This document contains **Supplementary Information** for the paper entitled:

General approach for in vivo recovery of cell type specific effector gene sets

Julius C. Barsi[1, 2, *], Qiang Tu[1, 2], Eric H. Davidson[1, *]

[1] Division of Biology, California Institute of Technology (mail-code 156-29) Pasadena, CA 91125 USA

[2] These authors contributed equally to this work.

* Correspondence should be addressed to J.C.B. (barsi@caltech.edu) or E.H.D. (davidson@caltech.edu)

Index

Supplementary Table 1. Genes assessed by RNA in situ
Supplementary Table 2. PMC genes reported in the literature
Supplementary Figure 1. PMC gene regulatory network
Supplementary Figure 2. High-throughput in vivo extraction of target cell types
Supplementary Figure 3. Arriving at a cell type specific transcriptome
Supplementary Figure 4. Referenced data points for RNA in situ data shown
Supplementary Figure 5. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 6. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 7. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 8. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 9. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 10. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 11. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 12. PMC transcriptome sub-set corroborated by RNA in situ hybridization
Supplementary Figure 13. Adhesion and Metabolism genes differentially expressed in PMCs
Supplementary Figure 14. Biomineralization and Calcium Toolkit genes differentially expressed in PMCs
Supplementary Figure 15. Cell Cycle and DNA Replication genes differentially expressed in PMCs
Supplementary Figure 16. Defensome and Immunity genes differentially expressed in PMCs
Supplementary Figure 17. GTPase and Kinase genes differentially expressed in PMCs
Supplementary Figure 18. Histone and Nervous genes differentially expressed in PMCs
Supplementary Figure 19. Metalloprotease and Signalling genes differentially expressed in PMCs
Supplementary Figure 20. Transcription Factor and Zinc Finger genes differentially expressed in PMCs
Supplementary Figure 21. Uncharacterized and Unclassified genes differentially expressed in PMCs
Supplementary Figure 22. Biomineralization gene expression profile dynamics
Supplementary Figure 23. Calcium Toolkit gene expression profile dynamics
Supplementary Note 1. PMC gene regulatory network
Supplementary Note 2. High throughput recovery of labeled cells
Supplementary Note 3. Skeletogenic cell-specific transcriptome
Supplementary Note 4. Gene expression profile dynamics
Supplementary Note 5. Embryo disaggregation

List of genes assessed by RNA in situ hybridization

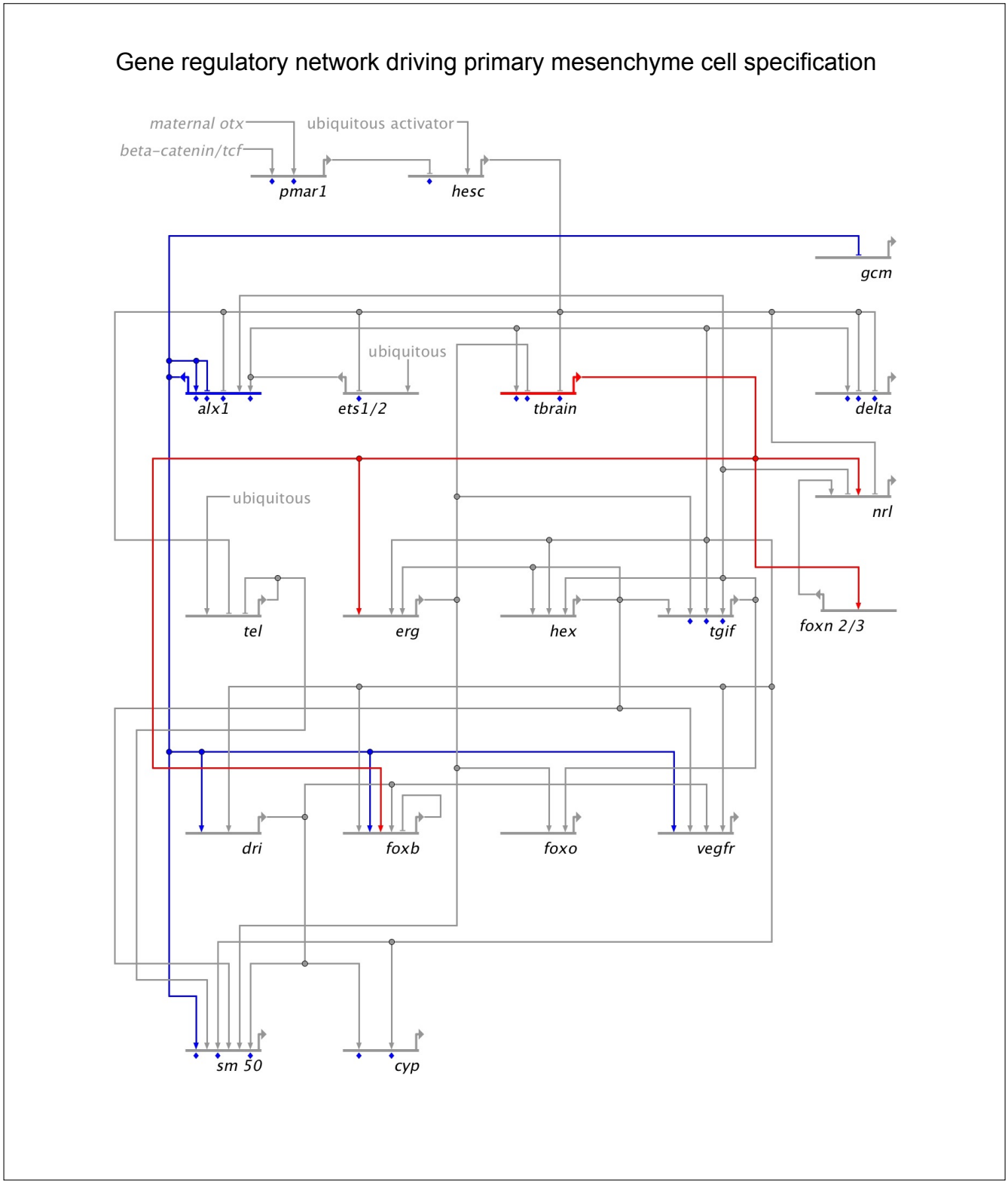
Scatterplot Ref. #	Gene Name	Synonyms	SPU ID	WHL ID
1	*3 apcol*		SPU_003768	WHL22.589988
2	*a2 (IV) l*		SPU_027250	WHL22.589096
3	*alx 1*		SPU_022817	WHL22.731056
4	*ank 2 h*	*ank 3-1*	SPU_000908	WHL22.361430
5	*astacin 1*	*anl 18*	SPU_003612	WHL22.642085
6	*b4 galt 6/8 **		SPU_007019	WHL22.385964
7	*bolal 3 **		SPU_006864	WHL22.66119
8	*c-lectin*		SPU_007882	WHL22.728834
9	*c 14 orf 124*		SPU_001333	WHL22.38
10	*cah 10 l*	*p 19*	SPU_004135	WHL22.642603
11	*calcneur l 1 **		SPU_002945	WHL22.520763
12	*cara 7 la*		SPU_012518	WHL22.446073
13	*cbpdel*		SPU_007682	WHL22.363067
14	*clect 14*	*sm 20*	SPU_005991	WHL22.39473
15	*coil l **		SPU_024897	WHL22.95489
16	*cp*		SPU_027885	WHL22.715503
17	*crocc **	*crocc 2/3/4*	SPU_008503	WHL22.100947
18	*cycpln **	*cyp 1*	SPU_007484	WHL22.733408
19	*deadringer*		SPU_017106	WHL22.544154
20	*deaf 1 l **	*samd 11*	SPU_022540	WHL22.158312
21	*enpep*		SPU_011978	WHL22.25917
22	*enpep 2 **		SPU_016020	WHL22.25901
23	*erg*		SPU_018483	WHL22.552472
24	*ets 1/2*		SPU_002874	WHL22.238821
25	*exoc 4-2 **		SPU_001083	WHL22.694798
26	*p 58 a*	*fcgbp l*	SPU_000439	WHL22.23776
27	*fred*	*frp / ficolin*	SPU_022598	WHL22.427938
28	*gcm*		SPU_006462	WHL22.54333
29	*lg*	*egf / vwc*	SPU_017202	WHL22.139301
30	*mitf*		SPU_008174	WHL22.677144
31	*msp 130*		SPU_002088	WHL22.438943
32	*msp 130 r1*		SPU_013822	WHL22.438976
33	*msp 130 r2*	*ap l*	SPU_016506	WHL22.451280
34	*msp 130 r3*		SPU_006387	WHL22.438994
35	*nadsyn 1*		SPU_003374	WHL22.167662
36	*npnt*		SPU_025235	WHL22.626493
37	*orct*		SPU_012646	WHL22.17128
38	*p 16*		SPU_018408	WHL22.239481
39	*pak (p 21) ph*		SPU_003534	WHL22.294876
40	*pdgfr*	*vegfr l*	SPU_000310	WHL22.240520
41	*ppap 2 b*		SPU_011127	WHL22.310985
42	*slc22a 18*		SPU_012704	WHL22.444466
43	*sm 32*	*sm 50*	SPU_018810	WHL22.742398
44	*tbr*		SPU_025584	WHL22.503644
45	*trpm 3/4*		SPU_011587	WHL22.652421
46	*uncharacterized*		SPU_018406	WHL22.239379
47	*uncharacterized*		SPU_005247	WHL22.501738
48	*uncharacterized*	*pks l*	SPU_028395	WHL22.22317
49	*uncharacterized*	*smtn l 2*	SPU_021845	WHL22.60330
50	*uncharacterized*		SPU_000152	WHL22.452609
51	*uncharacterized*			WHL22.441487
52	*uncharacterized*			WHL22.300872

Supplementary Table 1. Genes assessed by RNA in situ. List of fifty-two genes for which spatial expression in Sp blastula was assessed by RNA in situ hybridization. The first column labeled Scatterplot Reference # serves as a numeric identifier pertaining to Supplementary Figure 4 A2. The second and third columns provide the gene name and known synonyms, respectively. The fourth and fifth columns provide SPU and WHL identification numbers, respectively. Gene identification numbers may be used to retrieve locus coordinates, sequence information, expression data and cell-specific enrichment values from the following online resources (http://www.spbase.org:3838/cellspecific/) and (http://www.spbase.org:3838/quantdev/). (*) denotes genes transiently expressed in PMCs.

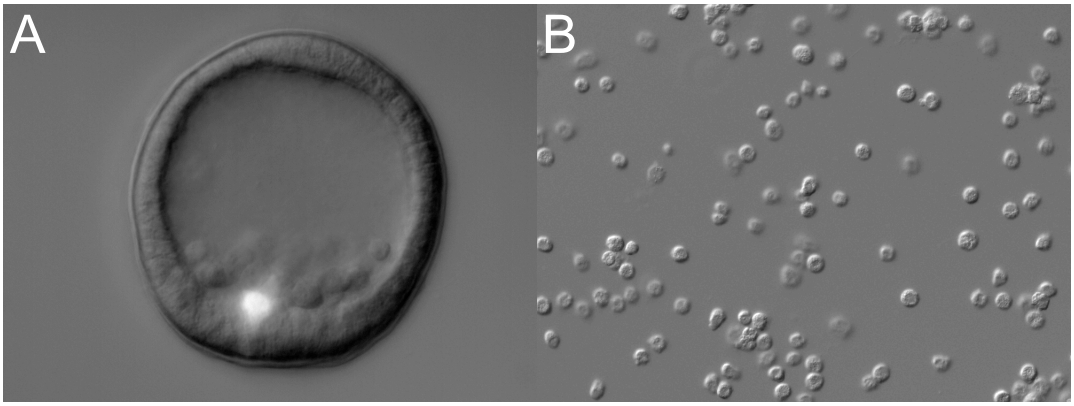
List of genes collated from the literature and the level of enrichment observed in this study

Fold Change	Gene Name	Synonyms	SPU ID	WHL ID
132	*tbr*		SPU_025584	WHL22.503644
34	*sm 30 e*		SPU_004867	WHL22.70610
29	*msp 130 r3*		SPU_006387	WHL22.438994
22	*clect 14*	*sm 20*	SPU_005991	WHL22.39473
19	*lamG*	*egff 3 / p 133*	SPU_012138	WHL22.441481
17	*sm 32*	*sm 50*	SPU_018810	WHL22.742398
16	*p 58 a*	*fcgbp l*	SPU_000439	WHL22.23776
15	*p 58 b*		SPU_000438	WHL22.23620
15	*cah 10 l*	*p 19*	SPU_004135	WHL22.642603
15	*msp 130 r1*		SPU_013822	WHL22.438976
14	*msp 130*		SPU_002088	WHL22.438943
14	*p 16 rel 1*		SPU_018403	WHL22.239320
13	*alx 1*		SPU_022817	WHL22.731056
13	*p 16 rel 2*		SPU_018406	WHL22.239379
13	*msp 130 r2*	*ap l*	SPU_016506	WHL22.451280
12	*ttrspn 19*	*net 7*	SPU_025068	WHL22.748368
12	*sm 29*		SPU_005990	WHL22.39470
12	*sm 37*		SPU_018813	WHL22.742403
11	*fred*	*frp / ficolin*	SPU_022598	WHL22.427938
10	*uncharacterized*	*pks l*	SPU_028395	WHL22.22317
10	*fn 3*	*fgfr 2 / igf 29*	SPU_004746	WHL22.147868
10	*c-lectin*		SPU_007882	WHL22.728834
10	*cara 7 la*		SPU_012518	WHL22.446073
9	*pdgfr*	*vegfr l*	SPU_000310	WHL22.240520
9	*deadringer*		SPU_017106	WHL22.544154
9	*clect 13*	*sm 21*	SPU_005989	WHL22.38332
9	*cycpln*	*cyp 1*	SPU_007484	WHL22.733408
9	*sm 30 b*		SPU_000826	WHL22.518124
9	*col 4 a 2*	*colp 4 alpha*	SPU_015708	WHL22.77073
7	*fox b*		SPU_004551	WHL22.743430
7	*sm 30 f*		SPU_004869	WHL22.70606
7	*pmc 1*	*sm 49*	SPU_027906	WHL22.411802
7	*ets 1/2*		SPU_002874	WHL22.238821
6	*a2 (IV) l*		SPU_027250	WHL22.589096
6	*elav l*		SPU_002324	WHL22.109107
6	*snail*		SPU_001113	WHL22.131363
6	*erg*		SPU_018483	WHL22.552472
6	*hhex*		SPU_027214	WHL22.626418
5	*p 16*		SPU_018408	WHL22.239394
5	*tgif*		SPU_018126	WHL22.614286
5	*3 apcol*		SPU_003768	WHL22.590901
5	*alx 4*		SPU_022816	WHL22.731149
5	*tsp*		SPU_027284	WHL22.553459
4	*delta*		SPU_016128	WHL22.423696
4	*casc 1*		SPU_022570	WHL22.38552
4	*astrnph*	*alpha-spectrin*	SPU_011699	WHL22.521351
4	*p 11*		SPU_017737	WHL22.462376
3	*ckip1*	*cdi*	SPU_005856	WHL22.16108
3	*calumenin*		SPU_018461	WHL22.13135
3	*bstrnph*	*beta-spectrin*	SPU_007720	WHL22.128681
3	*pmar 1*		SPU_014721	SPU_014721
3	*smad 1 / 5 / 8*		SPU_020722	WHL22.347626
3	*fbn 3.2*	*egf-rich / p 41*	SPU_014065	WHL22.145770
3	*arp 1*	*centph*	SPU_001366	WHL22.135434
3	*fox o*		SPU_009178	WHL22.241099
3	*gelslnph*	*advillin*	SPU_000296	WHL22.384598
3	*tel*		SPU_028479	WHL22.589178
2	*fox n 2 / 3*		SPU_015243	WHL22.607384
2	*jun*		SPU_003102	WHL22.318085
2	*stomatin*		SPU_014562	WHL22.527183
2	*hdlbp*	*vigilin*	SPU_016739	WHL22.338109
2	*ctd*		SPU_019434	WHL22.576168
2	*sm 30 C*		SPU_000827	WHL22.518121
2	*cofilin*		SPU_006172	WHL22.497183
2	*actinin*	*alpha-actinin*	SPU_020920	WHL22.675296
2	*fgf 9 / 16 / 20*	*fgf a*	SPU_006242	WHL22.343932
2	*nebnph*	*lasp 1*	SPU_009123	WHL22.586327

