

RESEARCH ARTICLE

Genome-wide assessment of differential effector gene use in embryogenesis

Julius C. Barsi^{1,†,§}, Qiang Tu^{1,†,*}, Cristina Calestani² and Eric H. Davidson¹

ABSTRACT

Six different populations of cells were isolated by fluorescence-activated cell sorting from disaggregated late blastula- and gastrula-stage sea urchin embryos according to the regulatory states expressed in these cells, as reported by recombinered bacterial artificial chromosomes producing fluorochromes. Transcriptomes recovered from these embryonic cell populations revealed striking, early differential expression of large cohorts of effector genes. The six cell populations were presumptive pigment cells, presumptive neurogenic cells, presumptive skeletogenic cells, cells from the stomodeal region of the oral ectoderm, ciliated band cells and cells from the endoderm/ectoderm boundary that will give rise both to hindgut and to border ectoderm. Transcriptome analysis revealed that each of these domains specifically expressed several hundred effector genes at significant levels. Annotation indicated the qualitative individuality of the functional nature of each cell population, even though they were isolated from embryos only 1-2 days old. In no case was more than a tiny fraction of the transcripts enriched in one population also enriched in any other of the six populations studied. As was particularly clear in the cases of the presumptive pigment, neurogenic and skeletogenic cells, all three of which represent precociously differentiating cell types of this embryo, most specifically expressed genes of given cell types are not significantly expressed at all in the other cell types. Thus, at the effector gene level, a dramatic, cell type-specific pattern of differential gene regulation is established well before any significant embryonic morphogenesis has occurred.

KEY WORDS: Effector gene expression, Embryonic gene regulation, FACS embryonic cell populations, Sea urchin embryos

INTRODUCTION

Embryonic development is driven by the progressive establishment of spatial regulatory states, generated by the differential zygotic expression of regulatory genes encoding transcription factors. This has been demonstrated in every adequately studied experimental system, by analysis of the underlying gene regulatory networks that encode these regulatory states. A comprehensive review has recently been published (Peter and Davidson, 2015). These networks of interacting regulatory genes cause expression of much larger cohorts of protein-coding effector genes, which ultimately

produce the cell's biological and morphogenetic functions, as well as the differentiated characteristics of the many cell types to which the embryo gives rise. However, we have as yet surprisingly little experimental evidence, beyond the behavior of individual genes, that directly addresses the spatial deployment of effector gene cohorts during early embryonic development. Temporally differential expression of thousands of genes during embryogenesis, i.e. of thousands of effector genes, was indicated by mRNA complexity studies as far back as the 1970s (Davidson, 1986). Evidence of the temporal progression of effector gene expression during embryogenesis has recently been fleshed out on a genome-wide scale, by staged embryo transcriptome measurements that are too numerous to mention individually, but system-level measurement of the spatial expression of effector gene cohorts has been lacking for any embryo. Large numbers of beautiful and detailed studies have illuminated the spatial activation of individual effector genes of interest in the specific domains of various embryos, but thus far the quantitative extent of this fundamental regulatory process has remained inferential.

The sea urchin embryo is an exemplar of a widespread mode of invertebrate embryogenesis particularly common in marine forms, but not confined to these, in which differentiated cell types are specified prior to gastrulation, even before any embryonic structures are manifested morphologically. Precocious (pregastrular) differentiation is characteristic of these 'mode 1' embryonic specification processes (Davidson, 1990; Peter and Davidson, 2015). It is a clear prediction from observations on individual effector genes that, early in development, batteries of such genes should be expressed differentially in the early differentiating cell types of mode 1 embryos. Nonetheless, until now, no system-level evidence on spatial effector gene deployment has been available for any kind of animal embryo.

Using a method pioneered in our laboratory, we exploit known regulatory states to mark cell types of interest fluorescently and separate them out by way of flow cytometry (akin to Defaye and Perrin, 2014; Barsi et al., 2014). Specifically, we use fluorescence-activated cell sorting (FACS) to separate out six different populations of cells from the pregastrular and early gastrula sea urchin embryo. cDNA was amplified from the mRNAs of these sorted cell populations and used to generate transcriptomes. After verification of all of the selected transcript populations, we carried out a comparative qualitative and quantitative analysis of their differential transcription of effector gene sets. A large-scale demonstration of spatially exclusive, early effector gene regulation emerges, particularly in cell populations that are fated to express given modes of differentiation.

RESULTS

Embryonic cell populations used in this study

The six cell populations isolated from sea urchin embryos and compared in this study were as follows. (1) Presumptive pigment

¹Division of Biology and Bioengineering, Caltech, 1200 East California Boulevard, Pasadena, CA 91125, USA. ²Department of Biology, Valdosta State University, Valdosta, GA 31698, USA.

*Present address: Institute of Genetics & Developmental Biology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, P. R. China.

[†]These authors contributed equally to this work

[§]Author for correspondence (barsi@caltech.edu)

cells were identified experimentally as cells expressing *gcm* at 45 hours postfertilization (hpf) (Ransick and Davidson, 2006; 2012). At this stage, these cells are in the blastocoel and in the process of embedding themselves in the aboral ectoderm. They were identified by sorting on the green fluorescent protein (GFP) generated by a *gcm* cis-regulatory reporter construct. (2) Presumptive neurogenic cells of the apical plate were identified as cells expressing *lhx2/9* at 35 hpf (Lundgren et al., 1995; Farfán et al., 2009). At this stage, these cells constitute a small patch within the apical neurogenic domain of the embryo. They were identified by sorting on the GFP generated by a recombinereed *lhx2/9*-GFP bacterial artificial chromosome (BAC). (3) Presumptive skeletogenic cells were identified as cells expressing *tbrain* or *alx1* recombinant BACs at 24 hpf, as described at length earlier (Barsi et al., 2014). (4) Cells of the stomodeal region of the oral ectoderm were identified by expression of *gsc* at 35 hpf (Li et al., 2013). At this stage, these cells include the future stomodeal area and surrounding regions of the upper central oral ectoderm. They were isolated by sorting on the GFP generated by a recombinereed *gsc*-GFP BAC. (5) Cells of the ciliated band were identified at 35 hpf as cells expressing *onecut* (Poustka et al., 2004; Fig. S1) and isolated by sorting on the GFP generated from a *onecut* GFP recombinant BAC. (6) Cells of the *veg1* lineage give rise to the most posterior endoderm and the adjacent ectoderm (Peter and Davidson, 2010, 2011). They were identified by expression of the *eve* homeodomain regulator at 24 hpf and isolated by sorting on GFP generated by an *eve*-GFP recombinant BAC.

These populations of spatially confined embryonic cells fall into two biological classes and, as can be expected, the results we present below are affected by this distinction. Populations 1-3 above are essentially of unitary fate. The descendants of all *gcm*⁺ cells at this stage will become pigment cells; so far as is known, the descendants of all *lhx2/9*⁺ cells will become neurons (or neuron accessory cells); and the descendants of all *alx1*⁺ and *tbrain*⁺ cells will become skeletogenic mesenchyme cells. However, at the respective selected stages, *gsc*⁺ cells (population 4) will give rise to stomodaeum, to neurons and to the squamous oral epithelium; *onecut*⁺ ciliated band cells (population 5) include neurogenic precursors, cells that will produce the eponymous cilia, cell types of the oral posterior ectoderm, including bilateral patches of *Vegf*-secreting cells; and *eve*⁺ cells (population 6) will give rise to posterior ectoderm, both oral and aboral, as well as to oral and aboral posterior endoderm. Given their mixed fates, it is remarkable that, as described below, each of the populations 4-6 nonetheless expresses a unique set of effector genes, although in general with less distinct separation and more mutual overlap. In one case, the *onecut*⁺ ciliated band and the *veg1* *eve*⁺ cells, some overlap could be expected because the *veg1* ectoderm eventually forms the vegetal strip of the ciliated band even though as analyzed here, the *veg1* cells were isolated at 24 hpf and the ciliated band cells much later, at 35 hpf. However, the cell populations of unitary fate, populations 1-3, provide clear and unequivocal results, and we focus mainly on these in the following.

