Zebrafish Transcription Factor ORFeome for Gene Discovery and Regulatory Network Elucidation

Justin King,^{1,*} Justin Foster,^{1,†} James M. Davison,^{2,‡}
John F. Rawls,² and Ghislain Breton¹

FIG. 1. (A, B) ChIP-seq is a TF-centric approach identifying DNA interaction sites for a single TF. Three promoters (genes 1–3) and 9 different TFs (TF1–9) are represented. *Arrows* represents TF–promoters interactions identified with the ChIP-seq method. (B) eY1H is a gene-centric approach identifying interactors for a single DNA fragment. Same representation as A except *arrows* represent promoter–TFs interactions identified with the eY1H method. (C) Workflow for Y1H and TF ORFeome. DNA fragment is cloned into reporter plasmid and subsequently integrated into yeast genome. Mating between the reporter strain and the ORF collection in compatible yeast is performed overnight. Diploid yeast is enriched using growth in selection media. At this point, the interaction can be revealed using direct reading substrate such as luciferin, or the reporter enzyme such as β -galactosidase can extract and quantify X-gal. ChIP, chromatin immunoprecipitation; ORF, open reading frame; TF, transcription factor; Y1H, yeast one hybrid.

¹Department of Integrative Biology and Pharmacology, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, Texas.

²Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, North Carolina.

*Current affiliation: Ascend Performance Materials, Houston, Texas.

†Current affiliation: Harris County Institute of Forensic Science, Houston, Texas.

‡Current affiliation: Department of Anesthesiology, Duke University School of Medicine, Durham, North Carolina.

Abstract

The completion of the zebrafish genome sequence and advances in miniaturization and multiplexing were essential to the creation of techniques such as RNA-seq, ChIP-seq, and high-throughput behavioral and chemical screens. Multiplexing was also instrumental in the recent enhancement of the classic yeast one-hybrid interaction techniques to provide unprecedented discovery capabilities for protein–DNA interactions. Unfortunately its use for zebrafish research is currently hampered by the lack of an open reading frame (ORF) clone collection. As a first step toward a complete collection, we describe a small library of transcriptional regulatory proteins comprising 142 ORFs and its potential applications.

Keywords: ORFeome, eY1H, yeast one hybrid

THE ZEBRAFISH GENOME CONTAINS 3326 unique genes encoding proteins with domains associated with transcriptional regulation.^{1–3} The elucidation of their role in complex regulatory landscapes composed of multiple factors acting on redundant enhancers will require an array of experimental approaches. Classic techniques such as chromatin immunoprecipitation (ChIP) and yeast one hybrid (Y1H) are often used to determine the interaction between a transcription factor (TF) and a putative cis-regulatory element.^{4–7} They were both recently enhanced into diametrically opposite applications: ChIP-seq and eY1H (Fig. 1A, B).

ChIP coupled with high-throughput sequencing (ChIP-seq) surveys the genomic landscape for a specific TF and was instrumental in revealing the continuous nature of most transcriptional networks^{8,9} (Fig. 1A). Depending on the TF, the genomic sites identified by ChIP-seq are numbered in the hundreds to the tens of thousands^{10,11} with a wide dynamic range of binding occupancy.⁹ Undoubtedly, the genome-wide perspective of ChIP-seq provides a more accurate representation of the redundant nature of cis-elements in proximal promoters and distal enhancers than a single TF/DNA approach.¹²

Although ChIP-seq is an extremely powerful technique, it is not well suited to survey the 3326 regulators on a selected promoter or enhancer. How many different TFs are involved in the regulation of a specific gene of interest? Extrapolation from available ChIP-seq data sets suggests that as many as 50 to 500 different TFs may regulate the expression of a single gene.^{10,11} After the identification of a regulatory region, the use of enhanced yeast one hybrid (eY1H) or high-throughput Y1H (HT-Y1H) provides the means to identify which ~50 TFs are potential regulators^{13,14} (Fig. 1B). Using a DNA fragment upstream of an enzymatic reporter such as β -galactosidase or cell-surface luciferase integrated in the yeast genome,^{14,15} a collection of TF open reading frames (ORFs) can be directly assayed for reporter activation. In follow-up experiments, this approach can be paired with reverse genetics for gene discovery or used on multiple DNA fragments in parallel for gene network reconstruction.^{16–18}

Use of ORFs (start to stop codons) provides advantages compared with full-length libraries^{1,19} because it eliminates additional amino acids that otherwise would be included in the Y1H fusion protein by noncoding regions. In addition, being restricted to the open reading frame guarantees the correct frame of translation for all clones. Open source collections covering several TF families, such as the present collection and the Arabidopsis TF ORFeome,²⁰ have the advantage of being usable in very different fields of investigation.^{20–24} In addition to providing a reliable copy of each clone, a gold standard ORFeome (fully sequenced) can be used to perform a comprehensive TF–TF interaction screen and subsequently build a second library of TF pairs to perform eY1H using heterodimers potentially harboring different DNA specificities.^{25,26}

