Identification of downstream effectors of retinoic acid specifying the zebrafish pancreas by integrative genomics

Ana R. López-Pérez1,5, Piotr J. Balwierz2, Boris Lenhard2, Ferenc Muller3, Fiona C. Wardle4, Isabelle Manfroid1, Marianne L. Voz1 & Bernard Peers1*

Retinoic acid (RA) is a key signal for the specification of the pancreas. Still, the gene regulatory cascade triggered by RA in the endoderm remains poorly characterized. In this study, we investigated this regulatory network in zebrafish by combining RNA-seq, RAR ChIP-seq and ATAC-seq assays. By analysing the effect of RA and of the RA receptor (RAR) inverse-agonist BMS493 on the transcriptome and on the chromatin accessibility of endodermal cells, we identified a large set of genes and regulatory regions regulated by RA signalling. RAR ChIP-seq further defined the direct RAR target genes in zebrafish, including hox genes as well as several pancreatic regulators like mnx1, insm1b, hnf1ba and gata6. Comparison of zebrafish and murine RAR ChIP-seq data highlighted the conserved direct target genes and revealed that some RAR sites are under strong evolutionary constraints. Among them, a novel highly conserved RAR-induced enhancer was identified downstream of the HoxB locus and driving expression in the nervous system and in the gut in a RA-dependent manner. Finally, ATAC-seq data unveiled the role of the RAR-direct targets Hnf1ba and Gata6 in opening chromatin at many regulatory loci upon RA treatment.

Retinoic acid (RA) is essential for the development of vertebrate embryos. This small lipophilic molecule controls the formation of many organs and tissues and is crucial for the anteroposterior (AP) patterning of the hindbrain and endoderm1–3. During gastrulation and early somitogenesis, RA forms a two-tailed gradient with highest levels in the mid-trunk region and progressive lower levels at the anterior and posterior regions of the embryo4. This gradient regionalizes the endoderm along the AP axis and specifies distinct organs at the appropriate location5, 6. RA is notably required for the specification of the pancreatic field in several animal models like zebrafish, xenopus, chick and mice7–12 and protocols for generating pancreatic cells in vitro from embryonic stem cells include a RA incubation step of endodermal cells13. Experiments in zebrafish have shown that, when RA signalling is blocked pharmacologically by incubating embryos with the pan-RAR inverse agonist BMS493, no pancreas develops; conversely, when embryos are incubated with exogenous RA, ectopic pancreatic cells are generated in the anterior endoderm. Such treatments performed at different stages revealed that RA instructs endoderm to form pancreas during late gastrulation (from about 8 to 12 h post-fertilization, hpf)7. Cell transplantation studies have also shown that RA, synthesized in the anterior paraxial mesoderm, acts directly on the mid-trunk endoderm to promote pancreas development14. Although several regulatory genes have been identified as important downstream targets of RA, such as Hox or Hnf1b genes15–17, the regulatory network triggered by RA in endoderm during gastrulation is still unclear. Many transcriptomic studies have identified RA-regulated genes in murine, xenopus or zebrafish embryos18–23; however these reports were often done on whole embryos and not done specifically on endodermal tissues. Still two studies performed on xenopus endoderm have nevertheless reported the involvement of Nدرگ، Fzd4 and Hnf1b genes as crucial mediators of RA signalling for pancreas development16, 24; however, it is not yet known if the role of these regulatory genes is conserved among vertebrates.

1Laboratory of Zebrafish Development and Disease Models (ZDDM), GIGA-R, SART TILMAN, University of Liège, Avenue de l’Hôpital 1, B34, 4000 Liège, Belgium. 2Institute of Clinical Sciences and MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, London W12 ONN, UK. 3Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK. 4Randall Centre for Cell and Molecular Biophysics, New Hunt’s House, Guy’s Campus, King’s College London, London SE1 1UL, UK. 5Present address: Umeå Centre for Molecular Medicine (UCMM), Umeå University, Umeå, Sweden. 5email: bpeers@uliege.be
RA levels are tightly controlled in embryos by the balance of its synthesis, controlled mainly by Aldh1a2 (Raldh2), and its degradation catalysed by Cyp26a enzymes. This synthesis/degradation balance is notably regulated by auto-regulatory loops where RA levels control expression of some of its metabolizing enzymes (e.g., Cyp26a1, Cyp26b1 and Dhrs3)23,27–28. RA controls gene expression through binding to the RAR (Retinoic Acid Receptors) which form heterodimers with Retinoid X Receptors (RXR). These RAR/RXR heterodimers bind to DNA regulatory sequences, called “Retinoic Acid Response Elements” (RAREs), located in promoters or enhancers. Genome-wide identification of RAR binding sites has been achieved by ChIP-seq experiments mostly using murine or human cell lines29, 30–32. Such studies confirmed that the RAR/RXR heterodimers bind to direct repeats of the RGKTC motif (R = A/G, K = G/T) usually separated by 5 bases (DR5), but also to repeats of this motif with other spacing and orientations. These data uncovered several thousand RAR/RXR binding sites in the murine and human genomes, some located near RA-induced genes like the Hox, Cyp26a, or Rar genes. However, it is not yet clear if all the identified RAR binding sites have a regulatory function.

In order to decipher the gene regulatory network triggered by RA during gastrulation in zebrafish endodermal cells, we combined RNA-seq, ChIP-seq and ATAC-seq approaches. We identified genes regulated by RA signalling by analysing the transcriptome of endodermal cells isolated from RA- or BMS493-treated embryos. Direct targets of RAR were identified by ChIP-seq assays. Then, ATAC-seq experiments29 allowed us to identify chromatin regions whose accessibility is modified upon RA treatment. By integrating all these data, we identify important transcription factors acting downstream RA signalling in endodermal cells and involved in pancreas development. Furthermore, comparison of the RAR sites detected in the zebrafish genome and in the murine genome unveiled RAR target genes which have been maintained during evolution and whose RAR site sequences are well conserved. By this approach, RAR sites with essential function can be identified in vertebrate genomes.

