren, W. J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N., and Stunnenberg, H. G. (2009) *Embo J* **28**, 1418-1428
24. Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., Wang, Q., Bekiranov, S., Sementchenko, V., Fox, E. A., Silver, P. A., Gingeras, T. R., Liu, X. S., and Brown, M. (2006) *Nat Genet* **38**, 1289-1297
25. Hua, S., Kittler, R., and White, K. P. (2009) *Cell* **137**, 1259-1271
26. Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A., and Brown, M. (2005) *Cell* **122**, 33-43
27. Ramagopalan, S. V., Heger, A., Berlanga, A. J., Maugeri, N. J., Lincoln, M. R., Burrell, A., Handunnetthi, L., Handel, A. E., Disanto, G., Orton, S. M., Watson, C. T., Morahan, J. M., Giovannoni, G., Ponting, C. P., Ebers, G. C., and Knight, J. C. (2010) *Genome Res* **20**, 1352-1360
28. Rochel, N. C., F. Godet, J. Roessle, M. Peluso-Iltis, C. Svergun, DI. Moras, D. . (2011) *Nature Structural and Molecular Biology* **5,** 564-70.
29. Whitington, T., Frith, M. C., Johnson, J., and Bailey, T. L. (2011) *Nucleic Acids Research* **39**, e98
30. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007) *Genome biology* **8**, R24
31. Devarakonda, S., Harp, J. M., Kim, Y., Ozyhar, A., and Rastinejad, F. (2003) *The EMBO journal* **22**, 5827-5840
32. Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M. G., Heyman, R. A., and Glass, C. K. (1994) *Nature* **371**, 528-531
33. Meijsing, S. H., Pufall, M. A., So, A. Y., Bates, D. L., Chen, L., and Yamamoto, K. R. (2009) *Science* **324**, 407-410
34. Gu, P., LeMenuet, D., Chung, A. C., Mancini, M., Wheeler, D. A., and Cooney, A. J. (2005) *Molecular and Cellular Biology* **25**, 8507-8519
35. Zechel, C. (2005) *Molecular reproduction and development* **72**, 550-556
36. Lee, C. H., and Wei, L. N. (1999) *Biochemistry* **38**, 8820-8825

37. Heery, D. M., Pierrat, B., Gronemeyer, H., Chambon, P., and Losson, R. (1994) *Nucleic Acids Res.* **22**, 726-731
38. Ramagopalan, S. V., Heger, A., Berlanga, A. J., Maugeri, N. J., Lincoln, M. R., Burrell, A., Handunnetthi, L., Handel, A. E., Disanto, G., Orton, S. M., Watson, C. T., Morahan, J. M., Giovannoni, G., Ponting, C. P., Ebers, G. C., and Knight, J. C. (2010) *Genome research* **20**, 1352-1360
39. Nielsen, R., Pedersen, T. A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., Denissov, S., Borgesen, M., Francoijs, K. J., Mandrup, S., and Stunnenberg, H. G. (2008) *Genes Dev* **22**, 2953-2967
40. Orlov, I., Rochel, N., Moras, D., and Klaholz, B. P. (2012) *The EMBO journal* **31**, 291-300
41. Chang, Y. S., Cho, J. Y., Cho, H. A., Kim, H. J., Chang, J., Ahn, C. M., Kim, S. K., and Kim, S. K. (2006) *Cancer Biol Ther* **5**, 586-592
42. Bouterfa, H. L., Piedrafita, F. J., Doenecke, D., and Pfahl, M. (1995) *DNA Cell Biol* **14**, 909-919
43. Cao, X., Teitelbaum, S. L., Zhu, H. J., Zhang, L., Feng, X., and Ross, F. P. (1996) *The Journal of biological chemistry* **271**, 20650-20654
44. Stephanou, A., and Handwerger, S. (1995) *Endocrinology* **136**, 933-938
45. Lee, M. O., Liu, Y., and Zhang, X. K. (1995) *Molecular and Cellular Biology* **15**, 4194-4207
46. Yang, Y., Chauhan, B. K., Cveklova, K., and Cvekl, A. (2004) *Journal of molecular biology* **344**, 351-368
47. Simandi, Z., Balint, B. L., Poliska, S., Ruhl, R., and Nagy, L. (2010) *FEBS letters* **584**, 3123-3130
48. Mahony, S., Mazzoni, E. O., McCuine, S., Young, R. A., Wichterle, H., and Gifford, D. K. (2011) *Genome Biol* **12**, R2

Legends to figures.

Figure 1. DR and IR frequencies at RAR-occupied sites in embryoid bodies and F9 cells. **A** Frequencies of the indicated half site spacings at the high, middle and low occupied sites in the embryoid body RAR ChIP-seq data set. **B.** DR/IR frequencies in the total data set of 13385 embryoid body sites. **C.** DR/IR frequencies at promoter proximal sites. **D.** DR/IR frequencies at the high, middle and low occupied sites in the RAR F9 cell RAR ChIP-seq data set. **E**. DR/IR frequencies at the top 13385 F9 sites. **F.** DR/IR frequencies at VDR-occupied sites from the data set of (27).

