8772–8787 *Nucleic Acids Research, 2018, Vol. 46, No. 17* *Published online 27 August 2018*
doi: 10.1093/nar/gky771

The Xenopus animal cap transcriptome: building a mucociliary epithelium

Alessandro Angerilli[1], Pawel Smialowski[2,3] and Ralph A.W. Rupp[1,*]

[1]Molecular Biology Division, Biomedical Center, Ludwig-Maximilians-University München, D-82152 Martinsried, Germany, [2]Bioinformatic Core Facility, Biomedical Center, Ludwig-Maximilians-University München, D-82152 Martinsried, Germany and [3]Helmholtz Zentrum München, Institute of Stem Cell Research, Ingolstädter Landstraße 1, D-85764 Neuherberg-München, Germany

Received May 11, 2018; Revised August 09, 2018; Editorial Decision August 13, 2018; Accepted August 24, 2018

ABSTRACT

With the availability of deep RNA sequencing, model organisms such as *Xenopus* offer an outstanding opportunity to investigate the genetic basis of vertebrate organ formation from its embryonic beginnings. Here we investigate dynamics of the RNA landscape during formation of the *Xenopus tropicalis* larval epidermis. Differentiation of non-neural ectoderm starts at gastrulation and takes about one day to produce a functional mucociliary epithelium, highly related to the one in human airways. To obtain RNA expression data, uncontaminated by non-epidermal tissues of the embryo, we use prospective ectodermal explants called Animal Caps (ACs), which differentiate autonomously into a ciliated epidermis. Their global transcriptome is investigated at three key timepoints, with a cumulative sequencing depth of ∼10^8 reads per developmental stage. This database is provided as online Web Tool to the scientific community. In this paper, we report on global changes in gene expression, an unanticipated diversity of mRNA splicing isoforms, expression patterns of repetitive DNA Elements, and the complexity of circular RNAs during this process. Computationally we derive transcription factor hubs from this data set, which may help in the future to define novel genetic drivers of epidermal differentiation in vertebrates.

INTRODUCTION

Tight control of gene transcription is essential for proper embryonic development. Destabilization of gene regulatory networks in the embryo is a common cause of disease and teratogenesis. Several transcriptomic studies have shed light on the dynamics of embryonic gene expression in the Western Clawed frog Xenopus tropicalis (1–5). While these studies have boosted our comprehension of the timing of genome activation, subsequent waves of zygotic transcription and regionalization of gene expression during gastrulation, many of them fall short on spatial information or face the increasing cellular complexity of the embryo, which makes it difficult to define the full extent of RNA regulation associated with the differentiation of individual tissues. Pluripotent stem cell-derived organoids (6,7) provide a solution to this problem. In Xenopus, Animal Caps represent a primary organoid culture system with remarkable properties. ACs consist of the central part of the blastocoel roof, explanted from embryos at the mid to late blastula stage, when the explants still consist of uncommitted embryonic cells. Their superficial cell layer forms an epithelium, the attached deep layer cells have a mesenchymal character. In isolation the explants round up into spheroid structures and survive with their maternally provided nutrients until tadpole stages. As shown by numerous studies since the 1980s, growth factors can induce in ACs distinct cell types from all germ layers, as well as the formation of complex organoids (8–10). To facilitate homogenous access for growth factor signals, ACs can be dissociated into single cell suspension by removing calcium from saline-based buffers. Adding back calcium leads to reaggregation of cells into spheres, which is necessary for subsequent differentiation. Interestingly, transient dissociation of ACs without exogenous inducing signals is sufficient to neuralize the explants, while intact animal caps form a mucociliary epidermis by default (11). Neural and non-neural ectoderm in the embryo experience different levels of BMP signalling (12). Cell dissociation of ACs is thought to disperse intrinsic BMP proteins, which promote the formation of mucociliary epithelia in vertebrates (13), thereby mimicking neural induction through BMP inhibitors, normally secreted by Spemanns organizer (12). Despite its similarity to the embryonic epidermis, it is important to note that the mucociliary epithelium produced from ACs is referred to as 'atypical epidermis', because neither its morphogenesis nor the stoichiometry of the different cell types have been quantitatively compared with the embryonic counterpart. Any difference in these aspects could in principle affect gene expression.

*To whom correspondence should be addressed. Tel: +49 89 2180 75438; Email: ralph.rupp@bmc.med.lmu.de

The formation of the Xenopus larval epidermis is of biomedical interest, since it represents a mucociliary epithelium similar to the human respiratory tract (14–17). It consists of merely five major cell types—goblet cells, representing the outer epithelial layer, with interspersed multiciliated cells (MCC), ionocytes (ICs) and small secretory cells (SSC). The sensorial layer of the epidermis contains a fifth cell type, characterised by expression of *tp63*. While *tp63*-positive basal cells are adult stem cells in the mammalian airway epithelium (18), the function of this fifth cell type in Xenopus is unclear (13). The latter four cell types are specified by BMP and Notch signalling in the sensorial layer of the epidermis (13,19). From neurulation on, MCC, IC and SSC precursor cells intercalate into the outer cell laye (13,20–23). Pioneering work in Xenopus has identified key regulatory factors controlling the differentiation of these cell types, but a genome-wide survey of the transcriptional changes accompanying the formation of the larval epidermis is missing up to date. Here we approach this goal by following the differentiation of unmanipulated ACs with deep RNA-Sequencing.

MATERIALS AND METHODS

Ethics statement

Animal work has been conducted in accordance with German Tierschutzgesetz. Production of Xenopus embryos by *in vitro* fertilization has been licensed by the Government of Oberbayern (Projekt/AZ ROB: 55.2.1.54-2532.6-3-11).

Embryos handling and AC preparation

Xenopus tropicalis eggs were collected, in vitro fertilized and handled as described by Showell and Conlon (24). ACs were manually dissected as described for *Xenopus laevis* by Sive *et al.* (25). Embryos were staged accordingly to Nieuwkoop and Faber (26). ACs were staged accordingly to sibling embryos. For the library preparation, approximately 30 ACs were collected for each sample, vortexed in TRIzol (Ambion) until complete dissolution and snap-frozen in liquid nitrogen. Each developmental stage is represented by 3 biological replicates. Each biological replicate derives from a different mating pair. For RNA *in situ* hybridization analysis, embryos were anesthesized in 0.05% Tricaine, fixed in MEMFA (0.1 M MOPS 2 mM EGTA 1 mM $MgSO_4$ 3.7% Formaldehyde) and stored in ethanol.

RNA *in situ* hybridization and immunocytochemistry

Antisense RNA probe sequences for analysis of splicing isoforms and repetitive DNA elements were selected from top entries in Supplementary Tables S5 and S7. Embryonic RNA from mixed developmental stages (NF10-24) was extracted as described above. Total RNA was then reverse transcribed using SuperScript III (Invitrogen) and random priming. The primers used to amplify the targets present overhang homologies to the pCS2+ vector. Primer sequences are detailed in the Supplementary Data (Supplementary Table S11). The resulting amplicons were gel purified and cloned into pCS2+ using the Gibson Assembly Method (27). Otogelin (otog), ATP6v1e1 and epidermal keratin (xk81a1) plasmids were kindly provided by Drs. Axel Schweickert (Universität Hohenheim) and Edgar Pera (Lund University). The plasmids containing BMP7.2 RNA 13254, BMP7.2 RNA 13257, CPNE1 RNA 0071, CPNE1 RNA 0068, ERV1-4-LTR_Xt, LTRX1-LTR_Xt and Rem2b were linearized using HindIII, Otogelin and XK81A1 with EcoRI, for EVPL X1 we used ClaI, while for XV-ATPase we used SacII. The linearized plasmids were subsequently transcribed overnight using T7 (antisense probe) and Sp6 (sense control) RNA Polymerases (Promega) and cleaned-up with RNeasy Mini-Kit (Qiagen). RNA in situ hybridizations were performed as in Hollemann *et al* (28). Immunocytochemistry was performed followed the protocol of Sive *et al.* (25). Monoclonal Anti-Acetylated Tubulin antibody (Sigma-Aldrich, T6793) was used with a dilution of 1:1000. Bleaching of pigment granules was achieved after the AP-staining reaction by incubating the dehydrated embryos in bleaching solution (1% hydrogen peroxide, 5% Formamide, 0.5% SSC) on a visible-light source, followed by re-fixation in MEMFA. Embryos were photographed using a Leica M205FA stereomicroscope. Pictures were obtained as Z-stack acquisitions.

RNA library preparation

Total RNA was extracted from the animal caps using TRIzol (Ambion) and phenol/chloroform extraction. On-column RNA clean-up, including DNA digestion, was performed using RNeasy Mini-Kit (Qiagen). RNA quality was checked on Agilent Bioanalyzer 2100 using Agilent RNA 6000 Pico Kit. Ribosomal RNA was depleted using Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat) from Illumina. The resulting Ribosomal-depleted RNA was cleaned and concentrated using RNA Clean & Concentrator™-5 from Zymo Research. Ribosomal depletion was verified using again the Agilent Bioanalyzer 2100. Libraries were prepared using NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina (New England Biolabs). Indexes and adaptors come from NEBNext Multiplex Oligos for Illumina from primer sets 1 and 2. DNA purification and size selection was performed using Agencourt AMPure XP from Beckman Coulter. Quality and size of the libraries was verified on the Agilent Bioanalyzer using Agilent DNA 1000 Kit.

Detection of circular RNAs

CircRNAs were identified with the method described by Westholm *et al.* (29). We used diluted RNA-Seq libraries as template for PCR-validation of circRNAs. Outward- and inward-facing Primer sets are listed in Supplementary Table S11. The resulting amplicons were visualized using Agilent Bioanalyzer with Agilent DNA 1000 Kit. In order to reach an appropriate length for DNA-sequencing a second PCR with primers containing SP6 promoter and pCS2+ derived backbone DNA sequences was performed for selected candidates. Sequencing of circRNA PCR products was conducted at MWG Eurofins.

RNA-sequencing and data analysis

Next generation sequencing of AC RNA-derived cDNA libraries (triplicates from three time points) was performed

on the Illumina HiSeq4000 platform at BGI (Hong Kong). At least 30 million reads per sample were generated. All sequencing data were aligned to version 9.0 of *X. tropicalis* genome as provided by Xenbase (30) using STAR algorithm version 2.4.2a (31). To calculate expression of splicing isoforms we used RSEM version 1.3.0 and NCBI *X. tropicalis* annotation release 103, which relies on genome version 9.1. Expression differences between time points and reference samples were analysed by the DESeq2 package version 1.18 (32). Analysis of repetitive elements was based on entries in RepBase version 19.10 (33).