Data relevant to the isolated cell populations are shown in Fig. 1. The domains of the embryo represented are indicated in row 1. In Fig. 1Aa, pigment cells are shown embedded in the aboral ectoderm wall of a late embryo, alone among all the cells in the embryo generating the orange/red echinochrome pigment. Fig. 1Ba shows a view of a late blastula embryo in which the neurogenic apical domain is marked by *foxq2* expression in orange; adjacent to this domain, on the oral side, is the *gsc* domain of the oral ectoderm stained blue. As described below, the *lhx2/9* domain is confined to a subregion of the *foxq2* apical plate. In Fig. 1Ca, the

now mesenchymal skeletogenic domain is highlighted by expression of *tbrain* (*pmc*, primary mesenchyme cells). In Fig. 1Da, a lateral view of an early gastrula embryo (35 hpf), the ciliated band domain that bounds the oral ectoderm can be seen in this optical section as two patches of *onecut* expression, one above and one below (orange), whereas the oral ectoderm at this stage is expressing *gsc* (blue). The view in Fig. 1Ea is oral, and the complete circular pattern of the ciliated band expressing *onecut* can be seen (blue), the oral ectoderm is unstained, and at the vegetal end of the embryo the anterior endoderm is marked as the domain of *foxa* expression (orange). Fig. 1Fa shows the *eve*-expressing *veg1* domain (orange), immediately abutting the *foxa* domain (blue), again in a 24 hpf embryo. In row 2 of Fig. 1 and in Fig. S2, examples are shown of expression of the respective recombinant BACs used for isolation of the cell populations that are our subject. Rows 3 and 4 of Fig. 1 show the respective FACS separation of cells from cell fragments and small cell aggregates by forward/side scatter, and the separations of live (in each case, the vast majority) from otherwise compromised cells. The panels in row 5 of Fig. 1 show the GFP gating by which active cells were separated from those not expressing the fluorochrome.

Genes enriched in the transcriptomes of each cell population

The initial project was to assess the efficacy of separating out cell populations. This was done *in extenso* in the preceding study on skeletogenic cells (Barsi et al., 2014). In that study, virtually every one of a large number of genes indicated by transcriptome analysis of the presumed skeletogenic population to be specifically expressed was indeed shown by *in situ* hybridization to be transcribed in skeletogenic cells. This result provided proof of technological principle for the extension to the additional five populations that are the subject of this paper.

Enrichment of specifically expressed transcripts was measured in the sorted cell populations by comparing the transcriptomes of the GFP-expressing populations with those of the non-GFP-expressing cells ('dark' cells) from the same FACS run. Given that incorporation of the recombinereed BACs generating the GFP is mosaic in the injected embryos that were used after disaggregation for the FACS separations, some cells of the same types as the GFP-expressing cells will be included in the dark control populations in these comparisons. However, although the selected cell types could constitute all of the respective GFP populations, they will constitute only a minor fraction (F) in each case of the dark populations: $F = (aT - G)/T$, where a is fraction of the total embryo accounted for by the given cell population, T is the number of cells in the total embryo at the relevant stage and G is the number of GFP-positive cells per embryo recovered. Thus, the maximal enrichment of specifically expressed transcripts that could be expected in comparing GFP-expressing with dark cell populations is F^{-1} . We know the approximate numbers of cells expressing these regulatory genes at the relevant stages from the *in situ* hybridizations and extensive other studies on expression of these genes (cf. references cited in first paragraph of Results). For the populations in this study, F values ranged from ~ 0.05 to ~ 0.085 , so the maximal enrichment that could be expected on the basis of passively enhanced concentration is of the order of 10- to 20-fold [excluding, of course, any particular gene(s) that are driven by the regulatory gene overexpressed in the recombinereed BAC used for selection of the experimental population]. This is completely consistent with what is observed for the overwhelming majority of the genes in the enriched populations in the scatter plots shown in the following figures, given the observed statistical spread.

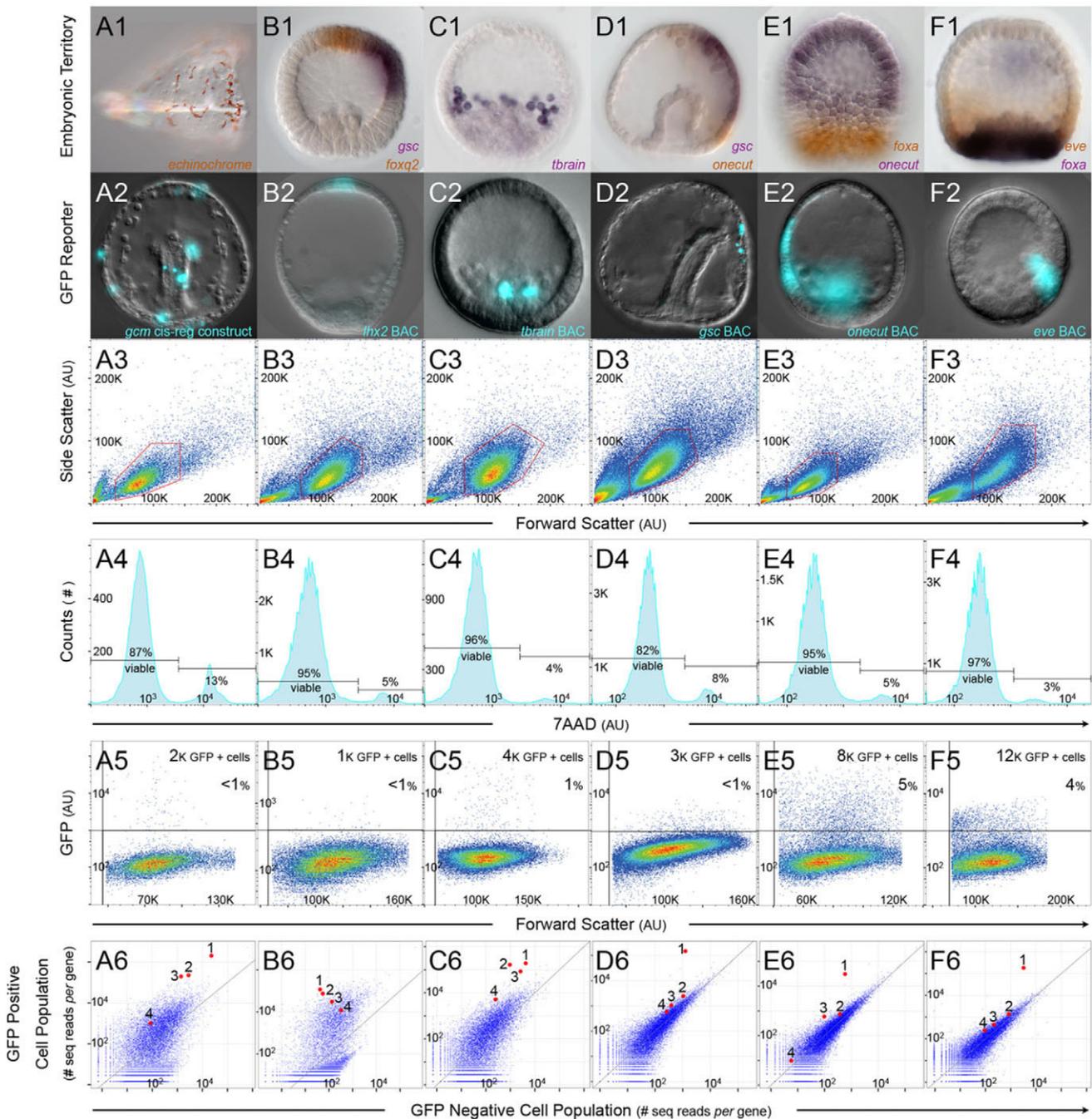


Fig. 1. See next page for legend.

