Mesendoderm Patterning by Vg1/Nodal Heterodimers

Mesendoderm patterning by Vg1/Nodal heterodimers

A dissertation presented

by

Tessa Montague

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of Biochemistry

Harvard University

Cambridge, Massachusetts

March 2018

Dissertation Advisor: Professor Alexander F. Schier Tessa Montague

Mesendoderm patterning by Vg1/Nodal heterodimers

Abstract

One of the key processes during development is the induction of the mesoderm and endoderm (mesendoderm), which generates the precursors of the heart, liver, gut, pancreas, kidney and other internal organs. Nodal, a member of the TGF-β family of secreted ligands, is considered the key inducer of mesendoderm in vertebrate embryos and embryonic stem cells. Another TGF-β family member, Vg1, has been implicated in mesoderm induction but its role has been unclear and controversial. My graduate work addressed the role of Vg1 during embryogenesis.

To interrogate Vg1's developmental role, we investigated strategies for interfering with mRNA function. First, we developed an antisense oligonucleotide (ASO) approach in zebrafish that allows RNase H-mediated degradation of target RNAs. Second, after the emergence of CRISPR/Cas9 genome editing, we developed a web tool for the selection and design of CRISPR/Cas9 targets called CHOPCHOP. Subsequent developments in the CRISPR/Cas9 technology, combined with large-scale studies of gRNA efficiency, illuminated additional rules for gRNA design that we incorporated into a second version of the web tool.

Finally, using CRISPR/Cas9, we knocked out *vg1* in zebrafish and uncovered a critical role in mesendoderm formation. *vg1* mutants fail to form mesendoderm, closely resembling *Nodal* loss-of-function mutants. We found that Nodal is processed and secreted without Vg1, but it requires

Vg1 for its endogenous activity. Conversely, Vg1 is unprocessed and resides in the endoplasmic reticulum without Nodal, and is only secreted, processed and active in the presence of Nodal. Co-expression of Nodal and Vg1 results in heterodimer formation and mesendoderm induction. Thus, we present a new model of mesendoderm induction that relies on the combination of two TGF-β-related signals: maternal and ubiquitous Vg1, and zygotic and localized Nodal. Modeling reveals that the pool of maternal Vg1 enables rapid signaling at low concentrations of zygotic Nodal.

Table of Contents

Acknowledgements

When I was 17 I had the incredible fortune to spend a summer working with Malcolm Logan at the National Institute for Medical Research (NIMR) in London, and that experience has had a lasting impact on my life. At that time, my life plan was to study Natural Sciences at Cambridge University then use my black belt and chemistry degree to become a CIA agent. First, to get into Cambridge, I needed some kind of academic experience, so I applied for a summer research position at NIMR. When I was asked in my interview which model organism I wanted to work with that summer, I stated “something cute and fluffy”. The NIMR panel informed me that doing experiments on cute, fluffy animals would likely mean killing them, so I changed my mind and asked for something ugly. As a result(?) I was placed in Malcolm Logan’s lab to study chick limb development. Malcolm was an amazing advisor. Even though I was a naïve high school student, he taught me every experimental technique himself, he gave me huge freedom to pursue the questions I wanted, and he spent hours filling in the holes in my biology education. When we attended lunchtime seminars, Malcolm would scribble notes on a piece of paper to help me understand what on Earth was going on. In exchange for the lessons in molecular and experimental biology, I made him a mixtape. Malcolm taught me more about biology than I learned during most of high school, and he instilled in me a passion for exploring the scientific unknown. I am eternally grateful for the dedication of time and energy he put into me and my scientific education. If it weren’t for that experience, I might be a CIA agent right now.

The second person who has had a huge scientific impact on me is Andrea Pauli. Andi was a postdoc in the Schier lab when I joined, and she was my lab mentor for ~3 years until she moved to her group leader position at the IMP Vienna. Andi is a force of nature. She has boundless

energy, enthusiasm, and troubleshooting ideas[1]. Andi showed me the kind of scientist I wanted to become: someone who is creative, determined, productive, and always holds herself to an incredibly high standard. I will always look up to Andi, figuratively and literally.

The third person who has had a major scientific influence on me is Alex. When I first met Alex, I thought he was a bit intimidating (and pretty cool, as exemplified by his faculty photo back then, set against an exposed brick background, wow). But I quickly learned what an unusually thoughtful and caring PI he is. Alex has created an amazing scientific environment to work in. The lab is supportive, energetic, and full of scientific ideas and diverse expertise. Through Alex's mentorship I have learned more than I could possibly have hoped for going into grad school. Alex taught me how to think deeply about a problem, how to rigorously approach it, how to ask questions, write, speak, and importantly, how to 'kill the bear'[2]. The Vg1 project was the perfect vehicle for my education. We carried out many rounds of the scientific cycle: formulating a hypothesis, testing it, updating the model, then repeat. One of my fondest memories of my PhD was the first time I got convincing evidence that Vg1 and Nodal form heterodimers – I co-injected *vg1-sfGFP* with *nodal,* and instead of Vg1 being localized inside the cell as usual, Vg1 relocated to bright puncta on the cell surface. When I excitedly showed Alex the data, he slowly sat back in his chair and looked out the window with a smile. There's no one I would have rather worked with on this project.

I would like to thank my committee members, Catherine Dulac, Cliff Tabin, Craig Hunter and Doug Melton for their ideas and input over the years. It's been a privilege to get their insight into

---

[1] AKA optimism

[2] https://en.wikipedia.org/wiki/Alexander_F._Schier

my project, and I'm also very grateful for the career advice. Cliff Tabin was Malcolm Logan's postdoc advisor, so his papers were some of the first scientific manuscripts I ever read. A special shout-out goes to Doug for the decade of work on Vg1 that established the foundations and mysteries of this project. Thank you for leaving a few things unsolved.

The Schier lab created a wonderful scientific home for me for the last 6 years. We had some great ski retreats, parties, and covert operations to uncover the undergrads who stole our food and hand soap. My compatriots, past and present, I will (and I do) miss you: Andi Pauli, James Gagnon, Jeff Farrell, Adam Carte, Shristi Pandey, Yiqun Wang, Phil Abitua, Vassilis Bitsikas, Mehdi Goudarzi, Martin Haesemeyer, Michal Rabani, Nate Lord, Will Joo, Bushra Raj, Owen Randlett, Constance Richter, Max Shafer, Summer Thyme, Jenna Sternberg, Dave Dingal, Kathryn Berg, Jess Leslie, Joo Won Choi, Amelia Hidalgo, Florian Merkle, Julien Dubrulle, David Schoppik, Eivind Valen, Peng Huang, Albert Pan, Patrick Mueller, Farhad Imam, Megan Norris, Alyson Ramirez, Laila Akhmetova, Guo-Liang Chew, Alix Lacoste, Katherine Rogers, Jennifer Li, Drew Robson, Lindsey Pieper, and last but certainly not least, Steve Zimmerman.

Thank you, Stella, my first mutant *vg1* fish. We made a good team.

11 Kent Court – Sriram Srikant, Ali Stoner and Katie Rich – it's been a great few years together. Thanks to Ali for being an elephant and a lamp with me, and to Katie for protecting me from the neighbors with the laundry machine.

It means a lot to me that so many of my family members (inc Uncle John, Aunt Becky, Aunt Cynthia, Julian and Emma) supported me at my defense, and have continually supported me in my career and life. Thanks to Julian for demonstrating that you can spend a Schier lab party

photographing the Schier-Mango book collection alone upstairs. Toby, big bro, thanks for the superior puns, typography lessons, fellow love of fluffy animals, and design influence on every aspect of my scientific life. I hope we can make a cephalopod lab website one day.