**Results**

**Retinoic acid affects the transcriptome of zebrafish endodermal cells.** To identify genes regulated by RA in zebrafish endodermal cells, we used the transgenic Tg(sox17:GFP) line which drives GFP expression in endodermal cells and allows their selection by fluorescence activated cell sorting (FACS). Tg(sox17:GFP) embryos were treated either with RA, BMS493 (pan-RAR inverse-agonist) or DMSO (control) from 1.25 to 11 hpf, a time window that covers the whole blastula and gastrula periods. Endodermal GFP+ cells were next selected by FACS from embryos at 3-somites (3-S) and 8-somites (8-S) stages (11 hpf and 13 hpf, respectively). Non-endodermal (GFP−) cells were also selected from the DMSO-treated control embryos in order to compare with GFP+ endodermal cells and identify genes displaying endodermal enriched expression. RNA-seq was performed on all these FACS-isolated cells prepared in triplicate (24 samples in total) and transcriptomes were analysed using the bioinformatic pipeline as described in “Materials and methods”. Principal component analysis of all RNA-seq data (Fig. 1A) shows (i) a tight clustering of all triplicate samples confirming a high reproducibility in the experiment, (ii) a strong difference between the transcriptome of endodermal and non-endodermal (NE) cells (discriminated along the first axis of the PCA plot), (iii) relatively similar transcriptomes of cells isolated at 3-S and 8-S stages, and (iv) a clustering of BMS493 samples near DMSO samples indicating a much weaker effect of BMS493 treatments compared to the RA treatments. These conclusions were further confirmed by the differential gene expression analyses between the different conditions. Indeed, 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 with FDR < 0.01) than between endodermal cells treated with RA versus DMSO (756 and 514 RA-regulated genes at 3-S and 8-S stages, respectively with FDR < 0.01) or with BMS493 versus DMSO (32 and 71 BMS493-regulated genes at 3-S and 8-S stages, respectively). We found a large overlap among the sets of genes having an endodermal-enriched expression at 3-S and at 8-S stages (Fig. S1A; list of genes given in Table S1) and these sets include all known endodermal markers including sox17, gata6/5/6 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 (Table S2), many of them being regulated at both 3-S and 8-S stages (Fig. S1B). These RA-regulated genes include known RAR-direct targets such as cyp26b1/a1, dhrs3a, nr2f2 and several hox genes, validating our protocol and the use of RA treatments. Interestingly, BMS493 treatment led mostly to down-regulation of gene expression. Indeed at 3-S stage, the 32 BMS493-regulated genes were all repressed, and, at 8-S stage, 68 genes were repressed while only 3 genes were up-regulated by BMS493 treatment (Tables S3). A large overlap was also observed between the genes down-regulated by BMS493 at 3-S and at 8-S stage (Fig. S1C). As expected, a large proportion of genes down-regulated by BMS493 were up-regulated by RA treatment, this observation being evident mostly at 3-S stage where 72% of genes repressed by the RAR inverse-agonist were induced by RA, while this proportion decreased to 40% at 8-S stage (Fig. 1B,C). Tables 1 and 2 show the genes significantly up-regulated by RA and down-regulated by BMS493 as well as those enriched in the endoderm at 3- and 8-somites stage, respectively (shown in bold). gata6, insm1a and ascl1b are the only known pancreatic regulatory genes which were regulated by both RA and BMS493 (Table 1). Other pancreatic transcription factors, like mnx1, insm1b, hnf1ba, nr5a2 or neuroD1, were induced by the RA-treatment but not significantly repressed by BMS493. Inversely, other pancreatic regulators were inhibited by BMS493 but not significantly induced by RA like pdx1, rfx6 and mtf1 (Tables S3 and S6). We can assume that the induction of pancreatic fate by RA (and the absence of pancreas upon BMS439 treatment) is mediated, at least in part, by the direct or indirect regulation of these pancreatic regulatory factors. In conclusion, these RNA-seq data highlight all the genes with an enriched expression in zebrafish endodermal cells and which are regulated by RA signalling. This gene set notably includes regulators involved in the AP patterning of the endoderm, such as hox genes, and known factors involved in the specification of pancreatic progenitors.
Identification of RAR binding sites in the zebrafish genome. To further identify the genes directly regulated by RAR and determine if some pancreatic regulatory genes are direct targets of RA signalling, we performed ChIP-seq experiments at the end of gastrulation. In absence of validated commercialized ChIP-grade antibody recognizing zebrafish RAR, we chose to express a tagged RARaa in zebrafish gastrulae by injecting zebrafish fertilized eggs with the 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 indicated that it is the most highly expressed RA receptor in zebrafish endodermal cells. We injected very small amount of this mRNA (about 30 pg) in order to avoid high non-physiological levels of RARa protein within embryos; the development of embryos were not affected by these injections. Chromatin was prepared at 11.5 hpf (3 somites stage) from about 2000 injected zebrafish embryos and immunoprecipitation was performed with a ChIP-grade Myc antibody. Comparison of reads obtained with the Myc-RAR ChIP and the input negative 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 (Table S4). By choosing such criteria, 2848 robust RAR binding sites were identified in the zebrafish genome. As shown in Fig. 2A, a majority of these sites are located near or within genes: 8% were found in promoters (i.e. 1 kb upstream of the gene TSS, Transcription Start Site), 30% in upstream sequences (from 1 to 10 kb), 22% in introns, 4% in exons, and only 33% in intergenic regions. Sequence analysis of all RAR peaks revealed that the highest represented motif corresponds to the Direct Repeat of the RGKTCA motif separated by 5 bases (reported as DR5) and being the RAR/RXR consensus binding sequence (present in 39% of identified RAR peaks) (Fig. 2B). The next most abundant motifs are also repetitions of the RGKTCA motif with different spacings (TR4 being a Direct Repeat separated by 1 and 2 bases : DR1/DR2) and orientations (Rara being an Inverted Repeat with no base separation : IR0). Furthermore, many RAR peaks were found near genes reported to be RAR-direct target genes in other species, such as cyp26a1, dhrs3, nr2f2 or the hoxb1a-hoxb4a genomic region (Fig. 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 (Fig. 2C and see below).

The RAR ChIP-seq peaks located within 250 kb from the TSS of a gene were assigned to this gene and when several genes were lying in the vicinity of a RAR site, the closest gene was considered as the putative
RAR-regulated gene. Using this strategy, amongst the 2848 RAR sites, 2144 were linked to a gene. Correlation analysis between RA gene expression regulated by RA and the number of RAR sites near the gene showed that RAR acts mainly as a transcriptional activator (Fig. 3A), supporting the classical model where RAR/RXR heterodimers mainly recruit co-activators upon RA ligand binding\(^1\). However, this correlation is not very high and genes down-regulated by RA can harbour nearby RAR sites. Indeed, from the gene set having RAR sites, 61 were down-regulated and 94 genes were up-regulated by RA at 3-S stage (Fig. 3B, Tables S5). Among the genes up-regulated by RA and harbouring a RAR site, we identified the pancreatic regulatory genes hnf1ba/b, gata6, insm1b, jag2b and mnx1 indicating that these genes are direct targets of RA. As for the genes down-regulated by BMS493, we found that a large number of them (i.e. 18 genes out of 32) contain RAR sites and amongst them 13 are also upregulated by RA (Fig. 3B, see legends for gene names). In conclusion, the ChIP-seq data allowed us to define the zebrafish RAR cistrome at the end of gastrulation and identify putative RAR direct target genes.

