Title: ATAC-seq of Low-Input and Cryopreserved Primordial Germ Cells Reveals Functional Enhancers

Akane Kawaguchi[1,2,8†], Mao Igari[3,8] , Yasuto Murayama[2,4], Hiroko Iikawa[3], Mika Sakamoto[5], Yasukazu Nakamura[2,5], Shigehiro Kuraku[1,2,], Keisuke Ishihara[6] and Daisuke Saito[7†]

Affiliations:

[1] Molecular Life History Laboratory, National Institute of Genetics, Mishima, Japan

[2] Department of Genetics, Sokendai (Graduate University for Advanced Studies), Mishima, Japan

[3] Graduate School of Systems Life Sciences, Kyushu University, Fukuoka, Japan

[4] Chromosome Biochemistry Laboratory, National Institute of Genetics, Mishima, Japan

[5] Genome informatics Laboratory, National Institute of Genetics, Mishima, Japan

[6] Department of Computational and Systems Biology, Pittsburgh, PA, USA

[7] Graduate School of Science, Kyushu University, Fukuoka, Japan

[8] These authors equality contributed to this study

[†] Correspondence: akane.kawaguchi@nig.ac.jp. Yata 1111, Mishima, Shizuoka 411-8540, Japan; saito.daisuke.036@m.kyushu-u.ac.jp. Moto-oka, 744, Nishi-ku, Fukuoka, Fukuoka 819-0395, Japan

Key words: PGCs enhancers, ATAC-seq, Small input and cryopreserved materials

Abstract

Cell fate specification during development is orchestrated by dynamic changes in gene expression, driven by the accessibility of cis-regulatory elements. To profile such functional regulatory elements at the genome-wide level, bulk ATAC-seq is a powerful method. Yet, its application has still been technically limited in rare cell types. Here, we optimized and titrated the ATAC-seq protocol for cultured chicken primordial germ cells (PGCs) and obtained high-quality chromatin accessibility profiles from as few as 200 cells, even after cryopreservation. Integrative analysis of our ATAC-seq with published RNA-seq and tissue ATAC-seq datasets identified over 10,000 PGC-specific accessible chromatin regions (ACRs), many absent from somatic tissues. Reporter assays demonstrated activity of selected enhancers in cultured PGCs, while transcriptome integration revealed preferential association of PGC-specific ACRs with genes highly expressed at embryonic day 2.5. *In vivo* transplantation confirmed that these enhancers are active in an early-stage-specific manner, but become silenced upon gonadal settlement. In conclusion, this streamlined method provides a practical and cost-effective approach to the epigenomic analysis of scarce cell populations, facilitating future investigations in developmental biology, germline research, and beyond.

Introduction

Cell fate specification and differentiation during development are governed by tightly regulated, dynamic changes in gene expression. Increasing evidence indicates that such transcriptional shifts are orchestrated through changes in chromatin accessibility at cis-regulatory elements. Therefore, capturing the chromatin landscape in a spatiotemporal manner is essential to understand and manipulate lineage decisions.

ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) enables genome-wide profiling of accessible chromatin with high resolution using far fewer input materials than other conventional profiling methods, such as the DNase I hypersensitivity assay (Ling and Waxman, 2013), allowing profiling from as few as 500 freshly isolated cells (Buenrostro *et al.*, 2013). The Omni-ATAC-seq protocol has further improved signal-to-noise ratios through simplified procedures and the use of additional solvents, and has successfully reduced mitochondrial DNA (mtDNA) contamination (Corces *et al.*, 2017). However, applying this method to rare embryonic cells remains challenging, as viable cell isolation is technically demanding and often yields limited material. Cells or tissue cryopreservation could address these issues (Fujiwara *et al.*, 2019; Halstead *et al.*, 2020; Peng *et al.*, 2021), but robust ATAC-seq protocols for frozen low-input samples are still underdeveloped. Although single-cell ATAC-seq offers an alternative (Buenrostro *et al.*, 2015), its cost and complexity limit its general accessibility (Pott and Lieb, 2015).

Primordial germ cells (PGCs) are an inherently small cell population in developing embryos. They originate outside the embryo proper and migrate dynamically to the gonads, where they undergo sex-specific programs of mitosis, meiosis, stem cell maintenance, and eventual differentiation into gametes (Wear, McPike and Watanabe, 2016; Yoshida, 2020; Barton *et al.*, 2024). Thus, changes in chromatin state within the germline are of broad biological interest, and substantial knowledge has been accumulated regarding DNA and histone modifications (Seki *et al.*, 2007; Nagano *et al.*, 2022). However, compared to these epigenetic marks, much less is known about the open–closed chromatin status of PGCs as revealed by ATAC-seq. Furthermore, because PGCs are known to have relatively high mitochondrial content (Motta *et al.*, 2000; Deniz Koç and Yüce, 2012), they provide an excellent benchmark model for evaluating and optimizing ATAC-seq protocols with respect to mitochondrial genome contamination—a common technical challenge in this assay.

In this study, we leveraged the unique advantages of cultured chicken PGCs and the chicken embryo model. Cultured PGCs can be expanded *in vitro* while preserving germline potential and an *in vivo*–competent migratory program, homing to the gonads via the

endogenous route after transplantation (van de Lavoir *et al.*, 2006; Whyte *et al.*, 2015; Morimoto and Saito, 2025). These properties enable rigorous benchmarking of ATAC-seq protocols and functional testing of cis-regulatory elements within their physiological context.

Here, we first optimized the Omni-ATAC-seq workflow for low-input and cryopreserved chicken PGCs, demonstrating that high-confidence chromatin accessibility profiles can be obtained from as few as 200 cells, even after freezing. This optimization also revealed sequence biases of Tn5 insertions in the chicken genome, underscoring the need for proper background controls. We then established an *in vitro* PGC differentiation system within this study, enabling a reporter assay to functionally evaluate PGC-specific and stage-specific enhancers identified from ATAC-seq data. Using transplantation into chicken embryos, we further demonstrated that selected enhancers are active in migrating PGCs *in vivo*. Together, these results show that our approach enables robust identification of functional cis-regulatory elements from minimal starting material, providing a powerful framework for studying chromatin regulation not only in germline biology but also in other developmentally dynamic or clinically relevant rare cell types.

Results and discussions

Cryopreservation impacts mtDNA content but not overall library quality

To evaluate the impact of cell number and cryopreservation on Omni-ATAC-seq quality, we prepared cultured chicken PGC samples with four different input sizes (200, 1,000, 5,000, and 25,000 cells), using either freshly harvested or cryopreserved cells stored at –80°C for one week (Fig. 1A). The protocol followed the original Omni-ATAC-seq method (Corces *et al.*, 2017), which employs a combination of detergents (Tween-20, NP-40, Digitonin) for effective membrane permeabilization, and PBS addition during tagmentation to enhance the signal-to-noise ratio by reducing mtDNA contamination. Cryopreserved samples exhibited high cell viability (>90%) based on trypan blue exclusion, and nuclear morphology remained comparable to that of live cells, as assessed by nuclear staining (Fig. 1B, C). These results indicate that both fresh and cryopreserved cells were suitable as starting materials. As a negative control, we included a sample of purified genomic DNA subjected to tagmentation. After library construction and sequencing, we processed fastq files through standard pre-processing steps, including adapter trimming, filtering, genome alignment, and duplicate removal. Cryopreserved samples exhibited higher duplication rates (19–27%) compared to live cells (9–14%) (Fig. 1D, Supplementary Table 1), suggesting that freeze–thaw steps may increase redundant amplification.

We next evaluated three key quality metrics of ATAC-seq data: (i) mtDNA (mtDNA) mapping rate, (ii) unmapped read rate, and (iii) FRiP (Fraction of Reads in Peaks) score. The mtDNA read contamination was notably higher in cryopreserved samples (50–60%) than in fresh samples (5–15%) (Fig. 1E, F), suggesting that cryopreservation may reduce the efficiency of mitochondrial removal during lysis. This observation is consistent with earlier reports on ATAC-seq protocols lacking detergent-based optimization (Fujiwara *et al.*, 2019). The proportion of unmapped reads was highest in 200-cell input samples (10–15%), regardless of cryopreservation, while samples with ≥1,000 cells showed stable mapping rates (~5–7%) (Fig. 1G). This suggests that input cell number, rather than freeze–thaw condition, is the major determinant of read mappability. FRiP scores, a standard measure of ATAC-seq signal quality, exceeded 0.2 in all libraries (Fig. 1H), thereby meeting ENCODE guidelines (https://www.encodeproject.org/atac-seq/). While FRiP scores showed a weak correlation with input cell number, they were not significantly influenced by cryopreservation. In the end, after removing duplicate reads and mtDNA-mapped reads, the consistent correlation signatures obtained from the BAM files do not exhibit large differences across conditions (Fig. 1H). Taken together, these results demonstrate that while cryopreservation increases mtDNA contamination and duplication rates, it does not substantially compromise library quality—even in PGC samples, which are known to contain relatively high levels of mtDNA (Motta *et al.*, 2000; Deniz Koç and Yüce, 2012). Importantly, reproducible and high-quality chromatin accessibility profiles can be obtained from as few as 200 cells, even after cryopreservation.

Peak Number and Sensitivity in Low-Input Cryopreserved Samples

We next examined how variations in input cell number and cryopreservation influence the number and quality of peaks detected by ATAC-seq. To begin, peak calling was performed using MACS3 without normalization (–broad, $q < 0.05$). In live-cell libraries, the number of identified peaks ranged from 31,018 to 48,786. In contrast, cryopreserved libraries yielded significantly fewer peaks, between 22,737 and 38,558, with a statistically significant reduction compared to live samples ($P < 0.05$; Fig. 2A, Supplementary Table 1). Peak counts showed a modest upward trend with increasing input cell numbers in both conditions ($P < 0.05$; Fig. 2B).

To assess whether these differences stemmed from sequencing depth, we next normalized all libraries to 10 million valid reads using computed scaling factors (Supplementary Table 2). After reading subsampling and re-calling peaks with identical MACS3 parameters, peak numbers across all libraries converged, with no statistically significant differences between groups ($P > 0.05$), except for the gDNA-derived negative

142 control (Fig. 2C). It suggests that the reduced peak counts in cryopreserved samples primarily
143 reflect lower sequencing depth rather than intrinsic biological differences.
144 We further assessed peak reproducibility by annotating all ATAC-seq peaks to genomic
145 features (promoter, exon, intron, intergenic, and transcription termination site regions) and
146 computing overlap rates. Promoter-associated peaks exhibited the highest reproducibility
147 across all samples, with overlap rates of 70–85%, regardless of preservation method or cell
148 number (Fig. 2D). Peaks located in intergenic, intronic, exonic, and TTS regions were more
149 variable (~65% overlap), likely due to the weaker and more heterogeneous nature of regulatory
150 elements in these regions (Fig. 2E, Supplementary Fig. 1A–C).
151 To evaluate signal robustness at key regulatory loci, we visualized the detected ACRs
152 at the *Nanog* promoter, a well-characterized marker of PGC identity and pluripotency (Choi *et*
153 *al.*, 2021). Consistent peaks were observed across all conditions, including cryopreserved 200-
154 cell samples (Fig. 2F). To determine whether increased sequencing depth could restore
155 detection of these weak peaks, we merged the reads in which after removing mtDNA and PCR-
156 duplicated reads across the same conditions considered as the biological replicates (Fig. 2G).
157 The merged datasets effectively recovered some of the weakly accessible regions, suggesting
158 that deeper sequencing can mitigate the sensitivity loss associated with cryopreservation.
159 Lastly, we investigated potential sequence bias introduced by the Tn5 transposase using
160 the gDNA control library, which can be utilized as a "blacklist." Peak calling and motif
161 enrichment analysis of this control yielded 1,023 peaks, disproportionately enriched in
162 intergenic regions containing the telomeric (TTAGGG)n hexamer, likely corresponding to
163 interstitial telomeric sequences (ITSs) (Alejandro D. Bolzán, 2017) (Fig. 2H, Supplementary
164 Fig. 1D–J). The chicken genome is characterized by a distinctive karyotype comprising macro-
165 and micro- chromosomes, and ITSs have been recognized as genomic signatures shaped by
166 meiotic recombination and chromosomal rearrangements throughout its evolutionary history
167 (I. Nadia et al., 2002). In light of such species-specific genomic features, assessing sequence
168 bias using a gDNA negative control provides a valuable means to account for these intrinsic
169 sequence properties when interpreting ATAC-seq results. This pattern most likely reflects
170 inherent sequence preferences of Tn5 rather than genuine chromatin accessibility, underscoring
171 the importance of distinguishing biological signals from technical artifacts in ATAC-seq
172 datasets.