My parents have supported and encouraged me to do and be whatever I wanted from a very early age, including astronaut, CIA agent and feminist. Their love and support mean the world to me. Thank you, Etienne and Alex, for supporting and caring for me like your own.

Finally, Fred, my best friend. You bring me so much joy.

“Inshallah we will see a shark.”

List of Figures

Appendix

List of Tables

Abbreviations

ASO	Antisense oligonucleotide
Cyc	Cyclops
Dpp	Decapentaplegic
GDF1	Growth differentiation factor 1
Gsc	Goosecoid
Lft1	Lefty1
LPM	Lateral plate mesoderm
MO	Morpholino
MZT	Maternal-to-zygotic transition
Oep	One-eyed pinhead
Spaw	Southpaw
Sqt	Squint
YSL	Yolk syncytial layer

Chapter 1: Introduction

Animals begin life as one cell, the zygote. Housed in the nucleus of the zygote are the instructions to make a complex, multicellular organism. Through the interpretation of those instructions, the embryo achieves an amazing feat: it undergoes rounds of cell division, induces specialized cell types, establishes its axes, and forms and patterns organs.

In my graduate work I focused on one of the earliest and most critical stages in development, the induction of endoderm and mesoderm (mesendoderm). In order to form tissues and organs animals must first generate three cell types, or germ layers: endoderm, which will produce the internal organs, mesoderm, which will form the muscle and bones, and ectoderm, which will generate the skin and nervous system. How the embryo induces the germ layers has engaged scientists for over a century.

History of mesoderm induction

The field of experimental embryology was established at the turn of the 20th century and led to the discovery of some of the fundamental principles of mesoderm induction. One of the earliest experiments, performed by Hans Driesch in 1892, provided the first example of regulative development. In this experiment, Driesch detached and separated the cells of a 2-cell sea urchin embryo and allowed each cell to develop. Rather than producing two incomplete embryos, to his surprise, each cell developed into a complete larva (Gilbert, 2016). This result revealed that blastomeres have a greater potential than their fate. Driesch, a physicist, was so troubled that

his biological findings couldn't be reconciled with his physics framework that he renounced science and became a philosopher. Luckily others didn't follow suit. Instead, a few years later, Hans Spemann cut a blastula-stage salamander embryo in half using a baby hair from his daughter. He noticed that each half was able to become a complete larva if, and only if, each half contained part of the 'gray crescent', a morphologically distinct region of the embryo (De Robertis, 2006; Gilbert, 2016). He noted that something important must lie in the gray crescent of the embryo that confers this ability to form a larva. After a fate map revealed that the gray crescent becomes the dorsal blastopore lip, Hans Spemann and his graduate student Hilde Mangold performed one of the most important experiments in the field of developmental biology. Believing that the dorsal blastopore lip contains the important ingredients for the development of the embryo, they transplanted the dorsal blastopore lip from one salamander embryo to the opposite side of another embryo. The transplant resulted in an embryo with two axes – conjoined salamander twins – and demonstrated the important concept of induction, that cells or tissues can produce a signal that induces the fates of other cells (Spemann and Mangold, 1924) (Figure 1-1). The tissue that causes induction was subsequently called an organizer, and this particular tissue was named the Spemann-Mangold organizer[3].

The next fundamental insight into induction, and specifically mesoderm induction, was provided by the work of Dutch embryologist Pieter Nieuwkoop. Nieuwkoop exploited a feature of amphibian embryos that they can be separated into pieces, or explants, which will continue to grow and develop in saline. To uncover the origin of embryonic mesoderm, he used blastula-stage axolotl embryos (before the organizer is induced) and separated the animal and vegetal

[3] One of the few known female scientists of the early 20th century, Hilde Mangold died tragically at age 26 before her seminal paper was published. Unfortunately, the term Spemann Organizer has become more commonplace than Spemann-Mangold Organizer.

caps in culture. In isolation, the animal cap developed into undifferentiated ectodermal tissue and the vegetal explant developed into undifferentiated endoderm. However, when he grafted the animal cap onto the vegetal explant, the 'recombinate' formed mesoderm, which developed into notochord, somites and blood cells (Nieuwkoop, 1969) (Figure 1-2). Nieuwkoop demonstrated that the mesoderm was entirely formed from the ectodermal animal cap, while the endoderm was formed from the vegetal explant. Thus, the endoderm produces an inductive signal that causes the overlying presumptive ectoderm to become mesoderm (Nieuwkoop and Ubbels, 1972). The dorsal-most cells of the presumptive endoderm, which will induce the Spemann-Mangold Organizer, were subsequently named the Nieuwkoop Center.

Figure 1-1. Spemann-Mangold organizer graft.
The dorsal lip from an albino *Xenopus* embryo (white tissue, above) is grafted to the opposite (ventral) side of a host blastula-stage embryo. Three days later, the embryo develops a second axis due to the inductive signals produced by the transplant. Figure from (De Robertis, 2006).

These embryological experiments indicated that there are molecularly distinct regions in the blastula-stage embryo that allow one population of cells to produce a signal to induce the fate of other cells. Thus, the next question was when is this regionalization established? A number of

groups started moving back in developmental time to identify when the embryo gains such asymmetries. After encouraging experiments by Nieuwkoop (Nieuwkoop, 1977), Robert Gimlich and John Gerhart performed an elegant experiment in which they irradiated *Xenopus* eggs to prevent the formation of axial mesodermal and neural structures. After growing the irradiated

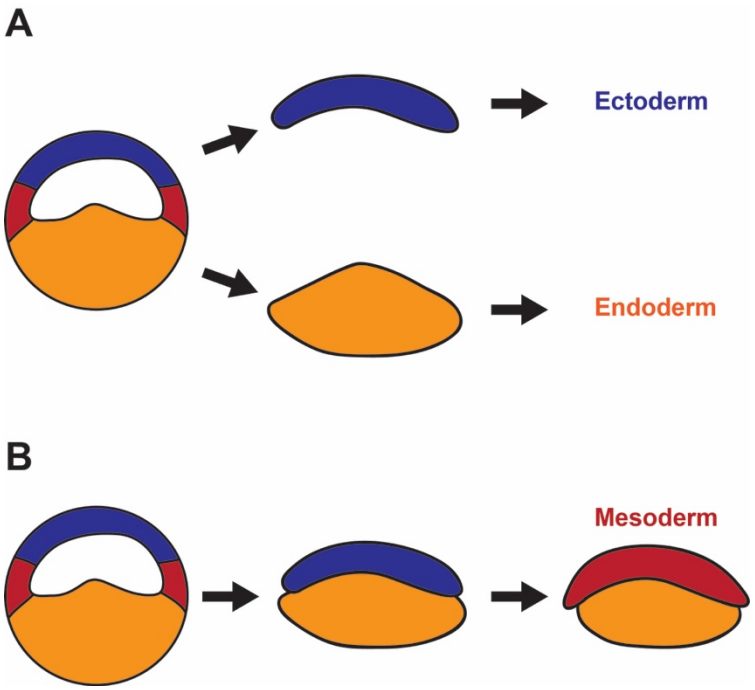


Figure 1-2. The Nieuwkoop explant experiment.
(A) In culture, animal cap explants differentiate into ectodermal tissues, while vegetal explants differentiate into endodermal tissues.
(B) Grafting an animal cap onto a vegetal explant results in mesoderm formation. The vegetal explant produces a mesoderm-inducing signal that causes the ectoderm to differentiate into mesoderm.

host embryos to the 64-cell stage, they transplanted two cells from synchronized, non-irradiated embryos into the vegetal half of the host. When the donor cells were taken from the vegetal and dorsal-most segment of the donor embryo, the host successfully generated axial structures; when donor cells were transplanted from any other vegetal region of the embryo, the host embryos failed to undergo axial development. This experiment revealed that the dorsal-most

vegetal cells are already molecularly distinct during early cleavage stages, and possess mesoderm-inducing properties (Gimlich and Gerhart, 1984).