Figure 2. RAR-RXR binding to DR0 elements. **A**. UCSC view of sequence tag density in .wig file format of the RAR-occupied site at the *Socs3* gene comprising a DR0 in EBs. **B**. EMSA analysis showing the ability of the indicated DR0 elements to compete with the labelled *Rarb* DR5 element for RAR-RXR complex formation. The sequences of the DR0 elements within the competing olgonucleotides are shown with the repeated half sites indicated by arrows. Variations form the consensus half site sequence are indicated in red. All competitors were used a 100-fold excess. **C.** Competition was performed with increasing quantities (10, 25, 50, 100-fold excess) of the oligonucleotides shown above each lane. Lane 1 is the oligonucleotide probe with no recombinant RAR-RXR and lane 2 with RAR-RXR, but no competitor. **D**. EMSA competition analysis of the indicated IR0 elements. The sequences of the IR0 motifs within the competing olgonucleotides are shown with the inverted half sites indicated by arrows. Mutated nucleotides are indicated in red. Competition was performed with increasing quantities as above.

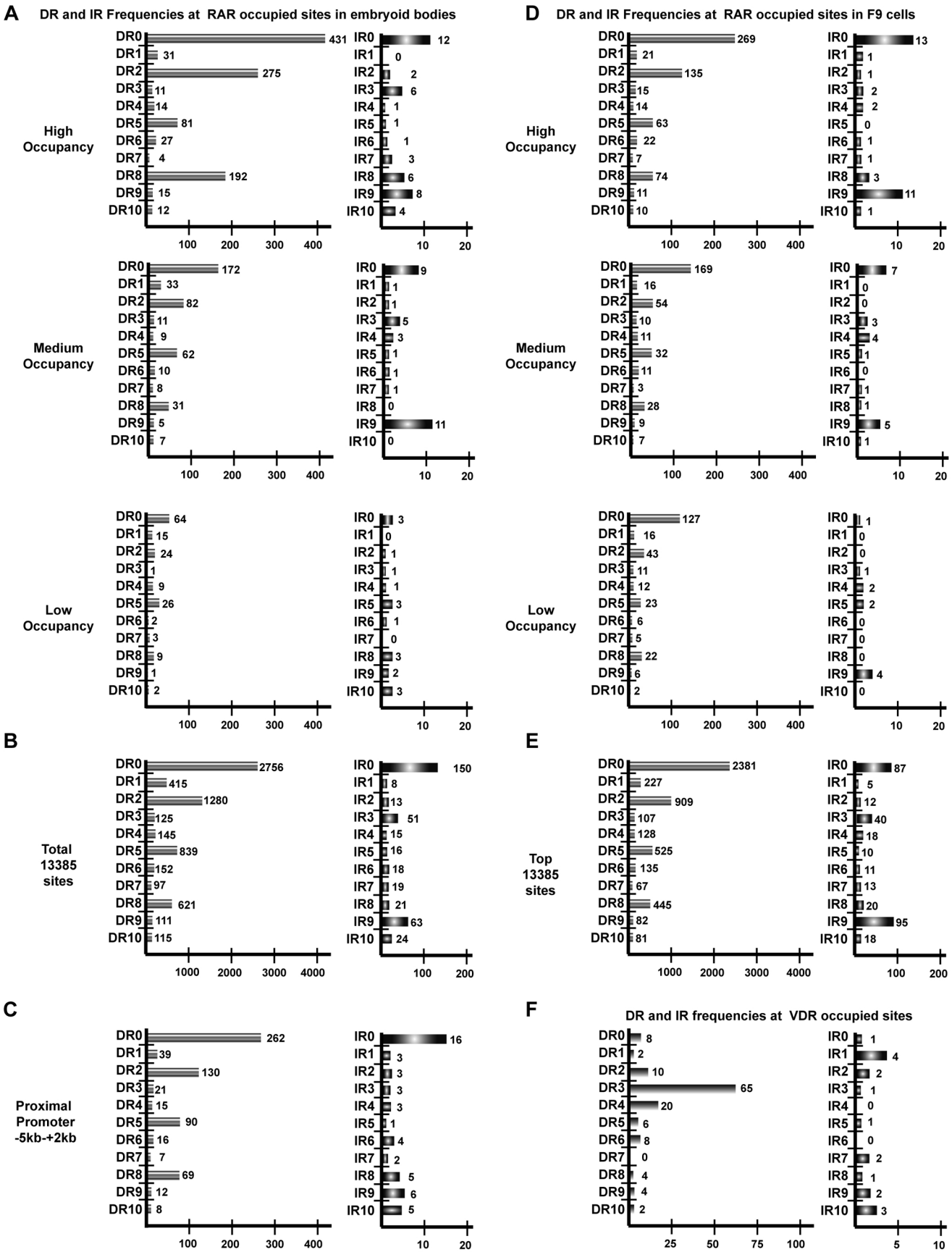
Figure 3. Quantification of the interaction between RAR-RXR and different RAR binding element oligonucleotides by ITC. **A.** Representative ITC titrations of RAR/RXR to the indicated DR/IR elemnts. RAR-RXR binds DR5, DR2, DR0 and IR0 with a stoichiometry of 1 DNA per heterodimer. **B**. Thermodynamic parameters (dissociation constant (Kd), enthalpy (ΔH) and entropy (-TΔS)) of DNA binding to RAR-RXR as determined by ITC.