173 These results demonstrate that Omni-ATAC-seq applied to cryopreserved, low-input
174 samples results in chromatin accessibility profiles largely comparable to those of live-cell
175 samples, particularly in promoter regions. Nonetheless, higher sequencing depth or replicate
176 integration may be required to capture low-abundance or weakly accessible regulatory
177 elements in frozen material. Importantly, we identified technical artifacts occurring by intrinsic
178 sequence preferences of the Tn5, which appear to be linked to species-specific chromosomal
179 features such as ITSs in the chicken genome. By incorporating a gDNA-based negative control
180 to detect and remove such artifacts, we provide a strategy to more accurately distinguish
181 genuine biological signals from sequence bias. This approach is especially relevant when
182 working with species that possess distinctive genomic architectures, and it expands the
183 applicability of Omni-ATAC-seq to a wider range of rare or previously inaccessible cell
184 populations in developmental and comparative genomics studies.
185
186 **Identification of PGC-Specific Accessible Chromatin Regions**
187 We next ask whether it is possible to identify functional ACRs either PGC-specific or
188 developmentally stage-specific, by integrative analysis with publicly available ATAC-seq
189 datasets (Fig. 3A). To identify PGC-specific regulatory elements, instead of differential peak
190 analysis we applied a binary comparison strategy, subtracting peaks shared with multiple
191 somatic tissues to remove commonly accessible ACRs (Fig. 3A). We merged our datasets by
192 input type (live or cryopreserved), each yielding 195–247 million mapped reads, so that the
193 library volume became comparable to other tissue libraries (Supplementary Table 2). Peak
194 calling produced 55,805 overlapping peaks between live and cryopreserved PGCs (Fig. 3B).
195 Live-specific peaks (~30,000) were far more numerous than cryopreserved-specific peaks
196 (~5,759), with shared peaks mainly in promoter regions and live-specific peaks enriched in
197 intronic and intergenic regions, that is consistent with the previous results at figure 2 (Fig. 3C,
198 Supplementary Table 3).
199 To define PGC-specific ACRs, we compared peaks shared by live and cryopreserved
200 samples with a binary matrix from 30 public ATAC-seq datasets of 8 somatic tissues and
201 embryonic organs (Fig. 3A). This comparison resulted in 10,550 PGC-specific ACRs (Fig. 3D,
202 Supplementary Table 4), distributed across promoters (10.62%), exons (13.79%), introns
203 (49.17%), intergenic (18.36%), and TTSs (8.03%). In contrast, 16,969 ACRs shared with
204 somatic tissues were enriched in promoters (16.56%) and exons (39.93%), with no intergenic
205 contribution. Overall, ACR numbers in PGCs were comparable to somatic tissues. These
206 findings suggest PGC-specific expression is largely controlled by enhancers in non-coding

207 regions rather than promoter-proximal sites, and intergenic ACRs at housekeeping genes are
208 not uniformly accessible across tissues or cell types.
209 To identify functional PGC-specific enhancers, we selected 16 representative candidate
210 regions from the >10,000 ACRs uniquely identified in PGCs by ATAC-seq (Supplementary
211 Table 5). First, we filtered for peaks that were exclusively detected in PGCs and absent in
212 somatic tissue samples. Next, we selected peaks that were consistently detected across various
213 PGC input conditions (ranging from low to high input, including both live and frozen samples),
214 aiming to evaluate whether functional enhancers could be robustly captured under suboptimal
215 or practical experimental conditions. The selected candidates spanned a broad range of
216 chromatin accessibility values (as determined by q-values and fold changes from peak calling)
217 and were located within 2,500 bp of annotated gene bodies. Because ACRs can vary in size—
218 from nucleosome-free regions (<100 bp) to regions spanning two or three nucleosomes—we
219 excluded regions smaller than 50 bp and larger than 550 bp (approximately the width of >3
220 nucleosomes) to avoid excessively broad peaks that may reflect open chromatin regions
221 without clear enhancer activity. This filtering strategy allowed us to include a representative
222 range of accessibility profiles while minimizing potential artifacts. Ultimately, candidate
223 enhancers were prioritized not only by accessibility rank but also by genomic context and
224 technical feasibility, enabling a focused evaluation of their regulatory potential in PGCs.
225
226 **Functional Validation of Candidate PGC-Specific ACRs in Cultured Cells**
227 To experimentally validate the enhancer activity of the selected PGC-specific ACRs, we
228 performed a fluorescence-based reporter assay in cultured PGCs. Reporter constructs
229 containing each candidate ACR upstream of a minimal tk promoter driving EGFP (Uchikawa
230 *et al.*, 2003) were co-transfected with a CAGGS-mCherry control plasmid to monitor
231 transfection efficiency (Fig. 3E). Of the 16 candidate ACRs prioritized as described above, 14
232 genomic regions were successfully amplified by genomic PCR and cloned into the reporter
233 vector. Two candidates could not be amplified and were excluded from further analysis
234 (Supplementary Table 5).
235 The tk-EGFP vector lacking any upstream enhancer served as a negative control and
236 produced only minimal EGFP signal (Fig. 3F). As a positive control, we used the Nanog
237 promoter, a well-characterized PGC-expressed regulatory element, which drove robust EGFP
238 expression (Supplementary Fig. 4), validating the sensitivity of our assay. Among the 14 tested
239 constructs, the majority showed EGFP fluorescence above the basal level (#1, #3, #4, #5, #6,

240 #8, #9, #12, #13, #14, and #15), and several candidates exhibited strong and consistent activity
241 comparable to the Nanog promoter (Supplementary Fig. 2). The co-expression of mCherry in
242 all samples confirmed successful transfection and enabled normalization of EGFP signal
243 intensities, excluding variation in transfection efficiency as a confounding factor. The activities
244 of the enhancer candidates were in strong agreement with the observation that the
245 corresponding peak loci exhibited robust and representative signals across PGC datasets
246 covering both low to high input amounts and live or frozen samples (Supplementary Fig. 3).
247 These results provide direct experimental evidence that a substantial fraction of the ATAC-
248 defined, PGC-specific ACRs, including those detectable even in frozen 200-cell datasets,
249 function as *bona fide* enhancers in their native germ cell context.
250
251 **Loss of Enhancer Activity upon PGC Differentiation**
252 To assess whether the enhancer activity of PGC-specific ACRs depends on the undifferentiated
253 state, we examined EGFP expression following *in vitro* differentiation of PGCs. In the course
254 of this study, we found that cultured PGCs spontaneously differentiate when transferred to a
255 medium lacking Activin A (Fig. 3E). Under these conditions, PGCs that typically remain in
256 suspension began to adhere to the culture dish and exhibited morphological changes
257 characteristic of differentiation. Notably, the promoter activity of established germline genes
258 such as *Nanog* and *DDX4* was markedly reduced in this condition, as evidenced by diminished
259 EGFP and mCherry fluorescence in reporter assays (Supplementary Fig. 4). To our knowledge,
260 this is the first report to define a culture condition that reliably induces spontaneous
261 differentiation of chicken PGCs *in vitro*.
262 Using this differentiation system, we cultured PGCs transfected with enhancer
263 reporter constructs in the absence of Activin A and monitored EGFP fluorescence before and
264 after differentiation. Strikingly, EGFP signals driven by all enhancer candidates—including
265 those with strong activity in undifferentiated PGCs such as #3, #4, #5, #6, #8, and #12—were
266 abolished following differentiation (Fig. 3F-H; Supplementary Fig. 5). In contrast, mCherry
267 fluorescence from the co-transfected control plasmid remained detectable in both
268 undifferentiated and differentiated cells, confirming that the observed loss of EGFP was not
269 due to differences in transfection efficiency or cell viability.
270 These findings indicate that the enhancer activity of the tested ACRs is tightly linked
271 to the undifferentiated state of PGCs and becomes silenced upon differentiation. Together,

these data support the interpretation that the tested ACRs represent germline-competent, state-dependent regulatory elements.

PGC-Specific Enhancer Activity is Restricted to Early Development

Based on the identification of PGC-specific enhancers using ATAC-seq (Fig. 3), we next sought to refine the selection of candidate elements by incorporating stage-specific transcriptomic data from endogenous chicken PGCs (Fig. 4A). To determine whether stage-specific ACRs could also be detected, we integrated our ATAC-seq data with RNA-seq datasets from Ichikawa *et al.* (Ichikawa et al. 2022) differential gene expression analysis (DESeq2; baseMean > 10, $q<0.01$, $|\log_2 FC|>2$) between migrating PGCs at embryonic day 2.5 (E2.5) and gonad-resident PGCs at E6 identified 525 genes significantly upregulated at E2.5 and 270 genes upregulated at E6 (Fig. 4B, Supplementary Table 6). We then examined whether PGC-specific ACRs were located near these differentially expressed genes. Among the 21,892 filtered peaks, 434 were associated with genes highly expressed at E2.5, and 321 genes contained PGC-specific ACRs within 2,500 bp of their gene bodies (Fig. 4C). In contrast, loci of genes highly expressed at E6—such as *LDHA*—tended to show small, constitutive peaks shared across all tissues, lacking PGC-specific ACRs (Fig. 4C). This pattern likely reflects the fact that the cultured PGCs used for ATAC-seq were originally derived from E2.5 embryos.

To assess stage-specific activity during PGC development, we performed *in vivo* transplantation of selected three candidate enhancers (#3, #4, #5) that had shown strong enhancer activity *in vitro* (Fig. 4D). Cultured PGCs transfected with the enhancer reporter constructs and a CAGGS-mCherry control plasmid were injected into the bloodstream of stage E2.5 chicken embryos. Transplanted embryos were collected and analyzed at two developmental time points: E3.5—when PGCs are migrating through the dorsal mesentery or have just arrived at the gonadal primordium—and E6.5, by which time PGCs have stably settled within the developing gonads. At E3.5, all tested enhancers exhibited detectable EGFP fluorescence, indicating that their activity is maintained during the migratory phase (Fig. 4 E–G; Supplementary Fig. 6). In contrast, at E6.5, EGFP signals were consistently absent, while mCherry fluorescence remained detectable, confirming successful transfection and survival of the transplanted PGCs. The peaks tested in these assays were not detected in any somatic tissue samples, yet showed highly robust accessibility in PGC ATAC-seq data. They were located near genes with high expression at E2 that markedly decreased by E6, providing consistent

305 evidence for their PGC specificity and stage-dependent activity (Fig. 4H-I). Notably, a *Nanog*
306 promoter construct continued to drive EGFP expression at both time points (Fig. 4E),
307 supporting the conclusion that the loss of enhancer activity was not due to global transcriptional
308 repression or cell death, but rather reflects stage-specific silencing of the tested PGC enhancers.
309 These results demonstrate that the activity of these enhancers is not only cell type–specific, but
310 also developmentally restricted—active during early migratory stages and silenced following
311 PGC colonization of the gonadal niche.

