Gata6 and Hnf1ba are the major downstream effectors of retinoic acid for the specification of zebrafish pancreas.

Ana López-Pérez[1,2], Isabelle Manfroid[1], Marianne L. Voz[1], Bernard Peers[1,*].

1. Laboratory of Zebrafish Development and Disease Models (ZDDM), GIGA, Avenue de l'Hôpital 1, B34, 4000 SART TILMAN, University of Liège, Belgium
2. Present address: Umeå Centre for Molecular Medicine (UCMM), Umeå University, Sweden

*. Corresponding author: BP (bpeers@uliege.be)

Abstract

Retinoic acid (RA) is a key signaling molecule required for the specification of the pancreatic field within the endodermal germ layer. Still, the gene regulatory cascade triggered by RA in endoderm remains poorly characterized. In this study, we investigated the gene regulatory network induced by RA signaling in zebrafish endodermal cells by a combination of RNA-seq, RAR ChIP-seq and ATAC-seq experiments. By analysing the effect of RA and BMS439 on the transcriptome and on the chromatin accessibility of endodermal cells, we identified a large set of genes and regulatory regions regulated by RA signaling. Localization of RAR binding sites in the zebrafish genome by ChIP-seq highlighted the putative direct RAR target genes. Among them, Hnf1ba and Gata6, two known pancreatic regulatory factors activated by RA treatment, play a crucial role in opening chromatin at many genomic loci as revealed by the strong enrichment of their sequence binding motifs in RA-induced nucleosome-free regions. Furthermore, comparison of RAR ChIP-seq data obtained in zebrafish and in mice highlights the evolutionary-conserved direct targets, comprising the well-known *Cyp26a* or *Hox* genes but also *Hnf1b* and *Gata6*. Some RAR binding sites are located in highly conserved noncoding regions revealing the strong evolutionary constraint to maintain the function of such regulatory sequences. Among them, we identify a novel RA-induced enhancer located far upstream from the *Hoxb* Locus. In conclusion, our data reveal the central role of HNF1ba and Gata6 as pioneer

transcription factors for the RA-dependent specification of the pancreatic field and highlight the RAR sites conserved from fish to mammals.

Keywords: Pancreas, retinoic acid, zebrafish, RNA-seq, ChIP-seq, ATAC-seq

Introduction

The endoderm is the embryonic germ layer that gives rise, among other organs, to the pancreas. During gastrulation and early somitogenesis, dynamic movements bring the endoderm into proximity with mesodermal tissues that will secrete different factors to pattern the endoderm, including FGF, Wnt, RA, and BMP ligands. Retinoic acid (RA) is important for the anterior-posterior patterning of the endoderm (Bayha et al., 2009; Martín et al., 2005; Stafford et al., 2004). It is notably required for the specification of the dorsal pancreatic bud in mice, chick, Xenopus and zebrafish embryos (Bayha et al., 2009; Molotkov et al., 2005; Stafford and Prince, 2002; Zeynali and Dixon, 1998), and used in protocols for generating insulin secreting beta-like cells *in vitro* from embryonic stem cells (Schiesser and Wells, 2014). The role of RA has been previously studied mostly in the anterior-posterior regionalization of the nervous system (Diez del Corral et al., 2014; Maden, 2007, 2002; Rhinn and Dollé, 2012; Schubert et al., 2006). RA signaling acts through the binding of intracellular receptor and regulates the expression of various genes including the Hox family to regionalize the forebrain, midbrain, hindbrain, and spinal cord during embryogenesis (Schubert et al., 2006). The precise role of RA in the induction of pancreas is still largely unclear. Previous studies in chick and Xenopus embryos showed that RA alone is not sufficient to induce ectopic pancreatic cells like in zebrafish, however inhibition of RA signaling at earlier stages of development prevents pancreas formation (Bayha et al., 2009; Zeynali and Dixon, 1998).

RA is synthesized from vitamin A in a two-step reaction that involve the action of RDH10 and RALDH2 enzymes (Das et al., 2014; Feng et al., 2010; Kam et al., 2012; Kedishvili, 2016) Sandell et al., 2007). The levels of endogenous RA are also tightly controlled through its degradation

catalysed by CYP26A and DHRS3 enzymes (Kinkel et al., 2009; Kam et al., 2012; Feng et al. 2010). In vertebrates, *CYP26A, DHRS3* and *RALDH2* are directly regulated by RA itself, creating an auto-regulatory loop that keeps RA levels within a normal range in the embryo (Feng et al., 2010; Kam et al., 2012). RA binds the Retinoic Acid Receptors (RAR) that are forming heterodimers with the Retinoic X Receptors (RXR). These RAR/RXR heterodimers regulate gene expression through their binding to "Retinoic Acid Response Elements" (RAREs) located in promoters or enhancers. Genome-wide identification of RAR binding sites has been achieved by ChIP-seq experiments on cell lines using antibodies raised against RAR and RXR (Chatagnon et al., 2015; Delacroix et al., 2010; Mahony et al., 2011; Moutier et al., 2012). Such studies confirmed that the RAR/RXR heterodimers bind to direct repeats of the RGKTCA motif (R=A/G, K=G/T) often separated by 5 bases (DR5)(Dilworth and Chambon, 2001; Kam et al., 2013), but also present with other spacing and orientations (DR0, DR1, DR2, DR8, IR0, IR3, …) (Chatagnon et al., 2015; Moutier et al., 2012). These data uncovered thousands of RAR/RXR binding sites in the murine genome, some located near RA-induced genes like the *Hox*, *Cyp26*, *Dhrs3* or *Rara* genes (Das et al., 2014; Tanoury et al., 2013). However, no data is available on the effects that RA might have on chromatin accessibility and whether RAR binding sites correlates with active enhancers.

Two zebrafish mutants demonstrated the requirement of RA for pancreas development. The *neckless* and *non-fin* mutants, containing a mutation in the *raldh2* enzyme (Begemann et al., 2001; Grandel et al., 2002), lack pancreatic progenitor cells, while the endodermal cells are still present, indicating that RA action is not required in endoderm formation but rather in pancreas specification. Furthermore, wild type embryos treated with BMS493, a molecule that binds the RARs blocking their activities, show the same phenotype as the *raldh2* mutants confirming that the lack of pancreas is due to a lack of RAR activity (and not to a secondary role of Raldh2). Besides, addition of RA into the growing media of *raldh2* mutants at different time points showed that RA acts at the end of gastrulation to induce pancreas (Stafford et al., 2004). Furthermore, treatment of wild type zebrafish embryos with RA during gastrulation is sufficient to induce ectopic pancreatic endocrine cells towards the anterior

part of the endoderm (Stafford and Prince, 2002). Cell transplantation studies have shown that RA synthesized in the anterior paraxial mesoderm act directly on the foregut endoderm to promote pancreas development (Stafford et al., 2006). Although many RAR direct target genes are known, the exact role of RA in the endoderm is still unclear and the gene regulatory network induced by RA in the endoderm is still poorly defined. Experiments performed on Xenopus explants have recently indicated the involvement of *Hnf1b* and *Fzd4* (Gere-Becker et al., 2018); however, many other regulatory genes induced by RA and mediating its effect are still to be identified.

In this study, we used a combination of RNA-seq, ChIP-seq and ATAC-seq experiments to investigate the gene regulatory network induced by RA in zebrafish endodermal cells. We identified by RNA-seq experiments all genes regulated by RA and BMS493 treatments. Direct targets of RA were also identified by performing RAR ChIP-seq assays. Then, ATAC-seq experiments (Buenrostro et al., 2015) allowed us to identify all chromatin regions whose accessibility is modified by RA signaling. By integrating all these transcriptomic and genomic data, we build a model describing the action of RA in endodermal cells. Furthermore, comparison of our RAR sites detected in the zebrafish genome with those of the murine genome highlight the RAR target genes which have been conserved from fish to mammals and which are probably crucial in RA action.

Material and methods

Sample preparation and cell purification

Fish were maintained in accordance with the national guidelines and all animal experiments described herein were approved by the ethical committee of the University of Liege (protocol number 1980).

Endodermal and non-endodermal cells were obtained using the transgenic line Tg(*sox17*:GFP). Embryos from this line were incubated in DMSO, 1 µM RA or 1 µM BMS943 from 8 cells stage (1.25 hpf) to 3 somites stage (11 hpf) and grown at 23ºC. Embryos were dechorionated and then deyolked and dissociated in FACS solution. Finally, cells were sorted by FACS Aria II based on the GFP expression. Sorting was performed as a single run in purity mode.

