Cell Reports, Volume *19*

Supplemental Information

Biotagging of Specific Cell Populations in Zebrafish Reveals Gene Regulatory Logic Encoded in the Nuclear Transcriptome

Le A. Trinh, Vanessa Chong-Morrison, Daria Gavriouchkina, Tatiana Hochgreb-Hägele, Upeka Senanayake, Scott E. Fraser, and Tatjana Sauka-Spengler

SUPPLEMENTAL INFORMATION

Supplemental Data

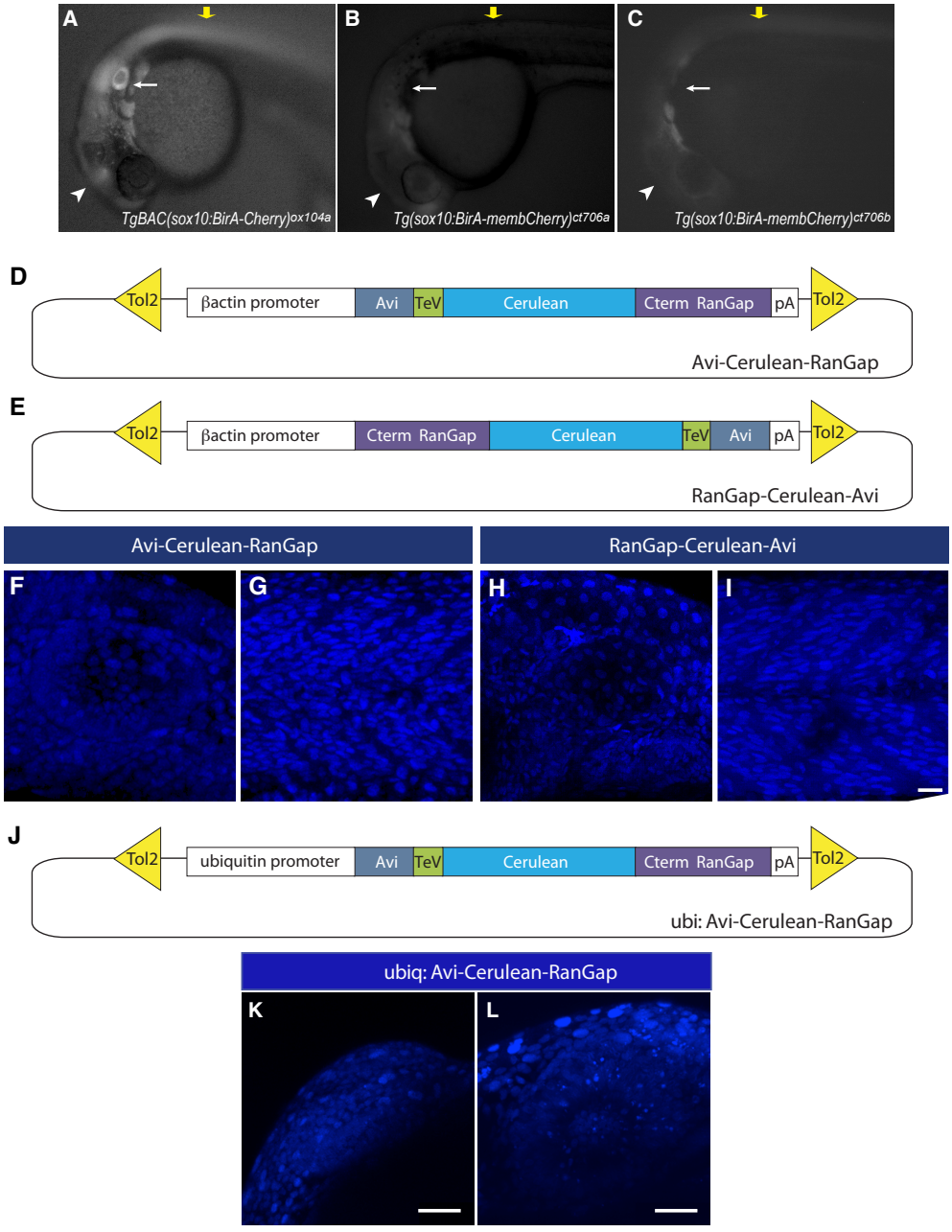


Figure S1. Biotagging *sox10* BirA drivers and Avi-tagged RanGap effectors. Related to Figures 1 and 2.

A-C Differential expression between *sox10* biotagging transgenic and BAC drivers. Wide-field image of *TgBAC(Sox10:BirA-mCherry)*^{ox104a} (ncBirA(BAC)) (A). *Tg(Sox10:BirA-membCherry)*^{ct706a} (ncBirA) and *Tg(Sox10:BirA-membCherry)*^{ct706b} (ncBirA(b)) transgenes at 24hpf (B, C). The two alleles (B, C) of *sox10* Biotagging transgenic exhibit expression in different neural crest derivatives that are included in the overall *sox10* expression pattern, by *sox10* Biotagging BAC (A). Arrow points to lack of expression in the otic vesicle, while arrowheads point to lack of expression in the midbrain of *sox10* transgenes.
D-L Biotagging Avi-tagged nuclear localized effectors. Schematic of two variant Avi-tagged RanGap constructs for generating Avi effector transgenes. N-terminal Avi-tag construct (Avi-Cerulean-Rangap) contains the beta-actin2 (*βactin*) promoter upstream of Avi-tag (steelblue), the Tobacco Etch Virus protease cleave site (TeV, green), Cerulean (turquoise), the C-terminal domain of RanGap (purple), and a polyA signal (D). C-terminal Avi-tag construct (RanGap-Cerulean-Avi) contains the beta-actin2 (*βactin*) promoter upstream of the C-terminal domain of RanGap (purple), Cerulean (turquoise), the Tobacco Etch Virus protease cleave site (TeV, green), Avi-tag (steelblue), and a polyA signal (E). Both constructs are flanked by tol2 elements (yellow) for transgenesis by Tol2 transposition. 3-D projection of confocal Z-stack of Avi-RanGap (F-G) and RanGap-Avi (H-I) of the developing inner ear (F, H) and somite (G, I), imaged at 32hpf. Both Avi-Cerulean-RanGap and Rangap-Cerulean-Avi proteins localize similarly to the nucleus of all cells in the embryo. Schematic of Avi-RanGap effector construct with ubiquitin promoter (*ubiq*) upstream of N-terminal Avi tagged RanGap elements (J). Confocal image of hindbrain (K) and eye (L) of *Tg(ubiq:Avi-RanGap)* embryo (nucAvi(*ubiq*)). Scale bars: $20\,\mu m$, except $50\,\mu m$ in (K, L).

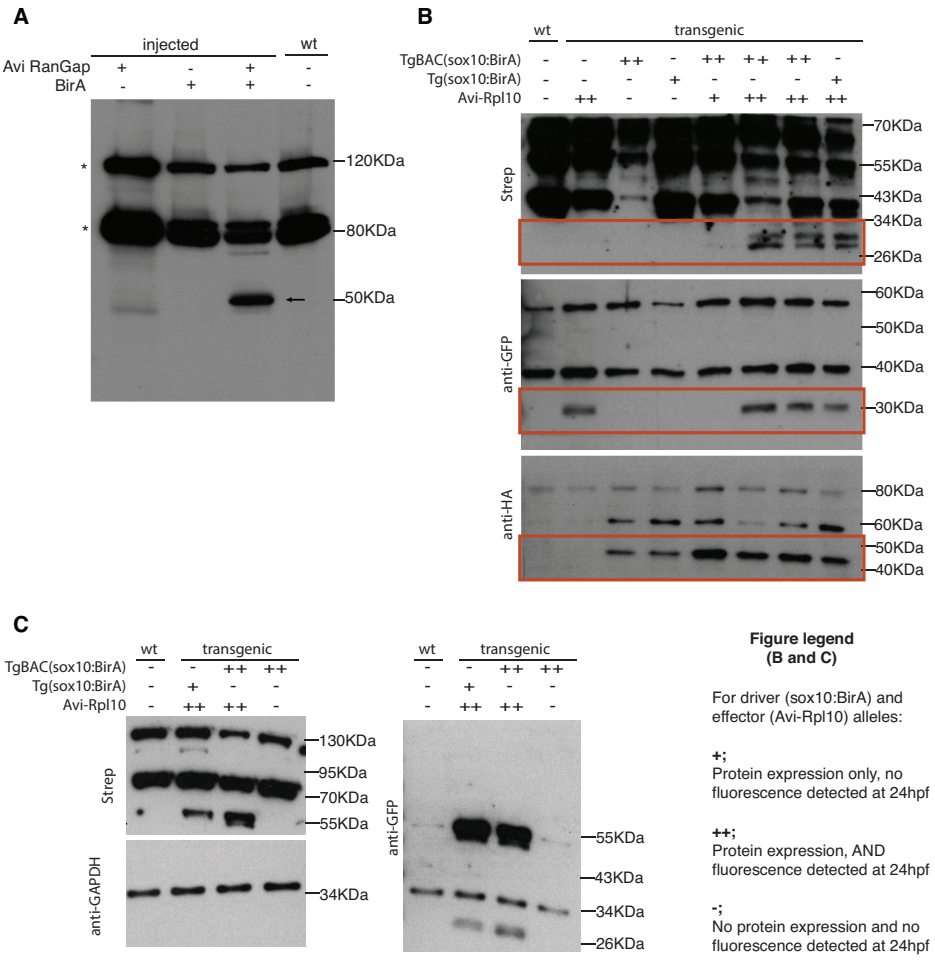
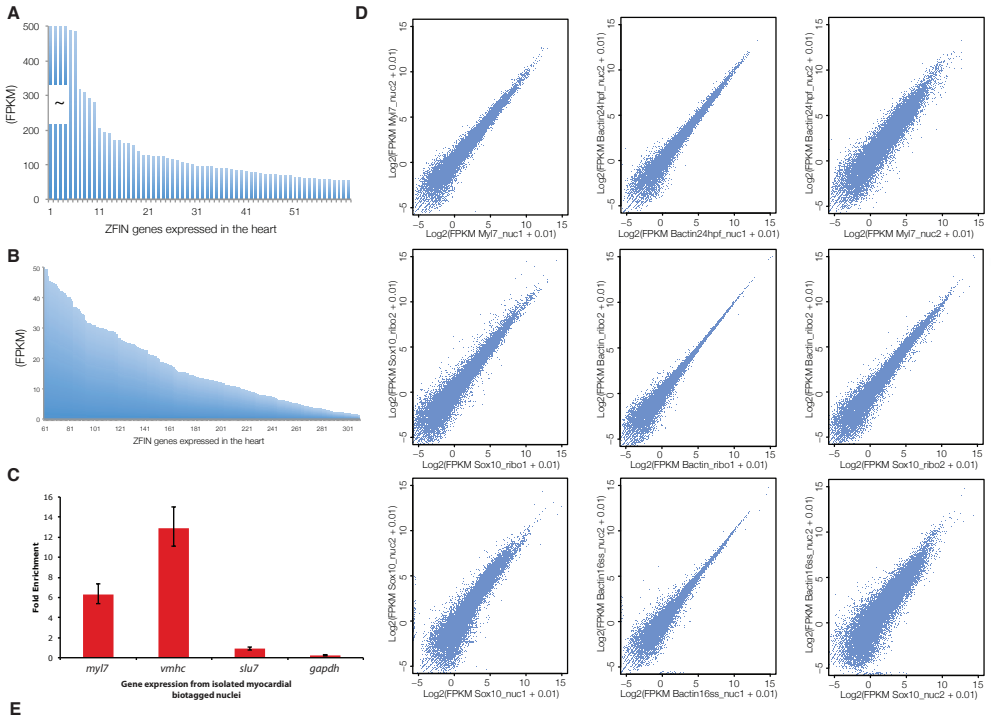


Figure S2. Biotinylation of Avi-tagged protein is dependent on the level of expression. Related to Figure 3.