An initial check is shown in row 6 of Fig. 1. Here, the positions of several transcripts known to be expressed specifically in the respective territories are highlighted on comparative transcriptome plots corresponding to each of the six cell populations (data points colored red). The only regulatory transcripts enriched $\gg F^{-1}$ will be those resulting from read-through of the recombineered BACs expressing the GFP, which are injected in multiple copies per egg [although the recombineered BACs include poly(A) sequences following the GFP, these do not function efficiently enough to induce 100% chain termination, and we invariably note some read-through products in transgenic embryos expressing these BACs]. The identities of the enumerated transcripts are specified in the legend of Fig. 1, and all of them appear on the upper edge, or

beyond, of the distributions in the comparative scatter plots. This indicates that, as should be the case, transcripts known to be expressed in the respective embryonic domains are enriched in the selected transcriptomes.

To obtain more comprehensive and unbiased evidence, we randomly chose transcripts that, according to the analysis, were enriched in each population, prepared *in situ* hybridization probes, and determined where in the embryo these transcripts are in fact located. These results are summarized in Fig. 2 and shown in detail in Fig. S3. The position of the data points representing transcripts chosen for whole-mount *in situ* hybridization (WMISH) examination can be seen in the scatter plots, outlined in black. Note that transcripts of different prevalence and different levels of

Fig. 1. Diversity of embryonic cell types analyzed. (A-F) Experiments that correspond to each of the six cell types recovered: (A) pigment cells; (B) apical subdomain; (C) skeletogenic cells; (D) oral ectoderm subdomain; (E) ciliated band; and (F) veg1. (Aa-Fa) The embryonic territory of each cell type according to endogenous gene expression of key marker genes or compounds. (Aa) Pigment cells are visible from the lateral view of a larva because of the echinochrome they express. (Ba-Fa) Select marker gene expression evidenced by way of RNA *in situ* hybridization. (Ba) Lateral view of a blastula shows the apical subdomain in orange as revealed by foxq2 mRNA localization, relative to gsc expression in blue. (Ca) Oral view of a mesenchyme blastula depicts skeletogenic cells in blue as revealed by tbrain mRNA localization. (Da) Lateral view of gastrula shows the oral ectoderm in blue as revealed by gsc mRNA localization, relative to onecut expression in orange. (Ea) Oral view of gastrula shows the ciliated band in blue as revealed by onecut mRNA localization, relative to foxa expression in orange. (Fa) Oral view of blastula shows veg1 in orange as revealed by eve mRNA localization, relative to foxa expression in blue. (Ab-Fb) *In vivo* GFP expression superimposed onto differential interference contrast micrographs, demonstrating that reporter expression faithfully recapitulates each of the corresponding embryonic territories. (Ab) gcm cis-regulatory construct drives GFP expression exclusively within pigment cells. (Bb) lhx2 BAC reporter drives GFP expression exclusively within a subdomain of the apical plate. (Cb) tbrain BAC reporter drives GFP expression exclusively within skeletogenic cells. (Db) gsc BAC reporter drives GFP expression exclusively within a subdomain of the oral ectoderm. (Eb) onecut BAC reporter drives GFP expression exclusively within the ciliated band. (Fb) eve BAC reporter drives GFP expression exclusively within veg1. (Ac-Fe) Flow cytometry used for the recovery of each cell type. (Ac-Fc) Color-coded data points (orange spectrum represents higher density) are correlated with cell volume; events contained within the red demarcation were visually corroborated to constitute individual cells and chosen for subsequent fluorescence-activated cell sorting analysis. (Ad-Fd) The fraction of cells for each experiment that remained viable after disaggregation and flow cytometry, according to 7AAD incorporation. (Ad) 87% of cells derived from gcm:GFP transgenic embryos remained viable. (Bd) 95% of cells derived from lhx2:GFP transgenic embryos remained viable. (Cd) 96% of cells derived from tbrain:GFP transgenic embryos remained viable. (Dd) 82% of cells derived from gsc:GFP transgenic embryos remained viable. (Ed) 95% of cells derived from onecut:GFP transgenic embryos remained viable. (Fd) 97% of cells derived from eve:GFP transgenic embryos remained viable. (Ae-Fe) Color-coded data points (orange spectrum represents higher density) show the total number of GFP-expressing cells recovered for each embryonic territory: (Ae) 2K pigment cells; (Be) 1K apical subdomain cells; (Ce) 4K skeletogenic cells; (De) 3K oral ectoderm subdomain cells; (Ee) 8K ciliated band cells; and (Fe) 12K veg1 cells. (Af-Ff) Comparative transcriptome analysis illustrates the abundance of every mRNA species expressed in the cell type of interest, relative to control. Select data points are highlighted in red and accompanied by a numeric identifier. (Af) 1, pks1; 2, fmo3; 3, sult1c; 4, betaLi. (Bf) 1, lhx2; 2, foxq2; 3, z133; 4, ankAT-1. (Cf) 1, msp130; 2, tbrain; 3, p19; 4, alx1. (Df) 1, gsc; 2, bra; 3, lefty; 4, nodal. (Ef) 1, onecut; 2, univin; 3, slsp1; 4, pax2/5/8. (Ff) 1, eve; 2, wnt8; 3, wnt1; 4, wnt16. 7AAD, 7-aminoactinomycin D; AU, arbitrary units; BAC, bacterial artificial chromosome; GFP, green fluorescent protein; K, thousand.

enrichment were chosen for each sample. In Fig. 2, the results of such corroboration are exemplified in the three WMISH images below each scatter plot (labeled 1, 2, and 3). The expression pattern of several other transcripts subjected to *in situ* hybridization can be found in Fig. S3 (see legend for identities of these transcripts). The result is clear; exactly as in the preceding study on transcripts enriched in skeletogenic cells (Barsi et al., 2014), transcript sets isolated by this method are, by direct observation, specifically expressed in the expected spatial domains. FACS isolation of the cells expressing the respective regulatory genes thus effectively worked to separate out these embryonic cell populations, irrespective of the very distinct cell types and embryonic locations sampled. ‘Specifically expressed’ here does not necessarily mean specifically and exclusively expressed, because in some cases the WMISHs revealed that the transcript also appears in another different specific location. This will, of course, increase the value of

a in the above expression, because the fraction of embryonic cells containing the given transcript would in such cases represent larger fractions of the embryo than for transcripts expressed in single differentiated cell types, thereby decreasing the possible degree of enrichment of such transcripts.

