

Investigating the molecular guts of endoderm formation using zebrafish

Daniela M. Figiel, Randa Elsayed and Andrew C. Nelson

Corresponding author: Andrew C. Nelson, School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry CV4 7AL, UK.
Tel.: +44 (0) 24 765 74523; E-mail: A.Nelson.1@Warwick.ac.uk

Abstract

The vertebrate endoderm makes major contributions to the respiratory and gastrointestinal tracts and all associated organs. Zebrafish and humans share a high degree of genetic homology and strikingly similar endodermal organ systems. Combined with a multitude of experimental advantages, zebrafish are an attractive model organism to study endoderm development and disease. Recent functional genomics studies have shed considerable light on the gene regulatory programs governing early zebrafish endoderm development, while advances in biological and technological approaches stand to further revolutionize our ability to investigate endoderm formation, function and disease. Here, we discuss the present understanding of endoderm specification in zebrafish compared to other vertebrates, how current and emerging methods will allow refined and enhanced analysis of endoderm formation, and how integration with human data will allow modeling of the link between non-coding sequence variants and human disease.

Key words: zebrafish; endoderm; development; chromatin; genomics

Introduction

A central question of biology is how a single fertilized egg can give rise to the many distinct cell types within a fully developed organism. By investigating this, we gain valuable knowledge of core principles governing metazoan development, and key distinctions that have driven speciation. In understanding how normal development occurs in model organisms, we gain a valuable blueprint allowing investigation of how genetic influences drive human developmental disorders.

Multiple classical model species have made invaluable contributions to our understanding of cell fate and organismal patterning, such as the fly *Drosophila melanogaster*, nematode *Caenorhabditis elegans*, rodents *Mus musculus* and *Rattus norvegicus* and frogs *Xenopus laevis* and *Xenopus tropicalis*. Each of these organisms gained popularity through their unique practical attributes and experimental tractability, each offering different advantages and drawbacks. Zebrafish is an increasingly

attractive model organism for studying vertebrate development and disease. Among its attributes are development external to the mother, large clutch sizes providing an abundance of embryonic material, optical transparency for ease of imaging, and also increasing genetic tractability. The provision of a fully-sequenced and well-annotated genome has allowed zebrafish to transition from an outstanding organism for studying vertebrate gene function [1, 2] to an ideal model organism for studying gene regulatory processes underlying cell fate decisions during embryogenesis [3]. Strikingly, 70% of human genes and 82% of known disease-related genes have a recognizable zebrafish homolog [4]. Zebrafish is consequently an increasingly popular model organism for modeling human genetic disease. Some other model organisms share key advantages with zebrafish, such as *ex utero* development and ease of microinjection in *Xenopus*. Others offer specific advantages over zebrafish, such as mice for which many relevant cell lines can be easily cultured and manipulated. Notably, both mouse and *Xenopus* are also

Daniela M. Figiel holds MSc degrees from the University of Nottingham and University of Warwick, UK and is undertaking PhD studies through the Medical Research Council Doctoral Training Partnership in Interdisciplinary Biomedical Research at Warwick Medical School.

Randa Elsayed holds MSc degrees from the King's College London and University of Warwick, UK and is undertaking PhD studies through the Medical Research Council Doctoral Training Partnership in Interdisciplinary Biomedical Research at Warwick Medical School.

Andrew C. Nelson is an associate professor at University of Warwick, UK. The Nelson lab studies chromatin-level control of cell fate and behavior with a particular focus on endoderm formation in zebrafish, and trophoblast differentiation during murine placenta development.

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

more similar to humans in morphological and evolutionary terms. Nevertheless, the unique combination of attributes zebrafish possess demands their use in studies designed to best exploit their advantages.

The endoderm is the innermost germ layer of the embryo proper, and makes major contributions to the respiratory and gastrointestinal tract and all associated organs including liver, pancreas, thyroid [5] and the pituitary gland [6]. Given the prevalence of developmental and metabolic disorders afflicting these organs, both the study of endoderm formation and function in zebrafish has clear relevance [7]. Here, we discuss use of functional genomics to understand gene regulatory control of zebrafish endoderm formation, similarities and evolutionary differences in vertebrate endoderm specification, and the future prospects for the application of zebrafish and functional genomics to advance our understanding of endoderm formation and human disease.

Part 1: endoderm specification—the more things change, the more they stay the same

The Nodal signaling pathway is necessary and sufficient to induce endoderm and mesoderm fates in all vertebrates, as discussed extensively elsewhere [8–11]. The process leading to endoderm specification therefore begins with maternally contributed factors that control Nodal induction and ends with stable expression of the master regulator of zebrafish endoderm fate—*sox32*. In this section, we discuss the current knowledge of transcriptional control of Nodal induction through to stable specification of endoderm.

Transcriptional control of Nodal induction

Endoderm specification occurs at the onset of gastrulation, predominantly in the first two cell tiers above the blastoderm margin [12, 13]. In zebrafish, there are three Nodal-related genes compared to one in mice and humans, two of which (*ndr1/sqt* and *ndr2/cyc*) participate in mesoderm induction. Binding of these TGF β family ligands to their cognate heterodimeric transmembrane receptors leads to intracellular phosphorylation and release of receptor-tethered effectors Smad2/3. Smad2/3 then associate with Smad4 and translocate into the nucleus where they associate with a range of other transcription factors at Nodal-responsive cis-regulatory modules (CRMs) to induce endoderm and mesoderm fates [8]. While maternal *ndr1* mRNA is detectable within eggs and its localization predicts the embryonic dorsal axis by the four-cell stage, it is translationally repressed (reviewed [10, 14]) and zygotically expressed *ndr1/2* are critical for endoderm induction.

Though Nodal signaling is key to endoderm specification, the upstream mechanisms that induce Nodal have undergone a degree of diversification during vertebrate evolution. A notable distinction between zebrafish and amniotes is the role of the yolk syncytial layer (YSL)—a teleost-specific structure suggested to be equivalent to the mouse primitive endoderm, which is the initial source of the endoderm-inducing Nodal activity. The YSL forms from collapse of marginal blastomeres into the yolk cell between the 9th and 10th cell cleavages, thus forming a syncytium that underlies the future endoderm (reviewed in [15]). Induction of *ndr1/2* expression at the blastoderm margin is controlled by multiple maternally supplied factors including β -catenin and transcription factors Nanog and Eomesodermin homolog A (Eomesa). Dorsal expression of *ndr1/2* and thus dorsal organizer formation is controlled by β -catenin [16–18],

though a direct requirement for β -catenin in zebrafish endoderm formation is yet to be established as we discuss later. Broader expression of *ndr1/2* is controlled during blastula stages by Nanog and Eomesa, which directly activate expression of *mxtx2*—a key regulator of YSL formation (Figure 1A) [19–22]. Eomesa and *Mxtx2* themselves directly bind *ndr1/2* to induce their expression [20, 22]. Of particular note is that our knowledge of the Eomesa/Nanog/*Mxtx2* regulatory arc upstream of *ndr1/2* was gained by exploiting the ability to express affinity-tagged transcription factors for chromatin immunoprecipitation with sequencing (ChIP-seq) analysis during early embryogenesis through microinjection of encoding mRNA, thus mitigating the paucity of commercial antibodies recognizing zebrafish proteins. Exploiting the *ex utero* development of zebrafish, fecundity and ease of microinjection to generate abundant embryonic material for ChIP-seq analysis during early development is virtually unparalleled and a key benefit of zebrafish.

Nanog is most notable for its role in mammalian embryonic stem cells, where it works alongside Pou5f1/Oct4, Sox2 and others to orchestrate the transcriptional network maintaining pluripotency [23]. In human embryonic stem cells (hESCs), differentiation towards definitive endoderm is initiated by positive regulation of EOMES expression by NANOG [24]. In zebrafish, Nanog functions alongside and directly regulates the closest fish homologs of Pou5f1 and Sox2 to control zygotic genome activation, while Nanog and Pou5f3 (the closest zebrafish homolog of Pou5f1) also control later aspects of endoderm specification, as we discuss later. Consequently, the role of Nanog is much broader than in endoderm induction alone. Strikingly, though Nanog has roles in YSL formation and in the equivalent mouse tissue, mechanistically these roles appear divergent since Nanog appears to non-autonomously control mouse primitive endoderm formation [25, 26] while zebrafish YSL formation is cell autonomously controlled via *Mxtx2*, which lacks a close mammalian homolog [20]. Similarly, Eomes has pivotal roles in endoderm formation in mice including in collaboration with Nodal, but acts later in the epiblast rather than having maternal functions in endoderm induction [27]. Consequently, initial activation of zygotic Nodal expression in zebrafish appears to be through divergent mechanisms driven by key conserved factors, rather than highly conserved mechanisms.

From Nodal induction to *sox32* expression

Ndr1/2 are transcriptional targets of Nodal, therefore, initial expression in the YSL spreads to overlying marginal blastomeres forming a spatiotemporal gradient of nuclear Smad2/3 in the five tiers overlying the YSL [28, 29]. Interestingly, Nodal signals from the YSL also appear to regulate endoderm specification in marginal blastomeres by driving asymmetric nuclear localization independent of Smad2/3-mediated transcription via JNK, and promoting Smad2 nuclear translocation [30]. Both ChIP-seq and ChIP-chip approaches have been successfully applied to explore Smad2 targets at blastula and early gastrula stages, respectively [31, 32]. Strikingly, comparison of Smad2 ChIP-seq targets in zebrafish with data from *Xenopus* embryos and differentiated human and mouse embryonic stem cells suggests broad conservation of Nodal target genes in early embryogenesis, validating use of zebrafish to explore core principles of vertebrate germ layer formation [32–35].