A Streptavidin Western blot of nuclear extracts from embryos injected with Avi-RanGap (lane 1), BirA (lane 2), both mRNA (lane 3) and wildtype un-injected embryos (lane 4). Arrow points to biotinylated Avi-RanGap that is specifically labelled with Streptavidin conjugate in lysates from embryos co-injected with Avi-RanGap and BirA (lane 3). Asterisks indicate endogenously biotinylated proteins also present in uninjected wildtype embryos. BirA expression (lane 2) does not elevate endogenous biotinylation over background level (lane 4). Avi-tag is insensitive to biotinylation by endogenous biotin ligases (line 1 compared to line 3). **B** Streptavidin (upper panel), Anti-GFP (middle panel) and Anti-HA (bottom panel) Western blot of whole cell extract from wildtype (lane 1); *TgBAC(sox10:BirA)* (ncBirA(BAC)) (lane 3); *Tg(sox10:BirA)* (ncBirA) (lane 4); *Tg(ubiq:AviRpl10)* (riboAvi(*ubiq*)) (lane 2); or double transgenic of Avi-tag and BirA (lanes 5, 6, 7, 8) embryos. Number of plus marks indicate level of expression described to the right of blots. Avi-Rpl10 is biotinylated only when highly expressed (lanes 6, 7, 8). Anti-GFP western blot detects a cleaved Avi-Rpl10 that is 30 kDa (red box in upper and middle panels). Anti-HA Western blot detects BirA that contains HA-tag (red box in bottom panel). **C** Streptavidin (upper left panel), Anti-GAPDH (lower left panel) and Anti-GFP (right panel) Western blot of whole cell extract from wildtype (lane 1); *TgBAC(sox10:BirA)* (lane 4); or double transgenic of Avi-Rpl10 and BirA (lanes 2 and 3) embryos in which BirA expression varies while "Avi-Rp10" expression is high. Anti-GFP Western blot detects both cleaved and uncleaved Avi-Rpl10 that is 30 kDa and 55 kDa, respectively.



E

Spearman's rank corr. coeff.	Bactin_16ss_nuc1	Bactin_16ss_nuc2	Bactin_16ss_ribo1	Bactin_16ss_ribo2	Bactin_24hpf_nuc1	Bactin_24hpf_nuc2	Myl7_nuc1	Myl7_nuc2	Sox10_nuc1	Sox10_nuc2	Sox10_ribo1	Sox10_ribo2
Bactin_16ss_nuc1	1.00000000	**0.96147120**	0.93190840	0.93322510	0.80551520	0.80939370	0.79682440	0.78933820	0.89679400	0.91170390	0.93297450	0.92500120
Bactin_16ss_nuc2	**0.96147120**	1.00000000	0.93032030	0.93049930	0.80736190	0.81140020	0.80014890	0.79242450	0.89602350	0.91039530	0.92886270	0.92418280
Bactin_16ss_ribo1	0.93190840	0.93032030	1.00000000	**0.97119760**	0.86360040	0.86788630	0.86045320	0.85084510	0.89221940	0.91360360	0.96306270	0.96682940
Bactin_16ss_ribo2	0.93322510	0.93049930	**0.97119760**	1.00000000	0.86375590	0.86786370	0.85991120	0.85067090	0.89257980	0.91366500	0.96297660	0.96695920
Bactin_24hpf_nuc1	0.80551520	0.80736190	0.86360040	0.86375590	1.00000000	**0.99346360**	0.97886880	0.97458280	0.78656150	0.80408370	0.85803000	0.86759210
Bactin_24hpf_nuc2	0.80939370	0.81140020	0.86788630	0.86786370	**0.99346360**	1.00000000	0.98355480	0.97969530	0.79191630	0.81029110	0.86078510	0.87168880
Myl7_nuc1	0.79682440	0.80014890	0.86045320	0.85991120	0.97886880	0.98355480	1.00000000	**0.99144250**	0.79052730	0.81392520	0.84911590	0.86340010
Myl7_nuc2	0.78933820	0.79242450	0.85084510	0.85067090	0.97458280	0.97969530	**0.99144250**	1.00000000	0.78336630	0.80933780	0.83950670	0.85311390
Sox10_nuc1	0.89679400	0.89602350	0.89221940	0.89257980	0.78656150	0.79191630	0.79052730	0.78336630	1.00000000	**0.93860340**	0.88907410	0.89310070
Sox10_nuc2	0.91170390	0.91039530	0.91360360	0.91366500	0.80408370	0.81029110	0.81392520	0.80933780	**0.93860340**	1.00000000	0.91182660	0.91200810
Sox10_ribo1	0.93297450	0.92886270	0.96306270	0.96297660	0.85803000	0.86078510	0.84911590	0.83950670	0.88907410	0.91182660	1.00000000	**0.96434400**
Sox10_ribo2	0.92500120	0.92418280	0.96682940	0.96695920	0.86759210	0.87168880	0.86340010	0.85311390	0.89310070	0.91200810	**0.96434400**	1.00000000

Figure S3. Enrichment of cell-type specific gene expression in developing cardiomyocytes and pairwise comparison of biological duplicates. Related to Figure 4.

A, B Barplot of FPKM values from *myl7* nuclear datasets for genes annotated in ZFIN expression database as detected in myocardium from 0-24hpf. **C** Quantification of relative enrichment of myocardial transcripts as determined by RT-qPCR using cDNA isolated from nuclei purified from embryos double transgenic for *Tg(myl7:BirA-membCherry)*^{ct704a}; *Tg(bactin:Avi-Cerulean-RanGap)*^{ct700a} (myoBirA;nucAvi(*bact*)) compared to cDNA extracted from whole embryos. Relative levels of transcripts for myosin light polypeptide 7 (*myl7*), ventricular myosin heavy chain (*vmhc*) and SLU7 splicing factor homolog (*slu7*) were normalized to glyceraldehyde 3-phosphate dehydrogenase (*gapdh*) transcripts from isolated nuclei or whole embryos. *Myl7* and *vmhc* are expressed exclusively in the myocardium, while *slu7* is expressed ubiquitously. Error bars represent standard deviations from triplicate RT-qPCR experiments.
D Scatterplots of log2 fold differences between biotagged biological duplicates for whole embryo (*bactin* nuclear and *bactin* ribo), cardiomyocyte (*myl7* nuclear) and neural crest (*sox10* nuclear and *sox10* ribo) samples. **E** Table presenting correlation coefficients to all possible pairwise comparisons of replicates/samples.

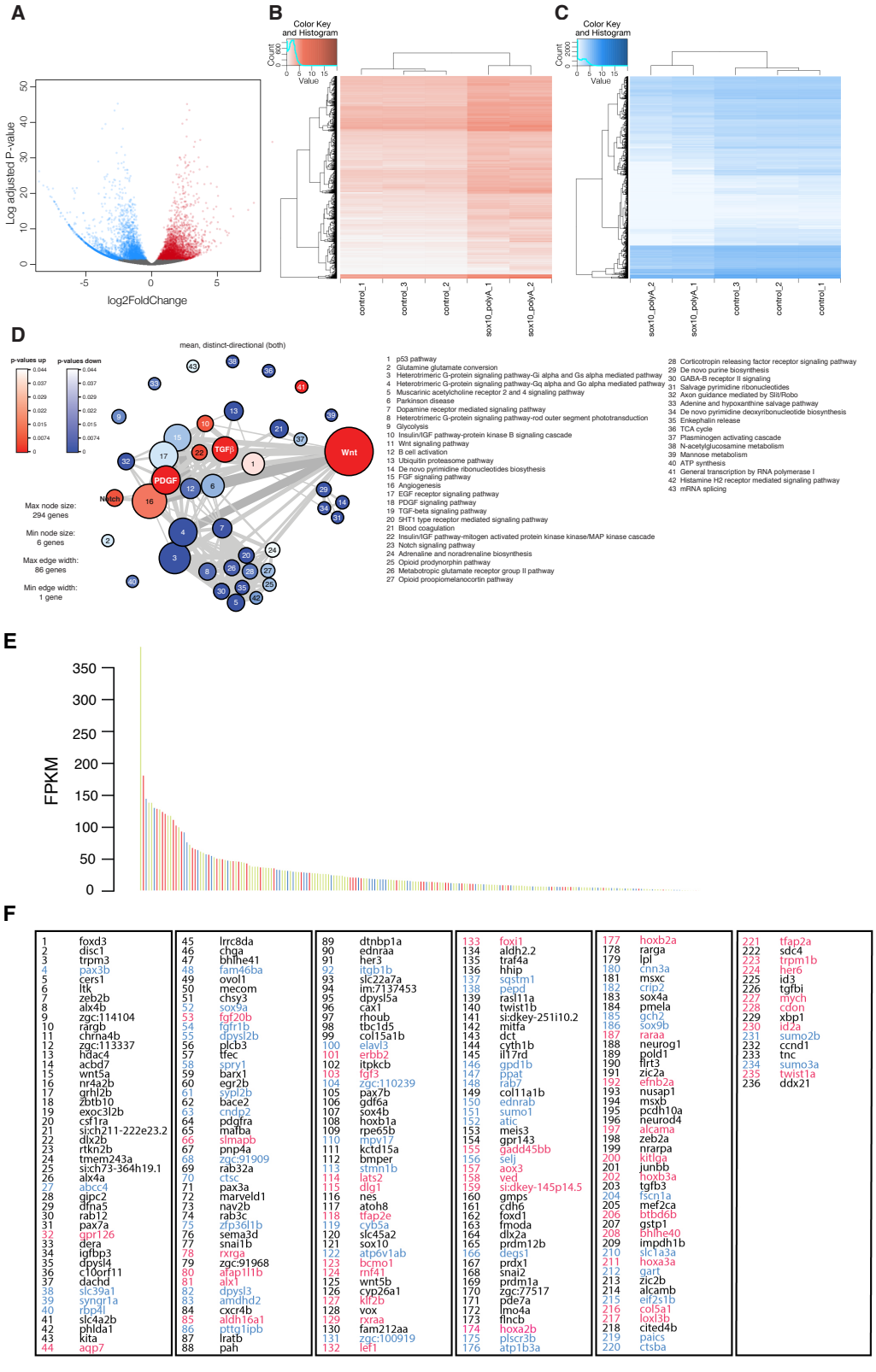


Figure S4. Neural crest identity of *sox10* Biotagged nuclei. Related to Figures 5 and 6.

A Volcano plot of differential expression analysis of ncBirA;nucAvi samples compared to whole embryo transcriptome, demonstrating the relationship between the *p*-value and log-fold change (red=enriched, blue= decreased in ncBirA;nucAvi samples) for the 9544 genes differentially expressed. **B** Heatmap of enriched genes. **C** Heatmap of depleted genes. Brackets highlight cluster of genes with variation between *sox10* biotagged samples. Heatmaps reveal differences in the *sox10* nuclear transcriptome compared to the whole embryo replicates. **D** Gene set enrichment analysis for the 3767 genes enriched (red) and the 5414 genes decreased (blue) in *sox10* nuclear transcriptome compared to whole embryo transcriptome. The maximum node contains 294 genes enriched for the Wnt signaling pathway, while the minimal node contains 6 genes decreased for general mRNA splicing machinery in the *sox10* nuclear samples. Size of node corresponds to number of genes in each gene set. The *p*-values are presented by color saturation; the numbers and their corresponding pathway for each node are listed below. **E** Barplot representation of average FPKM expression values across replicates of *sox10* nuclear polyA enriched transcriptome for all 236 neural crest genes as defined by *in situ* hybridization analysis from 0-24hpf in Zfin expression database. **F** List of 236 genes expressed in neural crest cells by 24hpf as defined by *in situ* hybridization as obtained from Zfin zebrafish gene expression database. Enriched genes (red text), decreased genes (blue text) and not differentially expressed (black text).