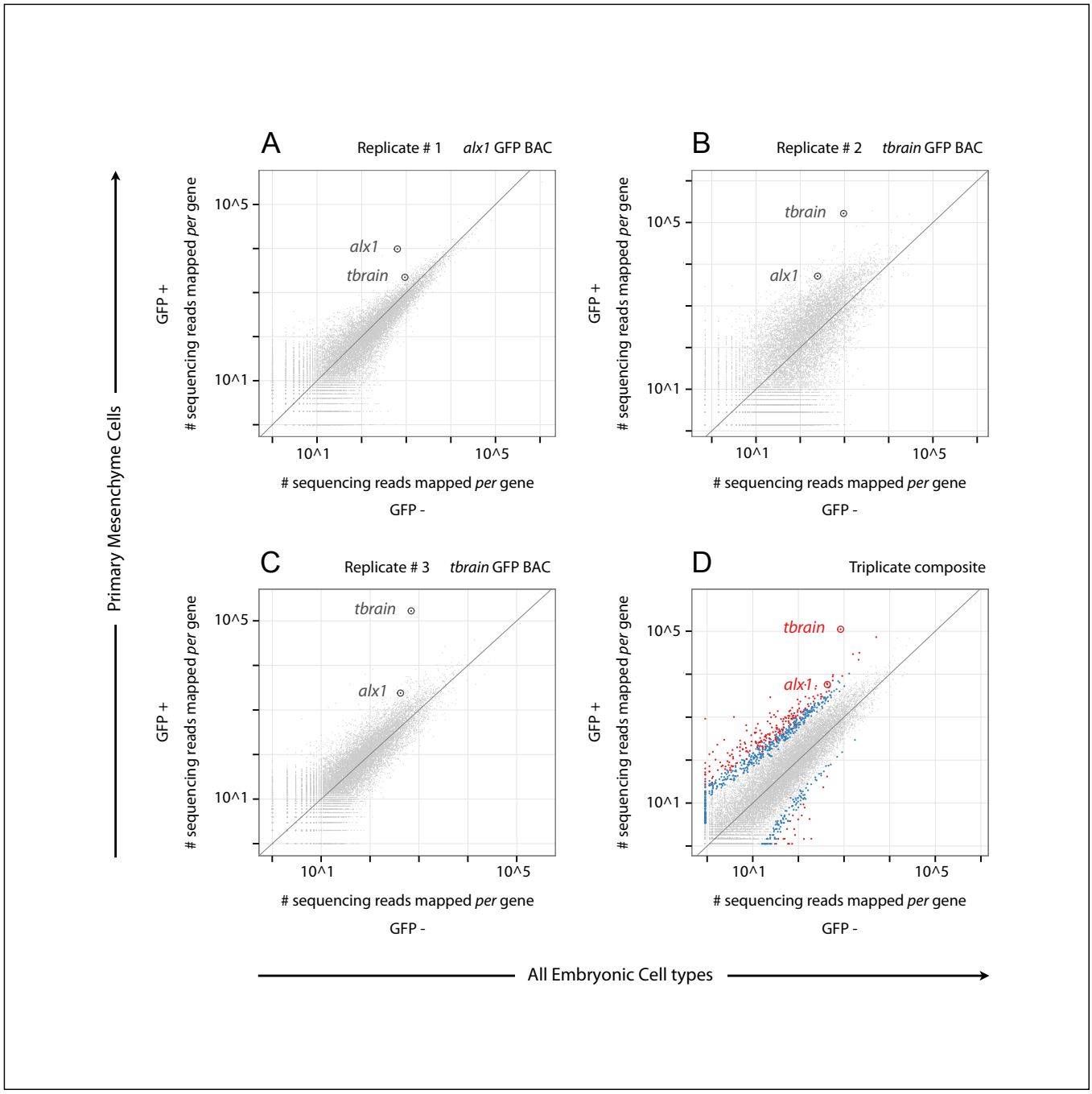
Supplementary Table 2. PMC genes reported in the literature. List of genes known to be differentially expressed in PMCs and their corresponding mRNA enrichment as observed in this study. The first column reflects PMC enrichment in terms of fold change, relative to all other embryonic cell types. The second and third columns provide the gene name and known synonyms, respectively. The fourth and fifth columns provide SPU and WHL identification numbers, respectively. Gene identification numbers may be used to retrieve locus coordinates, sequence information, expression data and cell-specific enrichment values from the following online resources (http://www.spbase.org:3838/cellspecific/) and (http://www.spbase.org:3838/quantdev/).



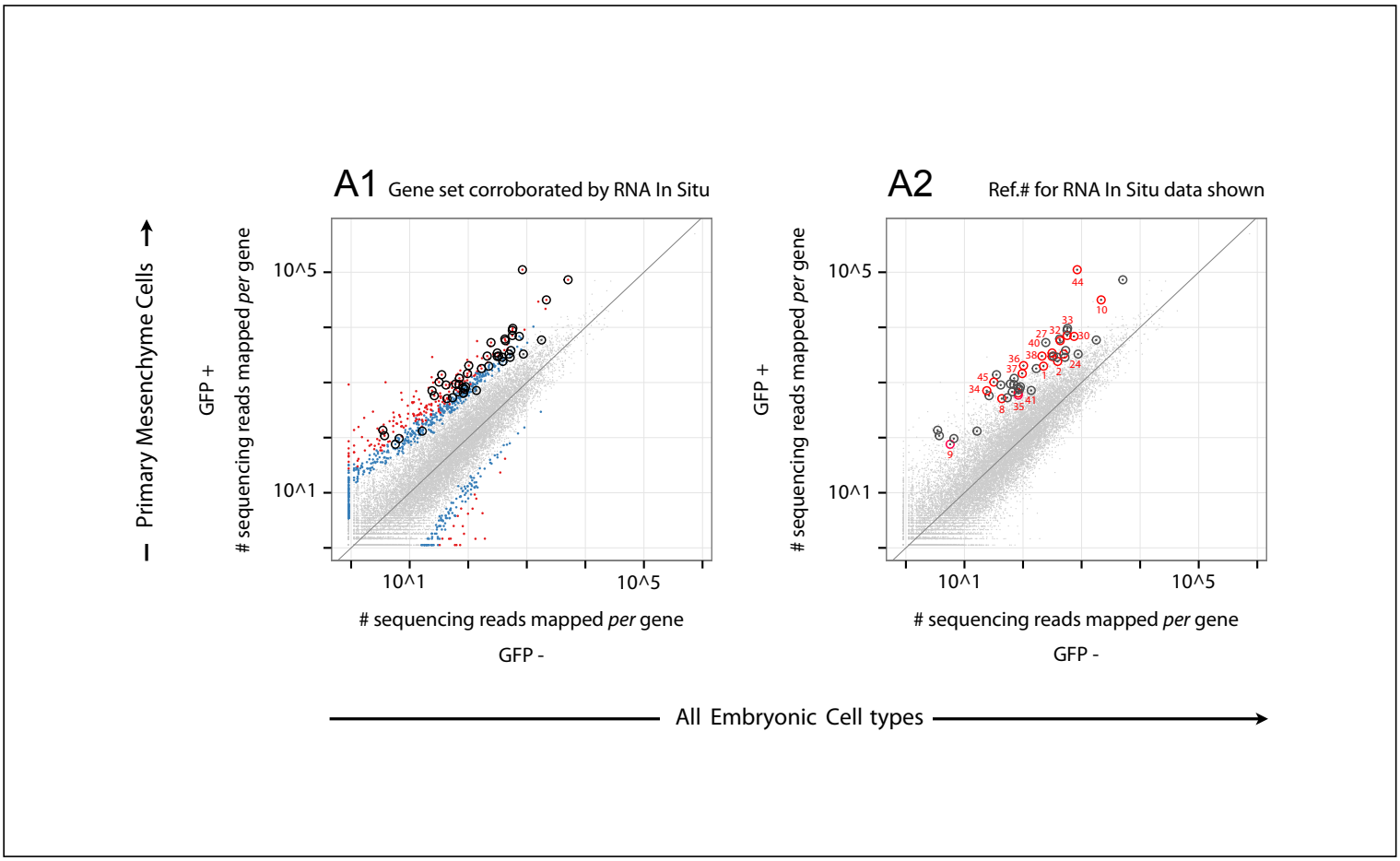
Supplementary Figure 1. **PMC gene regulatory network.** Directed graphical model summarizing decades of research, illustrates the regulatory gene interactions that drive PMC specification. Horizontal bars with a label symbolize the locus of a gene and its associated cis-regulatory apparatus. Edges originating from a bent arrow represent gene product. An arrow or footpad intersecting a horizontal bar signifies a functional protein/DNA interaction that causes activation or repression, respectively. A blue diamond located beneath such an intersection denotes that the DNA binding site has been experimentally identified. All other interactions shown are supported primarily through perturbation data. Ambiguity due to the crossing of edges is resolved through dots that populate every branch point. For an explanation of the developmental regulatory logic documented herein, see Supplementary Note 1.



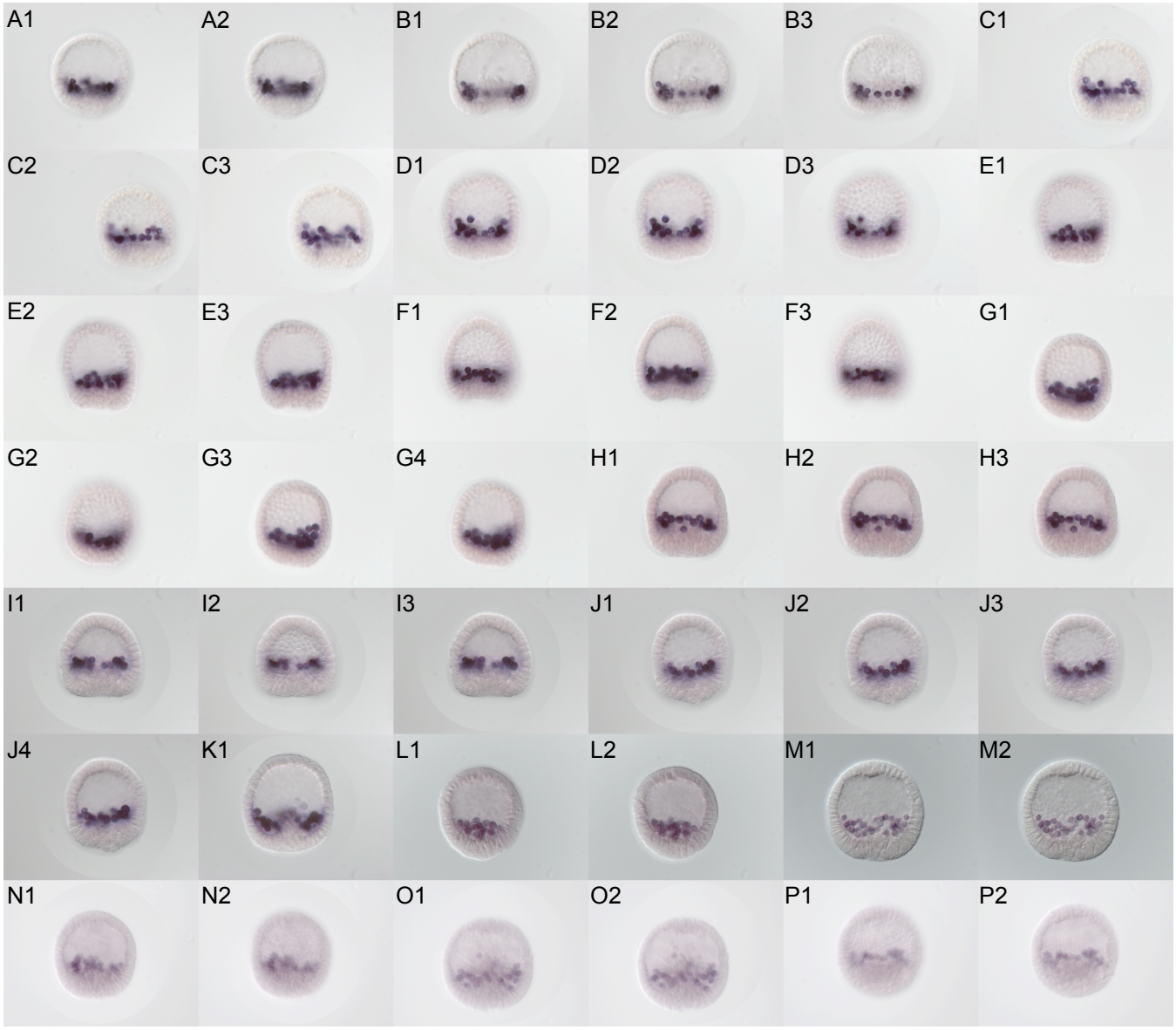
Supplementary Figure 2. **High-throughput in vivo extraction of target cell types.** (A) Live transgenic *Sp* blastula carrying an artificial chromosome that contains GFP under control of the cis-regulatory apparatus governing *alx1* expression. Epifluorescent image overlaid onto DIC background reveals that GFP is exclusively expressed in the target cell. (B) DIC image of individual live cells immediately following embryo disaggregation. Target cells are then segregated by way of FACS.