Putative lead transcription factors for epidermal differentiation were assessed by an approach adopted from Rackham *et al.* (34). To derive rank lists of transcription factors associated with change of transcript abundance in development we analysed RNA-seq results from pairwise comparisons (e.g. NF10 versus NF16) for differential expression by DESeq2. Results were filtered for transcription factors using a list from Blitz *et al.* (35). For each transcription factor we established a sphere of influence of up to three level depth using inter-gene relation scores based on STRING database version 10.5 (36). We considered only relations with total scores >300, where less then half of the value was attributed to text mining. Strength of influence and total score of each factor were calculated using DESeq2 logarithmic fold enrichment and adjusted *P*-values as described by Rackham *et al.* (32). To visualize results selected transcription factors were plotted together with closest STRING network nodes using cytoscape 3.6 (37) and stringApp1.2.1.

RESULTS

Epidermal gene expression profiles

We explanted the animal caps manually at blastula stage (NF8-9) and extracted RNA for transcriptome analysis at three time-points (Figure 1A), representing key events in epidermal differentiation. At early gastrula (NF10.5), AC cells have become committed to the epidermal fate (38,39). At mid neural fold (NF16), Notch/Delta signalling has specified multiciliated and inocyte cell fates in the deep sensorial layer and precursor cells are intercalating into the superficial layer of the epidermis (15,20). At the early tailbud stage (NF24), a mucociliary epithelium consisting of actively secreting goblet cells, ionocytes, and functional multiciliated cells (MCC) has formed (17). A fourth cell type, which secrets serotonin and provides anti-bacterial protection, appears one day later (21,22).

As shown in Figure 1B, acetylated alpha-tubulin protein (MCCs), as well as mRNAs for *otogelin* (goblet and small secretory cells), *atp6v1e1* (ionocytes), keratin *xk81a1* (pan-epidermal) and *tp63* were expressed in similar patterns in embryos and ACs. Moreover, ACs rotated autonomously within the culture dish and displace dye from their surface, both indications for coordinated and directional ciliary beating, similar to embryos (see Supplementary Data, Movies 1 and 2). The epidermal character of the ACs was also assessed by looking at non-epidermal genes. The expression of selected pluripotency associated (Supplementary Figure S1A), mesendodermal (Supplementary Figure S1B) and ectodermal genes (Supplementary Figure S1C)

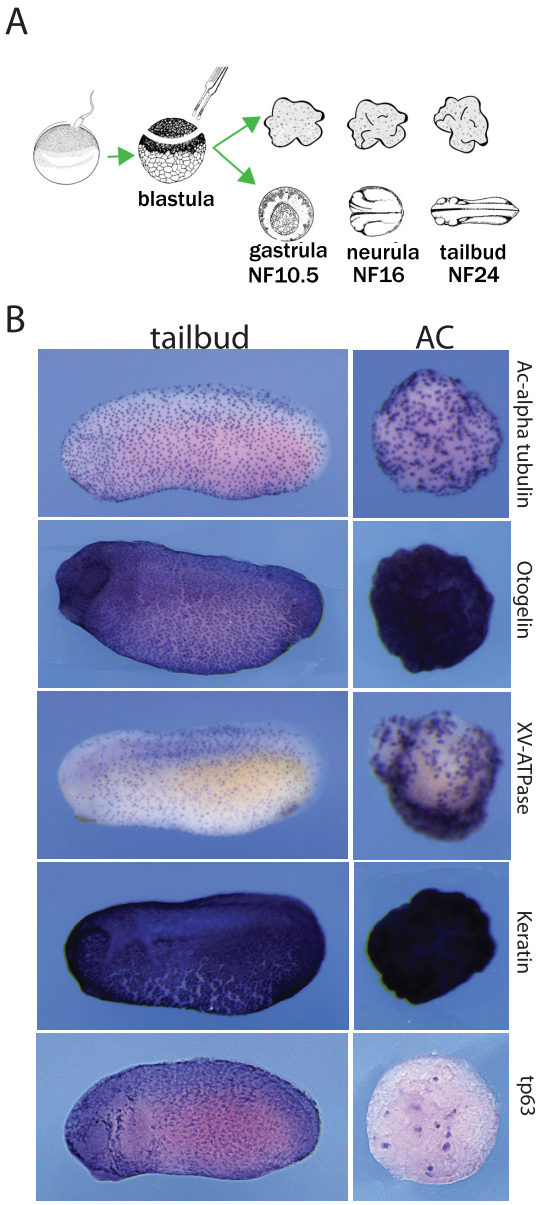

Figure 1. Autonomous differentiation of AC explants into a mucociliary epithelium. (**A**) Experimental procedure and timeline. (**B**) Early tailbud embryos and sibling ACs are stained for typical epidermal markers: acetylated-alpha tubulin protein (ICC), and for otogelin, xv-atpase, keratin and tp63 mRNA (RNA in situ hybridisation). ACs show a staining pattern comparable to the embryonic skin.

was plotted at the three timepoints. In general, we observed a decay in the expression levels of the former two groups, coupled to a strong up-regulation of ectoderm specific transcripts. Overall, these data indicate that our AC explants reliably describe the differentiation of the larval epidermis from germ layer determination to a differentiated epithelium with coordinated cilia movement.

The presence of mesendodermal transcripts in AC RNA samples—although at low abundance—raised the question of contamination. Their declining expression over the time-course is compatible with the assumption that regulatory genes of non-ectodermal differentiation programs could

simply reflect the pluripotent cell state of the tissue at the time of explantation, but it does not resolve the issue. Therefore, we extracted the reads for known pluripotency genes, including a set of genes conserved between mouse, human and *X. laevis* (40). The majority of these genes (30/39), notably all *oct4*-related *pou5f3* family members as well as *sox2* and *sox3*, were characterized by declining expression over the time course (Supplementary data, Table S1). Overall, this data confirms that ACs leave the pluripotent state as expected. However, the manual dissection of embryos inadvertently poses the risk of contaminating the tissue of interest with unwanted cells. Several studies in *X. laevis* have observed this problem, both with ACs and ectodermal explants (41,42). To obtain a quantitative estimate of such a contamination, we have measured by qRT/PCR the transcript levels of seven structural muscle genes, both in embryos and in the very same AC samples used for the RNA-Seq analysis. The selected genes are expressed at very high levels in Xenopus tropicalis (e.g. actc1: $\sim 3 \times 10^{10}$ TPE at NF24; Xenbase) and therefore represent sensitive means to evaluate the presence of non-ectodermal mRNAs. Based on calculations of their average expression levels in myocytes (see Supplementary Data, section 'Ectodermal purity of RNA-Seq data'), the degree of contamination with muscle-derived mRNAs in our data is around 0.1% and thus negligible.

At least 30 million reads were obtained from each ribozero library. Since sequences from the biological replicates cluster strongly (Figures 2D and 8), they sum up to approximately 100 million reads per developmental stage, providing a rich transcriptome resource for a simple, differentiating Xenopus tissue. At the gastrula stage, 74.92% of the reads mapped to unique gene loci on the *X. tropicalis* genome annotation 9.0. Of the remaining reads, 18.01% were from non-unique loci (mostly rRNA [13.3%], while 6.82% were too short to be aligned. Although the current X. tropicalis genome annotation is certainly still incomplete (43), it matches at least 93% of our reads. For the other stages, 66.19% (NF16) and 65.08% (NF24) of the reads were uniquely mapped. The transcripts were derived from 8525 genes (14932 RNA isoforms) at gastrula, 8403 genes (15 457 RNA isoforms) at neurula, and finally 8138 genes (15 046 RNA isoforms) at tailbud stage. The data does not contain small non-coding RNAs, which have been eliminated during library production (see materials and methods). We used this resource to define changes in gene expression between ACs and whole embryos and between ACs from different stages. While the former comparison describes the extent to which the forming epidermis contributes to the global embryonic transcriptome, the latter provides information on temporal changes in gene expression associated with epidermal cell fate specification and differentiation.

We first compared the transcriptome of neurula stage embryos (NF16) derived from Owens *et al.* (1) to the transcriptome of our corresponding AC libraries (Figure 2A, Supplementary Table S2) and found over 3800 transcripts to be differentially expressed. A bigger variance in mRNA levels was observed for the cohort of genes downregulated in AC versus embryos, compared to genes enriched in the AC (Figure 2A). The top one hundred genes upregulated in ACs have an average $\log_2$ fold-change of 3.68, while the bottom one hundred of downregulated genes have an average $\log_2$ fold change of –4.97. These findings reflect that epidermal RNAs are present, although diluted, in the whole embryo transcriptome, while transcripts from other tissues and germ layers are absent from ACs. Comparing the AC transcriptomes from different stages (Figure 2B and C; Supplementary Table S3), we deduced that the majority of gene expression changes are already happening between NF10.5 and NF16. In fact, more than half of the annotated transcriptome is differentially regulated between these two developmental stages, while only a minor portion of mRNAs are regulated between NF16 and NF24. In all cases, datasets from biological replicates grouped together (Figures 2D and 8), confirming the robustness of these conclusions.

While finalising this manuscript, a detailed catalogue of *X. tropicalis* cell states has become available from single cell RNA-Sequencing (43). We have compared our data to the potential fate regulators and marker genes for epidermal cell types reported by Briggs and colleagues. For this purpose we filtered our RNA data first for transcripts enriched in ACs over embryos at NF16, and secondly, for being differentially expressed during the timecourse. Our analysis discovers 98.8% of both potential fate regulators (160/162 genes, Supplementary Table S4/section 'Briggs 1') and epidermal cell type specific markers (418/423 genes; section 'Briggs 2') that were predicted from single cell transcriptomes. Among these genes, 18 did not fluctuate over time, and only seven are not expressed in animal caps, representing potential false positives (section 'Not expressed/not modulated'). Overall, our data independently validates the epidermal signature predicted by Briggs *et al.*, which represents some 7% of the total genes expressed in the epidermal data set presented here.

We also included a core set of over 800 genes from *X. laevis* ACs, which have been linked to MCC differentiation by gain and loss of function approaches (41). This gene set is derived from experimentally maximized differences in mRNA abundance targetting MCC differentiation in ACs, while our analysis detects natural fluctuations of mRNA abundance for all epidermal cell types. Despite this methodological difference, we retrieve 68% (552/813 genes) of the core multiciliogenic genes as being positively regulated during differentiation. An additional 135 genes from the core MCC markers are either not expressed in ACs, do not fluctuate or become downregulated in our timecourse (Supplementary Table S4, section 'Quigley'). These findings support the notion that isolated AC tissue recapitulates robustly the differentiation process of non-neural ectoderm. The inherent spatial information provided by the isolated organoid culture, simultanous RNA monitoring for all epidermal cell types under unmanipulated conditions, and the analysis of specialized RNA features, outlined below, distinguish our study from published reports and extend our knowledge about the transcriptional program of epidermal differentiation in Xenopus.