Smad2/3 genomic target site selection is in part dictated through physical interaction with sequence-specific transcription factors. Among the core Smad2/3 interacting factors across vertebrate species are Foxh1 and Eomes. There are two

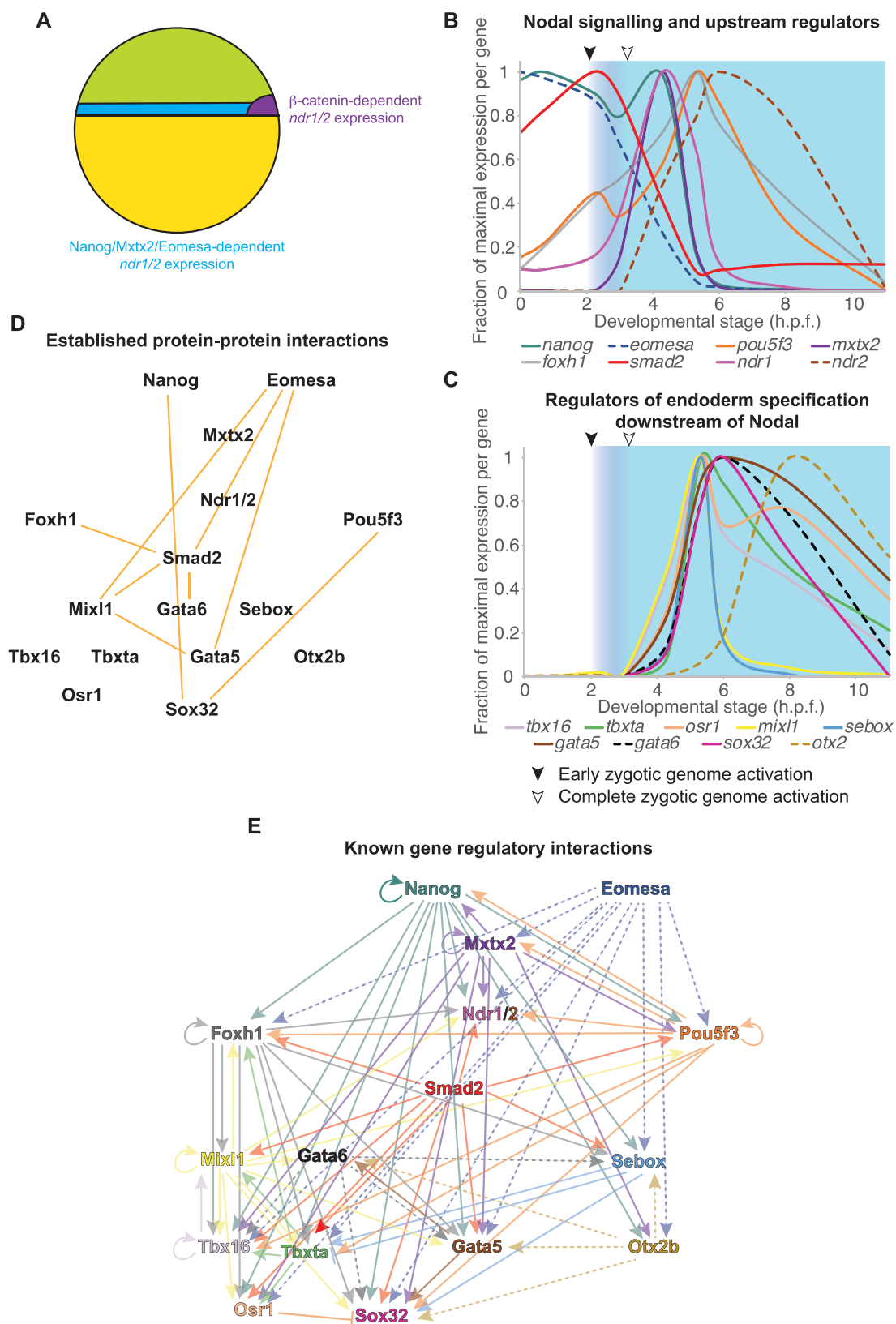


Figure 1. Expression and regulation of key genes during zebrafish endoderm specification. (A) Schematic of *ndr1/2* expression domains and their key regulators in late blastula stage embryos (labeled in purple and blue). The yolk cell is shown in yellow and *ndr1/2* negative domain in green. Dorsal is to the right. (B) RNAseq gene expression profiles of maternally contributed transcripts, *ndr1/2* and their upstream activators implicated in endoderm formation between fertilization and the end of gastrulation using data from White et al. [142]. Timepoints corresponding to the onset and complete zygotic genome activation are indicated. (C) As panel B indicating the expression of transcription factors downstream of Nodal implicated in endoderm specification. (D) Schematic indicating the known physical interactions between transcription factors implicated in zebrafish endoderm specification as discussed in the text. All factors present in panels B and C are shown. (E) Schematic indicating the likely regulatory relationships between factors present in panels B and C, based on ChIP-seq and/or genetic evidence from the cited literature.

paralogous Eomes genes in zebrafish owing to duplication of the ancestral gene. *Eomesb* is not expressed during early development and does not appear to participate in endoderm specification [32]. However, both *Eomesa* and *Foxh1* are maternally contributed in zebrafish (Figure 1B) [36, 37]. ChIP-seq analyses at blastula stages strongly suggest occupancy of *Eomesa* at Smad2-bound sites is a hallmark of Nodal-responsive genes, highlighting a key parallel with human endoderm [24, 32]. Moreover, these sites in zebrafish are enriched for the known *Foxh1* binding motif, and a more recent study exploiting microinjection of *foxh1-flag* mRNA followed by ChIP-seq further confirmed the co-occurrence of Smad2 and *Foxh1* at Nodal responsive sites [28], as observed in cell culture experiments [35].

At the onset of zygotic genome activation, Smad2/3-containing complexes initiate expression of various transcription factor genes (Figure 1C), the products of which form complexes with Smad2/3 as part of a feedforward mechanism. Notable examples include *Gata6* which has been shown to interact with Smad2 in zebrafish [31], *Mixer/Mixl1* in *Xenopus* [38], while other factors such as *Gata5* are likely to co-occupy complexes with Smad2 given they bind Smad2-interacting factors *Eomesa* and *Mixl1* (Figure 1D) [39]. Each of *Gata6*, *Mixl1* and *Gata5* have individually been shown to influence zebrafish endoderm master regulator *sox32* and other upstream factors (Figure 1E) [39–43]. However, *gata6* expression is not completely Nodal-dependent since, though expression in marginal blastomeres is lost in Nodal mutants, YSL expression is not [44].

While *mixl1* (also known as *bon* in zebrafish) has a conserved role in endoderm specification in vertebrates [45], in zebrafish another *Mix* family member *sebox* (also known as *mezzo* and *og9x*) acts in parallel with *mixl1*. *Sebox* is also a target of Nodal signaling and exhibits partial redundancy with *mixl1*, however, it also appears to function in mesoderm induction [46]. Though an identifiable *sebox* homolog exists in mammals and *Xenopus*, function during endoderm specification seems to be unique to fish [45, 47].

In zebrafish, *Foxh1* and *Eomesa* are individually dispensable for endoderm formation, with *eomesa* mutants exhibiting only minor loss of early endoderm marker expression [48, 49]. However, *Foxh1* and *Eomesa* are collectively required for endoderm formation, evidenced both through expression of a dominant negative *Eomesa* in maternal zygotic (MZ) *foxh1* mutants, and *Foxh1* knockdown in MZ*eomesa* mutants [32, 49]. This is in stark contrast to mammals where *Eomes* is individually required for definitive endoderm formation [24, 27]. However, recent analyses have revealed the likely reason for the different severities of *Eomes* mutant endoderm defects between zebrafish and mouse. *Eomes* is a member of the T-box family of transcription factors, which has undergone changes in complement in different lineages during vertebrate evolution [50]. Notably, this includes *tbx16* which is present in fish, frogs (where it is known as *vegt*), birds, marsupials and monotremes but lost in placental mammals [51]. Functional genomics analyses of *Eomesa* and *Tbx16* in zebrafish blastulas and gastrulas, respectively, suggests similar genomic target site selection, including at key regulator of endoderm formation *mixl1* [32, 52]. Expression of several mouse *Eomes* isoforms in zebrafish through mRNA microinjection also suggests functional conservation of *Eomes* across vertebrate evolution, while knockdown of *Tbx16* in *eomesa* mutants leads to profound loss of *mixl1* [53]. The different severities of endoderm phenotypes observed between mouse and zebrafish may therefore be due to T-box factor functional redundancy in

zebrafish rather than differences in *Eomes* molecular function. Given the loss of *tbx16* expression seen in *foxh1* mutants, it also seems likely that the loss of endoderm observed on abrogation of both *Eomesa* and *Foxh1* function is partially attributable to loss of *tbx16* [49]. The consistencies between zebrafish endoderm formation with that of other vertebrates, and the conservation of the Nodal transcriptional network therefore suggest that zebrafish is an ideal model organism to advance our knowledge of endoderm formation.