Conservation of RAR binding sites from fish to mammals. Functionally important regulatory regions are expected to be conserved during evolution. To determine which RAR binding sites are conserved in vertebrates, we compared our zebrafish ChIP-seq data with those of Chatagon and colleagues\(^{19}\) who identified RAR binding sites in the murine genome using the F9 embryonal carcinoma cells whose differentiation into primitive endodermal cell is induced by RA treatment. This comparison revealed that, among the 2144 zebrafish genes harbouring a RAR site, 722 have also a RAR site near the murine orthologous genes. This list of conserved RAR-bound genes comprises notably cyp26a1/b1, dhrs3a, nr2f2, many hbox genes, raraa/b as well as pancreatic genes gata6, insm1b, jag2b and mnx1 indicating that these genes are direct targets of RA. As for the genes down-regulated by BMS493, we found that a large number of them (i.e. 18 genes out of 32) contain RAR sites and amongst them 13 are also upregulated by RA (Fig. 3B, see legends for gene names). In conclusion, the ChIP-seq data allowed us to define the zebrafish RAR cistrome at the end of gastrulation and identify putative RA direct target genes.

Table 1. List of genes up-regulated by RA and down-regulated by BMS493 at 3-somites stage (with FDR < 0.01). The regulation of each gene is shown by the Fold Change expression ratio (FC) and with the False Discovery Rate (FDR) for the RA treatment (column 2 and 3) and for the BMS493 treatment (columns 4 and 5). The genes highlighted in bold are those showing enriched expression in endoderm (FC > 4 and FDR < 0.01).

| 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       |
| pdcdh10a  | 6.94           | 2.40E−06       | 11.17          | 3.42E−08       |
| mxt1      | 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       |
| hbox5b    | 2.62           | 2.16E−04       | 2.53           | 4.96E−03       |
| hbox5b    | 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       |
| sidkey-177n7.5 | 6.88      | 1.70E−03       | 154.57         | 8.30E−05       |
| hbox1a    | 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       | 7.87           | 1.25E−04       |
| ppm1h     | 5.03           | 3.56E−04       | 6.17           | 1.83E−03       |
| sidkey-177n7.2 | 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       |
| shh1      | 2.36           | 3.76E−03       | 4.03           | 1.32E−06       |
GFP levels were detected in the posterior hindbrain, in a pattern highly reminiscent of hoxb1b gene expression (shown on Fig. 5D). Furthermore, when transgenic embryos were treated with exogenous RA, GFP expression was drastically increased and detected in the whole morphologically affected embryos in a similar manner to hoxb1b (Fig. 5B,E). Conversely, treatments with the RA inverse-agonist BMS493 turned off the expression of this DR5-RAR-skap1:GFP transgene (Fig. 5C) as well as of hoxb1b, except in the tailbud region (Fig. 5F). This confirms the RARE function of this element and indicates a role in hoxbb cluster regulation.

We also verified the regulatory role of some other RAR sites identified near zebrafish pancreatic regulatory genes. For example, we tested the activity of a RAR site located near a CNE downstream of the zebrafish mnx1 gene. A RAR binding site is also found at a similar location downstream of the murine Mnx1 gene (Fig. S2A). This RAR element was able to target GFP expression to the posterior endoderm at low level and in the dorsal pancreatic bud at higher levels (Fig. S2B–D). When the RAR-mnx1:GFP transgenic embryos were treated with RA, GFP expression was slightly increased in whole endoderm; conversely, treatment with BMS493 abolished GFP expression (Fig. S2D–F).

In conclusion, these analyses show that a large fraction of RAR sites has been conserved during evolution and transgenic assays confirm that some elements are sufficient to drive expression in endoderm and confer a RA-response.

**RA affects chromatin accessibility in endodermal cells.** RAR–RXR complexes control gene expression through the recruitment of corepressors (e.g. NCoR/Smart) and coactivators (NCoA) which control chromatin compaction via HDACs and HATs, respectively. Thus, one possible strategy to identify functional RAREs is to perform a genome-wide analysis of chromatin accessibility modifications following RA or RA inverse-agonist treatments by ATAC-seq assays allowing the identification of open chromatin and nucleosome-free regions induced or repressed by RA signaling. As most nucleosome-free regions correspond to regulatory sequences, ATAC-seq can also highlight the enhancers or promoters whose accessibility is modified by RA signalling indirectly, i.e. through the binding of transcription factors whose expression is induced by RA.

| 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         |
| stich211-157b11.8 | 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         |
| pdih10a*   | 4.41            | 2.42E−04         | 9.77            | 8.08E−09         |
| epb411A    | 3.57            | 6.96E−06         | 7.06            | 6.52E−11         |
| nr2f1a*    | 3.53            | 4.28E−05         | 5.98            | 3.93E−09         |
| hmg2a      | 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         |
| foxl1b*    | 8.27            | 1.11E−06         | 25.02           | 2.99E−07         |
| sidkey-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         |
| sidkey-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         |
| iogap2     | 4.75            | 3.55E−04         | 4.80            | 3.76E−03         |
| zgc:113054 | 2.90            | 1.66E−04         | 3.36            | 8.39E−05         |
| sizhl1*    | 2.90            | 1.51E−03         | 2.93            | 7.29E−03         |

Table 2. List of genes up-regulated by RA and down-regulated by BMS493 at 8-somites stage (with FDR < 0.01). The regulation of each gene is shown by the Fold Change expression ratio (FC) and with the False Discovery Rate (FDR) for the RA treatment (column 2 and 3) and for the BMS493 treatment (columns 4 and 5). The genes highlighted in bold are those showing enriched expression in endoderm (FC > 4 and FDR < 0.01). Genes displaying the same regulation at 3 somites (Table 1) are shown by an asterisk.
Figure 2. Identification of RAR binding sites in the zebrafish endoderm. (A) Distribution of ChIP-seq peaks to the different regions of the zebrafish genome. (B) Top 3 motifs overrepresented in all ChIP-seq peak sequences with the percentage of sites containing the motifs and the p-value of enrichment. The three motifs consist to a repetition of the A/GGG TCA sequence; the first corresponds to the classical DR5 recognized by the RARA:RXR complex, the second (TR4) is a superposition of DR1 and DR2, and the third is a IR0. (C) Visualization of RARaa binding sites around the dhrs3a gene (upper panel), cyp26a1 (middle panel) and hoxb1a-hoxb4a genomic region (below panel). Tracks in gold correspond to RARaa ChIP-seq reads and identified RARaa peaks. The track in blue shows the location of conserved genomic sequences (from the UCSC Genome Browser obtained from comparison of 5 fish species).
Thus, sequence analysis of all RA-induced ATAC-seq peaks can give clues on the identity of transcription factors acting in the subsequent steps of the RA-induced regulatory cascade.