Figure 4. Composite DR8 elements. **A**. Consensus sequence of the composite DR8 element and the DR0 elements derived from MEME analysis. The number of DR0 elements in the 1000 highest occupied sites (total of 431) in EBs with the indicated additional 5' half sites, or without an additional half site are calculated. **B.** EMSA competition was performed with increasing quantities (10, 25, 50, 100-fold excess) of the oligonucleotides shown above each lane. Lane 1 is the oligonucleotide probe with no recombinant RAR-RXR and lane 2 with RAR-RXR, but no competitor. The wild type and mutated oligonucleotides are shown below the EMSA panel, and mutations are indicated in red. **C.** Location of the half sites 1, 2 and 3 (see panel A) of the DR8 elements from the 1000 highest occupied sites within the 150 bp window around the peak summit. The location relative to the summit is indicated on the X axis and the number of half sites on the Y axis. The red arrows indicate the major population where the DR2 spacing is closest to the peak center and the black arrows the minor population more consistent with DR8/DR0 occupancy. **D**. Location of the DR8, DR0 and DR2 elements within the 150 bp window around the peak summit in the 1000 high, medium, and low occupied classes of the EB data set. X and Y axis as indicated in panel B. Red and black arrows as described above.

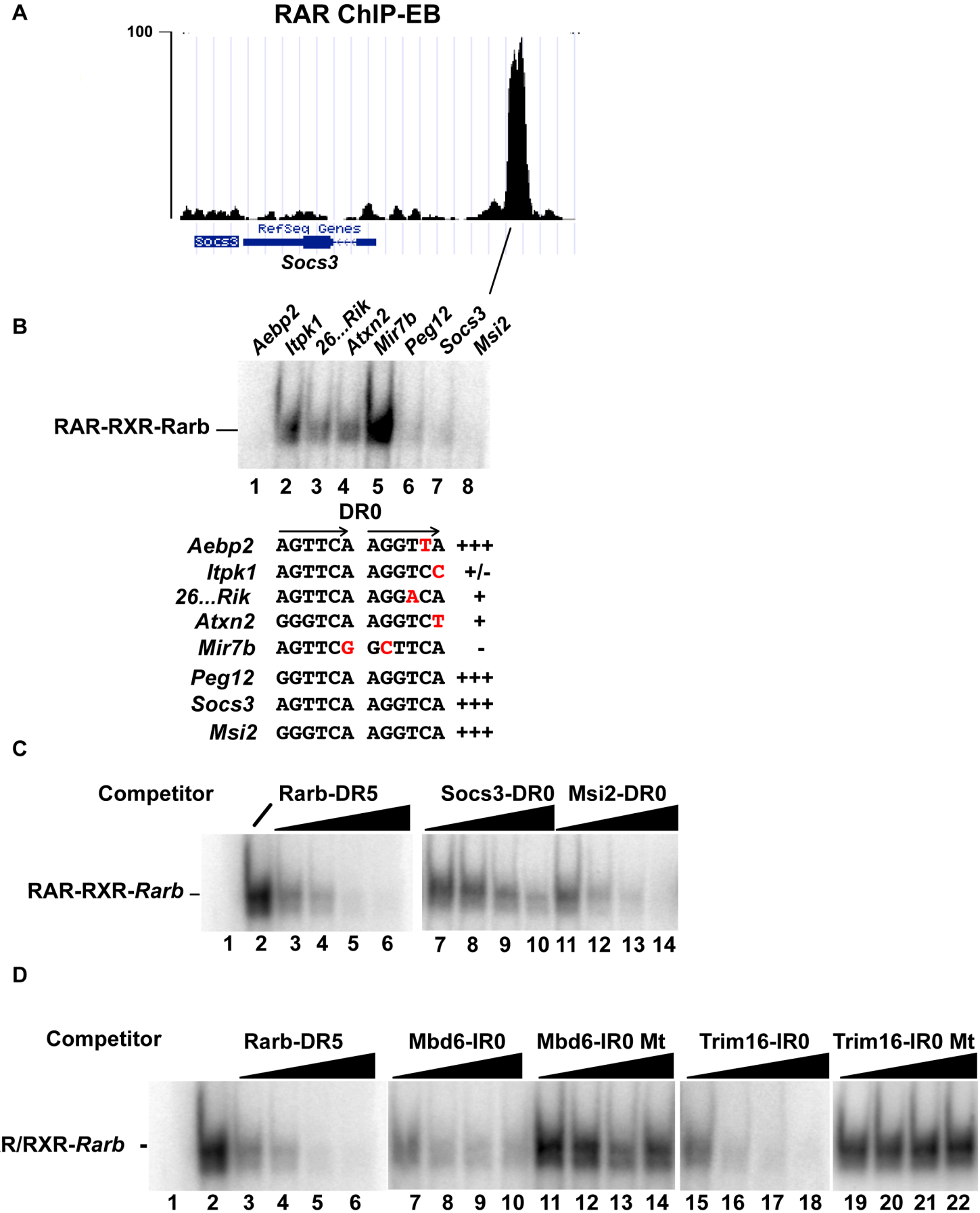
Figure 5. DR8, IR0, but not DR0 act as RAREs. **A**. Schematic representation of the reporter vector where three copies of the indicated motifs spaced by 12 nucleotides are inserted upstream of a TATA element and the CAT reporter gene. **B**. Relative CAT activity of the different reporter vectors 48 hours after transfection. EV is empty vector and the elements in the other vectors are shown below each lane. The DR5 is from the *Rarb* gene, DR0 is from the *Socs3* gene, the IR0 from the *Trim16*, gene, DR0 (P) is the pseudo DR0 motif from the *Hoxb13* locus, and DR8 (C) the composite DR8 element from the *Mafa* gene and DR8 the simple DR8 from the *Dedd* gene. **C**. Relative CAT activity from a second series of transfections showing the activity of the mutated composite *Mafa* DR8 element.