As implied above and summarized in Table 1, while a similar complement of transcription factors appears to control endoderm specification in all vertebrates, there are some differences in either the temporal expression of these factors or the alteration of the factors involved within gene families. For example, while functionally important *Eomesa* is maternally contributed in zebrafish, in other vertebrates it is zygotically expressed [27, 48, 54]. Similarly, recent evidence suggests that in *Xenopus* maternally contributed *Otx1*, *Vegt* and *Foxh1* orchestrate endoderm formation through coordinated binding to selected regulatory elements [55]. In contrast, in fish, the *vegt* orthologue *tbx16* is zygotically expressed in response to Nodal, and there is no evidence for *otx1* involvement in endoderm specification, though *otx2b* has been implicated [40]. Interestingly, in mice, *Otx1* and *Otx2* exhibit a degree of functional equivalence in neuroectoderm formation [56]. While not directly tested, it therefore seems likely that *otx1* and *otx2* orthologues may have similar functional potential with respect to endoderm formation, and that these evolutionary substitutions are effectively neutral. Furthermore, two-hybrid screening has suggested physical interaction between mammalian OTX2 and MIXL1 [57], raising the possibility that involvement of *otx2b* and *mixl1* in zebrafish endoderm formation may be conserved mechanism through direct physical association at target loci.

A notable difference between teleost and mammalian endoderm formation is the role of the zinc finger transcription factor *Osr1*, which is induced by Nodal [32] and limits endoderm differentiation from germ ring mesendoderm in zebrafish [58], though a similar role during mammalian endoderm induction has not been revealed.

On Sox32 and the establishment of endoderm fate

Integrated analyses of ChIP-seq and expression data in the early embryo, as well as classical analyses have revealed a complex gene regulatory network controlling *sox32* expression (discussed in [52]). Though *sox32* is a fish-specific member of the SoxF subgroup of SOX family transcription factors, it appears to have arisen from a small-scale tandem duplication of an ancestral gene hence it is adjacent to *sox17* in the zebrafish genome [59]. *Sox17* is a critical regulator of endoderm fate in other vertebrates, and capable of inducing endoderm fate on overexpression in embryonic stem cells [60, 61]. However, in zebrafish, *sox32* appears to act upstream of *sox17*, and *sox32* expression is sufficient to drive endoderm fate [62–64]. Furthermore, ChIP-seq analyses suggest the regulatory inputs into zebrafish *sox32* reflect those of human SOX17, evidenced by proximal binding of *Eomes* and Smad2 in both species [24, 52, 65]. *Sox32* therefore appears to have adopted the location of ancestral *Sox17* in the regulatory hierarchy. Importantly, while loss of *Sox17* in mice leads to substantial reduction of endoderm [66], *Sox17* depletion in zebrafish leads to laterality defects but not dramatic loss of endoderm [67]. Conversely, *Sox17* overexpression, while rescuing knockdown-induced laterality defects did not lead to increased pancreatic endoderm [67].

Table 1. Transcription factors implicated in zebrafish endoderm specification

Zebrafish factor	Role	Maternal	Early zygotic expression	ChIP-chip or ChIP-seq in zebrafish	Implicated in endoderm formation in:	
					<i>Xenopus</i>	Mammals
Nanog	Activates <i>mxtx2</i> expression to induce YSL formation and Nodal induction	Y	Y	Y	NA	Y
Pou5f3	Interacts with Sox32 and required for Sox32-mediated endoderm induction	Y	Y	Y	NA	Y
Mxtx2	Induces YSL formation and directly activates <i>ndr1/2</i>	N	Y	Y	NA	NA
Smad2	Main effector transcription factor activated by Nodal to induce Nodal target genes	Y	Y	Y	Y	Y
Foxh1	Constitutive Smad2 interactor redundantly required for endoderm gene expression with Eomesa	Y	Y	Y	Y	Y
Eomesa	Activates <i>mxtx2</i> expression and interacts with Smad2, Mixl1 and Gata5 to drive endoderm gene expression	Y	Y	Y	Y	Y
Osr1	Negative feedback regulator of Nodal-induced endoderm development	N	Y	N	N	N
Tbxta	Partially redundant with Tbx16 in regulating <i>mixl1</i> expression and thus endoderm induction	N	Y	Y	N	Y
Tbx16	Directly activates <i>mixl1</i> redundantly with Tbxta and Eomesa to promote endoderm fate	N	Y	Y	Y	NA
Otx2b	Regulator of transcription factor expression downstream of Nodal in endoderm specification	N	Y	N	Y ^a	Y ^b
Gata5	Interacts with Mixl1 and Eomesa to drive <i>sox32</i> expression	N	Y	N	Y	N
Gata6	Regulator of transcription factors downstream of Nodal in endoderm specification	N	Y	N	Y	Y
Mixl1	Interacts with Gata5, Smad2 and Eomesa to drive <i>sox32</i> expression. Regulates other genes implicated in endoderm formation including <i>dusp4</i> .	N	Y	Y	Y	Y
Sebox	Nodal induced regulator of <i>sox32</i> , partially redundant with Mixl1	N	Y	N	N	N
Sox32	Master regulator necessary and sufficient for zebrafish endoderm fate	N	Y	N	NA	NA

Note: other transcription factors implicated in endoderm formation are expressed in endoderm progenitors during gastrulation, but do not have identified/characterized phenotypes prior to segmentation stages: e.g. *foxa2*, *foxa3*, *foxd3*, *etv5a*, *cdx4*, *grhl3*, *her5*, *lmo4a*, *raraa*, *rarab*, *mx2b*, *fhl1b*, *tp53*. NA—No clear homolog is present.

^aOtx1 is implicated in endoderm formation in *Xenopus*.

^bOtx2 appears to act in the visceral endoderm in mouse rather than definitive.

How Sox32 acts to specify endoderm is not understood, however, there are key indicators suggesting significant parallels with mammalian and *Xenopus* Sox17. A key feature of SoxF transcription factors is conservation of a short peptide (EFE/DQYL) in the C-terminal transactivation domain which interacts with β -catenin [64, 68]. Indeed, a recent ChIP-seq study showed *Xenopus* Sox17 co-occupies Wnt-responsive endoderm CRMs with β -catenin [69]. Notably, though zebrafish *sox17* shows markedly higher conservation with Sox17 orthologues compared to *sox32*, the β -catenin interacting peptide shows greater conservation in *sox32* than *sox17* [64, 68]. Moreover, deletion of this peptide abrogates Sox32 ability to induce ectopic endoderm [68, 70]. However, though β -catenin is required to induce dorsal *ndr1/2*

as discussed above, no evidence for β -catenin requirement in conjunction with Sox32 has yet been presented. Zebrafish have two β -catenin genes—*ctnnb1* and *ctnnb2*, exhibiting considerable co-expression and sharing >92% sequence identity. It is therefore possible they have redundant function and perturbation of both may be needed to reveal a requirement in Sox32-mediated endoderm induction.

Sox32 also physically interacts with Pou5f3, the closest zebrafish homolog of mammalian Pou5f1/Oct4 (Figure 1D) [70, 71]. Quantitative imaging approaches have also revealed that Sox32 can associate with mouse Oct4 [72], mirroring interaction between Sox17 and Oct4 observed in mammals [73]. Maternal Pou5f3 is required for endoderm formation, and Sox32

overexpression in maternal zygotic Pou5f3 (*MZspg*) mutants fails to result in endoderm induction suggesting Sox32–Pou5f3 interactions are critical to endoderm fate [74, 75]. Furthermore, *Pou5f1* mRNA microinjection can rescue the loss of endoderm in *MZspg* mutants [76], though *pou5f3* cannot rescue mouse *Pou5f1* mutant embryonic stem cell self-renewal [77]. Taken together, our molecular and functional knowledge of zebrafish *sox32* and *sox17* strongly suggest that in teleosts *sox32* is the *de facto* equivalent of Sox17 in other vertebrates.

An outstanding question is precisely how Pou5f3 mediates Sox32 function and what the chromatin targets of Sox32–Pou5f3 complexes are likely to be. While key clues may be present in existing Pou5f3 ChIP-seq datasets, it is notable that widespread changes are observed in Pou5f3 genomic binding between 512 cell and late blastula stages [78]. Pou5f3 target site selection during endoderm formation is likely to be dependent on Sox32, which will only be present in a minority of Pou5f3+ cells contributing to existing ChIP-seq data [78]. Furthermore, use of heatshock-inducible dominant negative Pou5f3 indicates stage-specific pleiotropic effects on mesendoderm formation [79]. That Pou5f3 target genes appear to be timepoint- and cell type-specific means effectively exploring Pou5f3 targets during endoderm formation will require further functional genomics analysis in a pure endoderm population. However, this should be achievable by exploiting the potency of Sox32-mediated endoderm induction in zebrafish. Thus, co-injection of mRNAs encoding affinity tagged Sox32 and Pou5f3 would allow cell type-specific ChIP-seq interrogation of target genes during endoderm formation.

