

# **iDamIDseq and iDEAR: An improved method and a computational pipeline to profile chromatin-binding proteins of developing organisms.**

Jose Arturo Gutierrez-Triana<sup>1,3</sup>, Juan L. Mateo<sup>1,3</sup>, David Ibberson<sup>2</sup> and Joachim Wittbrodt<sup>1\*</sup>

<sup>1</sup> Centre for Organismal Studies (COS), University of Heidelberg, Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany

<sup>2</sup> Deep Sequencing Core Facility, Cell Networks, University of Heidelberg, Im Neuenheimer 267, D-69120 Heidelberg, Germany

<sup>3</sup> These authors contribute equally to this work.

\*Corresponding author:

Joachim Wittbrodt

jochen.wittbrodt@cos.uni-heidelberg.de

**Keywords: DamID, transcription factor, transcriptional regulation, chromatin profiling, epigenetics**

## Abstract

DNA adenine methyltransferase identification (DamID) has emerged as an alternative for profiling protein-DNA interactions, however critical issues in the method limit its applicability. Here we present iDamIDseq, a protocol that improves specificity and robustness making its use compatible with developing organisms. In addition, we present the analysis tool iDEAR (iDamIDseq Enrichment Analysis with R) to determine protein-DNA interactions genome wide. The combination of both allows establishing highly reliable transcription factor profiles, even in transient assays. For tissue specific expression we improved the Dam coding sequence to overcome predominant aberrant splicing of Dam fusions we discovered with the commonly used sequence.

## Introduction

Animal development is the result of an exquisite orchestration of changes in gene expression in time and space. Transcription factors (TF) and other chromatin-associated proteins are fundamental elements in these processes and the search for their targets and their logic of regulation in the genome is matter of deep investigation. Chromatin immunoprecipitation (ChIP) and DNA adenine methyltransferase identification (DamID) are the prevailing methods to profile transcription factor binding regions in the genome, reviewed in (Aughey and Southall, 2016; Furey, 2012). ChIP relies on antibody-based capture of protein-DNA complexes on crosslinked and sheared chromatin. Although the protocol is robust, it depends on highly specific precipitating antibodies. In particular, crossreacting antibodies may immunoprecipitate more than one TF simultaneously in a ChIP experiment. DamID offers a suitable solution to these problems. In DamID the fusion of a TF to the bacterial gene DNA adenine methyltransferase, *Dam*, allows for a restricted methylation of adenine residues of the GATC target sequences near the TF binding sites. These regions can then be selected and enriched by digesting the DNA with the restriction enzyme DpnI and linker-mediated PCR (LM-PCR). The PCR products are then hybridized to microarrays or used directly for deep sequencing. DamID is a technique that requires relatively low input material, less processing time and it is cost-effective. DamID has been used successfully in different model organisms such as *Drosophila melanogaster* (Southall et al., 2013), *Caenorhabditis elegans* (Schuster et al., 2010), *Arabidopsis thaliana* (Germann et al., 2006) and mammalian cell culture (Vogel et al., 2007). However, the current protocols require tight control to ensure low expression levels of the proteins fused to the *E.coli* Dam methylase. In addition, in the process of implementing DamID to developing medaka embryos, we faced serious problems such as lack of any DamID product and non-specific amplification (DpnI-independent amplification).

To overcome these drawbacks and allow wide and immediate application of the approach we have implemented iDamIDseq with a series of improvements to the original protocol resulting in a robust and easily applicable method. The approach presented here now allows transcription factor profiling even in

transient applications. We complement our experimental improvements with iDEAR (iDamID Enrichment Analysis with R), an associated analysis pipeline to swiftly establish highly reliable transcription factor profiles.

## Results and Discussion

The iDamIDseq protocol comprises modifications to critical steps of the original technique described below (Fig.1).

Since the *E. coli* Dam displays activity in near-cognate sequences (Horton et al., 2005), we used the mutant version DamL122A (herein referred as Dam) to increase the specificity of methylation on GATC sites.

Chimeric fusion proteins may compromise function due to steric hindrance (Arai et al., 2001). Therefore, we included a flexible linker between the *Dam* protein and the transcription factor, and tested different orientations. We observed that the methylation defect of *Dam*-deficient bacteria could be rescued differentially by the *Dam* fusions depending on the orientation of the fused protein (amino or carboxy terminal) and the lack or presence of the flexylinker (Supplementary Fig.1 and data not shown), these results may explain inefficient DamID amplification in some cases. Noteworthy, medaka embryos injected with mRNA of a fusion of the *Dam* gene and GFP are able to develop properly as assessed at 34 hours post fertilization (Fig.1B,C).