Alternative splicing shapes the epidermal transcriptome

The current *X. tropicalis* annotation release 103 documents more than 46 000 transcripts with a mean of 1.83 transcripts per gene (30). This raised the issue, whether and to what ex-

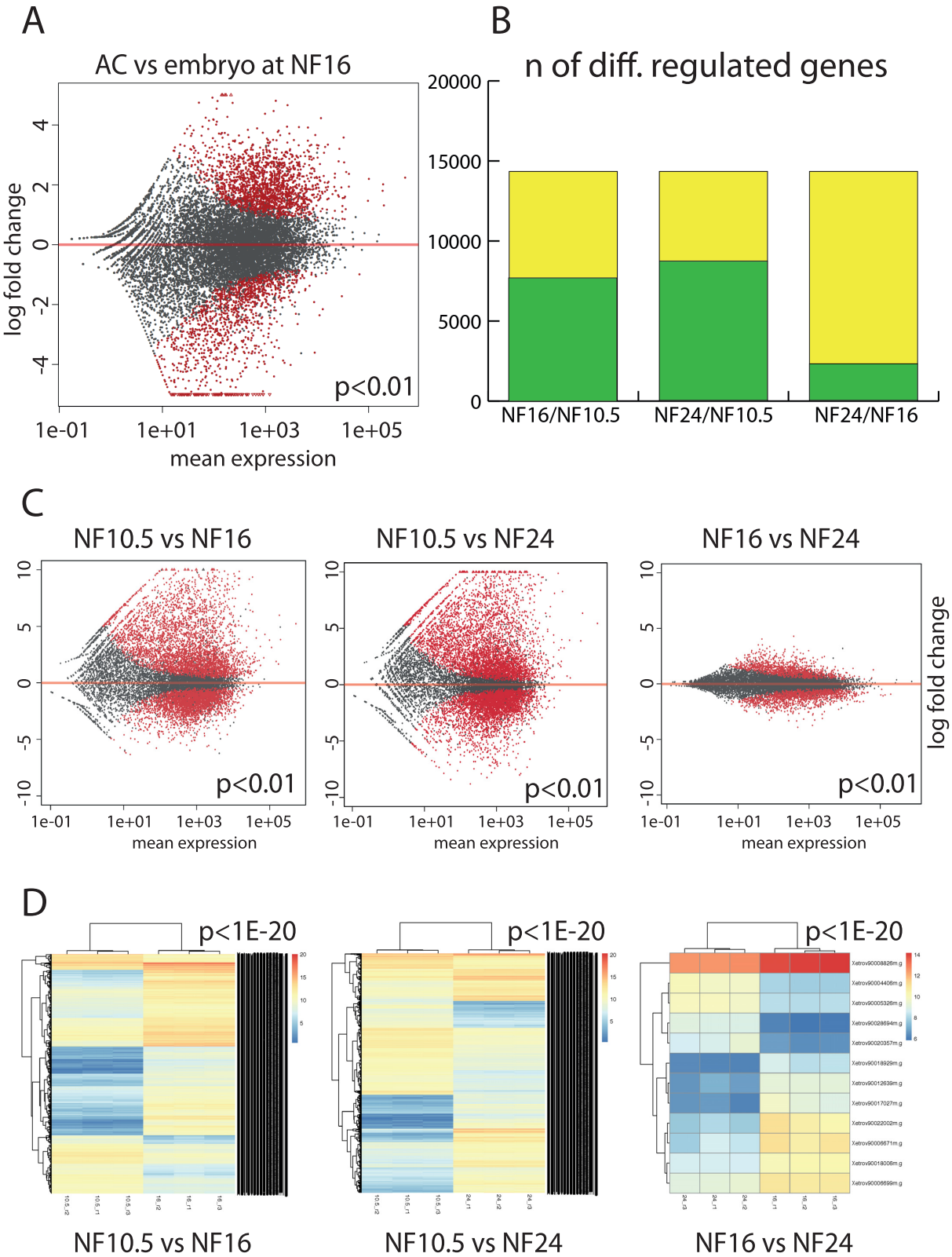

Figure 2. Comparative transcriptome analysis of epidermal differentiation. (**A**) Gene expression in the AC versus whole embryo at the neurula stage (NF16). Each dot represents a gene. In red, genes that are differentially expressed in a significant manner ($P < 0.01$), in grey genes with not-significant variation. (**B**) Plotting the extent of differential gene expression in AC. Yellow: total number of genes detected in the RNAseq. Green: genes that are differentially regulated between the given developmental stages ($P < 0.01$). The maximal number is defined by the annotated gene models present in XenTro9.1. (**C**) Scatter plots detailing the extent and magnitude of differential gene expression between indicated developmental stages ($P < 0.01$). (**D**) Unsupervised clustering of the differentially expressed genes from the biological replicates with a cutoff of $P <$ 1e–20.

tent the overall splicing of the transcriptome varies, both in the course of epidermal differentiation, and between the epidermis and the whole embryo (Figure 3). To address this question we used RSEM to map the reads of our libraries to known annotated splice junctions. At neurula stage, approximately 3200 splicing isoforms were differentially expressed between AC and embryo. Over 65% of these isoforms were up-regulated in prospective skin with a threshold of $P < 0.01$ (Figure 3 A; Supplementary Table S5). When comparing the different AC datasets with each other we observed an even higher variety in exon usage. In fact, the abundance of over 12 600 splicing isoforms (i.e. more than a quarter of all currently annotated splice variants) is modulated between gastrula and tailbud stages (Figure 3B; Supplementary Table S6, section 10vs24). The majority of the quantitative changes are happening again between stages NF10.5 and NF16, with 11054 splicing variants being differentially regulated ($P < 0.01$). Differences between stages NF16 and NF24 are smaller (2439 isoforms; Figure 3C and D), consistent with our findings for gene expression in general (see Figure 2C, D).

Transcript isoforms have also been detected in previous reports for *X. tropicalis* embryos (1–3), however, our dataset provides for the first time spatial information. To validate the significance of the observed splicing diversity, we looked for genes with at least two annotated splicing variants, whose expression varies strongly between AC (this study) and embryo (NF16-sample from (1)) and are represented by at least one thousand reads in our libraries (Supplementary Table S7). We selected three candidates with those properties. The *evpl* gene, encoding the desmosome component Envoplakin, has a skin specific exon that is not expressed in the embryo, but all other exons are ubiquitous; the bone morphogenetic protein *bmp7.2* gene has four splicing isoforms, of which only 1 (*bmp7.2* RNA13257) is expressed in the AC, while the others are expressed in the embryo; finally, the *cpne1* gene, coding for the Ca-dependent phospholipid binding protein Copine 1, has six splicing isoforms, whose ratios change dramatically during AC differentiation (Figure 4A). We then synthesized antisense probes from unique exons and performed RNA *in situ* hybridisations on ACs and embryos in parallel (Figure 4B). Confirming the predicted tissue-specific expression of mRNA splice variants, AC-specific probes gave staining on the ACs explants and on the epidermis of the embryos, while embryo-specific probes did not stain the ACs.

These results demonstrate that our AC database allows to deduce splice variants with a unique tissue-specific expression pattern. In fact, our findings suggest alternative splicing as a major regulatory mechanism in epidermal differentiation. Unfortunately, the current Xenopus protein annotation is not qualified to investigate systematically to which extent alternative splicing affects the protein coding capacity of the epidermal mRNA pool. For proof of principle, we have investigated four regulatory genes (*tp63, mcidas, grhl1* and *rfx2*) that have important functions in epidermal differentiation and come in multiple splicing variants. In each case, the observed splicing events affect the open reading frame of the mRNAs, although functional consequences are unknown (Supplementary Figure S3). Specifically, they predict five different TP63 protein variants, all due to C-terminal splicing events. Since the current *tp63* gene model misses promoter 1-derived transcripts, all identified mRNAs encode apparently δNp63 protein isoforms. For *mcidas*, we detected a new N-terminal domain, derived from a second start codon defining a unique aminoterminal peptide. The different grhl1 RNA isoforms seem to be produced by exon skipping. In the case of *rfx2* the transcript alignment suggests an exon-skipping for the Refseq transcript that removes 30 amino acid residues from an aminoterminal domain, which is present in two additional mRNAs (Supplementary Figure S4). This region shares similarity with the transactivation domain of human RFX1, suggesting that the two Rfx2 protein isoforms are functionally different. The abundance of alternatively spliced transcripts associated with the isolated ectoderm organoids make this dataset a promising starting point to search for novel splice variants.

To make this information available to the research community, we have generated an online Web Tool 'skin differentiation' (URL address provided below), where users can monitor expression levels of genes and splicing isoforms in ACs at the three developmental stages. Transcript abundance is expressed as fragments per Kilobase of transcript per Million mapped reads (FPKM). By typing in the name of a gene of interest, the tool retrieves all its annotated splicing variants, their relative expression in ACs at individual time-points, and their NCBI identifier.

Transcripts from repetitive DNA elements

For long times, repetitive DNA elements (REs) have been called *junk DNA* and their expression has been considered deleterious as provoking mutagenic insertional events (44,45). In recent times, several authors have proposed a functional role for transposable elements during embryogenesis (46–48). Host genes may be regulated by REs via readthru-transcription initiated at RE promoters, transactivation of genes by RE enhancers, and influences arising from epigenetic mechanisms targetted to REs in gene vicinity. Exploring the tissue-specific expression of such elements is therefore important to better understand genome activation and regulation during embryogenesis. Like in other vertebrates, the *X. tropicalis* genome contains up to 50% of REs (49–51). A small proportion of REs have retained transcriptional activity and are expressed in the wake of zygotic genome activation at the midblastula transition (49,52). Currently, 724 different REs have been annotated from *X. tropicalis* on RepBase (33). Since REs are suspected to impact gene expression during development, we characterized their expression profile. We first compared RE expression at gastrula stage between our ACs and the corresponding embryo dataset, derived from Owens *et al* (1). Nearly one fifth (137/724) of the known REs are differentially expressed in these two samples ($P < 0.01$; see Figure 5A, Supplementary Table S8). If one compares AC transcriptomes with each other, the variation in expression levels was even higher between stages NF10.5 and NF24 (261 REs, i.e. 36%; $P < 0.01$). On the other hand, little or no difference in RE transcript levels was observed between NF16 and NF24 (Figure 5, B-D; Supplementary Table S9). Overall these results are