It is notable that Sox32 can also physically interact with Nanog *in vivo*, disrupting Nanog–Pou5f3 interactions (Figure 1D) [72]. So far the functional importance of this has been considered in terms of spatially restricting Nanog–Pou5f3 complexes to the ventrolateral mesendoderm, rather than whether Sox32–Nanog complexes have direct functional importance [72]. Analysis of *Mnanog* and *MZnanog* mutants by Veil *et al.* [80] revealed that endoderm marker expression is substantially reduced during gastrulation compared to wild type animals. However, Veil *et al.* also confirmed the loss of *mxtx2* previously observed in morpholino knockdown experiments, therefore, it is currently unclear whether reduction in endoderm in *MZnanog* mutants is solely attributable to loss of upstream regulators of Nodal, or a combinatorial role with Sox32 [20, 80]. A recent study has revealed a role for Nanog in attenuating β -catenin-mediated transcriptional programs by competing for binding of β -catenin co-factor TCFs [81]. It is tempting to speculate that the role of Sox32–Nanog complex formation is to similarly attenuate Sox32– β -catenin interactions or Sox32–Pou5f3 interactions, allowing Nanog to both induce endoderm via the *Mxtx2*/Nodal pathway and restrict endoderm formation by disrupting Sox32-containing complexes. Such a model, however, requires further investigation including through mapping of Sox32–Nanog target sites, which could be achieved as proposed for Pou5f3 above.

In summary, while many of the transcription factors involved in endoderm formation in zebrafish and mammals are the same, there are nevertheless evolutionary changes in the complement of factors involved or the spatiotemporal expression patterns of common factors. The majority of these changes, however, appear to be largely neutral—either representing functional substitution or additional levels of redundancy. This is supported by studies exploiting functional genomics analyses and serves to highlight that evolutionary differences in the transcriptional programs leading to endoderm specification essentially cement the status quo.

Part 2: key challenges and future prospects for studying zebrafish endoderm formation

One of the major challenges of functional genomic and transcriptomic analyses of endoderm formation is that the cells of interest constitute a minor proportion of the embryo. As an indicative estimate, in recent single-cell RNA-seq analyses of the early embryo $3 \pm 1.4\%$ of cells at the six developmental stages profiled between specification at ~ 6 h post-fertilization (h.p.f.) and the onset of organogenesis at 24 h.p.f. were annotated as endoderm [82]. Consequently, endoderm-specific signals are likely to be masked by ensemble averaging across total cell populations in whole embryo analyses typically used to identify functionally active CRMs, such as ChIP-seq analyses of histone modifications or ATAC-seq [83]. As such, though published datasets have been invaluable resources for enhancer analysis during zebrafish embryogenesis [84], there is a clear need for enrichment for cells of interest to further probe the cis-regulatory landscape of endoderm formation. Similarly, while transcription factor ChIP-seq has been invaluable in uncovering the gene regulatory relationships underpinning endoderm specification, the majority of factors discussed above have additional expression domains beyond endoderm. Consequently, it is challenging to definitively ascribe signals to individual cell populations, while key endoderm-specific signals may be absent or below threshold.

Multiple possibilities exist to enrich for endoderm based on expression of specific markers such as *sox17*. The GFP reporter *Tg(sox17:GFP)* is a transgenic line [85, 86] widely used for imaging endoderm in a range of contexts, as well as enrichment by fluorescence activated cell sorting (FACS). Notable examples include FACS enrichment followed by expression analyses to probe mechanisms controlling endoderm migration and proliferation during gastrulation [87, 88], and at later stages to explore the role of prostaglandin E2 in liver versus pancreas specification [89], and sphingosine-1-phosphate/Yap1 signaling in endoderm cell survival and interactions with adjacent cardiac mesoderm [90]. This evidences that a more detailed characterization of endoderm gene expression and the cis-regulatory landscape is achievable using sorted populations combined with approaches such as single-cell RNA-seq and ATAC-seq [91–95]. This would offer great advantages, both in terms of providing a more comprehensive characterization of the endoderm [82, 96, 97], and by allowing exploration of how and when perturbation of known regulators of endoderm fate effect the chromatin landscape and gene expression.

While more complete exploitation of the *Tg(sox17:GFP)* line has the potential to revolutionize investigation of endoderm formation in zebrafish, it also has drawbacks. Since *sox17* is a downstream target of Sox32, reporter expression is not observed prior to endoderm specification. An alternative reporter is therefore required to enrich for presumptive endoderm for analysis of endoderm specification itself. Furthermore, though GFP expression in the endoderm persists in *Tg(sox17:GFP)* fish beyond gastrulation (when endogenous *sox17* is silenced), during segmentation stages *sox17* is appreciably expressed in mesoderm lineages including endothelial cells and hematopoietic precursors [98]. However, all such issues are surmountable using alternative transgenic lines. For example, *Tg(sebox:GFP)* marks the mesendoderm and has been used for imaging studies as early as 4 h.p.f. [99, 100]. It therefore offers a promising option for cell enrichment for detailed study of chromatin-level regulation of endoderm specification. Alternatively, for analysis of endoderm formation during late segmentation stages and beyond,

Tg(sox17:GFP) fish can be crossed with a range of other transgenic reporters to permit selection against contaminating mesoderm. There is also a wealth of fluorescent reporter strains that allow FACS purification of endoderm subpopulations. For example, Lavergne et al. [101] recently exploited the *Tg(pax6b:GFP)* line [102] to isolate and compare wild type and *pax6b* mutant pancreatic endocrine and enteroendocrine cells for RNA analysis. Alvers et al. [103] similarly created and used *TgBAC(cldn15la-GFP)* to enrich and compare wild type and *smo* mutant intestinal cells. Nissim et al. [89] exploited *elastase:GFP* [104] and *lfabp:GFP* [105] reporter fish to FACS purify exocrine pancreas and liver respectively, while Stuckenholz et al. [106] used gutGFP [107] combined with expression microarrays to profile liver, gut and pancreas progenitors across a developmental timecourse. There is therefore clearly scope to perform similar functional genomics profiling on defined sorted cell populations to elucidate the effects of genetic perturbation on chromatin-mediated regulation of gene expression. Conventional functional genomics techniques such as ChIP-seq typically require more input material than can be comfortably acquired by FACS purification of minor cell populations. However, should low-input ChIP-seq methods (e.g. [108, 109]) and emerging techniques such as Cut&Run [110] and CUT&Tag [111] fulfill their early promise, it will be possible to gain a more complete understanding of chromatin-level control of endoderm formation, function and defects.

While FACS-based approaches for cell enrichment offer an attractive method for cell type-specific analyses, they have drawbacks. These include the need for expensive capital equipment and potentially undesirable cell fate/state changes due to removing cells from their normal physiological context and prolonged processing times. However, zebrafish also offers alternative approaches such as by exploiting the Biotagging Toolkit for cell type-specific expression of proteins to biotinylate subcellular compartments such as nuclei for subsequent streptavidin pulldown and analysis. This therefore allows approaches such as ATAC-seq to profile chromatin accessibility in nuclei that have not undergone the prolonged FACS-purification [112, 113]. While great progress has been made in understanding gene regulatory control mechanisms underpinning endoderm specification and differentiation, the range of tools applicable to zebrafish now provide immense scope for both greater refinements and novel investigation of cell fate decisions throughout endoderm development.

Part 3: exploiting zebrafish to model human endoderm disorders and gene regulation

Not only is zebrafish an attractive model for studying fundamental aspects of vertebrate endoderm development, but also for modeling human genetic diseases [114, 115]. Furthermore, it is increasingly realized that many Mendelian and complex genetic disorders are driven by mutations in CRMs such as enhancers, rather than merely in protein coding sequence [116]. Understanding the extent to which cis-regulatory logic is conserved between zebrafish and human will inform what aspects of human genetic disease biology may be reliably modeled in zebrafish, and analysis of how mutations in CRMs influence phenotype. There is consequently a clear imperative both to relate CRMs identified in zebrafish to human genomic data, and to study consequences of human disease-associated mutations in a tractable experimental system such as zebrafish.

While the difficulty of studying minor cell populations can be overcome through enrichment procedures such as FACS and Biotagging, approaches combining the advantages of multiple

model systems offer valuable alternatives. Though one of the strengths of zebrafish is the ease of transgenesis, a relative weakness is our inability to expand and manipulate pluripotent cells *in vitro*. Where the primary motivation is to interrogate broader vertebrate or human biology, it is possible to exploit directed differentiation of hESCs to produce abundant endodermal populations for conventional functional genomics analysis. Since functional analysis of the identified CRMs cannot be performed *in vivo* in human embryos for technical and ethical reasons, transgenic reporter assays in zebrafish are a powerful alternative. Great progress has been made towards refining the efficient and representative application of transgenic reporter assays, such as development of the Tol2 system, targeted integration methods and insulation of transgenes [117–121]. The wealth of zebrafish fluorescent reporter lines also allows precision analysis of reporter expression patterns with reference to known anatomical landmarks (Figure 2).