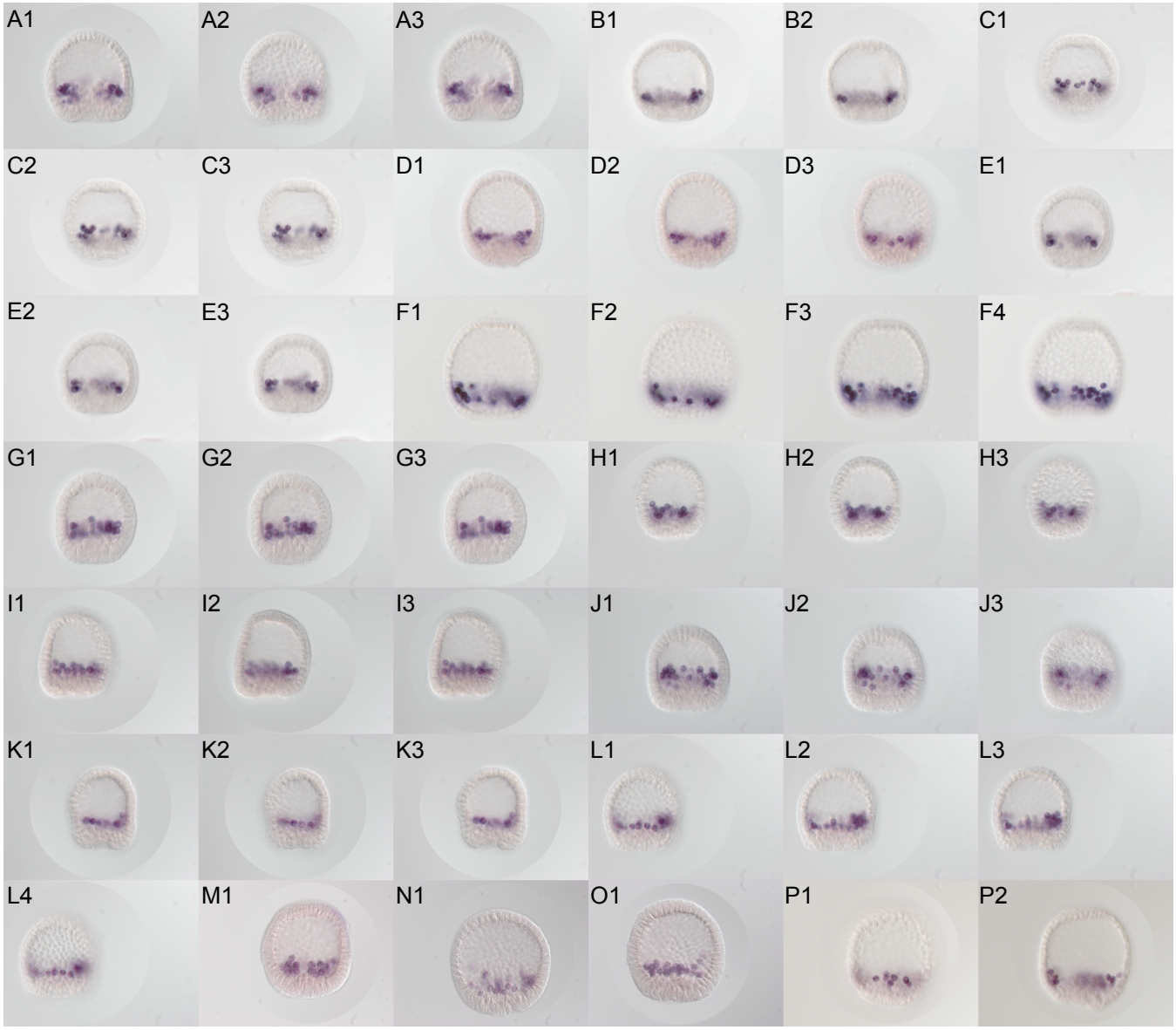
Supplementary Figure 3. **Arriving at a cell type specific transcriptome.** (A-D) Scatterplots compare transcriptome data unique to PMCs (GFP+) with that of all cell types (GFP-). Each data point reflects the relative mRNA abundance of a gene as estimated by the number of sequencing reads that map to its locus. Data points that fall above the diagonal indicate augmented levels of expression among PMCs, relative to other cell types. Data points corresponding to transcription factors *alx1* and *tbrain* have been encircled in order to distinguish them from the rest of the dataset and labeled accordingly. (A-C) Each scatterplot reflects data from one of the three replicates. (A) Replicate #1 obtained by using an *alx1*:GFP BAC. (B) Replicate #2 and (C) Replicate #3 obtained by using a *tbr*:GFP BAC. (D) The values in the graph reflect a consensus among all three replicates. Genes statistically determined to be differentially expressed are shown as data points colored either red or blue to indicate an associated P-value of less than 0.01 or 0.05, respectively.



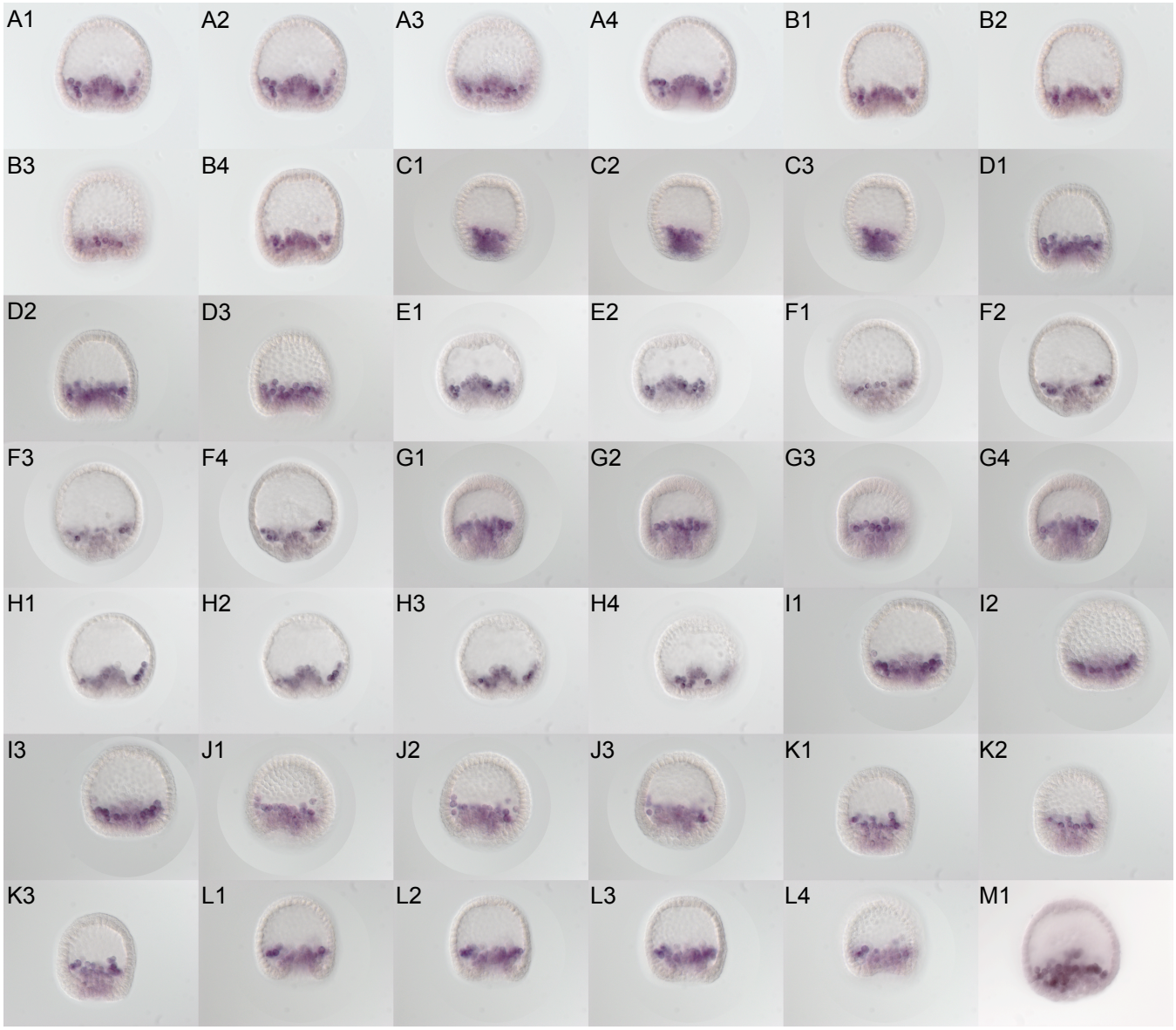
Supplementary Figure 4. **Referenced data points for RNA in situ data shown.** (A1-A2) Scatterplots compare transcriptome data unique to PMCs (GFP+) with that of all cell types (GFP-). Each data point reflects the relative mRNA abundance of a gene as estimated by the number of sequencing reads that map to its locus. The values in the graph reflect measurements taken across three independent replicates. Data points that fall above the diagonal indicate augmented levels of expression among PMCs, relative to other cell types. (A1) Genes statistically determined to be differentially expressed are shown as data points colored either red or blue to indicate an associated P-value of less than 0.01 or 0.05, respectively. Data points further assessed by RNA in situ hybridization have been encircled in black in order to distinguish them from the rest of the dataset. (A2) All data points assessed by RNA in situ hybridization have been encircled in order to distinguish them from the rest of the dataset. Transcripts for which spatial expression is shown in Supplementary Figures 5-12 are highlighted in red alongside a Scatterplot Reference # that may be looked up in Supplementary Table 1.



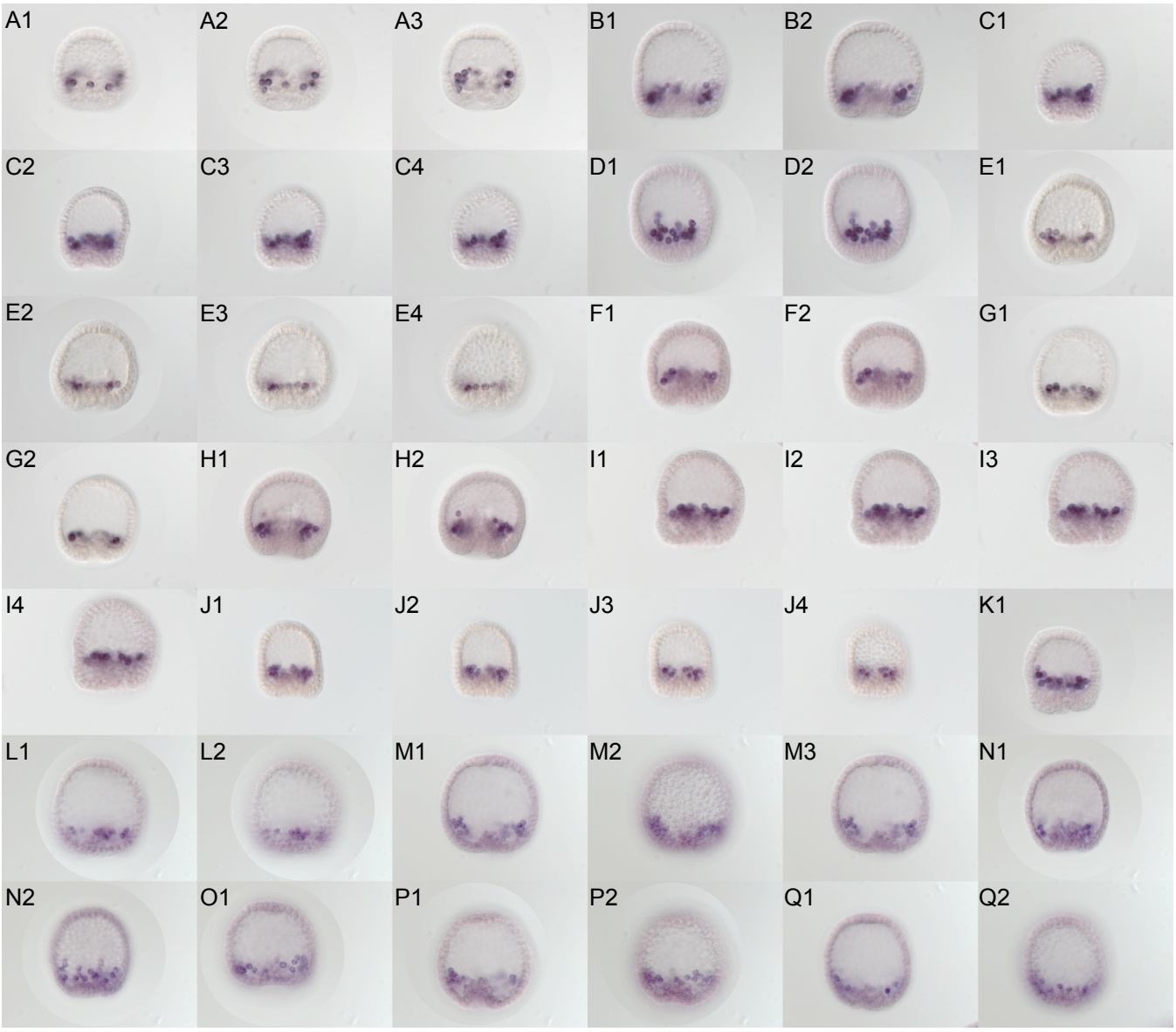
Supplementary Figure 5. **PMC transcriptome sub-set corroborated by RNA in situ hybridization**. Fixed *Sp* blastula exhibiting spatial gene expression with cellular resolution. Purple labeling reveals cells expressing a specific mRNA species. Labeling was achieved by whole mount RNA in situ hybridization. Multiple microphotographs were taken *per* embryo, at different focal planes, in order to visually resolve individually labeled cells. In turn, multiple replicates are shown in order to assess spatial variation in gene expression among biological specimens. (A-M) *msp130r1* mRNA. Expression pattern documented for (A1-A2) biological replicate 1, (B1-B3) biological replicate 2, (C1-C3) biological replicate 3, (D1-D3) biological replicate 4, (E1-E3) biological replicate 5, (F1-F3) biological replicate 6, (G1-G4) biological replicate 7, (H1-H3) biological replicate 8, (I1-I3) biological replicate 9, (J1-J4) biological replicate 10, (K1) biological replicate 11, (L1-L2) biological replicate 12, (M1-M2) biological replicate 13. (N-P) *c14orf124* mRNA. Expression pattern documented for (N1-N2) biological replicate 1, (O1-O2) biological replicate 2, (P1-P2) biological replicate 3.