Since we repeatedly obtained linker-mediated amplification independent of DpnI (data not shown), we reasoned that this problem is due to the ligation of the adaptors to free phosphorylated 5'ends resulting from the original genomic DNA preparation and not from the DpnI digestion. We dramatically enhanced the specificity of the adaptor ligation by switching the order of the DpnI and DpnII digestions and the addition of an alkaline phosphatase step. First, we reduced size-complexity by digesting the DNA with DpnII, which cuts GATC sites but is sensitive to adenine methylation. Then, we treated these fragments with alkaline phosphatase and next proceeded with the digestion using DpnI, which only cuts

methyated adenine GATC sites. LM-PCR amplification products were obtained only in samples treated with DpnI (Fig.1D).

In order to prepare the sample for deep sequencing, any contaminant genomic DNA must be removed. Therefore, we performed LM-PCR using primers protected with phosphorothioate modifications and next we treated the samples with T7 exonuclease.

As a proof of concept, we applied iDamIDseq to medaka (*Oryzias latipes*) to reveal a chromatin binding profile of the transcription factor Rx2, the homolog of the mammalian Rax homeodomain proteins involved in retina development. We injected mRNA coding for a nuclear localized Dam\_GFP or Dam\_Rx2 into one cell-stage medaka zygotes and allowed them to develop to Stage 22 (Iwamatsu, 2004) (34 hpf at 28°C, Fig.1B). We extracted the gDNA and processed the sample as described above with two biological replicates per condition. The correlation of read coverage over the genome is very high between replicates and quite distinct between Rx2 and GFP showing the robustness of this method (Fig.2A). We developed an R package, named iDEAR (iDamID Enrichment Analysis with R, available at <https://bitbucket.org/juanlmateo/idear>), to facilitate the straightforward analysis of differential methylation regions (see Methods). Using iDEAR, we were able to identify 7948 Rx2 target regions. Strikingly, we also identified 6255 regions with a significant depletion of signal in the Rx2 samples with respect to GFP. The average length of Rx2 enriched regions is 916bp while for the Rx2 depleted regions is 1011bp (median 731 and 830 respectively), indicating methylation of discrete and specific regions in the genome. The presence of the RAX binding motif correlated well with the enrichment level (Fig.2B), with over 70% of the top Rx2 target regions containing this motif. This suggests that these regions are *bona fide* Rx2 binding sites. Conversely, only 24% of the regions depleted of Rx2 contain this motif on average. This is clearly below the expected frequency of the homeodomain binding motif in random sequences, indicating a counter-selection against the Rx2 binding motif in these regions.

Inspection of regions of Rx2 enrichment revealed many known players involved in retinal development, such as Six3 (Loosli et al., 1998), Otx2 (Zuber et al., 2003), Pax6 (Loosli et al., 1998) and Sox2 (Reinhardt et al., 2015). Interestingly, we also

found enrichment of Rx2 on its own proximal upstream locus (Fig.2C). Part of this region overlaps with a previously characterized regulatory sequence labeled as the Rx2 promoter (Reinhardt et al., 2015), suggesting that Rx2 acts in a self-regulatory feedback loop.

Endogenous promoters can be used to generate transcription factor binding profiles in a tissue specific manner. We cloned *Dam*\_GFP after different promoters (ubiquitin (Mosimann et al., 2011), heat shock (Blechinger et al., 2002) and Rx2 (Reinhardt et al., 2015)) in plasmids carrying transgenesis markers such as *Cmlc2*:GFP or RFP. Surprisingly, none of the *Dam*\_GFP constructs showed GFP expression, whereas the unfused GFP construct did (Fig.3A and data not shown). Since the *Dam*\_GFP can be translated efficiently (see Fig.1C), we suspected of problems at the transcriptional/splicing level. RT-PCR of samples from the different Ubiquitin-driven constructs revealed the aberrant splicing of the *Dam* gene out of the final transcript (Fig.3A-B; Supplementary Fig.2). A customized optimization of the *Dam* gene (*oDam*) removed the cryptic splicing regulatory sites and restored the expression of the GFP in the larvae (Fig.3C; Supplementary Fig.3).

In conclusion, iDamIDseq generates reliable and robust transcription factor profiles when used in developing medaka embryos. Since the *Dam* is a prokaryotic gene, care must be taken to not only optimize codon usage but also eliminate any cryptic splicing sites to achieve proper expression of the fusion protein in different tissues of the fish. This technique can be extended to determine the binding profile of transcription factors of animals transiently via injection in early stages of either RNA or DNA plasmids. Furthermore, these improvements will allow implementing DamID in any organism that is amenable for transgenesis allowing a tissue and time specific profile.

## Materials and methods

### Fish maintenance

Medaka (*Oryzias latipes*) fish were bred and maintained as previously established (Loosli et al., 2000) (URL1). The animals used in the present study were from the inbred strain Cab. All experimental procedures were performed according to the guidelines of the German animal welfare law and approved by the local government (Tierschutzgesetz §11, Abs. 1, Nr. 1, husbandry permit number 35-9185.64/BH Wittbrodt).