RNA-seq library preparation and data analysis

cDNA was obtained using the SMARTer protocol with minor modifications. Briefly, cells were pelleted and lysed in Lysis Buffer by freezing in liquid nitrogen and stored at -80ºC. Synthesis of cDNA was performed directly on the lysed cells and the cDNA was amplified and purified before assessing its quality by the bioanalyzer (2100 high sensitivity DNA assay, Agilent Technologies). 150 pg. of cDNA were used as input to prepare the libraries using the Nextera XT DNA kit (Illumina). 75 bp single-end sequences were obtained using the NextSeq500 Illumina Sequencer with coverage of 20 million reads per library.

Raw reads were aligned to the zebrafish genome (Zv9, Ensembl genome version 79, ensembl.org) using STAR (Dobin et al., 2013). Genes with at least 1 read in at least 2 samples were kept for further analysis. Normalization and differential expression analysis were performed using DESeq2 (Love et al., 2014). Genes were considered differentially expressed with FDR < 0.01 (False Discovery Rate).

Synthesis of RARaa-myc

The whole zebrafish RARaa coding sequence was cloned into the pCS2MT vector that contains a myc-tag and mRNA was produced from this new construct by *in vitro* transcription (mMESSAGE mMACHINE sp6 transcription kit, Invitrogen). Validation of the tagged RARaa mRNA was done by western-blot using the ChIP-graded anti-myc antibody (ab9132, Abcam) in the cytoplasm and nuclei fractions of injected and non-injected embryos.

ChIP-seq library preparation and data analysis

Zebrafish fertilized eggs were injected with the myc-tagged RARaa mRNA at a concentration of 72 ng/µl and incubated until the bud stage was reached. RA acid was added to the media at a final concentration of 1 µM and embryos were grown for an extra hour. Around 2000 injected embryos were fixed for ChIP-seq analysis. A Diagenode Bioruptor sonicator was used for sonicating the chromatin. Dynal Protein A magnetic beads (Diagenode) and ChIP-graded anti-myc antibody (ab9132, Abcam) were used to precipitate the chromatin. Libraries were prepared with the NEBNext

Ultra II DNA Library Prep kit (Bioke). 42 bp pair-end sequences were obtained using the NextSeq500 Illumina Sequencer with coverage of 60 million reads per library.

Raw reads were mapped to the zebrafish genome (Zv9) using bowtie2. Enriched peaks were called using MACS2. Motif enrichment analysis were performed using HOMER adding a motif length from 6 to 18 bp, to ensure the inclusion of the different RARE. Annotation of peaks was done using ChIPseeker.

ATAC-seq library preparation and data analysis

After isolation by FACS, cells were pelleted and ATAC-Seq libraries were prepared as described elsewhere (Buenrostro et al., 2015). Libraries were sequenced 42 bp paired end using the NextSeq500 Illumina Sequencer with coverage of 40 million reads per library.

Raw reads were mapped to the zebrafish genome (Zv9) using bowtie2 (Langmead and Salzberg, 2012). Histone modification peaks were obtained from available datasets (GEO accession number GSE48254, (Paik et al., 2013)). Peaks were obtained by MACS2 (Zhang et al., 2008). For zCNEs, we downloaded previously described elements (Hiller et al., 2013).

Quality assessment of ATAC-seq libraries was performed with ATACseqQC (Ou et al., 2018) and ChIPpeakAnno (Zhu et al., 2010) was used to create the density and heatmap plots. Enriched peaks for ATAC-seq datasets were called using MACS2. Peaks showing different signal among ATAC-seq samples were identified with DiffBind (Stark and Brown, n.d.). Briefly, the density of mapped reads was calculated for each of the 156604 regions obtained by merging the peaks obtained for each of the 12 individual samples and differential analysis was performed in these regions to determined which are the regions with different open chromatin in each condition. Motif enrichment analysis for each set of treatment specific peaks were performed using HOMER (Heinz et al., 2010). Annotation of peaks was done using the ChIPseeker (Yu et al., 2015).

Data integration and visualization

To study the correlation of ATAC-seq/ChIP-seq and RNA-seq a student's t-test was performed to compare the read counts in RNA-seq of each gene and the transcription starting site (TSS) accessibility in ATAC-seq/ChIP-seq. Correlation of flanking treatment/cell specific and common ATAC-seq elements, as well as, ChIP-seq peaks and log2 fold change from RNA-seq was analysed by Spearman's rank correlation.

Results

Retinoic acid affects the transcriptome of zebrafish endodermal cells

Previous studies have shown that treating zebrafish embryos with 1 µM RA during gastrulation (from 5 to 10 hpf) induces ectopic pancreatic cells toward the anterior part of the embryo (Stafford and Prince, 2002). Inversely, blocking the RA action during this period, e.g. by adding the RA antagonist BMS493, blocks pancreas formation. In order to determine the effect of RA on the transcriptome of zebrafish endodermal cells, transgenic Tg(*sox17*:GFP) zebrafish embryos were treated during gastrulation either with 1µM RA, with 1µM BMS493 (RA-antagonist) or with DMSO (used as control) and endodermal GFP+ cells were next selected by FACS from embryos both at 3-somites (3-S) and 8-somites (8-S) stages (11-hpf and 13.5-hpf, respectively). Non-endodermal (GFP-) cells were also selected from the DMSO-treated control embryos in order to identify genes displaying endodermal enriched expression. RNA-seq was performed on all these FACS-isolated cells prepared in triplicates (24 samples in total) and transcriptomes were analysed using the bioinformatic pipelines as described in Methods. Principal component analysis of all RNA-seq data presented in Figure 1A indicates i) a tight clustering of all triplicate samples confirming a high reproducibility, ii) a strong difference between the transcriptome of endodermal and non-endodermal cells (discrimination along the first axis of the PCA plot), iii) relative similar transcriptomes of cells isolated at 3-S and 8-S stages, and iv) the clustering of BMS493 samples near DMSO samples indicates a much slighter

effect of BMS493 treatments compared to the RA treatments. These conclusions were further confirmed by the differential gene expression analyses between the different conditions : more differentially expressed genes were identified between endodermal and non-endodermal cells (1370 and 1410 differentially expressed genes at 3-S and 8-S stages, respectively) than between endodermal cells treated with RA versus DMSO (756 and 514 RA-regulated genes at 3-S and 8-S stages respectively) or with BMS493 versus DMSO (32 and 71 BMS493-regulated genes at 3-S and 8-S stages respectively). We found that there is a large overlap among the sets of genes having an endodermal-enriched expression at both 3-S and at 8-S stages (Supplementary Figure S1A; list of genes given in Supplementary Table S1 and S2) and these sets include all known endodermal markers including *sox17*, *gata6* and *foxa1/2/3*, validating the accurate sorting of endodermal cells. RA-regulated genes consist of a large set of up- and down-regulated genes (Supplementary Table S3 and S4), many of them being regulated at both 3-S and 8-S stages (Supplementary Figure S1B). These RA-regulated genes contain many known RAR-direct targets such as *cyp26b1/a1*, *dhrs3a* and several *hox* genes, validating the efficiency of RA-response. Interestingly, BMS493 treatment led mostly to a down-regulation of genes: at 3-S stage, all the 32 BMS493-regulated genes were repressed, and, at 8-S stage, 68 genes were repressed while only 3 genes were up-regulated by the BMS493 treatment (Supplementary Tables S5 and S6). A large overlap is also observed between the genes down-regulated by BMS493 at 3-S and at 8-S stage (Supplementary Figure S1C). As expected, a large proportion of genes down-regulated by BMS493 were up-regulated by RA treatments, this observation being evident mostly at 3-S stage (see Figures 1B and C). Tables 1 and 2 show the genes upregulated by RA and down regulated by BMS493 as well as those enriched in the endoderm at 3- and 8-somites stage respectively (highlighted in green). *gata6, insm1a* and *ascl1b* are the only 3 known pancreatic regulatory genes which are regulated by RA and BMS493 (Table 1). Other pancreatic transcription factors, like *mnx1*, *insm1b*, *hnf1ba*, *nr5a2, jag2* or *neuroD1*, were induced by the RA-treatment but were not significantly repressed by BMS493, and inversely, other pancreatic regulators are inhibited by BMS493 but not significantly induced by RA like *pdx1*, *rfx6* and *myt1b*

(Supplementary Tables S3 and S6). We can assume that the induction of pancreatic fate by RA is mediated, at least in part, by the direct or indirect regulation of these pancreatic regulatory factors. In conclusion, the analysis of all these RNA-seq data highlight all the genes with an enriched expression in zebrafish endodermal cells and which are regulated by the RA signaling pathways; this gene set should comprise regulators involved in the AP patterning of the endoderm and in the specification of the pancreatic field.

