

SUPPLEMENTAL METHODS

Embryo Collection

Adult breeding fish were housed in an aquatic animal facility in tanks of ~30 fish and maintained at 28.5C with 10-hour light and 14-hour dark cycles. To conduct breeding, approximately 90 adult fish were combined at a 2:1 female:male ratio in a large breeding tank the night before breeding. On the day of breeding, any eggs fertilized overnight were discarded, adults were allowed to breed for ~1 hour with embryos collected every 10 minutes. Embryos were placed in an incubator at 28.5C, visually inspected to ensure all embryos were at synchronous stages, and any dead or asynchronous embryos were removed. When embryos reached the appropriate developmental stage as determined by visual observation, they were dechorionated with pronase, dyed with ice-cold PBS, fixed in ice-cold 70% ethanol, then stored at -20C.

Cultured Cells

ZTF cells were generated according to a published protocol (Detrich et al., 2009). Briefly, adult zebrafish were anesthetized with tricane, a small piece of tail was removed with a scalpel, and fish were allowed to recover in water. Tail fin clips were washed 20 times with PBS containing pen/strep, then placed into individual wells of a 96-well plate in Amniomax-II complete medium. The plate was incubated at 28.5C and 5% CO₂ with media changes every 3-4 days, and monitored for cells migrating out of tissue (~1-4 weeks). Once cells migrated out from the tissue, the remaining tissue was removed and media changes continued until cells reached confluence. Cells were progressively plated in increasingly larger dishes until they filled 10cm plates. Once cells reached 6cm, the medium was changed to DMEM with 15% serum, L-glutamine, and pen/strep. Cells were screened to identify homogenous populations of cells morphologically resembling fibroblasts that proliferate at reasonable rates and give normal cell cycle profiles as determined by DNA content analyzed by flow cytometry. The cells were split 1-2 days before harvesting, harvested, centrifuged at 1200g, washed in PBS, centrifuged as before, fixed in 70% ethanol, and stored at -20C.

FACS purification

Ethanol-fixed 28 hpf embryos and ZTF cells were removed from -20C, centrifuged at 1400g and 4C, and placed on ice. The supernatant was removed and cell pellet was resuspended in PBS with 1%BSA to wash cells. Cells were centrifuged as before and resuspended in cold propidium iodide solution containing 20mg/ml RNase A, then incubated on ice for 5 minutes. The cells were then filtered through a 40um mesh then sorted on a FACSAria at a flowrate of 1.0ml/min. Cells are first gated by FCS/SSC to removed cell debris, then by FSC/PI to removed cell doublets. Cells are then sorted based on DNA content, sorting the left side of a G1 peak of 2N DNA content to prevent S-phase contamination, and cells in S-phase between the 2N and 4N DNA content peaks were sorted. More than 500,000 total cells were sorted for all G1 and S-phase samples and cells were kept at 4C or on ice during sorting. Sorted cells were then centrifuged at 1400g, washed in PBS-BSA, centrifuged again as before, and the dry pellet was frozen at -20C.

DNA purification

Cell pellets were removed from -20C and briefly thawed on ice. The cells were resuspended in SDS buffer and then proteinase K was added. The solution was quickly vortexed and incubated at 56C for 2hrs with periodic vortexing. After 2 hours of cell lysis, DNA was isolated by phenol-chloroform extraction and ethanol precipitation. Briefly, a 1:1 ratio of phenol-chloroform was added to the cell lysate, it was vortexed vigorously for 15sec, centrifuged at 14,000g and 4C for 10 minutes, and the top aqueous layer containing DNA was transferred to a new tube. A 1/10 volume of sodium acetate and 2.5x volume of ethanol was added to precipitate DNA, along with

glycogen as a carrier. The solution was placed at -80C for 20 minutes, centrifuged at 14,000g for 15 minutes, the pellet was washed with 70% ethanol and centrifuged as before and resuspended in water. The solution was RNase treated by adding 4ul of 2mg/ml DNase-free RNase A (Genelink) and placing it at 37°C for 1hr. DNA was then purified by phenol-choloroform extraction and ethanol precipitation as before, quantified using a Qubit dsDNA High Sensitivity assay, and stored at -20C.