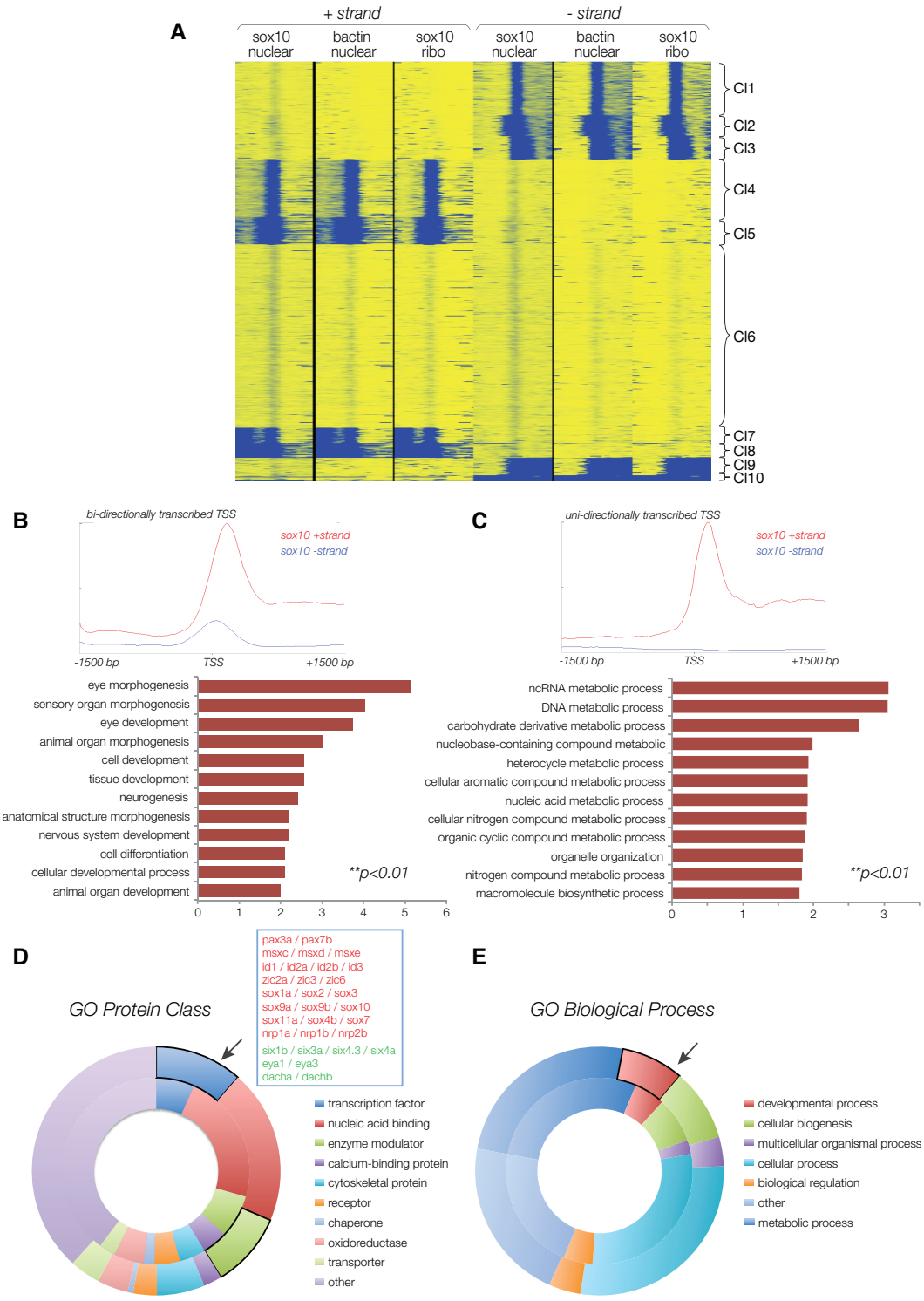


Figure S5: *K*-mean analysis identifies distinct clusters with bidirectional transcription at sites of open promoters. Related to Figure 5.

A Heatmap of *k*-mean clustering identifying 10 distinct clusters with varying levels of short bidirectional transcripts at open promoters. **B-C** Gene ontology (GO) terms for biological processes enriched for subclusters of genes with bidirectionally transcribed TSS in *sox10* nuclear dataset (B) and subclusters of genes with transcription at TSS only in sense direction (C). Bidirectionally transcribed loci associate with GO terms reflecting various developmental processes including eye and sensory morphogenesis, neurogenesis and cellular differentiation with high statistical significance ($p<0.01$), while loci not exhibiting bidirectional transcription at TSS associate with various metabolic processes and have no developmental feature ($p<0.01$). **D-E** Donut charts comparing functional classification according to protein class (D) and biological function (E) between genes with unidirectionally (inner donut chart) and bidirectionally transcribed TSSs (outer donut charts). Most significant difference between the two gene clusters is much larger number of transcription factors amongst bidirectionally transcribed loci (D) as well as increase in number of loci associated with developmental processes (E).

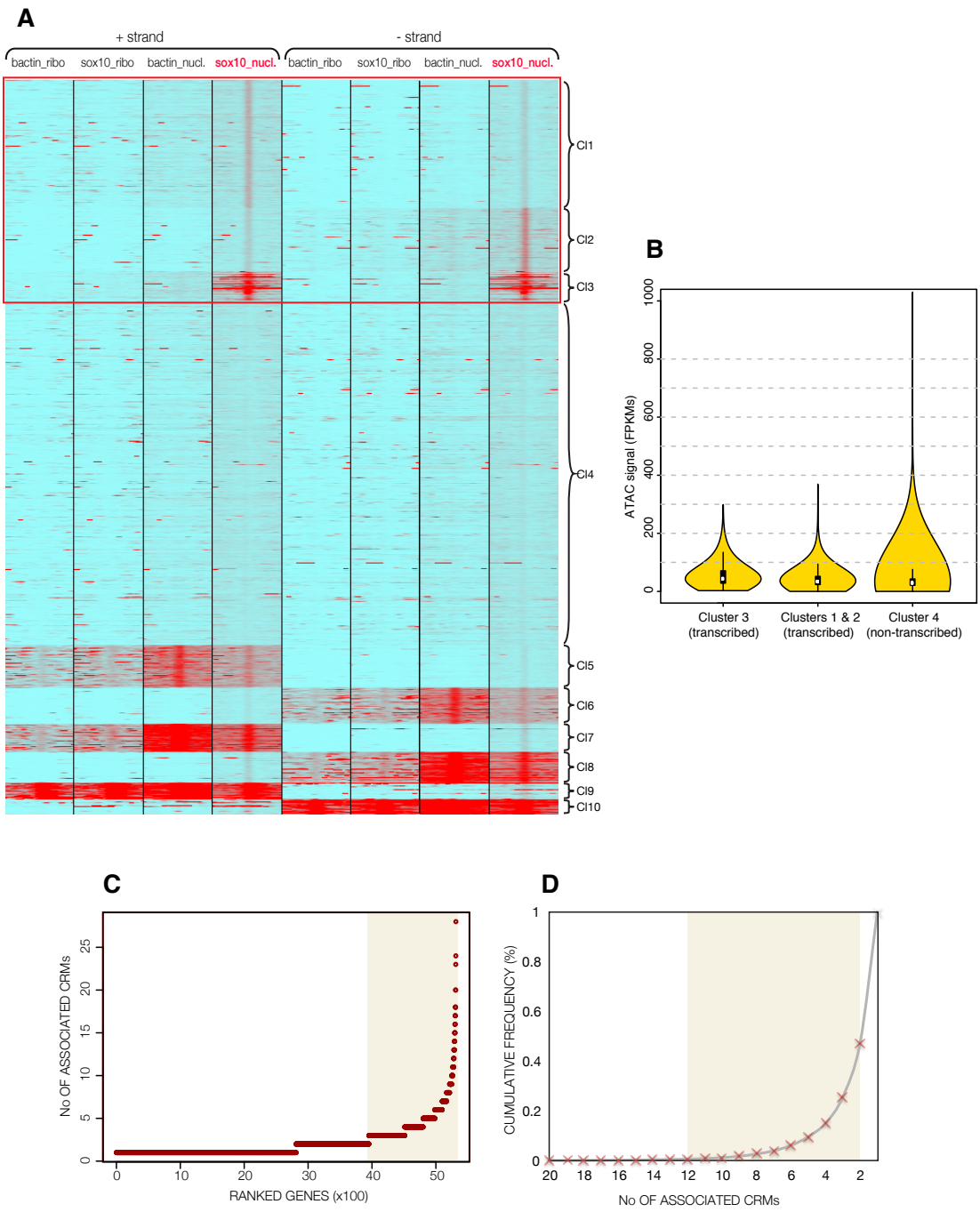


Figure S6. Identification of *cis*-regulatory modules (CRMs) through bidirectional transcription at sites of open chromatin (ATAC-peaks). Related to Figure 6.

A Heatmap showing all *k*-mean clusters from linear enrichment of mapped reads from *sox10* and *bactin* nuclear and ribosomal datasets associated across regions of open chromatin defined by ATAC-seq ($\pm$1.5kb per region). Ten clusters, totaling 65,458 distal open chromatin regions, were identified with approximately half of the ATAC-regions not associated with active transcription (cluster 4; 30,669 peaks). Three clusters show bidirectional transcription in *sox10* nuclear dataset (red box, also presented in Figure 6B of main text). Two groups of 2 reciprocal clusters each (clusters 5-6 and 7-8; total 12,230 CRMs) show associated bidirectional transcription in *sox10* and *bactin* nuclear dataset. These clusters most likely reveal ubiquitous enhancers. Two clusters contained elements with associated transcripts in both nuclear and ribosomal compartment (clusters 9-10; total 2,927 elements). **B** Violin plot visualising the distribution of ATAC-seq signal for NC-specific bidirectionally transcribed CRMs (Clusters 1, 2 and 3) and for non-transcribed accessible regions (Cluster 4). Although there's a greater variation in signal level distribution for the non-transcribed cluster 4, the median value of ATAC signal on transcribed and non-transcribed regions is similar. **C** Annotated genes ranked by the number of associated CRMs. **D** Cumulative frequency plot quantifying number of associated enhancers (identified from cluster 1 and 2, total of 6332 genes) per expressed gene (based on ATAC_TSS dataset). $\sim$47% of open annotated loci were associated with at least 2 CRMs, $\sim$25% with 3 or more elements and 15% with 4 or more elements.

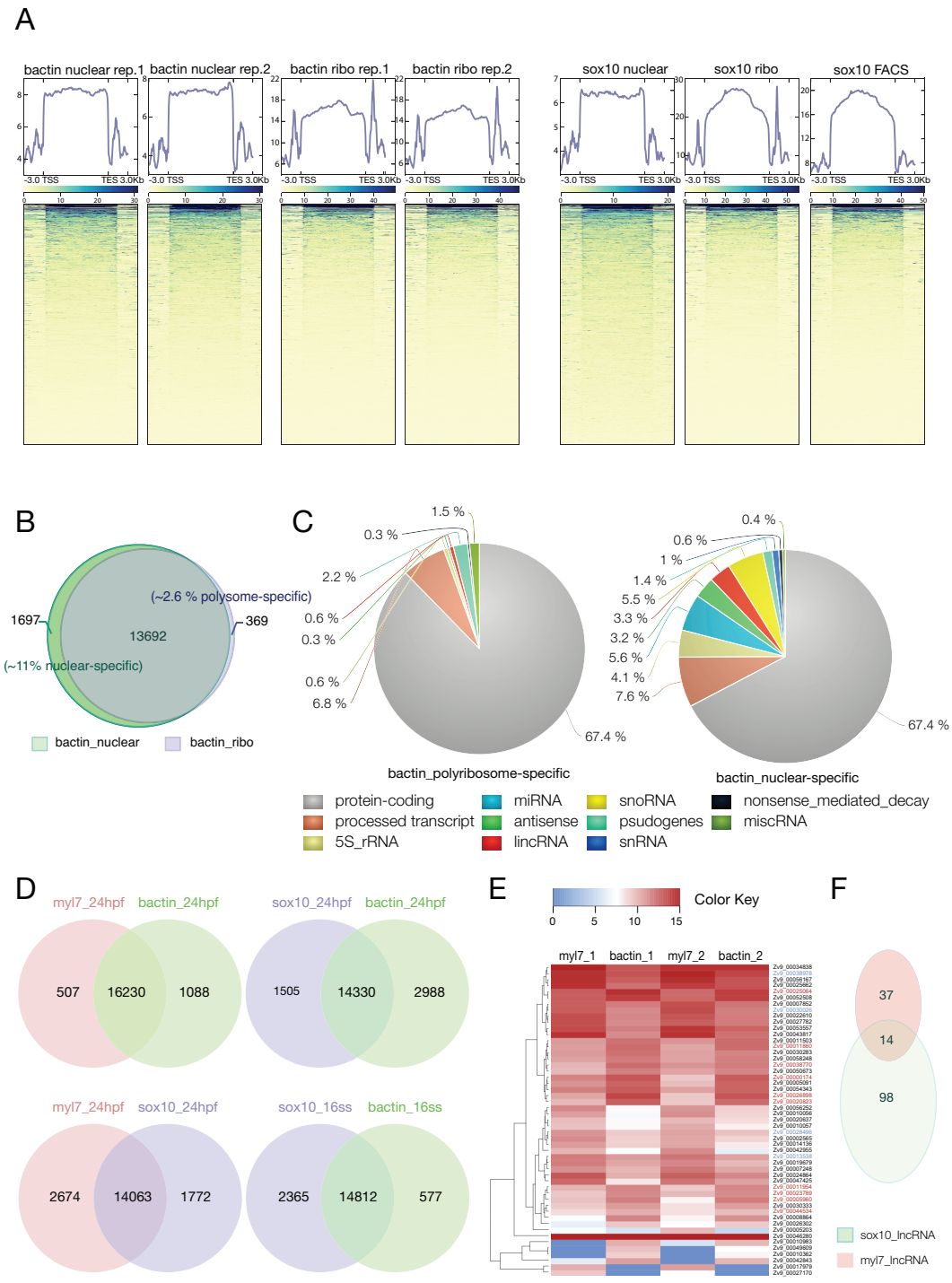


Figure S7. Comparative genome-wide profiles of nuclear and ribosomal transcripts in *sox10*, *myl7* subpopulation and whole embryo reveal subcellular compartment and cell-type differences. Related to Figure 7.