Genes directly regulated by RA have RAR binding sites near them

To further identify the genes directly regulated by RAR and determine if some pancreatic regulatory genes are direct targets of RA signaling, we performed ChIP-seq experiment. Since no ChIP grade RAR antibodies have been described for zebrafish to date, a tagged RARaa was expressed in zebrafish gastrulae by injecting zebrafish fertilized eggs with a mRNA coding for the zebrafish RARaa protein fused to a myc-tag at its C-terminal end. RARaa was chosen as the RNA-seq data indicate that it is the most highly expressed RA receptor in zebrafish endodermal cells. Injection of this mRNA did not disturb the development of embryos. Chromatin was prepared from about 2000 injected zebrafish embryos at 10hpf (end of gastrulation) and immunoprecipitation was performed with ChIP grade Myc antibody. Comparison of reads obtained with myc-RAR ChIP and input control led to the identification of 4858 RAR peaks. In order to identify *bona fide* RAR binding sites showing strong affinity, we selected all peaks with a height score above 50 (Supplementary Table S7). By choosing such criteria, 2848 RAR robust binding sites were identified. As shown in Figure 2A, a majority of these sites are located near or within genes: 8% in gene promoters (i.e. 1kb upstream gene TSS), 30% in upstream sequence (from 1 to 10 Kb), 22% in introns and 4% in exons while only 33% are intergenic. Sequence analysis of all RAR peaks revealed that the highest represented motif corresponds to the DR5 RAR/RXR consensus binding sequence (present in 39% of identified RAR peaks) (Figure 2B). Furthermore, several RAR peaks were observed around many genes known to be RAR direct target genes such as *cyp26a1*, *dhrs3* or the *hoxb1a-hoxb4a* genomic region (Figure 2C and data not shown). All these observations confirm the accuracy of the ChIP-seq data. Interestingly,

many identified zebrafish RAR sites are located in evolutionary conserved genomic sequences as shown by the fish PhastCons track of Figure 2C (and see below).

The RAR ChIP-seq peaks which were located less than 250 Kb from a gene TSS were assigned with the name of that gene and when several genes were lying in vicinity of a RAR site, the closest gene was considered as the putative RAR-regulated genes. Using that strategy, of the 2848 RAR sites, 2144 were linked to a gene. From this gene set, 94 were up-regulated by RA at 3-S stage and 61 were down-regulated (Figure 3B, Supplementary Tables S8 and S9). If we analyse the correlation of RA gene regulation and presence of RAR sites by plotting the fold change in gene expression induced by RA according to the number of RAR sites near the gene (Figure 3A), we observe that RAR acts mainly as a transcriptional activator. This analysis supports the classical model where the RAR/RXR heterodimer recruits co-activators upon RA ligand binding. Interestingly, a large part of the genes (i.e. 18 genes out of 32) down regulated by BMS493 treatment at 3S stage contain RAR sites (Figure 3B). Among the genes up-regulated by RA and harbouring a RAR site, we identified the pancreatic regulatory genes *hnf1ba/b, gata6, insm1b, jag2* and *mnx1* suggesting that these genes are direct target of RA.

RA affects the chromatin structure in endodermal cells

To determine whether the chromatin accessibility of endodermal cells is modified by RA or BMS493 treatment, we performed ATAC-seq experiments, a very sensitive technique allowing the identification of open chromatin and nucleosome-free regions (Buenrostro et al., 2013; Quillien et al., 2017). Furthermore, as most nucleosome-free regions corresponds to regulatory sequences, ATAC-seq can locate the enhancers or promoters whose accessibility is modified by the RA and BMS493 treatments. By this way, we can also assess whether the identified RAR binding sites are located in nucleosome-free regions and become more accessible after RA treatment. This assay can also identify enhancer regions, not bound by RAR, but which become more accessible upon RA treatment, thereby identifying regulatory regions indirectly activated by RA. The sequence analysis of such indirect RA targets can give clues on the transcription factors mediating the RA effect. Like

for the RNA-seq analyses, zebrafish embryos were treated with RA, BMS493 and DMSO (control) during gastrulation and about 10000 endodermal cells were selected by the FACS at 3-somites stage (11hpf). Non-endodermal cells from control DMSO-treated embryos were also analysed in parallel. Cell preparations and ATAC-seq were done in triplicates for each condition and analysed as described in Methods. To verify the accuracy of the ATAC-seq data, we performed several quality controls analyses. First, for all samples, the analysis of the ATAC-seq fragment size distribution reveals the expected pattern with abundant short (<150 bp) fragments corresponding to nucleosome-free regions and larger fragments of about 200 and 400 bp corresponding to mono- and bi-nucleosome regions, respectively (Supplementary Figure S2A). Secondly, as reported by Buenrostro and collaborators, genome mapping of the nucleosome-free fragments shows a clear enrichment in promoter regions immediately upstream of transcriptional start sites (TSSs) while mono-nucleosomes are depleted from TSSs and rather map just downstream of the TSSs in a periodic manner (Supplementary Figure S2B). Thirdly, we verified that the ATAC-seq fragments highlight zebrafish regulatory regions by comparing them with regions harbouring the histone modifications H3K4me3 and H3K27ac associated to promoters and enhancers which were identified from whole zebrafish embryos at 10 hpf (end of gastrulation, "bud stage") (Paik et al., 2013). Heat-maps of ATAC-seq reads from all samples show an obvious enrichment at loci harbouring these two histone modifications (Figure 4A and Supplementary Figure S3). As regulatory regions often display sequence conservation, we also compared our ATAC-seq reads to a collection of annotated zebrafish evolutionary-conserved non-coding elements (zCNEs) identified using multiple fish and tetrapod genomes (Hiller et al., 2013). Heat-maps of ATAC-seq reads from each sample also showed a strong correlation with zCNEs (Figure 4A, Supplementary Figure S3). These observations confirm that regions identified by ATAC-seq exhibited hallmarks of active regulatory elements. We next verified the reproducibility of ATAC-seq peaks by a principle component analysis (Figure 4B). The results show that i) triplicate samples are tightly clustered, and ii) endodermal and non-endodermal (NE) cell clusters are separated along the PC2 axis, representing 3% of the variability among data, while RA-treated cluster is separated

from the DMSO- and BMS-cell cluster along the PC3 axis representing only 1% of variability. So, as observed for the RNA-seq data, stronger differences are observed between GFP+ and GFP- cells compared to the differences between RA-treated and DMSO-treated endodermal cells. The samples corresponding to the BMS493-treated cells and DMSO-treated cells were not segregated, suggesting no significant effect of BMS493 on the nucleosome distribution. These conclusions were supported by the differential peak intensities analyses as described hereafter.

From the 12 samples analysed altogether, 156604 nucleosome-free regions were identified in the zebrafish genome. The density of reads was next calculated for all identified peaks in each cell sample in order to identify the elements which become accessible upon treatment (RA, BMS493 versus DMSO) or according to cell identity (GFP+ versus GFP-) with an FDR<0.05. This analysis revealed more differences in chromatin accessibility between cell types than between treatments. Indeed, among the 156604 identified ATAC-seq peaks, 22696 display significant differences in intensities between endodermal and non-endodermal cells : 9722 and 12974 regions are more accessible in endodermal and non-endodermal cells, respectively (Supplementary Tables S10 and S11), while only 1989 display differences between RA and DMSO treatments (1240 peaks have significant more reads upon RA treatment and 749 peaks have less reads on RA treatment, Supplementary Tables S12 and S13). No significant differences could be identified between the BMS493 samples and the DMSO controls. Interestingly, sequence analysis of all endodermal-specific ATAC-seq regions revealed a highly significant enrichment of the binding sites motif for the Gata, Fox and Sox protein families (Figure 4B), which can be explained by the action of the well-known endodermal regulatory factors Gata4/5/6, Foxa1/2/3 and Sox32/17. Regulatory regions opened specifically in endodermal cells were often identified near endoderm and pancreatic regulatory genes, such as in the *foxa2*, *nkx6.1*, *hnf4g, sox17* or *mnx1* loci (Supplementary Figure S4 and data not shown).