312 In summary, we establish that Omni-ATAC-seq can generate high-quality chromatin
313 accessibility profiles from as few as 200 cryopreserved chicken PGCs, enabling robust
314 identification of cis-regulatory elements even under low-input conditions. This methodological
315 advance removes a major technical barrier for epigenomic profiling of rare cell types and
316 facilitates sample storage, transport, and cross-laboratory collaboration. Beyond its technical
317 implications, our findings provide new opportunities to dissect the gene regulatory architecture
318 that underlies germline specification and developmental stage–specific transcriptional
319 programs in avian models. Given the unique experimental accessibility of chicken PGCs, this
320 approach will not only accelerate germline biology research but also open the door to
321 comparative and functional studies across diverse developmental systems.
322

323 **Materials and Methods**

324 **Animals, staging, and animal care**

325 Fertilized chicken (*Gallus gallus domesticus*, White leghorn) eggs were purchased from
326 Yamagishi poultry farm (Mie, Japan) and from Nagoya University through the National Bio-
327 Resource Project of the MEXT, Japan, respectively. Eggs were incubated at 38.5°C, and
328 embryos were staged by Hamburger and Hamilton's stage (Hamburger and Hamilton, 1951).
329 All animal experiments were performed with the approval of the Institutional Animal Care and
330 Use Committees at Kyushu University under the allowance number A25-189-1.
331

332 **Establishment of chicken cultured PGCs**

333 Circulating PGCs, along with blood cells, were harvested from blood of HH15 chicken
334 embryos and were cultured on non-coated (non-adhesive) plates in calcium-free DMEM
335 (Gibco) diluted with water, supplemented with Fibroblast Growth Factor Chimera (Wako),
336 Activin A (APRO Science group), and chicken serum (Biowest) (FAcs medium) according to
337 previously described methods (Chen *et al.*, 2019); (Whyte *et al.*, 2015)). After one month of

338 expansion, PGCs were cryopreserved at -80°C in Bambanker (NIPPON Genetics) until further
339 use in experiments.
340
341 **PGC frozen stock and ATAC-seq library preparation**
342 ATAC-seq libraries from living and cryopreserved PGC were prepared based on the Omni-
343 ATAC seq method with modifications as recently published ((Corces *et al.*, 2017)
344 ; (Kawaguchi *et al.*, 2024)). Cultured PGCs were harvested, counted the cell density and
345 aliquoted to the desired number of cells (200, 1000, 5000 and 25000) in individual DNA low
346 binding tubes (Eppendolf, 0030108051). Cells were collected with centrifugation (350 G for 5
347 min at room temperature) (KUBOTA, 2410). For the cryopreservation, collected cells were
348 individually resuspended into 50 µl of Banbanker (Nihon Genetics, CS-02-001). Tubes were
349 slowly frozen using Mr.Frosty (Thermofisher, 5100-0001) and kept at -80°C until library
350 preparation. When ATAC-preparations were conducted, cryopreserved cells were thrown at
351 37°C for 5 min, and the cryopreserved samples were prepared together with living cultured
352 PGC samples. Concurrently, living cultured PGCs were harvested as described above. In the
353 cryopreserved and living cultured cells, 500 µl Lysis buffer 1 was added (10 mM Tris-HCl
354 pH7.5, 3 mM MgCl2, 2 mM NaCl, x50 Protease inhibitor), and gently mixed and harvested
355 (350 G for 5 min at 4°C) (KUBOTA, 2410) . The cell pellets were resuspended in 50 µl Lysis
356 buffer 2 (10 mM Tris-HCl pH7.5, 3 mM MgCl2, 2 mM NaCl, 0.1% Tween-20, 0.1% NP-40,
357 0.01% Digitonin, x50 Protease inhibitor) and incubated on ice for 15 min. 450 ul of Lysis buffer
358 3 (10 mM Tris-HCl pH7.5, 3 mM MgCl2, 2 mM NaCl, 0.1% Tween-20) was added to the cells
359 and mixed gently. Immediately, cells were spun down (350 G for 5 min at 4°C) and pellets
360 were resuspended with 25 µl of Tn5 Transposase reaction for the input of 200 and 1000 cells,
361 and 50 µl Tn5 Transposase reaction for the input on 5000 and 25000 cells (Transposase
362 reaction: 12.5 µl of the 2x Transposase buffer (20 mM Tris-HCl (pH7.5), 10 mM MgCl2, 20%
363 DMF), 16.5 µl of PBS, 0.5 µl of 10% Tween-20 (0.1% f. c.), 0.5 µl of 1% Digitonin (0.01% f.
364 c.) and 2.5 µl of assembled in-house Tn5 (0.5 µg of in-house Tn5 stock was diluted in 1:10)).
365 Tagmentation proceeded at 37°C in water-bath for 1 hour with occasional agitation. The
366 reaction was stopped by adding a 5 times volume of PB (QIAGEN, 19065) and vertexing for
367 30 seconds. Tn5 treated DNA was purified with MinElute PCR purification Kit (QIAGEN,
368 28004) and eluted with 20 µl of EB buffer. 10 µl of purified DNA solution was used for library
369 amplification, and final amplification cycles were defined with intermediate qPCR as described
370 in the original ATAC-seq method (Buenrostro et al., 2013). Nonetheless, all libraries were
371 amplified with 9 -16 total PCR cycles in the deserved number. Final libraries were amplified

372 with NEBNext® High-Fidelity 2XPCR Master Mix (NEB, M0541)) with Nextera Unique Dual
373 index (UDI) adapter plate A (illumina, #1000000002694). After PCR amplification, the
374 libraries were purified with magnetic beads for DNA isolation. Removing Adapters were
375 prepared following the manufacturing protocol. Sequencing was performed using Nova-
376 seq6000 PE150 at the third party.
377
378 **Computational analysis and data visualization**
379 **Reference genome and annotation versions**
380 Throughout the analysis, the reference genome assembly bGalGal1.mat.broiler.GRCg7b
381 (NCBI accession; GCF_016699485.2) and gene annotation from Gallus Enriched Gene
382 Annotation were used as reference gene model (Degalez *et al.*, 2024) (https://gega.sigenae.org).
383
384 **ATAC-seq processing and peak analysis**
385 Trimming of Nextera adapters and filtering of low-quality reads were performed using Fastp
386 (v0.23.2) with the parameters (-q 15 -n 10 -t 1 -T 1 -l 20) (Chen *et al.*, 2018). Bowtie2 (v2.5.4)
387 (Langmead *et al.*, 2019) was used for read mapping with the parameters (--no-mixed --no-
388 discordant -X 1000). SAMtools (v1.17) (Danecek *et al.*, 2021) was used to convert SAM files
389 to BAM files, and sorting and indexing were performed with default settings. Picard (v3.0.0)
390 (https://broadinstitute.github.io/picard/), was used for duplicate removal and fragment size
391 distribution analysis with the MarkDuplicates and CollectInsertSizeMetrics tools. Duplicate-
392 removed BAM files were further used to assess mtDNA and unmapped reads using SAMtools
393 with the parameters (view -f 4). Peaks were detected with MACS3 peakcall (v3.0.0a6), and
394 used the --broad option to account for the nature of accessible regions spanning multiple
395 nucleosomes with the parameters (-f BAMPE -q 0.05 -g 1053332251 --broad) (Zhang *et al.*,
396 2008). Peaks were ranked based on p-values and q-values obtained from MACS3. Peak
397 comparisons were performed using Bedtools (v2.31.0) (Quinlan and Hall, 2010), and
398 annotations were carried out using annotatePeaks.pl from HOMER (v4.9.1) (Heinz et al. 2010;
399 http://homer.ucsd.edu/homer/). To identify potential PGC-specific ACRs, we focused on
400 ACRs detected in both live and cryopreserved samples. For binary comparisons, peaks across
401 all samples were intersected using Bedtools (v2.31.0). To exclude gDNA-derived peaks from
402 the PGC ATAC-seq data, overlapping peaks were identified with Bedtools intersect with the
403 parameters (-u -f 0.5), and subsequently subtracted. For the data visualization, generating
404 bigwig files, correlation heatmaps were carried out with multiBamSummary, plotCorrelation
405 and bamCoverage by DeepTools (v3.1.2) with the parameters (--binSize 10 --normalizeUsing

406 RPKM --smoothLength 30 --effectiveGenomeSize 1053332251). The upsert plot was
407 generated by intervene (v0.6.5) (Quinlan and Hall, 2010; Khan and Mathelier, 2017). For motif
408 analysis, 1,023 peaks from gDNA tagmentation and 1,023 ACRs from merged live PGC
409 samples (randomly selected using Bedtools) were converted to FASTA format and analyzed
410 for enriched motifs using MEME (Bailey *et al.*, 2015).
411
412 **RNA-seq analysis**
413 RNA-seq datasets were obtained from publicly available data from K. Ichikawa et al. (2022),
414 consisting of four biological replicates each of *in vivo* PGCs at E2.5 and E6. The datasets were
415 obtained from the Gene Expression Omnibus under accession number GSE188689. Trimming
416 of Nextera adapters and filtering of low-quality reads were performed using Fastp (v0.23.2)
417 with the parameters (-q 15 -n 10 -t 1 -T 1 -l 20 -w 16) (Chen *et al.*, 2018). Mapping and
418 generation of count matrices were conducted using the STAR-RSEM pipeline with default
419 parameter (Li and Dewey, 2011; Dobin *et al.*, 2013). Differential expression analysis was
420 carried out using PyDESeq2 (cutoff: baseMean >10, q>0.01) (Muzellec *et al.*, 2023). For
421 visualization, fastq files processed by Fastp were mapped using STAR to generate
422 "Aligned.toTranscriptome.bam" files. Final bigWig files were generated using deepTools
423 (v3.1.2) with the parameters (--binSize 10 --normalizeUsing RPKM --smoothLength 30 --
424 effectiveGenomeSize 1053332251).
425
426 **Tn5 transposase purification and assembly**
427 The pTXB1-Tn5 expression plasmid (Addgene, #60240) was digested with XbaI and BamHI,
428 and fragment of Tn5-intein-CBD was subcloned into the pE21a (+) vector (NovoPro,
429 V011023) as described by Sato et al., (Sato *et al.*, 2019). Tn5 was expressed and purified as
430 described (Picelli *et al.*, 2014; Sato *et al.*, 2019), with the following modification. The
431 *Escherichia coli* Rosetta 2 cells (DE3) carrying the Tn5 expression plasmid were grown in 3
432 linters (1.5 L x 2) of Luria-Broth (LB) medium containing 100 μg/mL ampicillin and 34 μg/mL
433 chloramphenicol at 37 °C until the optical density reached 0.9 at 600 nm. The cell cultures
434 were cooled to 10 ºC and further cultivated for 15 h after addition of isopropyl β-D-1-
435 thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM to induce protein expression.
436 Following a further 4-hour cultivation at 23 °C, the cells were harvested by centrifugation and
437 resuspended in 50 mL of Lysis buffer (20 mM HEPES-KOH pH 7.2, 0.8 M NaCl, 1 mM EDTA,
438 0.2% (w/v) Triton X-100, 10% (w/v) glycerol) supplemented with cOmplete protease inhibitor
439 cocktail (EDTA-free, Merck). The cells were disrupted by sonication and clarified by

440 centrifugation at 28,000 g at 4 °C for 45 min. 50 mL of the cell lysate (supernatant fraction)
441 was gradually mixed with 5 mL of 10% (v/v) polyethyleneimine (pH 7.5) after which the
442 mixture was incubated at 4 °C for 20 minutes. It was then clarified by centrifugation at 20,000
443 g at 4 °C for 20 min. The resulting supernatant fraction was mixed with 85 mL of Lysis buffer
444 containing protease inhibitor and then applied to a 17 mL bed volume of chitin resin (New
445 England BioLabs). The resin was then washed with 250 mL of Lysis buffer followed by 20 mL
446 of lysis buffer containing protease inhibitor and 100 mM dithiothreitol (DTT). The resin was
447 resuspended in the same buffer and incubated at 4 °C for 36 hours. The eluate was dialyzed
448 against D1 buffer (25 mM HEPES-KOH pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 0.2% (w/v)
449 Triton X-100, 20% (w/v) glycerol) for 15 hours. The dialyzed fraction was applied to a 1 mL
450 Resource S column (Cytiva). The column was developed with a linear gradient of 0.2 to 1 M
451 NaCl in H buffer (20 mM HEPES-KOH pH 7.2, 0.2 mM EDTA, 20% (w/v) glycerol) totaling
452 15 mL. The peak fractions were collected and dialyzed against D2 buffer (100 mM HEPES-
453 KOH pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 10% (w/v) glycerol) at 4 °C for 15 hours and then
454 mixed with glycerol equivalent to 0.8 volume of the dialyzed fraction. The aliquots were snap-
455 frozen in liquid nitrogen and stored at -80 ºC. Oligo and Tn5 assembly was produced according
456 to the publication (Picelli *et al.*, 2014). Tn5 assembled with Oligo was kept at -20C up to one
457 month.
458
459 **Plasmid constructions**
460 ***pT2A-tk-EGFP***
461 The pT2AL200R150G vector (Kawakami, 2005) was digested with XhoI and BglII. A DNA
462 fragment containing multiple cloning sites (MCS), the tk promoter, EGFP coding sequence,
463 and polyadenylation (polyA) sequence was amplified by PCR from ptkEGFP (Uchikawa *et al.*,
464 2003; Kawakami, 2005). This fragment was then recombined with the digested the
465 pT2AL200R150G vector using the In-Fusion® HD cloning kit (TAKARA).
466
467 ***pT2A-CAGGS-mCheery***
468 The pT2A-CAGGS-MCS1-2A-MCS2 vector (Saito *et al.*, 2022) was digested with NotI and
469 MluI. A mCheery coding sequence was amplified by PCR. This fragment was ligated into the
470 digested the pT2A-CAGGS-MCS1-2A-MCS2 by Mighty Mix DNA Ligation Kit (TAKARA).
471
472 ***pT2A-Genomic fragment:tkEGFP***