As for the RNA-seq assays, zebrafish embryos were treated with RA, BMS493 and DMSO during blastula and gastrulation and about 10,000 endodermal cells were selected by FACS at 3-somites stage (11 hpf). Non-endodermal cells from control DMSO-treated embryos were also analysed in parallel. Cell preparations and ATAC-seq experiments were done in triplicate for each condition and analysed as described in "Materials and methods". We first verified the accuracy of the data by several quality control analyses. First, for all samples, the analysis of the ATAC-seq fragment size distribution revealed 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 (Fig. S3A). Secondly, as reported previously, genome mapping of the nucleosome-free fragments showed a clear enrichment in promoter regions immediately upstream of transcriptional start sites (TSSs), while mono-nucleosomes were depleted from TSSs and rather mapped just downstream of the TSSs in a periodic manner (Fig. S3B). Thirdly, we verified that the ATAC-seq fragments correspond to many zebrafish regulatory regions by comparing them with regions harbouring the histone modifications H3K4me3 and H3K27ac marks identified in zebrafish embryos at 10 hpf. Heatmap plots of ATAC-seq reads from all samples showed an obvious enrichment at loci harbouring these two histone modifications (Fig. 6A and Fig. S4). As regulatory regions often display sequence conservation, we also compared our ATAC-seq reads to the collection of zebrafish evolutionary-conserved non-coding elements (zCNEs). Heat-maps of ATAC-seq reads from each sample also showed a strong correlation with zCNEs (Fig. 6A and Fig. S4). These observations confirm that regions identified by ATAC-seq exhibited hallmarks of active regulatory elements. The reproducibility of ATAC-seq data was also analysed by PCA (Fig. 6B). This PCA showed that (i) triplicate samples are tightly clustered, and (ii) endodermal and non-endodermal (NE) cell clusters are separated along the PC2 axis, while the RA-treated cluster is separated from the DMSO- and BMS-cell cluster along the PC3 axis. Thus, as observed for the RNA-seq data (Fig. 1A), stronger differences are observed between GFP+ and GFP− cells compared to the differences between RA-treated and DMSO-treated endodermal cells.

Figure 3. Integrated analysis of the ChIP-seq and RNA-seq data. (A) Correlation of RA gene regulation (log2 fold change of expression RA versus DMSO) according to the number of neighbouring RARα ChIP-seq peaks. Only RA-regulated genes were included in the plot. (B) Venn diagram showing the overlap of genes harbouring nearby RARα binding sites (yellow) and those up-regulated by RA (purple), or down-regulated by RA (blue). The below panel also shows the overlap with the genes down-regulated by BMS493 (red). The 13 genes showing up-regulation by RA, down-regulation by BMS493 and harbouring a RAR site are tshz1, nr6a1b, foxg1b, nr2f5, gata6, dhrs3a, hoxb1b, slc22a3, ppm1b, nrp1a, col7a1l, hoxc1a and hoxb5b.
The samples corresponding to the BMS493-treated cells and DMSO-treated cells were not clearly segregated and did not reveal significant effects of BMS493 on the chromatin accessibility.

From all 12 ATAC-seq samples, a total of 156,604 nucleosome-free regions were identified within the zebrafish genome. Differential peak intensity analyses revealed that 9722 and 12,974 regions are more accessible in endodermal and non-endodermal cells, respectively (with FDR < 0.05, Tables S7). Interestingly, sequence analysis of all endodermal-specific ATAC-seq regions highlight a significant enrichment of the binding site motifs for the Gata, Fox and Sox protein families (Fig. 6C), in accordance with the well-known function of Gata4/5/6, Foxa1/2/3 and Sox32/17 in endodermal cell differentiation. Also, these endoderm-enriched ATAC-seq peaks are often identified near endodermal and pancreatic regulatory genes, such as in the foxa2, nkx6.1, hnf4g, sox17 or mnx1 loci (Fig. S5 and data not shown), suggesting the presence of endoderm-specific enhancers at these locations.

As expected, the RA treatments had less influence on ATAC-seq peaks compared to the cell type identity (i.e. endoderm versus non-endoderm). Still, 1240 genomic regions were found to be more accessible and 749 regions were less accessible in the RA-treated endoderm compared to the DMSO-treated controls (with FDR < 0.05). If the RA-treated samples were directly compared with the BMS493-treated samples, more peaks were detected as treatment-dependent: 3277 regions were more accessible in RA-treated while 1762 were more accessible in the BMS493-treated cells (Tables S8). Annotation of the RA-induced ATAC-seq peaks to the closest gene revealed
that they were often located near RA-upregulated genes identified above by RNA-seq. Moreover, we found a significant correlation between the level of RA-induced gene expression and the number of RA-induced ATAC-seq elements (Fig. 6E). Interestingly, 11% of RA-induced ATAC-seq peaks corresponded to RAR binding sites identified by ChIP-seq and possess the DR5 motif recognized by the RAR/RXR complex (Fig. 6D). This was the case for RA-induced peaks in the dhrs3a, cyp26a1/b1 and insm1b genes (see black boxes in Fig. S6 and data not shown). However, many identified RAR sites did not display a significant increase in accessibility, such as those located in the hoxba locus (see green boxes in Fig. S6). Furthermore, the majority of RA-induced ATAC-seq peaks did not harbour RAR binding sites although they were usually found near RA-upregulated genes; instead, such peaks often harbour Gata or Hnf1b binding motifs (38% and 13% of all RA-induced elements, respectively; Fig. 6D). Thus, it can reasonably be assumed that these genes are indirectly stimulated by RA signalling. For example, the pancreatic regulatory genes pdx1, insm1a, rfx6 or neurod1, all upregulated by RA-treatment but devoid of RAR binding sites, had RA-induced ATAC-seq peaks located in their genomic neighbourhood and contained either Hnf1b or Gata binding motifs in such peaks (Fig. S7 and data not shown). To identify the GATA and Hnf1b family members involved in these indirect RA-regulations, we searched in the list of RA-stimulated genes (Table S2) for these 2 types of transcription factors and we also determined if RAR sites are detected in their genomic loci (Tables S4 and S5).