### Plasmids

The variant DamL122A was created by site directed mutagenesis of the *E. coli* *Dam* gene using mutagenesis primers and flanking primers (Heckman and Pease, 2007) (Supplementary Table 1). The flexible linker was cloned as a dsOligo that encodes four GGGS amino acid repeats. The repeat sequences are flanked by NheI and SpeI sites. The mmGFP was amplified from plasmid pT2-*otp*ECR6\_E1B::mmGFP (Gutierrez-Triana et al., 2014). All fragments were cloned into the pCS2+ vector (Rupp et al., 1994) as either amino or carboxy terminal fusions, followed by the SV40\_polyadenylation signal of the pCS2 vector. Plasmid integrity was confirmed by sequencing.

We used the gene synthesis service of GeneArt (ThermoFisher Scientific) to obtain the optimized Dam sequence (oDam). In addition to codon optimization, cryptic splice sites were avoided (the DNA sequence is shown as an alignment to the unmodified DamL122A in Supplementary Fig.3). We replaced the DamL122A with the optimized *Dam* in the pCS2+ plasmids described above.

The DamL122A or oDam cassettes were excised using AgeI and NotI from the pCS2+ plasmids and subcloned downstream of the 3.5 kb zebrafish *ubiquitin* promoter (Mosimann et al., 2011) in a Tol2\_based plasmid (Kawakami et al., 1998) having *cmlc2*::EGFP as the insertional reporter (Rembold et al., 2006).

## Microinjection

mRNA was synthesized using mMessage\_mMachine SP6 kit (ThermoFisher Scientific, AM1340) on linearized pCS2+ templates. One cell stage zygotes were injected with mRNA at 10 ng/μl. The progenies of injected fish were maintained in ERM medium (Loosli et al., 2000) supplemented with penicillin streptomycin (P0781, Sigma-Aldrich, 200 units/200 μg per ml of ERM). Embryos were collected at stage 22 (34 hpf at 28°C). Unfertilized and dead embryos were removed. Recombinant Ubiquitin promoter plasmids were injected into one-cell stage zygotes at 10 ng/μl in the presence of 10 ng/μl Tol2 transposase mRNA. The injected embryos were maintained in ERM medium supplemented with 0.2 mM N-phenylthiourea (Sigma-Aldrich, catalogue no. P7629) until hatching.

## DpnI protection assay

pCS2+ plasmids carrying the cassettes coding for every particular *Dam* fusion protein were used to transform the *E. coli* strain C2925, deficient in the *dam/dcm* methylation system. As control, the pUC19 plasmid was used to transform C2925 cells and One Shot® TOP10 cells, which have a normal methylation system. Bacterial genomic DNA was isolated from 3 ml LBamp cultures from individual colonies using the DNeasy Tissue kit (QIAGEN, Cat. No. 69504). 1 μg of gDNA was digested with 10 units of DpnI (NEB, R0176S) for 1h at 37°C. The products were run in a 1% agarose gel.

## iDamIDseq protocol

*gDNA isolation:* Wash 20-30 embryos or tissue with 1xERM or 1xPBS respectively. Remove as much media as possible. Homogenize using pestle in 400 μl of TEN buffer (100 mM Tris-HCl pH8.5, 10 mM EDTA, 200 mM NaCl, 1% SDS) plus 20 μl of 20mg/ml Proteinase\_K. Incubate O/N at 50-60°C. Cool down at RT for 5 min. Add 20 μl of 10 mg/ml RNase A (DNase and Proteinase-free, ThermoScientific, EN0531). Incubate 15 min at RT. Add 600 μl of phenol:chloroform:isoamylalcohol (25:24:1)(Roth, A156.1). Mix well by inversion. Incubate for 10 min at RT. Centrifuge at 10.000 rpm at RT for 20 min. Transfer the aqueous phase to a tube containing 600 μl of chloroform. Mix and centrifuge at 10.000 rpm for 10 min. Transfer the aqueous phase to a tube



containing 600 µl of isopropanol. Mix and put at -20°C for 30 min. Centrifuge at 10.000 rpm at 4°C for 20 min. Remove supernatant and add 800 µl of cold 70% ethanol. Centrifuge at 14.000 rpm at RT for 10 min. Remove supernatant as much as possible and dry the pellet at 60°C for 10 min. Add 50 µl of prewarmed water (60°C). Incubate the tubes for 10-20 min at 60°C, gently flicking the tube sporadically until pellet is dissolved. Check OD260 and quality in a gel.

*DpnII Digestion and Alkaline phosphatase treatment:* In a 20 µl reaction, add 2 µl 10x NEB3.1 buffer, 1 µg of gDNA and 10 units of DpnII (NEB, R0543S). Incubate 6 hours at 37°C. Inactivate the enzyme at 65°C for 20min. To the inactive DpnII reaction add 23 µl H<sub>2</sub>O, 5 µl 10xAP buffer and 5 units of Antarctic phosphatase (NEB, M0289). Incubate reaction 1h at 37°C and inactivate at 70°C for 10 min. cleanup using column and elute in 12 µl of H<sub>2</sub>O.