There are multiple examples of studies where hESCs have been used to produce differentiated endoderm subpopulations combined with functional genomics analysis, e.g. [121–124]. Such studies identified putative enhancers of interest based on ChIP-seq analysis of commonly studied histone modifications such as H3K4me1 and H3K27ac [125, 126]. For example, Loh et al. [123] differentiated hESCs to definitive endoderm, and then to three distinct subpopulations along the anterior–posterior axis of the developing endoderm–anterior foregut (AFG), posterior foregut (PFG) and midgut/hindgut (MHG). Potential use of resulting functional genomics data could involve comparison with multi-species genomic conservation data, cross-referencing with emerging zebrafish functional genomics data, and ultimately functional investigation through transgenic reporter analysis in zebrafish.

While published zebrafish endoderm functional genomics datasets from gastrulation onwards are currently scarce, undoubtedly exploitation of methods discussed above will lead to complementary human and zebrafish functional data for direct comparison. However, analysis of zebrafish-human conservation within CRMs identified in differentiated hESCs can currently easily be performed using multiple databases. One example is ANCORA [127]. ANCORA provides sets of highly conserved non-coding elements (HCNEs) between multiple species pairs and genome assemblies, allowing identification of conserved components of CRMs at a range of size and percentage conservation cut-offs. While the relationship between sequence and functional conservation of CRMs is still emerging, as discussed elsewhere [128, 129], putative CRMs can be prioritized for further study based on human-zebrafish sequence conservation as an indicator of potential conserved function. For example, comparison of H3K27ac ChIP-seq data from Loh et al. [123] with human-zebrafish HCNE data from ANCORA reveals conservations within a range of putative human CRMs including at genes implicated in disease such as *FOXA2*, *BMPRI1A* and *TCF7L2* (Figures 2 and 3). *Foxa2* is implicated in liver and pancreas formation in zebrafish [130], and human genetic variants in CRMs proximal to *FOXA2* have been shown to affect glucose regulation and risk of type 2 diabetes mellitus (T2D) [131, 132]. *TCF7L2* has also been implicated in T2D with a profound effect on disease susceptibility [133]. Analysis of *tcf7l2* mutant zebrafish recapitulates aspects of the human pathology indicating a highly useful model for studying the pleiotropic effects of *TCF7L2* loss-of-function [134]. Other genes associated with human disease with conserved CRMs, such as *BMPRI1A* which is implicated in polyposis syndromes of the gut, are yet to be completely studied in zebrafish in the context of disease

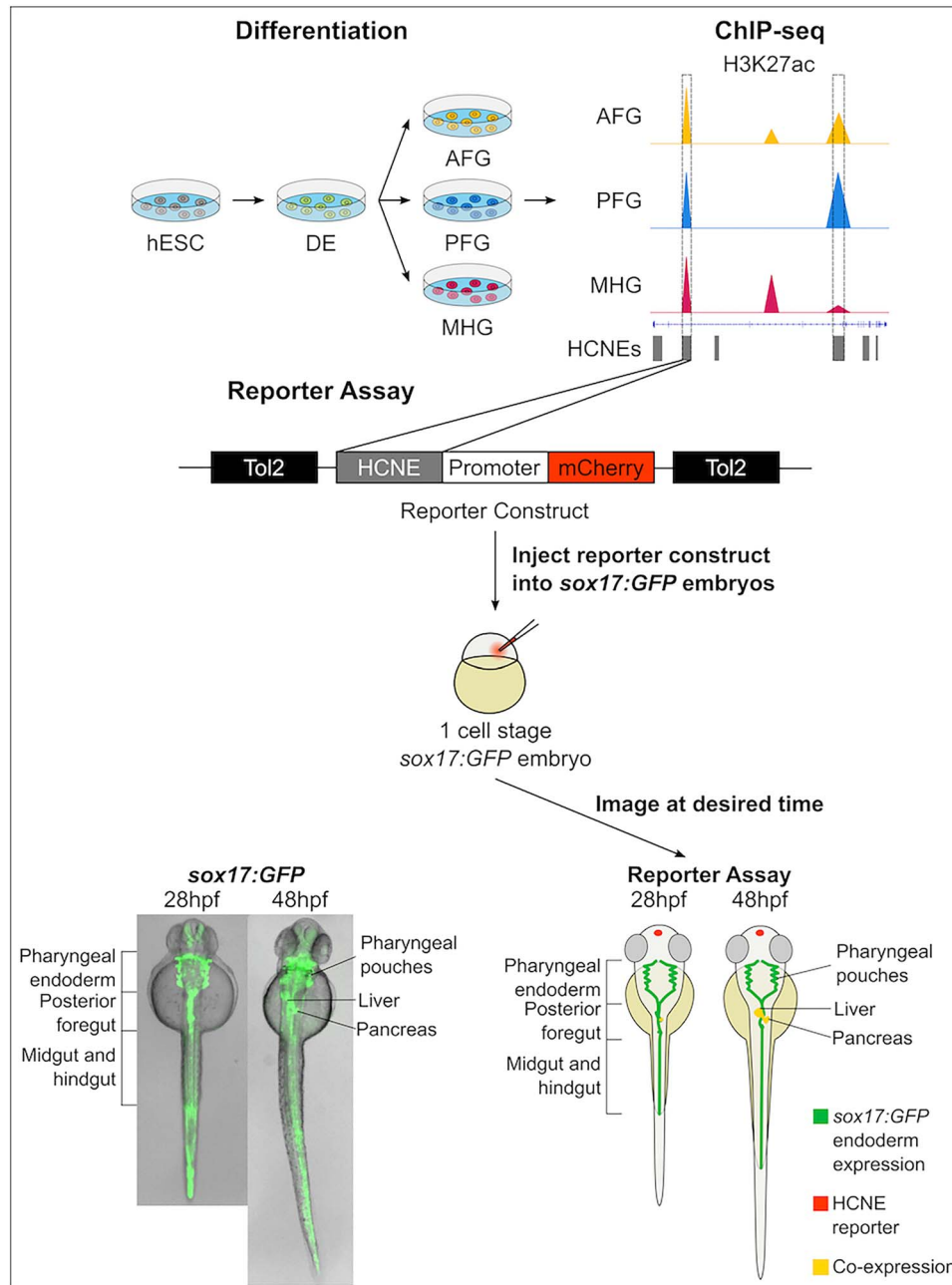


Figure 2. Example procedure for the study of human putative CRMs in zebrafish reporter assays. Genomic regions of interest can be identified by functional genomics approaches such as ChIP-seq in human samples either from directed differentiation of embryonic stem cells or primary tissue. Sub-selection can be performed based on attributes such as presence of highly conserved non-coding elements (HCNEs) and cloned into mCherry fluorescent reporter constructs for microinjection into transgenic zebrafish embryos e.g. *Tg(sox17:GFP)* at the 1-cell stage followed by imaging of both mCherry and GFP throughout development to relate novel reporter expression patterns to the established reporter line. Example epifluorescence images of *Tg(sox17:GFP)* larvae at 28 and 48 h.p.f. are shown with anatomical structures labeled equivalently to cartoons.

[135, 136]. Use of functional genomics data from human and/or zebrafish combined with knowledge of sequence conservation and/or genetic data will allow us to study how CRM mutations influence gene expression and phenotype in zebrafish disease models.

In a recent example of the integration of human genetic and functional genomic data and zebrafish reporter assays, Eufrásio *et al.* [137] cross-referenced single nucleotide polymorphisms (SNPs) associated with T2D identified in genome-wide association studies with putative pancreatic enhancers identified

through transcription factor ChIP-seq in human endocrine pancreas samples. The resulting 10 human enhancers were examined in transient reporter assays in zebrafish, confirming both endocrine pancreatic enhancer activity for six, and revealing the functional consequences of numerous identified SNPs [137]. Notably, refined methods for analysis of wild type and mutant CRMs are being developed such as Q-STARZ, which uses a dual-CRM dual-reporter system combined with targeted integration to allow comparative analysis of CRMs without technical bias [138]. The continued development of zebrafish genetic tools to