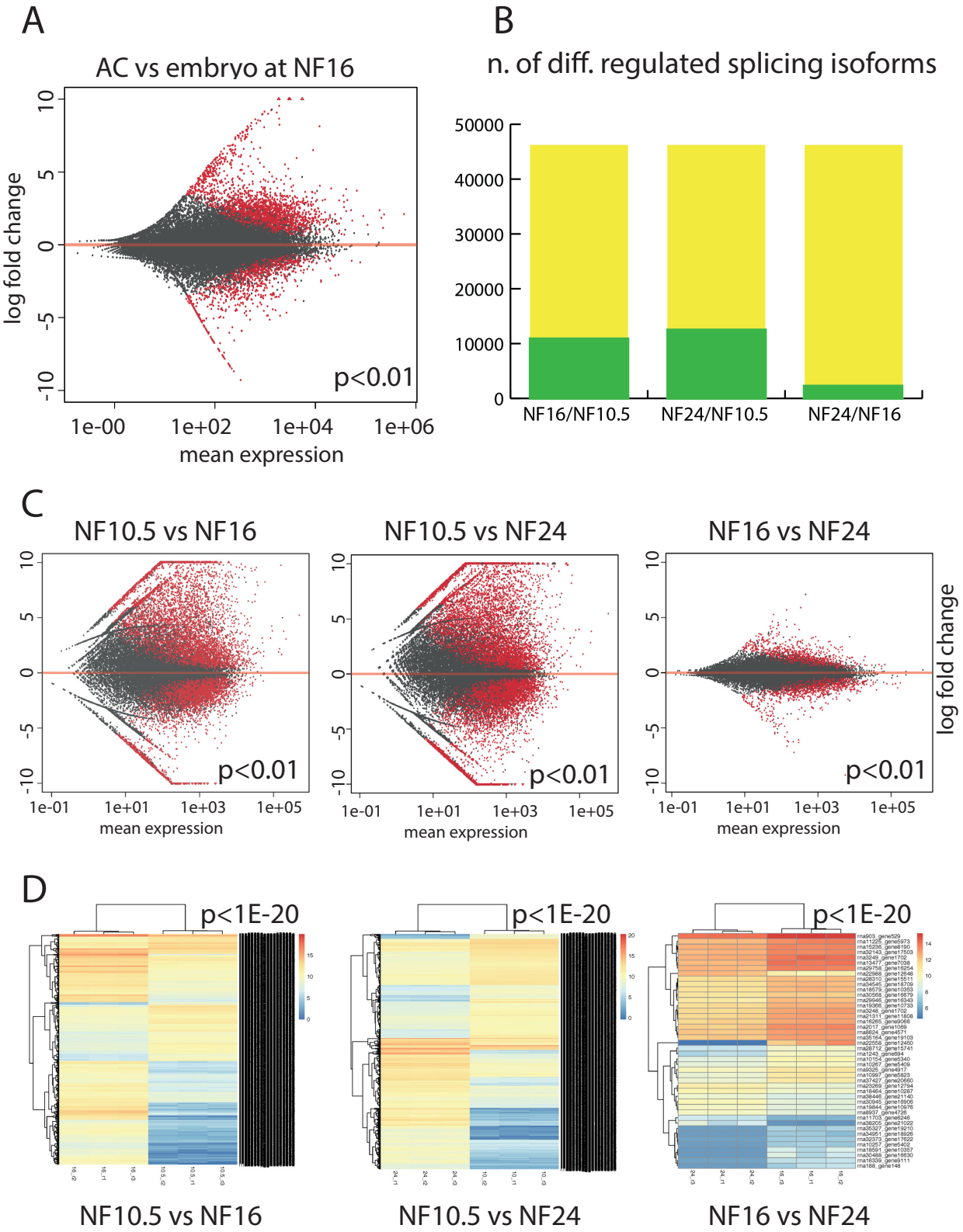

Figure 3. Epidermis-specific splicing isoforms. (**A**) Expression of splicing isoforms in the AC vs whole embryo at the neurula stage (NF16). Each dot represents an annotated splicing isoform. In red, splice isoforms that are differentially expressed in a significant manner ($P < 0.01$), in grey transcripts with not-significant variation. (**B**) Plotting the extent of differentially expressed splice isoforms in AC. Green: isoforms that are differentially regulated between the given developmental stages ($P < 0.01$). Yellow: the maximal number is defined by the annotated transcript models present in NCBI *Xenopus tropicalis* annotation release 103. (**C**) Scatter plots detailing the extent and magnitude of differential splice isoform expression between indicated developmental stages ($P < 0.01$). D) Unsupervised clustering of the differentially expressed splicing isoforms from the biological replicates with a cutoff of $P <$ 1e–20.

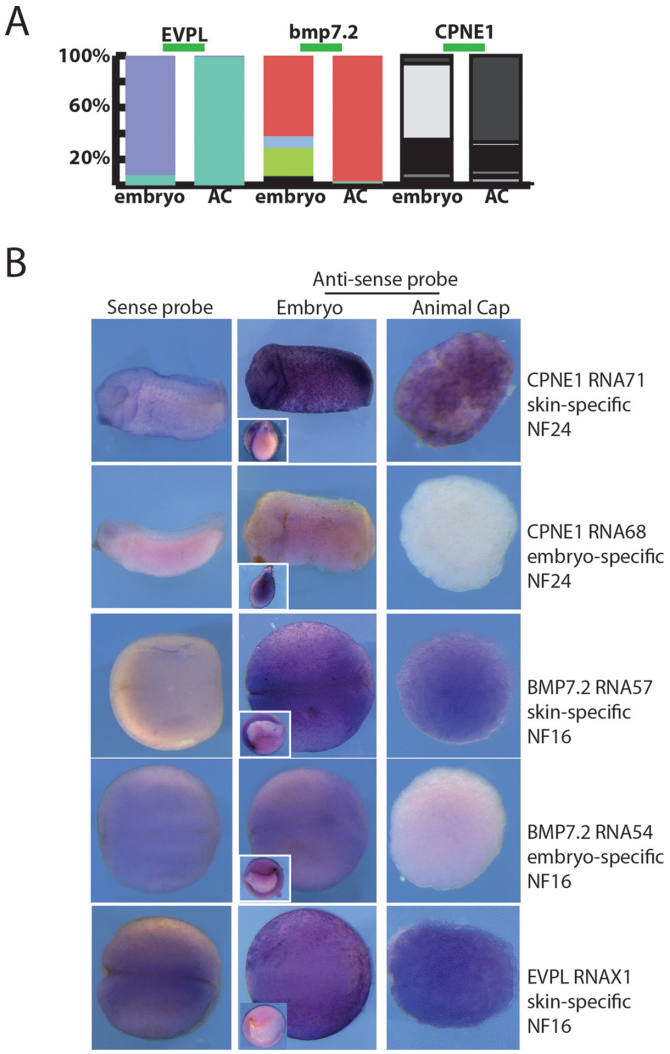

Figure 4. Spatial expression patterns of embryonic and epidermal RNA splicing isoforms. (**A**) Representations of the variety of splicing isoforms expressed in the ACs or in whole embryo for the indicated genes. (**B**) RNA *in situ* hybridizations for age-matched embryos and ACs at indicated stages, hybridized either with the skin specific splicing isoform probes (CPNE1 RNA0071, BMP7.2 RNA13257 and EVPL X1) or with the embryo-specific isoform probes (CPNE1 RNA0068, BMP7.2 RNA13254). Genes were picked from Supplementary Table S7; bmp7.2 is not on the list, since ACs and embryo share the same prevalent isoform. It was nevertheless chosen because of the importance of BMP signaling in ectodermal cell fate choice. Inserts show embryos hemisectioned prior to hybridization. Staining reactions have been conducted in parallel for each row.

in agreement with the reported global silencing of RE transcription during gastrulation (52).

In order to expand this information, we decided to concentrate our attention on specific REs, which reveal large expression changes between gastrula and tailbud stages and are represented by at least 10K reads in the RNA-Seq libraries. To this point, we have selected 3 different kinds of elements for further analysis: ERV1-4-LTR_Xt, an endogenous retrovirus; LTRX1-LTR_Xt, a LTR retrotransposon; and REM2B transcripts, derived from centromeric repeats. We cloned these transcripts by RT/PCR to produce probes for RNA in situ hybridisation analysis (Figure 6). At the gastrula stage, the three REs were transcribed in both embryos and ACs, predominantly in a unidirectional manner. Hybridisations performed on halved embryos identified spatial differences in expression. Rem2B transcripts were uniformly expressed in the animal and equatorial region of the embryo, except for the outermost cell layer and the vegetal pole area. LTRX1 transcripts were abundant in deep cell layers of the equatorial region, representing the mesoderm, as well as in a punctated fashion in superficial ectodermal cells. Finally, ERV1-4-LTR_Xt transcripts were broadly expressed throughout the embryo, but always in a punctated manner. By the tailbud stage, RE transcripts have largely disappeared from both embryos and ACs, except for Rem2B, which is expressed at reduced levels in the embryonic CNS and tailbud tissues.

While the underlying mechanisms dictating these expression patterns remain unclear, our results indicate that RE subclasses are subject to specific transcriptional regulation, compatible with the emerging view of a developmental role for repetitive DNA elements.

Circular RNAs

CircRNAs are covalently closed RNA molecules that originate via back splicing events (53,54). This class of transcripts was for long considered as products of erroneous splicing (55,56), but in recent years new functions have been proposed. CircRNAs seem to act as miRNA sponges, to interact with the POL II machinery, and to induce alternative splicing events (57). Recently it has been shown that the cytoplasm of Xenopus oocytes contains hundreds of circRNAs originating mainly from intronic sequences (58). Thus, we wondered, whether their expression is developmentally regulated.