*DpnI digestion:* In a 10 µl reaction, add 1 µl CutSmart buffer, 5 µl of DpnII/AP treated sample and 10 units of DpnI enzyme (NEB, R0176S). Exclude DpnI from the control tube. Incubate reaction at 37°C for 12 hours. Inactivate at 80°C for 20 min.

*Adaptor ligation:* To the DpnI +/- treated sample add 2µl 10x T4 ligase buffer, 1 µl of 50 µM dsOligos AdRt/b, 2.5 units of T4 DNA ligase (Thermo Scientific, EL0011) and 6.5 µl H<sub>2</sub>O. Incubate O/N at 16-18°C. Clean up using column and elute in 50 µl of H<sub>2</sub>O.

*LM-PCR:* In a 25 µl reaction, add 2.5 µl of 10x ThermoPol Buffer, 1 µl 10 mM dNTPS, 1 µl 10µM AdR\_PCR primer (Supplementary Table 1), 5 µl of ligation sample and 1.25 units of Taq polymerase (NEB, M0267S). PCR program is as follows: 68°C 10 min, 1 cycle of 94°C 15 sec, 65°C 30 sec, 68°C 5 min and 20 to 30 cycles of 94°C 15 sec, 65°C 30 sec, 68°C 2 min. The number of cycles needs to be determined experimentally. Run 5 µl of PCR in a 1% agarose gel. A smear should appear only in the DpnI+ samples. Smear is typically from 200 bp to 2Kb.

*T7 exonuclease treatment:* To 30  $\mu$ l of clean LM-PCR\*, add 5  $\mu$ l 10xCutSmart buffer, 10units T7 exonuclease (NEB, M0263S) and 14  $\mu$ l H<sub>2</sub>O. Incubate 1h at 25°C and cleanup using column in 20  $\mu$ l of H<sub>2</sub>O. The sample is then ready for library preparation.

In order to obtain the minimum amount of DNA for deep sequencing it might be necessary to repeat the PCR step and pool the DNA samples.

### **Sequencing**

DNA samples were fragmented using the Covaris S2 sonicator in AFA microtubes. The library was prepared using the NEBNext Ultra DNA Library Prep kit for Illumina (E7370, NEB). Sequencing was performed in an Illumina HiSeq 2000 sequencing system.

### **Data analysis**

*Sequencing data processing:* Reads were mapped to the medaka genome (Kasahara et al., 2007) (oryLat2 assembly) using Bowtie2 (Langmead and Salzberg, 2012) with default parameters. Then the mapped reads were filtered with samtools (Li et al., 2009) to keep only those with a minimum mapping quality of 20.

*Identification of enriched regions:* First, the set of potential DpnI fragments was built from a BSgenome object for the oryLat2 assembly using the function `vMatchPattern` with the restriction site “GATC”. Only fragments spanning adjacent predicted restriction sites with length in the range from 200 to 2000 bases were considered. Next, the number of reads falling into each predicted fragment was counted for each one of the samples with the function `summarizeOverlaps` setting the parameter `ignore.strand` to `TRUE`. These counts were used to produce the correlation heatmap in Figure 1E. In order to discard fragments with spurious mapped reads, only fragments with a minimum number of reads relative to the fragment length were kept. This threshold was computed

as three times the total number of reads in all the fragments considered divided by the total length of them. After this selection, fragments that were not further apart than the smallest fragment length were joined together. With the resulting set of genomic regions, the read count was computed again with `summarizeOverlaps` and the resulting matrix was used to compute significant differences between samples using DESeq2 (Love et al., 2014). The R package `iDEAR` implements this data analysis pipeline and is available at <https://bitbucket.org/juanlmateo/idear>.

*Motif enrichment:* FIMO (Grant et al., 2011) was used to identify motif matches of the RAX binding motif (Jolma et al., 2013) (RAX\_DBD) in the sequence within the coordinates of each identified region. The resulting regions were sorted by significance (from highest to lowest) and divided in 50 bins. This was done independently for each set of enriched or depleted Rx2 regions. For each bin, the ratio of regions with at least one match of the motif divided by the number of regions in the bin was computed. The shuffled sequences were generated by randomly permuting dinucleotides for each sequence extracted from the Rx2 regions.