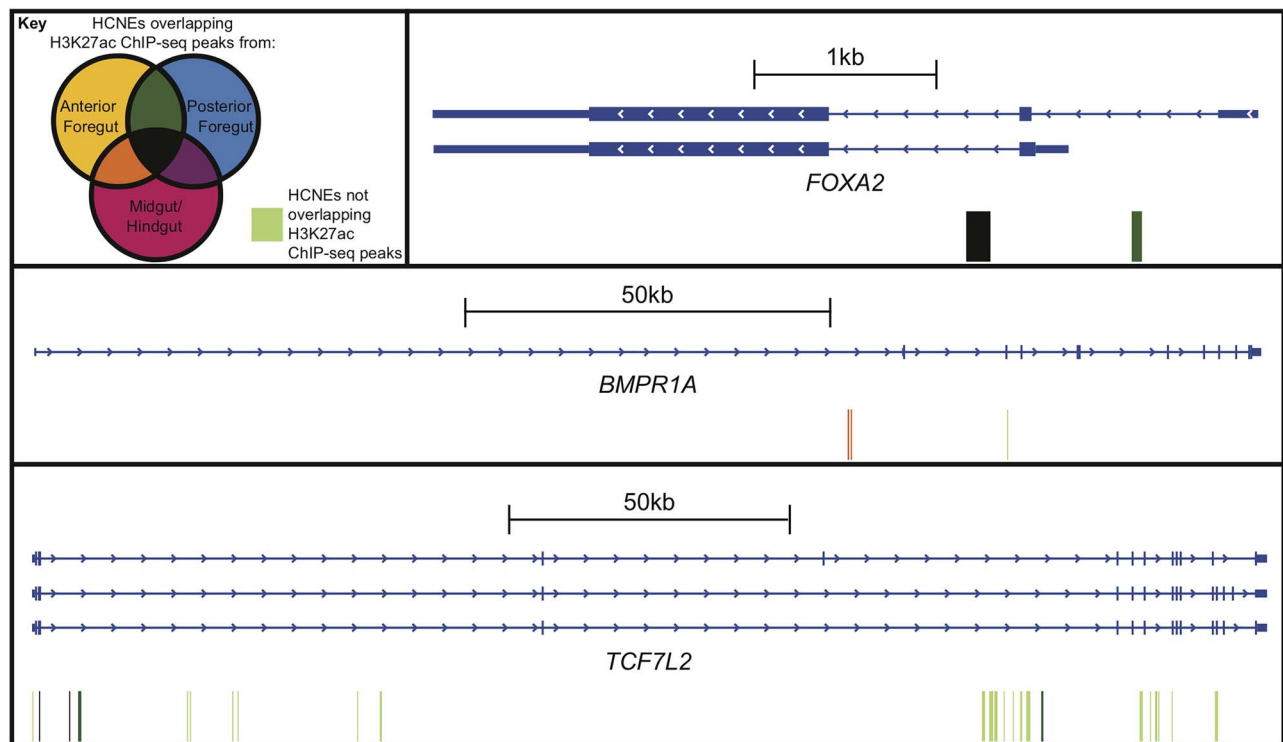


Figure 3. Example HCNEs within H3K27ac ChIP-seq peaks from hESC differentiated towards distinct endoderm cell populations. Depicted are human *FOXA2*, *BMPR1A* and *TCF7L2* loci with highly conserved non-coding elements (HCNEs) color-coded based on overlap with H3K27ac histone ChIP-seq data for anterior foregut, posterior foregut and midgut/hindgut endoderm populations from Loh et al. [123]. The Venn diagram indicates color-coding of HCNEs overlapping H3K27ac ChIP-seq peaks. HCNEs were downloaded from ANCORA (<http://ancora.genereg.net>) comparing human (hg19) to zebrafish (danRer10) genome builds using a window size of 30 bp and 70–100% identity threshold.

study gene and CRM function will therefore greatly facilitate disease modeling and the progression from functional genomics analysis to functional understanding of individual CRMs, and the consequences of variation and mutation to phenotype.

A further notable example of endoderm disease-causing long-range enhancer mutations is pancreatic agenesis, a congenital disease where the pancreas fails to develop during embryogenesis [139]. Linkage and whole-genome sequencing of individuals with isolated pancreas agenesis and no causal protein-coding mutations uncovered six different mutations in a previously uncharacterized putative enhancer region location ~25 kb downstream of *PTF1A* (pancreas-specific transcription factor) in 10 families with pancreatic agenesis [140]. A recent unpublished study suggests that this human putative pancreatic enhancer is able to drive expression in acinar, duct and pancreas progenitor cells in zebrafish reporter assays [141]. Strikingly, the same study identified a novel zebrafish *ptf1a* enhancer that was capable of driving the equivalent expression pattern in reporter assays. CRISPR deletion of this enhancer also led to pancreatic agenesis in zebrafish suggesting a conserved requirement. Importantly, though this zebrafish enhancer was not highly conserved in human, the functionally equivalent human and zebrafish enhancers appeared to contain similar transcription factor binding sites such as for key pancreatic regulators *Foxa2* and *Pdx1* [141]. This strongly suggests that integrated analysis of human and zebrafish functional genomics data combined with interrogation of disease associated CRMs can reveal and confirm the causal link between enhancer mutations and developmental disorders, independent of a high degree of sequence conservation.

Conclusions

Though there are notable differences, current evidence indicates that gene regulatory control of early endoderm formation is largely conserved between zebrafish and other vertebrates. The logical exploitation of transgenic lines combined with emerging technologies will allow more detailed analysis of gene regulatory control of endoderm formation throughout ontogenesis, providing a blueprint for understanding normal development and disease. Integration of functional genomic and sequence data from human and zebrafish studies will allow the further development of disease modeling strategies in zebrafish to understand how mutations and sequence variants in non-coding regions of the genome influence developmental and metabolic defects.

Key Points

- The core gene regulatory program controlling endoderm specification is largely conserved between teleosts and other vertebrates
- An expanding toolkit of transgenic lines and methods will allow more detailed analyses on minor endoderm cell populations
- Zebrafish is emerging as an effective organism to study the relationship between non-coding variants/-mutations and human disease

Funding

Daniela Figiel and Randa Elsayed are funded by the Medical Research Council Doctoral Training Partnership in

Interdisciplinary Biomedical Research (MR/N014294/1). Andrew Nelson's research on endoderm formation in zebrafish is funded by a Wellcome Seed Award (210177/Z/18/Z).

References

- Driever W, Solnica-Krezel L, Schier AF, et al. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 1996;**123**:37–46.
- Haffter P, Granato M, Brand M, et al. The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 1996;**123**:1–36.
- Horsfield JA. Packaging development: how chromatin controls transcription in zebrafish embryogenesis. *Biochem Soc Trans* 2019;**47**:713–24.
- 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.
- Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* 2009;**25**:221–51.
- Fabian P, Tseng KC, Smeeton J, et al. Lineage analysis reveals an endodermal contribution to the vertebrate pituitary. *Science* 2020;**370**:463–7.
- Seth A, Stemple DL, Barroso I. The emerging use of zebrafish to model metabolic disease. *Dis Model Mech* 2013;**6**:1080–8.
- Schier AF. Nodal morphogens. *Cold Spring Harb Perspect Biol* 2009;**1**:a003459.
- Robertson EJ. Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Semin Cell Dev Biol* 2014;**32**:73–9.
- Sampath K, Robertson EJ. Keeping a lid on Nodal: transcriptional and translational repression of Nodal signalling. *Open Biol* 2016;**6**:150200.
- Hill CS. Spatial and temporal control of NODAL signaling. *Curr Opin Cell Biol* 2018;**51**:50–7.
- Schier AF, Talbot WS. Molecular genetics of axis formation in zebrafish. *Annu Rev Genet* 2005;**39**:561–613.
- Ober EA, Field HA, Stainier DY. From endoderm formation to liver and pancreas development in zebrafish. *Mech Dev* 2003;**120**:5–18.
- Zaucker A, Kumari P, Sampath K. Zebrafish embryogenesis—a framework to study regulatory RNA elements in development and disease. *Dev Biol* 2020;**457**:172–80.
- Carvalho L, Heisenberg CP. The yolk syncytial layer in early zebrafish development. *Trends Cell Biol* 2010;**20**:586–92.
- Kelly C, Chin AJ, Leatherman JL, et al. Maternally controlled (beta)-catenin-mediated signaling is required for organizer formation in the zebrafish. *Development* 2000;**127**:3899–911.
- Dougan ST, Warga RM, Kane DA, et al. The role of the zebrafish Nodal-related genes *squint* and *cyclops* in patterning of mesendoderm. *Development* 2003;**130**:1837–51.
- Bellipanni G, Varga M, Maegawa S, et al. Essential and opposing roles of zebrafish beta-catenins in the formation of dorsal axial structures and neurectoderm. *Development* 2006;**133**:1299–309.
- Hong SK, Jang MK, Brown JL, et al. Embryonic mesoderm and endoderm induction requires the actions of non-embryonic Nodal-related ligands and Mxtx2. *Development* 2011;**138**:787–95.
- Xu C, Fan ZP, Muller P, et al. Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway. *Dev Cell* 2012;**22**:625–38.
- Bruce AE, Howley C, Dixon Fox M, et al. T-box gene *eomesodermin* and the homeobox-containing *Mix/Bix* gene *mtx2* regulate epiboly movements in the zebrafish. *Dev Dyn* 2005;**233**:105–14.
- Xu P, Zhu G, Wang Y, et al. Maternal *Eomesodermin* regulates zygotic *Nodal* gene expression for mesendoderm induction in zebrafish embryos. *J Mol Cell Biol* 2014;**6**:272–85.
- Onichtchouk D, Driever W. Zygotic genome activators, developmental timing, and pluripotency. *Curr Top Dev Biol* 2016;**116**:273–97.
- Teo AK, Arnold SJ, Trotter MW, et al. Pluripotency factors regulate definitive endoderm specification through *eomesodermin*. *Genes Dev* 2011;**25**:238–50.
- Frankenberg S, Gerbe F, Bessonard S, et al. Primitive endoderm differentiates via a three-step mechanism involving *Nanog* and RTK signaling. *Dev Cell* 2011;**21**:1005–13.
- Messerschmidt DM, Kemler R. *Nanog* is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev Biol* 2010;**344**:129–37.
- Arnold SJ, Hofmann UK, Bikoff EK, et al. Pivotal roles for *eomesodermin* during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* 2008;**135**:501–11.
- Dubrulle J, Jordan BM, Akhmetova L, et al. Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. *Elife* 2015;**4**:e0504.
- van Boxtel AL, Chesebro JE, Heliot C, et al. A temporal window for signal activation dictates the dimensions of a Nodal signaling domain. *Dev Cell* 2015;**35**:175–85.
- Hozumi S, Aoki S, Kikuchi Y. Nuclear movement regulated by non-Smad Nodal signaling via JNK is associated with Smad signaling during zebrafish endoderm specification. *Development* 2017;**144**:4015–25.
- Liu Z, Lin X, Cai Z, et al. Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. *J Biol Chem* 2011;**286**:28520–32.
- Nelson AC, Cutty SJ, Niini M, et al. Global identification of Smad2 and *Eomesodermin* targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for *Eomesodermin* in repression of ectodermal gene expression. *BMC Biol* 2014;**12**:81.
- Lee KL, Lim SK, Orlov YL, et al. Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet* 2011;**7**:e1002130.
- Kim SW, Yoon SJ, Chuong E, et al. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* 2011;**357**:492–504.
- Yoon SJ, Wills AE, Chuong E, et al. HEB and E2A function as SMAD/FOXH1 cofactors. *Genes Dev* 2011;**25**:1654–61.
- Bruce AE, Howley C, Zhou Y, et al. The maternally expressed zebrafish T-box gene *eomesodermin* regulates organizer formation. *Development* 2003;**130**:5503–17.
- Pogoda HM, Solnica-Krezel L, Driever W, et al. The zebrafish forkhead transcription factor *FoxH1/Fast1* is a modulator of Nodal signaling required for organizer formation. *Curr Biol* 2000;**10**:1041–9.
- Germain S, Howell M, Esslemont GM, et al. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. *Genes Dev* 2000;**14**:435–51.
- Bjornson CR, Griffin KJ, Farr GH, 3rd, et al. *Eomesodermin* is a localized maternal determinant required for endoderm induction in zebrafish. *Dev Cell* 2005;**9**:523–33.