In order to identify circRNAs in our dataset we used the method described by Westholm *et al* (29). Only reads overlapping with the circular junction can be unambiguously attributed to a circRNA. For example, in the second biological replicate of NF16 ACs, we have identified 4770 circular junctions. Approximately 382 of them are flanked by a GT-AG splicing site and 314 align to genome annotation 9.0 (Figure 7A). With this method we have been able to identify ∼250 circRNAs per developmental stage (Figure 7B), the majority of which being represented by few reads only, while the top abundant ones are represented by low double-digit numbers of reads per library (Supplementary Table S10). Remarkably, some circRNAs are expressed in all three biological replicates of one developmental stage, while others accumulate constantly throughout epidermal differentiation (Figure 7C). Such circRNAs may be produced in a stage- or cell state-dependent manner. To validate the existence of circRNAs, we performed end-point RT-PCR analysis for three of them (circ6, circ27 and circ1, highlighted in Supplementary Table S10) using two pairs of primers: one pair spanning the circular junction and the second pair amplifying DNA in inverse orientation. As predicted for circRNAs, the first primer pairs gave amplicons of the expected size (Figure 7D), while the second primer pairs resulted in longer amplicons. The predicted products spanning the circular junctions were then confirmed by sequencing. This

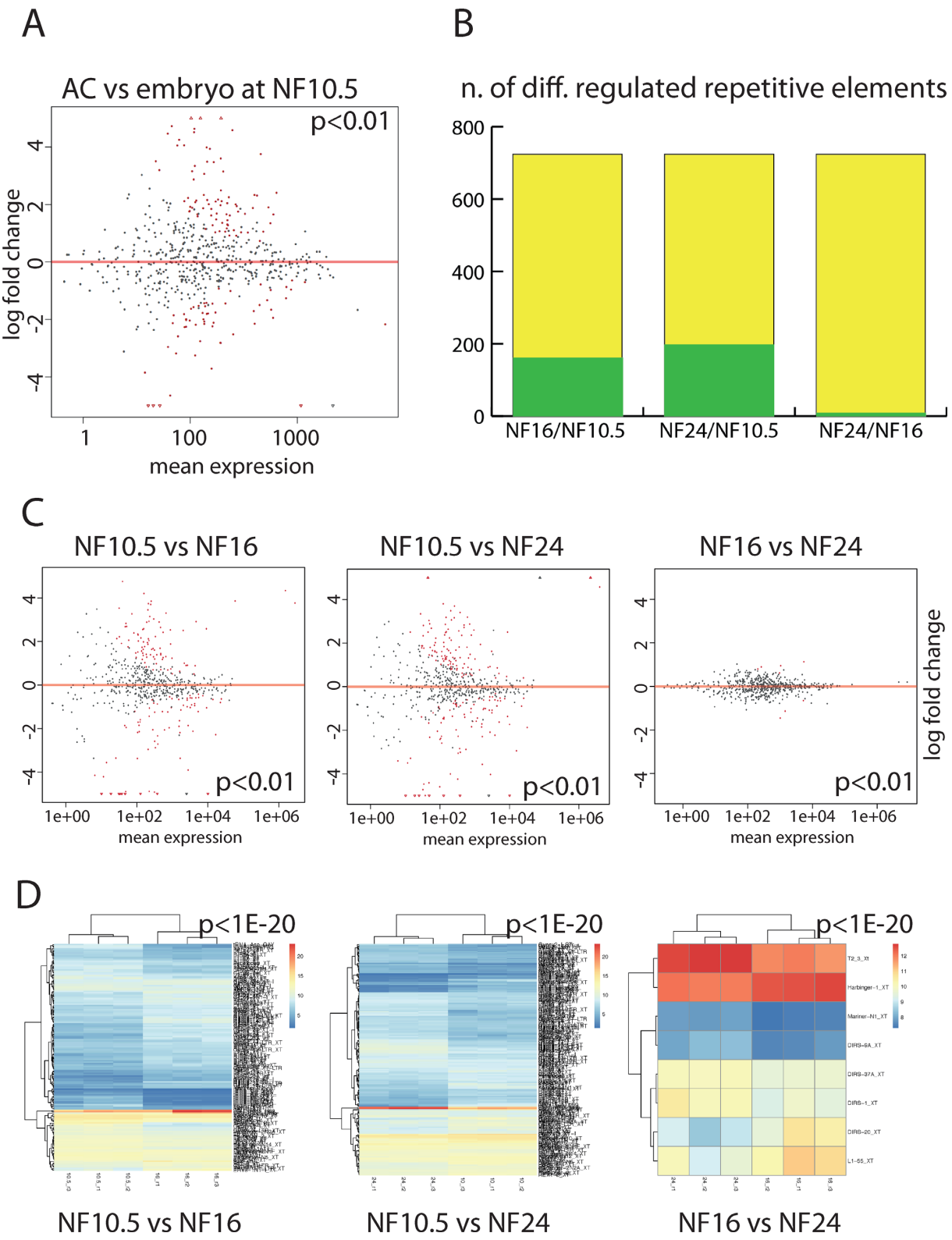

Figure 5. Transcripts from repetitive DNA elements during epidermal differentiation. (**A**) Expression of RE transcripts in the AC vs whole embryo at the gastrula stage (NF10.5). Each dot represents an annotated RE transcript. RE RNAs that are differentially expressed in a significant manner ($P <$ 0.01) are shown in red, while transcripts with not-significant variation are shown in gray. (**B**) Plotting the extent of differentially expressed RE-derived transcripts in AC. Green: differentially regulated RE RNAs between the given developmental stages ($P < 0.01$). Yellow: the maximal number defined by REPBASE 19.10. (**C**) Scatter plots detailing the extent and magnitude of differential RE expression between indicated developmental stages ($P < 0.01$). D) Unsupervised clustering of the differentially expressed RE transcripts from the biological replicates with a cutoff of $P <$ 1e–20.

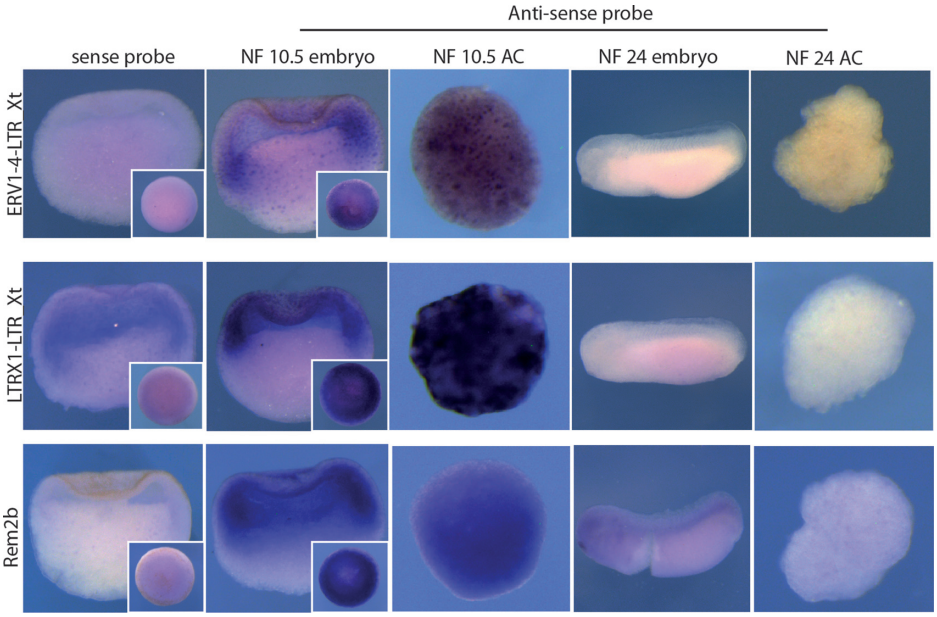

Figure 6. Region and stage-specific expression of repetitive DNA elements via RNA *in situ* hybridisation analysis of ERV1_4-LTR, LTRX1_LTR_Xt and Rem2b. Pictures display results from sense and antisense probe hybridisations performed and stained in parallel. Pictures of NF10.5 embryos represent sagittal sections; inserts show whole mount view of the animal hemisphere.

first identification of circRNAs in a Xenopus embryonic tissue should aid future functional studies.

Guiding principles in epidermal differentiation

Our AC RNA-Seq database provides rich information about Xenopus epidermal differentiation. Can we derive predictive information from this database? As a first step towards this goal, we carried out separate principal component analyses for gene expression, splicing isoforms and RE transcripts at the three investigated time points, selecting the two principal components that accommodate the majority of the observed variance for each transcript class (Figure 8). The PCA on gene expression is capable of resolving all three time points. The different biological replicates of a given developmental stage cluster together. The PCA for splicing isoforms and RE transcripts can only distinguish the gastrula timepoint from the older samples, but does not fully resolve neurula (NF16) and tailbud (NF24) stages. The obtained RNA-Seq data on gene expression is therefore a reliable resource to investigate stage-specific processes of mucociliary differentiation in Xenopus.

In addition, we made use of the Mogrify algorithm that has been developed to predict candidates for direct reprogramming of human cell types (34). In our case, the algorithm takes into account the known interaction network of single transcription factors (TF) as derived from the STRING database (59) and pairs this information with their expression dynamics in Xenopus ACs between two developmental timepoints. By this approach, we evaluate the importance of transcription factors as potential regulators of the observed transcriptional changes. The algorithm ranks TFs based on the measured degree of change of their sphere of influence as described in materials and methods. For each pairwise comparison, the method generates two lists of TFs predicted to control the transcriptome of the earlier (neg_rank list) or the later stage (pos_rank list), respectively (Supplementary Table S11). From the group of confirmed epidermal TFs, this method predicts regulatory roles for E2F4/5 and KLF17 from the gastrula timepoint on, followed at the neurula stage by tp63, gata2, grhl1/3 and members of the Fox (foxj1/1.2, foxi1, foxa1, foxn4) and Rfx (rfx1, -2, -3) protein families, compatible with other reports (41,43,60).

The additional TFs in these rank lists represent potentially novel regulators of stage-specific events in mucociliary epithelium differentiation. To illustrate relations among listed TF candidates, we have compared gastrula and tailbud AC transcriptomes and plotted the results for the top groups from the corresponding pos_rank and neg_rank lists (Supplementary Table S11, section '10vs24'). This deliberate choice to compare the youngest with the oldest of our samples should increase the chance of identifying novel regulatory candidates, since expression of known epidermal TFs peak at neurula. Based on the functional annotation of TFs provided by the GeneCards database (61), the predicted TF networks can be functionally interpreted in terms of stage-specific regulatory signatures. As shown in Figure 9A, the early gastrula stage transcriptome is predicted to be under control of TFs known to affect ES cell differentiation (SALL1, ZNF281), cell fate selection (FoxR1), epithelial differentiation (KLF17) and cell proliferation/cell survival (ZFAT, DMTF1, E2F4, TERF1, PMS1). Moreover, SIM1 and several members of the ZIC protein family (i.e. Zic1, -3, and -4) define sub-hubs within the predicted TF network, which may reflect the known neural differentiation bias of blastula ACs within their prospective dorsal halves (62). Indeed, early vertebrate ectoderm is prone to neuralization (11), and this bias is suppressed in Xenopus ACs by intrinsic BMP signalling (13). All these early