## **RT-PCR**

10 hatchling larvae (10 dpf) per experimental group, expressing EGFP in the heart, were transferred to 1.5 ml tube and total RNA was isolated using Trizol reagent (ThermoFisher Scientific, 15596026) according to the manufacturer's instructions. 1 µg of DNase-treated total RNA was used to synthesize cDNA. PCR was done using Q5 polymerase (NEB, M0491) using 1xQ5 buffer, 200 µM dNTPs, 0.5 µM RT\_Ubi fwd primer, 0.5 µM RT\_mmGFP rev primer (Supplementary Table 1), 2 µl of cDNA and 0.02 U/µl Q5 polymerase. Cycling parameters were: 94°C 2min; 35 cycles, 94°C 15sec, 60°C 30 sec and 72°C 1min; 72°C 5 min. PCR products were analyzed in an agarose gel and sequenced.

## Acknowledgements

We thank Soojin Ryu for support during the initial DamID experiments. We highly appreciate the important input by Lazaro Centanin and Daigo Inoue, and the supporting feedback by the Wittbrodt lab.

## Author contributions

JAGT, JLM and JW designed the study. JAGT and DI performed the experiments. JLM developed iDEAR. JAGT, JLM and JW analyzed the data and wrote the manuscript.

## Funding

This work was financially supported by the DFG (CRC 873, JW) and the ERC (JW).

## Data availability

The sequencing raw data are available at ArrayExpress with accession E-MTAB-4610.

## References

- Arai, R., Ueda, H., Kitayama, A., Kamiya, N. and Nagamune, T.** (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* **14**, 529–32.
- Aughey, G. N. and Southall, T. D.** (2016). Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 25–37.
- Blechinger, S. R., Evans, T. G., Tang, P. T., Kuwada, J. Y., Warren, J. T. and Krone, P. H.** (2002). The heat-inducible zebrafish hsp70 gene is expressed during normal lens development under non-stress conditions. *Mech. Dev.* **112**, 213–5.
- Furey, T. S.** (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13**, 840–52.

- Germann, S., Juul-Jensen, T., Letarnec, B. and Gaudin, V.** (2006). DamID, a new tool for studying plant chromatin profiling in vivo, and its use to identify putative LHP1 target loci. *Plant J. Cell Mol. Biol.* **48**, 153–163.
- Grant, C. E., Bailey, T. L. and Noble, W. S.** (2011). FIMO: scanning for occurrences of a given motif. *Bioinforma. Oxf. Engl.* **27**, 1017–8.
- Gutierrez-Triana, J. A., Herget, U., Lichtner, P., Castillo-Ramírez, L. A. and Ryu, S.** (2014). A vertebrate-conserved cis-regulatory module for targeted expression in the main hypothalamic regulatory region for the stress response. *BMC Dev. Biol.* **14**, 41.
- Heckman, K. L. and Pease, L. R.** (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2**, 924–932.
- Horton, J. R., Liebert, K., Hattman, S., Jeltsch, A. and Cheng, X.** (2005). Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. *Cell* **121**, 349–61.
- Iwamatsu, T.** (2004). Stages of normal development in the medaka *Oryzias latipes*. *Mech. Dev.* **121**, 605–618.
- Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al.** (2013). DNA-binding specificities of human transcription factors. *Cell* **152**, 327–39.
- Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., et al.** (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714–9.
- Kawakami, K., Koga, A., Hori, H. and Shima, A.** (1998). Excision of the tol2 transposable element of the medaka fish, *Oryzias latipes*, in zebrafish, *Danio rerio*. *Gene* **225**, 17–22.
- Langmead, B. and Salzberg, S. L.** (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup** (2009). The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–9.
- Loosli, F., Köster, R. W., Carl, M., Krone, a and Wittbrodt, J.** (1998). Six3, a medaka homologue of the *Drosophila* homeobox gene *sine oculis* is expressed in the anterior embryonic shield and the developing eye. *Mech. Dev.* **74**, 159–64.
- Loosli, F., Köster, R. W., Carl, M., Kühnlein, R., Henrich, T., Mücke, M., Krone, A. and Wittbrodt, J.** (2000). A genetic screen for mutations affecting

- embryonic development in medaka fish (*Oryzias latipes*). *Mech. Dev.* **97**, 133–9.
- Love, M. I., Huber, W. and Anders, S.** (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Mosimann, C., Kaufman, C. K., Li, P., Pugach, E. K., Tamplin, O. J. and Zon, L. I.** (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177.
- Reinhardt, R., Centanin, L., Tavhelidse, T., Inoue, D., Wittbrodt, B., Concordet, J.-P., Martinez-Morales, J. R. and Wittbrodt, J.** (2015). Sox2, Tlx, Gli3, and Her9 converge on Rx2 to define retinal stem cells in vivo. *EMBO J.* **34**, 1572–88.
- Rembold, M., Lahiri, K., Foulkes, N. S. and Wittbrodt, J.** (2006). Transgenesis in fish: efficient selection of transgenic fish by co-injection with a fluorescent reporter construct. *Nat. Protoc.* **1**, 1133–9.
- Rupp, R. A., Snider, L. and Weintraub, H.** (1994). *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8**, 1311–23.
- Schuster, E., McElwee, J. J., Tullet, J. M. A., Doonan, R., Matthijssens, F., Reece-Hoyes, J. S., Hope, I. A., Vanfleteren, J. R., Thornton, J. M. and Gems, D.** (2010). DamID in *C. elegans* reveals longevity-associated targets of DAF-16/FoxO. *Mol. Syst. Biol.* **6**, 399.
- Southall, T. D., Gold, K. S., Egger, B., Davidson, C. M., Caygill, E. E., Marshall, O. J. and Brand, A. H.** (2013). Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. *Dev. Cell* **26**, 101–12.
- Vogel, M. J., Peric-Hupkes, D. and van Steensel, B.** (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478.
- Zuber, M. E., Gestri, G., Viczian, A. S., Barsacchi, G. and Harris, W. A.** (2003). Specification of the vertebrate eye by a network of eye field transcription factors. *Dev. Camb. Engl.* **130**, 5155–67.