Annotation of the RA-induced ATAC-seq elements to the closest gene revealed that they are often located near RA-upregulated genes identified above by RNA-seq. We found that there is a significant correlation between the level of RA-induced gene expression and the number of RA-induced ATAC-

seq elements (Figure 4E). Interestingly, about 10% (136 regions) of RA-induced ATAC-seq peaks contain RAR binding sites as identified above by ChIP-seq and 101 regions (~74%) possess the DR5 motif recognized by the RAR/RXR complex (Figure 5D). This is the case for RA-induced peaks in the *dhrs3a*, *cyp26a1*/*b1* and *insm1b* genes (Supplementary Figure S5 and data not shown). In contrast, about 90% of RA-induced ATAC-seq peaks do not harbour RAR binding sites although they are usually found near RA-upregulated genes (Supplementary Table S14). Such genes are presumably indirect targets of RA signaling. For example, the pancreatic regulatory genes *pdx1* and *insm1a*, both regulated by RA signaling but not harbouring RAR binding site in their genomic loci, have both one RA-induced ATAC-seq peak located about 50 Kb upstream their TSS (see Supplementary Figure S6). Motif analysis of all RA-induced elements showed a significant enrichment for Gata and Hnf1b binding motifs, present in ~38 and ~13% of the RA-induced elements respectively (Figure 4B). To identify which GATA and Hnf1b proteins are opening these chromatin regions, we searched in the list of the direct RAR-target genes (Supplementary Table S8) those coding for members of these transcription factor families. *hnf1ba* and *hnf1bb* are both induced by RA, however the level of expression of *hnf1ba* is about 200-fold higher than *hnf1bb* in endodermal cells. Furthermore, *hnf1ba* contains three RAR binding sites located in nucleosome-free and evolutionary-conserved regions (Figure 5), while *hnf1bb* has only one RAR site of weaker affinity (data not shown). As for the Gata family, *gata6* is strongly induced by RA and has a high affinity RAR binding site in a nucleosome-free region located about 30 kb upstream from its TSS (Figure 5). Thus, all these analyses strongly suggest that Gata6 and Hnf1ba are the major effector of the RA signaling in zebrafish. As these two transcription factors are known to be crucial for pancreas development (Carrasco et al., 2012; Xuan et al., 2012), it indicates that they are the key effectors of RA action to induce the pancreatic field.

Some RAR genomic binding sites are well conserved among vertebrate species.

As many zebrafish RAR sites are located in evolutionary conserved genomic regions (Figure 2C and 5), we next determined the proportion of RAR sites which are also found in mammalian genomes. Chatagnon et al. published the chromatin binding of both RAR and RXR proteins in the murine F9

embryonal carcinoma cells during differentiation into primitive endodermal cells after induction with RA (Chatagnon et al., 2015). Comparison of our datasets with the one obtained by Chatagnon et al. showed that, among the 2144 zebrafish genes harbouring a RAR site, 722 have also a RAR site in the murine orthologous gene. This list of conserved RAR-bound genes comprises notably the *cyp26a1/b1*, *dhrs3a*, several *hox* genes, *raraa/b* as well as the pancreatic *gata6* and *hnf1ba/b* genes. We next determined whether the RAR binding sites are themselves conserved. For that, we determine which RAR sites are located in highly conserved non coding elements (HCNEs) identified through the comparison of multiple vertebrate genomes (Engström et al., 2008). This analysis revealed that 235 HCNEs regions carry RAR binding sites, of which 116 are zebrafish and 146 murine RAR binding sites. Interestingly, in 27 of them, the location and the binding motifs of these RAR sites are conserved within these HCNE (see Figure 6 and data not shown) revealing that the organization of the enhancer has been maintained throughout vertebrate evolution. This confirms that the RA regulation of a subset of genes have been well conserved throughout vertebrate's evolution. Among these well conserved RAR sites, we have identified an extremely well conserved enhancer located in the last intron of *skap1* gene (Figure 4) which is located far downstream from the *hoxba* locus and which could act at long distance to regulate not only *skap1* gene but also several *Hoxba* genes.

Discussion

It is known that RA is required for the proper development of the pancreas as well as for the anterior-posterior regionalization of the endoderm (Chen et al., 2004; Molotkov et al., 2005), but the mechanism by which it acts on the endoderm to set up the anterior-posterior axis and to induce pancreas are not well known. Here, we have deciphered the genome-wide transcriptional mechanisms of RA action in the endodermal cells during specification of the pancreas. To this end, we have determined the transcriptome and the nucleosome-free regions of endodermal cells collected from embryos that were treated with DSMO, RA and BMS493, and the binding sites of RARaa in whole embryos.

RA affected much more the transcriptome of endodermal cells than BMS493 treatment; furthermore, while the genes affected by RA were either up or down regulated, the ones affected by BMS493 were mostly down regulated. The fact that much fewer genes are affected by the loss of function (BMS493 treatment) compared to the gain of function (RA treatment) could be explained by the number of endodermal cells affected by these 2 treatments: RA treatment acts on a large proportion of the endoderm (anterior endoderm) while the BMS493 treatment acts on a small set of endodermal cells (posterior foregut) where the endogenous RA is active. The observation that most of the BMS493-affected genes are down-regulated strongly suggests that the RARs act as transcriptional activators. It is thus possible that the genes down regulated by RA corresponds to indirect targets. 23 genes were up-regulated by RA and down-regulated by BMS493 at 3 somites stages and 27 at 8 somites stages. Among them, some were already known to be regulated by RA and involved in its synthesis or signalling, such as *cyp26b1*, *cyp26a1*, *dhrs3a,* and *raraa*. Besides, we found some that are expressed during pancreas development, such as *gata6, hnf1ba,* and *nr5a2*.

When the location of the genomic RARaa binding sites is analysed and compared to the list of RA-regulated genes, it is obvious that only a minority of RA-regulated genes have RARaa binding sites around their loci; we can propose that a large part of these genes are the actual direct RARaa target genes, the other genes probably corresponding to indirect RAR targets. Also, it is interesting to notice that the majority of RARaa binding sites could not be linked to nearby RA-regulated genes (see Figure 3B). This observation raises the question of the function of such RARaa binding sites. Several hypotheses can be proposed about their putative role: 1) some of these sites could be functional but the regulation of the nearby genes by RA did not reach the statistical threshold (FC>2 fold, FDR<0.01); 2) some of these RAR sites are involved in the regulation of genes located at much longer distances (>250 kb); 3) some of these RAR sites are involved in the RA-regulation of nearby genes but in other tissues or at later developmental stages; or 4) some of these RAR are not functional and do not participate in the transcriptional regulation by RA.

Comparison of the zebrafish RARaa binding sites with the murine RAR ones showed that sequence where the RARs bind is conserved among species. Besides, many orthologous genes have binding sites for the RAR in both species, indicating that the regulation of those genes by RA is conserved among species. However, when we compared the binding sites location by using the HCNE, we could not detect a high overlap. Only a small proportion of genes have their RAR binding sites within HCNE. Furthermore, when RAR binding sites are found near the same orthologous murine and zebrafish gene, these RAR sites are not often located at the same place near this gene, being located upstream in one species and in introns or downstream in the other. Thus, this suggests high plasticity or remodelling during vertebrate evolution allowing relocation of RARE around genes. These analyses suggest that although genes regulated by RA are conserved between both species, the binding site for this receptor is often not located in the same conserved regions, only 23 genes harbouring their RARE in the same HCNE in mouse and zebrafish.

Moreover, by comparing chromatin accessibility upon RA and BMS493 treatment, we characterized regions of the chromatin that are either induced or blocked by RA. However, we could not detect any region of the chromatin affected by BMS493. The regions modified by RA are mainly located in introns, regions downstream of the TSS and distal intergenic regions of the chromatin, which indicates that they might correspond to enhancers. Furthermore, the correlation of the regions induced by RA with gene expression showed that genes up-regulated by RA have several regions induced by RA around them. Although, these regions generally do not correspond to their TSS, supporting again the hypothesis that the regions induced by RA are mainly enhancers. Indeed, we have identified the motifs enriched in these regions and they correspond to the binding sites of three known TFs: RAR:RXR, Gata6, and Hnf1b. Interestingly, the genes coding for these TFs are also up-regulated by RA and two of them, Gata6 and Hnf1b, are involved in pancreas development (Carrasco et al., 2012; Decker et al., 2006; Gere-Becker et al., 2018, p.; Lancman et al., 2013).