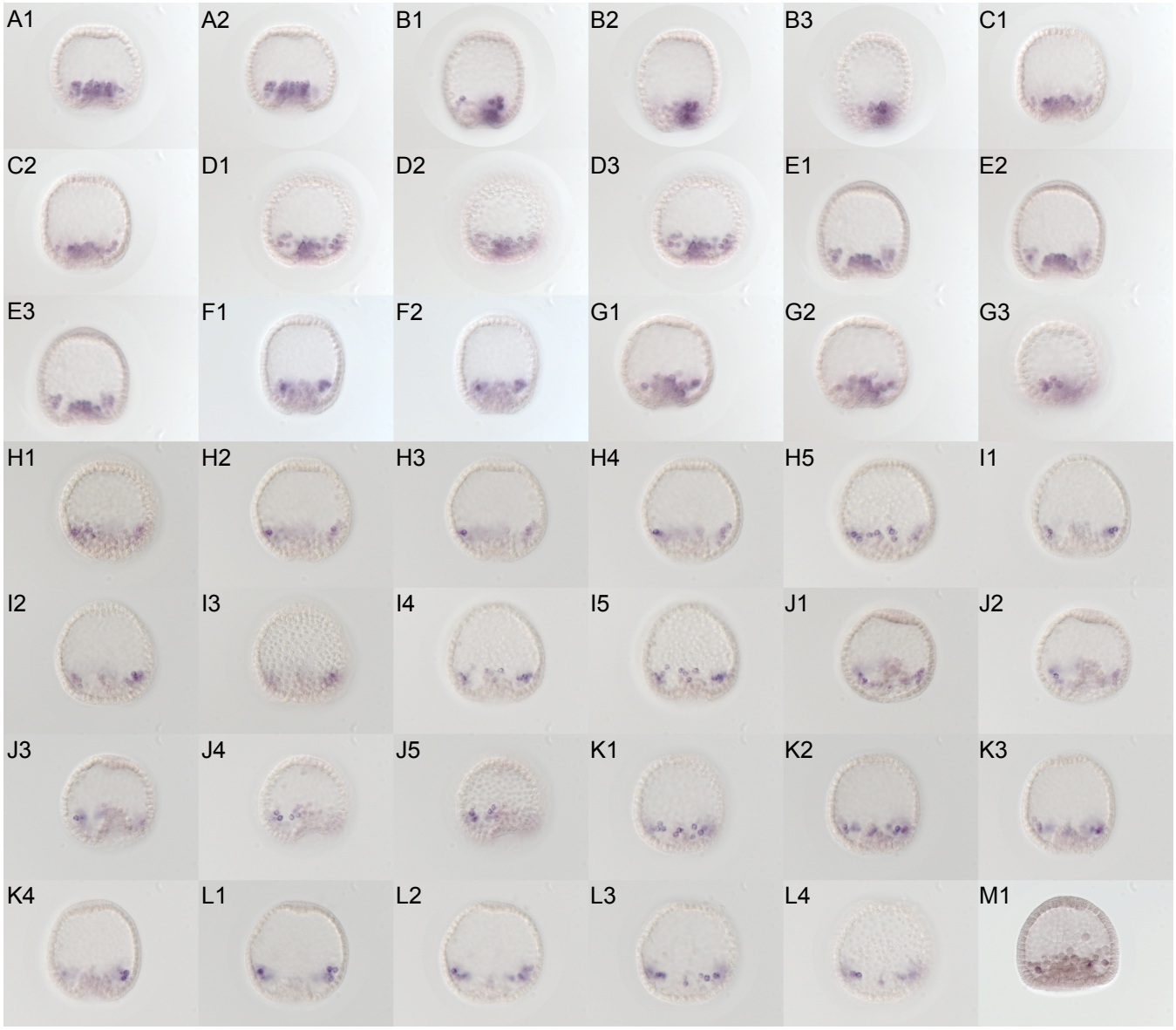
Supplementary Figure 6. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-P) *cah10l* mRNA. Expression pattern documented for (A1-A3) biological replicate 1, (B1-B2) biological replicate 2, (C1-C3) biological replicate 3, (D1-D3) biological replicate 4, (E1-E3) biological replicate 5, (F1-F4) biological replicate 6, (G1-G3) biological replicate 7, (H1-H3) biological replicate 8, (I1-I3) biological replicate 9, (J1-J3) biological replicate 10, (K1-K3) biological replicate 11, (L1-L4) biological replicate 12, (M1) biological replicate 13, (N1) biological replicate 14, (O1) biological replicate 15, (P1-P2) biological replicate 16.



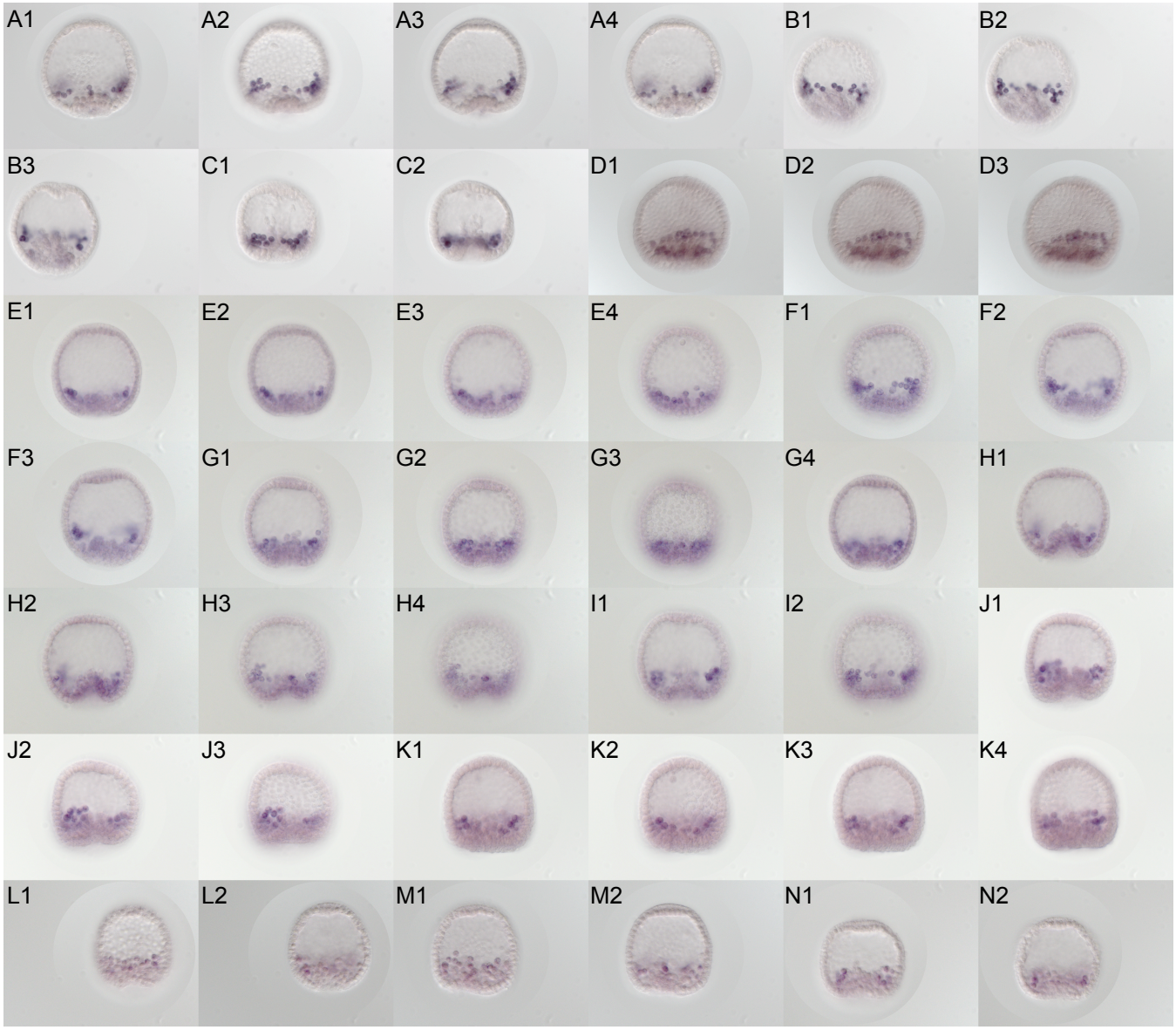
Supplementary Figure 7. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-M) *3apcol* mRNA. Expression pattern documented for (A1-A4) biological replicate 1, (B1-B4) biological replicate 2, (C1-C3) biological replicate 3, (D1-D3) biological replicate 4, (E1-E2) biological replicate 5, (F1-F4) biological replicate 6, (G1-G4) biological replicate 7, (H1-H4) biological replicate 8, (I1-I3) biological replicate 9, (J1-J3) biological replicate 10, (K1-K3) biological replicate 11, (L1-L4) biological replicate 12, (M1) biological replicate 13.



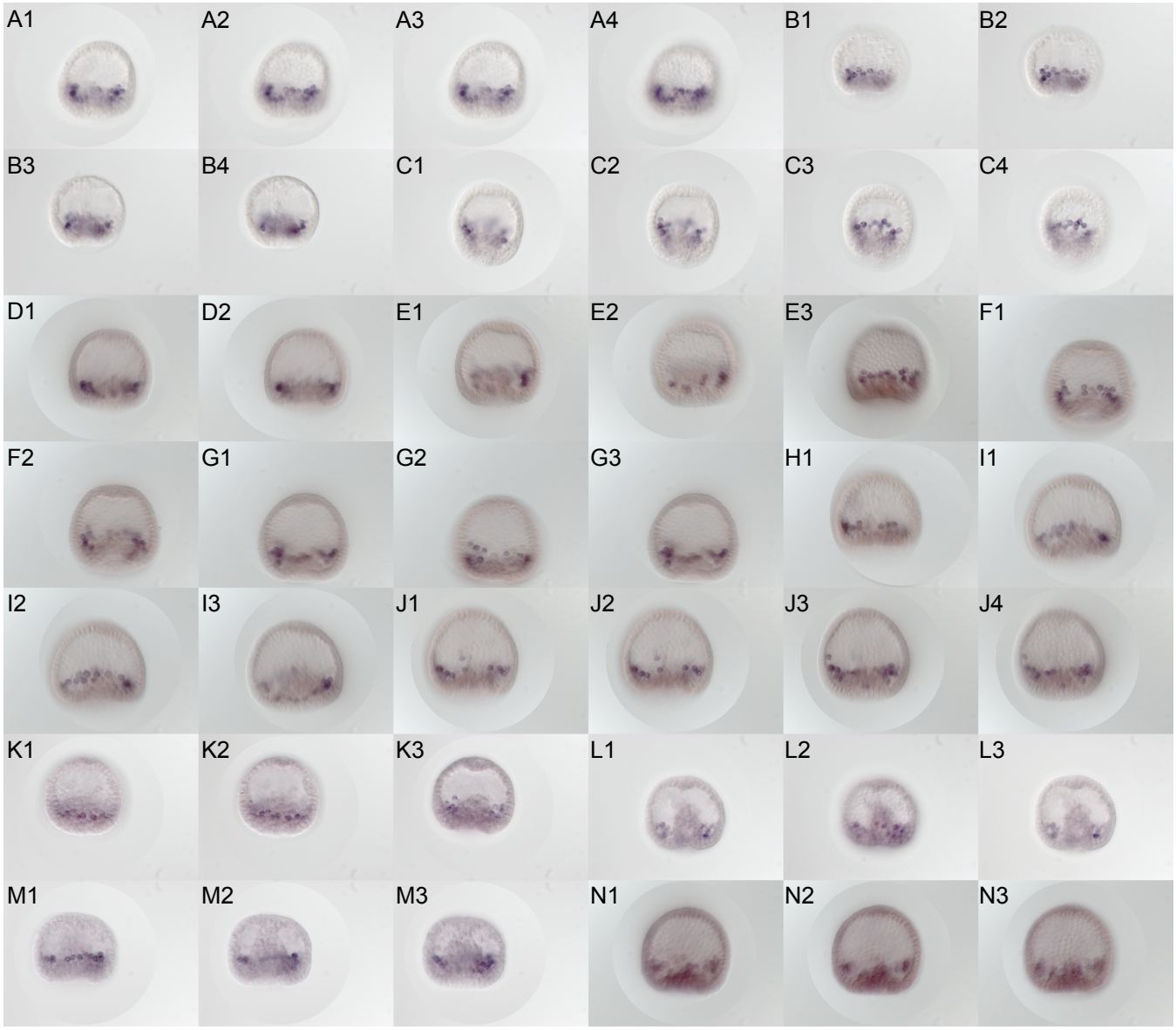
Supplementary Figure 8. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-K) *p16* mRNA. Expression pattern documented for (A1-A3) biological replicate 1, (B1-B2) biological replicate 2, (C1-C4) biological replicate 3, (D1-D2) biological replicate 4, (E1-E4) biological replicate 5, (F1-F2) biological replicate 6, (G1-G2) biological replicate 7, (H1-H2) biological replicate 8, (I1-I4) biological replicate 9, (J1-J4) biological replicate 10, (K1) biological replicate 11. (L-Q) *nadsyn* mRNA. Expression pattern documented for (L1-L2) biological replicate 1, (M1-M3) biological replicate 2, (N1-N2) biological replicate 3, (O1) biological replicate 4, (P1-P2) biological replicate 5, (Q1-Q2) biological replicate 6.