A research resource: lists of effector genes whose transcripts are significantly enriched in each isolated cell population

Preferentially expressed genes in each sorted population are listed in order of statistical improbability that their transcripts are not enriched, given the transcript read count distribution, as computed using the R packages identified in the Materials and Methods. The *P* value distributions shown range from $<10^{-10}$ to $<10^{-2}$ for each data set. The abbreviated gene names (for synonyms see the Echinobase genome database <http://www.echinobase.org/Echinobase/Search/SpSearch/index.php>), ID numbers, crude ontology assignment and enrichment significances are listed in Table S1A-E. As we show explicitly in the following description, the enriched transcript sets for all six of these preparations are almost entirely non-overlapping; that is, only a minute fraction of the genes represented in any one enriched transcript population is also represented in the enriched transcript population of another. Thus, within the confines of the measurements on these particular cell populations, these transcript sets are enriched uniquely; of course, this does not preclude preferential expression of any of these transcripts in another tissue or cell type or developmental stage than those studied here. Nonetheless, Table S1A-E provides a qualitative resource of a particular kind. For example, Table S1 could be useful for an investigator who wishes to discover what genes are specifically expressed in ciliary band cells, which generate characteristic long cilia, but are not expressed in cells that make ‘normal’ cilia, such as the epithelial veg1 cells or oral cells immediately surrounding the stomodeum among the populations studied here. Or it might be of interest to explore further the genes encoding signaling effector components that are specifically represented in the eve-expressing (veg1) cell population (Table S1E), because this domain includes the future ectoderm-endoderm boundary of the embryo.

Tables S1A and S1B represents differentiated or differentiating cell types (pigment cells and neurogenic cells, respectively), and all the effector gene transcripts that are expressed specifically in each of these cell types should be included in the enriched transcript sets, provided the genes are expressed above the one to a few transcripts per gene-cell, our minimal possible effective cut-off. Sea urchin skeletogenic effector gene sets have been previously reported by Barsi et al. (2014) and thus, are not included Table S1. Additionally, alternative methods have produced similar, albeit smaller-scale catalogs (Zhu et al., 2001; Livingston et al., 2006; Rafiq et al., 2012). Transcriptomes of various classes of neurons have been much studied elsewhere (Momčilović et al., 2014; Tu et al., 2014), although the developmental functions in neurogenesis of the apical lhx2/9 domain are scarcely resolved. However, our understanding of the overall differentiated nature of the sea urchin embryonic cells producing the naphthoquinone pigment echinochrome is likely to profit greatly from the resource represented by the enriched pigment cell transcript set in Table S1A.

A relatively ‘universal’ set of effector gene transcripts

In Fig. 3 is shown the exact obverse of the enriched and depleted transcript sets indicated in red and green in Fig. 2. Here is illustrated a set of 495 transcripts no member of which is ever either enriched or underrepresented, to any statistically significant extent, in any of

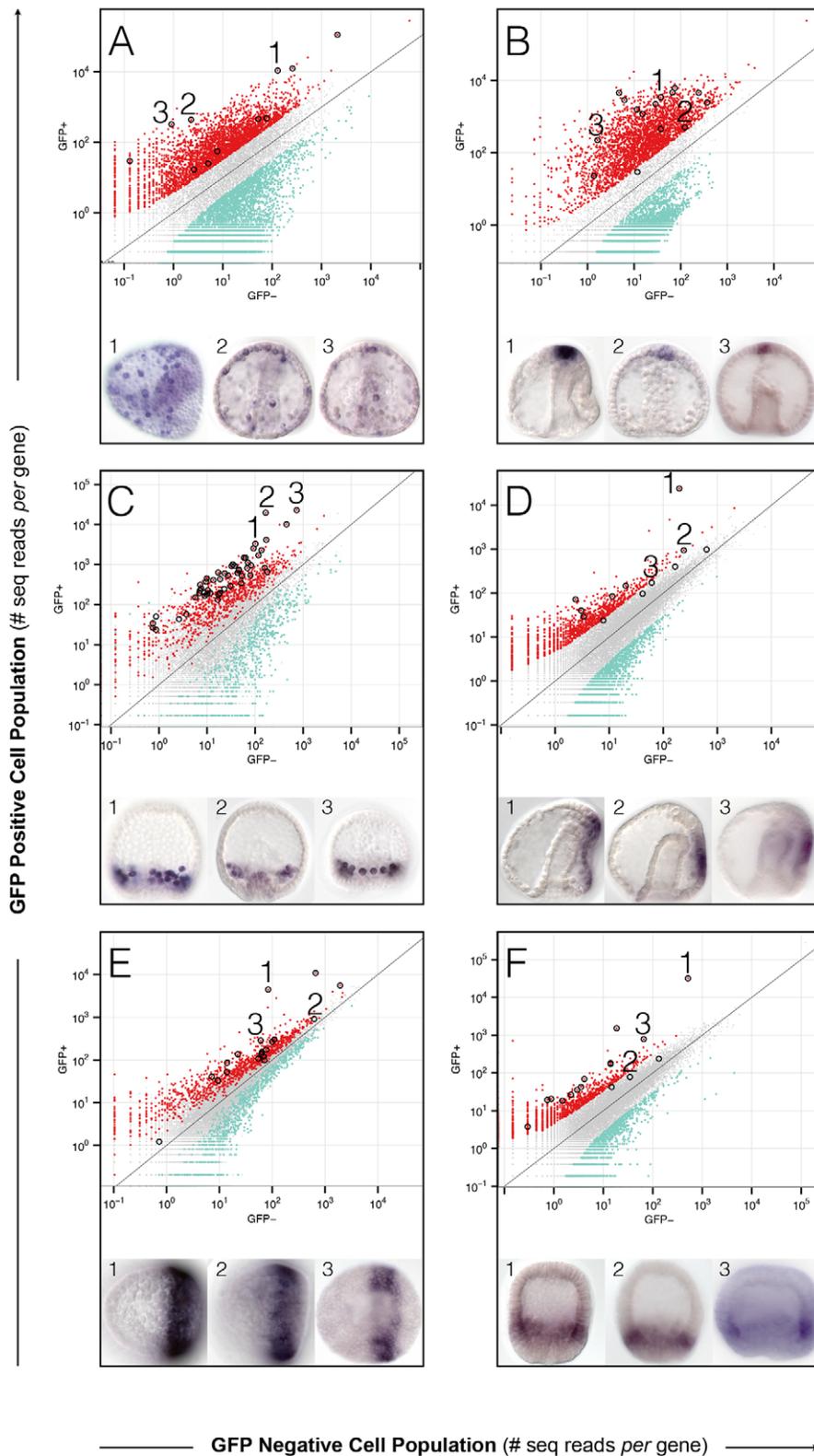


Fig. 2. Enriched populations of transcripts and WMISH corroboration. (A-F) Experiments that correspond to each of the six cell types recovered: (A) pigment cells; (B) apical subdomain; (C) skeletogenic cells; (D) oral ectoderm subdomain; (E) ciliated band; and (F) veg1. Comparative transcriptome analysis for each case is shown in the form of a scatterplot, illustrating the abundance of every mRNA species enriched to the cell type of interest (red data points), relative to control. For all cell types, underrepresented transcripts have been colored green. Data points representing mRNA species corroborated by WMISH have been outlined in black, three of which are enumerated and their corresponding pattern of expression shown immediately below. (A) 1, *gcm*; 2, *whl22.493883*; 3, *gpr54l_3*. (B) 1, *foxq2*; 2, *whl22.510486*; 3, *kifc3L3*. (C) 1, *p19*; 2, *tbr*; 3, *msp130*. (D) 1, *gsc*; 2, *bra*; 3, *nodal*. (E) 1, *onecut*; 2, *z166*; 3, *slsp1*. (F) 1, *eve*; 2, *wnt1*; 3, *wnt16*. GFP, green fluorescent protein. Note that color codes reflect the consensus among replicates, whereas the individual data points shown represent transcript abundance as observed for a single replicate.