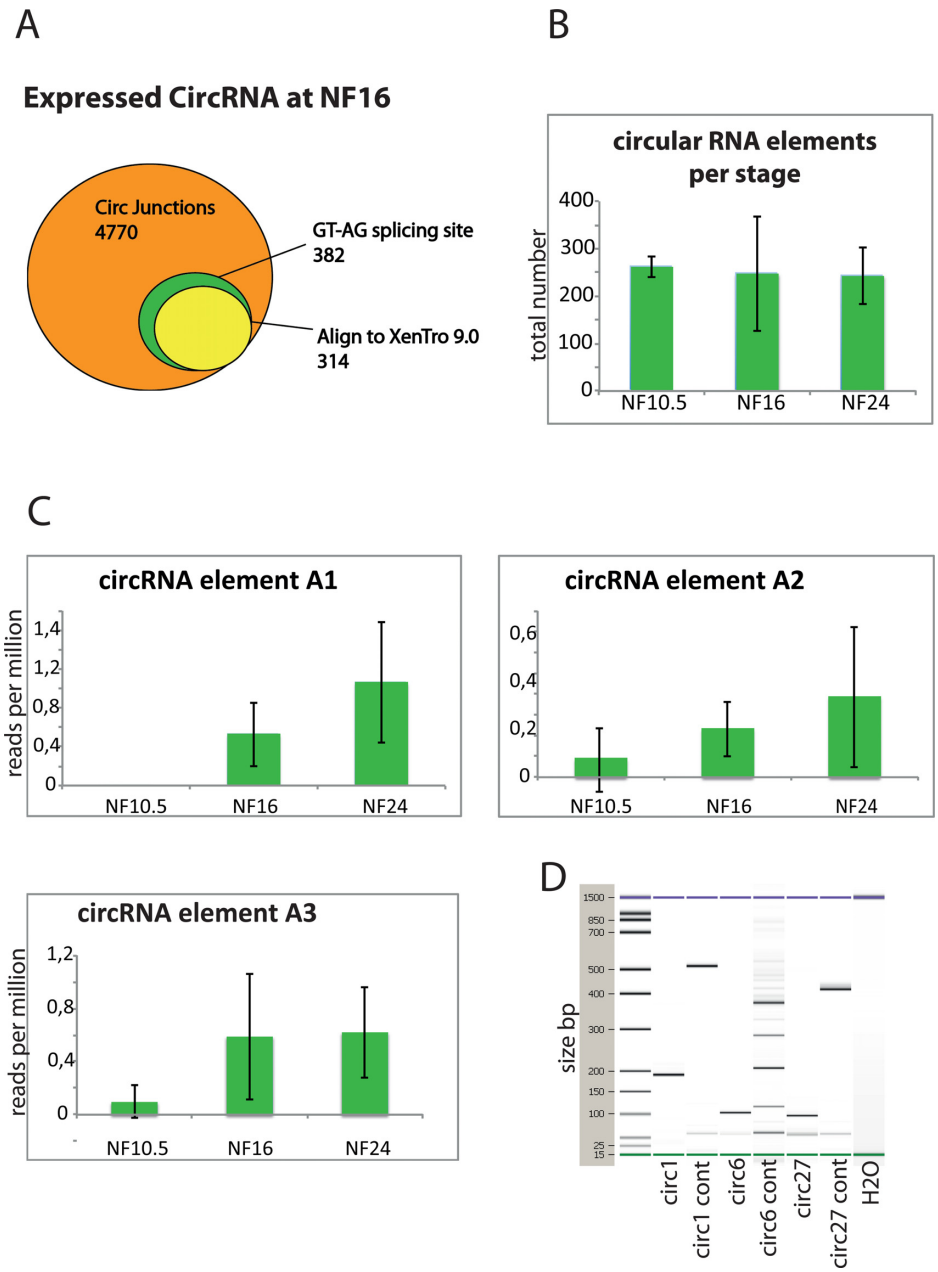

Figure 7. Analysis of circular RNAs. (**A**) Graphical representation of the filtered circular junctions from one single NF16 AC biological sample. (**B**) Total number of circRNAs detected in each developmental stage, error bars represent standard deviation. (**C**) Three examples of circRNAs that accumulate during the developmental timecourse. (**D**) Bioanalyzer gel of specific PCR-amplified circRNAs (highlighted in Supplementary Table S8). Each element has been amplified with a specific pair of primers, resulting in the predicted amplicon size (circ1 = 188bp, circ6 = 100bp and circ27 = 93bp) and a control reverse-complement primer pair, labelled as 'cont', which is expected to produce larger or multiple amplicons.

factors are absent from the predicted TF network controlling the tailbud epidermis (Figure 9B; Supplementary Table S9, pos_rank list). Here, the algorithm selects several TFs, which have not been recognized before as epidermal regulators in *Xenopus* (41–43). Two of them are connected to epidermal differentiation only by expression (ELF1, KLF3; this paper and (28)), but also the broadly expressed proto-oncogene c-fos belongs to this group. Unexpectedly, some of the predicted TF candidates have been connected to gene regulation in lymphoid cells (IKZF2, ZBTB1, KLF13; (61). This may indicate that epidermal and lymphoid cells share some cellular or metabolic functions. Further experiments are needed to investigate the roles of these candidates.

DISCUSSION

The intrinsic capacity of naïve Xenopus AC explants to differentiate into ciliated epidermis (8,9,11) offers a remarkably simple and accessible system to investigate the transcriptional program unfolding in the process of mucociliary epithelium formation. The three investigated time-points cover major transitions in epidermal differentiation, in-

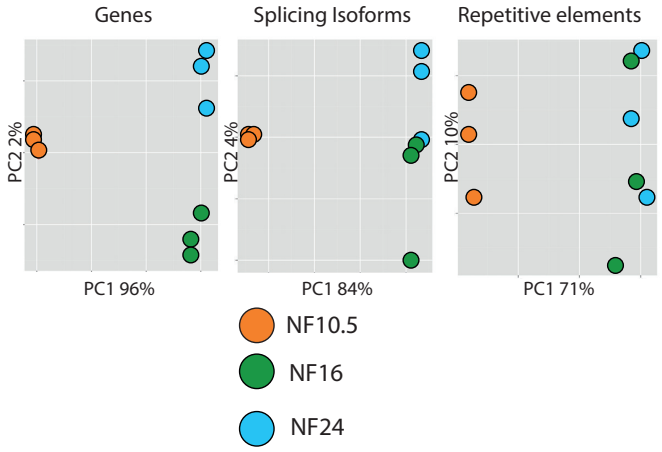

Figure 8. Prognostic value of AC transcriptomes. Principal component analysis of the AC samples for transcript classes indicated on top. Each dot represents a single biological replicate, each color represents a stage. On the X and Y-axis are plotted the first and second principal components. NF10.5 separates always very well from the other two stages. The transcriptomes of NF16 and 24 can be resolved mainly by gene expression and to less extent by RNA splicing.

cluding ectodermal committment, specification of deep cell layer-derived precursor cells, and differentiation into goblet, ionocyte and MCC cell types. SSCs, a third cell type derived from deep cells, appear on the surface with one day delay, compared to ionocytes and MCCs (21,22). While our data set will miss genes being activated late in SSCs (i.e. after stage NF24), it does contain transcripts from known SSC-specific genes such as foxa1, itpkb and ddc (see Supplementary Table S3; Web-Tool 'Skin differentiation'). With regard to single cell-transcriptome derived SSC markers (38), we find 38/38 genes in ACs (Supplementary Table S4, sections Briggs 1 and 2). This confirms the observation by Briggs and colleagues that the different ectodermal cell types appear and can be distinguished earlier than previously anticipated (43). In total, our data contains information for the five major cell types, from which the Xenopus larval skin is built from. This data may also be helpful to compare and filter gene expression profiles for other tissues and organs derived from growth factor-induced AC organoid-cultures (8).

Recently, the spatial and temporal transcriptomes of distinct ectodermal domains have been assessed by microdissection of *X. laevis* embryos in the course of neurulation (stages NF12.5 and NF14; (42). This so-called EctoMap has generated a molecular atlas, filtering gene signatures for five ectodermal domains. While concentrating mainly on the spatial analysis of neural, neural crest and placodal domains, an RNA signature is reported also for non-neural ectoderm. This signature consists of some 50 genes, all except three (htr7, casd1, sidt1) we find to be expressed in neurula-stage ACs (see online Web Tool 'skin differentiation'). This overlap in gene expression between cultured explants and freshly microdissected ectoderm suggests that most genes involved in epidermal differentiation are autonomously transcribed in ACs, although we cannot exclude that a few genes require signals from neigbhoring embryonic regions.

In comparison to the existing databases, our study includes both earlier (NF10.5) and later (NF24) timepoints and details both the differential expression of over 3800 genes between AC and embryos, as well as their stage-specific regulation in epidermis. With over 8000 genes and 15 000 splice isoforms identified in total, it exceeds the number of epidermal genes from single cell transcriptomic analysis by more than a factor of 15. This suggests ample room for future discoveries, specifically for genes with modulated expression and/or complex activities. Notably, it reveals for the first time the extent of tissue-specific alternative mRNA splicing during vertebrate tissue differentiation. In addition, we have investigated the temporal and spatial expression of repetitive DNA elements and described a compendium of epidermal circRNAs derived from exonic sequences. Therefore, the AC transcriptome reported here extends significantly our knowledge about the genetic basis underlying formation of a mucociliary epithelium in Xenopus, and provides a complementary dataset to single-cell analysis of the murine airway epithelium (63).

Principal component analysis of differentially expressed genes completely separated the three time points, while clustering the biological replicates. This indicates that the pattern of gene expression is predictive of the AC developmental age. In general, the majority of transcriptional changes happen between gastrula (NF10.5) and neurula (NF16) stages, which capture the transition from ectoderm committment to epidermal cell fate specification. Consistently, ACs initially express TFs that regulate pluripotency (Pou5f3 members, Sox2/3, Eomes), panectodermal (e.g. Zic1, Zic3, Hes7.1) and mesendodermal specification (Sox17b, Cdx1). As these latter RNAs disppear from the transcriptome of older ACs, they probably constitute remnants of the blastula pluripotent state. In contrast, the neurula stage transcriptome is enriched for TFs involved in the specification of the various skin cell types (tp63, foxj1, foxi1, foxa1, klf17 and grhl3). Unexpectedly, the pos_rank lists from the comparisons between gastrula and neurula against the tailbud transcriptome suggest strong contributions from Fos and Rel/NF-KB protein families during differentiation of the larval skin. This is reminiscent of results from Rackham and colleagues, who identified Fos and Rel as two of five TFs required to transdifferentiate human fibroblasts into keratinocytes (34). The other three TFs predicted to help in keratinocyte transdifferentiation, i.e. Foxq1, Sox9, Mafb, are also expressed in ACs (see online Web Tool 'skin differentiation'). Although these findings are speculative at the moment, they may promote interest in searches for novel TFs involved in epidermal gene networks by the method applied here.