473 The genomic region of interest was isolated and amplified by PCR from the genome of BL-E
474 strain chickens maintained at Nagoya University. The amplified fragments were digested with
475 XhoI-EcoRV, XhoI, or EcoRV-BglII and subsequently ligated into XhoI-EcoRV or EcoRV-
476 BglII digested pT2A-tkEGFP.
477
478 ***pT2A-Nanog:EGFP***
479 The pT2AL200R150G vector was digested with XhoI and BglII. A DNA fragment containing
480 multiple cloning sites (MCS), EGFP coding sequence, and polyadenylation (polyA) sequence
481 was recombined with the digested the pT2AL200R150G vector using the In-Fusion® HD
482 cloning kit (TAKARA), resulting in the pT2A-MCS-EGFP plasmid. A 4,000 bp upstream
483 region from the Nanog start codon was isolated and amplified by genomic PCR from BL-E
484 strain chickens. This fragment was ligated into the MluI-EcoRV site of pT2A-MCS-EGFP.
485
486 ***pT2A-MVH(DDX4):mCheery***
487 The pT2AL200R150G vector was digested with XhoI and BglII. A DNA fragment containing
488 multiple cloning sites (MCS), mCheery coding sequence, and polyadenylation (polyA)
489 sequence was recombined with the digested the pT2AL200R150G vector using the In-Fusion®
490 HD cloning kit (TAKARA), resulting in the pT2A-MCS-mCheery plasmid. An approximately
491 2,000 bp upstream genomic region from the start codon of the mouse VASA homolog (MVH),
492 kindly provided from Dr. Bertrand Pain, was amplified and recombined with the MluI-EcoRV-
493 digested pT2A-MCS-mCheery vector using the In-Fusion® HD cloning kit. KOD One®
494 (Toyobo) was used for all PCR reactions. Primers and amplified genomic fragments are listed
495 in Supplementary Table 5.
496
497 **Plasmid transfection**
498 A total of 5 x 10^4 cultured PGCs were washed with OPTI-MEM (Gibco) and placed into a 96-
499 well coated plate containing 100 µl of FAot medium (adding Ovo-transferrin instead of chicken
500 serum) (Whyte *et al.*, 2015) without heparin and antibiotics (penicillin, streptomycin, and
501 amphotericin) for 3 hrs. A transfection mixture containing 0.2 µg of total plasmids and 0.5 µl
502 of Lipofectamine 2000 (ThermoFisher Scientific) in 50 µl of OPTI-MEM was then added to
503 the wells. The ratio of plasmid concentrations used was 3:1 for reporter vector : CAGGS-
504 mCherry vector, and an appropriate amount of Tol2 transposase vector (Sato *et al.*, 2007) was
505 co-transfected accordingly. 6 hrs after transfection, the medium was replaced with conventional
506 FAcs medium (maintenance medium) containing heparin and antibiotics.

507

508 **Induction of PGC differentiation**

509 After transfection, PGCs were cultured for one week in a maintenance medium. Subsequently,
510 the cells were transferred to tissue-culture treated plates containing differentiation medium,
511 which was prepared by omitting Activin A from the standard maintenance medium. The
512 medium was refreshed by replacing half of the volume with fresh differentiation medium every
513 two days.

514

515 **PGC transplantation**

516 Transfected PGCs were resuspended in Opti-MEM (Gibco) at a concentration of 5,000 cells/µl
517 and injected into host chicken embryos at HH15-16. The embryos were incubated at 38.5°C
518 and collected after 1 or 4 days of development, corresponding to E3.5 and E6.5, respectively.

519

520 **Imaging**

521 Fluorescence images of cultured cells were acquired using an inverted microscope (IX83,
522 Olympus) equipped with an sCMOS camera (Zyla-4.2 plus, ANDOR) and operated with
523 cellSens software (Olympus). Embryo images were captured using a cooled CCD camera
524 (ORCA-R2, HAMAMATSU Photonics) mounted on a macro zoom microscope (MVX10,
525 Olympus) and controlled by High Speed Recording (HSR) software (HAMAMATSU
526 Photonics).

527

528 **Author contributions**

529 AK and DS: Study design, data analysis, investigation, methodology, writing-original draft,
530 writing-final manuscript and funding acquisition. MI: Data analysis, *in vitro* culture
531 experiments, investigation. HI, YM, MS, YN, SK and KI: Data analysis, methodology and
532 discussion. All authors reviewed and approved the manuscript.

533 **Competing interests**

534 The authors declare that they have no competing interests.

535 **Acknowledgment**

536 We thank Dr. Haruki Ochi from Kobe university for the critical reading of the manuscript and
537 comments. Dr. Yohey Ogawa from National Institute of Genetics, and Dr. Yuya Okusaki and

538 Dr. Kenichi Nishijima from University of Nagoya for the fruitful discussion. This work is
539 supported by the following foundations, AK, DS and KI: "Strategic Research Projects" grant
540 from ROIS (Research Organization of Information and Systems). AK: Grant-in-Aid for
541 Scientific Research KAKENHI (C) 23K05798, Grant-in-Aid for Transformative Research
542 Areas (A) 25H02584 and The Naito Science & Engineering Foundation. DS: Grant-in-Aid for
543 Scientific Research KAKENHI (B) 23K23897.

544 **Data and resource availability**

545 All new sequence reads were deposited in the DNA Data Bank of Japan (DDBJ) Sequence
546 Read Archive under BioProjectID under accession PRJDB37400, in the Genome Sequence
547 Archive under Accession SAMD01620406-SAMD01620426. All data generated in this study
548 are available within the article and its Supplementary Data files.
549

550 **Figure Legends**

551 Figure 1. Quality assessments for Omni-ATAC-seq of live and frozen cultured PGCs.
552 (A) Schema showed the harvesting of PGCs from E2.5 chicken embryos and the workflow of
553 the Omni-ATAC-seq benchmarking setup. (B–C) Cell viability analysis using NucBlue
554 staining in live PGCs (B–B″) and frozen PGCs (C–C″). Frozen PGCs retained nuclear
555 morphology comparable to that of live cells. (D–H) Quality metrics across input conditions,
556 (D) duplication rates detected after initial quality control and adapter trimming using Fastp,
557 and (E) proportion of unmapped reads after alignment. (F-G) contamination rates of
558 mitochondrial reads. (H) FRiP score for the individual libraries. In E to I, live and frozen
559 samples marked with magenta and blue, respectively. (I) Heatmap comparing all Omni-ATAC-
560 seq libraries across a range of input cell numbers for both live and frozen conditions.
561

562 Figure 2. Omni-ATAC-seq from frozen PGCs reduced sensitivity for weak peaks in
563 cryopreserved samples, whereas live and frozen samples still exhibit consistent promoter
564 accessibility at strong peaks.
565 (A) Number of peaks detected by MACS3 (--broad, --q 0.05) without read downscaling, across
566 input sizes and preservation conditions. Peak numbers were significantly lower in frozen
567 samples (blue) compared to live samples (magenta) (t-test, $P < 0.05$). (B) Relationship between
568 input cell number and peak counts (without downscaling). Slight increases in peak counts were
569 observed with higher input, regardless of preservation. (C) Number of peaks detected after
570 normalizing all libraries to 10 million valid reads using scaling factors (no significant

571 difference, $P > 0.05$). (D–E) Heatmaps showing pairwise peak overlaps after downscaling, for
572 promoter-associated peaks (D) and intergenic peaks (E). Promoter peaks were highly
573 concordant across all conditions, while intergenic peaks exhibited greater variability. (F) IGV
574 tracks at *Nanog* locus show consistent peak detection across conditions and input levels at
575 transcription start site (TSS) and transcription termination site (TTS), with merged replicates
576 enhancing signal clarity. (G) Representative genome browser views showing accessible
577 chromatin regions in live, frozen, and gDNA control samples. Live-specific low-intensity
578 peaks are reduced or absent in frozen and gDNA samples. (H) Representative genome browser
579 showed that frozen PGC ATAC-seq libraries fail to detect weak ACRs visible in live samples.
580 Peaks in gDNA samples are enriched at interstitial telomeric sequences (ITSs), reflecting
581 sequence preference of the Tn5 transposase.
582
583 Figure 3. Comparison of PGC and somatic tissue ATAC-seq samples enabled the identification
584 of cell-type-specific cis-regulatory elements, even from low-input frozen PGC ATAC-seq
585 library.
586 Schema of the analysis pipeline: ACRs identified in cultured PGCs were compared with
587 published ATAC-seq datasets from somatic tissues to define PGC-specific ACRs. (B) Venn
588 diagram showing the number of peaks detected in live and cryopreserved PGCs, and their
589 overlap. A total of 55,805 peaks were shared between conditions, with live samples
590 contributing a greater number of unique peaks. (C) Peak annotation of ACRs categorized by
591 feature (Promoter, exon, intron, intergenic, TTS). Many populations of intronic and intergenic
592 peaks were detected and biased in live samples. (D) Upset plot illustrating peak overlaps among
593 8 somatic tissue datasets. Bar plots indicate total peak numbers; pie charts show genomic
594 feature distributions of commonly shared ACRs (left) and PGC-specific ACRs (right). (E)
595 Schematic diagram of the transplantation assay with cultured PGCs. Cultured PGCs transfected
596 with enhancer reporter constructs, CAGGS-mCherry, and CAGGS-T2TP plasmids were co-
597 transfected in cultured PGCs. After the expression of EGFP was confirmed, the culture medium
598 containing Activin was replaced with a differentiation medium (without Activin). (F) Negative
599 control: tk-EGFP construct lacking an upstream enhancer (Only tk). (G) Enhancer construct
600 #6. (H) IgV tracks showed genomic loci at ACR candidate #6. The peak tracks were shown
601 ATAC-seq signals from merged live PGC (mazenta), merged frozen PGC (light blue) and
602 individual 8 tissues (blue).
603
604 Figure 4. PGC-specific ATAC peaks exhibit enhancer activity *in vitro* and are downregulated

605 upon differentiation.
606 Schematic overview of the analysis workflow. (B) Volcano plots showing transcriptome re-
607 analysis comparing *in vivo* PGCs at E2.5 and E6. Genes with absolute $\log_2$ fold change > 2, q-
608 value < 0.01, and baseMean > 10 were highlighted. Genes highly expressed at E2.5 are marked
609 in light orange, and those at E6 in orange. The top 50 most significantly expressed genes are
610 labeled. (C) IGV tracks showing genomic regions around ACR candidate #6. The top two
611 tracks show transcript data from E2.5 and E6, respectively. ATAC-seq signals from merged
612 live PGCs, merged frozen PGCs, and individual somatic tissues are displayed. The left panel
613 shows high expression of LOC107055200 at E2.5, while the right panel shows high expression
614 of LDHA at E6. (D) Schematic diagram of the transplantation assay. Cultured PGCs transfected
615 with enhancer reporter constructs, CAGGS-mCherry, and CAGGS-T2TP plasmids were
616 injected into the bloodstream of E2.5 chicken embryos. Embryos were harvested at either E3.5
617 or E6.5. (E-G) Fluorescence images of transplanted PGCs at E3.5 and E6.5, showing ventral
618 views of the regions outlined in (D). (E) Positive control: Nanog promoter–EGFP construct.
619 (F) Enhancer construct #4. (G) Negative control: tk-EGFP construct lacking an upstream
620 enhancer (Only tk). Scale bars: 50 μm. All experiments were performed with n=5 embryos per
621 condition. (H–I) IGV tracks showing transcripts from E2.5 and E6, along with ATAC-seq
622 signals from merged live PGCs and frozen PGCs across input amounts. At the Nanog locus
623 and INRAGALG00000002296, a highly expressed gene at E2.5, strong ATAC-seq signals
624 were observed around the TSS and/or TTS. Nanog transcripts were detectable in both E2.5 and
625 E6.
626
627 Supplementary Figure 1: Evaluation of peak overlap, sequence bias, and motif enrichment in
628 control and experimental samples. (A–C) Heatmaps showing the percentage of peak overlap
629 across samples for intronic and exonic regions, and transcription termination sites (TTS) based
630 on HOMER annotations. These features exhibited lower reproducibility (~60–65%) compared
631 to promoter peaks, regardless of input condition. (D-E) Nucleotide composition analysis of
632 live-cell peaks and gDNA control peaks. For the comparison, 1,023 peaks were randomly
633 sampled from the live-cell dataset to match the number of peaks identified in the gDNA library.
634 The live-cell peaks showed a mild GC bias, likely reflecting promoter enrichment. In contrast,
635 gDNA peaks displayed no strong nucleotide bias. (F-G) K-mer analysis (k=15) revealed strong
636 sequence preference in the gDNA peaks, indicative of Tn5 transposase sequence bias, which
637 was not observed in live-cell peaks. (H) Frequency distribution of control-biased sequences
638 found in gDNA peaks, confirming that gDNA libraries tend to capture non-specific, sequence-

driven Tn5 activity. (I) MEME motif analysis of gDNA-derived peaks revealed highly enriched tandem repeats of the canonical telomeric sequence (TTAGGG)n, with 6.5 – 7.8 repeats per motif. These sequences were frequently located in intergenic regions and likely correspond to interstitial telomeric sequences (ITSs), emphasizing that peaks from gDNA reflect intrinsic sequence preferences rather than chromatin accessibility. (J) In contrast, MEME analysis of live-cell ATAC-seq peaks showed no significant enrichment for repetitive telomeric motifs, further supporting that accessible chromatin profiles in live samples reflect true biological regulation.