hnf1ba and hnf1bb were both induced by RA but the level of expression of hnf1ba was about 200-fold higher than hnf1bb in endodermal cells. Furthermore, hnf1ba contains three RAR binding sites located in evolutionary-conserved and nucleosome-free regions (Fig. 7A), while hnf1bb has only one weak RAR site (data not shown). Out of the 10 members of the GATA family, only gata4 and gata6 were significantly induced by RA but gata6 was expressed in the endoderm at about 40-fold higher level compared to gata4. Furthermore, gata6 has a high affinity RAR binding site in a nucleosome-free region located about 30 kb upstream from its TSS (Fig. 7B) while no such RAR site was present around the gata4 gene locus (data not shown). Thus, all these analyses strongly suggest that, upon RA induction, RAR–RXR complexes directly activate expression of Gata6 and Hnf1ba, which in turn will open regulatory chromatin regions of many genes, including those coding for several pancreatic regulatory factors.

Discussion
In this study, we have used an integrative genomic approach to identify the genes regulated directly and indirectly by RA signalling in the endoderm at the end of gastrulation and notably involved in the specification of pancreatic progenitors. RNA-seq data performed on FACs-selected endodermal cells identified genes with enriched expression in endoderm and regulated by RA signalling. This gene set includes many genes previously shown to be regulated by RA in other germ layers like hox, nr2f, cyp26a1, dhrs3 or hnf1b genes, indicating that the regulatory network triggered by RA is shared, at least partially, between the three germ layers. Analysis of the zebrafish RAR cistrome at the end of gastrulation confirms hox genes as direct RAR targets, acting on the antero-posterior patterning of the endoderm, but the data also pinpoint 4 direct RAR target genes (i.e. mnx1, hnf1ba, gata6 and insm1b) known to play a crucial role in pancreas development. Analysis of the nucleosome-free

![Figure 5.](https://www.nature.com/scientificreports/) The conserved RAR site from skap1 gene 4th intron is a functional RARE. Pictures of the DR5-RAR-skap1::GFP transgenic embryos treated with DMSO (control, panel A), with RA (panel B) or with BMS493 (panel C). Upper panels display GFP fluorescent expression and lower panels (A’, B’ and C’) display embryo morphology. GFP Expression in the posterior endoderm gut is indicated by yellow arrows.
regions upon RA/BMS493 treatments reveals that, in addition to RAR, Gata and Hnf1b factors are the major transcription factors driving the increase in chromatin accessibility at numerous regulatory regions in endodermal cells. All together, these data indicate that Gata6 and Hnf1ba, which are directly regulated by RAR, activate in a second step the expression of many genes. Previous studies on Xenopus embryos showed the role of Ndrg1, Hnf1b and Fzd4 downstream of RA signalling for the specification of pancreatic field\textsuperscript{18, 26}. Our study confirms the conserved role of hnf1ba in zebrafish endoderm; in contrast, the zebrafish fzd4 ortholog was not found to be regulated by RA in endoderm nor other fzd paralogs. As for the zebrafish ndrg1a gene, it is induced by RA in zebrafish embryos but does not harbour any RAR site in its genomic loci revealing an indirect regulation by RA signalling in zebrafish.

The treatment of embryos with exogenous RA leads to a similar number of up- and down-regulated genes; in contrast the treatment with the pan-RAR inverse-agonist BMS493 mostly causes down-regulation of genes. Such difference can be explained by the mode of action of BMS493 which stabilizes the association of RAR with its co-repressors NcoR/Smrt\textsuperscript{19, 28}. In contrast, RA causes mainly a release of co-repressors and the recruitment of co-activators. Most RA-downregulated genes detected in the present study likely represent indirect RAR targets such as genes expressed in anterior endodermal cells whose fate is modified upon RA treatment. This could be the case for the nkx2.1, nkx2.3, nkx2.7 and tbx1 genes known to be expressed in the anterior endoderm and which are downregulated upon RA treatment and do not contain any RAR site in their genomic loci. We also observed
that the number of genes regulated by BMS493 was much lower compared to the number of RA-regulated genes. This raises the concern that the RAR activity could not be totally blocked by the BMS493 treatments. However, at the BMS493 dose used in this study (1 µM), most of the RAR activity should be blocked according to previous dose–response studies. The action of BMS493 as inverse-agonist should repress all direct target genes. Such difference in the number of BMS493- versus RA-regulated genes can be explained by considering that a large proportion of RAR direct target genes are still in a repressed state in the (DMSO-treated) control embryos. This could be due to the low proportion of endodermal cells receiving the endogenous RA signal. Indeed, BMS493 will affect only the cells under endogenous RA signalling (i.e. mid-trunk) while ectopic RA treatment affects all the other endodermal cells. In addition, a down-regulation is more difficult to detect for the genes having a low basal expression level or for genes expressed in a few endodermal cells. Another study comparing gene expression upon agonist (RA) versus inverse-agonist (AGN193109) also revealed less genes affected by the inverse-agonist, likely for the same reasons.

In our study, we also determined the RAR cistrome at the end of zebrafish gastrulation to identify RAR direct targets. Comparison of the zebrafish RAR cistrome with the RAR cistrome of murine embryonic F9 cells identified 722 out of 2144 genes harbouring RAR sites in both zebrafish and mice, among which are several Hox, Cyp26a1, Meis and Nr2f genes, as well as the 4 pancreatic regulatory genes Mnx1, Gata6, Hnf1b and Insul1. This conservation is further supported by the location of many RAR sites in CNEs revealing an evolutionary constraint.

Figure 7. Location of RARα binding sites and of nucleosome-free regions in the hnf1ba (panel A) and gata6 (panel B) gene loci. Visualization of ATAC-seq reads from the merged 3 replicates obtained from endoderm treated with BMS493, DMSO or RA and from non-endodermal cells (NE), in addition to tracks showing RARα binding sites, H3K27Ac and H3K3me3 marks determined by ChIP-seq. Regions showing conservation of genomic sequences among 5 fish species are also shown by blue boxes below the tracks. The RAR binding sites are highlighted by gold boxes.
to maintain these RAR sites. However, the majority of the zebrafish and murine RAR binding sequences could not be aligned. For example, the two RAR sites located near the zebrafish hnf1ba and the murine Hnf1b genes are both located in fish and mammalian CNEs, respectively; these two RAR sites have also similar location in the murine and zebrafish genes (i.e. in the 4th intron and downstream the Hnf1b gene) but the zebrafish and murine RAR site sequences are too divergent to be aligned. We can nevertheless reasonably assume that these RAR sites have a common origin and function.