Figure 6. Identification of a pseudo DR0 motif. **A**. Sequence of the pseudo DR0 motif identified from the MEME analysis. **B**. UCSC view of sequence tag density in .wig file format of the RAR-occupied sites at the *Hoxb13* and *Wsb2* genes comprising the pseudo DR0 in embryoid bodies. **C**. EMSA competition analysis of the various indicated pseudo DR0 to compete for RAR/RXR complex formation. The sequences of the pseudo DR0 elements within the competing oligonucleotides are shown with the half sites indicated by arrows. Mutated nucleotides are indicated in red and the conserved G at position 2 of the half sites are boxed in blue. All competitors were used a 100-fold excess.

Figure 7. Association of RAR-occupied sites with gene expression. **A.** Venn diagramme representing the overlap between the total number of transcripts associated with RAR binding sites and those regulated by RA. **B-C.** DR and IR frequencies at the RAR-occupied sites associated with RA-induced and repressed transcripts, respectively. **D**. DR5/DR0 ratios in the indicated classes of RAR-occupied sites.

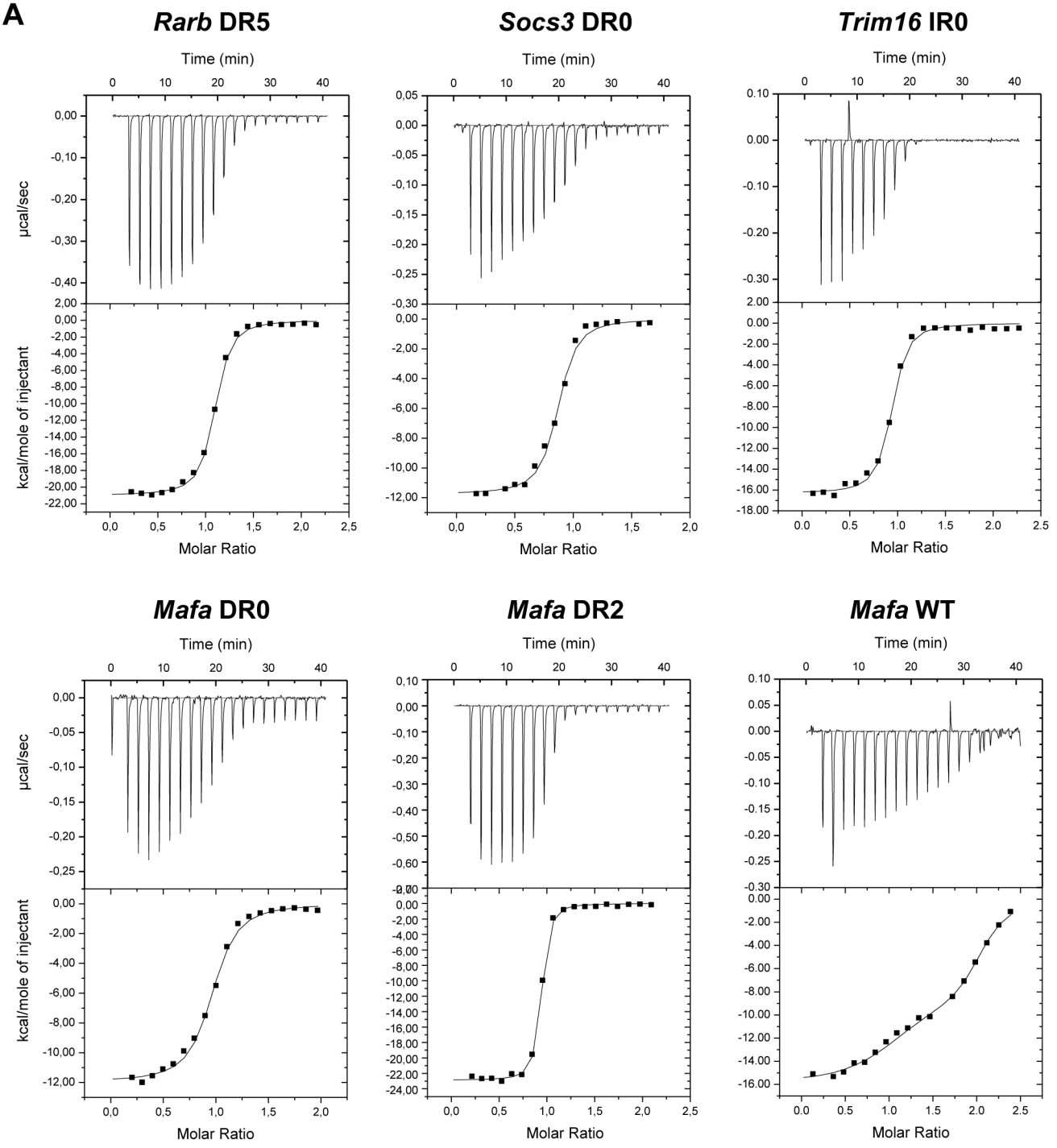


Moutier et al., Fig. 1



Moutier et al., Fig. 2

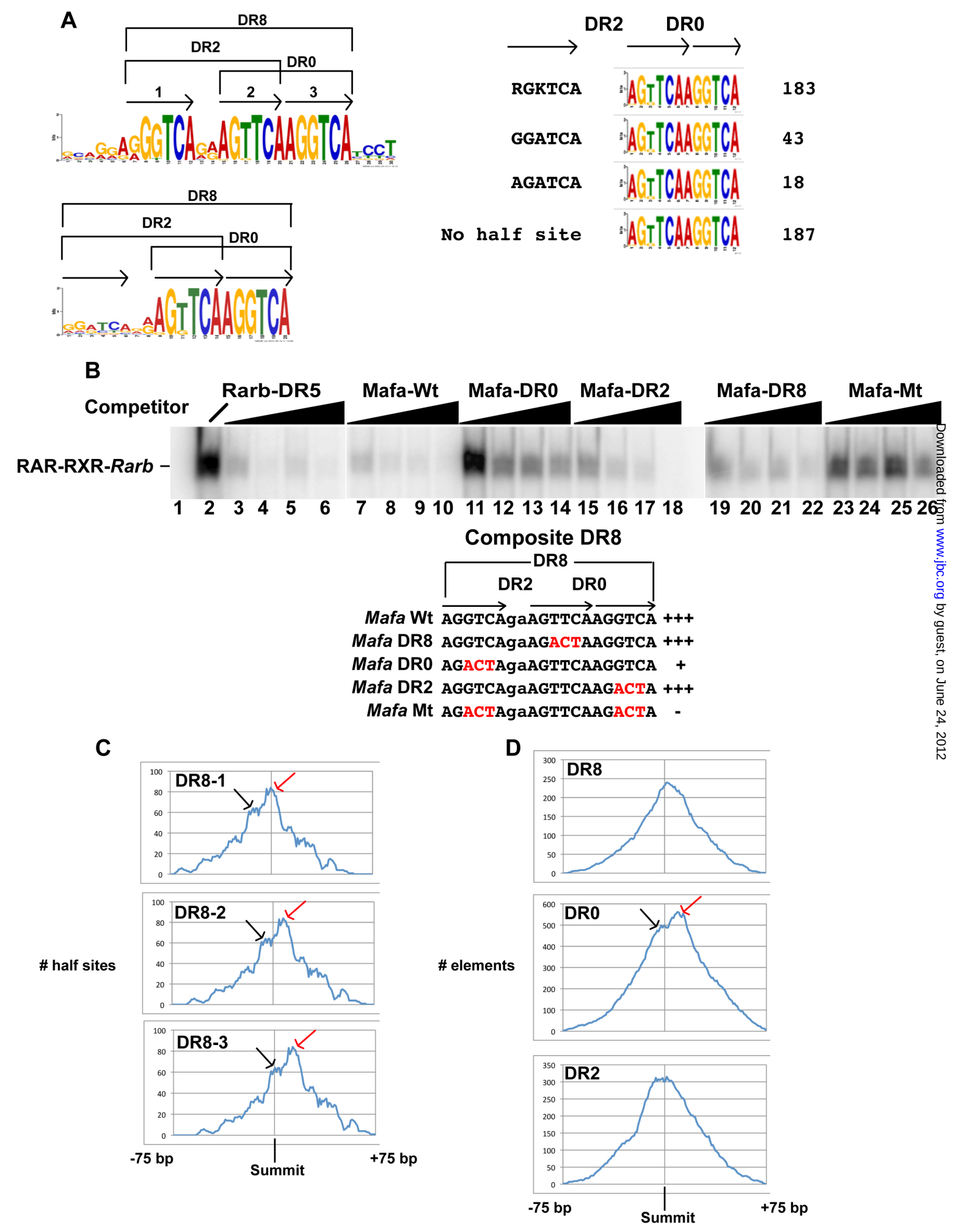
A