the six selected GFP⁺ transcript preparations. The relative prevalence or number of sequencing reads (normalized against total reads) per locus for each transcript species varies by small factors in comparing the six different GFP⁺ preparations, and likewise when comparing such values for each transcript across the six control GFP⁻ preparations. These numbers can be seen in

Table S2. Most importantly, Table S2 shows no systematic enrichment or depletion in comparing the prevalence between GFP⁺ and GFP⁻ preparations of the same population isolate, nor does it indicate the concentration of any particular functional category in this gene set (which could reflect the limit of such ontological categories). These details are all implied by the

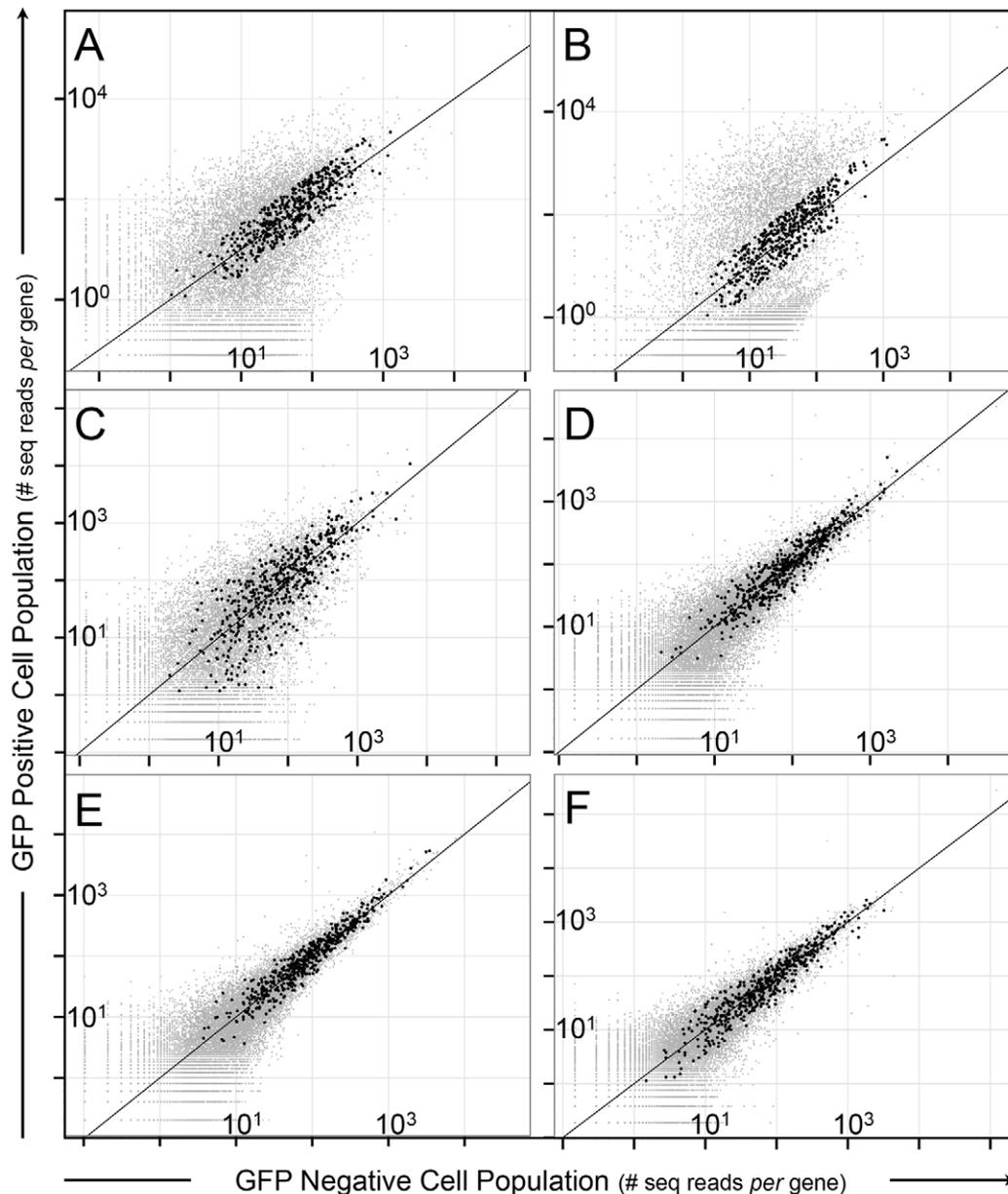


Fig. 3. Cohort of universally expressed transcripts. (A-F) Comparative transcriptome analysis for each of the six cell types recovered: (A) pigment cells; (B) apical subdomain; (C) skeletogenic cells; (D) oral ectoderm subdomain; (E) ciliated band; and (F) *veg1*. Illustrated in each scatterplot is the abundance of every mRNA species expressed in the cell type of interest (plotted along the ordinate), relative to control (plotted along the abscissa). Data points colored black represent 495 transcripts expressed across all six cell types analyzed.

distributions seen in Fig. 3, of which the simplest descriptor is that in every sorted sample the members of this gene set map with the unsorted majority distribution.

This is not meant to indicate the fraction of effector genes that are universally expressed, because we made no attempt to estimate the total number of transcript species behaving like those of Fig. 3, even within the six populations compared here. Furthermore, had we examined other transcript populations, some of these 495 transcripts might have displayed sharp depletion or enrichment profiles. All that can be said is that if there were genes expressed in all cells in all spatial domains of the embryo similarly (per gene, except for minor quantitative variations), their behavior would be that seen in Fig. 3 for the samples we looked at. In fact, a large fraction of the 495 transcripts are likely to be distributed universally in the embryo,

with regard to spatial expression, based on another kind of comparison. Maternal transcripts in sea urchin embryos are in general universally distributed, and they are often universally expressed as ‘housekeeping’ proteins. The maternal transcriptome of these eggs has been studied in detail (Tu et al., 2014; Peter and Davidson, 2015). Out of the 495 transcript species in the set discussed here, 421 turn out also to belong to the maternal transcript component.

Mutual specificity of effector gene sets

The most informative demonstration of differential effector gene use is direct comparison of the specifically enriched populations of transcripts expressed in each of the three differentiated cell types. In the six panels of Fig. 4, such comparisons are shown between the