Finally, by integrating all these data, a model can be proposed were Gata6 and Hnf1ba TFs are the key mediators of RA action in endodermal cells to induce pancreas. Previous studies in different

species have shown that both of them are involved in pancreas development (Gere-Becker et al., 2018; Holtzinger and Evans, 2005; Lancman et al., 2013). A zebrafish mutant line where the *hnf1ba* gene is inactive has been identified and display severe defects in pancreas development (Sun and Hopkins, 2001). Concerning *gata6*, no zebrafish mutant line has been described so far but a loss of function experiments has been reported using morpholinos and indicate *gata4/6* involvement in the liver as well as exocrine pancreas development, although the phenotype analyses were quite limited (Holtzinger and Evans, 2005). Our data suggest that both factors, Gata6 and Hnf1b, should act as pioneer factors by opening the chromatin near target genes, as we confirmed for the RAR (data not shown). The pioneer activity of Gata6 has been previously reported during mouse liver development (Iwafuchi-Doi and Zaret, 2014; Zaret and Carroll, 2011), however, Hnf1b activity as a pioneer factor has not yet been reported. Furthermore, *hnf1ba* has been identified as a key mediator of RA action in the zebrafish hindbrain (Ghosh et al., 2018), indicating that at least one part of RA action seems very similar in the endoderm and hindbrain. This is also confirmed by the GO analysis performed for the 355 genes up-regulated by RA, where some of the terms were associated with hindbrain, brain and head development (Data not shown).

Declarations

All animal experiments were conducted according to national guidelines and were approved by the ethical committee of the University of Liège (protocol numbers 1980). The raw datasets generated in the current study will be deposited on ENA (in progress). The authors declare that they have no competing interests.

Acknowledgments

We are grateful for Fiona Wardle for the zebrafish transgenic line Tg (sox17:GFP). We thank the following technical platforms: GIGA-Zebrafish (H Pendeville), GIGA-Cell Imaging and Flow Cytometry platform (S Ormenese and S Raafat), GIGA-Genotranscriptomic (W Coppieters, and L Karim).

ALP was supported by ZENCODE-ITN European project 643062 and the Léon Fredericq fund. BP, IM, and MLV are associate researchers from FRS/FNRS (Fonds National pour la Recherche Scientifique).

Authors' contributions

ALP carried out the RNA-seq, ChIP-seq and ATAC-seq experiments and performed the bioinformatic analyses. BP conceived the study. BP, ALP, MLV and IM participated in the interpretation of the data and wrote the manuscript. All authors read and approved the final manuscript.

Bibliography

Bayha, E., Jørgensen, M.C., Serup, P., Grapin-Botton, A., 2009. Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis. PLoS ONE 4, e5845. https://doi.org/10.1371/journal.pone.0005845

Begemann, G., Schilling, T.F., Rauch, G.-J., Geisler, R., Ingham, P.W., 2001. The zebrafish neckless mutation reveals a requirement for raldh2 in mesodermal signals that pattern the hindbrain. Development 128, 3081–3094.

Buenrostro, J., Wu, B., Chang, H., Greenleaf, W., 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218. https://doi.org/10.1038/nmeth.2688

Carrasco, M., Delgado, I., Soria, B., Martín, F., Rojas, A., 2012. GATA4 and GATA6 control mouse pancreas organogenesis. J Clin Invest 122, 3504–3515. https://doi.org/10.1172/JCI63240

Chatagnon, A., Veber, P., Morin, V., Bedo, J., Triqueneaux, G., Sémon, M., Laudet, V., d'Alché-Buc, F., Benoit, G., 2015. RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. Nucleic Acids Res. 43, 4833–4854. https://doi.org/10.1093/nar/gkv370

Chen, Y., Pan, F.C., Brandes, N., Afelik, S., Sölter, M., Pieler, T., 2004. Retinoic acid signaling is essential for pancreas development and promotes endocrine at the expense of exocrine cell differentiation in Xenopus. Developmental Biology 271, 144–160. https://doi.org/10.1016/j.ydbio.2004.03.030

Das, B.C., Thapa, P., Karki, R., Das, S., Mahapatra, S., Liu, T.-C., Torregroza, I., Wallace, D.P., Kambhampati, S., Van Veldhuizen, P., Verma, A., Ray, S.K., Evans, T., 2014. Retinoic Acid Signaling Pathways in Development and Diseases. Bioorg Med Chem 22, 673–683. https://doi.org/10.1016/j.bmc.2013.11.025

Decker, K., Goldman, D.C., Grasch, C.L., Sussel, L., 2006. Gata6 is an important regulator of mouse pancreas development. Dev Biol 298, 415–429. https://doi.org/10.1016/j.ydbio.2006.06.046

Delacroix, L., Moutier, E., Altobelli, G., Legras, S., Poch, O., Choukrallah, M.-A., Bertin, I., Jost, B., Davidson, I., 2010. Cell-Specific Interaction of Retinoic Acid Receptors with Target Genes in Mouse Embryonic Fibroblasts and Embryonic Stem Cells. Mol Cell Biol 30, 231–244. https://doi.org/10.1128/MCB.00756-09

Diez del Corral, R., Morales, A., Diez del Corral, R., Morales, A.V., 2014. Retinoic Acid Signaling during Early Spinal Cord Development. Journal of Developmental Biology 2, 174–197. https://doi.org/10.3390/jdb2030174

Dilworth, F.J., Chambon, P., 2001. Nuclear receptors coordinate the activities of chromatin remodelling complexes and coactivators to facilitate initiation of transcription. Oncogene 20, 3047–3054. https://doi.org/10.1038/sj.onc.1204329

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Engström, P.G., Fredman, D., Lenhard, B., 2008. Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. Genome Biology 9, R34. https://doi.org/10.1186/gb-2008-9-2-r34

Feng, L., Hernandez, R.E., Waxman, J.S., Yelon, D., Moens, C.B., 2010. Dhrs3a Regulates Retinoic Acid Biosynthesis through a Feedback Inhibition Mechanism. Dev Biol 338, 1–14. https://doi.org/10.1016/j.ydbio.2009.10.029

Gere-Becker, M.B., Pommerenke, C., Lingner, T., Pieler, T., 2018. Retinoic acid-induced expression of Hnf1b and Fzd4 is required for pancreas development in Xenopus laevis. Development 145, dev161372. https://doi.org/10.1242/dev.161372

Ghosh, P., Maurer, J.M., Sagerström, C.G., 2018. Analysis of novel caudal hindbrain genes reveals different regulatory logic for gene expression in rhombomere 4 versus 5/6 in embryonic zebrafish. Neural Development 13, 13. https://doi.org/10.1186/s13064-018-0112-y

Grandel, H., Lun, K., Rauch, G.-J., Rhinn, M., Piotrowski, T., Houart, C., Sordino, P., Küchler, A.M., Schulte-Merker, S., Geisler, R., Holder, N., Wilson, S.W., Brand, M., 2002. Retinoic acid signalling in the zebrafish embryo is necessary during pre-segmentation stages to pattern the anterior-posterior axis of the CNS and to induce a pectoral fin bud. Development 129, 2851–2865.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hiller, M., Agarwal, S., Notwell, J.H., Parikh, R., Guturu, H., Wenger, A.M., Bejerano, G., 2013. Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: application to zebrafish. Nucleic Acids Res. 41, e151. https://doi.org/10.1093/nar/gkt557

Holtzinger, A., Evans, T., 2005. Gata4 regulates the formation of multiple organs. Development 132, 4005–4014. https://doi.org/10.1242/dev.01978

Iwafuchi-Doi, M., Zaret, K.S., 2014. Pioneer transcription factors in cell reprogramming. Genes Dev 28, 2679–2692. https://doi.org/10.1101/gad.253443.114

Kam, R.K.T., Deng, Y., Chen, Y., Zhao, H., 2012. Retinoic acid synthesis and functions in early embryonic development. Cell Biosci 2, 11. https://doi.org/10.1186/2045-3701-2-11

Kam, R.K.T., Shi, W., Chan, S.O., Chen, Y., Xu, G., Lau, C.B.-S., Fung, K.P., Chan, W.Y., Zhao, H., 2013. Dhrs3 protein attenuates retinoic acid signaling and is required for early embryonic patterning. J. Biol. Chem. 288, 31477–31487. https://doi.org/10.1074/jbc.M113.514984

Kedishvili, N.Y., 2016. RETINOIC ACID SYNTHESIS AND DEGRADATION. Subcell Biochem 81, 127–161. https://doi.org/10.1007/978-94-024-0945-1_5