40. Tseng WF, Jang TH, Huang CB, et al. An evolutionarily conserved kernel of *gata5*, *gata6*, *otx2* and *prdm1a* operates in the formation of endoderm in zebrafish. *Dev Biol* 2011;**357**:541–57.
41. Kikuchi Y, Trinh LA, Reiter JF, et al. The zebrafish *bonnie* and *clyde* gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors. *Genes Dev* 2000;**14**:1279–89.
42. Reiter JF, Kikuchi Y, Stainier DY. Multiple roles for *Gata5* in zebrafish endoderm formation. *Development* 2001;**128**:125–35.
43. Reiter JF, Alexander J, Rodaway A, et al. *Gata5* is required for the development of the heart and endoderm in zebrafish. *Genes Dev* 1999;**13**:2983–95.
44. Hong SK, Levin CS, Brown JL, et al. Pre-gastrula expression of zebrafish extraembryonic genes. *BMC Dev Biol* 2010;**10**:42.
45. Pereira LA, Wong MS, Mei Lim S, et al. The Mix family of homeobox genes—key regulators of mesendoderm formation during vertebrate development. *Dev Biol* 2012;**367**:163–77.
46. Poulain M, Lepage T. Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish. *Development* 2002;**129**:4901–14.
47. Chen G, Tan R, Tao Q. *Sebox* regulates mesoderm formation in early amphibian embryos. *Dev Dyn* 2015;**244**:1415–26.
48. Du S, Draper BW, Mione M, et al. Differential regulation of epiboly initiation and progression by zebrafish *Eomesodermin A*. *Dev Biol* 2012;**362**:11–23.
49. Slagle CE, Aoki T, Burdine RD. Nodal-dependent mesendoderm specification requires the combinatorial activities of *FoxH1* and *Eomesodermin*. *PLoS Genet* 2011;**7**:e1002072.
50. Papaioannou VE. The T-box gene family: emerging roles in development, stem cells and cancer. *Development* 2014;**141**:3819–33.
51. Ahn D, You KH, Kim CH. Evolution of the *tbx6/16* subfamily genes in vertebrates: insights from zebrafish. *Mol Biol Evol* 2012;**29**:3959–83.
52. Nelson AC, Cutty SJ, Gasiunas SN, et al. In vivo regulation of the zebrafish endoderm progenitor niche by T-box transcription factors. *Cell Rep* 2017;**19**:2782–95.
53. Talbot CD, Walsh MD, Cutty SJ, et al. *Eomesodermin* is functionally conserved between zebrafish and mouse in spite of different mutant phenotypic severities, and controls left/right organiser formation via interlocking feedforward loops. *bioRxiv* 2020. doi:10.1101/2020.10.02.324244.
54. Russ AP, Wattler S, Colledge WH, et al. *Eomesodermin* is required for mouse trophoblast development and mesoderm formation. *Nature* 2000;**404**:95–9.
55. Paraiso KD, Blitz IL, Coley M, et al. Endodermal maternal transcription factors establish super-enhancers during zygotic genome activation. *Cell Rep* 2019;**27**:2962–2977 e2965.
56. Suda Y, Nakabayashi J, Matsuo I, et al. Functional equivalency between *Otx2* and *Otx1* in development of the rostral head. *Development* 1999;**126**:743–57.
57. Ravasi T, Suzuki H, Cannistraci CV, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* 2010;**140**:744–52.
58. Mudumana SP, Hentschel D, Liu Y, et al. Odd skipped related1 reveals a novel role for endoderm in regulating kidney versus vascular cell fate. *Development* 2008;**135**:3355–67.
59. Voltaire E, Brunet F, Naville M, et al. Expansion by whole genome duplication and evolution of the *sox* gene family in teleost fish. *PLoS One* 2017;**12**:e0180936.
60. Niakan KK, Ji H, Maeher R, et al. *Sox17* promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* 2010;**24**:312–26.
61. Seguin CA, Draper JS, Nagy A, et al. Establishment of endoderm progenitors by *SOX* transcription factor expression in human embryonic stem cells. *Cell Stem Cell* 2008;**3**:182–95.
62. Alexander J, Stainier DY. A molecular pathway leading to endoderm formation in zebrafish. *Curr Biol* 1999;**9**:1147–57.
63. Dickmeis T, Mourrain P, Saint-Etienne L, et al. A crucial component of the endoderm formation pathway, *CASANOVA*, is encoded by a novel *sox*-related gene. *Genes Dev* 2001;**15**:1487–92.
64. Kikuchi Y, Agathon A, Alexander J, et al. *Casanova* encodes a novel *Sox*-related protein necessary and sufficient for early endoderm formation in zebrafish. *Genes Dev* 2001;**15**:1493–505.
65. Brown S, Teo A, Pauklin S, et al. *Activin/Nodal* signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* 2011;**29**:1176–85.
66. Kanai-Azuma M, Kanai Y, Gad JM, et al. Depletion of definitive gut endoderm in *Sox17*-null mutant mice. *Development* 2002;**129**:2367–79.
67. Aamar E, Dawid IB. *Sox17* and *chordin* are required for formation of Kupffer's vesicle and left-right asymmetry determination in zebrafish. *Dev Dyn* 2010;**239**:2980–8.
68. Sinner D, Rankin S, Lee M, et al. *Sox17* and *beta-catenin* cooperate to regulate the transcription of endodermal genes. *Development* 2004;**131**:3069–80.
69. Mukherjee S, Chaturvedi P, Rankin SA, et al. *Sox17* and *beta-catenin* co-occupy *Wnt*-responsive enhancers to govern the endoderm gene regulatory network. *Elife* 2020;**9**:e58029.
70. Zhao J, Lambert G, Meijer AH, et al. The transcription factor *Vox* represses endoderm development by interacting with *Casanova* and *Pou2*. *Development* 2013;**140**:1090–9.
71. Frankenberg SR, Frank D, Harland R, et al. The POU-er of gene nomenclature. *Development* 2014;**141**:2921–3.
72. Perez-Camps M, Tian J, Chng SC, et al. Quantitative imaging reveals real-time *Pou5f3-Nanog* complexes driving dorsoventral mesendoderm patterning in zebrafish. *Elife* 2016;**5**:e11475.
73. Aksoy I, Jauch R, Chen J, et al. *Oct4* switches partnering from *Sox2* to *Sox17* to reinterpret the enhancer code and specify endoderm. *EMBO J* 2013;**32**:938–53.
74. Lunde K, Belting HG, Driever W. Zebrafish *pou5f1/pou2*, homolog of mammalian *Oct4*, functions in the endoderm specification cascade. *Curr Biol* 2004;**14**:48–55.
75. Reim G, Mizoguchi T, Stainier DY, et al. The POU domain protein *spg (pou2/Oct4)* is essential for endoderm formation in cooperation with the HMG domain protein *casanova*. *Dev Cell* 2004;**6**:91–101.
76. Onichtchouk D, Geier F, Polok B, et al. Zebrafish *Pou5f1*-dependent transcriptional networks in temporal control of early development. *Mol Syst Biol* 2010;**6**:354.
77. Morrison GM, Brickman JM. Conserved roles for *Oct4* homologues in maintaining multipotency during early vertebrate development. *Development* 2006;**133**:2011–22.
78. Leichsenring M, Maes J, Mossner R, et al. *Pou5f1* transcription factor controls zygotic gene activation in vertebrates. *Science* 2013;**341**:1005–9.