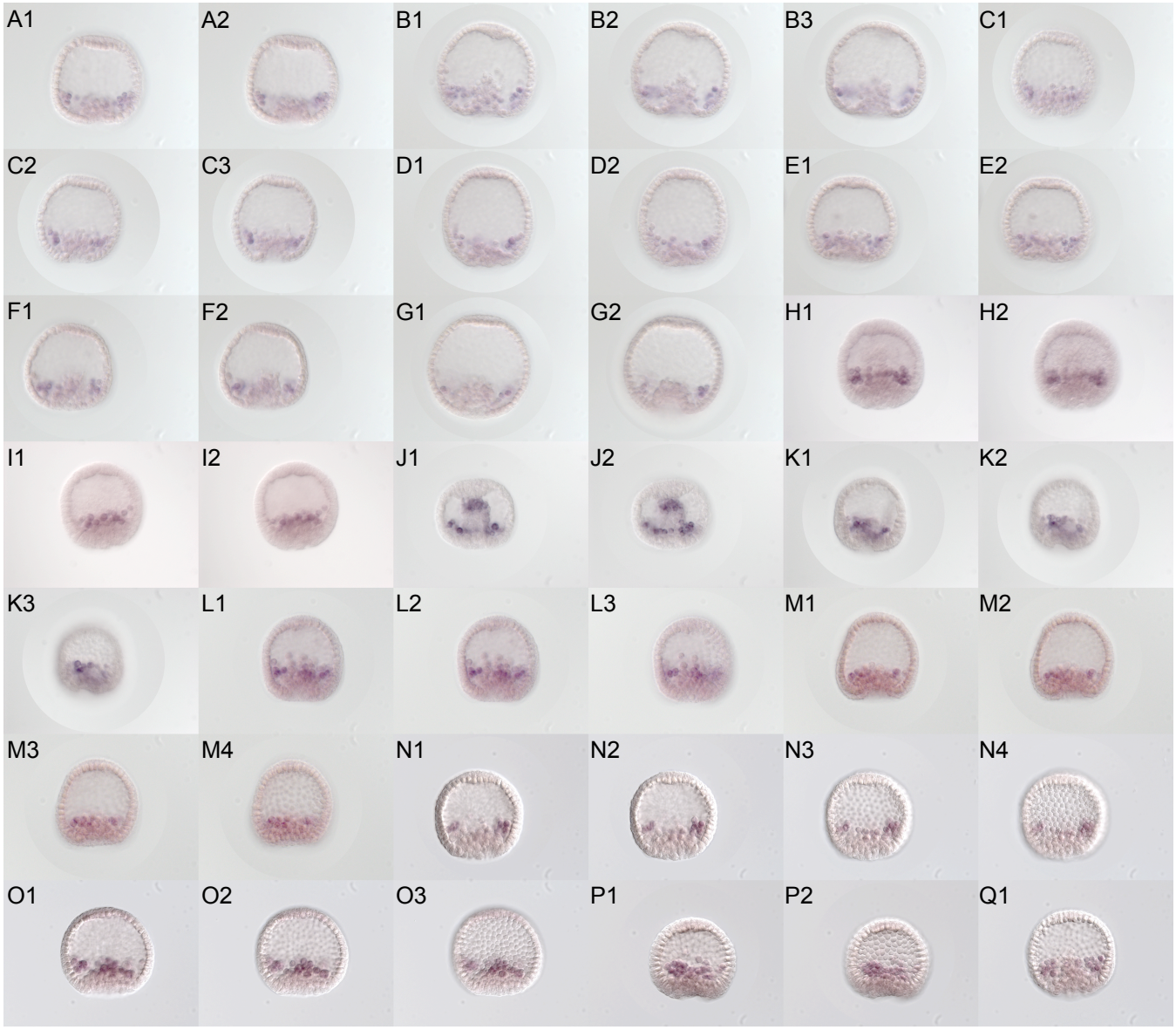
Supplementary Figure 9. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-G) *ets1/2* mRNA. Expression pattern documented for (A1-A2) biological replicate 1, (B1-B3) biological replicate 2, (C1-C2) biological replicate 3, (D1-D3) biological replicate 4, (E1-E3) biological replicate 5, (F1-F2) biological replicate 6, (G1-G3) biological replicate 7. (H-M) *msp130r3* mRNA. Expression pattern documented for (H1-H5) biological replicate 1, (I1-I5) biological replicate 2, (J1-J5) biological replicate 3, (K1-K4) biological replicate 4, (L1-L4) biological replicate 5, (M1) biological replicate 6.



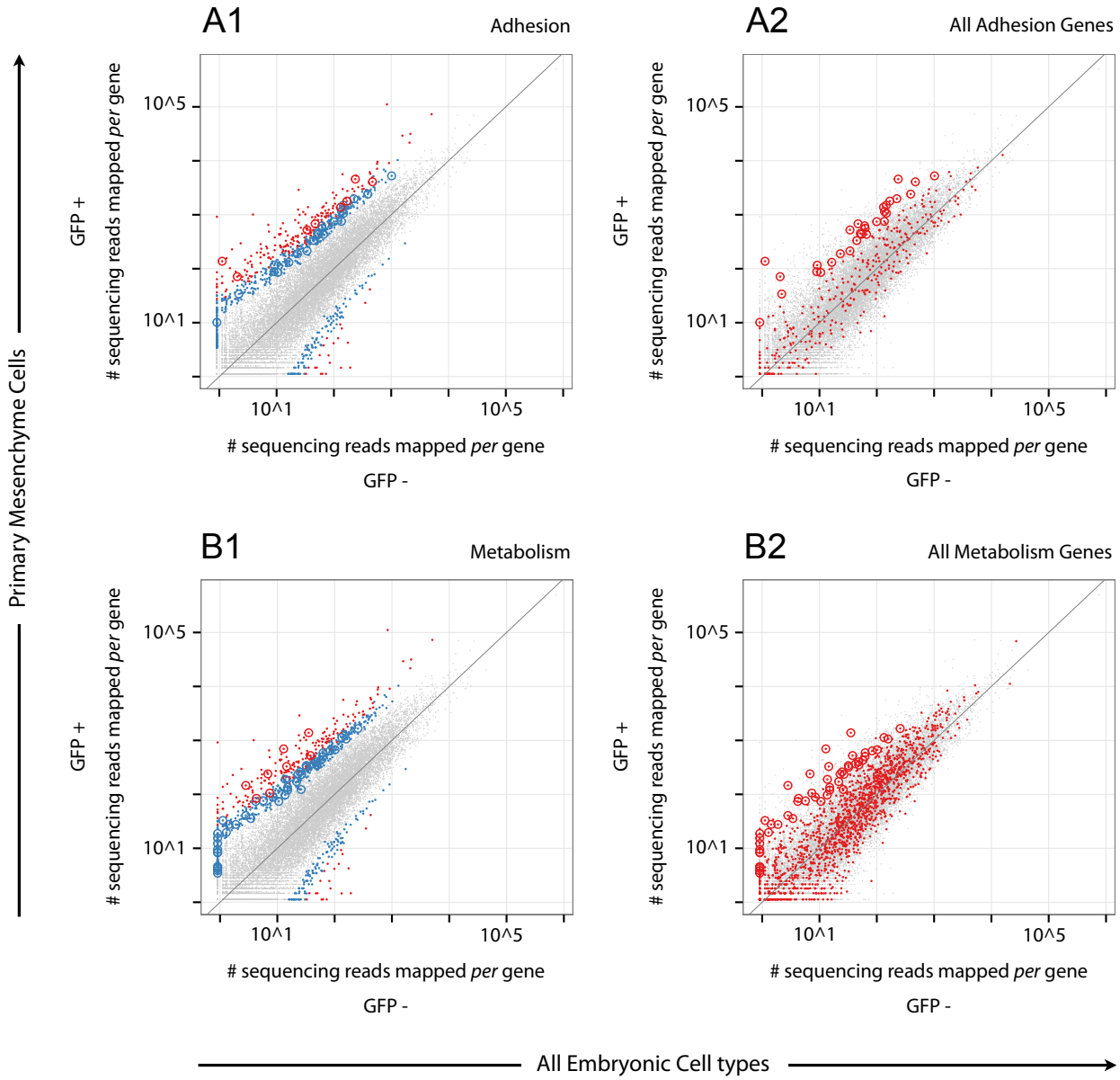
Supplementary Figure 10. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-D) *tbr* mRNA. Expression pattern documented for (A1-A4) biological replicate 1, (B1-B3) biological replicate 2, (C1-C2) biological replicate 3, (D1-D3) biological replicate 4. (E-K) *mitf* mRNA. Expression pattern documented for (E1-E4) biological replicate 1, (F1-F3) biological replicate 2, (G1-G4) biological replicate 3, (H1-H4) biological replicate 4, (I1-I2) biological replicate 5, (J1-J3) biological replicate 6, (K1-K4) biological replicate 7. (L-N) *ppap2b* mRNA. Expression pattern documented for (L1-L2) biological replicate 1, (M1-M2) biological replicate 2, (N1-N2) biological replicate 3.



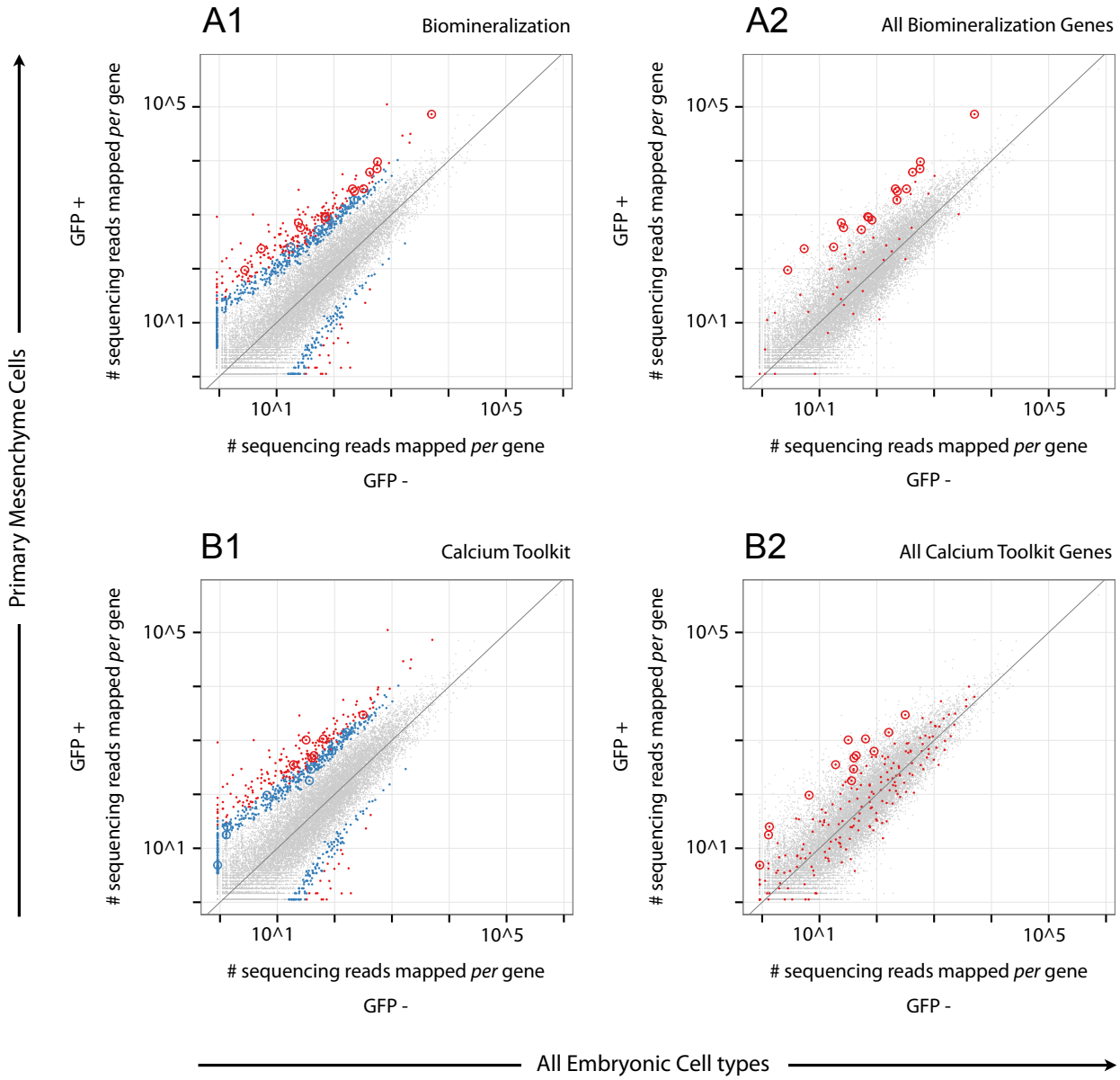
Supplementary Figure 11. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-C) *pdgfr* mRNA. Expression pattern documented for (A1-A4) biological replicate 1, (B1-B4) biological replicate 2, (C1-C4) biological replicate 3. (D-J) *msp130r2* mRNA. Expression pattern documented for (D1-D2) biological replicate 1, (E1-E3) biological replicate 2, (F1-F2) biological replicate 3, (G1-G3) biological replicate 4, (H1) biological replicate 5, (I1-I3) biological replicate 6, (J1-J4) biological replicate 7. (K-N) *orct* mRNA. Expression pattern documented for (K1-K3) biological replicate 1, (L1-L3) biological replicate 2, (M1-M3) biological replicate 3, (N1-N3) biological replicate 4.