## Figure legends

Figure 1. iDamIDseq to robustly analyze protein-DNA interactions. A) Overview of the iDamIDseq procedure. Critical modifications are highlighted in red. B)

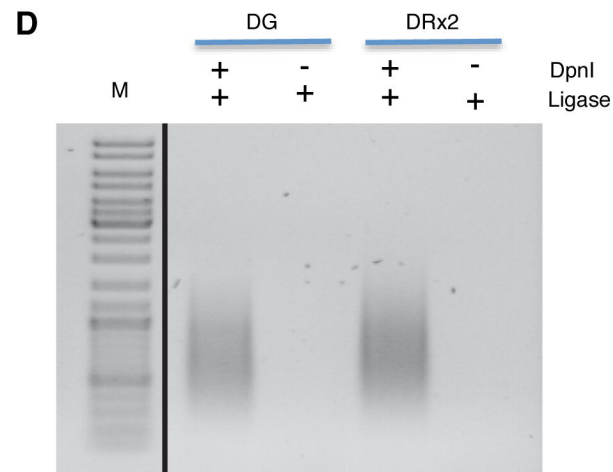
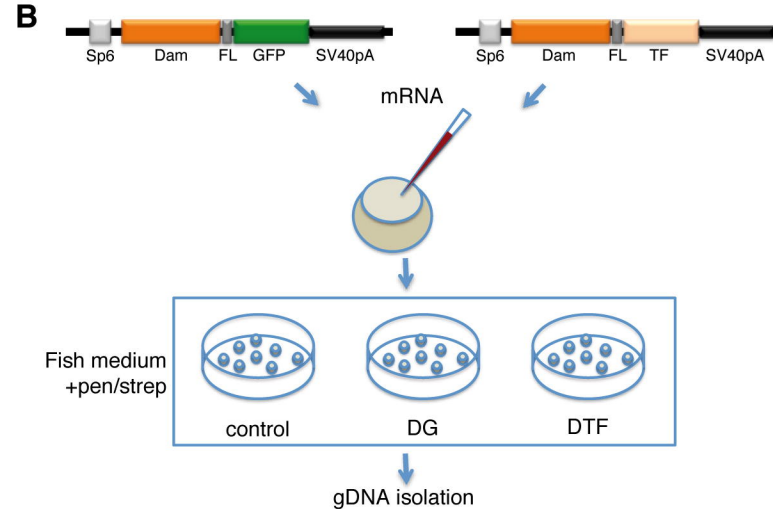
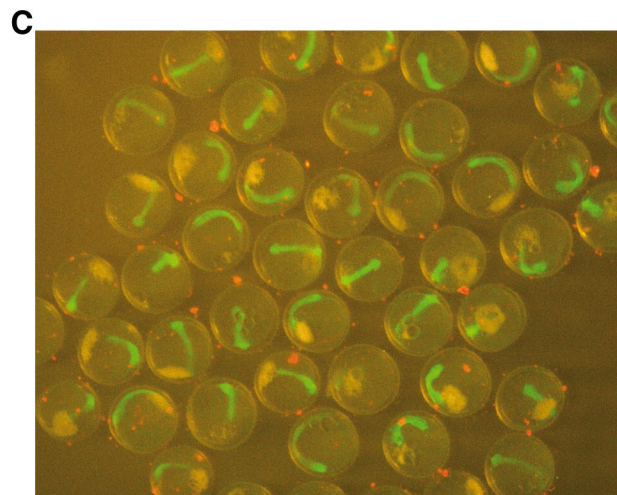
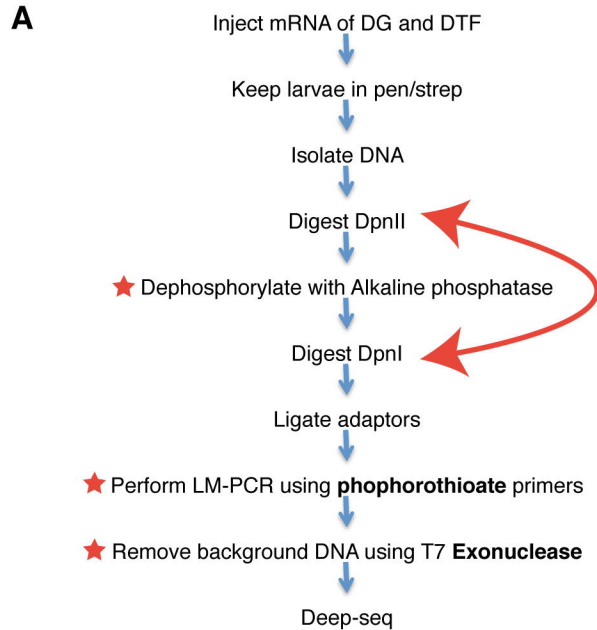
Delivery of Dam fusion proteins. In vitro synthesized mRNA encoding Dam\_GFP (DG) or Dam\_TF (DTF) were injected into 1 cell stage medaka zygotes. C) Dam\_GFP expression in medaka embryos (stage 22) did not affect embryonic development. D) iDamID amplification at 25 cycles from embryos injected with DG or DRx2 mRNA. Amplified DNA fragments were observed only in samples treated with DpnI. The black line marks intervening lanes removed.