With the embryonic regions and tissues required to induce mesoderm identified, the next step was to identify the signals that induce the mesoderm. In the late 1970s after the discovery of reverse transcriptases (Baltimore, 1970; Temin and Mizutani, 1970), the first mRNAs were successfully cloned (Efstratiadis et al., 1975; 1976; Rougeon and Mach, 1976; Rougeon et al., 1975) and strategies to develop cDNA libraries were developed (Gubler and Hoffman, 1983; Okayama and Berg, 1982). This enabled groups to subdivide the embryo, extract mRNA, generate a cDNA library, and identify the genes expressed in that region of the embryo. One of the central questions of the mid-1980s was whether embryos inherit localized factors from their mother, which instruct the fate of future cells. This idea was proposed by Edmund B. Wilson in 1925 (Wilson, 1925) after Edwin Conklin observed localized substances in ascidian oocytes (Conklin, 1905). It gathered momentum after actin mRNA was found to be asymmetrically localized in an ascidian embryo (Jeffery et al., 1983) and John Gurdon and colleagues demonstrated that the components required for the expression of actin mRNA were asymmetrically distributed in the uncleaved *Xenopus* embryo (Gurdon et al., 1985). Doug Melton's group took a systematic and unbiased strategy to identify localized mRNAs in the frog egg by separating the vegetal and animal thirds of *Xenopus* eggs, extracting total RNA, generating a radiolabelled cDNA library, and hybridizing it to a whole-embryo cDNA library (Rebagliati et al., 1985). The vast majority of detectable mRNAs were uniformly represented in the embryo, but the authors identified three mRNAs that were enriched in the animal hemisphere and one mRNA that was enriched in the vegetal hemisphere of the egg and early embryo. They named the vegetal mRNA *Vg1.*

The identification of asymmetrically localized mRNAs in the *Xenopus* egg caused great excitement. It provided the tantalizing possibility that the blastomeres of an embryo possess different identities because they asymmetrically inherited instructive mRNAs from the egg. The localized *Xenopus* mRNAs were followed by the discovery of localized mRNAs in other systems, such as *Drosophila bicoid* (Frigerio et al., 1986), in which the localization of the mRNA was shown to be critical for its function (Berleth et al., 1988).

Discovery of Vg1 and Activin

Following the identification of *Vg1* with a cDNA screening approach, Melton and colleagues generated an *in situ* probe to examine its localization. Consistent with their previous findings, *Vg1* mRNA was exquisitely localized to a vegetal crescent in the *Xenopus* egg (Melton, 1987; Weeks and Melton, 1987) (Figure 1-3). Given the subcortical localization of the mRNA, they hypothesized that Vg1 might function to specify the vegetal cells to become endoderm, or it could encode the instructive signal produced by vegetal cells that induces the overlying marginal cells to become mesoderm (Melton, 1987; Nieuwkoop, 1969). By cloning the full-length mRNA and comparing its amino acid sequence to a protein sequence data bank, they found that *Xenopus Vg1* possesses 40% homology with a TGF-β family member *β-A-inhibin* (Weeks and Melton, 1987). Vg1's predicted protein sequence included a signal sequence at the N-terminus (indicating that it enters the endoplasmic reticulum), and like other TGF-β family members, a basic cleavage site. The region of shared homology with β-A-inhibin was predicted to lie downstream of the cleavage site. This analysis suggested that Vg1 is a secreted factor.

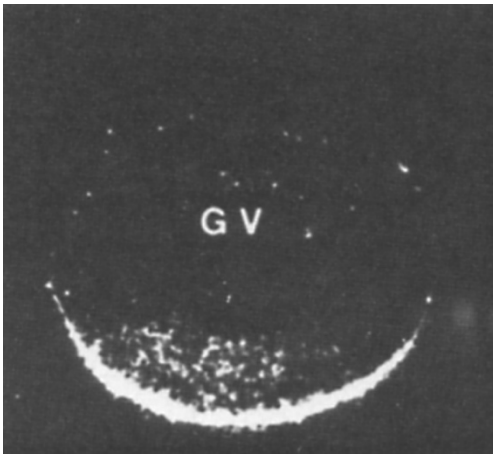


Figure 1-3. *vg1* mRNA is localized in a vegetal crescent in the *Xenopus* oocyte.
From (Weeks and Melton, 1987).

In the same year, Jim Smith identified a bona fide *Xenopus* mesoderm-inducing factor. Using the *Xenopus* XTC cell line he found that animal caps cultured in XTC-conditioned media differentiated into notochord and muscle (Smith, 1987). The factor was named XTC-MIF (mesoderm inducing factor), although the factor's molecular identity, and whether it was found endogenously in the embryo, was unknown.

Two years later, Leslie Dale and Alan Colman provided the first analysis of Vg1 protein. They raised an antibody against Vg1 and detected it in the vegetal half of the *Xenopus* egg and embryo. Vg1 segregated with cytosolic but not membrane fractions indicating that it is likely secreted (Dale et al., 1989). However, this marked the beginning of a series of findings over multiple years that could not be reconciled with Vg1's predicted function as a mesoderm inducer. Although Vg1 possessed sequence homology with TGF-β family members, which are cleaved to produce an active ligand, Dale and colleagues did not detect the predicted cleaved product of Vg1 (Dale et al., 1989). In the same year, David Tannahill and Doug Melton undertook a similar

strategy to investigate the localization, secretion and processing of Vg1 protein. Like Dale and colleagues, they detected only full-length, endogenous Vg1 protein in the egg and embryo, although upon substantial overexpression of *vg1* mRNA they sometimes detected a small amount of protein close to the predicted size of cleaved Vg1. Consistent with the Colman group's results, Vg1 was primarily restricted to the vegetal hemisphere of the embryo, and it localized to a cytosolic fraction, indicating it was not a membrane protein (Tannahill and Melton, 1989). However, inconsistent with its predicted secretory function, expression of Vg1 in oocytes (which readily secrete proteins) did not result in detectable, secreted Vg1 (Tannahill and Melton, 1989). Tannahill and Melton provided some possible explanations. They suggested that perhaps only a relatively small amount of cleaved Vg1 is made, causing it to be undetectable. The secretion of Vg1 might be tightly regulated, causing a failure to detect oocyte-secreted Vg1. They also noted that the conservation of cysteines across Vg1 and other TGF-β family members suggests that Vg1 forms dimers, which it could do with itself or with another TGF-β family member (Tannahill and Melton, 1989).

The following year, the identity of XTC-MIF was reported by Smith and others to be Activin A, another member of the TGF-β family (Smith et al., 1990; van den Eijnden-van Raaij et al., 1990). Focus in the field shifted to Activin, which held the promise that it was the long sought-after mesoderm induction factor. However, Northern blot experiments by Gerald Thomsen and Doug Melton demonstrated that Activin A is not expressed until late gastrula stages – long after mesoderm induction has occurred. They showed that a related protein, Activin B, also possesses potent mesoderm-inducing activity, but the transcripts aren't detected until late blastula stages after the mesoderm induction event has taken place (Thomsen et al., 1990). Subsequent analyses using the more sensitive technique of RT-PCR also failed to detect Activin transcripts in the egg (Dohrmann et al., 1993).