Lancman, J.J., Zvenigorodsky, N., Gates, K.P., Zhang, D., Solomon, K., Humphrey, R.K., Kuo, T., Setiawan, L., Verkade, H., Chi, Y.-I., Jhala, U.S., Wright, C.V.E., Stainier, D.Y.R., Dong, P.D.S., 2013. Specification of hepatopancreas progenitors in zebrafish by hnf1ba and wnt2bb. Development 140, 2669–2679. https://doi.org/10.1242/dev.090993

Langmead, B., Salzberg, S., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–9.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Maden, M., 2007. Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat. Rev. Neurosci. 8, 755–765. https://doi.org/10.1038/nrn2212

Maden, M., 2002. Retinoid signalling in the development of the central nervous system. Nature Reviews Neuroscience 3, 843–853. https://doi.org/10.1038/nrn963

Mahony, S., Mazzoni, E.O., McCuine, S., Young, R.A., Wichterle, H., Gifford, D.K., 2011. Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. Genome Biol 12, R2. https://doi.org/10.1186/gb-2011-12-1-r2

Martín, M., Gallego-Llamas, J., Ribes, V., Kedinger, M., Niederreither, K., Chambon, P., Dollé, P., Gradwohl, G., 2005. Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. Developmental Biology 284, 399–411. https://doi.org/10.1016/j.ydbio.2005.05.035

Molotkov, A., Molotkova, N., Duester, G., 2005. Retinoic acid generated by Raldh2 in mesoderm is required for mouse dorsal endodermal pancreas development. Developmental Dynamics 232, 950–957. https://doi.org/10.1002/dvdy.20256

Moutier, E., Ye, T., Choukrallah, M.-A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D., Benoit, G., Davidson, I., 2012. Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology. J. Biol. Chem. 287, 26328–26341. https://doi.org/10.1074/jbc.M112.361790

Ou, J., Liu, H., Yu, J., Kelliher, M.A., Castilla, L.H., Lawson, N.D., Zhu, L.J., 2018. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. BMC Genomics 19, 169. https://doi.org/10.1186/s12864-018-4559-3

Paik, E.J., Mahony, S., White, R.M., Price, E.N., DiBiase, A., Dorjsuren, B., Mosimann, C., Davidson, A.J., Gifford, D., Zon, L.I., 2013. A Cdx4-Sall4 Regulatory Module Controls the Transition from Mesoderm Formation to Embryonic Hematopoiesis. Stem Cell Reports 1, 425–436. https://doi.org/10.1016/j.stemcr.2013.10.001

Quillien, A., Abdalla, M., Yu, J., Ou, J., Zhu, L.J., Lawson, N.D., 2017. Robust identification of developmentally active endothelial enhancers in zebrafish using FANS-assisted ATAC-Seq. Cell Rep 20, 709–720. https://doi.org/10.1016/j.celrep.2017.06.070

Rhinn, M., Dollé, P., 2012. Retinoic acid signalling during development. Development 139, 843–858. https://doi.org/10.1242/dev.065938

Schiesser, J.V., Wells, J.M., 2014. Generation of β cells from human pluripotent stem cells: Are we there yet? Ann N Y Acad Sci 1311, 124–137. https://doi.org/10.1111/nyas.12369

Schubert, M., Holland, N.D., Laudet, V., Holland, L.Z., 2006. A retinoic acid-Hox hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. Developmental Biology 296, 190–202. https://doi.org/10.1016/j.ydbio.2006.04.457

Stafford, D., Hornbruch, A., Mueller, P.R., Prince, V.E., 2004. A conserved role for retinoid signaling in vertebrate pancreas development. Dev Genes Evol 214, 432–441. https://doi.org/10.1007/s00427-004-0420-6

Stafford, D., Prince, V.E., 2002. Retinoic Acid Signaling Is Required for a Critical Early Step in Zebrafish Pancreatic Development. Current Biology 12, 1215–1220. https://doi.org/10.1016/S0960-9822(02)00929-6

Stafford, D., White, R.J., Kinkel, M.D., Linville, A., Schilling, T.F., Prince, V.E., 2006. Retinoids signal directly to zebrafish endoderm to specify insulin-expressing beta-cells. Development 133, 949–956. https://doi.org/10.1242/dev.02263

Stark, R., Brown, G., n.d. DiffBind: Differential binding analysis of ChIP-Seq peak data 34.

Sun, Z., Hopkins, N., 2001. vhnf1, the MODY5 and familial GCKD-associated gene, regulates regional specification of the zebrafish gut, pronephros, and hindbrain. Genes Dev. 15, 3217–3229. https://doi.org/10.1101/gad946701

Tanoury, Z.A., Piskunov, A., Rochette-Egly, C., 2013. Vitamin A and retinoid signaling: genomic and nongenomic effects Thematic Review Series: Fat-Soluble Vitamins: Vitamin A. J. Lipid Res. 54, 1761–1775. https://doi.org/10.1194/jlr.R030833

Xuan, S., Borok, M.J., Decker, K.J., Battle, M.A., Duncan, S.A., Hale, M.A., Macdonald, R.J., Sussel, L., 2012. Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. J Clin Invest 122, 3516–3528. https://doi.org/10.1172/JCI63352

Yu, G., Wang, L., He, Q., 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–3.

Zaret, K.S., Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25, 2227–2241. https://doi.org/10.1101/gad.176826.111

Zeynali, B., Dixon, K.E., 1998. Effects of retinoic acid on the endoderm in Xenopus embryos. Dev. Genes Evol. 208, 318–326.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhu, L., Gazin, C., Lawson, N., Pages, H., Lin, S., Lapointe, D., Green, M., 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 11, 237.

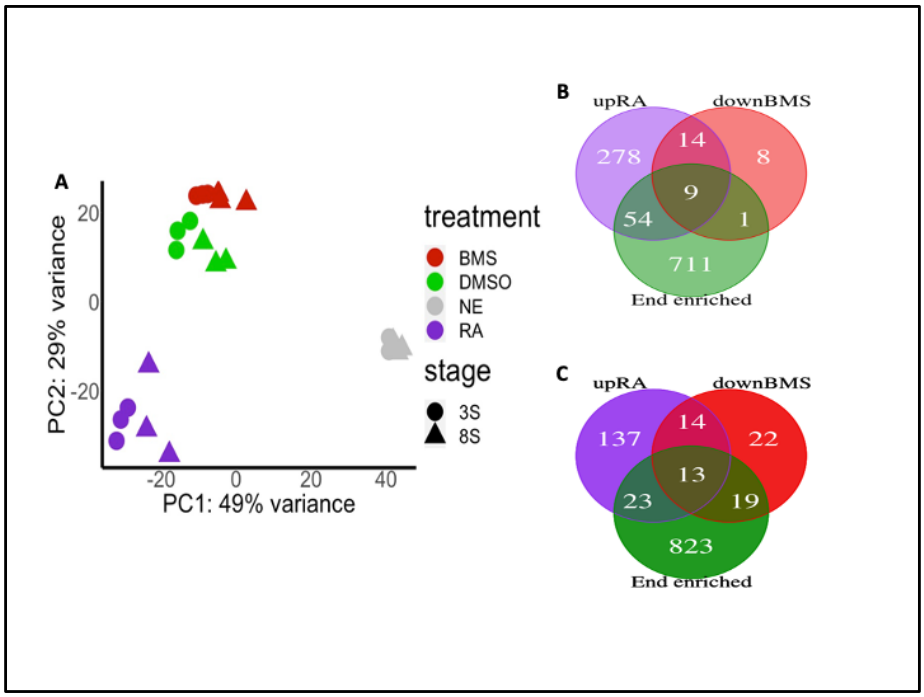

Figure 1. RA has a bigger effect on the transcriptome of endodermal cells than BMS493.

(A) PCA of rlog calculated by DESeq2 for cells at 3 and 8 somites stage. We can see a strong clustering among RA-treated and BMS-treated and control cells. A second cluster can be define based on the stage, 3 and 8 somites stage for each treatment. (B-C) Venn diagram displaying the number of genes up-regulated by RA (purple), down-regulated by BMS493 (red) and endodermal enriched (green) at 3 somites (B) and 8 somites stage (C).