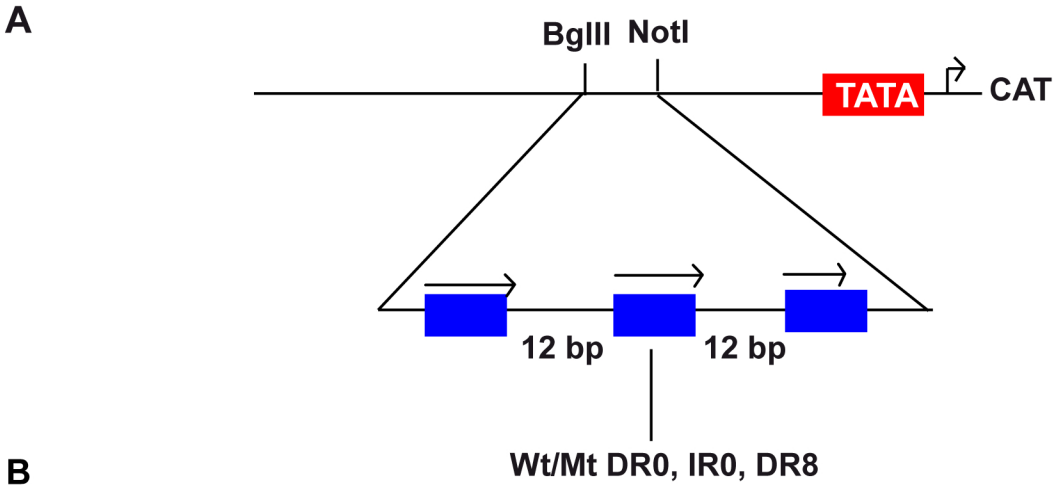
B

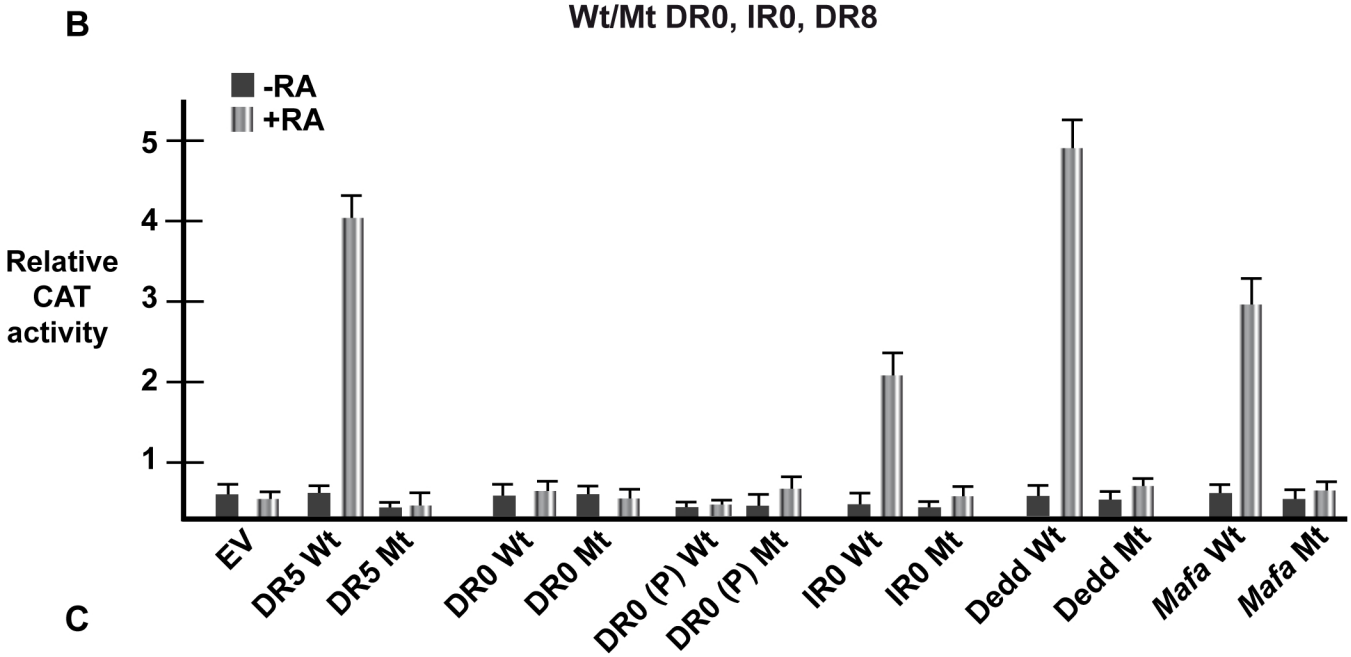
Element	Kd (nM)	ΔH (cal/mol)	-TΔS (cal/mol)
Rarb **DR5**	**73±8**	**-20970**	**947**
Hoxa10 **DR2**	**79±9**	**-19420**	**817**
Msi2 **DR0**	**76±13**	**-12140**	**205**
Socs3 **DR0**	**110±16**	**-11760**	**190**
Mafa **DR0**	**180±18**	**-11960**	**230**
Trim16 **IR0**	**58±8**	**-16300**	**540**
Vat1 **IR0**	**97±13**	**-14450**	**410**
Hoxb13 **(p)DR0**	**116±15**	**-8338**	**-92**
Rnd3 **(p)DR0**	**93±10**	**-13190**	**302**
Mafa **DR2**	**25±2**	**-22870**	**1050**
Mafa **WT**	**13 ± 2**	**-1544**	**395**
	180 ± 18	**-9500**	**24**