A Visualizations of the *bactin* and *sox10* nuclear transcriptomes on a global scale using deepTools heatmapper and profiler demonstrate a unique architecture characteristic of pervasive transcription that is enriched (with *bactin* ubiquitous control shown in replicates) evenly across untranslated and translated regions of annotated gene bodies demarcated by transcription start sites (TSS) and transcription end sites (TES). This enrichment is lost in the *sox10* ribosomal transcriptome signature (again, with *bactin* ubiquitous control shown in replicates), where higher levels of transcription in the central, translated regions of gene bodies define the transcriptional structure. As expected, the ribosomal transcriptional signature is highly similar to the profile of whole *sox10*-expressing cells that were obtained via FACS, where majority of transcripts ($\sim$90%) recovered in this manner are cytosolic. **B** Venn diagram showing RNA species distribution between nuclear and ribosomal dataset from whole embryo and **C** the corresponding RNA species found in the respective subcellular compartments with color-code referring to the different RNA species in lower panel. **D** Venn diagram comparing transcript differences between *myl7*, *sox10* and whole embryo nuclear datasets. **E** Heatmap of 51 differentially expressed lncRNAs ($p<0.05$) in the *myl7* versus *bactin* nuclear datasets (26-30hpf). 14 non-coding transcripts common between NC and myocardial differentially expressed lncRNAs are labelled in red (depleted) and blue (enriched) in the *myl7*- versus *bactin*-expressing nuclei. **F** Venn diagram comparing *NC* and *myl7* differentially expressed lncRNAs.

Table S1. Biotagging driver and effector lines. Related to Figures 1 and 2.

Biotagging BirA driver	Allele	Abbreviation	Expression	Note
Tg(bactin:BirA-membCherry)	ct709a	ubBirA	ubiquitous	
Tg(bactin:NLS-BirA-membCherry)	ct710a	ubBirA(NLS)	ubiquitous	
Tg(ubiq:NLS-BirA-Cherry)	ox114	ubBirA(ubiq-NSL)	ubiquitous	
Tg(ubiq:BirA-Cherry)	ox115	ubBirA(ubiq)	ubiquitous	
Tg(sox10:BirA-membCherry)	ct706a	ncBirA	cranial neural crest derivatives	female-exhibit ubiquitous maternal expression
Tg(sox10:BirA-membCherry)	ct706b	ncBirA(b)	cranial and trunk neural crest derivatives	
Tg(zic2a:NLS-BirA-membCherry)	ct708a	hbBirA	presumptive telencephalon and diencephalon, hindbrain primary neurons	expression gone by 2dpf
Tg(myl7:BirA-membCherry)	ct704a	myoBirA	myocardium	
Tg(myl7:NLS-BirA-membCherry)	ct705a	myoBirA(NLS)	myocardium	
Tg(kdrl:BirA-membCherry)	ct703a	endoBirA	vascular endothelium	
TgBAC(sox10:BirA-Cherry)	ox104a	ncBirA(BAC)	neural crest and derivatives	female may exhibit ubiquitous maternal expression

Biotagging Avi effector	Allele		Expression	Note
Tg(bactin:Avi-Cerulean-RanGap)	ct700a	nucAvi(bact)	ubiquitous	
Tg(bactin:RanGap-Cerulean-Avi)	ct701a	nucAvi(bact-Cterm)	ubiquitous	
Tg(ubiq:Avi-Cerulean-RanGap)	ox113a/b	nucAvi(ubiq)	ubiquitous	
Tg(bactin:Avi-Cerulean-Rpl10)	ox111	riboAvi(bact)	ubiquitous	
Tg(ubiq:Avi-Cerulean-Rpl10)	ox112	riboAvi(ubiq)	ubiquitous	

Table S2. Genes overrepresented in *myl7* vs *bactin* nuclei. Related to Figure 4.

1	ENSDARG00000077777	tmsb4x
2	ENSDARG00000055618	acta1b
3	ENSDARG00000021124	cfl1
4	ENSDARG00000021184	rbfox1l
5	ENSDARG00000052150	pbx4
6	ENSDARG00000017219	pabpc1a
7	ENSDARG00000023290	fabp3
8	ENSDARG00000002710	ncl
9	ENSDARG00000023963	tpm4a
10	ENSDARG00000029764	mef2ca
11	ENSDARG00000012820	nop56
12	ENSDARG00000059654	eif3ba
13	ENSDARG00000040334	mat2aa
14	ENSDARG00000057317	nexn
15	ENSDARG00000058115	fgfr2
16	ENSDARG00000011648	snrpd1
17	ENSDARG00000055026	ptch2
18	ENSDARG00000018460	mbnl2
19	ENSDARG00000020242	fermt2
20	ENSDARG00000037539	tnnc1b
21	ENSDARG00000052182	rbfox2
22	ENSDARG00000052004	cav1
23	ENSDARG00000058549	smyd1b
24	ENSDARG00000008433	unc45b
25	ENSDARG00000002731	sdc2
26	ENSDARG00000045843	apex1
27	ENSDARG00000070787	jupa
28	ENSDARG00000005333	tln1
29	ENSDARG00000071395	camk2g1
30	ENSDARG00000016744	fbn2b
31	ENSDARG00000030638	pdlim7
32	ENSDARG00000040237	mef2d
33	ENSDARG00000030213	sap30l
34	ENSDARG00000010721	septin6
35	ENSDARG00000008867	rap1b
36	ENSDARG00000093281	nipbla
37	ENSDARG00000013007	adrbk2
38	ENSDARG00000015681	gsk3ab
39	ENSDARG00000010844	kras
40	ENSDARG00000043923	sox9b
41	ENSDARG00000022968	fxr1
42	ENSDARG00000005629	smyd2b
43	ENSDARG00000017803	gsk3b
44	ENSDARG00000058449	arnt2
45	ENSDARG00000070670	cnip2
46	ENSDARG00000070721	vdrb
47	ENSDARG00000040001	fryl
48	ENSDARG00000009789	sgcd
49	ENSDARG00000014986	acvr1l
50	ENSDARG00000002791	atp1a1a.1
51	ENSDARG00000034893	rarab
52	ENSDARG00000044841	notch1a
53	ENSDARG00000021509	tab2
54	ENSDARG00000068028	trps1
55	ENSDARG00000012553	rap1aa
56	ENSDARG00000056964	ilk
57	ENSDARG00000011027	fgfr1a
58	ENSDARG00000059794	kdm6al
59	ENSDARG00000018693	cdh2
60	ENSDARG00000003938	rpa1
61	ENSDARG00000042535	actc1a
62	ENSDARG00000018881	apobec2a
63	ENSDARG00000071865	nrp1a
64	ENSDARG00000023878	stk24b
65	ENSDARG00000056322	ldb3a
66	ENSDARG00000053255	itgb1b
67	ENSDARG00000062909	furina
68	ENSDARG00000031756	mef2aa
69	ENSDARG00000042151	skia
70	ENSDARG00000006524	supt6h
71	ENSDARG00000061052	nipblb
72	ENSDARG00000091481	foxc1a
73	ENSDARG00000026531	alcama
74	ENSDARG00000038159	tcf7l1a
75	ENSDARG00000019728	bmpr1aa
76	ENSDARG00000020527	nup62l
77	ENSDARG00000077129	notch1a
78	ENSDARG00000006272	mpp5a
79	ENSDARG00000008947	rtf1
80	ENSDARG00000076393	tmem65
81	ENSDARG00000007369	tcf7l1b
82	ENSDARG00000056800	med12
83	ENSDARG00000037238	smad5
84	ENSDARG00000008034	skib
85	ENSDARG00000086515	rad21a
86	ENSDARG00000055357	leo1
87	ENSDARG00000058868	apc
88	ENSDARG00000077840	meis2b
89	ENSDARG00000002952	smo
90	ENSDARG00000068567	shha
91	ENSDARG00000019709	rhoua
92	ENSDARG00000053279	apln
93	ENSDARG00000074271	wrb
94	ENSDARG00000078366	robo2
95	ENSDARG00000033733	ntn1a
96	ENSDARG00000035054	pdcd10a
97	ENSDARG00000089368	hopx
98	ENSDARG00000016200	trib3
99	ENSDARG00000013705	ccm2
100	ENSDARG00000035957	gmnn
101	ENSDARG00000096282	notch1a
102	ENSDARG00000025309	dpf3
103	ENSDARG00000010655	ppm1k
104	ENSDARG00000040649	prickle1a
105	ENSDARG00000021225	prkci
106	ENSDARG00000019969	smarcd3b
107	ENSDARG00000096578	cul4a
108	ENSDARG00000010472	atp1a2a
109	ENSDARG00000004151	rab23
110	ENSDARG00000043137	cdca8
111	ENSDARG00000053209	smad6a
112	ENSDARG00000009418	mef2cb
113	ENSDARG00000011615	mybpc3
114	ENSDARG00000024452	cul4a
115	ENSDARG00000019717	pbx2
116	ENSDARG00000043886	paf1
117	ENSDARG00000020334	ptpn11a
118	ENSDARG00000006603	csrp1a
119	ENSDARG00000078293	rbm24a
120	ENSDARG00000034624	nuf2
121	ENSDARG00000078784	mgaa
122	ENSDARG00000040387	mmd
123	ENSDARG00000006092	rad21a
124	ENSDARG00000079027	chsy1
125	ENSDARG00000020053	nkd1
126	ENSDARG00000029370	ankrd6b
127	ENSDARG00000020201	cdc73
128	ENSDARG00000045971	vegfaa
129	ENSDARG00000037924	gna13b
130	ENSDARG00000055639	ruvbl2
131	ENSDARG00000009170	ctr9
132	ENSDARG00000054003	rargb
133	ENSDARG00000039022	stk25b
134	ENSDARG00000025826	gna12a
135	ENSDARG00000027423	igf1ra
136	ENSDARG00000007993	csnk1a1
137	ENSDARG00000019615	fn1a
138	ENSDARG00000017115	kctd10
139	ENSDARG00000002591	ruvbl1
140	ENSDARG00000008052	nudc
141	ENSDARG00000003398	rbpja
142	ENSDARG00000000943	dzip1
143	ENSDARG00000014685	esco2
144	ENSDARG00000014796	wnt11r
145	ENSDARG00000005150	tbx20
146	ENSDARG00000055151	smyd2a
147	ENSDARG00000019838	ugdh
148	ENSDARG00000038847	gins3
149	ENSDARG00000052091	rbpjb
150	ENSDARG00000026294	erbb2
151	ENSDARG00000056029	cyp26c1
152	ENSDARG00000041799	cx43
153	ENSDARG00000010432	eaf2
154	ENSDARG00000019008	erp44
155	ENSDARG00000052674	csnk1a1
156	ENSDARG00000006526	fn1b
157	ENSDARG00000045097	bmpr1ab
158	ENSDARG00000061600	tmem2
159	ENSDARG00000026784	robo1
160	ENSDARG00000031809	rbm24b
161	ENSDARG00000012591	pdcd10b
162	ENSDARG00000054030	hoxb5b
163	ENSDARG00000020057	bmpr2b
164	ENSDARG00000014891	robo2
165	ENSDARG00000030722	xirp1
166	ENSDARG00000059301	hmgcrb
167	ENSDARG00000014098	pkd2
168	ENSDARG00000012397	eya4
169	ENSDARG00000041421	taz
170	ENSDARG00000039677	dsc2l
171	ENSDARG00000036548	s1pr2
172	ENSDARG00000069150	rit1
173	ENSDARG00000011941	bmpr2a
174	ENSDARG00000035831	gata6
175	ENSDARG00000026348	csad
176	ENSDARG00000004251	dhfr
177	ENSDARG00000054438	fzd2
178	ENSDARG00000041022	pdcd4b
179	ENSDARG00000062236	ndrg4
180	ENSDARG00000036115	krit1
181	ENSDARG00000036194	pitx2
182	ENSDARG00000073765	gna13a
183	ENSDARG00000069090	atp6v0d1
184	ENSDARG00000018025	tbx2a
185	ENSDARG00000033999	cyp26a1
186	ENSDARG00000028067	bnip3lb
187	ENSDARG00000029738	fbln1
188	ENSDARG00000071465	lrrc39
189	ENSDARG00000019995	bmp4
190	ENSDARG00000002216	tbx3a
191	ENSDARG00000075621	birc5a
192	ENSDARG00000004532	cobl
193	ENSDARG00000070061	gfer
194	ENSDARG00000061017	spns2
195	ENSDARG00000034434	igf1rb
196	ENSDARG00000004246	slit2
197	ENSDARG00000036090	pard6gb
198	ENSDARG00000035427	surf6
199	ENSDARG00000000529	ofd1
200	ENSDARG00000019930	tal1
201	ENSDARG00000032976	cmlc1
202	ENSDARG00000042977	nfil3
203	ENSDARG00000016526	gata3
204	ENSDARG00000035253	npr3
205	ENSDARG00000060813	plekha7a
206	ENSDARG00000077004	aldh1l1
207	ENSDARG00000079564	vmhc
208	ENSDARG00000070623	med10
209	ENSDARG00000053535	lmo7b
210	ENSDARG00000031751	npr1a
211	ENSDARG00000055158	prox1a
212	ENSDARG00000070795	eaf1
213	ENSDARG00000009875	raf1a
214	ENSDARG00000011538	hspb7
215	ENSDARG00000018721	npnt
216	ENSDARG00000006120	tbx2b
217	ENSDARG00000078261	nphp3
218	ENSDARG00000006110	chd
219	ENSDARG00000029439	atp2a2a
220	ENSDARG00000079906	cacnb2a
221	ENSDARG00000041430	bmp2b
222	ENSDARG00000019096	myl7
223	ENSDARG00000069922	popdc2
224	ENSDARG00000038858	setd7
225	ENSDARG00000020610	tnnt2a
226	ENSDARG00000058548	bves
227	ENSDARG00000036670	aplnrb
228	ENSDARG00000004023	isl1
229	ENSDARG00000008305	hand2
230	ENSDARG00000017369	sema3d
231	ENSDARG00000087446	mmd2a
232	ENSDARG00000040291	cadm4
233	ENSDARG00000040085	ccsapb
234	ENSDARG00000046019	snai1b
235	ENSDARG00000052470	igfbp2a
236	ENSDARG00000044689	ttc8
237	ENSDARG00000044405	ttc4
238	ENSDARG00000062262	ednrab
239	ENSDARG00000052708	tnni1b
240	ENSDARG00000017821	gata5
241	ENSDARG00000038107	sgcg
242	ENSDARG00000035682	ltbp3
243	ENSDARG00000002213	invs
244	ENSDARG00000021110	poc1b
245	ENSDARG00000015815	kdrl
246	ENSDARG00000059774	dub
247	ENSDARG00000007025	ttc9c
248	ENSDARG00000053868	etv2
249	ENSDARG00000010191	ttc39c
250	ENSDARG00000067626	ywhag1
251	ENSDARG00000059327	gata2a
252	ENSDARG00000090124	alcama
253	ENSDARG00000059911	bbs7
254	ENSDARG00000036987	has2
255	ENSDARG00000075549	cdh5
256	ENSDARG00000003399	fgf8a
257	ENSDARG00000061527	nxt2
258	ENSDARG00000042624	pak1
259	ENSDARG00000041572	zfpm1
260	ENSDARG00000044339	rp2
261	ENSDARG00000004256	wnt11
262	ENSDARG00000060002	ogg1
263	ENSDARG00000035759	gata4
264	ENSDARG00000032801	grk5
265	ENSDARG00000042667	klf2a
266	ENSDARG00000019920	lft1
267	ENSDARG00000010591	foxn4
268	ENSDARG00000071208	wnt4a
269	ENSDARG00000091006	cobl
270	ENSDARG00000063522	bbs4
271	ENSDARG00000078055	acd
272	ENSDARG00000059545	foxj1a
273	ENSDARG00000023026	pkp2
274	ENSDARG00000090096	sgcg
275	ENSDARG00000075713	shox2
276	ENSDARG00000013441	hey2
277	ENSDARG00000036865	slc2a12
278	ENSDARG00000045626	nek8
279	ENSDARG00000011400	tnnc1a
280	ENSDARG00000037677	fgf24
281	ENSDARG00000076348	mylk3
282	ENSDARG00000028663	tek
283	ENSDARG00000053617	camk2a
284	ENSDARG00000075169	bbs1
285	ENSDARG00000058050	smyd3
286	ENSDARG00000016813	ccdc40
287	ENSDARG00000090637	myh6
288	ENSDARG00000045219	dkk1b
289	ENSDARG00000002210	stk25a
290	ENSDARG00000062978	ccdc151
291	ENSDARG00000078356	aplnra
292	ENSDARG00000042631	glrx2
293	ENSDARG00000039827	bbs5
294	ENSDARG00000041952	prox2
295	ENSDARG00000021232	nkx2.7
296	ENSDARG00000038867	shhb
297	ENSDARG00000016139	mkks
298	ENSDARG00000008398	cacna1c
299	ENSDARG00000024894	tbx5a
300	ENSDARG00000036116	eif3hb
301	ENSDARG00000088810	prox1b
302	ENSDARG00000071498	fhl1a
303	ENSDARG00000004405	snx10a
304	ENSDARG00000052960	nppa
305	ENSDARG00000019260	dhrs9
306	ENSDARG00000053455	ccdc103
307	ENSDARG00000053318	lrrc6
308	ENSDARG00000018004	nkx2.5
309	ENSDARG00000021889	gja3