Gene Name	FC(RA vs DMSO)	FDR(RA vs DMSO)	FC(DMSO vs BMS)	FDR(DMSO vs BMS)
lgals2b	16,24	1,24E-15	5,71	1,27E-03
SLC22A3	13,96	3,19E-12	31,26	1,36E-14
foxg1b	11,51	1,90E-17	9,93	1,05E-08
pcdh10a	6,94	2,40E-06	11,17	3,42E-08
mxtx1	5,57	1,81E-07	5,43	5,80E-05
nr2f1a	5,26	2,21E-08	7,83	3,61E-11
gata6	3,78	1,08E-03	9,62	4,21E-09
insm1a	3,57	3,23E-03	36,84	1,85E-10
hoxb1b	2,62	2,16E-04	2,53	4,96E-03
hoxb5b	21,17	2,96E-40	3,91	1,32E-06
nr6a1b	12,67	2,26E-04	38,63	2,89E-03
dhrs3a	11,09	4,78E-08	59,32	8,76E-20
si:dkey-117n7.5	6,88	1,70E-03	154,57	8,30E-05
hoxc1a	6,86	2,25E-10	5,86	1,19E-06
GADD45G	6,09	1,53E-03	21,42	1,73E-06
col7a1l	6,04	8,18E-04	74,87	1,25E-04
ppm1h	5,03	3,56E-04	6,17	1,83E-03
si:dkey-117n7.2	4,81	4,28E-03	66,54	2,28E-03
trabd2b	4,29	5,45E-05	4,17	2,11E-03
ascl1b	3,95	5,22E-03	16,67	1,51E-09
nrip1a	3,93	3,93E-03	6,94	1,15E-03
nr2f5	3,76	6,19E-08	3,66	2,22E-06
tshz1	2,36	3,76E-03	4,03	1,32E-06

Table 1. Genes up-regulated by RA and down-regulated by BMS493 at 3 somties stage. In green are the genes that are also enriched in endoderm.

Gene Name	FC(RA vs DMSO)	FDR(RA vs DMSO)	FC(DMSO vs BMS)	FDR(DMSO vs BMS)
hoxb5b *	12,07	1,55E-32	6,26	7,24E-16
slc13a2	8,58	2,14E-07	11,18	7,32E-07
dhrs3a *	8,13	2,40E-18	100,51	1,57E-56
hoxb6b	7,46	4,15E-10	3,40	5,40E-03
si:ch211-157b11.8	7,65	1,62E-06	5,32	3,25E-03
SLC22A3 *	5,43	1,09E-08	30,78	1,85E-27
ezra	5,25	4,97E-04	7,89	3,59E-05
tmem30c	5,15	9,33E-05	8,01	1,85E-05
cldn15a	4,66	7,70E-04	10,12	1,86E-07
pcdh10a *	4,41	2,42E-04	9,77	8,08E-09
epb41l4a	3,57	6,96E-06	7,06	6,52E-11
nr2f1a *	3,53	4,28E-05	5,98	3,93E-09
hmga2	2,54	5,11E-04	3,18	6,39E-05
hoxa3a	18,66	9,70E-40	3,41	1,59E-05
skap1	12,49	8,28E-17	5,17	1,66E-04
masp1	9,61	2,19E-10	17,11	2,99E-07
foxg1b *	8,27	1,11E-06	25,02	2,99E-07
si:dkey-172f14.2	8,00	7,40E-08	14,27	1,77E-04
zgc:194209	7,70	1,29E-07	6,84	6,66E-04
cldn2	7,40	2,18E-06	11,58	1,85E-05
si:dkey-25e11.10	5,68	8,30E-06	6,95	1,05E-03
ascl1a	5,61	7,72E-08	11,93	2,35E-07
cldnc	5,19	1,96E-06	10,72	8,81E-12
hoxc1a *	5,00	1,10E-05	5,90	1,09E-05
iqgap2	4,75	1,55E-04	4,80	3,76E-03
zgc:113054	2,90	1,66E-04	3,36	8,39E-05
tshz1 *	2,90	1,51E-03	2,93	7,29E-03

Table 2. Genes up-regulated by RA and down-regulated by BMS493 at 8 somties stage. In green are the genes that are also enriched in endoderm. Genes that are also regulated at 3 somites are marked with an *

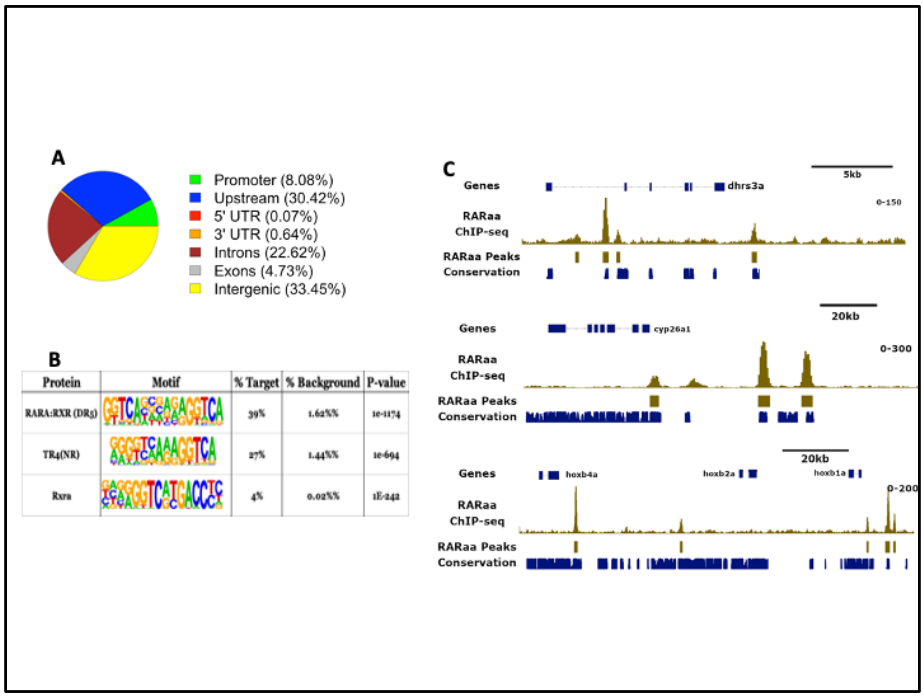

Figure 2. RAR binding sites correspond to cis-regulatory elements. (a)Proportion of ChIP-seq peaks that correspond to different regions of the chromatin. (B) Top 3 motifs enriched in ChIP-seq elements. The first one correspond to a DR6 motif, the second one to a DR2 and the third one to a IR0. The three of them are motifs where the heterodimer RAR:RXR can be bound. (C) Visualization of RARaa binding sites around the dhrs3a (up), cyp26a1 (middle) and hoxb1a-hoxb4a genomic region (down). Tracks corresponding RARaa binding sites and RARaa peaks (gold). Additional track corresponding conservation of genomic sequences from the UCSC Genome Browser among 5 fish species are also shown.

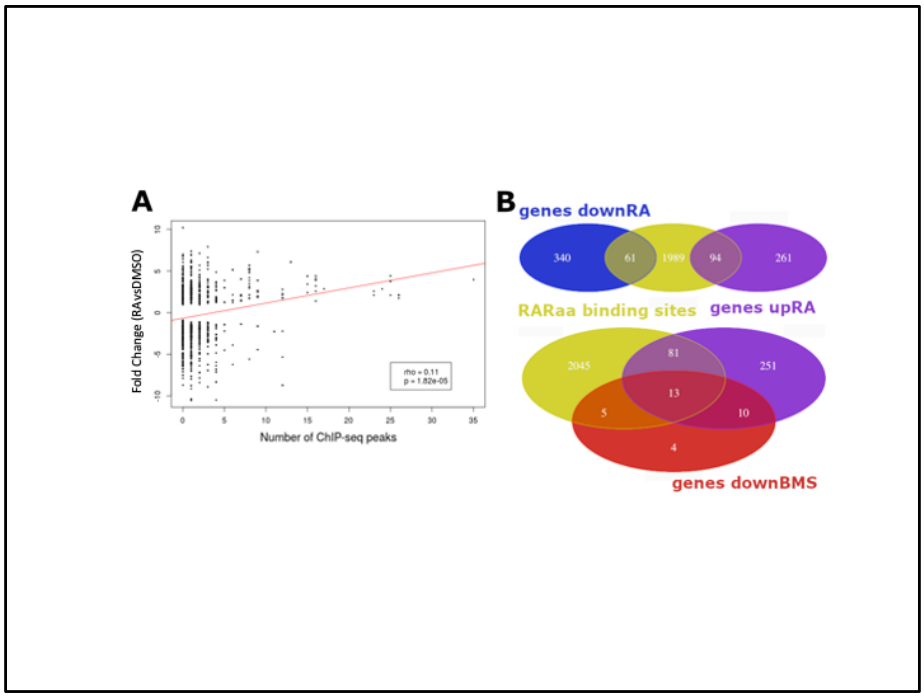

Figure 3. Most of the RAR binding sites are located around genes regulated by RA. (A)Correlation of gene expression (log2 fold change) between endodermal cells treated with DMSO or RA to number of flanking RARaa ChIP-seq peaks. Only significantly expressed genes were included. (B) Venn diagram showing the number of genes around peaks (yellow) from ChIP-seq experiments that correlate with genes up-regulated by RA (purple), down-regulated by RA (blue) and donw-regulated by BMS (red).