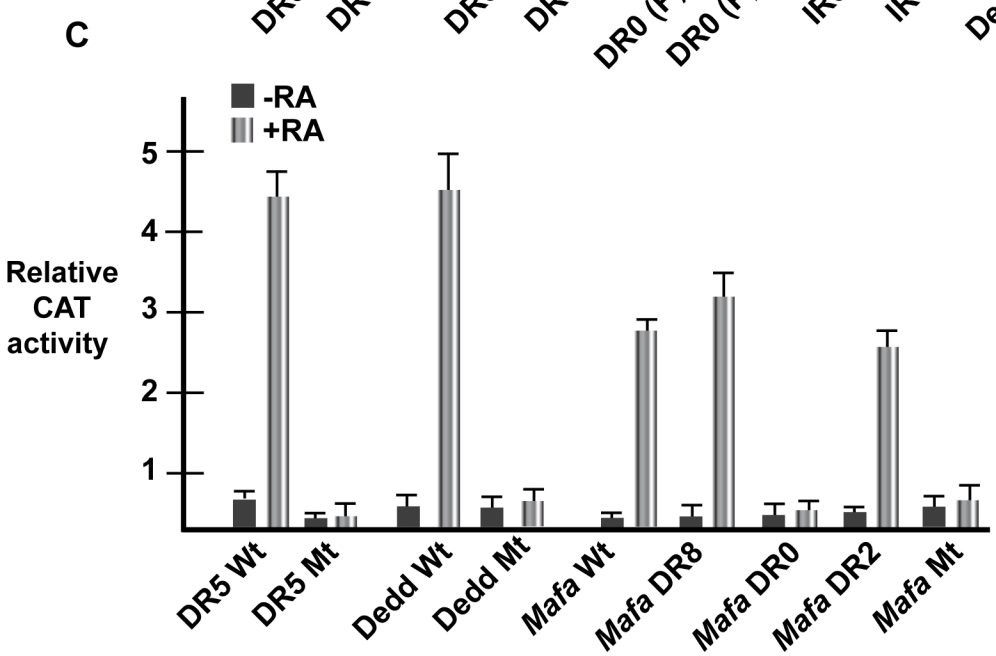
Moutier et al., Fig. 3

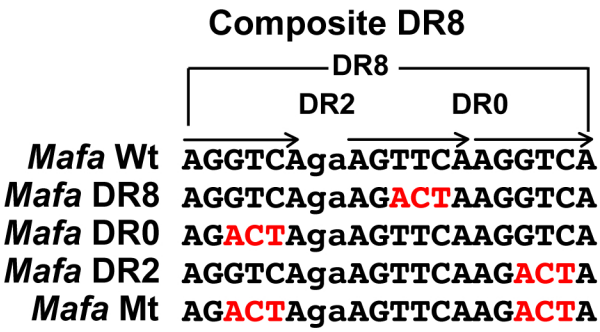


Moutier et al., Fig. 4

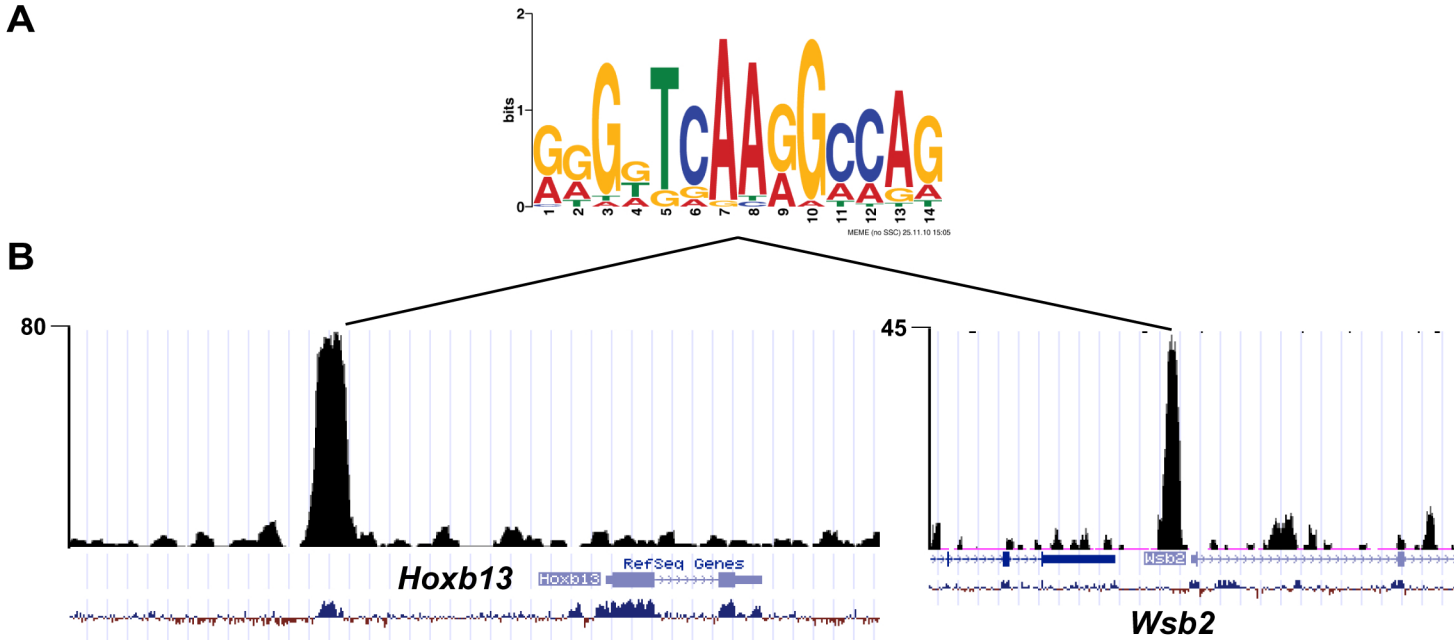


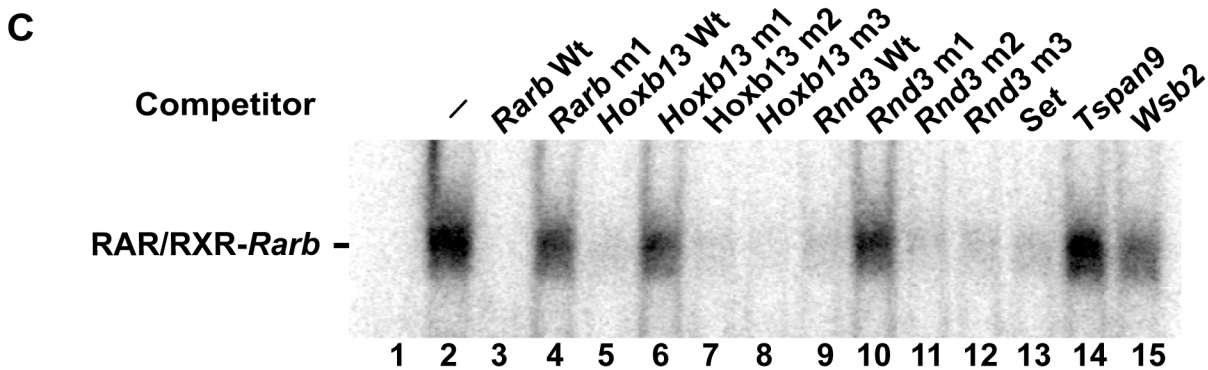






Moutier et al., Fig. 5



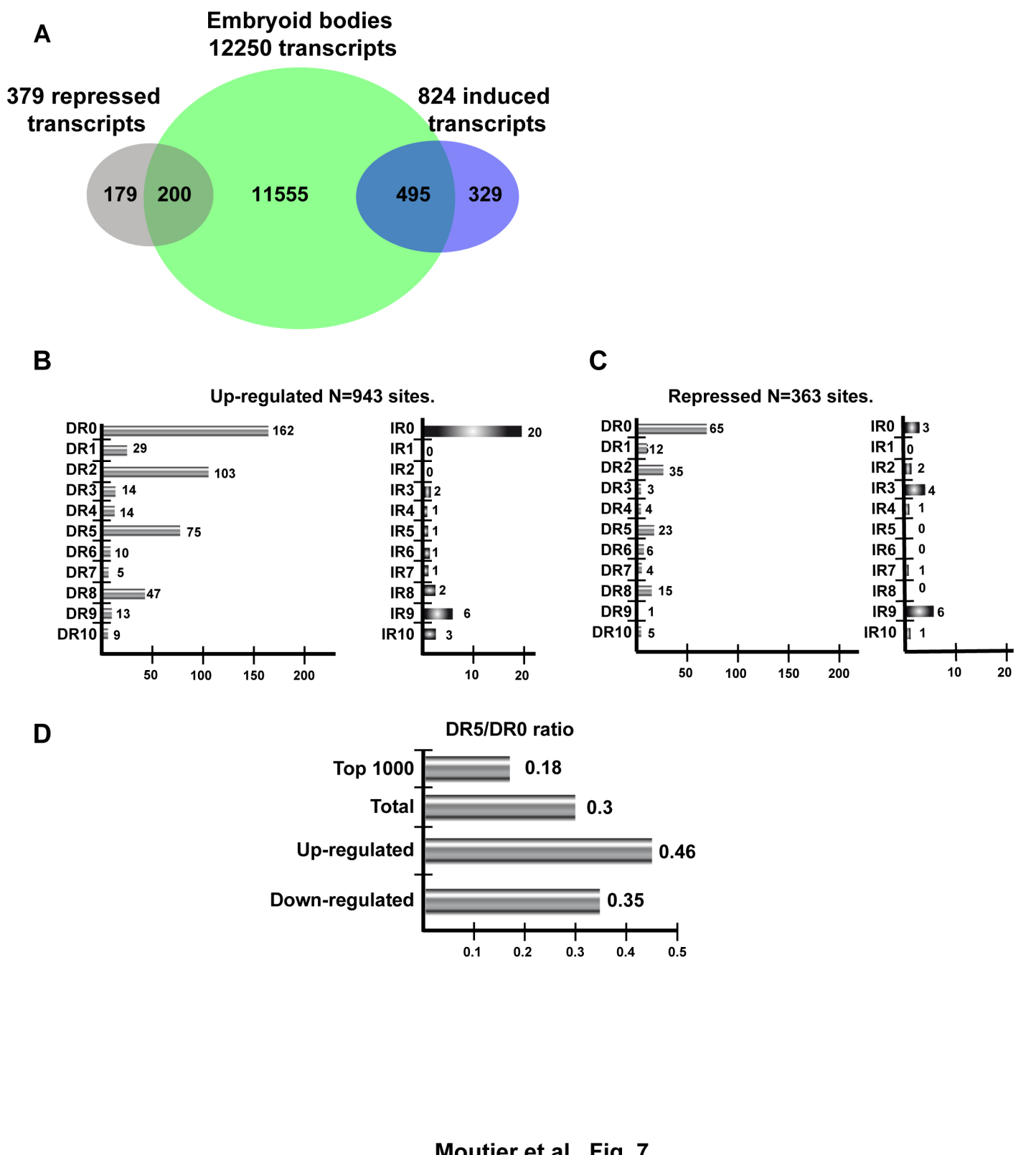


Hoxb13 Wt ggaAGGTCAAGGCCAagt +++
Hoxb13 m1 ggaAGGGATCGGCCAagt –
Hoxb13 m2 ggaAGGTCAAGTCCAagt +++
Hoxb13 m2 ggaAGGTCAAGGTCAagt +++

Rnd3 Wt tagGGTTCAAGGCCGggt +++
Rnd3 m1 tagGGGATCAGGCCGggt –
Rnd3 m2 tagGGTTCAAGTCCGggt +++
Rnd3 m3 tagGGTTCAAGGTCGggt +++

Set gagAGGTCAAGGCCAgcc +++
Tspan9 tagGAATCAAAGCAAgag –
Wsb2 ccgTTGTCAAGGCCAgca +

Moutier et al., Fig. 6



Moutier et al., Fig. 7