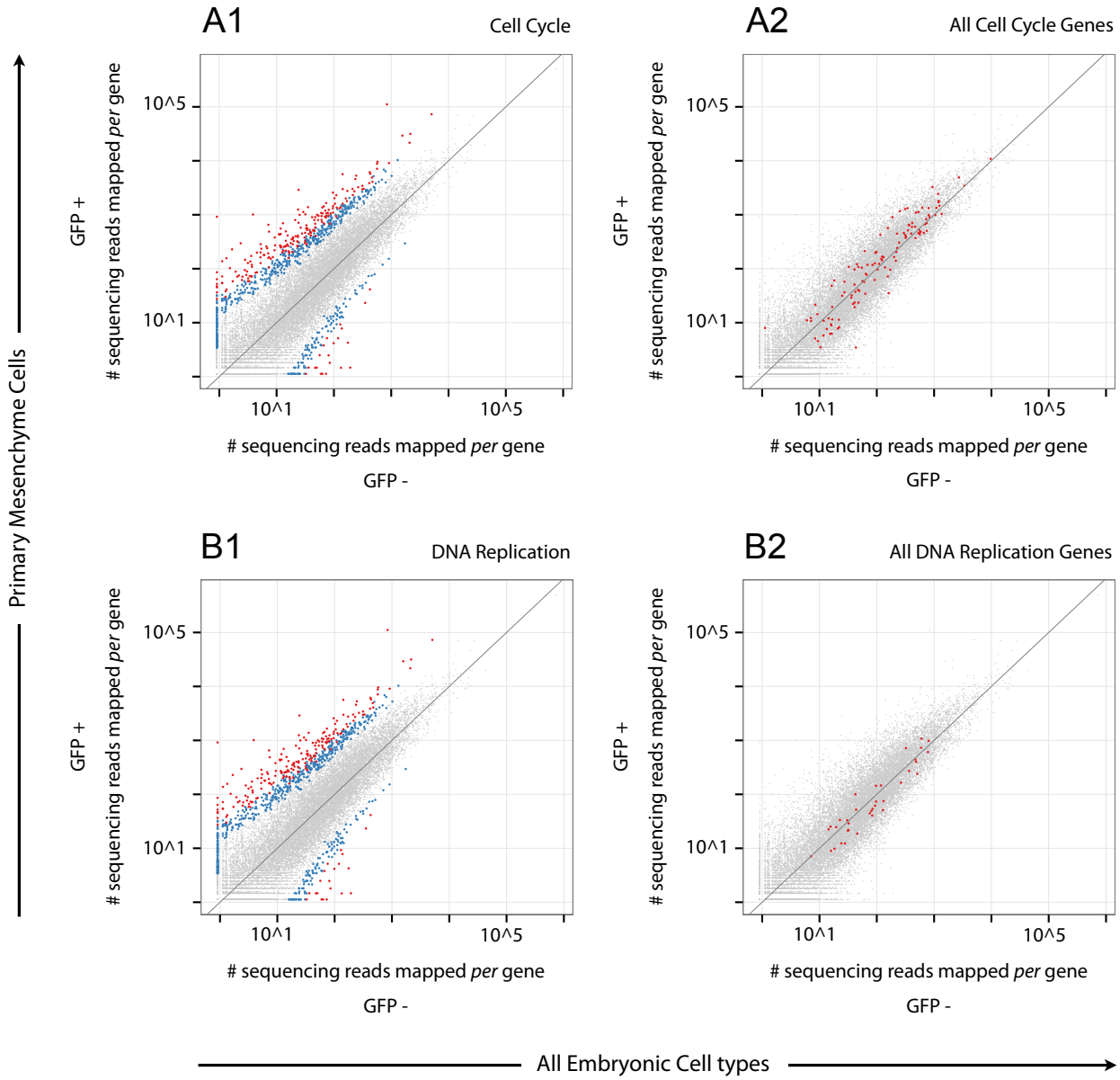
Supplementary Figure 12. **PMC transcriptome sub-set corroborated by RNA in situ hybridization.** Overview as described in the figure legend of Supplementary Figure 5. (A-G) *trpm3* mRNA. Expression pattern documented for (A1-A2) biological replicate 1, (B1-B3) biological replicate 2, (C1-C3) biological replicate 3, (D1-D2) biological replicate 4, (E1-E2) biological replicate 5, (F1-F2) biological replicate 6, (G1-G2) biological replicate 7. (H-I) *c-lectin* mRNA. Expression pattern documented for (H1-H2) biological replicate 1, (I1-I2) biological replicate 2. (J-K) *a2(IV)l* mRNA. Expression pattern documented for (J1-J2) biological replicate 1 taken at a more advanced developmental stage, (K1-K3) biological replicate 2. (L-M) *npnt* mRNA. Expression pattern documented for (L1-L3) biological replicate 1, (M1-M4) biological replicate 2. (N-Q) *fred* mRNA. Expression pattern documented for (N1-N4) biological replicate 1, (O1-O3) biological replicate 2, (P1-P2) biological replicate 3, (Q1) biological replicate 4.



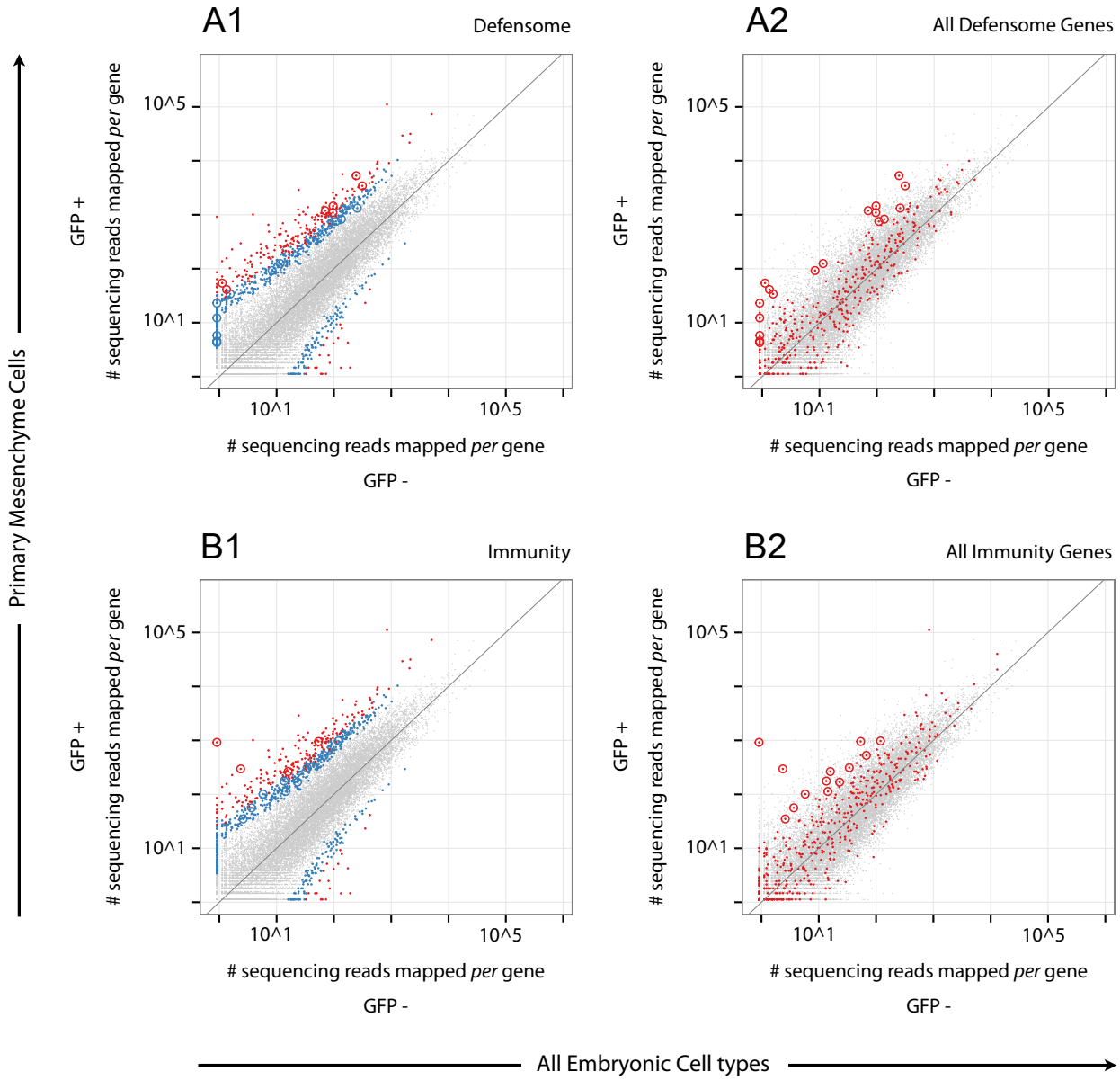
Supplementary Figure 13. **Adhesion and Metabolism genes differentially expressed in PMCs**. (A1-B2) Scatterplots compare transcriptome data unique to PMCs (GFP+) with that of all cell types (GFP-). Each data point reflects the relative mRNA abundance of a gene as estimated by the number of sequencing reads that map to its locus. The values in the graph reflect measurements taken across three independent replicates. Data points that fall above the diagonal indicate augmented levels of expression among PMCs, relative to other cell types. PMC enriched genes associated to the category specified at the top of each scatterplot are shown as encircled data points. (A1, B1) Genes statistically determined to be differentially expressed are shown as data points colored either red or blue to indicate an associated P-value of less than 0.01 or 0.05, respectively. (A2, B2) All genes associated to the category specified at the top of each scatterplot are shown as red colored data points. (A1) Subset of PMC enriched transcripts that encode for adhesion related proteins. (A2) 8% of all adhesion genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for metabolism related proteins. (B2) 3% of all metabolism genes are differentially expressed within PMCs.



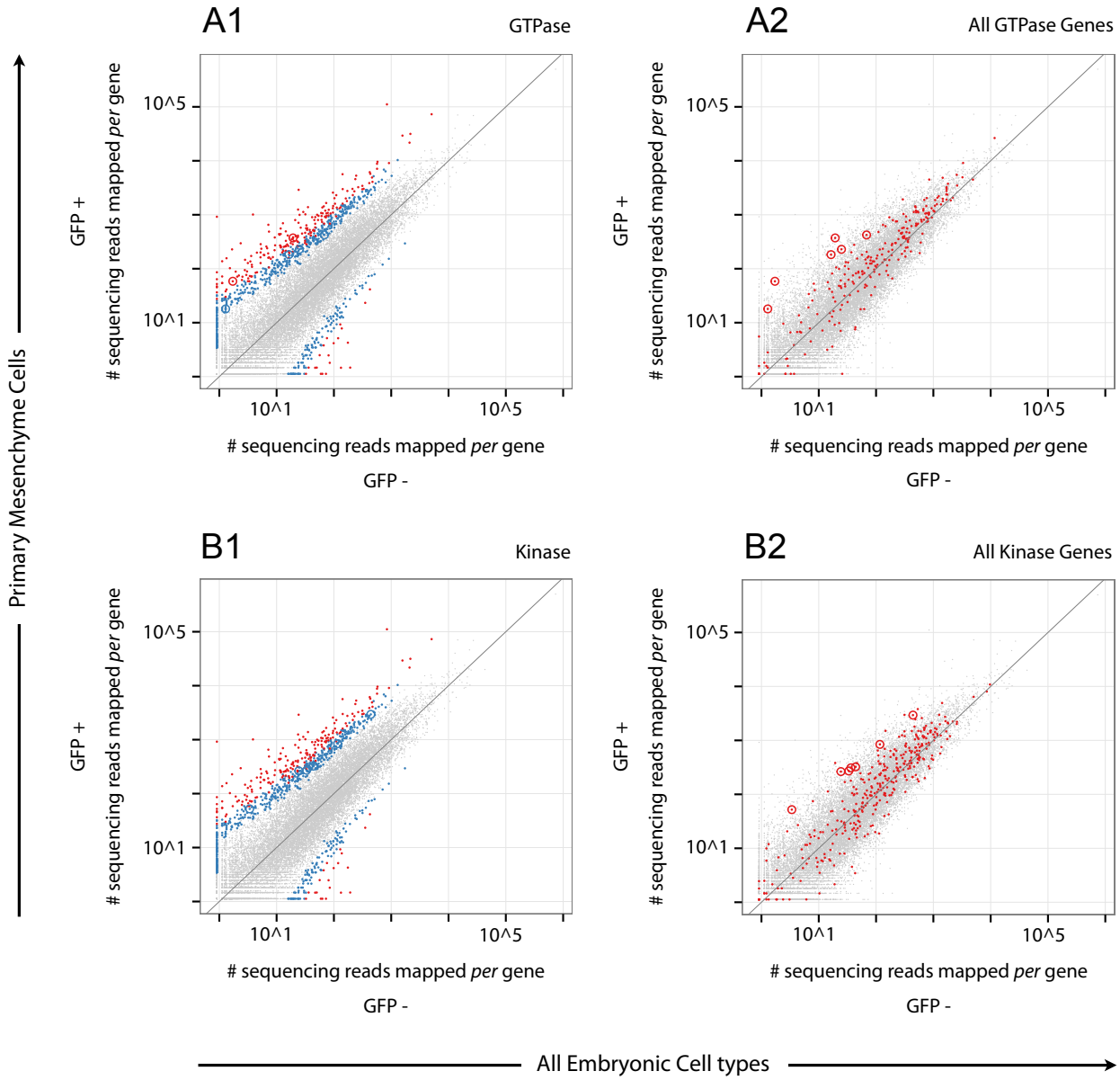
Supplementary Figure 14. **Biomineralization and Calcium Toolkit genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for biomineralization related proteins. (A2) 40% of all biomineralization genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for calcium toolkit related proteins. (B2) 7% of all calcium toolkit genes are differentially expressed within PMCs.