Supplemental Experimental Procedures

Zebrafish maintenance and strains.

This study was carried out in accordance to procedures authorized by the UK Home Office in accordance with UK law (i.e. Animals (Scientific Procedures) Act 1986) and the recommendations in the Guide for the Care and Use of Laboratory Animals (US). Adult fish were maintained as described (Westerfield, 2000). Wild-type embryos for transgenesis were obtained from AB or AB/TL mix strains.

BirA cassette design.

The tripartite biotinylation driver expression cassette consists of open reading frame (ORF) of bacterial biotin ligase, BirA, preceded by 3xHA sequence for protein detection and separated from the membrane-tethered mCherry fluorescent reporter (membCherry) by a short sequence encoding the ribosome-skipping peptide of *Thosea asigna* virus (2A, Fig.1B-D). The membrane localization signal for membCherry was derived from the last 20 amino acids of human Ras (Apolloni et al., 2000). The biotinylation driver construct was generated by fusion PCR with 3XHA-BirA and membCherry templates, using intervening overlap sequence between the two to encode 2A sequence. The 20 bp overlap was built into the primers used for amplification of the templates. We have generated BirA drivers with both cytoplasmic and nuclear cellular localization (NLS) to enable biotinylation of both cellular component-associated Avi-tagged proteins (nuclear envelope and cell membrane) and Avi-tagged intra-nuclear factors, respectively. The full list of generated biotinylation drivers is shown in Table S1. Full sequences of the plasmids are available through NCBI and Addgene (https://www.addgene.org/Tatjana_Sauka-Spengler/).

Generation of Biotagging transgenic drivers.

Tol2-mediated transgenesis for driver lines: Biotagging transgenic drivers were created using conventional Tol2-mediated zebrafish transgenesis (Kawakami, 2004) . BirA expression cassette was placed under the control of previously published proximal enhancers and tissue- or cell-specific promoters (Table S1) to create defined expression patterns with the entire expression module flanked by Tol2 transposable elements. The presence of the 2A peptide allowed for simultaneous expression of BirA and the fluorescent membrane-Cherry (membCherry) reporter for screening and imaging purposes (Fig.1B-D, B'-D'). The transgenes were generated by co-injecting 80 pg DNA expression constructs and 40 pg of *tol2* mRNA into single cell embryos. The injected embryos (mosaic F_0 generation) were raised to reproductive age, out-crossed to wildtype adults and the F_1 offspring screened for proper expression of fluorescent reporter. F_1 carriers were raised for future experiments. Subsequent generations of transgenes are maintained as out-crosses to wildtype adults to ensure single copy transgenic propagation.

Tol2-mediated Biotagging transgenic drivers for expression of BirA produced in this study include four tissue-specific lines: *Tg(sox10:BirA-2A-membCherry)*^{ct706a} (ncBirA) expressing BirA in delaminating and migrating neural crest under control of the *sox10* promoter (Carney et al., 2006) (Fig.1B,B'; Fig.2G); *Tg(zic2a:BirA-2A-membCherry)*^{ct708a} (hbBirA) in the neural plate border cells under the control of the *zic2a/zic5* enhancer (Nyholm et al., 2007) (Fig.1D,D'); *Tg(myl7:BirA-2A-membCherry)*^{ct704a} (myoBirA) in the myocardium of the developing heart (Huang et al., 2003) (Fig.1C,C') and *Tg(kdrl:BirA-2A-membCherry)*^{ct703a} (endoBirA) in endothelium of the circulatory system (Jin et al., 2005) (not shown). In addition, ubiquitous BirA driver lines included *Tg(bactin:BirA-2A-membCherry)*^{ct709a} (ubBirA) and *Tg(ubiq:BirA-2A-mCherry)*^{ox115} (ubBirA(*ubiq*)) (Higashijima et al., 1997; Mosimann et al., 2011) (Fig.1E). Test experiments, using homozygote *Tg(bactin:BirA-2A-membCherry)*^{ct709a}(ubBirA)embryos (Fig.1E), which express BirA at very high levels, show no developmental defects and can reproduce, indicating that expression of BirA is not toxic in zebrafish.

BAC recombineering for driver lines: The availability of well characterized *cis*-regulatory modules (CRMs) for BirA drivers can limit the application of the Tol2-mediated transgenes driver approach. Recombineered BACs, containing gene-associated regulatory elements, can serve as an alternative to transgenic BirA drivers using known CRMs. To generate Biotagging BAC drivers we replaced the first coding exon of the gene of interest with a BirA cassette (Fig.1G) and used Tol2-mediated transgenesis to integrate recombineered BACs into zebrafish genome

(Bussmann and Schulte-Merker, 2011; Suster et al., 2009). To achieve this, we generated the donor cassette containing HA-tagged BirA ORF, separated from mCherry reporter by ribosomal-skipping peptide (2A) and terminating with polyA, followed by FRT site-flanked Kanamycin selection gene and recombined it into the selected BAC backbone using lambda prophage homologous recombination system available in the SW105 bacterial background (http://ncifrederick.cancer.gov/research/brb/productDataSheets/recombineering/bacterialStrains.aspx), according to the previously published protocol (Yu et al., 2000). To increase the efficiency of transgenesis and enable single-copy integration into zebrafish genome, the BACs were also modified to include the long terminal repeats (LTRs) of the Tol2 transposon. The iTol2-Amp cassette, containing the Ampicillin expression construct flanked by inverted Tol2 recombination arms was amplified according to published protocols (Abe et al., 2011), using iTol2-Amp plasmid and previously described primers:

pIndigobac_itol2_fw:
TTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTCCCTGCTCGAGCCG
GGCCCAAGTG
pIndigobac_itol2_rev:
CCCGCCAACACCCGCTGACGCGAACCCCTTGCGGCCGCATATTATGATCCTCTAGATCAGATCT

and recombined into the *sox10* locus-containing BAC DKEY-201F15, with *pIndigo* backbone. We have also created an extended iTol2-Amp cassette for integration of Tol2 arms into BAC clones with the *pTARBAC* backbone. The new iTol2-Amp cassette containing long homologous recombination arms (5'arm-224 bp and 3'arm-221 bp) that flank loxP sites on the pTARBAC backbone can be amplified using pTARBAC_loxP_5': GCTGTCGGAATGGACGATA and pTARBAC_loxP_3': GCAAGTATTGACATGTCGTCGT primers and recombined using procedures described above. 100-200 pg of recombinant BAC DNA was co-injected with 50-100 pg of *tol2* mRNA into one cell-stage embryos to generate F_0 generation. Potential F_0 founders were raised, outcrossed and the F_1 clutches screened for mCherry expression. Selected positive F_1 embryos were raised for future experiments.

Using a BAC containing the *sox10* locus we successfully overcame variation in expression patterns obtained from the conventional *sox10* transgenics (Fig.S1A-C). We detected strong neural crest expression of mCherry in the F_1 offspring from 3 out of 9 screened *TgBAC(sox10:BirA-Cherry)*^{ox104a} (ncBirA(BAC)) F_0 adults (Fig.1G-H, Fig.S1A). All the F_1 embryos from the three independent founders showed consistent neural crest-specific mCherry expression (Fig.1G-H, Fig.2I). This is in contrast to *sox10* BirA (ncBirA) drivers, obtained using conventional proximal promoter transgenesis, which often exhibited variability in expression patterns between different founders due to integration position effects (Fig.S1B,C). However, we exploited this variability to generate a number of biotagging drivers that enable profiling of specific subpopulations of *sox10*-expressing cells (Fig.S1B,C).