We identified several diurnally rhythmic TF transcripts in adult zebrafish liver² and established a cloning strategy to generate a small TF ORFeome. To take advantage of the insert size flexibility of Gibson cloning technology,²⁷ we engineered a vector (pENTR/D-A/eI) (Supplementary Fig. S1; Supplementary Data are available online at www.liebertpub.com/zeb), allowing the production of Gateway™ compatible clones

identical to the commonly used directional TOPO™ cloning (ThermoFisher Scientific). After digestion of two flanking *AleI* sites, the linear vector is compatible with the Gibson protocol, assuming compatible ends are being used (see protocol in Supplementary Data). A library of 142 TF ORFs were generated in pENTR and entirely sequenced using internal primers when necessary. Clones were then transferred using Gateway LR clones™ (ThermoFisher Scientific) into the eY1H-compatible pDEST22 and transformed into yeast (Supplementary Table S1). After a 1 day mating procedure with compatible yeast harboring the DNA fragment of interest, this unique collection can be used to identify novel TF regulators using well-established yeast-based protocol from our laboratory and others (Fig. 1C).^{28,29} We recently used this collection to identify, in an unbiased manner, a novel and important role for the hepatocyte nuclear factor 4 family of TFs in mediating transcriptional responses to commensal microbiota.³⁰ We believe that considering the complexity of vertebrate transcriptional networks and the advantages provided by the zebrafish model, this and subsequent versions of the zebrafish TF ORFeome will likely provide invaluable tools for network reconstruction.³¹

Disclosure Statement

No competing financial interests exist.

References

1. Armant O, März M, Schmidt R, *et al.* Genome-wide, whole mount in situ analysis of transcriptional regulators in zebrafish embryos. *Dev Biol* 2013;380:351–362.
2. Boyle G, Richter K, Priest HD, *et al.* Comparative analysis of vertebrate diurnal/circadian transcriptomes. *PLoS One* 2017;12:e0169923.
3. Howe K, Clark MD, Torroja CF, *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 2013;496:498–503.
4. Gilmour DS, Lis JT. In vivo interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol Cell Biol* 1985;5:2009–2018.
5. Gilmour DS, Lis JT. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci USA* 1984;81:4275–4279.
6. Li JJ, Herskowitz I. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* 1993;262:1870–1874.
7. Wang MM, Reed RR. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* 1993;364:121–126.
8. Johnson DS, Mortazavi A, Myers RM, *et al.* Genome-wide mapping of in vivo protein-DNA interactions. *Science* 2007;316:1497–1502.
9. Biggin MD. Animal transcription networks as highly connected, quantitative continua. *Dev Cell* 2011;21:611–626.
10. Wang J, Zhuang J, Iyer S, *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* 2012;22:1798–1812.
11. MacQuarrie KL, Fong AP, Morse RH, *et al.* Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* 2011;27:141–148.
12. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 2012;13:613–626.
13. Hens K, Feuz J-D, Isakova A, *et al.* Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nat Methods* 2011;8:1065–1070.
14. Reece-Hoyes JS, Diallo A, Lajoie B, *et al.* Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat Methods* 2011;8:1059–1064.
15. Bonaldi K, Li Z, Kang SE, *et al.* Novel cell surface luciferase reporter for high-throughput yeast one-hybrid screens. *Nucleic Acids Res* 2017;45:e157.
16. Gaudinier A, Zhang L, Reece-Hoyes JS, *et al.* Enhanced Y1H assays for Arabidopsis. *Nat Methods* 2011;8:1053–1055.
17. Taylor-Teeple M, Lin L, de Lucas M, *et al.* An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 2015;517:571–575.
18. Fuxman Bass JI, Sahni N, Shrestha S, *et al.* Human gene-centered transcription factor networks for enhancers and disease variants. *Cell* 2015;161:661–673.
19. MGC Project Team, Temple G, Gerhard DS, *et al.* The completion of the Mammalian Gene Collection (MGC). *Genome Res* 2009;19:2324–2333.

20. Pruneda-Paz JL, Breton G, Nagel DH, *et al.* A genome-scale resource for the functional characterization of Arabidopsis transcription factors. *Cell Reports* 2014;8:622–632.
21. Niwa M, Daimon Y, Kurotani K-I, *et al.* BRANCHED1 interacts with FLOWERING LOCUS T to repress the floral transition of the axillary meristems in Arabidopsis. *Plant Cell* 2013;25:1228–1242.
22. Li L, Ljung K, Breton G, *et al.* Linking photoreceptor excitation to changes in plant architecture. *Genes Dev* 2012;26:785–790.
23. Guan P, Wang R, Nacry P, *et al.* Nitrate foraging by Arabidopsis roots is mediated by the transcription factor TCP20 through the systemic signaling pathway. *Proc Natl Acad Sci USA* 2014;111:15267–15272.
24. Zhou Y, Liu X, Engstrom EM, *et al.* Control of plant stem cell function by conserved interacting transcriptional regulators. *Nature* 2015;517:377–380.
25. Grove CA, De Masi F, Barrasa MI, *et al.* A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell* 2009;138:314–327.
26. Trigg SA, Garza RM, MacWilliams A, *et al.* CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat Methods* 2017;333:601.
27. Gibson DG. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* 2011;498:349–361.
28. Fuxman Bass JI, Reece-Hoyes JS, Walhout AJM. Gene-centered yeast one-hybrid assays. *Cold Spring Harb Protoc* 2016;2016:pdb.top077669.
29. Breton G, Kay SA, Pruneda-Paz JL. Identification of Arabidopsis transcriptional regulators by yeast one-hybrid screens using a transcription factor ORFeome. *Methods Mol Biol* 2016;1398:107–118.
30. Davison JM, Lickwar CR, Song L, *et al.* Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor Hepatocyte nuclear factor 4 alpha. *Genome Res* 2017;27:1195–1206.
31. Sivasubbu S, Sachidanandan C, Scaria V. Time for the zebrafish ENCODE. *J Genet* 2013;92:695–701.

Address correspondence to:

Ghislain Breton, PhD

Department of Integrative Biology and Pharmacology

McGovern Medical School

University of Texas Health Science Center at Houston

6431 Fannin Street, MSB 4.216

Houston, TX 77030

E-mail: ghislain.breton@uth.tmc.edu