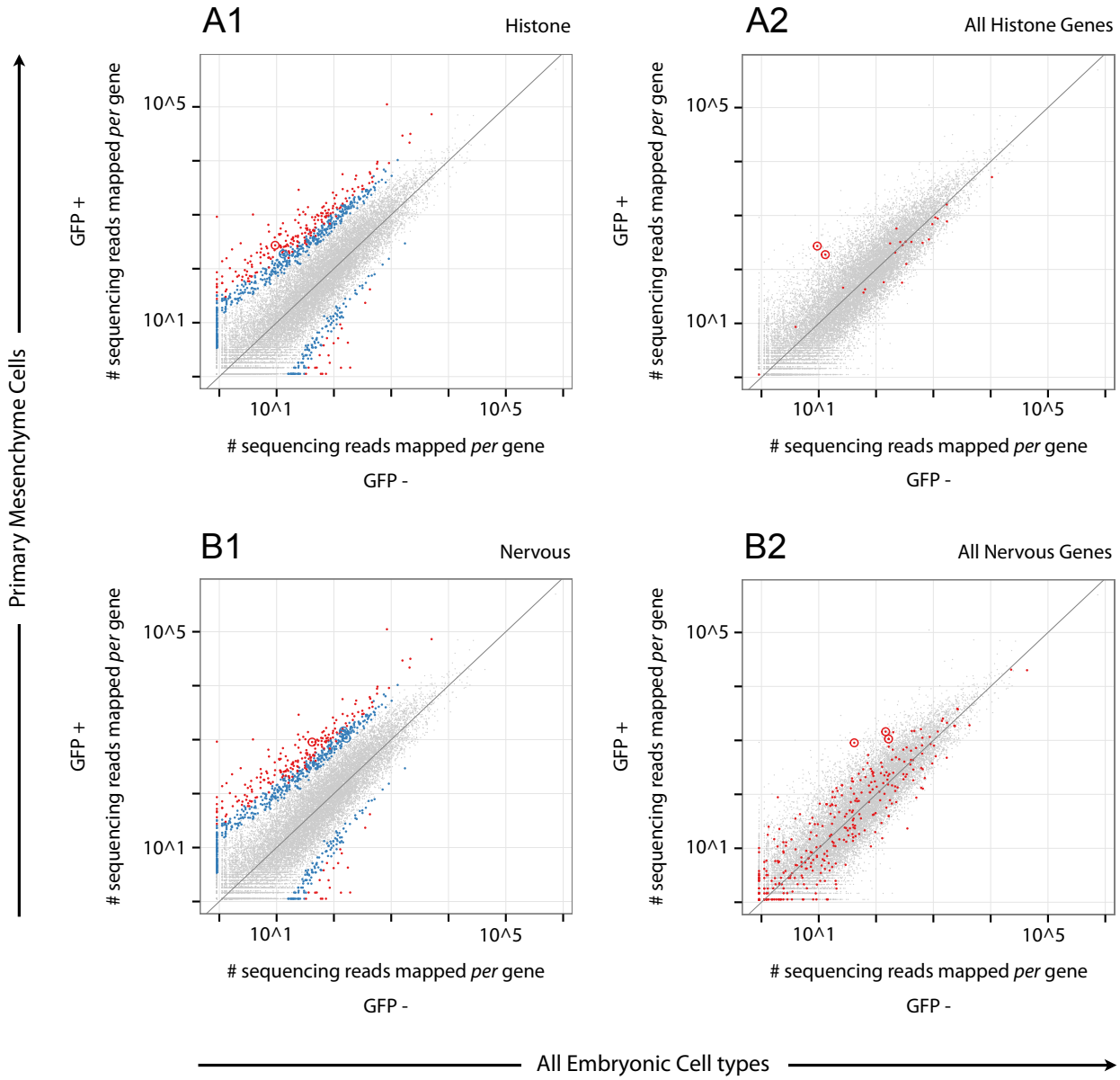
Supplementary Figure 15. **Cell Cycle and DNA Replication genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) No PMC enriched transcripts encode for cell cycle related proteins. (A2) All cell cycle genes. (B1) No PMC enriched transcripts encode for DNA replication related proteins. (B2) All DNA replication genes.



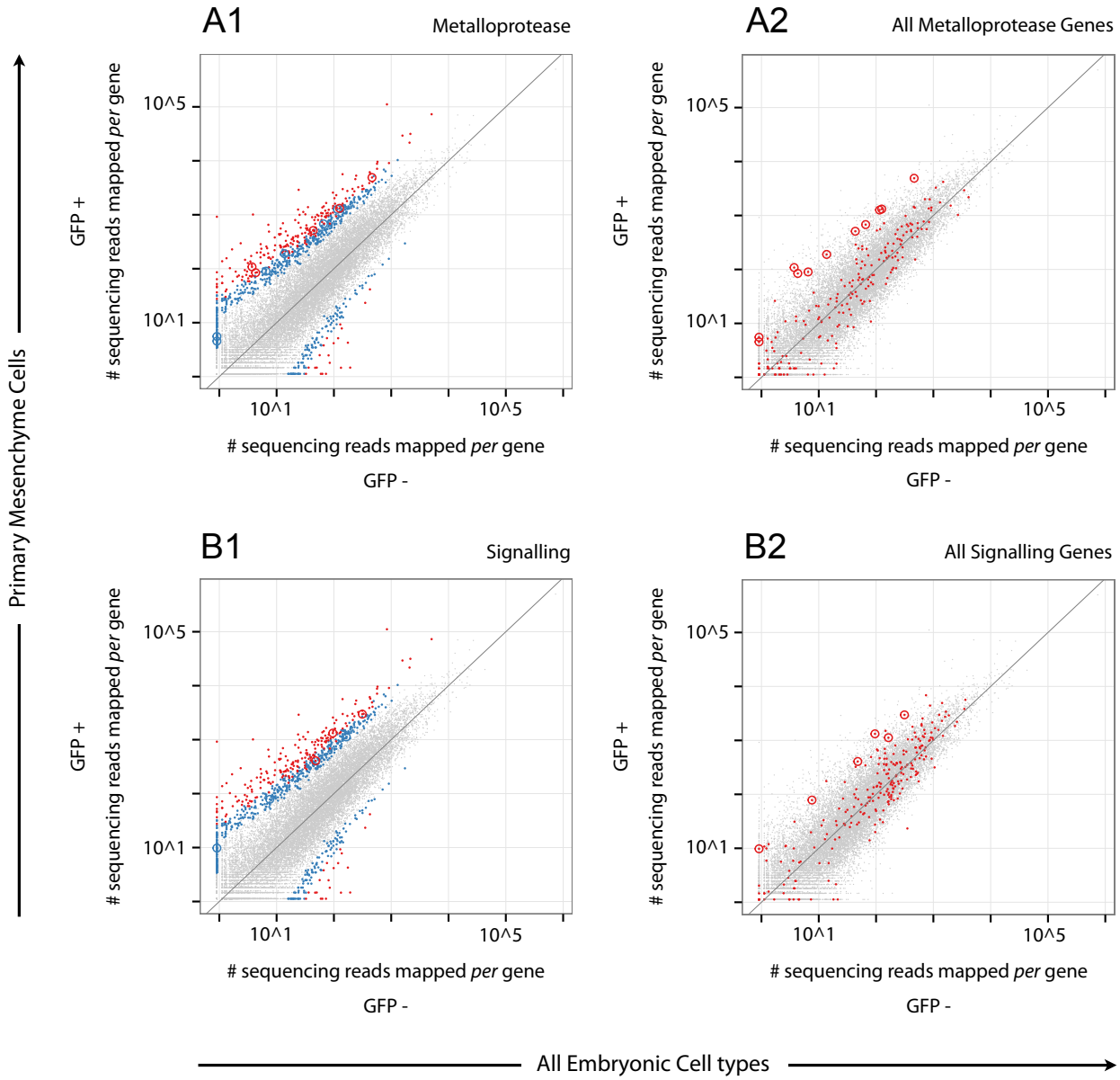
Supplementary Figure 16. **Defensome and Immunity genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for defensome related proteins. (A2) 4% of all defensome genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for immunity related proteins. (B2) 6% of all immunity genes are differentially expressed within PMCs.



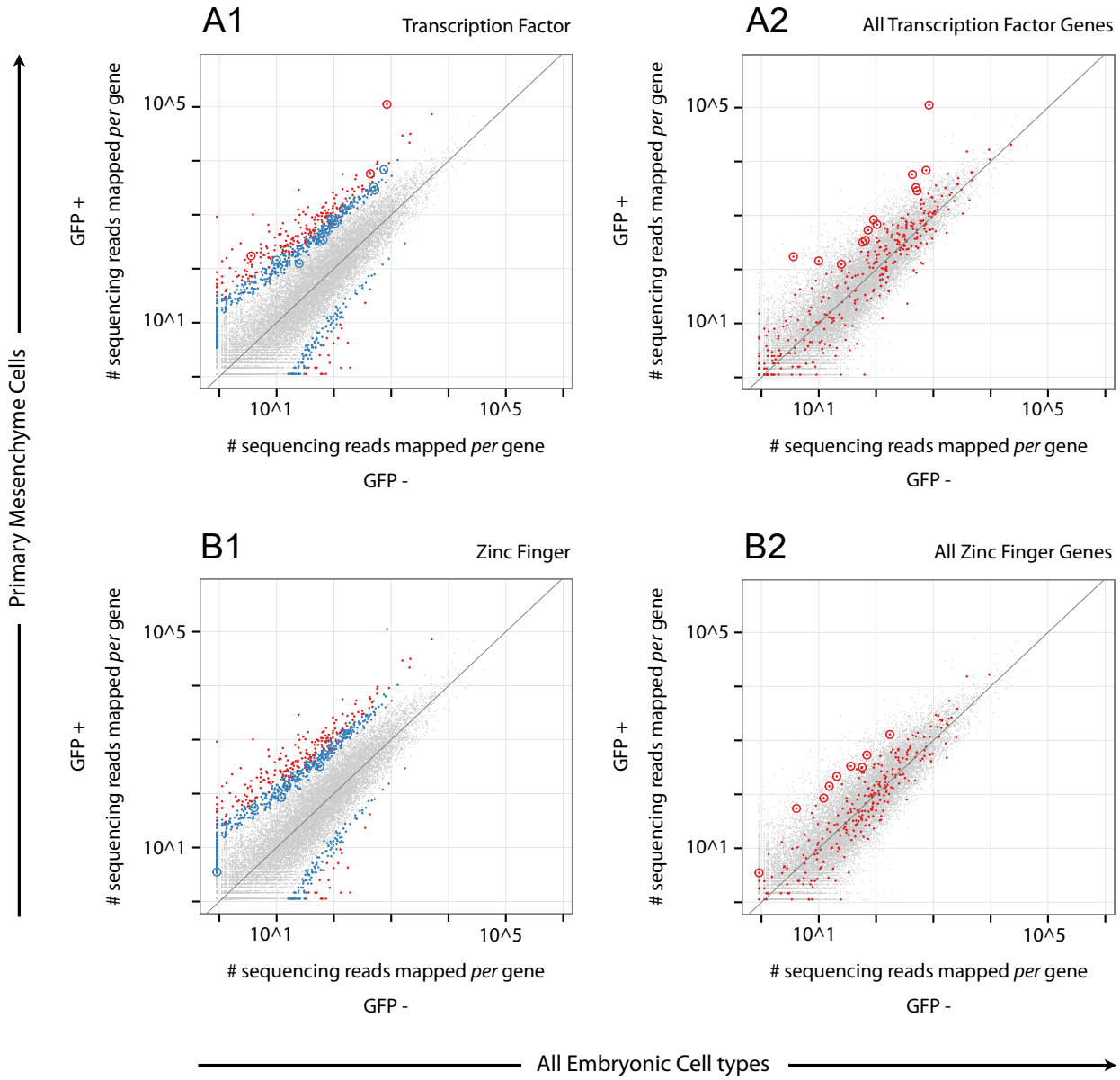
Supplementary Figure 17. **GTPase and Kinase genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for GTPase related proteins. (A2) 3% of all GTPase genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for kinase related proteins. (B2) 5% of all kinase genes are differentially expressed within PMCs.



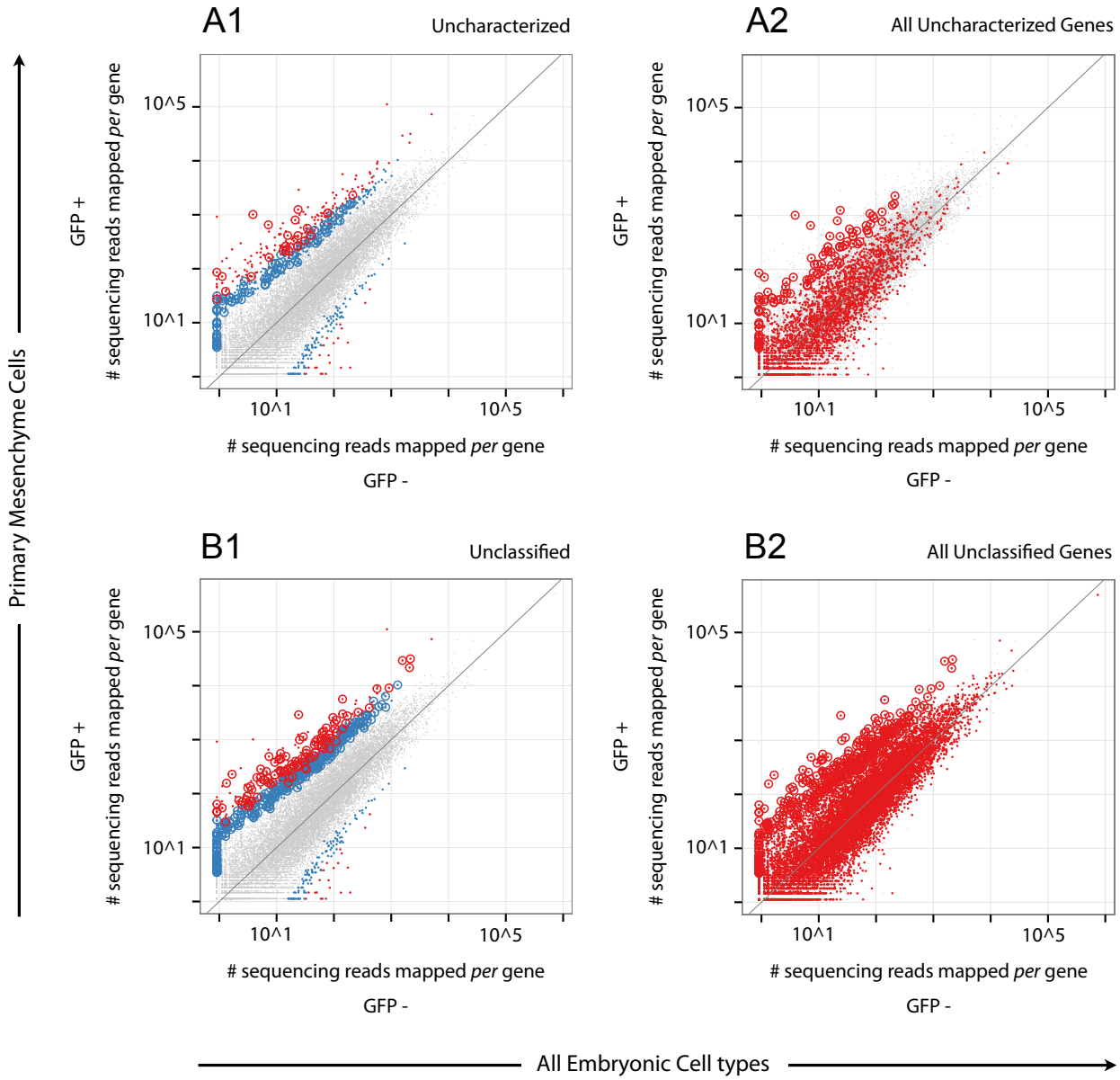
Supplementary Figure 18. **Histone and Nervous genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for histone related proteins. (A2) 8% of all histone genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for nervous related proteins. (B2) 4% of all nervous genes are differentially expressed within PMCs.