Differences between Biotagging with Transgenic and with BAC drivers: Biotagging transgenic drivers generated via routine transgenesis approaches in zebrafish express BirA-membCherry cassette under the control of the minimal promoter and proximal enhancers (Fig.1B-D). Using this approach, it is typical to observe a large variability in expression patterns in different founders produced with the same expression construct (Fig.S1B,C), suggesting that such expression cassettes are very sensitive to position effects, and that their activity is strongly influenced by their genomic integration sites. Exploiting this variability, we generated a number of Biotagging drivers that enable profiling of specific subpopulations of *sox10*-expressing cells (Fig.S1B,C). However, many of the *sox10* drivers do not recapitulate the full *sox10* gene expression pattern characterized using *in situ* hybridization.

Biotagging BAC drivers are much more consistent with all founders always showing reproducible expression patterns (Fig.S1A). Moreover, they are virtually insensitive to position effects and identified genomic integrations were never found to be silent. In majority of cases, the expression of BirA transgene from the BAC backbone results in robust and strong endogenous-like level of expression.

Generation of Avi effector lines.

To generate zebrafish transgenic effector lines, we used the ubiquitous zebrafish beta-actin2 (*βactin*) and ubiquitin (*ubiq*) promoters to drive expression of Avi-tagged fusions that associate with different cellular compartments: (i) Avi-tagged protein containing Cerulean protein fused to the carboxy-terminal domain of avian Ran GTPase-activating protein 1 (RanGap1), targeted to the outer nuclear envelope, for use in INTACT procedure and (ii) Avi-tagged Rpl10a to biotinylate the polyribosomes, for use in TRAP approach.

Generating Avi-RanGap lines: To generate Avi-RanGap (nucAvi) lines, in which the nuclear envelope is specifically Avi-tagged we used C-terminal domain of the avian RanGap protein because fusions with the equivalent region from the zebrafish RanGap resulted in recombinant protein that inconsistently associated with nuclear envelope, displaying much broader cellular distribution, and affecting normal development (construct resulted in embryonic lethality when injected at high concentrations of $100\,\mathrm{pg}$ of mRNA per embryo, data not shown). Protein domain analysis of chicken and zebrafish RanGap domains using SMART tools (http://smart.embl-heidelberg.de) indicated that the chicken but not the zebrafish RanGap C-terminal region contains a Ran Binding Domain, which associates with nuclear pore complexes (Mahajan et al., 1997; Rose and Meier, 2001), to directly bind Avi-Cerulean-RanGap fusion and localize it to the outer nuclear envelope (Fig.3A, B). We generated transgenic lines expressing the both N and C-terminal Avi fusion of avian RanGap (*Tg(bactin:Avi-Cerulean-RanGap)*^{ct700a} (nucAvi(*bact*) and *Tg(bactin:RanGap-Cerulean-Avi)*^{ct701a} (nucAvi(*bact*-Cterm)). They show similar localization to the outer nuclear envelope (Fig.S1F-I).

All N-terminal effector constructs employed a modified Avi tag (14 aa), followed by a 7 aa-peptide specifically recognized and cleaved by Tobacco Etch Virus protease, generated by Strouboulis lab (Driegen et al., 2005) (Fig.S1D, TeV in green). Inclusion of a protease cleavage sequence adjacent to biotin acceptor peptide helps reduce the non-specific background resulting from streptavidin bead pull-down of endogenously biotinylated proteins. This is particularly useful in analysis of protein complexes by Mass Spectrometry, where the biotin-tagged target protein and its interacting partners can be specifically released from the streptavidin beads by TeV cleavage. C-terminal Avi-tag is preceded by TeV sequence (Fig.S1E).

To assess the possible steric effects of the Avi-tag on the localization of the RanGap fusion protein and its availability for biotinylation, we have created two versions of Avi-tagged RanGap effector lines, Avi-Cerulean-RanGap and RanGap-Cerulean-Avi (Fig.S1D,E). Both versions localized to the outer nuclear envelope (Fig.S1F-I) and are interchangeable; however, we preferentially use the *Tg(bactin:Avi-Cerulean-RanGap)*^{ct700a} (nucAvi(*bact*)) line, where the RanGAP domain in the protein fusion is located at the C-terminus, similar to the full-length RanGap protein. To select for the most ubiquitously, even expressing Avi-tagged RanGap effectors, the offspring of multiple founders was screened by confocal microscopy.

Generating Avi-Rpl10 lines: To enable isolation of the polyribosomes, we established lines that label a component of polyribosomes, Rpl10a (Tryon et al., 2013), with the Avi-tag *Tg(bactin:Avi-Cerulean-Rpl10)*^{ox111} (riboAvi(*bact*)) and *Tg(ubiq:Avi-Cerulean-Rpl10)*^{ox112} (riboAvi(*ubiq*)). While lines established with the *ubiquitin* promoter resulted in higher expression of Avi-Cerulean-Rpl10, we also found that high expression of Rpl10 resulted in embryonic lethality (data not shown). Only founders that resulted in Avi-Rpl10 lines showing no developmental defects that could reproduce were maintained and used for subsequent profiling.

Preparation of Streptavidin beads for RNA procedures.

As Streptavidin Dynabeads (Invitrogen, cat. no.11205D) are not supplied in RNase-free solutions, $250\,\mu\mathrm{g}$ of M-280 or MyOne T1 Streptavidin-coated beads was transferred to a microcentrifuge tube, separated from supernatant using magnetic stand (DynaMag™-2 magnet from Invitrogen, cat. no.12321D) and washed twice with $1\,\mathrm{mL}$ of Solution A (DEPC-treated $0.1\,\mathrm{M}$ NaOH, DEPC-treated $0.05\,\mathrm{M}$ NaCl), with 2-3 minutes on a nutator, followed by at least 3 minutes on the magnetic stand. The beads were subsequently washed once in Solution B (DEPC-treated $0.01\,\mathrm{M}$ NaCl), re-suspended in NPB and moved to a new RNase-free low-binding tube, until the nuclei suspension was ready. Prior to incubation with cell nuclei/polysomes, the beads were captured using the magnetic stand, the supernatant removed, and replaced with the nuclei/polysomes suspension.

Nuclei Isolation.

We optimized the nuclei purification protocol for the highest yield by testing a number of lysis buffers and found the following to give the highest consistent yield of nuclei per embryo. Zebrafish embryos (~100-350 embryos per pulldown experiment) expressing both biotinylation driver and Avi-RanGap effector alleles (nucAvi) in specific cell types were dechorionated and washed in hypotonic Buffer H (20 mM HEPES (pH 7.4), 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT and 1X cOmplete™ protease inhibitor), supplemented with 0.01% Tricaine. Embryos then were re-suspended in 1 mL/50 embryos Buffer H and transferred to a Dounce homogenizer (2 or 7 mL Kontes Glass Co, Vineland, NJ). Embryos were dissociated with a sequence of 20 strokes using loose fitting pestle A, incubated on ice for 15 minutes, followed by 60 strokes of tight fitting pestle B (3 x 20 strokes, pausing 5 minutes on ice after each set of 20 strokes) to allow for lysis of cell membranes. Cells were checked for lysis by visualizing cells with 1:1 dilution of Trypan blue on microscope. Nuclei were collected by centrifugation for 10 minutes at 2,000g and re-suspended in 1 mL of nuclei purification buffer (NPB: 10 mM HEPES (pH 7.4), 40 mM NaCl, 90 mM KCl, 0.5 mM EDTA, 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT and 1X cOmplete™ protease inhibitor (Roche, cat. no.05892791001)). To purify nuclei, the suspension was incubated with 250 µg (1.5 x 10^7 beads) of M-280 Streptavidin-coated Dynabeads prepared for RNA procedures (see preparation of beads) with rotation at 4 °C for 30 minutes. The nucleus-bead suspension was further diluted with NPB containing 0.1% Triton X-100 (NPBt) to a final volume of 20 mL. A flow-based setup was devised as previously published (Deal and Henikoff, 2011) using 10 mL plastic serological pipette (VWR, cat. no.89130-898 or BD Falcon, cat. no.357551) attached to a 1 mL micropipette tip (Gentaur Reach Pipet Tip, cat. no.24-165R or Rainin, cat. no.RT-L1000S), both pre-treated with NPB + 1% BSA for 10 minutes to avoid material loss. To capture the specifically biotinylated nuclei, the 1 mL pre-coated pipette tip was attached to the MiniMACS separator magnet (OctoMACS Separator, Miltenyl Biotec, cat. no.130-042-109) and the diluted suspension of nuclei-beads was allowed to flow through the setup. The flow rate was set to ~0.75 mL min^{-1} using a two-way stopcock (Bio-Rad, cat. no.732-8102) or a T-valve (from Bio-Rad Low-pressure fittings kit, cat. no.731-8220), connected at the end of the tip via a short piece of Tygon tubing (Fisher Scientific, cat. no.14-169-1C). For flow rates faster than 0.75 mL min^{-1}, the flow-through was drawn up with the same serological pipette, without removing the pipet tip from the magnetic separator and the suspension was allowed to drip through the tip one more time to maximize recovery efficiency. Subsequently, pipet tip was removed from the stand, attached to p1000 pipetman and beads and nuclei were released from the wall of the tip by repeated drawing of fresh 20 mL NPBt into and out of the tip and the magnetic purification was repeated as described above. At the end, beads and nuclei were eluted in 1 mL of NPBt, placed into the 1.7 mL microfuge tube and onto the magnetic stand (DynaMag™-2 magnet from Invitrogen, cat. no.12321D) to remove the supernatant. Purified nuclei-beads were re-suspended in 20 µL NPB, stained with DAPI and imaged or frozen for future use. Nuclei yield after purification were determined by staining 1:20 of the total nucleus prep with DAPI and subsequent counting of the number of bead-bound nuclei. When used for RNA extraction, the 250 µg bead-nuclei pellets were immediately dissolved in 100 µL of RNA lysis buffer, incubated at room temperature for 10 minutes and replaced onto the magnetic stand. The RNA lysis buffer containing total nuclear RNA is then removed into a fresh tube and snap-frozen for future use or for immediate extraction (see RNA extraction and library preparation).

Polysomal Isolation.

We adapted the polysomal isolation protocol from TRAP, **T**ranslating **R**ibosome **A**ffinity **P**urification (Heiman et al., 2014) with modifications optimised for the Biotagging system. Zebrafish embryos (350 embryos) expressing both biotinylation driver and Avi-Rpl10a effector alleles (riboAvi) in specific cell types were dechorionated and washed in Cell Lysis Buffer (20 mM HEPES (pH 7.4), 150 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, rRNasin, RNaseOUT, SUPERaseIN and cOmplete™-EDTA-free protease inhibitor). Embryos were lysed in 7 mL of Cell Lysis Buffer without cycloheximide in a Dounce homogeniser as described in the Nuclei Isolation protocol above. The key difference is that cycloheximide was added to the cell suspension after pestle A to a final concentration of 100 µg mL^{-1} and incubated at RT for 15 minutes before proceeding with pestle B strokes. Homogenised embryos were cleared by centrifugation at 2000g for 10 minutes at 4 °C and the post-nuclear supernatant removed into clean RNase-free 1.7 mL microfuge tubes (1 mL supernatant per tube). IGEPAL CA-630 and 07:0 DHPC were added to the supernatant to a final concentration of 1% each and mixed by inverting tubes gently 10 times. Following incubation on ice for 5 minutes, the supernatant was cleared by centrifugation at 20,000g for 10 minutes at 4 °C.