Interestingly, 24 RAR sites were found in extremely conserved non-coding regions and sequences could be aligned from zebrafish to human. This is the case for RAR sites located in some hox clusters, near meis25, nrpia1, ncoa3 and in the skap1 gene (this study). Such high conservation suggests that the level of expression of these RA target genes must be precisely controlled. We show in the present study that the highly conserved RAR site in the 4th intron of the zebrafish skap1 gene drives expression in the neural tube and gut in a pattern reminiscent to hoxbb1 expression, suggesting that this RARE controls the expression of the neighbouring hoxbb cluster. This is further supported by recent chromosome conformation capture (Hi-C) data obtained from adult zebrafish muscle tissue; mining these data show that the 4th intron skap1 CNE interacts with several hoxbb genes (not shown). Thus, this indicates that hox regulation by RA involved additional RAR sites located very far downstream from the Hoxb cluster (i.e. more than 200 kb in the murine and human genome). This observation also reveals that the assignment of RAR sites to the closest neighbour could result in some incorrect links between RAR sites and their putative target genes. To improve RAR-gene assignment, ChIP-seq should ideally be combined with HiC on zebrafish gastrula. Still, we think that our strategy enabled to accurately define the majority of RAR-gene assignments as revealed by the correlation found between RA gene induction and the number of nearby RAR sites.

Our study also highlights that RA modifies the accessibility of thousands of genomic regions in zebrafish endodermal cells. The modest effect of BMS493 on chromatin is probably due to the small proportion of endodermal cells under the influence of endogenous RA, as already discussed above for the RNA-seq data. Interestingly, about 10% of the sequences showing increased accessibility upon RA treatment correspond to RAR binding sites, while no RA site shows reduced accessibility upon RA treatment. This observation supports the classical model where, upon RA binding, the RAR–RXR complex interacts with coactivators recruiting HAT which open chromatin structure. Still, many RAR sites display the same chromatin accessibility despite strong activation of nearby genes by RA treatment. In such case, the RAR–RXR complex is often bound to a nucleosome-free region, even without RA signalling (i.e. upon DMSO- or BMS493-treatments), indicating that RAR–RXR regulates transcription through a mechanism other than changing chromatin structure. The increase of accessibility at 10% of RAR sites still suggests that the RAR–RXR complex contributes to the destabilization of nucleosomal structure at these sites and could act with (or as) pioneering transcription factors.

About 90% of chromatin regions with increased accessibility upon RA treatment do not harbour RAR binding sites but rather contain GATA or HNF1b binding motifs. Both in zebrafish endoderm and in murine embryonic F9 cells, Gata6 and Hnf1b genes harbour RAR sites and are strongly induced by RA. Thus, these two factors seem to act as important effectors of RA signalling by opening chromatin at numerous regulatory regions. Gata6 and Hnf1b have previously been shown to be involved in pancreas development and GATA factors have also been shown to act as pioneering factors. Whether Hnf1b has an intrinsic ability to bind nucleosomal DNA and to open chromatin or whether this requires the involvement of other co-factors remain to be determined.

By determining the transcriptomic landscape and genome-wide chromatin accessibility of endodermal and non-endodermal cells at the end of gastrulation, the present study also provide the lists of genes and regulatory cis-elements specifically active in the endoderm. For example, the RA site located downstream from the mnx1 gene is located in a region which is nucleosome-free only in the endoderm. Accordingly, our transgenic reporter assay shows that this region drives expression in the posterior endoderm and in the dorsal pancreas. Other endoderm-specific ATAC-seq peaks showed similar endoderm-specific activity (data not shown). Thus, our ATAC-seq data and RNA-seq data allows the genome-wide identification of endoderm specific enhancers. This strategy can also help to assign a regulatory region to its target gene by correlating expression and chromatin structure; the accuracy of such analyses can be increased by including more data from different cell types or from different developmental stages.

In conclusion, the present study provides interesting resources not only to decipher the regulatory network triggered by RA but also to unveil regulatory regions involved in endoderm formation and patterning in zebrafish. Our data confirm the reassignment of anterior endoderm into midtrunk endoderm by RA treatment, as revealed by the downregulation of anterior markers such as nkx2.1, nkx2.3 and nkx2.7 and the up-regulation of most features triggered by RA but also to unveil regulatory regions involved in endoderm formation and patterning in zebrafish.

Materials and methods
Zebrafish strains, pharmacological treatments and FACS sample preparation. 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). The efficiency of pharmacological treatments was tested on Tg(pax6:GFP) embryos, allowing a direct visualisation of pancreatic GFP+ cells in...
living transgenic embryos. These experiments demonstrated the absence of pancreatic cells at 1 µM BMS493 while the most efficient induction of ectopic pancreatic cells in anterior endoderm were observed at 1 µM RA as previously reported. No ectopic pancreatic cells were observed at 25 or 50 nM RA, concentrations usually used in cell culture. The requirement of high RA concentration in zebrafish is probably due to the degradation of exogenous RA by the endogenous Cyp26 enzyme whose expression is induced by RA. Endodermal and non-endodermal cells were obtained using the transgenic line Tg(sox17:GFP). Tg(sox17:GFP) transgenic embryos were incubated with DMSO (0.1% volume) as control, 1 µM RA or 1 µM BMS493 from 1.25 to 11 hpf. Embryos were used either directly at 11 hpf (3-somite stage) or either washed in embryos medium, raised until 13 hpf (8-somite stage) in embryos medium. Embryos were dechorionated, deyolked, dissociated in FACS solution (HBSS without Ca²⁺ and Mg²⁺ and containing 1% BSA) by gentle pipetting and cells were directly sorted by FACS Aria II based on the GFP expression. Sorting was performed as a single run in “purity mode” and three replicates were prepared for each treatment.