This result coincided with a refocus on Vg1. In 1993, the Melton and Colman labs reported one of the most discouraging results for Vg1 as an endogenous mesoderm inducer, along with a possible solution. Both groups showed that *vg1* injection into animal caps or UV-irradiated embryos failed to generate mesoderm (Dale et al., 1993; Thomsen and Melton, 1993). However, aware that TGF-β family members are active as a cleaved dimer (Massagué, 1987), yet the cleaved product of Vg1 is undetectable, they fused the Vg1 mature domain onto the prodomain of other TGF-β family members, BMP4 and BMP2 (Dale et al., 1993; Thomsen and Melton, 1993). These fusion proteins produced a cleaved mature Vg1 domain that formed secreted dimers, and induced the formation of mesoderm in UV-irradiation, embryonic injection and animal cap assays (Dale et al., 1993; Thomsen and Melton, 1993). Thus, both groups concluded that Vg1 is a potent mesoderm inducer when successfully processed to its mature form, which might only occur under tight control in the embryo (Dale et al., 1993; Thomsen and Melton, 1993). Dale and colleagues made three other interesting observations. First, using reducing and non-reducing western blots they noted that Vg1 dimers couldn't be detected endogenously or under conditions of overexpression, suggesting that the vast majority of Vg1 in the embryo exists in a monomeric form. Second, they demonstrated that the signal peptide at the N-terminus of Vg1 is not cleaved upon entry into the ER, contrary to what happens to other secreted proteins (Heijne, 1990). They interpreted this to mean that Vg1 might be retained in the ER because it still contains the hydrophobic signal or because it is unable to correctly fold. Third, they showed that when BMP-Vg1 is secreted in an oocyte assay it is in a dimeric form, suggesting that Vg1 requires dimer formation to be secreted (Dale et al., 1993).

At the same time that mesoderm-inducing activity was demonstrated for the *Xenopus* Vg1 chimeric proteins, Kathryn Helde and David Grunwald reported *vg1* expression in zebrafish. Like

Xenopus, zebrafish *vg1* was also detected as a maternal transcript, but unlike *Xenopus* it lacked the exquisite localization pattern, and was instead ubiquitous in the early embryo (Helde and Grunwald, 1993). One explanation the authors proposed was that maternal transcripts aren't localized in the early blastula embryo because region-specific gene expression is only initiated after the maternal-to-zygotic transition (MZT) (Helde and Grunwald, 1993).

Discovery of Nodal

The field underwent a shift in 1993. Xunlei Zhou and Michael Kuehn mapped a retroviral insertion mutation in mice that caused embryonic ectoderm to fail to differentiate into mesoderm (Conlon et al., 1991; Iannaccone et al., 1992). The mutation was found to lie in a gene encoding another TGF-β family member that's expressed at the time of mesoderm formation and later in the node, the mouse equivalent of the Spemann-Mangold Organizer. The authors named the gene *Nodal* (Zhou et al., 1993). Nodal quickly became the favored endogenous mesoderm inducer in vertebrates. While Vg1 was expressed in the right place and time to be a mesoderm inducer, it appeared to lack endogenous mesoderm-inducing activity. By contrast, Activin possessed potent mesoderm-inducing activity in ectopic *Xenopus* assays but was not expressed at the right time to be an endogenous mesoderm inducer[4]. Nodal provided a solution to the puzzle: it was expressed at the right time, in the correct place, and genetic evidence proved it was required to form mesoderm (Conlon et al., 1991; 1994; Iannaccone et al., 1992; Zhou et al., 1993). Zebrafish *nodal* mutants demonstrated that Nodal is required for the formation of mesoderm and endoderm *in vivo* (Feldman et al., 1998).

[4] And subsequent mutation of the Activins demonstrated they are not required for mesoderm formation in mice (Matzuk et al., 1995)

Subsequent studies in *Xenopus* (Jones et al., 1995; Joseph and Melton, 1997; Lowe et al., 1996; Lustig et al., 1996; Smith et al., 1995; Takahashi et al., 2000), chick (Levin et al., 1995; Pagán-Westphal and Tabin, 1998), mice (Brennan et al., 2002; Collignon et al., 1996; Conlon et al., 1991; 1994; Kumar et al., 2015; Lowe et al., 1996; Saijoh et al., 2003; Zhou et al., 1993) and zebrafish (Erter et al., 1998; Feldman et al., 1998; Long et al., 2003; Noël et al., 2013; Rebagliati et al., 1998a; 1998b; Sampath et al., 1998; Toyama et al., 1995) demonstrated that *Nodal* orthologs possess mesoderm-inducing activity and direct left-right axis formation in a variety of systems (see 'The Nodal Signaling Pathway').

Vg1 in *Xenopus*, zebrafish and chick

With the discovery that multiple members of the TGF-β family can regulate cell growth and differentiation, focus turned to the receptors that are activated by these ligands. A series of studies in the early 1990s identified a class of serine/threonine transmembrane kinases that are activated by Activin, named the type I (Attisano, 1993) and type II Activin receptors (Attisano et al., 1992; Mathews et al., 1992; Mathews and Vale, 1991). Expression of a truncated, dominant-negative Activin type IIB receptor in *Xenopus* inhibited mesoderm formation, which led the authors to conclude that Activin is required for embryonic mesoderm induction (Hemmati-Brivanlou and Melton, 1992). However, subsequent experiments demonstrated that BMP-Vg1 activity is also inhibited by the truncated Activin receptor, suggesting that type II Activin receptors can mediate signaling by multiple TGF-β family members (Schulte-Merker et al., 1994). Receptor binding assays showed that Activin, but not mature Vg1, directly binds type II Activin

receptors, indicating that Vg1 might activate the type II receptors through an indirect mechanism (Kessler and Melton, 1995).

Further evidence suggested that the mature region of Vg1 is itself a potent mesoderm inducer. Daniel Kessler and Doug Melton noted that secretion of mature Vg1 is inefficient from BMP2- or BMP4-Vg1-injected oocytes, so they generated a new chimeric protein with the prodomain of activin βB and the mature domain of Vg1. Expression of this protein in oocytes resulted in a large amount of secreted mature Vg1. Incubation of animal cap explants with the supernatant of activin βB-Vg1-injected oocytes resulted in mesoderm formation, with a concentration-dependent effect of mature Vg1 on gene expression (Kessler and Melton, 1995)[5]. The authors thus proposed a model for Vg1-mediated mesoderm formation in which cortical rotation of the early *Xenopus* embryo directs the proteolytic activation of Vg1, either by locally activating a protease or by inversely restricting the release of a cleavage inhibitor (Kessler and Melton, 1995).

Important mysteries remained: Why can endogenous Vg1 cleavage not be visualized? Why is there no evidence that endogenous Vg1 is secreted? And why does endogenous Vg1 not induce mesoderm formation?

Studies of Vg1 moved to other animal systems. The Melton lab demonstrated that the *Xenopus* anti-Vg1 antibody cross-reacts with zebrafish Vg1 (denoted here as zVg1). This revealed that zVg1 is expressed ubiquitously in the early zebrafish embryo until the 10-somite stage, and is only detected as a full-length protein, akin to *Xenopus.* Unexpectedly, when they injected zVg1

[5] It should be noted that the Melton Vg1 chimeric proteins used the cleavage site and first 4 amino acids of the mature domain of BMP2/Activin. It is possible that those 4 amino acids conferred ectopic function on the mature domain of Vg1.

into *Xenopus* embryos, zVg1 was processed, secreted and could induce the formation of mesoderm (Dohrmann et al., 1996)[6]. By generating fusion proteins of the Vg1 prodomain and zVg1 mature domain, the authors demonstrated that the zVg1 cleavage site, but not the Vg1 cleavage site, allowed zVg1 to be cleaved in *Xenopus* (Dohrmann et al., 1996). They concluded that *Xenopus* and zebrafish Vg1 are both regulated at the post-translational level, and noted that the trans-acting molecule required for Vg1 processing might be spatially separated from Vg1 in the embryo (Dohrmann et al., 1996).