79. Khan A, Nakamoto A, Tai M, et al. Mesendoderm specification depends on the function of Pou2, the class V POU-type transcription factor, during zebrafish embryogenesis. *Dev Growth Differ* 2012;**54**:686–701.
80. Veil M, Schaechtle MA, Gao M, et al. Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. *Development* 2018;**145**:dev155366.
81. He M, Zhang R, Jiao S, et al. Nanog safeguards early embryogenesis against global activation of maternal beta-catenin activity by interfering with TCF factors. *PLoS Biol* 2020;**18**:e3000561.
82. Wagner DE, Weinreb C, Collins ZM, et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 2018;**360**:981–7.
83. Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;**10**:1213–8.
84. Bogdanovic O, Fernandez-Minan A, Tena JJ, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* 2012;**22**:2043–53.
85. Mizoguchi T, Verkade H, Heath JK, et al. Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* 2008;**135**:2521–9.
86. Sakaguchi T, Kikuchi Y, Kuroiwa A, et al. The yolk syncytial layer regulates myocardial migration by influencing extracellular matrix assembly in zebrafish. *Development* 2006;**133**:4063–72.
87. Stuckemann T, Wegleiter T, Stefan E, et al. Zebrafish Cxcr4a determines the proliferative response to Hedgehog signalling. *Development* 2012;**139**:2711–20.
88. Woo S, Housley MP, Weiner OD, et al. Nodal signaling regulates endodermal cell motility and actin dynamics via Rac1 and Prex1. *J Cell Biol* 2012;**198**:941–52.
89. Nissim S, Sherwood RI, Wucherpfennig J, et al. Prostaglandin E2 regulates liver versus pancreas cell-fate decisions and endodermal outgrowth. *Dev Cell* 2014;**28**:423–37.
90. Fukui H, Terai K, Nakajima H, et al. S1P-Yap1 signaling regulates endoderm formation required for cardiac precursor cell migration in zebrafish. *Dev Cell* 2014;**31**:128–36.
91. Buenrostro JD, Wu B, Litzenburger UM, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;**523**:486–90.
92. Cusanovich DA, Daza R, Adey A, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 2015;**348**:910–4.
93. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015;**161**:1202–14.
94. Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009;**6**:377–82.
95. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 2015;**161**:1187–201.
96. Farrell JA, Wang Y, Riesenfeld SJ, et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 2018;**360**:eaar3131.
97. Satija R, Farrell JA, Gennert D, et al. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;**33**:495–502.
98. Chung MI, Ma AC, Fung TK, et al. Characterization of Sry-related HMG box group F genes in zebrafish hematopoiesis. *Exp Hematol* 2011;**39**:986–998 e985.
99. Ruprecht V, Wieser S, Callan-Jones A, et al. Cortical contractility triggers a stochastic switch to fast amoeboid cell motility. *Cell* 2015;**160**:673–85.
100. Shah G, Thierbach K, Schmid B, et al. Multi-scale imaging and analysis identify pan-embryo cell dynamics of germlayer formation in zebrafish. *Nat Commun* 2019;**10**:5753.
101. Lavergne A, Tarifeno-Saldivia E, Pirson J, et al. Pancreatic and intestinal endocrine cells in zebrafish share common transcriptomic signatures and regulatory programmes. *BMC Biol* 2020;**18**:109.
102. Delporte FM, Pasque V, Devos N, et al. Expression of zebrafish pax6b in pancreas is regulated by two enhancers containing highly conserved cis-elements bound by PDX1, PBX and PREP factors. *BMC Dev Biol* 2008;**8**:53.
103. Alvers AL, Ryan S, Scherz PJ, et al. Single continuous lumen formation in the zebrafish gut is mediated by smoothed-dependent tissue remodeling. *Development* 2014;**141**:1110–9.
104. Dong PD, Provost E, Leach SD, et al. Graded levels of Ptf1a differentially regulate endocrine and exocrine fates in the developing pancreas. *Genes Dev* 2008;**22**:1445–50.
105. Her GM, Chiang CC, Chen WY, et al. In vivo studies of liver-type fatty acid binding protein (L-FABP) gene expression in liver of transgenic zebrafish (*Danio rerio*). *FEBS Lett* 2003;**538**:125–33.
106. Stuckenzholz C, Lu L, Thakur P, et al. FACS-assisted microarray profiling implicates novel genes and pathways in zebrafish gastrointestinal tract development. *Gastroenterology* 2009;**137**:1321–32.
107. Field HA, Ober EA, Roeser T, et al. Formation of the digestive system in zebrafish. I. Liver morphogenesis. *Dev Biol* 2003;**253**:279–90.
108. Zhu B, Hsieh YP, Murphy TW, et al. MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nat Protoc* 2019;**14**:3366–94.
109. Brind'Amour J, Liu S, Hudson M, et al. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun* 2015;**6**:6033.
110. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* 2018;**13**:1006–19.
111. Kaya-Okur HS, Janssens DH, Henikoff JG, et al. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* 2020;**15**:3264–83.
112. Trinh LA, Chong-Morrison V, Sauka-Spengler T. Biotagging, an in vivo biotinylation approach for cell-type specific subcellular profiling in zebrafish. *Methods* 2018;**150**:24–31.
113. Trinh LA, Chong-Morrison V, Gavriouchkina D, et al. Biotagging of specific cell populations in zebrafish reveals gene regulatory logic encoded in the nuclear transcriptome. *Cell Rep* 2017;**19**:425–40.
114. Patton EE, Tobin DM. Spotlight on zebrafish: the next wave of translational research. *Dis Model Mech* 2019;**12**:dmm039370.
115. Adamson KI, Sheridan E, Grierson AJ. Use of zebrafish models to investigate rare human disease. *J Med Genet* 2018;**55**:641–9.
116. Karnuta JM, Scacheri PC. Enhancers: bridging the gap between gene control and human disease. *Hum Mol Genet* 2018;**27**:R219–27.

117. Hadzhiev Y, Miguel-Escalada I, Balciunas D, et al. Testing of cis-regulatory elements by targeted transgene integration in zebrafish using PhiC31 integrase. *Methods Mol Biol* 2016;**1451**:81–91.
118. Roberts JA, Miguel-Escalada I, Slovik KJ, et al. Targeted transgene integration overcomes variability of position effects in zebrafish. *Development* 2014;**141**:715–24.
119. Villefranc JA, Amigo J, Lawson ND. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn* 2007;**236**:3077–87.
120. Bessa J, Tena JJ, de la Calle-Mustienes E, et al. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn* 2009;**238**:2409–17.
121. Kawakami K. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol* 2007;**8**(Suppl 1):S7.
122. Lee K, Cho H, Rickert RW, et al. FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep* 2019;**28**:382–393.e387.
123. Loh KM, Ang LT, Zhang J, et al. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* 2014;**14**:237–52.
124. Wang A, Yue F, Li Y, et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* 2015;**16**:386–99.
125. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007;**39**:311–8.
126. Creighton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 2010;**107**:21931–6.
127. Engström PG, Fredman D, Lenhard B. Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. *Genome Biology* 2008;**9**:R34.
128. Polychronopoulos D, King JWD, Nash AJ, et al. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res* 2017;**45**:12611–24.
129. Nelson AC, Wardle FC. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* 2013;**140**:1385–95.
130. Shin CH, Chung WS, Hong SK, et al. Multiple roles for Med12 in vertebrate endoderm development. *Dev Biol* 2008;**317**:467–79.
131. Gaulton KJ, Ferreira T, Lee Y, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nature Genetics* 2015;**47**:1415–25.
132. Tabassum R, Chavali S, Dwivedi OP, et al. Genetic variants of FOXA2: risk of type 2 diabetes and effect on metabolic traits in north Indians. *Journal of Human Genetics* 2008;**53**:957–65.
133. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006;**38**:320–3.
134. Facchinello N, Tarifeno-Saldivia E, Grisan E, et al. Tcf7l2 plays pleiotropic roles in the control of glucose homeostasis, pancreas morphology, vascularization and regeneration. *Sci Rep* 2017;**7**:9605.
135. Howe JR, Bair JL, Sayed MG, et al. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet* 2001;**28**:184–7.
136. Brosens LA, Langeveld D, van Hattem WA, et al. Juvenile polyposis syndrome. *World J Gastroenterol* 2011;**17**:4839–44.
137. Eufrásio A, Perrod C, Ferreira FJ, et al. In vivo reporter assays uncover changes in enhancer activity caused by type 2 diabetes associated SNPs. *Diabetes* 2020;**69**:2794–805.
138. Bhatia S, Kleinjan DJ, Uttley K, et al. Q-STARZ: quantitative spatial and temporal assessment of regulatory element activity in zebrafish. *bioRxiv* 2020. doi:[10.1101/2020.09.13.290460](https://doi.org/10.1101/2020.09.13.290460).
139. Miguel-Escalada I, Pasquali L, Ferrer J. Transcriptional enhancers: functional insights and role in human disease. *Curr Opin Genet Dev* 2015;**33**:71–6.
140. Weedon MN, Cebola I, Patch AM, et al. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* 2014;**46**:61–4.
141. Bordeira-Carriço R, Teixeira J, Duque M, et al. Cis-regulatory similarities in the zebrafish and human pancreas uncover potential disease-related enhancers. *bioRxiv* 2020. doi:[10.1101/2020.04.27.064220](https://doi.org/10.1101/2020.04.27.064220).
142. White RJ, Collins JE, Sealy IM, et al. A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife* 2017;**6**:e30860.