In the course of this analysis, we have become aware of reads, which map to protein coding genes, but represent alternative splice junctions. A systematic search for such reads revealed that almost a quarter of all annotated Xenopus splice variants are differentially expressed in differentiating ACs. This may have important implications both for embryonic development and experimentation. Many genes that have been deeply studied for their function in the skin are expressed in the ACs as multiple splicing isoforms, potentially broadening their functional repertoire. This may also affect gene expression analysis by RNA in situ hybridisa-

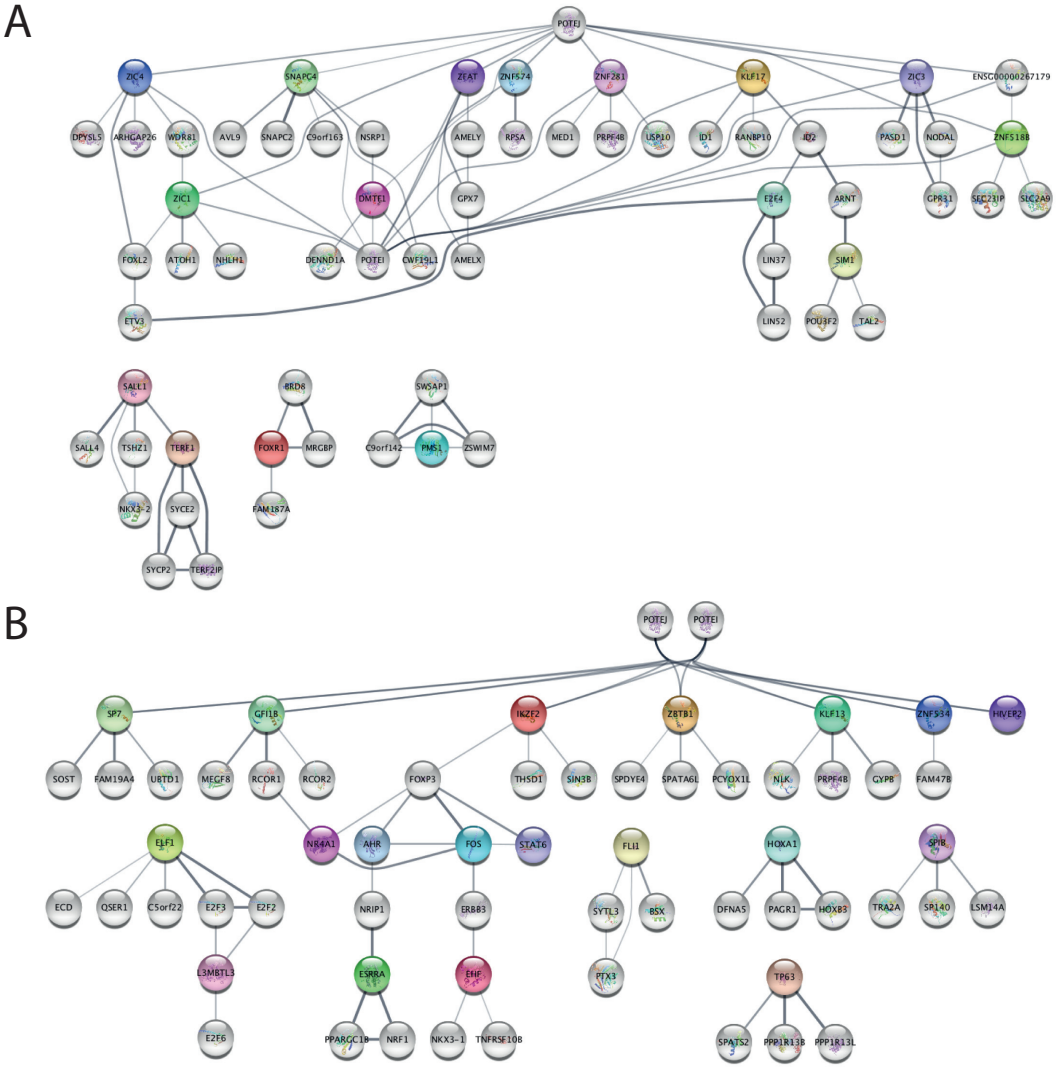

Figure 9. Transcription factor networks for the comparison of NF10.5 vs NF24 based on Mogrify. The top 15 TFs from each list (Supplementary Table S10) have been plotted for comparison. Each TF coming from our dataset appears in a coloured circle. The algorithm assembles hubs connecting all the factors using interactions deposited in the STRING database (in gray). The thickness of connecting lines is indicative of the strength of the known interaction. (**A**) Hubs derived from neg_rank list, naming factors specific for gastrula stage (NF10.5). (**B**) Hubs derived from pos_rank list, naming factors specific for tailbud stage (NF24).

tion or real time PCR. Although the current annotation of Xenopus protein variants has not allowed to investigate this insight globally, we have demonstrated that some epidermal splice variants change the protein capacity of important regulatory mRNAs. Furthermore, about 80 genes produce mRNA variants that are expressed in ACs and embryos in mutually exclusive manners. Some gene functions could rely on tissue-specific protein variants from alternatively spliced mRNAs. In non-mammalian vertebrate models, gene functions are frequently assessed through targetted protein knockdown by Morpholino oligonucleotides, for which rescue experiments are recommended to exclude OFF-target effects. The outcome of gain-of-function experiments could also be affected, when in specific tissues non-native cDNA variants are expressed. We noticed that Refseq cDNAs may not represent the major tissue-specific mRNA for a gene. For instance, Refseq mRNAs for δNtp63 constitute only about 25% of their total reads at NF16 (see online Web Tool 'skin differentiation'). The grhl1 Refseq mRNA is basically not expressed in ACs. These cases illustrate that it will be helpful to define the predominant splice isoforms for functional analysis.

Another intriguing result has to do with the expression of repetitive DNA sequences, in particular from transposable elements. Data from mice and humans have indicated that RE expression in normal cells is restricted to an early developmental period (10–12). In Xenopus, this lasts from zygotic genome activation at MBT through gastrula stages, when most transcribed REs have become silenced (52). Here we provide for the first time evidence for additional

spatial regulation of RE transcription. In fact, over 100 RE-derived transcripts are differentially expressed between ACs and embryos. Specifically, RNA in situ hybridisation has revealed an unexpected regionalization in the expression of some of these elements in the early-gastrula embryo. The most striking example is presented by the endogenous retrovirus ERV1-4LTR-X, which is expressed in a salt-and-pepper pattern distributed throughout the embryo. Together, these findings strongly argue for region- and possibly cell type- or cell cycle-specific regulatory mechanisms, which impact RE transcription. Our data also suggest that the production of circular RNAs might be developmentally regulated. In Xenopus, circRNAs derived from intronic sequences, were first described in the oocyte nucleus and cytoplasm. These maternal lariat RNAs are known to persist up to blastula stages and have been proposed to play a role in mRNA translation (58). A recent study shows that circRNA synthesis in Xenopus testis is affected by Atrazine, a potent endocrine disruptor responsible for developmental anomalies (64). Our finding that validated circRNAs accumulate in the course of AC differentiation is indicative of their *de novo* production. The newly discovered zygotic circRNAs should be helpful for functional investigation of this unusual RNA class in future.

In summary, this study has revealed an extraordinary RNA complexity associated with the formation of the mucociliary epithelium represented by the Xenopus larval skin. This compendium will be helpful for future delineation of individual cell type-specific transcriptomes, identification of conserved genetic pathways, and the elucidation of the functional impact of alternative splicing for vertebrate differentiation.

DATA AVAILABILITY

The RNA-Seq data is deposited at Sequence Read Archive under the IDs: SRR1795649, SRR651127, SRR1795663. The online WEB-Tool 'skin differentiation' can be accessed via the link: http://www.molekularbiologie.abi.med.uni-muenchen.de/ueber_uns/rupp/databases/skin_differentiation.html

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We apologize to all authors, whose work has not been able to cite appropriately due to space constraints. We express our gratitude to Edith Mentele and Barbara Hölscher for their expert technical support in AC dissection and RNA in situ Hybridisation, and to Peter Walentek for critical reading of the manuscript and sharing unpublished results. We also thank Drs A. Schweickert and E. Pera for providing molecular reagents and advice.

FUNDING

German Research Council DFG [CRC1064 project A12 to R.A.W.R.]. Funding for open access charge: Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

1. Owens,N.D., Blitz,I.L., Lane,M.A., Patrushev,I., Overton,J.D., Gilchrist,M.J., Cho,K.W. and Khokha,M.K. (2016) Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Rep.*, **14**, 632–647.
2. Paranjpe,S.S., Jacobi,U.G., van Heeringen,S.J. and Veenstra,G.J. (2013) A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development. *BMC Genomics*, **14**, 762.
3. Tan,M.H., Au,K.F., Yablonovitch,A.L., Wills,A.E., Chuang,J., Baker,J.C., Wong,W.H. and Li,J.B. (2013) RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. *Genome Res.*, **23**, 201–216.
4. Peshkin,L., Wuhr,M., Pearl,E., Haas,W., Freeman,R.M. Jr, Gerhart,J.C., Klein,A.M., Horb,M., Gygi,S.P. and Kirschner,M.W. (2015) On the relationship of protein and mRNA dynamics in vertebrate embryonic development. *Dev. Cell*, **35**, 383–394.
5. Popov,I.K., Kwon,T., Crossman,D.K., Crowley,M.R., Wallingford,J.B. and Chang,C. (2017) Identification of new regulators of embryonic patterning and morphogenesis in Xenopus gastrulae by RNA sequencing. *Dev. Biol.*, **426**, 429–441.
6. McCauley,H.A. and Wells,J.M. (2017) Pluripotent stem cell-derived organoids: using principles of developmental biology to grow human tissues in a dish. *Development*, **144**, 958–962.
7. Eiraku,M., Watanabe,K., Matsuo-Takasaki,M., Kawada,M., Yonemura,S., Matsumura,M., Wataya,T., Nishiyama,A., Muguruma,K. and Sasai,Y. (2008) Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell*, **3**, 519–532.
8. Asashima,M., Ito,Y., Chan,T., Michiue,T., Nakanishi,M., Suzuki,K., Hitachi,K., Okabayashi,K., Kondow,A. and Ariizumi,T. (2009) In vitro organogenesis from undifferentiated cells in Xenopus. *Dev. Dyn.*, **238**, 1309–1320.
9. Green,J. (1999) The animal cap assay. *Methods Mol. Biol.*, **127**, 1–13.
10. Green,J.B. and Smith,J.C. (1990) Graded changes in dose of a Xenopus activin A homologue elicit stepwise transitions in embryonic cell fate. *Nature*, **347**, 391–394.
11. Grunz,H. and Tacke,L. (1989) Neural differentiation of Xenopus laevis ectoderm takes place after disaggregation and delayed reaggregation without inducer. *Cell Differ. Dev.*, **28**, 211–217.
12. De Robertis,E.M. and Kuroda,H. (2004) Dorsal-ventral patterning and neural induction in Xenopus embryos. *Annu. Rev. Cell Dev. Biol.*, **20**, 285–308.
13. Cibois,M., Luxardi,G., Chevalier,B., Thome,V., Mercey,O., Zaragosi,L.E., Barbry,P., Pasini,A., Marcet,B. and Kodjabachian,L. (2015) BMP signalling controls the construction of vertebrate mucociliary epithelia. *Development*, **142**, 2352–2363.
14. Boon,M., Wallmeier,J., Ma,L., Loges,N.T., Jaspers,M., Olbrich,H., Dougherty,G.W., Raidt,J., Werner,C., Amirav,I. *et al.* (2014) MCIDAS mutations result in a mucociliary clearance disorder with reduced generation of multiple motile cilia. *Nat. Commun.*, **5**, 4418.
15. Dubaissi,E. and Papalopulu,N. (2011) Embryonic frog epidermis: a model for the study of cell-cell interactions in the development of mucociliary disease. *Dis. Model. Mech.*, **4**, 179–192.
16. Walentek,P. and Quigley,I.K. (2017) What we can learn from a tadpole about ciliopathies and airway diseases: using systems biology in Xenopus to study cilia and mucociliary epithelia. *Genesis*, **55**, e23001.
17. Werner,M.E. and Mitchell,B.J. (2012) Understanding ciliated epithelia: the power of Xenopus. *Genesis*, **50**, 176–185.
18. Hogan,B.L., Barkauskas,C.E., Chapman,H.A., Epstein,J.A., Jain,R., Hsia,C.C., Niklason,L., Calle,E., Le,A., Randell,S.H. *et al.* (2014) Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell*, **15**, 123–138.
19. Sirour,C., Hidalgo,M., Bello,V., Buisson,N., Darribere,T. and Moreau,N. (2011) Dystroglycan is involved in skin morphogenesis downstream of the Notch signaling pathway. *Mol. Biol. Cell*, **22**, 2957–2969.
20. Deblandre,G.A., Wettstein,D.A., Koyano-Nakagawa,N. and Kintner,C. (1999) A two-step mechanism generates the spacing