In 1996 a Vg1 ortholog was cloned in chicken (Seleiro et al., 1996). *cVg1* was found to be expressed at the site of axis (primitive streak) formation prior to gastrulation and mesoderm formation. Anti-Vg1 western blots of chick tissue detected full-length cVg1 but no cleavage product and injection of *cVg1* mRNA into *Xenopus* embryos resulted in no phenotype. However, expression of a chimera of the chick BMP4 prodomain with the cVg1 mature domain induced mesoderm induction and axis formation in *Xenopus* embryos (Seleiro et al., 1996). COS cells transfected with *cVg1* produced cleaved cVg1, and grafts of these cells in pre-streak chick embryos resulted in an ectopic streak (Seleiro et al., 1996). Thus, it was concluded that cVg1 is required for axis formation, and it was suggested that cVg1 might act upstream of Nodal (Seleiro et al., 1996). A subsequent study by Jane Dodd and Claudio Stern's groups revealed similar results. Transfection of COS cells with a *dorsalin-1-cVg1* chimera resulted in secreted, mature cVg1 that could induce mesoderm formation in *Xenopus* animal caps and chick explants (Shah et al., 1997). Grafts of cells expressing *dsl-cVg1* at ectopic locations caused ectopic expression of *cNodal*, suggesting that cVg1 acts upstream of cNodal (Skromne and Stern, 2001).

[6] The authors injected 5 ng of *zVg1* to visualize a cleaved product, which is *a lot* of mRNA. Mesoderm induction occurred at lower concentrations.

Vg1 in left-right asymmetry

Progress on the mesoderm-inducing capabilities of *Xenopus* and zebrafish Vg1 was hampered by the inability to uncover an endogenous function for the gene. Meanwhile, Joseph Yost's group noticed that the *Xenopus* Nieuwkoop center possessed left-right asymmetries in its mesoderm-inducing capabilities. In vertebrates, Nodal activity is restricted to the left side of the node during somite stages, and subsequently extends up the left lateral plate mesoderm (LPM) through auto-induction and repression from Lefty. Nodal induces expression of target genes that cause the left-right asymmetric development of organs, such as the asymmetric jogging and looping of the heart (Figure 1-4) (Hamada et al., 2002). The Yost lab tested the capability of maternal mRNAs to influence left-right asymmetry in the embryo, and reported that injection of *BMP-Vg1* into the right side of the Nieuwkoop center results in randomization of cardiac and visceral left-right asymmetry (Hyatt et al., 1996). Upon injection of *BMP-Vg1* into the right side of the Nieuwkoop center, *Xnr1* (*Nodal)* was now expressed on both the left and right sides of the LPM. They concluded that maternal Vg1 induces the expression of *Xnr1* on the left side, which induces the correct asymmetric positioning and looping of the heart and gut (Hyatt et al., 1996). Consistent with this idea, injection of a truncated Activin receptor into the left side of embryos, which had been shown to block BMP-Vg1 activity (Schulte-Merker et al., 1994), caused randomization of left-right asymmetry (Hyatt and Yost, 1998).

The idea that Vg1 is the endogenous mesoderm inducer lost more traction in 1996 with the discovery of VegT. *VegT* was found to localize to a vegetal crescent in the *Xenopus* oocyte; it encodes a transcription factor whose overexpression resulted in mesoderm formation; and depletion of *VegT* using an antisense oligonucleotide strategy resulted in transformation of the

presumptive mesoderm into ectoderm instead of mesoderm (Zhang and King, 1996; Zhang et al., 1998a).

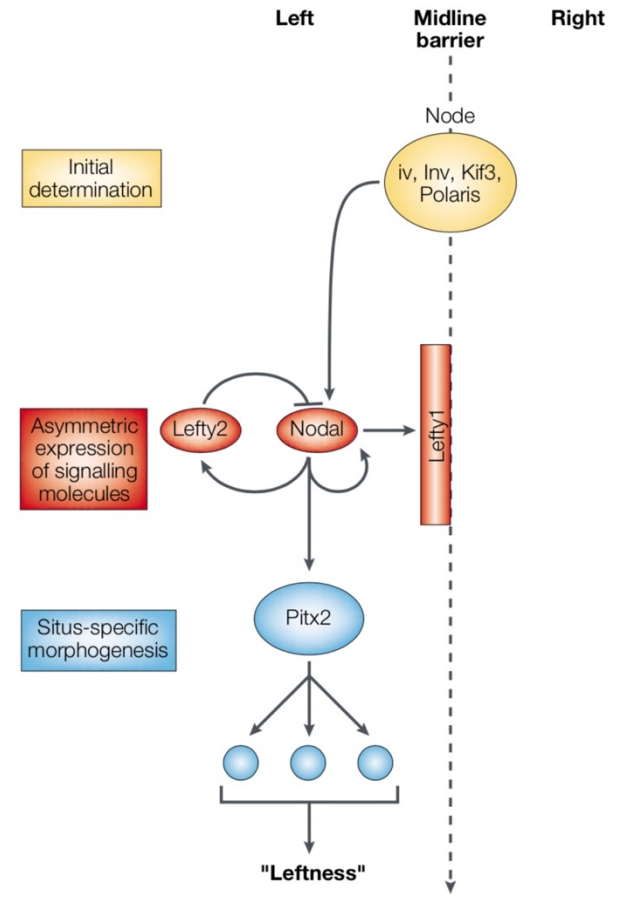


Figure 1-4. Establishment of the left-right axis in mice.
Nodal is expressed at the node and becomes restricted to the left side of the lateral plate mesoderm (LPM). It induces its own expression as well as the expression of the Lefty inhibitors. This allows *Nodal* to extend along the left LPM until it induces expression of the transcription factor Pitx2. Subsequent signaling events cause the left-right asymmetric morphogenesis of organs. From (Hamada et al., 2002).

At this point it was clear that a genetic or depletion strategy was necessary to interrogate the endogenous function of *Vg1*, since the primary evidence for Vg1 mesoderm-induction activity came from overexpression studies. Elaine Joseph and Doug Melton developed a mutant ligand approach in which they designed BMP-*Vg1* mRNAs with mutations that would theoretically

interfere with dimer formation or receptor binding. Three of the mutated ligands successfully failed to induce mesoderm gene expression when injected into embryos, and in the most extreme cases, the resulting embryos failed to generate dorsal mesoderm (Joseph and Melton, 1998). The authors cited this as evidence that endogenous Vg1 is required for mesoderm induction. Meanwhile, other dominant-negative approaches were used to argue that Activin has an endogenous role in mesoderm induction (Wittbrodt and Rosa, 1994). The lack of specificity (Hemmati-Brivanlou and Melton, 1992; Meno et al., 1999; Schulte-Merker et al., 1994) and range of phenotypes (Dyson and Gurdon, 1997; Hawley et al., 1995; Hemmati-Brivanlou and Melton, 1992; Kessler and Melton, 1995; Schulte-Merker et al., 1994; Sun et al., 1999) produced by dominant-negative approaches underscored the need for a genetic strategy.

A few years later Bilge Birsoy, Chris Wylie and Janet Heasman designed antisense oligonucleotides (ASOs) against *Vg1* to cause depletion of the endogenous mRNA. Injection of *Vg1* ASOs caused concentration-dependent effects on mesoderm formation, and although all three germ layers were still present, two mesoderm structures (notochord and somites) were severely affected (Birsoy et al., 2006). The authors noted that depletion of Vg1 causes reduction in Smad2 phosphorylation and an increase in Smad1 phosphorylation, suggesting that Vg1 induces phosphorylation of Smad2 and expression of BMP antagonists (Birsoy et al., 2006).