**RNA-seq library preparation and data analysis.** cDNA was obtained using the SMART-seq2 protocol as recently described. Briefly, each cell preparation was pelleted and lysed in Lysis Buffer by freezing in liquid nitrogen and stored at − 80 °C. Synthesis of cDNA was performed directly on the lysed cells; cDNA were amplified by 10 PCR cycles and purified before assessing their quality on the bioanalyzer (2100 high sensitivity DNA assay, Agilent Technologies). One hundred and fifty pg of cDNA were used as input to prepare the libraries using the Nextera XT DNA kit (Illumina). Seventy five bp single-end sequences were obtained using the NextSeq500 Illumina Sequencer with coverage of about 20 million reads per library.

Raw reads were aligned to the zebrafish genome (Zv9, Ensembl genome version 79, ensembl.org) using STAR software. Normalization and differential expression analysis were performed using DESeq2. Genes were considered differentially expressed with FDR < 0.01 (False Discovery Rate).

**Expression of RARaa-myc in zebrafish embryos.** The whole zebrafish RARaa coding sequence was amplified by RT-PCR and inserted into the pCS2MT vector in frame with Myc-epitope tags located at the C-terminal end. The Myc-RARaa mRNA was synthesized from this plasmid by in vitro transcription (mMESSAGE mMACHINE sp6 transcription kit, Invitrogen) and injected into fertilized zebrafish eggs. Validation of Myc-tagged RARaa protein expression was done by western-blotting on cytosolic and nuclear lysate fractions from injected embryos using a ChIP-grade anti-myc antibody (ab9132, Abcam).

**ChIP-seq library preparation and data analysis.** About 0.5 mL of myc-tagged RARaa mRNA was injected in zebrafish fertilized eggs at a concentration of 70 ng/µL and embryos were incubated until they reached bud stage (10.5 hpf, end of gastrulation). RA was added to the medium at that stage for inducing RAR binding (Chatagnon et al. 2015) and embryos were grown for an extra hour. Around 2000 injected embryos were used for chromatin immunoprecipitation essentially as previously described and fixed with 1.85% PFA (Diagenode Bioruptor sonicator was used for chromatin shearing). Dynal Protein A magnetic beads (Diagenone) and ChIP-graded anti-myc antibody (ab9132, Abcam) were used for the immunoprecipitation, an aliquot of chromatin being taken before the immunoprecipitation step and used as negative control. Libraries were prepared with the NEBNext Ultra II DNA Library Prep kit (Bioke). Forty two bp pair-end sequences were obtained using the NextSeq500 Illumina Sequencer with coverage of about 20 million reads per library.

**ATAC-seq library preparation and data analysis.** After isolation by FACS, cells were pelleted and ATAC-Seq libraries were prepared as described elsewhere. Libraries were sequenced 42 bp paired end using the NextSeq500 Illumina Sequencer with coverage of 40 million reads per library. Three replicates were prepared for each condition.

**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 start 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.
**References**