Supplementary Figure 2: Fluorescence images of PGCs transfected with individual enhancer reporter constructs. Cultured PGCs were co-transfected with individual Tol2-based enhancer–EGFP reporter constructs, CAGGS-mCherry, and CAGGS-T2TP plasmids. Each genomic fragment was cloned upstream of a minimal tk promoter driving EGFP. Images were taken under maintenance conditions (+Activin A). For each construct, EGFP and mCherry fluorescence are shown alongside bright-field images. The tk-EGFP vector lacking an enhancer (Only tk) served as a negative control. The Nanog promoter–EGFP construct was included as a positive control. Scale bars: 20 μm. All experiments were performed with n=5 embryos per condition.

Supplementary Figure 3: Genome tracks of PGC ACRs which were validated in this study. Representative IgV tracks at the genomic loci of 14 ACR were identified as uniquely accessible in PGCs, and its absence from somatic tissue datasets. Tracks showed ATAC-seq signal intensities for live and cryopreserved PGCs (low to high input amounts), genomic DNA controls, and merged somatic tissue ATAC-seq libraries (color-coded by sample type). Gene models (bottom) indicate annotated loci near each candidate region. Each candidate region exhibits a robust accessibility peak in PGCs, with consistent signal across live and frozen conditions, and no detectable signal in somatic tissues. Scale bar: 1 Kb.

Supplementary Figure 4: Loss of Nanog- and DDX4/MVH-promoter reporter signals upon differentiation of cultured chicken PGCs.
Cultured chicken PGCs were co-transfected with a Nanog promoter–EGFP reporter (Nanog:EGFP) and a DDX4 promoter–mCherry reporter (DDX4:mCherry), then maintained in maintenance medium (+Activin A) or switched to differentiation medium (−Activin A). Under maintenance conditions, both Nanog:EGFP and DDX4:mCherry signals were readily

673 detected, whereas under differentiation conditions both signals were abolished. Bright-field
674 images are shown for reference. Scale bar: 30 µm.
675
676 Supplementary Figure 5: Fluorescence images of enhancer-reporter PGCs before and after *in*
677 *vitro* differentiation. PGCs were transfected with individual Tol2-based enhancer–EGFP
678 reporter constructs, along with CAGGS-mCherry and CAGGS-T2TP plasmids, and cultured
679 under maintenance conditions (+Activin A) or differentiation conditions (−Activin A) for 3
680 weeks. For each construct, EGFP and mCherry fluorescence were observed and compared
681 between conditions. All enhancer constructs that exhibited EGFP fluorescence under
682 maintenance conditions showed reduced or undetectable EGFP signal upon differentiation.
683 mCherry signal remained detectable across both conditions. Scale bars: 25 µm. All experiments
684 were performed with n=5 embryos per condition.
685
686 Supplementary Figure 6: Fluorescence images of transplanted PGCs expressing individual
687 enhancer constructs at E3.5 and E6.5.
688 Cultured PGCs were transfected with individual Tol2-based enhancer–EGFP reporter
689 constructs, CAGGS-mCherry, and CAGGS-T2TP plasmids, and transplanted into the
690 bloodstream of E2.5 chicken embryos. Fluorescence images were acquired at E3.5 and E6.5,
691 showing ventral views of the regions where PGCs were located. For each enhancer construct,
692 EGFP and mCherry fluorescence are shown. Scale bars: 50 µm. All experiments were
693 performed with n=5 embryos per condition. Individual IGV tracks displaying the
694 corresponding genomic loci and ATAC-seq peaks from merged live PGCs and frozen PGCs
695 across a range of input amounts (low to high).
696
697 Supplementary Table 1: Details of ATAC-seq libraries prepared in this study
698 Supplementary Table 2: SRA data details and accession numbers utilized in the analysis
699 Supplementary Table 3: Annotated peak features using HOMER annotatePeaks.pl
700 Supplementary Table 4: Genomic location of PGC-specific ACRs
701 Supplementary Table 5: ACR information validated in vivo and in vitro
702 Supplementary Table 6: Differential expression genes in E2.5 vs E6 In vivo PGCs
703
704 **References**

705 Bailey, T.L. *et al.* (2015) 'The MEME Suite', *Nucleic acids research*, 43(W1), pp. W39–49.

Barton, L.J. *et al.* (2024) 'The journey of a generation: advances and promises in the study of primordial germ cell migration', *Development (Cambridge, England)*, 151(7). Available at: https://doi.org/10.1242/dev.201102.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics (Oxford, England)*, 34(17), pp. i884–i890.

Chen, Y.-C. *et al.* (2019) 'In vitro culture and characterization of duck primordial germ cells', *Poultry science*, 98(4), pp. 1820–1832.

Choi, H.J. *et al.* (2021) 'Differential transcriptional regulation of the NANOG gene in chicken primordial germ cells and embryonic stem cells', *Journal of animal science and biotechnology*, 12(1), p. 40.

Corces, M.R. *et al.* (2017) 'An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues', *Nature methods*, 14(10), pp. 959–962.

Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2). Available at: https://doi.org/10.1093/gigascience/giab008.

Degalez, F. *et al.* (2024) 'Enriched atlas of lncRNA and protein-coding genes for the GRCg7b chicken assembly and its functional annotation across 47 tissues', *Scientific reports*, 14(1), p. 6588.

Deniz Koç, N. and Yüce, R. (2012) 'A light- and electron microscopic study of primordial germ cells in the zebra fish (Danio rerio)', *Biological research*, 45(4), pp. 331–336.

Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics (Oxford, England)*, 29(1), pp. 15–21.

Fujiwara, S. *et al.* (2019) 'High Quality ATAC-Seq Data Recovered from Cryopreserved Breast Cell Lines and Tissue', *Scientific reports*, 9(1), p. 516.

Halstead, M.M. *et al.* (2020) 'Systematic alteration of ATAC-seq for profiling open chromatin in cryopreserved nuclei preparations from livestock tissues', *Scientific reports*, 10(1), p. 5230.

Hamburger, V. and Hamilton, H.L. (1951) *A Series of Normal Stages in the Development of the Chick Embryo*.

Kawaguchi, A. *et al.* (2024) 'A chromatin code for limb segment identity in axolotl limb regeneration', *Developmental cell*, 59(16), pp. 2239–2253.e9.

Kawakami, K. (2005) 'Transposon tools and methods in zebrafish', *Developmental dynamics : an official publication of the American Association of Anatomists*, 234(2), pp. 244–254.

Khan, A. and Mathelier, A. (2017) 'Intervene: a tool for intersection and visualization of multiple gene or genomic region sets', *BMC bioinformatics*, 18(1), p. 287.

Langmead, B. *et al.* (2019) 'Scaling read aligners to hundreds of threads on general-purpose processors', *Bioinformatics (Oxford, England)*, 35(3), pp. 421–432.

van de Lavoir, M.-C. *et al.* (2006) 'Germline transmission of genetically modified primordial germ cells', *Nature*, 441(7094), pp. 766–769.

Li, B. and Dewey, C.N. (2011) 'RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome', *BMC bioinformatics*, 12, p. 323.

Ling, G. and Waxman, D.J. (2013) 'DNase I digestion of isolated nulcei for genome-wide mapping of DNase hypersensitivity sites in chromatin', *Methods in molecular biology (Clifton, N.J.)*, 977, pp. 21–33.

Morimoto, M. and Saito, D. (2025) 'Avian primordial germ cell migration: History, mechanisms, applications, and unanswered questions', *Protein expression and purification*, 235, p. 106788.

Motta, P.M. *et al.* (2000) 'Mitochondrial morphology in human fetal and adult female germ cells', *Human reproduction (Oxford, England)*, 15 Suppl 2, pp. 129–147.

Muzellec, B. *et al.* (2023) 'PyDESeq2: a python package for bulk RNA-seq differential expression analysis', *Bioinformatics (Oxford, England)*, 39(9). Available at: https://doi.org/10.1093/bioinformatics/btad547.

Nagano, M. *et al.* (2022) 'Nucleome programming is required for the foundation of totipotency in mammalian germline development', *The EMBO journal*, 41(13), p. e110600.

Peng, S. *et al.* (2021) 'Successful ATAC-Seq From Snap-Frozen Equine Tissues', *Frontiers in genetics*, 12, p. 641788.

760 Picelli, S. *et al.* (2014) 'Tn5 transposase and tagmentation procedures for massively scaled
761 sequencing projects', *Genome research*, 24(12), pp. 2033–2040.

762 Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing
763 genomic features', *Bioinformatics (Oxford, England)*, 26(6), pp. 841–842.

764 Saito, D. *et al.* (2022) 'Stiffness of primordial germ cells is required for their extravasation in
765 avian embryos', *iScience*, 25(12), p. 105629.

766 Sato, S. *et al.* (2019) 'Biochemical analysis of nucleosome targeting by Tn5 transposase', *Open*
767 *biology*, 9(8), p. 190116.

768 Sato, Y. *et al.* (2007) 'Stable integration and conditional expression of electroporated
769 transgenes in chicken embryos', *Developmental biology*, 305(2), pp. 616–624.

770 Seki, Y. *et al.* (2007) 'Cellular dynamics associated with the genome-wide epigenetic
771 reprogramming in migrating primordial germ cells in mice', *Development (Cambridge,*
772 *England)*, 134(14), pp. 2627–2638.

773 Uchikawa, M. *et al.* (2003) 'Functional analysis of chicken Sox2 enhancers highlights an array
774 of diverse regulatory elements that are conserved in mammals', *Developmental cell*, 4(4), pp.
775 509–519.

776 Wear, H.M., McPike, M.J. and Watanabe, K.H. (2016) 'From primordial germ cells to
777 primordial follicles: a review and visual representation of early ovarian development in mice',
778 *Journal of ovarian research*, 9(1), p. 36.

779 Whyte, J. *et al.* (2015) 'FGF, Insulin, and SMAD Signaling Cooperate for Avian Primordial
780 Germ Cell Self-Renewal', *Stem cell reports*, 5(6), pp. 1171–1182.

781 Yoshida, S. (2020) 'Mouse Spermatogenesis Reflects the Unity and Diversity of Tissue Stem
782 Cell Niche Systems', *Cold Spring Harbor perspectives in biology*, 12(12). Available at:
783 https://doi.org/10.1101/cshperspect.a036186.