All studies on Vg1 to date had used the *Vg1* allele identified in the original cDNA screen, which contains a proline 20 amino acids from the N-terminus of the protein (Rebagliati et al., 1985). Birsoy and colleagues identified a second *Vg1* allele that is equally represented in the oocyte and embryo, with a serine substituted for the proline (Birsoy, 2005). The authors claimed that the *Vg1(ser)* allele can rescue the ASO-induced phenotype, Vg1(ser) is processed more readily than Vg1(pro) and, unlike Vg1(pro), it possesses mesoderm-inducing activity (Birsoy et al., 2006).

However, while this paper claimed to resolve the mystery of Vg1's function, the knockdown phenotypes were mild, and it suffered from a number of major experimental flaws[7]. Nonetheless, this was the last major paper published on *Xenopus* Vg1.

GDF1 and GDF3

A Vg1 ortholog was noticeably absent from mammals until Daniel Kessler's group screened a mouse expression library with the Vg1 monoclonal antibody and achieved cross-reactivity with a clone encoding GDF1 (Growth Differentiation Factor 1) (Wall et al., 2000). In the same year, a group published a *Gdf1* mouse mutant, which had defects in left-right patterning (Rankin et al., 2000). Interestingly, although *Gdf1* had been cloned 9 years earlier and was found to possess its greatest homology with *Xenopus Vg1*, the author had concluded that it was unlikely to be the Vg1 mammalian ortholog since the majority of homology lay downstream of the predicted cleavage site (Lee, 1990).

Northern blot and *in situ* analysis of *Gdf1* revealed that it is expressed throughout the embryo proper, and then in the primitive node, ventral neural tube and lateral plate mesoderm (Rankin et al., 2000). The *Gdf1* mutant mouse demonstrated that GDF1 is required for establishing the left-right axis, and in its absence, *Nodal, Lefty* and *Pitx2* are not expressed, suggesting that GDF1 is upstream of *Nodal* expression (Rankin et al., 2000). Meanwhile, Wall and colleagues showed that, although GDF1 is expressed after mesoderm induction, it shares a number of properties

[7] First, the authors claim that Vg1(ser) is processed more readily than Vg1(pro), but the figure only shows a smaller version of the full-length protein (35 kDa) rather than the cleaved product of Vg1 (18 kDa). Ridiculous. Second, they claim that Vg1(ser) has mesoderm-inducing activity whereas Vg1(pro) (in the same experiment) does not. However, they don't show the Vg1(pro) data. Third, the 'mesoderm-inducing activity' is a qPCR result with no error bars or evidence that it was ever repeated.

with *Xenopus* Vg1. For instance, expression of *Gdf1* in *Xenopus* results in no detectable cleavage or embryonic activity, but upon fusing the GDF1 mature domain to the BMP2 prodomain, BMP-GDF1 is successfully cleaved and induces mesoderm in *Xenopus* (Wall et al., 2000). In addition, consistent with the mouse mutant, expression of *Gdf1* on the right but not left side of the embryo resulted in a reversal of left-right asymmetry (Wall et al., 2000). These studies presented a promising ortholog of *Xenopus*, zebrafish and chick *vg1*, although they suggested that mammalian *vg1* was required for left-right asymmetry, and not for mesoderm induction.

The Schier lab provided evidence that Vg1 and GDF1 function through a Nodal-like pathway in the embryo. Using zebrafish with a mutation in the Nodal coreceptor One-eyed pinhead (Oep), Simon Cheng and colleagues showed that BMP-GDF1 and Activin-Vg1 activity are dependent on functional Oep, but independent of the Nodal ligands Cyclops and Squint. They showed that BMP-GDF1 and Activin-Vg1 also activate the same type I and type II Activin receptors as Nodal, their receptor binding ability is dependent on Oep (Cheng et al., 2003), and Activin-Vg1 activity is inhibited by Lefty through its interaction with Oep (Cheng et al., 2004). By contrast, Activin activity was found to be independent of Oep (Cheng et al., 2003) and was not inhibited by Lefty (Cheng et al., 2004). These studies, combined with other work that showed Vg1 and Nodal signal via activation of Smad2 (Graff et al., 1996), suggested that Nodal and Vg1 share signaling components with each other, and not with Activin.

A second TGF-β family member was also predicted to serve as a mammalian ortholog of Vg1 (Jones et al., 1992), but its role in development wasn't interrogated until years later. Mouse *Gdf3* expression was found to be highly localized to the pluripotent cells of the inner cell mass and epiblast, with its expression ceasing prior to gastrulation (Chen et al., 2006). To interrogate the function of *Gdf3* the authors generated knockout mice, the majority of which developed into

fertile, healthy adults. However, a subset didn't make it through development. A proportion of mutant embryos had absent or reduced expression of endoderm and mesoderm markers, and displayed defects in anterior visceral endoderm formation, akin to some of the mutants in the Nodal signaling pathway (Brennan et al., 2001; 2002). The authors thus investigated *Nodal* expression in the mutants, and noted that *Nodal* was absent or upregulated in a proportion of embryos, suggesting some synergy between Nodal and GDF3. *In vitro* studies showed that GDF3 is processed to its mature form when fused to the BMP2 prodomain, and is active in a Nodal-responsive luciferase assay in the presence of Cripto, the mouse Nodal coreceptor (Chen et al., 2006). Likewise, expression of *BMP-Gdf3* in *Xenopus* assays resulted in mesoderm induction, causing the authors to conclude that GDF3 acts in a Nodal signaling pathway in the pre-gastrulation embryo (Chen et al., 2006).

The same year, Olav Andersson and Carlos Ibáñez generated *Gdf1;Nodal* mutants to investigate possible interactions between the GDF1 and Nodal pathways. While *Nodal* mutants died at gastrulation (Lowe et al., 2001) and *Gdf1* mutants had defects in left-right axis formation (Rankin et al., 2000), $Gdf1^{-/-};Nodal^{+/-}$ compound mutants had defects in notochord and prechordal plate development, two structures associated with differentiation of the axial mesoderm (Andersson et al., 2006). The authors examined the expression of *Nodal* over development in the absence of *Gdf1* to see if Nodal and GDF1 act synergistically to control *Nodal* expression. In agreement with previous studies (Rankin et al., 2000), they found that *Nodal* expression was absent in the LPM in the absence of *Gdf1*, indicating that GDF1 is upstream of *Nodal* expression. However, *Nodal* expression was unaffected by the loss of *Gdf1* in other spatial and temporal domains, including the epiblast and node (Andersson et al., 2006).

The partially penetrant phenotype of *Gdf3* mutants suggested that it might be acting in concert with other TGF-β family members to pattern the mesoderm. Andersson and Ibáñez generated *Gdf1;Gdf3* double mutants to investigate cooperation between the ligands. A subset of *Gdf3* single mutants showed defects in anterior visceral endoderm formation and mesoderm formation (Andersson et al., 2007; Chen et al., 2006). The *Gdf1;Gdf3* double mutants showed more severe defects in anterior visceral endoderm formation and mesoderm formation, consistent with the idea of cooperativity between the ligands (Andersson et al., 2007). In addition, *in vitro* expression of *Gdf1 or Gdf3* and *Cripto* in the presence or absence of *Nodal* revealed that Nodal has synergistic effects on GDF1 and 3 activity in a Nodal-responsive luciferase assay, suggesting synergism between the GDF ligands and Nodal (Andersson et al., 2007).