1. Rhinn, M. & Dollé, P. Retinoic acid signalling during development. *Development* 139, 843–858 (2012).
2. Niederreither, K. & Dollé, P. Retinoic acid in development: Towards an integrated view. *Nat. Rev. Genet.* 9, 541–553 (2008).
3. Ghyselinck, N. B. & Duester, G. Retinoic acid signaling pathways. *Development* 146, dev167502 (2019).
4. Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S. & Miyawaki, A. Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature* 496, 363–366 (2013).
5. Larsen, H. L. & Grapin-Botton, A. The molecular and morphogenetic basis of pancreas organogenesis. *Semin. Cell Dev. Biol.* 66, 51–68 (2017).
6. Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Am. J. Anat.* 232, 284–298 (2013).
7. Stafford, D. & Prince, V. E. Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development. *Curr. Biol.* 12, 1215–1220 (2002).
8. Stafford, D., Hornbruch, A., Mueller, P. R. & Prince, V. E. A conserved role for retinoid signaling in vertebrate pancreas development. *Dev. Genes Evol.* 214, 432–441 (2004).
9. Bayha, E., Jørgensen, M. C., Serup, P. & Grapin-Botton, A. Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis. *PLoS One* 4, e5845 (2009).
10. Chen, Y. et al. Retinoic acid signaling is essential for pancreas development and promotes endoderm at the expense of exocrine cell differentiation in Xenopus. *Dev. Biol.* 271, 144–160 (2004).
11. Molotkov, A., Molotkova, N. & Duester, G. Retinoic acid generated by Raldh2 in mesoderm is required for mouse dorsal endodermal pancreatic development. *Dev. Dyn.* 232, 950–957 (2005).
12. Martín, M. et al. Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. *Dev. Biol.* 284, 399–411 (2005).
13. Pagliuca, F. W. et al. Generation of functional human pancreatic β cells in vitro. *Cell* 159, 428–439 (2014).
14. Stafford, D. et al. Retinoids signal directly to zebrafish endoderm to specify insulin-expressing beta-cells. *Development* 133, 949–956 (2006).
15. Song, J., Kim, H. J., Gong, Z., Liu, N.-A. & Lin, S. Vhnf1 acts downstream of Bmp, Fgf, and RA signals to regulate endocrine beta cell development in zebrafish. *Dev. Biol.* 303, 561–575 (2007).
16. Hernandez, R. E. et al. Vhnf1 integrates global RA patterning and local FGF signals to direct posterior hindbrain development in zebrafish. *Development* 131, 4511–4520 (2004).
17. Nolte, C., De Kumor, B. & Krumlauf, R. Hox genes: Downstream “effectors” of retinoic acid signaling in vertebrate embryogenesis. *Genes Dev.* 5, 1–17 (1999).
18. Gere-Becker, M. B., Pommerenke, C., Lingner, T. & Pieler, T. Retinoic acid-induced expression of Hnf1b and Fzd4 is required for pancreas development in Xenopus laevis. *Development* 145, dev161372 (2018).
19. Chatagnon, A. et al. Retinoid binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. *Nucleic Acids Res.* 43, 4833–4854 (2015).
20. Paschaki, M. et al. Transcriptomic analysis of murine embryos lacking endogenous retinoic acid signaling. *PLoS One* 8, e62274 (2013).
21. Arima, K. et al. Global analysis of RAR-responsive genes in the Xenopus neurula using cDNA microarrays. *Dev. Dyn.* 232, 414–431 (2005).
22. Samarut, E. et al. Retinoic acid receptor subtype-specific transcriptotypes in the early zebrafish embryo. *Mol. Endocrinol.* 28, 260–272 (2014).
23. Feng, L., Hernandez, R. E., Waxman, J. S., Yelon, D. & Moens, C. B. Dhrs3a regulates retinoic acid biosynthesis through a feedback inhibition mechanism. *Dev. Biol.* 338, 1–14 (2010).
24. Cunningham, T. J., Colas, A. & Duester, G. Early molecular events during retinoic acid induced differentiation of neuromesodermal progenitors. *Biomed. Open* 5, 1821–1833 (2016).
25. Berenguer, M., Meyer, K. F., Yin, J. & Duester, G. Discovery of genes required for body axis and limb formation by global identification of retinoic acid-regulated epigenetic marks. *PLoS Biol.* 18, e3000719 (2020).
26. Zhang, T., Guo, X. & Chen, Y. Retinoic acid-activated Ndrg1a represses Wnt/β-catenin signaling to allow Xenopus pancreas, oesophagus, stomach, and duodenum specification. *PLoS One* 8, e65058 (2013).
27. White, R. J., Nie, Q., Lander, A. D. & Schilling, T. F. Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo. *PLoS Biol.* 5, e304 (2007).
28. Schilling, T. E., Nie, Q. & Lander, A. D. Dynamics and precision in retinoic acid morphogen gradients. *Curr. Opin. Genet. Dev.* **22**, 562–569 (2012).
29. Kinkel, M. D. et al. Cyp26 enzymes function in endoderm to regulate pancreatic field size. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7864–7869 (2009).
30. Delacroux, L. et al. Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. *Mol. Cell. Biol.* **30**, 231–244 (2010).
31. Mahony, S. et al. Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* **12**, R2 (2011).
32. Mouttier, E. et al. Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *J. Biol. Chem.* **287**, 26328 (2012).
33. Buenrostro, J. D., Gresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
34. Hiller, M. et al. Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: Application to zebrasfish. *Nucleic Acids Res.* **41**, e151 (2013).
35. Quillien, A. et al. Robust identification of developmentally active endothelial enhancers in zebrasfish using FANS-assisted ATAC-Seq. *Cell Rep.* **20**, 709–720 (2017).
36. Puik, E. J. et al. A Cdx4-Sall4 regulatory module controls the transition from mesoderm formation to embryonic hematopoiesis. *Stem Cell Rep.* **1**, 425–436 (2013).
37. Love, C. E. & Prince, V. E. Expression and retinoic acid regulation of the zebrasfish nr2f orphan nuclear receptor genes. *Dev. Dyn.* **241**, 1603–1615 (2012).
38. Germain, P. et al. Differential action on coregulator interaction defines inverse retinoid agonists and neutral antagonists. *Chem. Biol.* **16**, 479–489 (2009).
39. Nadendla, E. et al. An unexpected mode of binding defines BMS948 as a full retinoic acid receptor β (RARβ, RNR1B2) selective agonist. *PloS One* **10**, e0123195 (2015).
40. Yang, H. et al. A map of cis-regulatory elements and 3D genome structures in zebrasfish. *Nature* **588**, 337–343 (2020).
41. Zaret, K. S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
42. Sun, Z. & Hopkins, N. vhnf1, the MOD5 and familial GCKD-associated gene, regulates regional specification of the zebrafish gut, proepiblasts, and hindbrain. *Genes Dev.* **15**, 3217–3229 (2001).
43. Haumaître, C. et al. Lack of Tcf21/VHNF1 in mice leads to pancreas agenesis. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1490–1495 (2005).
44. Carrasco, M., Delgado, L., Soria, B., Martin, F. & Rojas, A. GATA4 and GATA6 control mouse pancreas organogenesis. *J. Clin. Investig.* **122**, 3504–3515 (2012).
45. Allen, H. L. et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat. Genet.* **44**, 20–22 (2011).
46. Bossard, P. & Zaret, K. S. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125**, 4909–4917 (1998).
47. Mizoguchi, T. et al. Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* **135**, 2521–2529 (2008).
48. Lawongwe, A. et al. Pancreatic and intestinal endocrine cells in zebrasfish share common transcriptional signatures and regulatory programmes. *BMC Biol.* **18**, 109 (2020).
49. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 530 (2014).
51. Morley, R. H. et al. A gene regulatory network directed by zebrasfish No tail accounts for its roles in mesoderm formation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3829–3834 (2009).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **10**, 357–359 (2012).
53. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
54. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
55. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform.* **11**, 237 (2010).
56. Bossard, P. & Zaret, K. S. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125**, 4909–4917 (1998).
57. Mizoguchi, T. et al. Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* **135**, 2521–2529 (2008).
58. Lawongwe, A. et al. Pancreatic and intestinal endocrine cells in zebrasfish share common transcriptional signatures and regulatory programmes. *BMC Biol.* **18**, 109 (2020).
59. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 530 (2014).
61. Morley, R. H. et al. A gene regulatory network directed by zebrasfish No tail accounts for its roles in mesoderm formation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3829–3834 (2009).
62. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **10**, 357–359 (2012).
63. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
64. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
65. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPpeakAnno: A Bioconductor package to annotate ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
66. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
67. Ou, J. et al. ATACseqQC: A bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* **19**, 169 (2018).
68. Zhu, L. et al. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform.* **11**, 237 (2010).
69. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
70. Fisher, S. et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrasfish. *Nat. Protoc.* **1**, 1297–1305 (2006).

**Acknowledgements**

We thank the following technical platforms of GIGA from Liège University: GIGA-Zebrafish (H Pendeville), GIGA-Cell Imaging and Flow Cytometry platform (S Ormenese and S Raafat), GIGA-Genotranscriptomic (W Coppieters, and L Karim). We are indebted to Virginie Vonberg for technical help in transgene constructions and for zebrasfish egg injections. We also thank Fabio D’Orazio and Nan Li for assistance with the ATAC-seq protocol. We acknowledge all members of the ZENCODE-ITN project for helpful discussions and feedback.

**Author contributions**

A.R.L.-P. carried out the RNA-seq, ChIP-seq and ATAC-seq experiments and performed the bioinformatic analyses with help from P.B. and B.L. F.M. and F.W helped in the setting of ATAC and ChIP experiments and for funding acquisition. B.P., A.R.L.-P., M.L.V. and I.M. conceived the study, interpreted the data and wrote the manuscript. All authors read, edited and approved the final manuscript.

**Funding**

This work was supported by ZENCODE-ITN European project 643062 (A.R.L., B.L., F.M., F.C.W. and B.P.) and by the Léon Frederic fund (A.R.L. and B.P.). B.P., I.M., and M.L.V. are associate researchers from FRS/FNRS (Fonds National pour la Recherche Scientifique).
Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-02039-y.
Correspondence and requests for materials should be addressed to B.P.
Reprints and permissions information is available at www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021