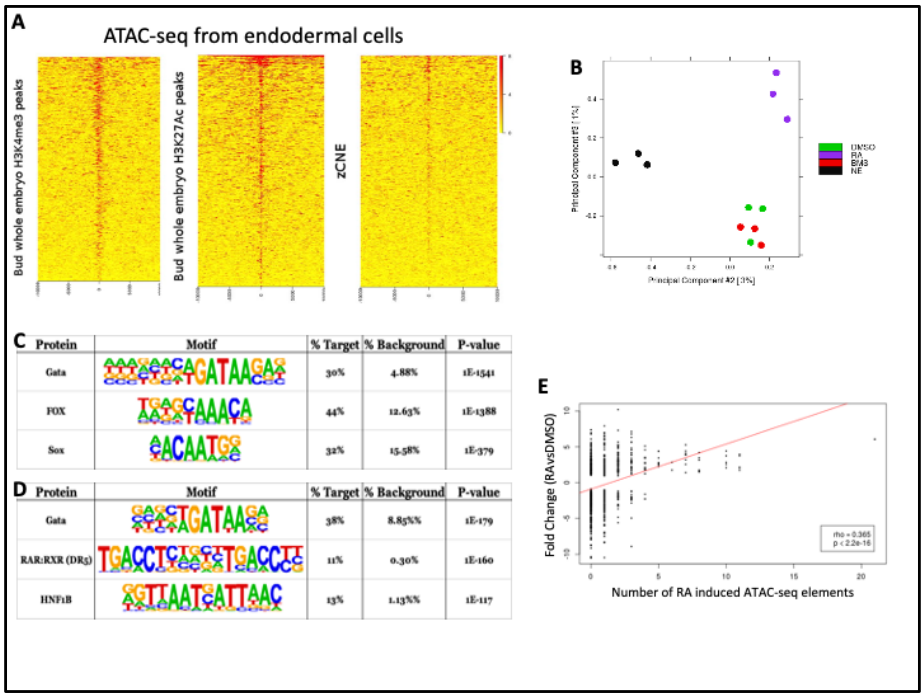

Figure 4. Open chromatin analysis. (A) Heat maps showing enrichment of ATAC-seq reads at the middle of the histone mark peaks and at areas that correspond to zCNE. The maps display intervals flanking 10 kb up and downstream of the features. The heat map plots shown on this figure corresponds to the ATAC-seq data obtained with control endodermal cells (3SDMSO). Similar results were obtained for the other samples (see Supplementary Figure 2). (B) PCA plots obtained for the ATAC-seq libraries. The one on the left represent the PC1 vs PC2 and the one on the right the PC2 vs PC3. Endodermal vs non-endodermal cells are separated based on PC2, while RA-treated vs control and BMS493 are separated based on the PC3. However, no separation can be distinguished for DMSO and BMS493 treated cells. (C-D) Top 3 enriched motifs found in endodermal specific elements (C) and RA-induced elements (D). (E) Plots comparing gene expression (log2 fold change), between RA treated and control endodermal cells to number of flanking inducedRA elements (up) and repressedRA elements (down). Only significantly expressed genes were included.

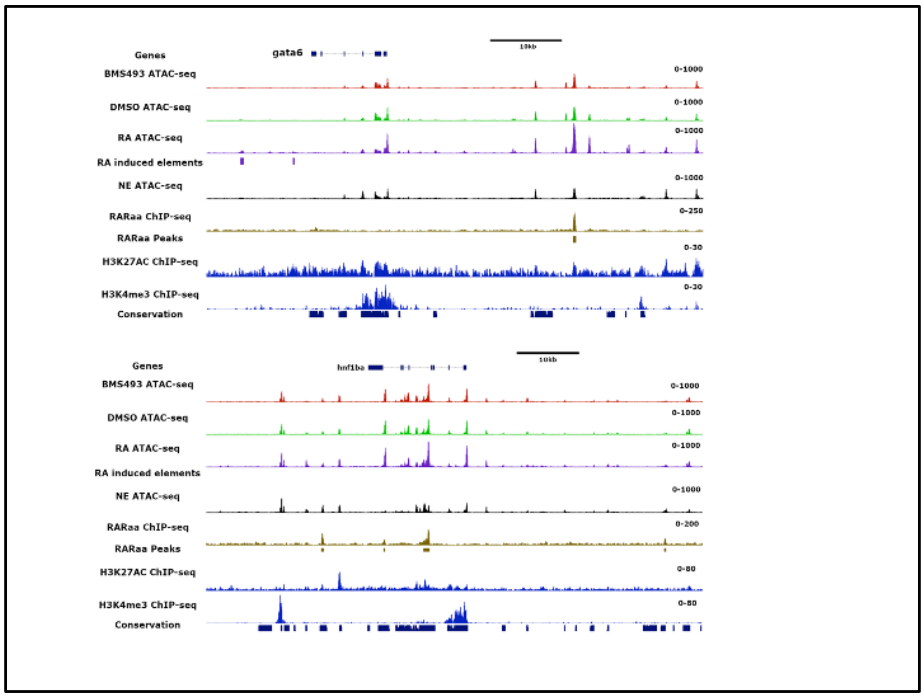

Figure 5. Visualization of RARaa binding sites and of nucleosome-free regions (ATAC-seq reads) around the gata6 (up) and hnf1ba (down).
(A) Visualization of ATAC-seq reads from the merged 3 replicates from endoderm and non-endodermal cells, RARaa binding sites, H3K27Ac and H3K3me3 marks determined by ChIP-seq. Additional track corresponding conservation of genomic sequences from the UCSC Genome Browser among 5 fish species are also shown.

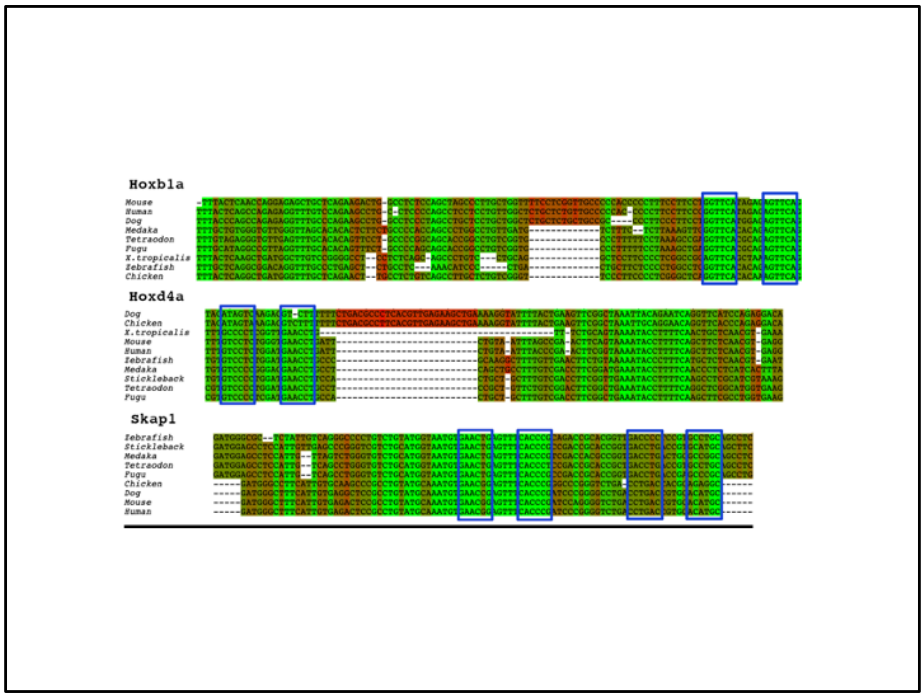

Figure 6. RARE found by ChIP-seq analysis are conserved among vertebrates. Conservation of RARE near *hoxb1a, hoxd4a* and *skap1* genes in 10 species. The RARE around *hxb1a* and *skap1* were not conserved in *Stickleback*. Green to red, conservation of nucleotides among species, in green nucleatides conserved in all the specides and in red nucleotides not conserved.