784 Zhang, Y. *et al.* (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome biology*, 9(9),
785 p. R137.

786

787

Figure 1. Kawaguchi *et al.*

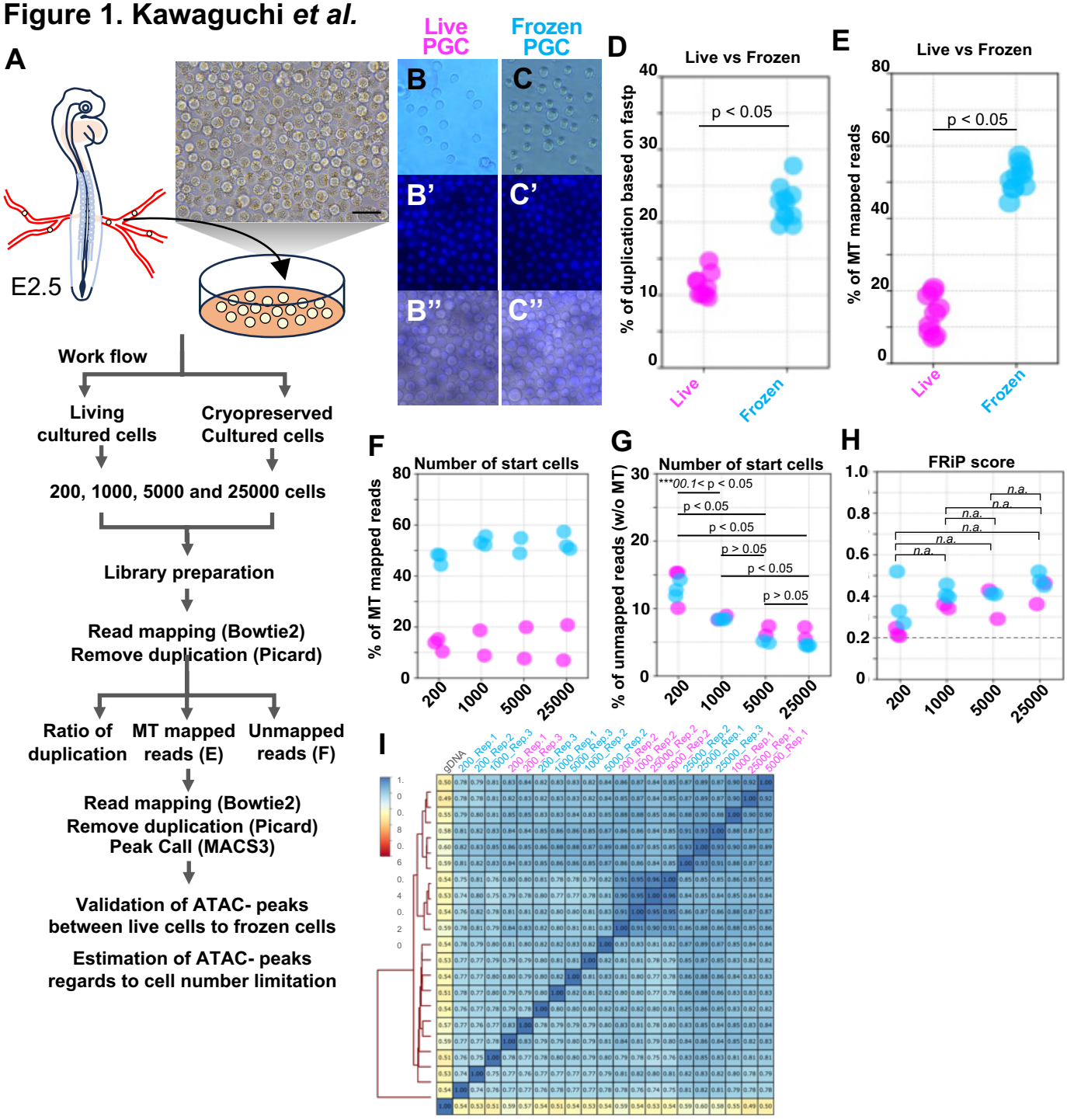


Figure 2. Kawaguchi *et al.*

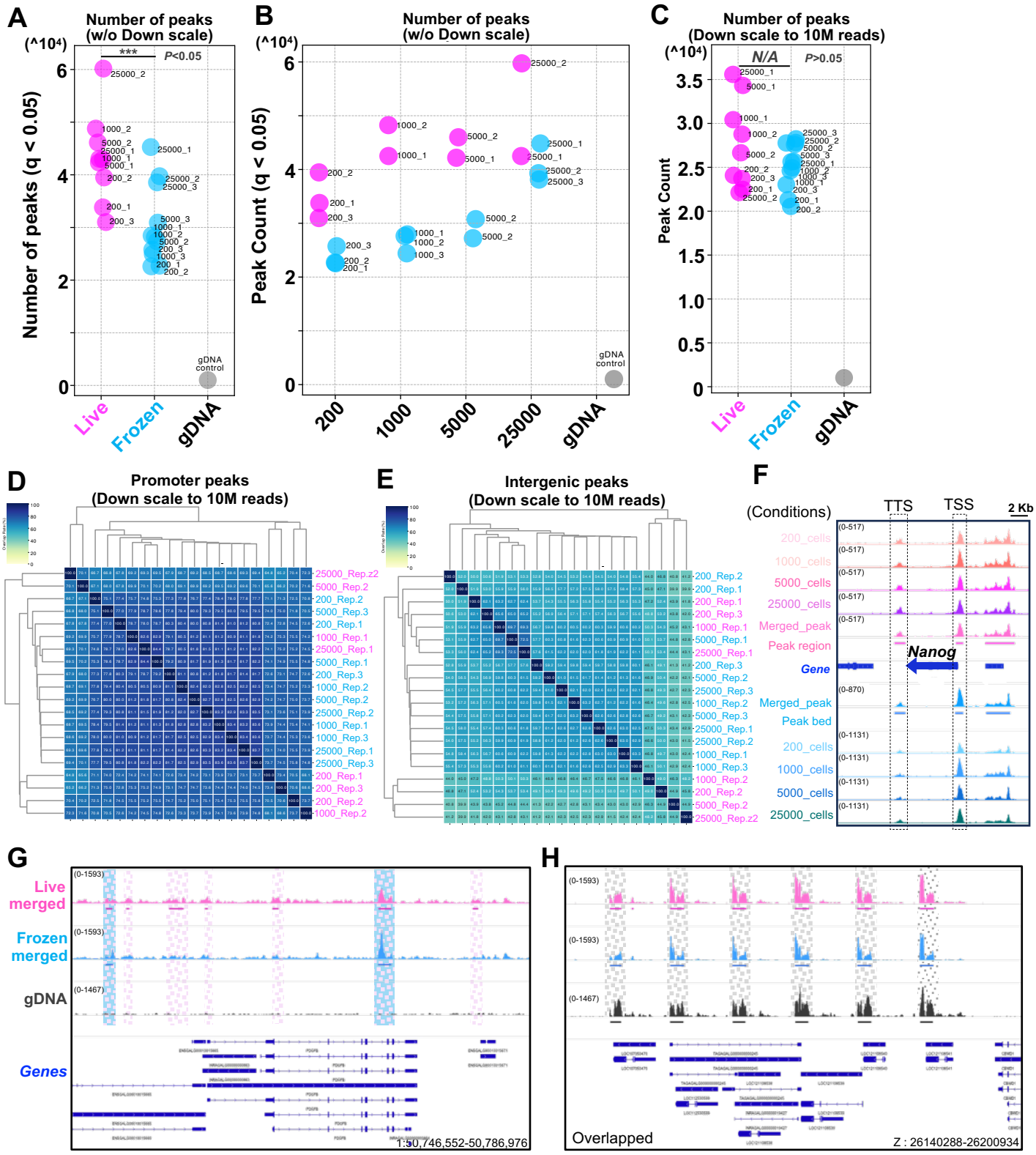


Figure 3. Kawaguchi *et al.*

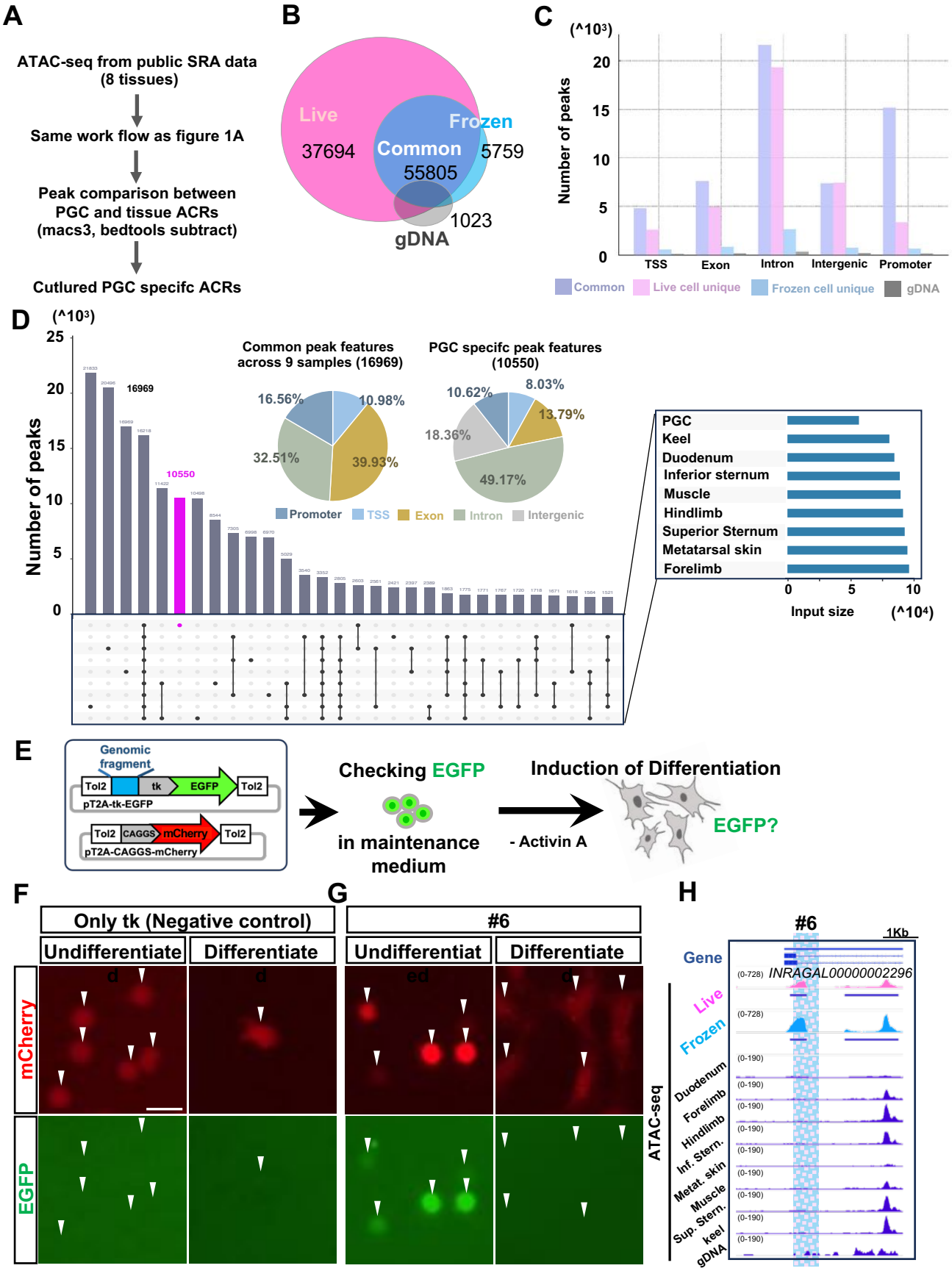


Figure 4. Kawaguchi *et al.*

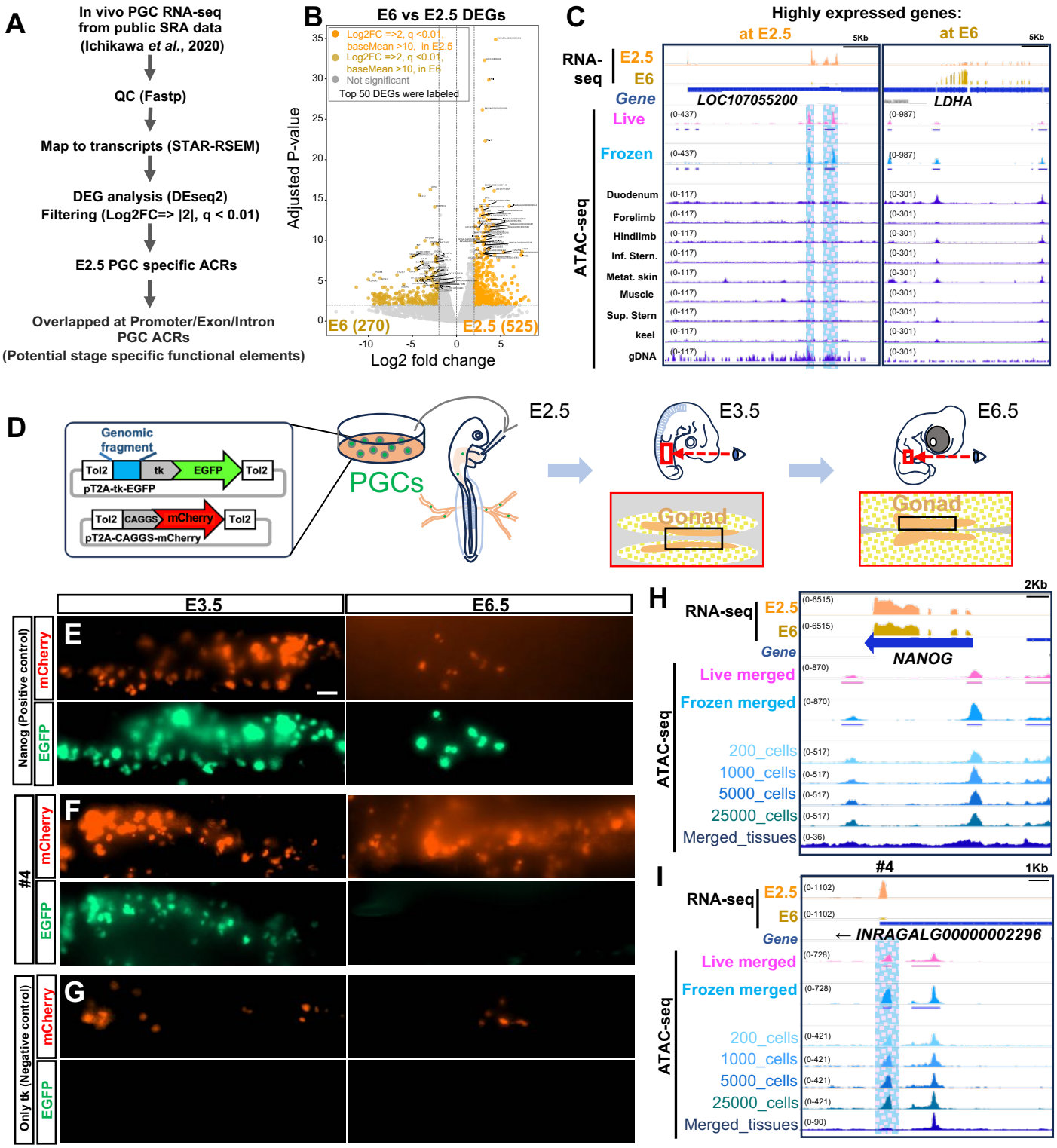
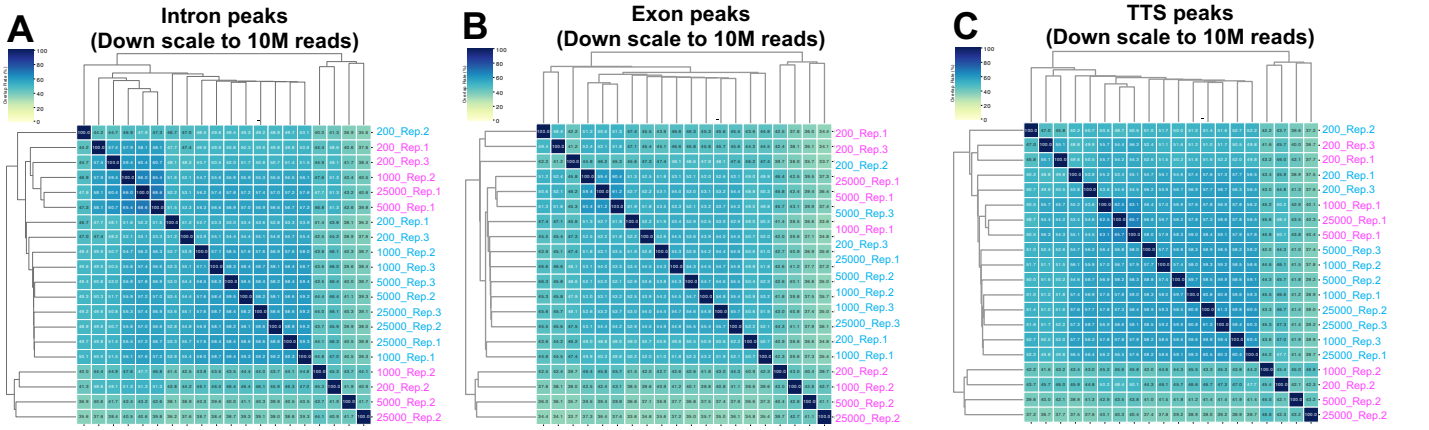
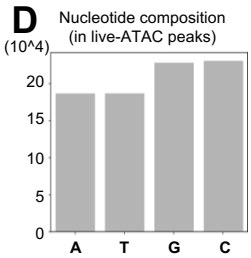