A convincing mechanism for GDF1 and Nodal's synergistic activity was subsequently provided by the work of Chinatsu Tanaka, Hiroshi Hamada and Yukio Saijoh. By rescuing *Nodal* expression in the lateral plate mesoderm (LPM) of *Gdf1* mutants using tissue-specific *Gdf1* expression, they demonstrated that *Gdf1* is required in the node along with *Nodal* to transfer Nodal activity to the LPM. Given the shared signaling components of the GDF1 and Nodal pathways (Cheng et al., 2003; 2004; Wall et al., 2000), the authors used a *Xenopus* assay to measure the effect of GDF1 on a Nodal-responsive luciferase reporter. They found that *Gdf1* injection into *Xenopus* animal caps resulted in the production of full-length and processed GDF1, but yielded no activation of the luciferase reporter. However, co-injection of *Gdf1* and *Nodal* resulted in significantly greater activity than *Nodal* alone. A similar result was achieved for co-expression of *zVg1* and *cyclops* or *squint,* the zebrafish orthologs of *Nodal* (Tanaka et al., 2007). To understand the nature of this enhanced activity, the authors performed co-immunoprecipitation experiments with GDF1 and Nodal and detected an interaction, suggesting that the proteins form a heterodimer. To test the properties of the heterodimers, *Xenopus*

oocytes were co-injected with either *Gdf1* or *Nodal* mRNAs, or a combination of *Gdf1* and *Nodal* mRNAs, and the supernatant was collected 3 days later. *Xenopus* animal caps exposed to the Nodal/GDF1 media had substantial luciferase reporter activity, whereas animal caps exposed to a mixture of Nodal media and GDF1 media failed to activate the reporter. This indicated that the heterodimers likely form intracellularly. Further experiments in *Xenopus* and mice demonstrated that GDF1 increases the range of Nodal, which could allow the transfer of Nodal activity from the node to the LPM. The authors suggested that GDF1 might increase Nodal activity by increasing its affinity for the receptor (Tanaka et al., 2007). GDF1/Nodal heterodimers were subsequently purified from HEK cells, and used to induce embryonic stem cells to differentiate into endoderm (Fuerer et al., 2014).

A Unifying Model for Vg1?

A few years after the mammalian ortholog of Vg1 was shown to increase Nodal's range and activity during the establishment of the left-right axis, a study suggested a similar role for zebrafish Vg1 (also called Dvr1/Gdf3). Annita Peterson and Joseph Yost targeted the zebrafish *vg1* gene with morpholinos, which resulted in zebrafish embryos with a randomization in left-right organ patterning (Peterson et al., 2013). Expression of *southpaw* (a zebrafish ortholog of *Nodal*) and its targets were reduced in the morphants, indicating that Nodal signaling was affected by the reduction in *vg1*. While the original discovery of *vg1* in zebrafish indicated that *vg1* expression is ubiquitous in the early embryo until mid-gastrulation (Helde and Grunwald, 1993), Peterson and colleagues uncovered a second period of expression in the node and LPM (akin to *Gdf1),* which could explain the phenotype (Peterson et al., 2013; Rankin et al., 2000; Wall et al., 2000). The authors ectopically expressed *vg1* and *southpaw* in *Xenopus* animal caps and

found that Vg1 and Southpaw in combination created a larger domain of ectopic mesodermal gene expression than Southpaw alone, causing the authors to conclude that Vg1 and Southpaw mutually enhance each other's activity. However, co-immunoprecipitation experiments failed to detect an interaction between Southpaw and Vg1, and contrary to previous results (Cheng et al., 2004), the Nodal inhibitors Charon and Lefty did not inhibit BMP-Vg1 activity (Peterson et al., 2013). The authors concluded that Southpaw and Vg1 'cooperate' to asymmetrically transfer a signal from the node to the LPM, which ultimately induces the asymmetric development of organs (Peterson et al., 2013).

These studies suggested a unifying model for Vg1's function: Vg1 cooperates with Nodal to pattern the left-right axis by increasing the range of Nodal. Vg1 might induce mesoderm induction when artificially fused to a different prodomain that facilitates its cleavage in the early embryo, however its endogenous function appeared to lie later in development. The only data missing for this model was genetic evidence of Vg1's function in zebrafish, *Xenopus* and chick.

The Nodal Signaling Pathway

Nodal is a member of the TGF-β family of ligands and is a key inducer of mesendoderm in vertebrate embryos and embryonic stem cells. Mouse and zebrafish *Nodal* mutants fail to generate mesendoderm (Conlon et al., 1991; 1994; Feldman et al., 1998; Zhou et al., 1993), and *Nodal* expression induces mesendodermal cell types in *Xenopus* (Jones et al., 1995; Joseph and Melton, 1997; Lustig et al., 1996; Smith et al., 1995; Takahashi et al., 2000), chick (Bertocchini and Stern, 2002), zebrafish (Bisgrove et al., 1999; Erter et al., 1998; Feldman et al., 1998; Gritsman et al., 2000; 1999; Meno et al., 1999; Sampath et al., 1998; Toyama et al., 1995) and

embryonic stem cells (Brandenberger et al., 2004; Camus et al., 2006; D'Amour et al., 2005; Hoveizi et al., 2014; Kubo, 2004; Parisi et al., 2003; Smith et al., 2008; Takenaga et al., 2007; Vallier et al., 2004; Yasunaga et al., 2005).

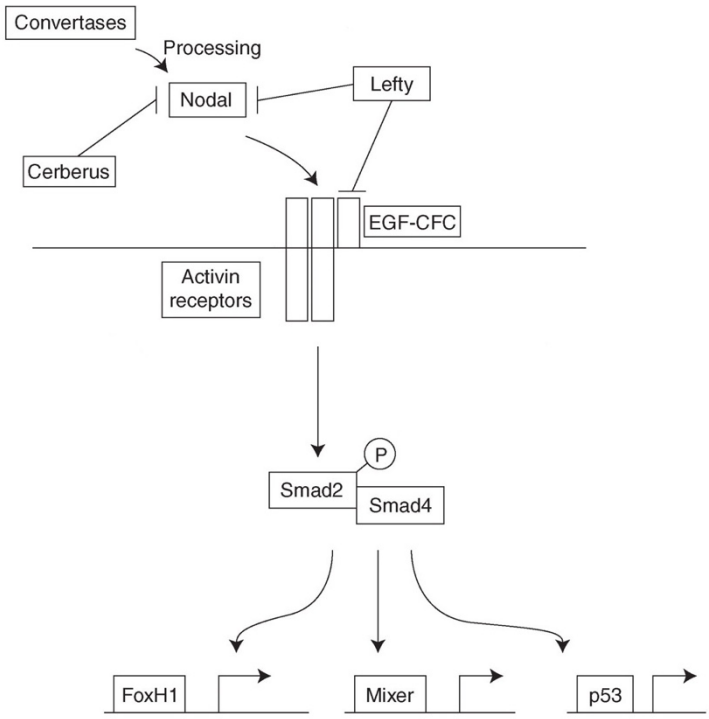


Figure 1-5. The Nodal signaling pathway.
Nodal is an extracellular ligand of the TGF-β family of proteins. It is processed to its mature form by convertases and inhibited by Lefty and Cerberus proteins. Nodal binds to a complex of Activin type I and type II receptors and an EGF-CFC coreceptor. This causes phosphorylation of Smad2 and nuclear entry of a pSmad2-Smad4 complex. Transcription of Nodal target genes is controlled in part by transcription factors FoxH1, Mixer and p53. Adapted from (Schier, 2009).

There is a second period of *Nodal* expression during the formation of the left-right axis (Levin et al., 1995; Long et al., 2003; Lowe et al., 1996; Pagán-Westphal and Tabin, 1998). *Nodal* is required for correct left-right axis patterning in mice (Brennan et al., 2002; Kumar et al., 2008; Saijoh et al., 2003) and zebrafish (Noël et al., 2013). In addition, *Nodal* expression correlates with the mispositioning of organs in mouse laterality mutants (Collignon et al., 1996; Lowe et al., 1996)

and UV-treated *Xenopus* embryos (Lohr et al., 1997). Misexpression of *Nodal* in chicks (Levin et al., 1997) and *Xenopus* (Sampath et al., 1997) randomizes the positioning of the heart and viscera.