pattern of the ciliated cells in the skin of Xenopus embryos. *Development*, **126**, 4715–4728.
21. Dubaissi,E., Rousseau,K., Lea,R., Soto,X., Nardeosingh,S., Schweickert,A., Amaya,E., Thornton,D.J. and Papalopulu,N. (2014) A secretory cell type develops alongside multiciliated cells, ionocytes and goblet cells, and provides a protective, anti-infective function in the frog embryonic mucociliary epidermis. *Development*, **141**, 1514–1525.
22. Walentek,P., Bogusch,S., Thumberger,T., Vick,P., Dubaissi,E., Beyer,T., Blum,M. and Schweickert,A. (2014) A novel serotonin-secreting cell type regulates ciliary motility in the mucociliary epidermis of Xenopus tadpoles. *Development*, **141**, 1526–1533.
23. Quigley,I.K., Stubbs,J.L. and Kintner,C. (2011) Specification of ion transport cells in the Xenopus larval skin. *Development*, **138**, 705–714.
24. Showell,C. and Conlon,F.L. (2009) Egg collection and in vitro fertilization of the western clawed frog Xenopus tropicalis. *Cold Spring Harb. Protoc.*, **2009**, pdb prot5293.
25. Sive,H.L., Grainger,R.M. and Harland,R.M. (2000) *Early Development of Xenopus laevis: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
26. Nieuwkoop,P.D. and Faber,J. (1994) *Normal table of Xenopus Laevis (Daudin)*. 2nd edn. Garland Publishing Inc, NY.
27. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. 3rd and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
28. Hollemann,T., Chen,Y., Grunz,H. and Pieler,T. (1998) Regionalized metabolic activity establishes boundaries of retinoic acid signalling. *EMBO J.*, **17**, 7361–7372.
29. Westholm,J.O., Miura,P., Olson,S., Shenker,S., Joseph,B., Sanfilippo,P., Celniker,S.E., Graveley,B.R. and Lai,E.C. (2014) Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.*, **9**, 1966–1980.
30. Karimi,K., Fortriede,J.D., Lotay,V.S., Burns,K.A., Wang,D.Z., Fisher,M.E., Pells,T.J., James-Zorn,C., Wang,Y., Ponferrada,V.G. *et al.* (2018) Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Res.*, **46**, D861–D868.
31. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
32. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
33. Bao,W., Kojima,K.K. and Kohany,O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*, **6**, 11.
34. Rackham,O.J., Firas,J., Fang,H., Oates,M.E., Holmes,M.L., Knaupp,A.S., Consortium,F., Suzuki,H., Nefzger,C.M., Daub,C.O. *et al.* (2016) A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.*, **48**, 331–335.
35. Blitz,I.L., Paraiso,K.D., Patrushev,I., Chiu,W.T.Y., Cho,K.W.Y. and Gilchrist,M.J. (2017) A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. *Dev. Biol.*, **426**, 409–417.
36. Szklarczyk,D., Morris,J.H., Cook,H., Kuhn,M., Wyder,S., Simonovic,M., Santos,A., Doncheva,N.T., Roth,A., Bork,P. *et al.* (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.*, **45**, D362–D368.
37. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
38. Jones,E.A. and Woodland,H.R. (1986) Development of the ectoderm in Xenopus: tissue specification and the role of cell association and division. *Cell*, **44**, 345–355.
39. Jones,E.A. and Woodland,H.R. (1987) The development of animal cap cells in Xenopus: a measure of the start of animal cap competence to form mesoderm. *Development*, **101**, 557–563.
40. Livigni,A. and Brickman,J.M. (2013) Oct4: the final frontier, differentiation defining pluripotency. *Dev. Cell*, **25**, 547–548.
41. Quigley,I.K. and Kintner,C. (2017) Rfx2 stabilizes Foxj1 binding at chromatin loops to enable multiciliated cell gene expression. *PLos Genet.*, **13**, e1006538.
42. Plouhinec,J.L., Medina-Ruiz,S., Borday,C., Bernard,E., Vert,J.P., Eisen,M.B., Harland,R.M. and Monsoro-Burq,A.H. (2017) A molecular atlas of the developing ectoderm defines neural, neural crest, placode, and nonneural progenitor identity in vertebrates. *PLoS Biol.*, **15**, e2004045.
43. Briggs,J.A., Weinreb,C., Wagner,D.E., Megason,S., Peshkin,L., Kirschner,M.W. and Klein,A.M. (2018) The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, **360**, eaar5780.
44. Gregory,T.R. (2005) *The Evolution of the Genome*. Elsevier Academic Press, Amsterdam.
45. Ohno,S. (1972) So much "junk" DNA in our genome. *Brookhaven Symp. Biol.*, **23**, 366–370.
46. Bodega,B. and Orlando,V. (2014) Repetitive elements dynamics in cell identity programming, maintenance and disease. *Curr. Opin. Cell Biol.*, **31**, 67–73.
47. Erwin,J.A., Marchetto,M.C. and Gage,F.H. (2014) Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.*, **15**, 497–506.
48. Faulkner,G.J., Kimura,Y., Daub,C.O., Wani,S., Plessy,C., Irvine,K.M., Schroder,K., Cloonan,N., Steptoe,A.L., Lassmann,T. *et al.* (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.*, **41**, 563–571.
49. Hellsten,U., Harland,R.M., Gilchrist,M.J., Hendrix,D., Jurka,J., Kapitonov,V., Ovcharenko,I., Putnam,N.H., Shu,S., Taher,L. *et al.* (2010) The genome of the Western clawed frog Xenopus tropicalis. *Science*, **328**, 633–636.
50. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
51. Mouse Genome Sequencing,C., Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
52. van Kruijsbergen,I., Hontelez,S., Elurbe,D.M., van Heeringen,S.J., Huynen,M.A. and Veenstra,G.J.C. (2017) Heterochromatic histone modifications at transposons in Xenopus tropicalis embryos. *Dev. Biol.*, **426**, 460–471.
53. Ivanov,A., Memczak,S., Wyler,E., Torti,F., Porath,H.T., Orejuela,M.R., Piechotta,M., Levanon,E.Y., Landthaler,M., Dieterich,C. *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.*, **10**, 170–177.
54. Zhang,X.O., Wang,H.B., Zhang,Y., Lu,X., Chen,L.L. and Yang,L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134–147.
55. Cocquerelle,C., Mascrez,B., Hetuin,D. and Bailleul,B. (1993) Mis-splicing yields circular RNA molecules. *FASEB J.*, **7**, 155–160.
56. Nigro,J.M., Cho,K.R., Fearon,E.R., Kern,S.E., Ruppert,J.M., Oliner,J.D., Kinzler,K.W. and Vogelstein,B. (1991) Scrambled exons. *Cell*, **64**, 607–613.
57. Chen,L.L. (2016) The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.*, **17**, 205–211.
58. Talhouarne,G.J. and Gall,J.G. (2014) Lariat intronic RNAs in the cytoplasm of Xenopus tropicalis oocytes. *RNA*, **20**, 1476–1487.
59. Jensen,L.J., Kuhn,M., Stark,M., Chaffron,S., Creevey,C., Muller,J., Doerks,T., Julien,P., Roth,A., Simonovic,M. *et al.* (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic. Acids. Res.*, **37**, D412–D416.
60. Kwon,T., Chung,M.I., Gupta,R., Baker,J.C., Wallingford,J.B. and Marcotte,E.M. (2014) Identifying direct targets of transcription factor Rfx2 that coordinate ciliogenesis and cell movement. *Genom Data*, **2**, 192–194.
61. Rebhan,M., Chalifa-Caspi,V., Prilusky,J. and Lancet,D. (1998) GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics*, **14**, 656–664.
62. Gaur,S., Mandelbaum,M., Herold,M., Majumdar,H.D., Neilson,K.M., Maynard,T.M., Mood,K., Daar,I.O. and Moody,S.A.

(2016) Neural transcription factors bias cleavage stage blastomeres to give rise to neural ectoderm. *Genesis*, **54**, 334–349.
63. Treutlein,B., Brownfield,D.G., Wu,A.R., Neff,N.F., Mantalas,G.L., Espinoza,F.H., Desai,T.J., Krasnow,M.A. and Quake,S.R. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, **509**, 371–375.
64. Sai,L., Li,L., Hu,C., Qu,B., Guo,Q., Jia,Q., Zhang,Y., Bo,C., Li,X., Shao,H. *et al.* (2018) Identification of circular RNAs and their alterations involved in developing male Xenopus laevis chronically exposed to atrazine. *Chemosphere*, **200**, 295–301.