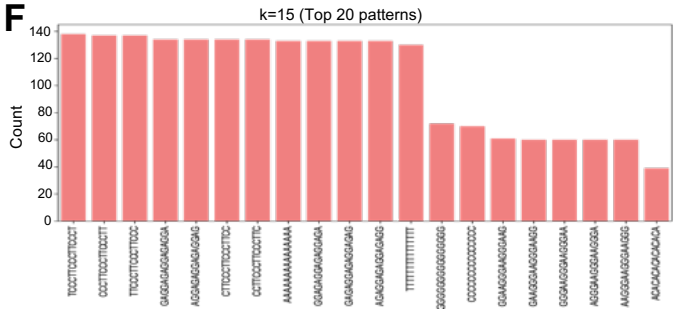


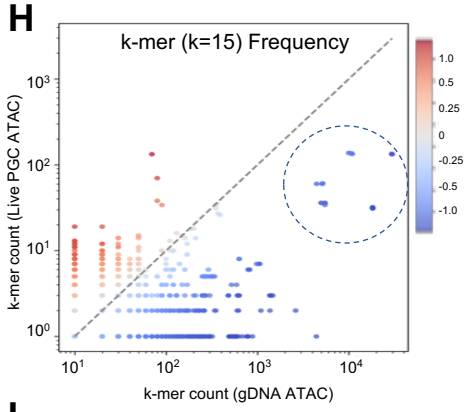
Figure 4 Kawaguchi *et al.*

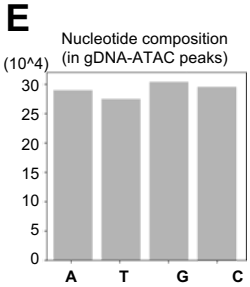
Supplementary Figure 1. Kawaguchi *et al.*

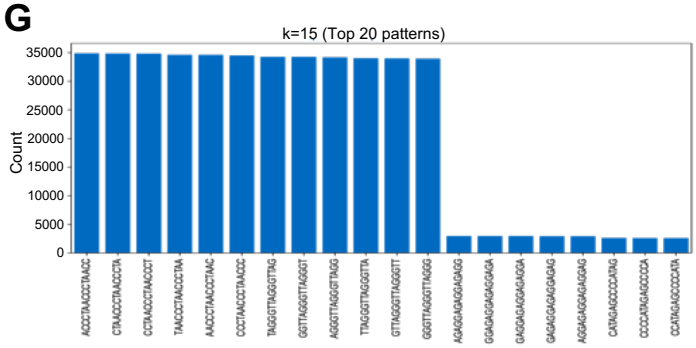






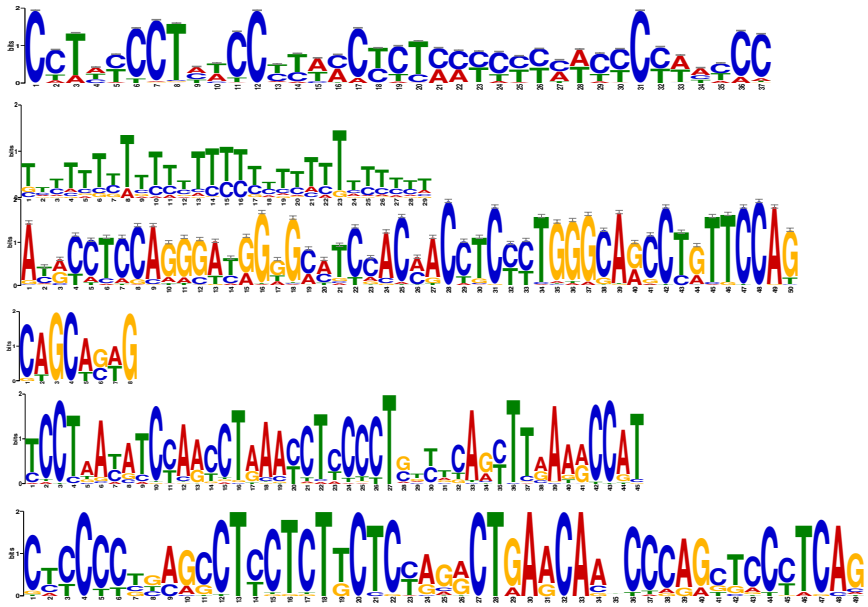






E-value	Mtofi with	Rank
$p < 3.3e-120$	n = 6.5 ~ 7.83	1-10

J



E-value	Mtofi with	sites (1023)	Motif rank
8.7e-116	43	193	13

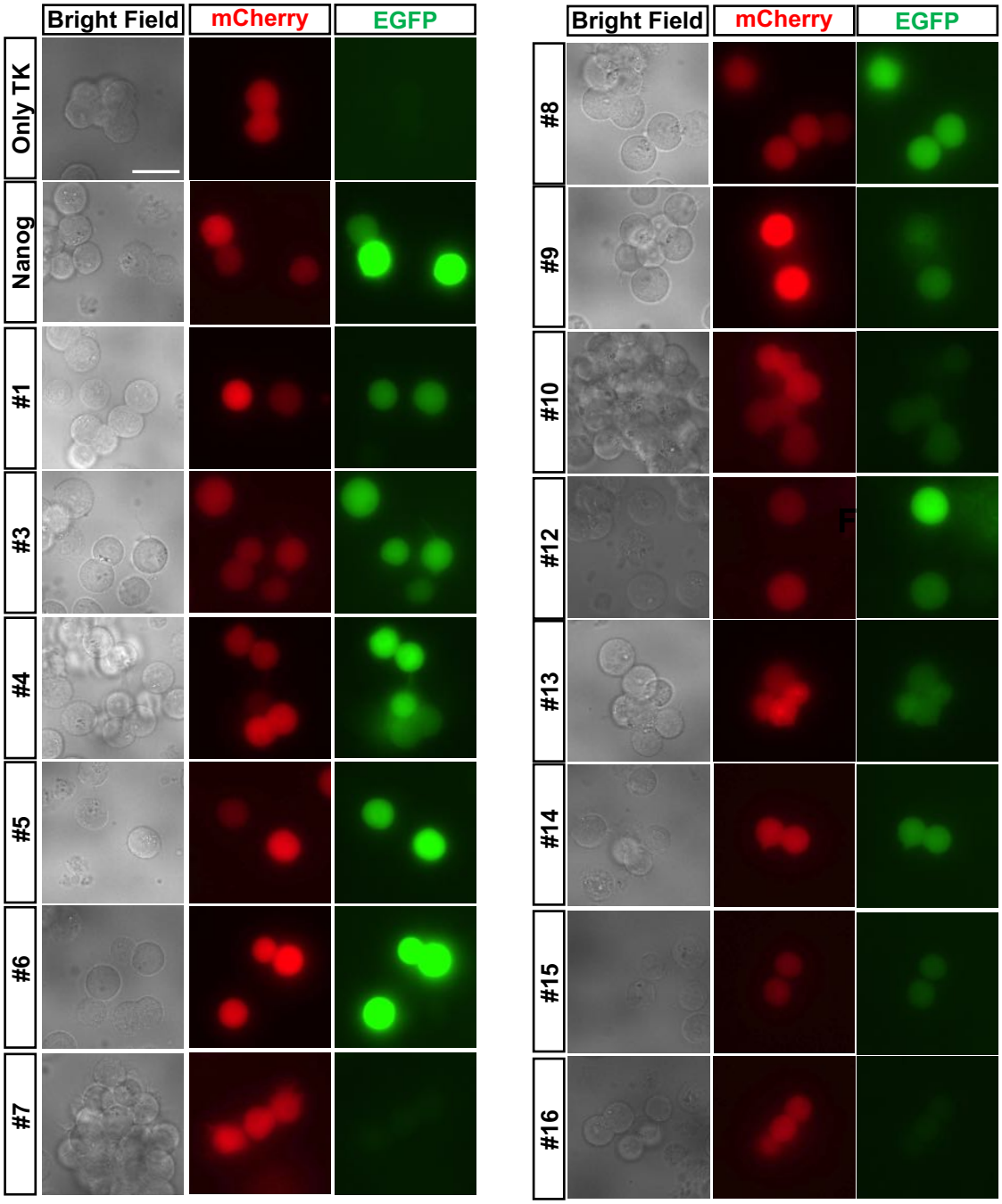
E-value	Mtofi with	sites (1000)	Motif rank
9.5e-050	29	215	1