To purify polysomes, the post-mitochondrial supernatant was removed and added to $250\,\mu g$ (2.5 x 10^8 beads) of MyOne T1 Streptavidin-coated Dynabeads ($1\,mL$ supernatant per $250\,\mu g$ beads) prepared for RNA procedures (see preparation of beads) with rotation at $4\,^{\circ}C$ for 1 hour. The tubes containing polysomes-beads suspension were placed onto a magnetic stand (DynaMag™-2 magnet from Invitrogen, cat. no.12321D) to remove the unbound lysate. The pelleted polysomes-beads were washed four times (changing tubes in between washes to minimise background) in the cold room with High Salt Buffer ($20\,mM$ HEPES (pH 7.4), $350\,mM$ KCl, $10\,mM$ $MgCl_2$, $0.5\,mM$ DTT, rRNasin, RNaseOUT, SUPERaseIN, $100\,\mu g\,mL^{-1}$ cycloheximide and 1% IGEPAL CA-630) by pooling $500\,\mu g$ beads per $1\,mL$ High Salt Buffer. After the final wash, the tubes with polysomes-beads were placed onto a magnetic stand to remove the High Salt Buffer. When used for RNA extraction, the $500\,\mu g$ polysomes-beads pellets were immediately dissolved in $200\,\mu L$ of RNA lysis buffer, incubated at RT for 10 minutes and replaced onto the magnetic stand. The RNA lysis buffer containing polysomal-bound RNA is then removed into a fresh tube and snap-frozen for future use or for immediate extraction (see RNA extraction and library preparation).

FACS and ATAC.

The *sox10*-expressing cells were isolated from *TgBAC(sox10:BirA-mCherry)*^{ox104a} (ncBirA(BAC)) embryos at 16ss using Fluorescence Activated Cell Sorting (FACS). Prior to FACS embryos were dissociated using $20\,mg\,mL^{-1}$ collagenase in 0.05% Trypsin/$0.53\,mM$ EDTA/1xHBSS buffer to obtain single cell suspensions. Reaction was stopped in $10\,mM$ HEPES/0.25% BSA/1xHBSS buffer and mCherry-positive neural crest cells were sorted using BD FACSARIA Fusion System. Sorted cells were spun down and washed in PBS and immediately used in ATAC procedure. Tagmentation was performed as previously described (Buenrostro et al., 2015). Fragment size was verified using Tapestation (Agilent) and libraries were quantified using KAPA Library Quant Kit for Illumina Sequencing Platforms (KAPABiosystems).

Western blot analysis.

Protein extract was obtained from zebrafish embryos at specified time-points. Embryos were de-yolked, lysed with a Dounce homogenizer (Pestle A) in hypotonic buffer ($20\,mM$ HEPES (pH 7.9), $1.5\,mM$ $MgCl_2$, $10\,mM$ KCl, and $1\,mM$ DTT) with protease inhibitors, and centrifuged at maximum speed to obtain cytoplasmic fraction. Nuclear fraction was obtained by lysis of remaining nuclei pellet in nuclear lysis buffer ($20\,mM$ HEPES (pH 7.9), $1.5\,mM$ $MgCl_2$, $0.2\,mM$ EDTA, 20% glycerol, $420\,mM$ KCl, $0.4\,mM$ PMSF) with protease inhibitors using Pestle B followed by centrifugation at maximum speed. Detection was performed with anti-HA antibody (Roche (cat. no.12CA5), 1:1000), rabbit anti-GFP antibody (1:1000, Torrey Pines Biolabs, Houston/TX, www. chemokine.com, used for detection of Cerulean) and Streptavidin-HRP Conjugate (used for detection of biotin). GFP and HA-tag were detected using standard Western blot procedure, while biotinylated proteins were detected using a modified procedure. After the transfer, the blots were blocked for 1 hr in 5% BSA/1X TBST ($20\,mM$ Tris, $137\,mM$ NaCl, 0.2% TWEEN-20) and incubated for 1 hr at room temperature with Streptavidin–HRP conjugate (NEL750, Perkin Elmer,1:10,000). Filters were then washed 6 times for 20 minutes in 1X TBS ($20\,mM$ Tris, $137\,mM$ NaCl) + 0.3% Triton X-100 and signal was detected using ECL Plus Western Blot Detection Reagent (GE Healthcare Life Sciences, cat. no. RPN2132).

Quantitative Real-Time PCR analysis.

Two-step qRT-PCR was performed using ABI's Sybr-Green RT-PCR system (Applied Biosystems). Briefly, RNA was extracted from nuclei isolated from 48hpf embryos double transgenic for the Biotagging myocardial driver allele, *Tg(myl7:BirA-2a-membCherry)*^{ct704a} (myoBirA) and the Avi effector allele, *Tg(bactin:Avi-Cerulean-RanGap)*^{ct700a} (nucAvi(*bact*)) using RNAqueous Micro kit (see RNA extraction and library preparation). cDNA was synthesized with reverse transcriptase (SuperScript II RT, Invitrogen) using random hexamers for priming. Reverse transcription reactions were diluted in series (1-10,000-fold) and $1\,\mu L$ was amplified in triplicates on a 7000 Sequence Detection System (Applied Biosystems). Quantification was performed using the delta-delta Ct ($\Delta\Delta$Ct) method (Livak and Schmittgen, 2001). Primers used for qRT-PCR are as follows: *myl7* (myl7_fw: AGGGGGAAAACTGCTCAAAG and myl7_rev: TGATAACTCCATCCCGGTTC), *vmhc* (vmhc_fw: TCGTCAGTCGTGAAGGTGAC and vmhc_rev: GGCTCATGAAGGAAGGTGAA), *slu7*

(slu7_fw: AGAAAAGGAGCATGCGAAAA and slu7_rev: atgcctgtgccagaaaactt) and *gapdh* (gapdh_fw: GATACACGGAGCACCAGGTT and gapdh_rev: CGTTGAGAGCAATACCAGCA).

Quality control for Nuclei and Polysomal Isolation, RNA extraction, and library preparation.

Maximal efficiency of nuclei and polysomal isolation is highly dependent on complete lysis of cells with Buffer H (nuclei isolation) and Cell Lysis Buffer (polysomal isolation). Hence, it is essential that embryos are lysed at a ratio of 50 embryos (26hpf or younger) per $1\,\mathrm{mL}$ of buffer with 20 strokes of pestle A and 60 strokes of pestle B. Care must be undertaken that pestle B is not faulty and has the appropriate small clearance for efficient cell lysis. Quality of cell lysis for each experiment was determined by RNA extraction of the unbound fraction. For nuclei isolation, the flow-through containing unbound nuclei (the 'lysate') was pelleted by centrifugation at 2000g for 10 minutes at $4\,^{\circ}\mathrm{C}$ and dissolved in $400\,\mu\mathrm{L}$ RNA lysis buffer. The sample was placed onto a magnetic stand to remove residual Dynabeads; RNA lysis buffer containing total RNA from the lysate was processed in the same manner. Similarly, for polysomal isolation, the supernatants containing unbound polysomes were cleared by centrifugation at 7000g for 10 minutes at $4\,^{\circ}\mathrm{C}$, the supernatant removed until about $200\,\mu\mathrm{L}$ is left, followed by addition of $800\,\mu\mathrm{L}$ RNA lysis buffer.

Both purified and unbound nuclei and polysomes were lysed and RNA pools extracted using RNAqueous Micro Scale Total RNA Isolation Kit (Ambion cat. no.AM1931), genomic DNA was removed by 20 minutes of rDNaseI (provided with Ambion cat.no.AM1931) treatment. Before library production, quality of the RNA was assayed using Agilent RNA 6000 Pico kit (Agilent Technologies, cat. no.5067-1513) on the Agilent 2100 Bioanalyzer, as specified by manufacturer. We ensured that in both nuclei and polysomal isolation experiments, the Bioanalyzer profiles for experimental and lysate samples were highly similar and displayed the expected quantity ratios (i.e. much higher amount of RNA in the lysate sample compared to experimental sample). Non-directional sequencing libraries after polyA-selection of RNA transcripts (NEBNext® Poly(A) mRNA Magnetic Isolation Module, NEB) were built using NEBNext Ultra RNA library kit for Illumina (NEB). For directional RNA-sequencing, 30-$50\,\mathrm{ng}$ of total nuclear RNA and 40-$50\,\mathrm{ng}$ of total polysomal RNA were first enriched by ribodepletion using Ribo-Zero™ Magnetic Kit (Epicentre). Subsequently sequencing libraries were prepared using Stranded RNA-Seq Library Preparation Kit (KAPABiosystems), according to manufacturer's instructions. Deep sequencing was performed on HiSeq2500 or Nextseq500 Illumina platforms. Biological duplicates were generated for each experimental condition and pairwise comparison performed on biological duplicates to ensure high quality of sequence data for analysis (Fig.S3D). cDNA libraries for RT-PCR validation were generated using Superscript II RT and random hexamer priming (Life Technology, cat. no.18064-014). To directly compare different total RNA isolation protocols - biotagged nuclei, biotagged ribosomes and FACS total RNA yield was quantified per batches of 100 embryos. Number of positive cells/organelles per embryos recovered was deduced and calculated using previously defined standards of $1\,\mathrm{pg}$ RNA/cell for Avi-RanGap, $0.05\,\mathrm{pg}$ RNA/cell for Avi-Rpl10 and cell counts from FACS experiment.

Bioinformatics Processing.

ATAC-Seq: ATAC-seq data was sequenced using paired-end 40 bp run on the NextSeq500 platform. Reads were trimmed for quality using sickle (v 1.33) (Joshi and Fass, 2011) and mapped using bowtie (v.1.0.0). Bigwig files were generated using an enhanced Perl script courtesy of Jim Hughes. Only paired reads with insert sizes larger than 100 bp were selected and reads were displaced by +4 bp and -5 bp as described previously (Buenrostro et al., 2013) and extended to a read length of 100bp. Peak calling was performed using MACS2 with –nomodel and –slocal 1000 parameters (Zhang et al., 2008). Zebrafish Ensembl gene models were extended by 100 bp in 5' of the TSS to account for gene mis-annotation. ATAC-seq peaks overlapping with extended TSSs were used to define open promoter set (**ATAC_TSS**). Putative *cis*-regulatory element set (**ATAC_enhancer**), was identified as ATAC peaks not overlapping with Ensembl-annotated promoter regions or exons.

RNA-seq analysis: RNA-Seq data was sequenced using 50 bp paired-end reads on HiSeq2000 and HiSeq2500 platforms. Whole embryo polyadenylated transcriptome at 24hpf generated by Armant and colleagues (Armant

et al., 2013) was downloaded from SRA (Accession SRP014596). Reads were mapped to the zebrafish genome (Jul.2010 Zv9/danRer7 assembly) with STAR (v.2.4.2a) using default parameters (Dobin et al., 2013). Sets of BAM files incorporating reads belonging to either DNA strand were generated using custom scripts available at https://github.com/tsslab/biotagging/. Count tables were produced for Ensembl gene models using subread featureCount v1.4.5 (Liao et al., 2014) or htseq-count for strand-specific quantification. Differential expression analysis for different gene models (ENSEMBL gene models, custom gene models for intron quantification and published lncRNA models) was performed using DESeq2 (Anders et al., 2012; Love et al., 2014). Enriched genes were selected at a p-value of 0.05 after a Benjamini-Hochberg adjustment for multiple testing. Gene set enrichment analysis was performed using the Piano package (Varemo et al., 2013) and the Panther pathway classification downloaded for zv9 version of the genome (ftp://ftp.pantherdb.org/). Transcript levels were quantified in RPKM and FPKM, as previously described (Mortazavi et al., 2008). Genes expressed at FPKM$>$1 were deemed expressed. Data generated in this study submitted to GEO (GSE89670) and are also available via Daniocode consortium (http://danio-code.zfin.org/daniocode/).

Genome-wide analysis of polyA-enriched neural crest nuclear transcripts validates the Biotagging approach: Previous studies using the INTACT system employed polyA-based enrichment of RNA, thus harvesting the spliced portion of the nuclear transcriptome. To cross-validate our approach, we applied similar analyses to the nuclear RNA pool isolated from 24hpf neural crest cells, biotagged by crossing the *Tg(sox10: BirA-2A-membCherry)*^{ct706a} line (ncBirA) with the Avi-RanGAP effector line (nucAvi(*bact*)). RNA-seq libraries were prepared from polyA-selected nuclear transcripts, sequenced and analyzed. Differential expression analysis comparing polyA-selected nuclear neural crest to the whole embryo transcriptomes at 24hpf (Armant et al., 2013), identified 6580 differentially expressed genes (p<0.05), with 2918 genes significantly enriched and 3662 decreased in the *sox10* nuclear samples (Fig.S4A). Biological replicates of *sox10* biotagged nuclei samples were strikingly similar to each other for both enriched (Fig.S4B) and decreased genes (Fig.S4C) as shown by heat map representations of their gene expression levels and by scatter plot comparison of complete datasets indicating our purification and library production approach are highly reproducible (Fig.S3D). We found that 209 genes out of the 236, reported in Zfin as expressed in neural crest cells by 24hpf (Bradford et al., 2011), were expressed in the nuclear samples at 2 FPKMs or higher (Fig.S4E,F). Gene set enrichment (GSE) analysis revealed the presence of neural crest-relevant pathways implicated in the formation of neural crest derivatives, such as Wnt, PDGF, TGFβ and Notch (Fig.S4D, red nodes). In particular, the largest node from the GSE analysis consisted of 294 Wnt pathway genes, in line with previous evidence for its major involvement in migratory crest (Dorsky et al., 1998) and its primary role in differentiation of pigment and sensory neuron lineages (Pavan and Raible, 2012). The reduced, but not absent, representation of general metabolic pathways such as the TCA cycle, *de novo* purine biosynthesis and glycolysis (Fig.S4D) confirmed the value of profiling of small, specifically defined cell populations. Notably, we found a significantly decreased representation of genes involved in neuronal differentiation (axon guidance, opioid prodynorphin and GABA-B receptor II signalling) (Fig.S4D, blue nodes).As neural cell-types are intimately mingled with migratory neural crest cells, this reduction indicates that we can cleanly dissociate targeted and non-targeted cell-types by our purification protocol.