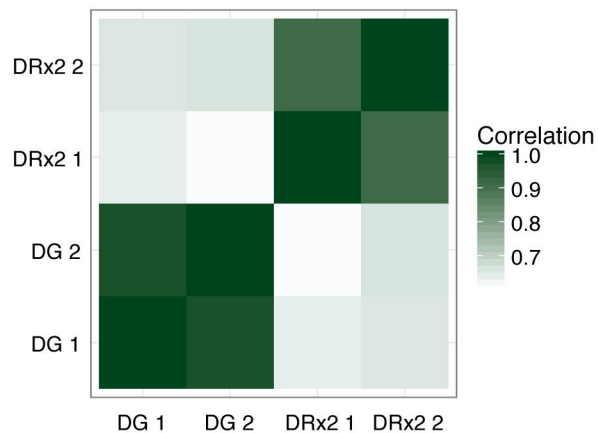
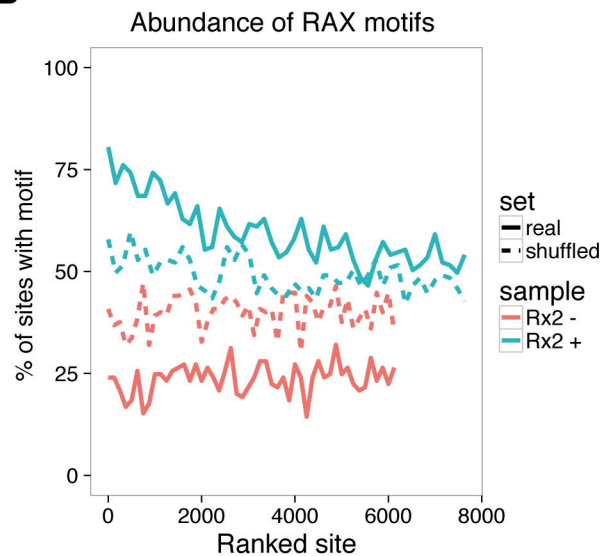
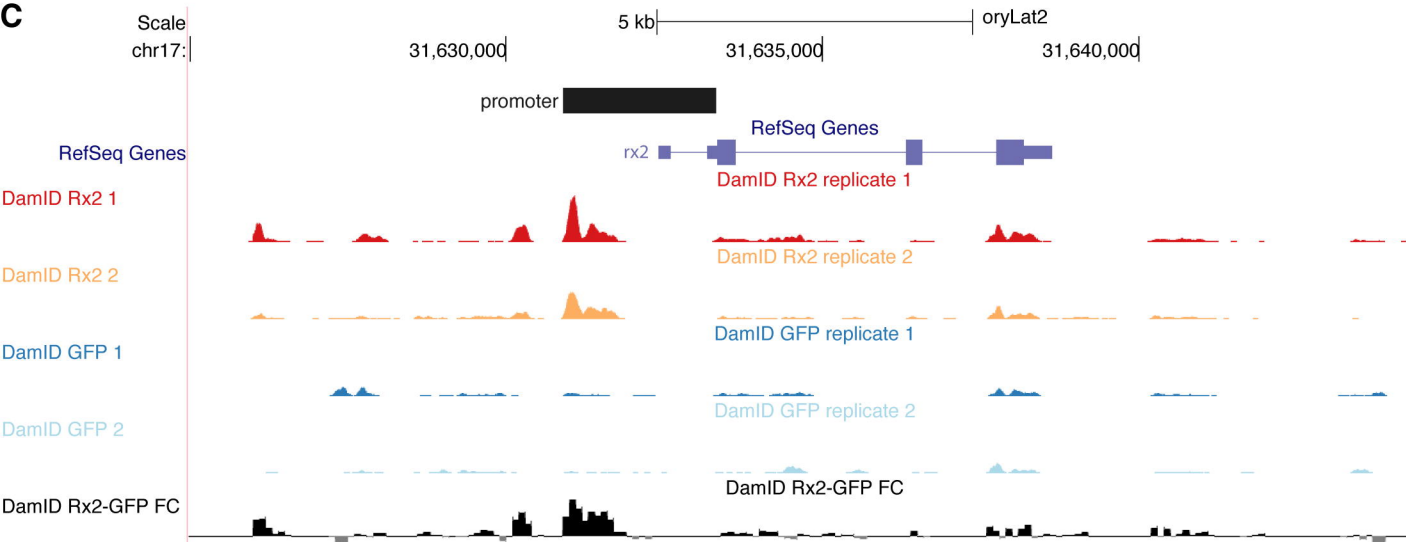
Figure 2. Analysis of iDamIDseq on Rx2. A) Samples showed high correlation between replicates and low correlation between DRx2 and DG based on the genome wide read coverage. B) Representation of the abundance of the RAX binding motif with respect to the level of significance of the identified regions. In blue and red are the Rx2 enriched and depleted regions respectively, dashed lines represent the background established by random shuffling of the same sequence. C) Representation of the iDamIDseq signal around the Rx2 locus showing the normalized read coverage for each one of the replicates of Rx2 and GFP fusions and the log2 fold change of the average signal per replica (track in black at the bottom). The black bar over the RefSeq Genes track represents the previously characterized Rx2 promoter. The area of maximum Rx2 enrichment is included in the 5' end of this promoter.