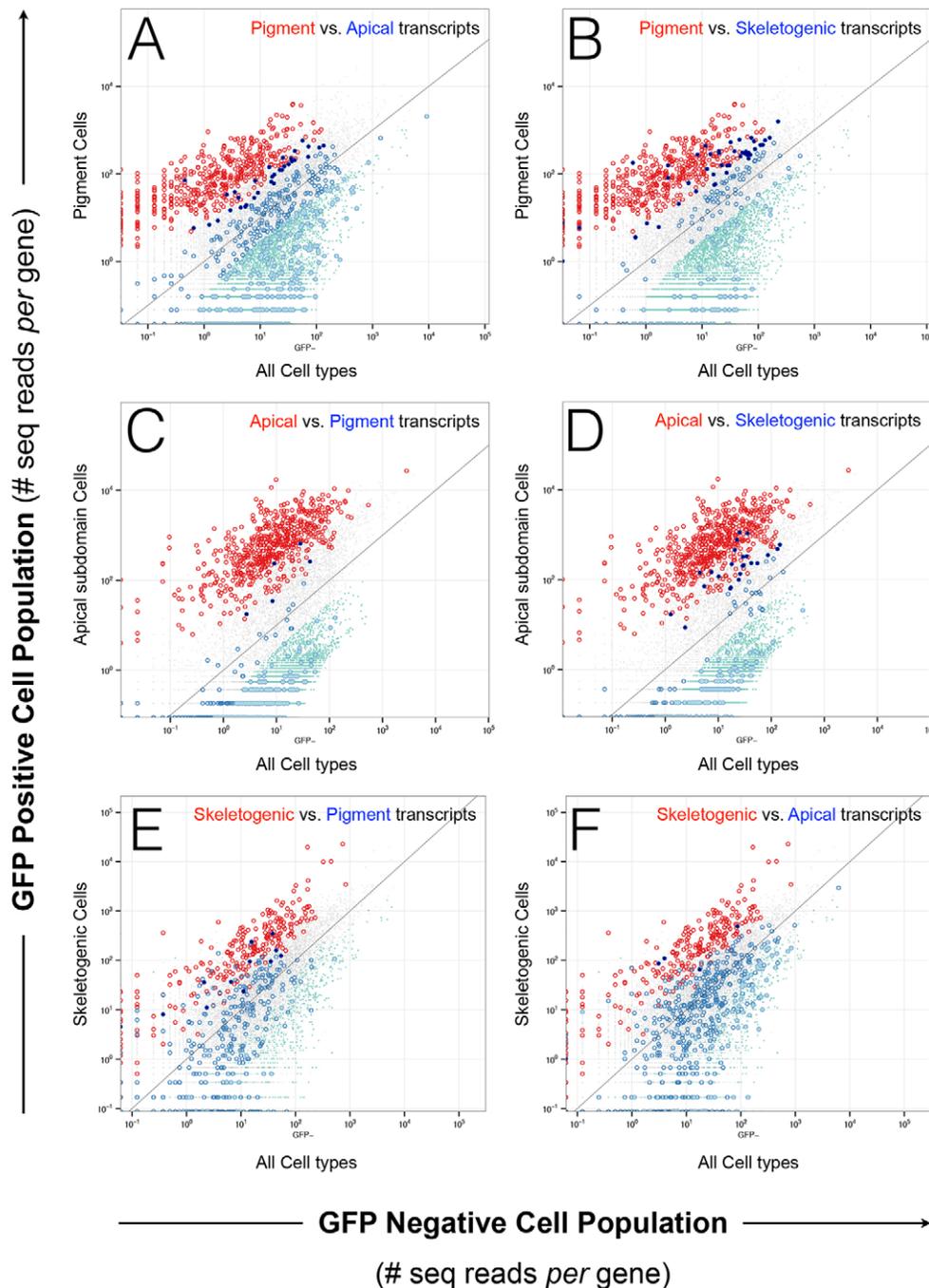


Fig. 4. Specificity of effector gene sets. (A-F) Comparative transcriptome analysis for three differentiated cell types: (A,B) pigment cells; (C,D) apical subdomain; and (E,F) skeletogenic cells. Illustrated in each scatterplot is the abundance of every mRNA species expressed in the cell type of interest (plotted and labeled along the ordinate), relative to control (plotted and labeled along the abscissa). Data points outlined in red represent transcripts uniquely expressed in the cell type specified at the top of each plot (also in red). Underrepresented transcripts are shown as green data points. Data points outlined in blue represent transcripts uniquely expressed in a different cell type, specified at the top of each plot (also in blue). Of these, those that are depleted relative to the data set shown have been shaded light blue. Likewise, those that are enriched have been shaded dark blue. All remaining data points have been colored light gray. As described in detail in the Materials and Methods, these assignments are the result of statistical prevalence analysis; the main quantitative importance of this figure is that transcripts shown in red are enriched beyond the upper bound of a ± 0.05 probability envelope; depleted transcripts in light green lie beyond the lower boundary of this envelope; and the gray transcript population gives the distribution of insignificantly enriched or depleted transcripts within this envelope. Note that color codes reflect consensus among the replicates, whereas the individual data points shown represent transcript abundance as observed for a single replicate.

enriched cohorts of transcripts for each cell type (red) co-plotted in this same transcriptome with the enriched transcript cohorts of each of the other two cell types (blue). Thus, in Fig. 4A we see the distribution of the pigment cell-enriched cohort (red) compared in the same sorted pigment cell scatter plot with the distribution of those transcripts that were identified as enriched specifically in the apical neurogenic cells (blue). Likewise, in Fig. 4B the distribution of the enriched transcript cohort of the skeletogenic cells is co-plotted on the sorted pigment cell scatter plot. In Fig. 4C, the distribution of the enriched apical neurogenic transcript cohort, as seen in the sorted *lhx2/9+* cell transcriptome, is compared with that of the enriched transcript cohort of the pigment cells in the same neurogenic cell transcriptome; and in Fig. 4D the enriched apical

neurogenic transcript cohort is compared with the transcript cohort identified as enriched in the skeletogenic cells, again plotted on the neurogenic cell transcriptome distribution. In Fig. 4E, the distribution of transcripts identified as enriched in pigment cells is plotted on the transcript distribution of the sorted skeletogenic cells; and in Fig. 4F the neurogenic cell enriched cohort is plotted on the same sorted skeletogenic cell transcriptome. With minor variations, all six comparisons reveal the striking individuality of these sets of enriched transcripts with respect to one another. In each panel of Fig. 4, the dark blue dots represent transcripts recovered in the comparator enriched transcript set that co-mingle with the red dots displaying the population of enriched transcripts in that sorted sample. This would mean that the genes encoding these transcript

species are preferentially transcribed in both cell types. Compared with the hundreds of transcripts enriched in each cell type, for example, there are only five transcript species of the pigment cell-enriched cohort also identified in the apical neurogenic cell-enriched cohort (Fig. 4C), and only 22 skeletogenic cell-enriched transcripts also enriched in the apical neurogenic cells (Fig. 4D). By this analysis, the enriched transcript sets of skeletogenic and pigment cells are slightly more similar (larger numbers of dark blue dots), which as we note in the Discussion is to be expected, because aspects of their functions are similar at the stages compared. Most striking, however, is the sharp depletion in each sorted population of transcripts present in the enriched transcript sets of the other two cell types, in every comparison of Fig. 4. These are indicated by the light blue symbols located below the statistical envelope within which reside the mass of gray dots representing non-selected transcripts. Many, in some cases most, of the comparator enriched population is in fact absent entirely, or these severely depleted transcripts are seen only along the abscissa or ordinate axes of the selected cell population. Thus the genes encoding the severely depleted transcripts are not transcribed significantly in each of the specific cell types or are represented at levels too low to be meaningful biologically.

DISCUSSION

The mechanistic essence of embryogenesis is installation of spatially differential regulatory gene expression. This is the fundamental process by which diverse sets of genes encoding transcription factors are respectively activated in the appropriate regions of the early embryo. Regulatory states are thus regionally formulated, and they serve as the causal determinants of embryonic fates; all developmental events downstream of the progression of regulatory states depend in each region either directly or indirectly on effector genes activated in response to these different regulatory states, i.e. on signaling genes, cell biology genes and differentiation effector genes. Much beautiful experimental analysis has now closed the circle, at least in fortuitous model systems, and demonstrated that the progression of determinant regulatory states in embryogenesis is in turn the direct read-out of the network of regulatory gene interactions encoded in the genomes of the embryo (Peter and Davidson, 2015). However, the exact course of the regulatory events that culminate in the appearance of differentiated embryonic cell types remains a challenging gray area. Considered embryo-wide, this aspect of the process is asynchronous and spatially complex, because some embryonic regions give rise uniquely and quickly to terminal cell types, whereas other domains of the same embryo are engaged in progressive specification of spatial regulatory states long in advance of resolution of terminally differentiated gene expression.