Intron quantifications: Starting from Ensembl gene models, we incorporated intron positions in a custom GTF file and quantified total read count for all introns and exons of a gene model, respectively. We reduced intron positions by 10% of total intron length from 5' and 3' ends to account for mis-annotated splice sites. Only introns located within genes whose exonic sequence wasexpressed at $>$1 FPKM in nuclear samples were selected for analysis. Moreover, genes whose introns contain another gene or transcript were excluded. After differential expression, analysis comparing nuclear and polysomal beta-actin samples at 16ss, genome wide additive expression profile of all differentially enriched introns larger than 30 kb were obtained using ngsplot (Shen et al., 2014).

Global transcriptional patterns in multiple datasets: Global distribution profiles and genebody plots were obtained using deepTools package profiler and heatmapper tools (Ramirez et al., 2014).

Clustering of transcriptional patterns using *k*-means algorithm: To characterise bidirectional transcription at active gene promoters in nuclear *sox10* datasets, we employed *k*-means clustering using seqMINER (Ye et al.,

2011). The clustering was performed using linear normalization, k=10 clusters and a window of $\pm$1.5 kb from the active TSS set (ATAC_TSS). This procedure was applied to read associated with each strand of the *sox10* nuclear, *bactin* nuclear and *sox10* polysomal RNA-seq datasets. Corresponding heat maps were generated using seqMINER. *K*-means clustering was applied in a similar strand-specific fashion, using linear enrichment parameter, to identify and classify enhancers active specifically in neural crest cell nuclei, but using the active enhancer set (ATAC_enhancer). For *cis*-regulatory signature analysis, clusters 1 and 2 enriched in bidirectionally transcribed elements were selected and used to quantify specific bidirectional *cis*-regulatory modules transcription in neural crest (*sox10* samples).

Scatterplot quantification of transcribed elements: Count tables for each *k*-cluster identified from ATAC_TSS and ATAC_enhancer datasets were generated using subread featureCount (Liao et al., 2014) for every two samples being compared (e.g. sample A and B). To generate dot plots in R, counts were normalized to reflect difference in sequencing depth between the two samples by dividing each count value of the sample with a greater number of reads (e.g. sample A) to a "downsampling factor" (no. of sample A reads/no. of sample B reads). To plot and obtain Pearson correlation coefficient values of ATAC_enhancer *k*-clusters whose transcription is more enriched in *sox10* nuclear datasets, non-strand specific count tables for all clusters were generated and normalized for *sox10* nuclear and *bactin* nuclear samples. Cluster 4 containing non-transcribed elements was excluded. For ATAC_TSS *k*-clusters, strand-specific count tables were generated and normalized to highlight *k*-clusters whose transcription is more enriched in *sox10* nuclear datasets, according to negative or positive strand. To quantify the proportion of TSSs that are bidirectionally transcribed in *sox10* nuclear compared to *bactin* nuclear, the number of elements that have a normalized count value of >0 on both strands were determined. R package dplyr "anti_join" function was used to identify bidirectional elements within a given cluster that are found only in *sox10* nuclear or *bactin* nuclear.

Ranking neural crest (NC)-specific *cis*-regulatory modules (CRMs): To rank specific NC *cis*-regulatory modules from cluster 1 and 2, *sox10* nuclear replicates and *bactin* replicate BAM files were merged and read counts for 11,655 elements ATAC_enhancer features were obtained using subread featureCount (Liao et al., 2014) and features expressed at FPKM$>$1 in *bactin* and *sox10* samples were considered. The ratio of FPKM values between *sox10* nuclear and *bactin* nuclear is termed fold-change (FC) for 11,655 non-null elements from Clusters 1 and 2. These CRMs were ranked according to FC value. CRMs were assigned to the proximal genes targets based on distance in whose regulation they are putatively involved. Using bedtools, a total of 11,655 CRMs were assigned to 4,767 genes. To quantify the additive effect of multiple enhancers on a single locus, we computed the Additive Fold Change (AFC), as a sum of FCs of all active CRMs assigned to a given locus. We ranked all loci according to their AFC values to identify a critical set of highly regulated loci, defined as genes whose relative specific enhancer transcription (measured by AFC) falls beyond this inflexion point. We assumed that these genes constitute neural crest transcriptional signature at 16-18ss.

GREAT Analysis: To annotate and assign biological significance to the identified CRMs with intermediate to high FC value ($1<FC<5$), bidirectionally transcribed specifically in the neural crest nuclei, we applied the GREAT tool (**G**enomic **R**egions **E**nrichment of **A**nnotations **T**ool) (Hiller et al., 2013; McLean et al., 2010), which allows prediction of functional *cis*-regulatory regions by analysing the annotations of the genes lying proximal to them. The analysis was performed using GREAT default parameters: basal regulatory region extending 5 kb upstream and 1 kb downstream from TSS, with maximal 1 Mb extension and using the whole genome as a reference. We retained results obtained with a with a $p<0.001$ according to both binomial and hypergeometric tests.

Repeated elements quantification: Repeat analysis was carried out as previously described (Goke et al., 2015). Briefly, coordinates for repeats were downloaded from UCSC Genome Browser (v. Jan 19 2011) for danRer7 (Zv9). Read count for repeat positions were obtained using featureCount and FPKM values calculated accordingly (Liao et al., 2014). Differential expression was carried out using the rank product non-parametric method (Breitling et al., 2004) using the R package RankProd, with $pfp<0.05$ (Hong et al., 2006).

References

G. Abe, M. L. Suster, and K. Kawakami. Tol2-mediated transgenesis, gene trapping, enhancer trapping, and the Gal4-UAS system. *Methods in Cell Biology*, 104:23–49, 2011.

S. Anders, A. Reyes, and W. Huber. Detecting differential usage of exons from RNA-seq data. *Genome Research*, 22:2008–2017, 2012.

A. Apolloni, I. A. Prior, M. Lindsay, R. G. Parton, and J. F. Hancock. H-ras but not K-ras traffics to the plasma membrane through the exocytic pathway. *Molecular and Cellular Biology*, 20:2475–2487, 2000.

O. Armant, M. Marz, R. Schmidt, M. Ferg, N. Diotel, R. Ertzer, J. C. Bryne, L. Yang, I. Baader, M. Reischl, and et al. Genome-wide, whole mount *in situ* analysis of transcriptional regulators in zebrafish embryos. *Developmental Biology*, 380:351–362, 2013.

Y. Bradford, T. Conlin, N. Dunn, D. Fashena, K. Frazer, D. G. Howe, J. Knight, P. Mani, R. Martin, S. A. Moxon, and et al. ZFIN: enhancements and updates to the zebrafish model organism database. *Nucleic Acids Research*, 39:D822–829, 2011.

R. Breitling, P. Armengaud, A. Amtmann, and P. Herzyk. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS letters*, 573:83–92, 2004.

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

J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf. ATAC-seq: A method for assaying chromatin accessibility genome-wide. In F. M. A. et al., editor, *Current Protocols in Molecular Biology 109*, pages 21–29, 2015.

J. Bussmann and S. Schulte-Merker. Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development*, 138:4327–4332, 2011.

T. J. Carney, K. A. Dutton, E. Greenhill, M. Delfino-Machin, P. Dufourcq, P. Blader, and R. N. Kelsh. A direct role for Sox10 in specification of neural crest-derived sensory neurons. *Development*, 133:4619–4630, 2006.

R. B. Deal and S. Henikoff. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nature Protocols*, 6:56–68, 2011.

A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29:15–21, 2013.

R. I. Dorsky, R. T. Moon, and D. W. Raible. Control of neural crest cell fate by the Wnt signalling pathway. *Nature*, 396:370–373, 1998.

S. Driegen, R. Ferreira, A. van Zon, J. Strouboulis, M. Jaegle, F. Grosveld, S. Philipsen, and D. Meijer. A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Research*, 14:477–482, 2005.

J. Goke, X. Lu, Y. S. Chan, H. H. Ng, L. H. Ly, F. Sachs, and I. Szczerbinska. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell*, 16:135–141, 2015.

M. Heiman, R. Kulicke, R. J. Fenster, P. Greengard, and N. Heintz. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nature Protocols*, 9:1282–1291, 2014.

S. Higashijima, H. Okamoto, N. Ueno, Y. Hotta, and G. Eguchi. High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. *Developmental Biology*, 192:289–299, 1997.

M. Hiller, S. Agarwal, J. H. Notwell, R. Parikh, H. Guturu, A. M. Wenger, and G. Bejerano. Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: application to zebrafish. *Nucleic Acids Research*, 41(e151), 2013.

F. Hong, R. Breitling, C. W. McEntee, B. S. Wittner, J. L. Nemhauser, and J. Chory. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*, 22:2825–2827, 2006.

C. J. Huang, C. T. Tu, C. D. Hsiao, F. J. Hsieh, and H. J. Tsai. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Developmental Dynamics*, 228:30–40, 2003.

S. W. Jin, D. Beis, T. Mitchell, J. N. Chen, and D. Y. Stainier. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development*, 132:5199–5209, 2005.

N. A. Joshi and J. N. Fass. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). 2011. URL `https://githubcom/najoshi/sickle`.

K. Kawakami. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. *Methods in Cell Biology*, 77:201–222, 2004.

Y. Liao, G. K. Smyth, and W. Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30:923–930, 2014.

K. J. Livak and T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, 25:402–408, 2001.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15:550, 2014.

R. Mahajan, C. Delphin, T. Guan, L. Gerace, and F. Melchior. A small ubiquitin-related polypeptide involved in targeting ranGAP1 to nuclear pore complex protein ranBP2. *Cell*, 88:97–107, 1997.

C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, and G. Bejerano. GREAT improves functional interpretation of *cis*-regulatory regions. *Nature Biotechnology*, 28:495–501, 2010.

A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods*, 5:621–628, 2008.

C. Mosimann, C. K. Kaufman, P. Li, E. K. Pugach, O. J. Tamplin, and L. I. Zon. Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development*, 138:169–177, 2011.

M. K. Nyholm, S. F. Wu, R. I. Dorsky, and Y. Grinblat. The zebrafish *zic2a-zic5* gene pair acts downstream of canonical Wnt signaling to control cell proliferation in the developing tectum. *Development*, 134:735–746, 2007.

W. J. Pavan and D. W. Raible. Specification of neural crest into sensory neuron and melanocyte lineages. *Developmental Biology*, 366:55–63, 2012.

F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, and T. Manke. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, 42:W187–191, 2014.

A. Rose and I. Meier. A domain unique to plant ranGAP is responsible for its targeting to the plant nuclear rim. *Proceedings of the National Academy of Sciences of the United States of America*, 98:15377–15382, 2001.

L. Shen, N. Shao, X. Liu, and E. Nestler. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics*, 15:284, 2014.

M. L. Suster, K. Sumiyama, and K. Kawakami. Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics*, 10:477, 2009.

R. C. Tryon, N. Pisat, S. L. Johnson, and J. D. Dougherty. Development of Translating Ribosome Affinity Purification for zebrafish. *Genesis*, 51:187–192, 2013.

L. Varemo, J. Nielsen, and I. Nookaew. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, 41: 4378–4391, 2013.

M. Westerfield. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. University of Oregon Press, 4th edition, 2000.

T. Ye, A. R. Krebs, M. A. Choukrallah, C. Keime, F. Plewniak, I. Davidson, and L. Tora. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Research*, 39:e35, 2011.

D. Yu, H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, and D. L. Court. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 97:5978–5983, 2000.

Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and et al. Model-based analysis of ChIP-seq (MACS). *Genome Biology*, 9:R137, 2008.