Figure 3. Optimization of the bacterial Dam gene is necessary for proper expression of Dam fusion proteins, avoiding aberrant splicing. A) Plasmids containing GFP, DG and cMyc\_DG cassettes driven by the 3.5Kb Ubiquitin promoter (Ubi) were co-injected with Tol2 transposase into one cell stage zygotes. Successfully injected larvae expressing EGFP in the heart were selected for further studies. Note that only Ubi::GFP expressed ubiquitously in the body of the larvae. B) RNA was isolated from pools of larvae from the experimental groups. RT-PCR was performed using a forward primer annealing in the non-coding exon included in Ubiquitin promoter (NoE) and the reverse primer in the body of the GFP coding sequence. Proper splicing occurs between NoE and GFP in the Ubi::GFP larvae. In Ubi::Dam\_mGFP larvae, incorrect splicing occurs between the NoE and a cryptic acceptor site in the GFP coding region (red arrowhead). In the Ubi::cMyc\_Dam\_GFP, NoE is spliced to the proper acceptor

upstream of the cMyc sequence, but it is miss-spliced after the cMyc sequence, using a cryptic donor site, to the same cryptic acceptor sequence in the mGFP as for Ubi::Dam\_GFP (see also supplementary Fig.3). The prokaryotic Dam ORF carries a strong splicing potentiator recognized in the eukaryotic context. C) Optimization of the Dam ORF removed this potential, facilitating proper expression of the fusion proteins.

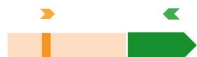




**A****B****C**

**A**

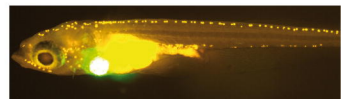
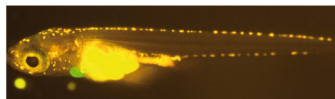
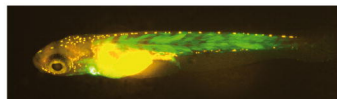
Ubi::GFP



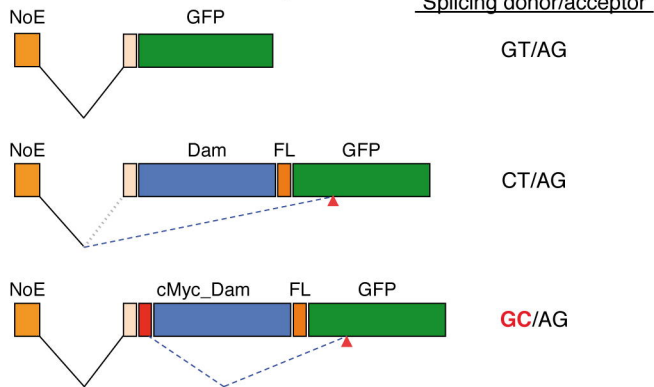
Ubi::Dam\_FL\_GFP



Ubi::cMyc\_Dam\_FL\_GFP



RNA isolation/RT-PCR

**B****C**