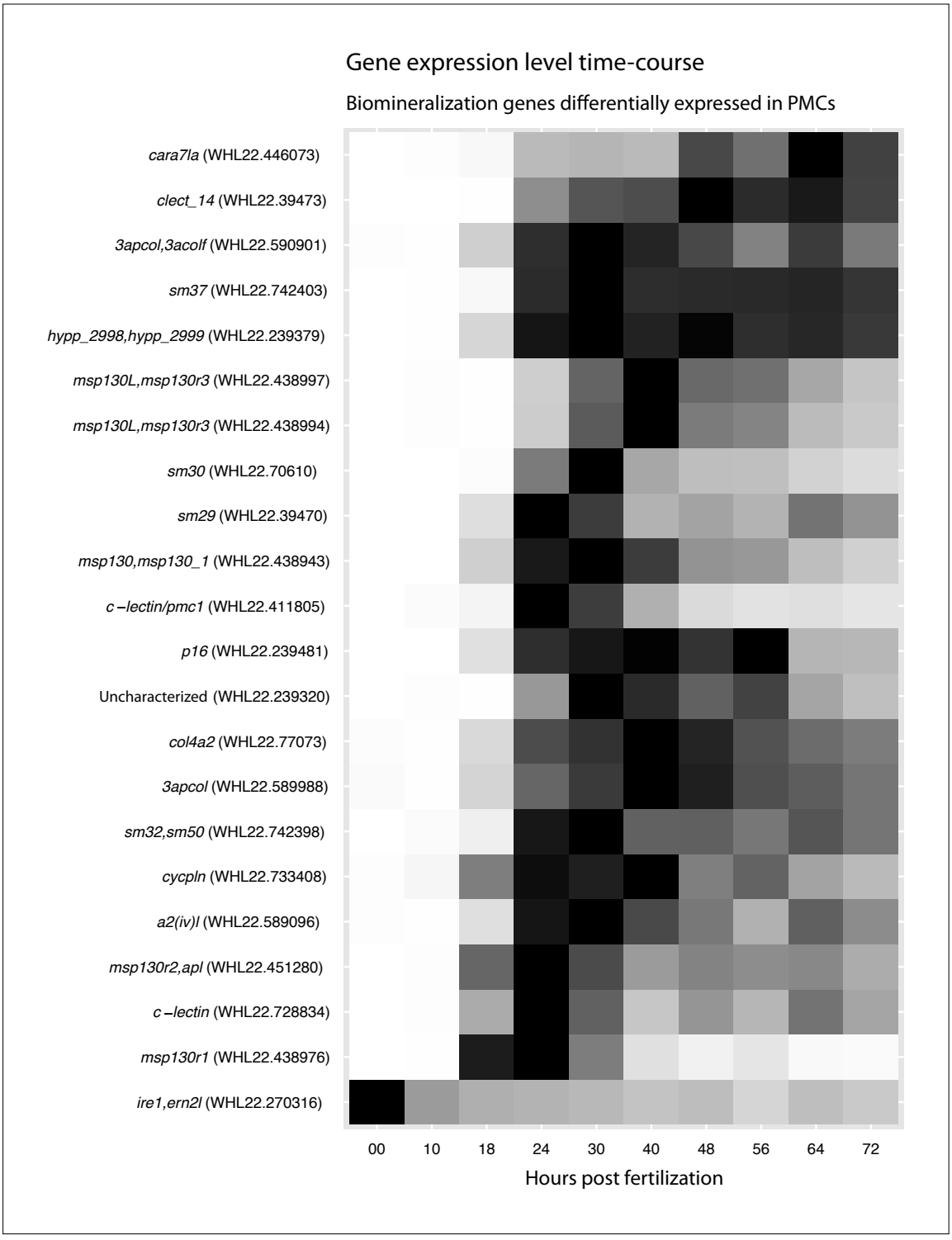
Supplementary Figure 19. **Metalloprotease and Signalling genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for metalloprotease related proteins. (A2) 5% of all metalloprotease genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for signalling related proteins. (B2) 5% of all signalling genes are differentially expressed within PMCs.



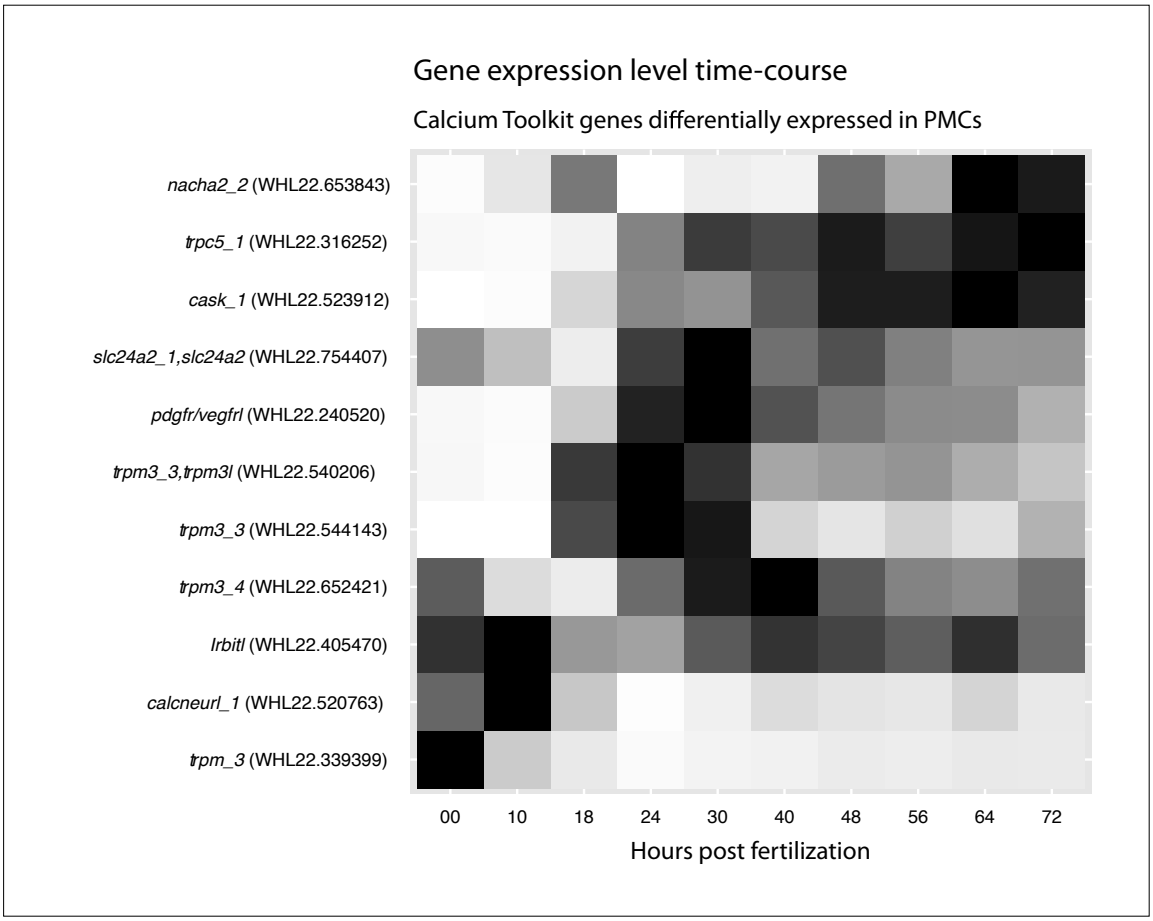
Supplementary Figure 20. **Transcription Factor and Zinc Finger genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for transcription factor related proteins. (A2) 5% of all transcription factor genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for zinc finger related proteins. (B2) 3% of all zinc finger genes are differentially expressed within PMCs.



Supplementary Figure 21. **Uncharacterized and Unclassified genes differentially expressed in PMCs**. Overview as described in the figure legend of Supplementary Figure 13. (A1) Subset of PMC enriched transcripts that encode for uncharacterized proteins. (A2) 3% of all uncharacterized genes are differentially expressed within PMCs. (B1) Subset of PMC enriched transcripts that encode for unclassified proteins. (B2) 3% of all unclassified genes are differentially expressed within PMCs.



Supplementary Figure 22. **Biomineralization gene expression profile dynamics**. Gene expression level time-course data for biomineralization genes differentially expressed in PMCs. Measurements have been transformed to color according to a grey-scale gradient and are presented in the form of a heat map. The degree, to which each element in the matrix has been shaded, reflects percentage of expression relative to each gene's peak value; whereby a solid black square denotes maximum expression among the time points assessed. Of these, the 24 hpf time point best approximates the developmental stage for which cell type specific transcriptome data was attained. Low abundance genes have been excluded from the analysis. For additional information, see Supplementary Note 4.



Supplementary Figure 23. **Calcium Toolkit gene expression profile dynamics**. Overview as described in the figure legend of Supplementary Figure 22. For additional information, see Supplementary Note 4.

Supplementary Note 1. **PMC gene regulatory network**

The consecutive partition of maternal anisotropies that occurs as a consequence of unequal cleavage events, leads to *pmar1* expression exclusively within the PMC founder cells. *pmar1* encodes a paired-class regulatory factor that functions as an obligatory transcriptional repressor. From among the genes that are directly repressed by *pmar1*, *hesC* is the most relevant with respect to specification, for it similarly encodes a transcriptional repressor that is otherwise ubiquitously expressed. Consequently, *hesc* prevents the expression of its target genes in every cell of the embryo, except in the PMC founder cells where its own transcription is prohibited. Localized relief of *hesc* repression results in the transcription of its immediate downstream targets within the PMC founder cells. These in turn, encode for transcription factors driving PMC identity, e.g. *alx1* (colored blue), *ets1/2* and *tbr* (colored red). A feed-forward sub-circuit pivots around *ets1/2* and serves to reinforce the expression of these regulatory genes. Ultimately, specification proceeds autonomously as the process is emancipated from initial inputs. In part, due to the auto-regulatory feedback characteristics of *alx1* and later, thanks to the multiple cross-regulatory interactions between the next tier of transcription factors, e.g. *erg*, *hex*, *tgif*. The wiring described above, belongs to a class of circuitry not uncommon in developmental GRNs, the function of which is to install a regulatory state in one domain (here the PMC founder cells) while prohibiting this same state elsewhere. All genes depicted are regulatory in nature, the last two being the exception (*sm50*, *cyP*); these represent members of the differentiation gene battery responsible for skeletogenesis.

Supplementary Note 2. **High throughput recovery of labeled cells**

A total of thirty-six thousand zygotes were microinjected with an artificial chromosome genetically engineered to express GFP in PMCs. Twelve thousand embryos were assessed *per* replicate. The *cis*-regulatory control apparatus that governs *alx1* transcription was utilized to drive GFP expression in the first replicate (Figure 3 A1-3), whereas the other two replicates employed an analogous reporter driven instead, by the *cis*-regulatory control apparatus that governs *tbr* expression (Figure 3 B1-C3). The fraction of cells acquired as a result of embryo disaggregation varied across replicates with respect to the number of total events detected by flow cytometry (Figures 3 A1, B1, C1). From among the population of cells selected for further analysis, cell viability never dropped below 92% for any given replicate (Figures 3 A2, B2, C2). FACS was then used to recover a subset of PMCs, on account of their GFP expression. 10400 PMCs were recovered in the first replicate, 4156 in the second and 2959 in the third (uppermost right quadrant of Figures 3 A3, B3, C3, respectively). For each of the replicates, 50000 GFP negative cells were collected alongside PMCs in order to serve as an internal control (lowermost right quadrant of Figures 3 A3, B3, C3). In transgenic *Sp* embryos, not all cells incorporate the exogenous DNA delivered to the zygote. By the end of the blastula stage there are 32 PMCs *per Sp* embryo of which typically 2 express GFP when low concentrations are used (Figure 2 compare B with C1-3). This suggests that prior to the formation of a syncytium, approximately 6% of PMCs become labeled. Given that each replicate on average, recovered 5838 PMCs, this would imply a 25% recovery of labeled cells.

Supplementary Note 3. **Skeletogenic cell-specific transcriptome**

A cell type specific transcriptome has been attained for PMCs of sea urchin blastula. This resource is available online in its entirety, through an interactive portal at (http://www.spbase.org:3838/cellspecific/). The values publicized by this database reflect gene expression levels averaged across all replicates, in terms of sequencing reads mapped *per* gene. For each gene, expression levels from the cell type of interest are compared to that of the embryo. Relative enrichment is conveyed in terms of gene expression fold change; whereby positive fold change reflects a higher concentration of transcripts among PMCs, relative to all embryonic cell types. Conversely, a negative fold change value reflects mRNA depletion. Ultimately, in an effort to catalog genes of functional import for PMC biology, we suggest focusing exclusively on genes that are enriched in a statistically significant manner (P value ≤ 0.05). To achieve this and more, the online portal boasts a suite of tools that facilitate data mining. For instance, genes may be categorized according to their biological function and sorted by cell-specific enrichment, statistical significance, or a number of other parameters. Selected gene sets automatically are mapped onto a comprehensive scatterplot, such that data points of particular interest may be visualized, relative to the entire dataset.

Supplementary Note 4. **Gene expression profile dynamics**

Biomineralization and Calcium Toolkit gene cohorts are particularly relevant to primary mesenchyme cell-biology and thus, were monitored throughout embryonic development. Specifically, the level of expression for each of the selected genes was assessed across ten time-points, covering 72 hours post fertilization. Supplementary Figures 22-23 capture expression profile dynamics for PMC enriched genes involved in the terminal differentiation that culminates with the formation of skeletal rods.

By measuring gene activity throughout developmental time, we acquire an additional parameter with which to estimate functional relevance. By virtue of the fact that the genes analyzed are coexpreesed within the same cell, specific aspects of gene regulation may also be inferred from these heat maps. For instance, synchronized activation of terminal differentiation genes as visualized in Supplementary Figure 22, strongly suggests that they are governed by the same combination of transcription factors. In turn, this likely reflects a similarity in the *cis*-regulatory apparatus of each gene depicted.

Supplementary Note 5. **Embryo disaggregation**

Hyalin Extraction Medium (1 liter)

17.5 grams NaCl

0.75 grams KCl

1.25 grams MgSO4 or 2.5g MgSO4.7H2O

22.5 grams Glycine

1.21 grams Tris

0.76 grams EGTA

pH to 8.2, salinity should be in the range of 34-36 parts *per* thousand

Calcium-Free Sea Water (1 liter)

26.5 grams NaCl

0.70 grams KCl

5.81 grams MgSO4 or 11.9 grams MgSO4.7H2O

0.50 grams NaHCO3

pH to 8.2, salinity should be in the range of 34-36 parts *per* thousand