The sea urchin embryo remains the exemplar of extensive embryonic gene regulatory network analysis. Nonetheless, only two subregions of the solved networks extend to activation of at least some immediate differentiation gene drivers, and thereafter to transcription of known differentiation effector genes. One of these is the skeletogenic domain, where the specification gene regulatory network is apparently complete and where several in-depth studies have addressed activation of biomineralization and other effector genes (reflected in the tbrain domain of Fig. S4). The other is specification of pigment cells, another early differentiating lineage of this embryo, in which several genes encoding enzymes engaged in synthesis of the naphthoquinone pigment are known. These two differentiating embryonic cell types are isolated and characterized in the present study. A third, also isolated and comparatively analyzed

here, is the neurogenic cell population expressing the *lhx2/9* gene. This domain is one of the few embryonic domains remaining for which we still lack comprehensive knowledge of the transcriptional interactions responsible for establishment of its regulatory state (including *lhx2/9* activation).

The assays, sequence identities and validations presented in Figs 1-3 and the supplementary figures and tables of the present paper show that the transcriptomes of the six populations of embryonic cells studied here are indeed mutually unique. Three of these populations, the ciliary band cells, the *veg1* cells and the circumstomodae cells, are each a heterogeneous mixture of different future cell fates, including diverse incipient differentiations. As is to be expected, the results of intercomparisons among these populations (Fig. S5) are more nuanced than those that are so strikingly revealed in the comparisons of the enriched transcript sets shown for the three differentiated cell populations in Fig. 4. In Fig. 4, as described above, the enriched effector gene transcript set of each differentiating cell type is compared against that of each of the other two cell types. The results in each panel provide a textbook illustration of differential effector gene expression early in development. This is illustrated in three ways in each of the six panels of Fig. 4. First, each expresses several hundred transcript species in highly specific ways in respect to all the other cells in the embryo, shown in red. Second, only minute fractions of the enriched transcript set of each cell type are ever found in the enriched transcript set of another cell type, shown by the few superimposed dark blue and red points in each panel. The only minor exception is seen in comparisons of the pigment cell- and skeletogenic cell-enriched transcript populations, which is to be expected because these are both migratory cells of mesodermal embryological origin. Third, again with the quantitatively minor exception of the pigment cell- and skeletogenic cell-enriched populations, most of the genes encoding the enriched transcript populations of each cell type are not expressed at all in the other two cell types, so that their relative prevalence is negligible, or are otherwise significantly depleted in the log probability distributions, shown in light blue. Taking into account the fact that the comparisons are with unsorted cells, which include the cell types in question, these essentially mean the same thing, i.e. that the genes giving rise to the enriched transcript populations of each cell type are in general simply not being expressed in the other cell types.

The pigment cell and skeletogenic cell gene regulatory networks show that the effector genes are expressed if and when their driver regulatory genes are expressed in the differentiating cell-regulatory state. In this situation, absence of expression unequivocally means absence of the requisite sequence-specific transcriptional activators. The non-expressed genes are never turned on in the lineages leading to the pigment cells, skeletogenic cells and neurogenic cells of this study. Thus, the present study shows explicitly how the speed and power of spatially differential regulatory gene expression causes large-scale, cell type-specific deployment of effector gene cohorts in this embryo within little more than a day after fertilization.

MATERIALS AND METHODS

BAC reporters

Bacterial artificial chromosomes have been genetically engineered to express GFP in lieu of an endogenous gene, the first exon of which is replaced with a GFP cassette by way of homologous recombination. Each BAC harbors the entire locus of a gene that is exclusively expressed in the cell type of interest. The conceptual basis for using BACs as a means to label individual cell types is the assumption that they are large enough to harbor

the complete cis-regulatory apparatus that governs the expression of a marker gene. This is the principle that confers spatial and temporal precision of expression to a BAC reporter. In this study, we take advantage of six BACs that have been determined to recapitulate endogenous gene expression. A library containing hundreds of genetically engineered BAC reporters can be accessed at http://www.echinobase.org/Echinobase/bac_table/bac_table.php. All BAC reporters discussed herein are publicly available upon request.

Computational procedures

Data visualization

The results were visualized predominantly in the form of scatterplots made possible by R software (R Development Core Team, 2013; version 3.1.2; <http://www.r-project.org>) and the ggplot2 package (Wickham, 2009; version 1.0.0; <http://ggplot2.org>). The online resource was generated using Shiny software (RStudio Inc. 2013; server version v1.0.0.42, R package version 0.8.0; <http://www.rstudio.com/shiny/>).

Differential gene expression analysis

Raw read counts for each gene locus were used to calculate differential gene expression by way of the R package known as edgeR (version 3.8.5; <http://www.r-project.org>). First, the effective library size for each sample was calculated by the trimmed mean of M-values method, provided in the package. Then, pairs of replicates were used to estimate the biological coefficient of variation (BCV) by way of the generalized linear model (GLM) method, also provided in the package. Differential gene expression was calculated using the GLM likelihood ratio test. A fixed value of 0.4 was found to account accurately for the BCV observed across all samples. The substitution of alternative BCV values had no effect on the vast majority of differentially expressed genes. To facilitate distinguishing between enriched and depleted gene cohorts, we expanded P values from [0.1] to [0.2]: for enriched gene cohorts (Count_GFP+>Count_GFP-), the P values were maintained; for depleted gene cohorts (Count_GFP+<Count_GFP-), the P values were replaced by $[2 - p]$. Thus, a P value close to 0 indicates very significantly enriched, whereas a P value close to 2 indicates very significantly depleted. In the case of samples with multiple replicates, the replicates were considered together for the estimation of enrichment or depletion. Conversely, the universal gene cohorts were defined by having their expanded P values between 0.05 and 1.95 across all cell types analyzed, and their read counts were always >1.

Mapping Illumina sequencing reads

The same pipeline described in our previous work (Barsi et al., 2014; Tu et al., 2014) was used, except that all software was upgraded to the latest version when performing analysis: STAR version 2.4.0b (Dobin et al., 2013) and HTSeq version 0.6.0 (<http://www-huber.embl.de/users/anders/HTSeq/>), the statistics of which can be found in Fig. S6.

Developmental model organism

Adult sea urchins were sourced locally off the coast of Southern California. They were kept at Caltech's Kerckhoff Marine Laboratory before being transferred to Caltech's main campus for experimentation purposes.

Flow cytometry

A FACS Aria Flow Cytometer Cell Sorter (BD Biosciences) was used to isolate individual cells immediately after embryo disaggregation. The only distinction from the standard operating protocol was the use of twice-filtered seawater (0.2 μ m) in lieu of the regular sample diluent. This operational alternative is of biological importance when assessing live cells derived from marine model organisms.