Next Generation Sequencing

Purified DNA solutions were thawed and 1ug of DNA was sheared on a Covaris S220 Focused-ultrasonicator optimized to produce 300bp fragments. The quality of sheared DNA was analyzed using an Agilent TapeStation 2200 with High Sensitivity D1000 ScreenTape. Libraries were prepared using an NEBNext DNA Library Prep Kit for Illumina with NEBNext Multiplex Oligos for Illumina, according to manufacturer's instructions. Briefly, DNA fragment ends are prepared for adapter ligation, adapters are ligated onto DNA, adapter-ligated DNA is size-selected and cleaned up, 5 cycles of PCR amplification of adapter-ligated DNA was performed with different multiplex primers for each sample, and the PCR product was purified. The quality of the library prep was analyzed on an Agilent 2200 TapeStation and quantified using Qubit. DNA libraries were pooled and sequenced at a depth of 300 million total reads per lane in 100bp paired-end read fragments on an Illumina HiSeq 2500. Genome coverage for developmental samples ranged from 0.44x to 2.95x, with a mean of 1.8x. Coverage for 28 hpf Exp 2 was ~6x, and coverage for ZTF cells was ~2.75x.

Data Processing

Base calls were performed automatically during sequencing using bcl2fastq. Paired-end sequencing reads were aligned to the GRCz10/danRer10 genome build (downloaded from Illumina iGenomes project site) using BWA-Mem with standard parameters, and duplicates were marked using Picard (<https://broadinstitute.github.io/picard/>). Data filtering and processing were performed using Matlab (<http://www.mathworks.com/products/matlab/>). Data were filtered to remove duplicates as well as 100bp regions with greater than 5 standard deviations from the genome median of the G1 control in fixed windows. Two biological replicates of the 28 hpf S/G1 sample were combined into one higher coverage consensus sample ("28 hpf Exp1") and processed together with all developmental embryo samples using the same 28 hpf G1-read windows (n=265255, median=9431bp, IQR=2532bp). Two biological replicates for each developmental sample were averaged into one higher coverage consensus sample and re-processed as described above. A third biological and experimental replicate of the 28 hpf sample ("28 hpf Exp2") was combined with the two original 28 hpf biological replicates for determination of embryo replication timing properties and comparisons to the sequence features data sets ("28 hpf embryos") using combined G1 windows (n=631168, median=3865bp, IQR=1379). Processing of ZTF cells was performed as described above using a ZTF G1 reference to determine windows (n=191234, median=12964, IQR=3261. Zebrafish genome (GRCz10/danRer10) sequence was obtained from the UCSC Genome browser and used to determine GC content in 10kb fixed windows. Repeats were determined from repeatmasker files downloaded from the UCSC Genome Browser Annotation Database and analyzed in 100kb fixed windows. Genes were determined from RefSeq gene lists downloaded from Ensemble Biomart and analyzed in 250kb windows.

Replication Timing Analysis

Chromosome 4 was removed from the whole genome analysis for the determination of genome mean, median, standard deviation, autocorrelation and correlation, as well as comparisons to GC content and repeat density, since the long arm is an outlier in these respects. Origins were identified using difference in approximate derivatives (diff) function in Matlab to identify areas

of local maxima. Replication timing of features including repeats, genes, transcription start sites (TSSs), epigenetic marks (H3K4me3, H3K4me1, H3K27ac), enhancers, PDREs, Ac-PDREs, DARs, and DMRs were performed by interpolating timing values from the indicated timepoint to the midpoint of the feature. Autocorrelation was performed in 300bp increments and plotted at 3kb resolution. Randomized autocorrelation was performed as before on 10 iterations of randomly permuted data.

Segmentation and Classification of the Genome Based on Replication Timing

To segment the zebrafish genome based on replication timing changes, the maximum replication timing difference between any two developmental samples was calculated for every 20kb point, and the hmmSeq (<http://www.stat.missouri.edu/~guhasu/software.html>) package in R was used to partition the zebrafish genome based on those values with a two state hidden Markov model (Cui et al., 2015). For each genomic segment identified by the HMM model, a single replication timing value was calculated for each time point and each biological replicate by averaging all values within the segment. The HMM segments with statistically significant changes in replication timing across development were identified using ANOVA, and the false discovery rate was controlled using the Benjamini–Hochberg procedure ($\alpha=0.05$). K-means clustering was used to classify the regions with statistically significant timing changes. The optimal number of clusters was estimated by examining “elbow” plots of the sum of squared errors versus the number of clusters.

SUPPLEMENTAL REFERENCES

- CUI, S., GUHA, S., FERREIRA, M. & TEGGE, A. 2015. hmmSeq: A hidden Markov model for detecting differentially expressed genes from RNA-seq data. *The Annals of Applied Statistics*, Volume 9, No.2, 901-925.
- DETRICH, H. W., WESTERFIELD, M. & ZON, L. I. 2009. *Essential zebrafish methods : cell and developmental biology*, Amsterdam, Elsevier.