E-value	Mtofi with	sites (1000)	Motif rank
3.8e-047	50	53	2

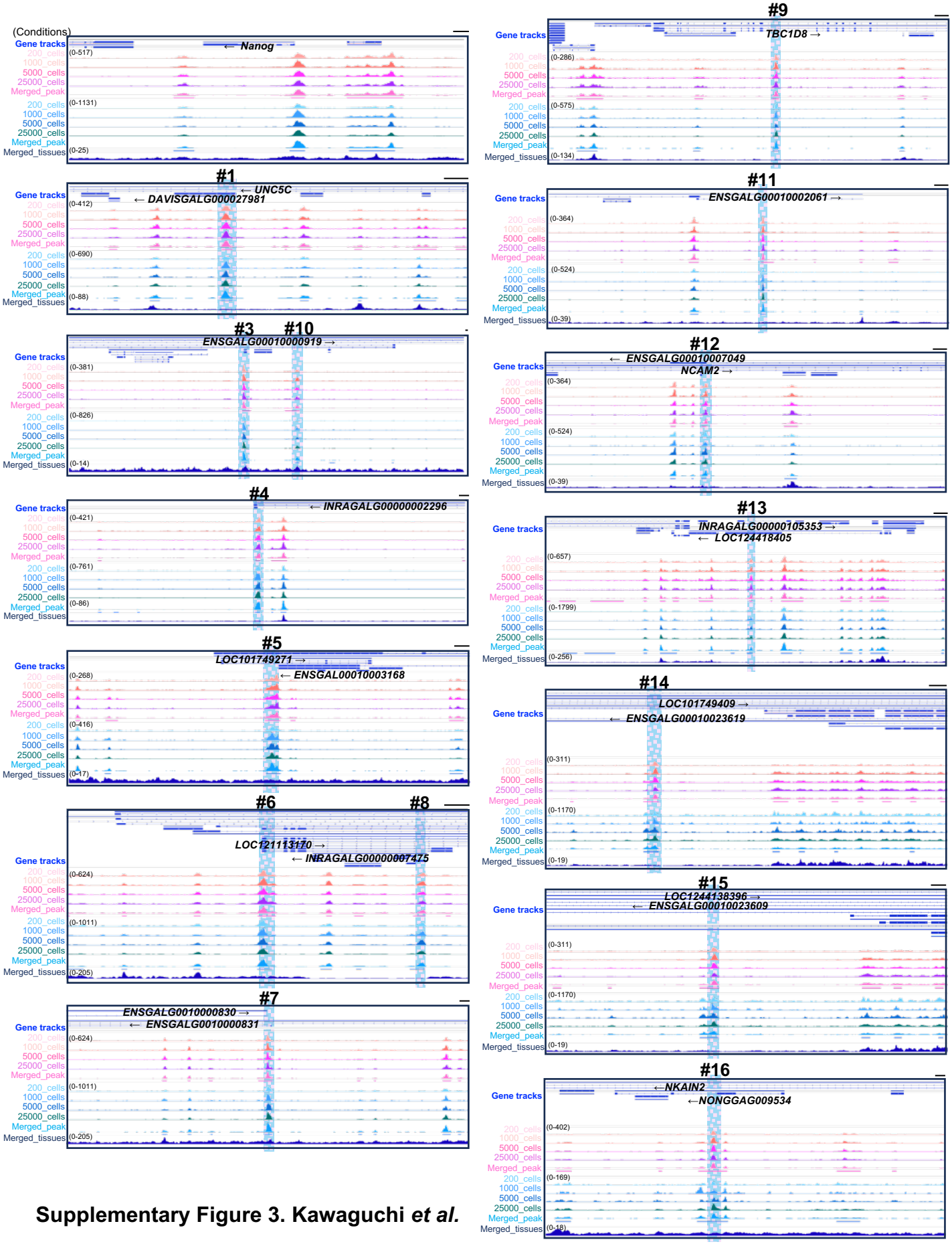
E-value	Mtofi with	sites (1000)	Motif rank
4.1e-009	8	817	3

E-value	Mtofi with	sites (1000)	Motif rank
1.4e-041	45	58	4

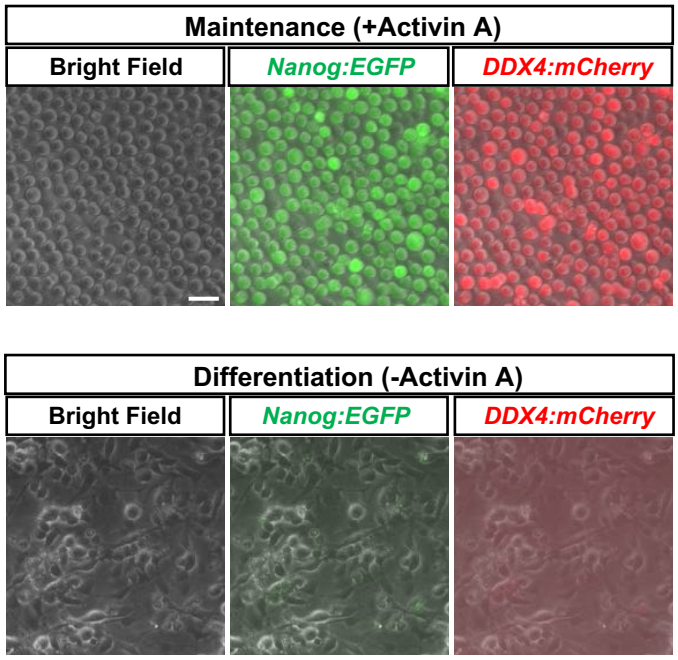
E-value	Mtofi with	sites (1023)	Motif rank
8.7e-116	43	193	5



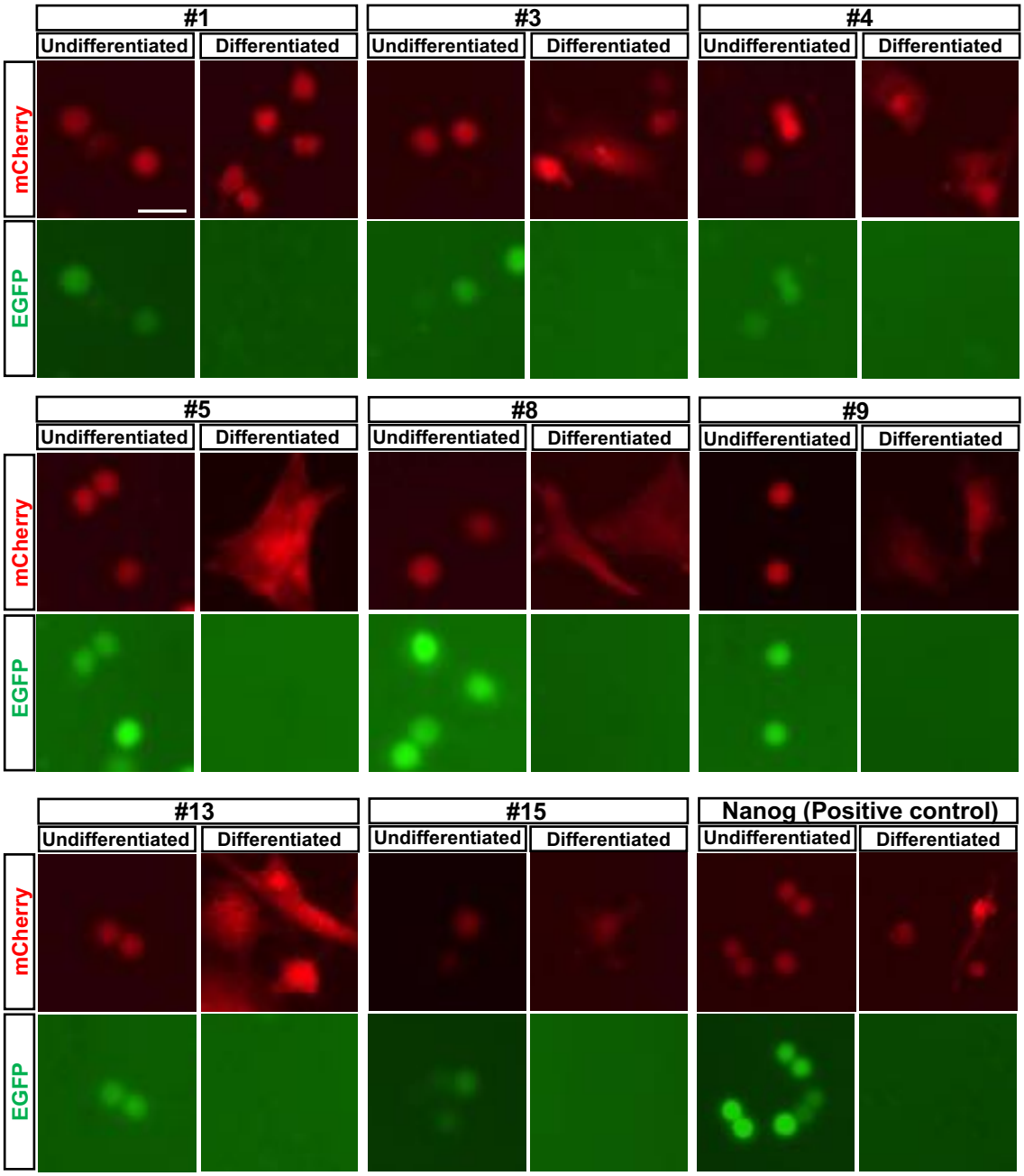
Supplementary Figure 2. Kawaguchi *et al.*



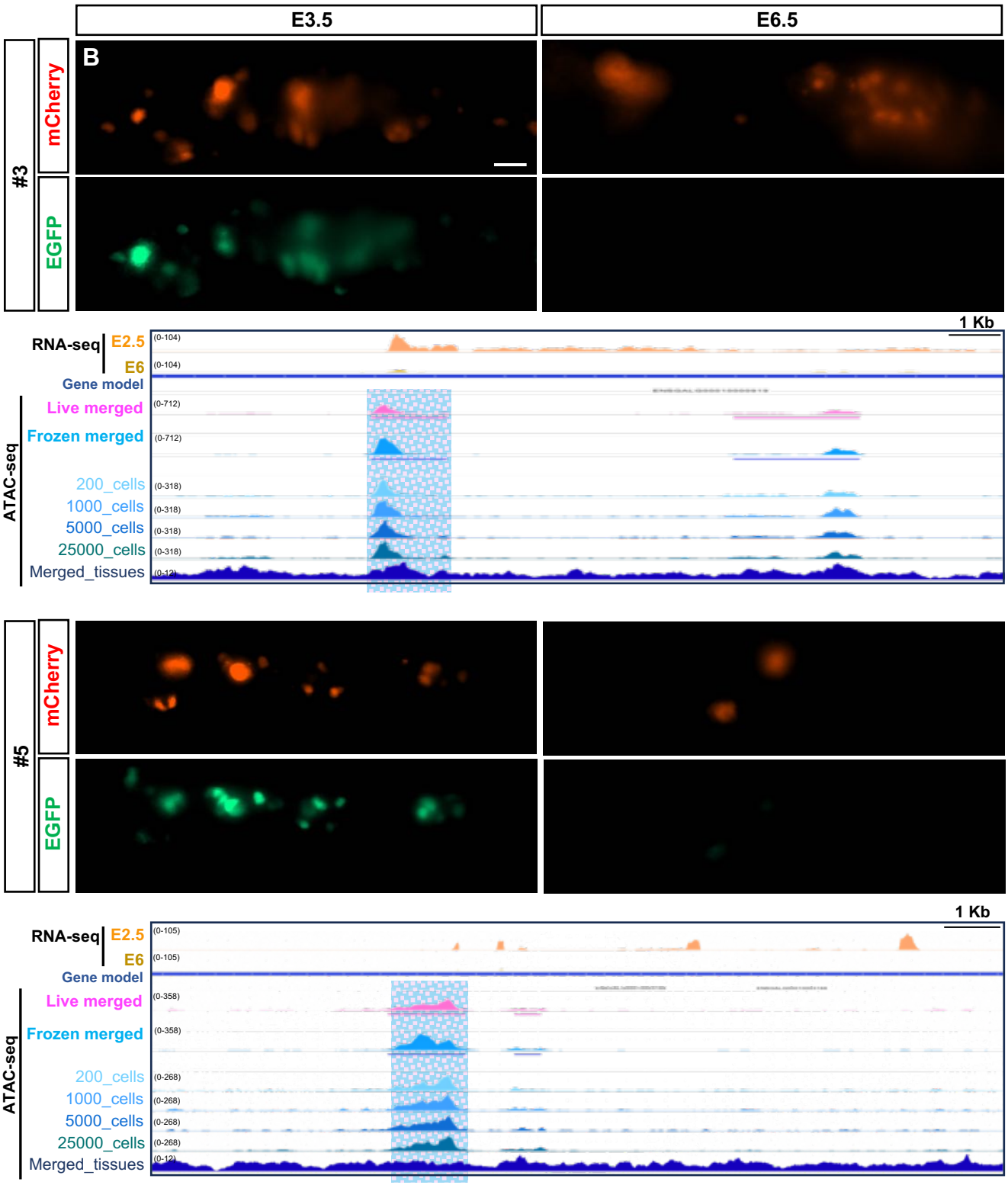
Supplementary Figure 3. Kawaguchi *et al.*



Supplementary Figure 4. Kawaguchi *et al.*



Supplementary Figure 5. Kawaguchi *et al.*



Supplementary Figure 6. Kawaguchi *et al.*