Gene transfer

Sea urchin eggs were briefly treated in filtered seawater (FSW) containing citric acid (0.5 M) and aligned on protamine-coated Petri dishes. FSW containing para-aminobenzoic acid (300 mg/ml) was used in order to facilitate injection. Eggs were fertilized *in situ*, and the resulting zygotes were injected (1 pl/zygote) with reporter BACs (50 ng of DNA per ml of

nuclease-free water). Injection needles were fabricated in house from borosilicate glass capillary tubing (1 mm outer diameter \times 0.75 mm inner diameter \times 100 mm long) using a Flaming/Brown P-80 (Sutter Instruments) micropipette puller. The consecutive micromanipulation of thousands of embryos was achieved on an Axiovert 40 C (Zeiss) compound microscope equipped with a single-axis oil hydraulic MM0-220 (Narishige) micromanipulator and a picospritzer III (Parker) microinjection dispense system. Transgenic embryos were cultured at 15°C in FSW containing trace amounts of penicillin and streptomycin.

Isolation and handling of specific sea urchin embryonic cell populations

The steps of this procedure are all presented in detail in the preceding publication (Barsi et al., 2014). Briefly, the sequence of steps is as follows: (1) injection of recombiner BAC reporters expressing GFP under control of specific regulatory gene cis-regulatory apparatus; (2) disaggregation of the injected embryos after they had attained the appropriate developmental stage; (3) FACS sorting to isolate the desired cell populations; (4) isolation of the sorted cell mRNA and cDNA amplification; and (5) transcriptome sequencing and analysis by standard mapping and statistical methods (Barsi et al., 2014). In the earlier work, we showed that the disaggregation and FACS procedures per se have no quantitative or qualitative effect on the transcriptomes, by comparisons with control embryos. At the time of injection, the spatial accuracy of expression of the incorporated BAC expression constructs was checked microscopically (cf. Fig. 1). Numbers of embryos used to obtain each of the populations of sorted cells analyzed in this work, and other procedural statistics, are also presented in the preceding study.

Microscopy

Both live and fixed transgenic embryos were monitored for accurate reporter expression using an Axioskop 2 plus (Zeiss) compound microscope equipped for fluorescence and differential interference contrast microscopy. Digital images were taken using an AxioCam MRm (Zeiss) camera. Embryos shown were visualized through a 20 \times objective lens, whereas individual cells were imaged using a 40 \times objective lens.

RNA *in situ* hybridization

WMISH was performed on Sp embryos following a published method optimized in our laboratory (Ransick, 2004).

RNA processing

Total RNA was extracted from each of the various cell populations isolated by FACS using an RNeasy Plus Mini Kit (Qiagen). The only distinction from the manufacturer's recommended protocol was a twofold increase in the DNase incubation time.

Availability of raw data

All data supporting the findings communicated in this study have been submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under accession number SRP052830.

Online resource for the research community

All cell-specific data originating from this study can be interrogated using purpose-built query/visualization tools or downloaded in their entirety via Echinobase (http://www.echinobase.org/SpBase/rnaseq/embryonic_territory.html).

Acknowledgements

This work is dedicated to the memory of our mentor, E.H.D. (1937-2015). Many collaborators contributed their invaluable expertise to this project. We would particularly like to acknowledge the expert assistance of Shelley Diamond and her assistants Diana Perez and Pat Koen in developing an efficacious FACS protocol for sea urchin embryonic cells; the generous provision by Dr Enhu Li of the *in situ* hybridization images in Fig. 1; Dr Andrew Ransick for kindly providing the gcm cis-reg construct used to isolate pigment cells and related micrographs used in Figs 1Aa and 2A #1; and the multiple contributions of Erika Vielmas, who authenticated the expression patterns of the injected BACs used to isolate the various cell populations by microscopic examination and assisted with miscellaneous aspects of this study.

Competing interests

The authors declare no competing or financial interests.

Author contributions

J.C.B. and E.H.D. designed the research; J.C.B. and C.C. performed the research; J.C.B., Q.T. and E.H.D. analyzed the data; and J.C.B. and E.H.D. wrote the paper.

Funding

This work was supported by a grant from the National Institute of Child Health and Development [HD067454]; and from the National Institute of General Medical Sciences [RR015044]. Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.127746/-DC1>

References

- Barsi, J. C., Tu, Q. and Davidson, E. H.** (2014). General approach for in vivo recovery of cell type-specific effector gene sets. *Genome Res.* **24**, 860-868.
- Davidson, E.** (1986). *Gene Activity in Early Development*, 3rd edn. New York: Academic Press.
- Davidson, E. H.** (1990). How embryos work: a comparative view of diverse modes of cell fate specification. *Development* **108**, 365-389.
- Defaye, A. and Perrin, L.** (2014). Tissue specific RNA isolation in *Drosophila* embryos: a strategy to analyze context dependent transcriptome landscapes using FACS. *Methods Mol. Biol.* **1196**, 183-195.
- 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.
- Farfán, C., Shigeno, S., Nödl, M.-T. and de Couet, H. G.** (2009). Developmental expression of *apterous/Lhx2/9* in the sepiolid squid *Euprymna scolopes* supports an ancestral role in neural development. *Evol. Dev.* **11**, 354-362.
- Li, E., Materna, S. C. and Davidson, E. H.** (2013). New regulatory circuit controlling spatial and temporal gene expression in the sea urchin embryo oral ectoderm GRN. *Dev. Biol.* **382**, 268-279.
- Livingston, B. T., Killian, C. E., Wilt, F., Cameron, A., Landrum, M. J., Ermolaeva, O., Sapojnikov, V., Maglott, D. R., Buchanan, A. M. and Etensohn, C. A.** (2006). A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Dev. Biol.* **300**, 335-348.
- Lundgren, S. E., Callahan, C. A., Thor, S. and Thomas, J. B.** (1995). Control of neuronal pathway selection by the *Drosophila* LIM homeodomain gene *apterous*. *Development* **121**, 1769-1773.
- Momčilović, O., Liu, Q., Swistowski, A., Russo-Tait, T., Zhao, Y., Rao, M. S. and Zeng, X.** (2014). Genome wide profiling of dopaminergic neurons derived from human embryonic and induced pluripotent stem cells. *Stem Cells Dev.* **23**, 406-420.
- Peter, I. S. and Davidson, E. H.** (2010). The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.* **340**, 188-199.
- Peter, I. S. and Davidson, E. H.** (2011). A gene regulatory network controlling the embryonic specification of endoderm. *Nature* **474**, 635-639.
- Peter, I. S. and Davidson, E. H.** (2015). *Genomic Control Process, Development and Evolution*. Oxford: Academic Press, Elsevier.
- Poustka, A. J., Kühn, A., Radosavljevic, V., Wellenreuther, R., Lehrach, H. and Panopoulou, G.** (2004). On the origin of the chordate central nervous system: expression of *onecut* in the sea urchin embryo. *Evol. Dev.* **6**, 227-236.
- Rafiq, K., Cheers, M. S. and Etensohn, C. A.** (2012). The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development* **139**, 579-590.
- Ransick, A. J.** (2004). Detection of mRNA by in situ hybridization and RT-PCR. *Methods Cell Biol.* **74**, 601-620.
- Ransick, A. and Davidson, E. H.** (2006). cis-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. *Dev. Biol.* **297**, 587-602.
- Ransick, A. and Davidson, E. H.** (2012). Cis-regulatory logic driving glial cells missing: self-sustaining circuitry in later embryogenesis. *Dev. Biol.* **364**, 259-267.
- R Development Core Team** (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Tu, Q., Cameron, R. A. and Davidson, E. H.** (2014). Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*. *Dev. Biol.* **385**, 160-167.
- Wickham, H.** (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Zhu, X., Mahairas, G., Illies, M., Cameron, R. A., Davidson, E. H. and Etensohn, C. A.** (2001). A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development* **128**, 2615-2627.