In humans and mice one Nodal ligand induces both mesoderm formation and left-right patterning, but in zebrafish the task is divided between three ligands: Cyclops and Squint pattern the mesendoderm and Southpaw patterns the left-right axis. *Xenopus* possess 5 nodal-related genes: *Xnr1* to *5*.

Nodal proteins are secreted, extracellular ligands that bind to complexes of Activin type I and type II receptors. Activation of the serine/threonine kinases is dependent on the presence of the EGF-CFC coreceptor Cripto/Oep (Schier, 2003). Upon activation and phosphorylation of the receptors, the intracellular protein Smad2 is phosphorylated, pSmad2 binds to Smad4, pSmad2-Smad4 translocates to the nucleus, and it complexes with proteins such as FoxH1 and Mixer to bind to and activate a subset of Nodal target genes including *Nodal* itself (Shen, 2007) (Figure 1-5). Nodal activity is regulated by extracellular inhibitors, including the Lefty proteins, which are divergent members of the TGF-β family, and Cerberus/Charon.

The TGF-β family of proteins

TGF-β family members are synthesized as proproteins that dimerize through conserved cysteines and stabilize a characteristic cysteine knot structure (Constam, 2014) (Figure 1-6). Dimerization is mediated by intra-chain disulfide bonds (Thompson, 2003; Wang et al., 2016) and inter-chain disulfide bonds between prodomains (Shi et al., 2011; Wang et al., 2016) and/or

mature domains (Mi et al., 2015). TGF-β family members were originally thought to assemble as homodimers (Gray and Mason, 1990), but there are now multiple examples of proteins in this family that form heterodimers (Aono et al., 1995; Eimon and Harland, 2002; Fuerer et al., 2014; Guo and Wu, 2012; Hazama et al., 1995; Israel et al., 1996; Little and Mullins, 2009; Nishimatsu and Thomsen, 1998; Schmid et al., 2000; Shimmi et al., 2005; Suzuki et al., 1997; Tanaka et al., 2007). These include BMP2-7 heterodimers, which pattern the dorsoventral axis in zebrafish (Little and Mullins, 2009), Nodal-GDF1 heterodimers that pattern the left-right axis in mice (Fuerer et al., 2014; Tanaka et al., 2007), and the *Drosophila* BMP proteins Decapentaplegic (Dpp) and Screw that form broadly distributed homodimers and spatially restricted heterodimers, creating a biphasic signal for patterning the blastoderm embryo (Shimmi et al., 2005).

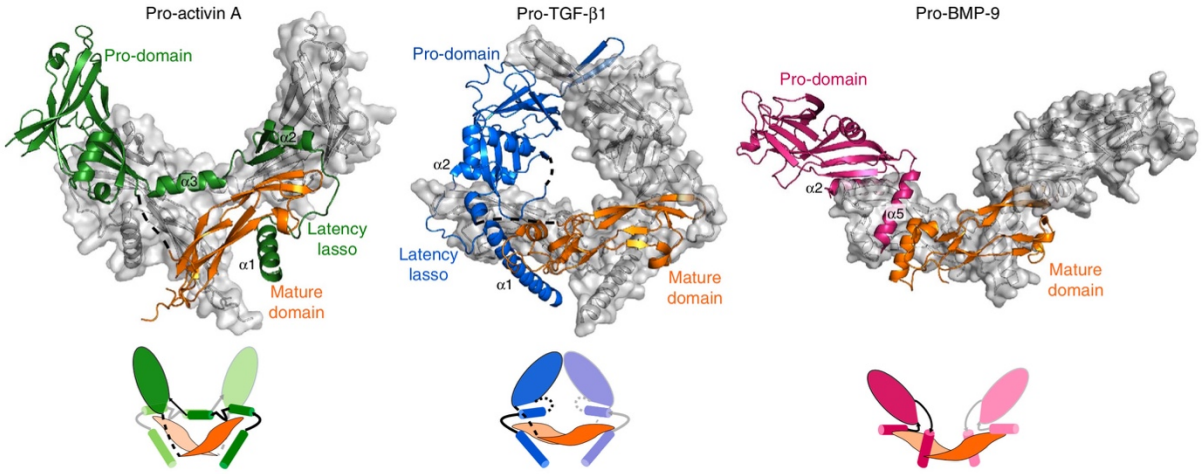


Figure 1-6. The structure of three TGF-β family proteins.
Each superfamily member consists of a prodomain and a mature domain separated by a basic cleavage site. The proteins dimerize through disulfide bonds either within or between peptide chains. From (Wang et al., 2016).

TGF-β-related proteins comprise a prodomain and a mature domain separated by a basic cleavage site. The mature domain is the signaling component of the dimer and the prodomain is required for folding, disulfide bond formation and export of mature dimers (Gray and Mason,

1990). Additional roles have been ascribed to specific TGF-β family prodomains. For instance, the BMP4 prodomain can remain non-covalently associated with the mature domain after cleavage, causing transport to the lysosome for degradation (Degnin et al., 2004). In contrast, cleavage of the Nodal prodomain reduces the stability of Nodal and decreases its signaling range (Le Good et al., 2005). The BMP4 prodomain promotes assembly of BMP4 with BMP7 and is necessary and sufficient for the formation of BMP4/7 heterodimers in mice (Neugebauer et al., 2015).

Proprotein convertases such as Furin and the serine protease PACE4 recognize cleavage sites with the motif RXXR (Constam, 2014). Furin and PACE4 cleave an RQRR site in Nodal, and this cleavage is required for Nodal signaling *in vivo* (Beck et al., 2002). However, under some conditions full-length, uncleavable Nodal is still able to signal (Ben Haim et al., 2006). The activity of proprotein convertases can occur extracellularly, for instance cleaving Nodal in the mouse embryo (Beck et al., 2002) or intracellularly, for instance processing zebrafish Southpaw during left-right patterning (Tessadori et al., 2015).

There are multiple mechanisms by which processing by proprotein convertases can be regulated. First, the expression of the enzymes can be restricted temporally or spatially. This occurs in the mouse embryo, where *Furin* and *Pace4* are expressed in the extraembryonic ectoderm while *Nodal* is expressed in the epiblast cells; only where the two protein pools meet does processing occur (Beck et al., 2002). Second, proprotein convertases can regulate their own maturation. For instance, convertases carry out an auto-cleavage event to mature from an inactive zymogen to an active enzyme. Furin undergoes a second auto-cleavage event that is pH-dependent and changes its conformation (Dillon et al., 2012). Third, to regulate processing of specific substrates rather than all proprotein targets, TGF-β-related proteins can utilize tissue-

specific cleavage motifs. This occurs in *Drosophila*, where one Dpp cleavage motif is required for the protein's function in the wing disc while another is required for activity in the midgut (Sopory et al., 2010). Fourth, proteins that bind the proprotein convertase cleavage site of TGF-β-related proteins can inhibit processing, for instance in the case of E-selectin, which inhibits cleavage of TGF-β (Yang et al., 2010). Finally, control of processing can be achieved with spatial compartmentalization of ligands and convertases. For instance, the Nodal coreceptor Cripto binds both Nodal and Furin, and is thought to localize Nodal processing to the cell surface (Blanchet et al., 2008; Constam, 2014).

Figure 1-7. *cyclops* and *squint* single mutants exhibit mild floorplate and cyclopia phenotypes.
Adapted from (Dougan et al., 2003).

Zebrafish mesendoderm patterning

The first major patterning event in zebrafish development is the specification of the mesendoderm, which occurs after the maternal-to-zygotic transition (MZT). The initiation of

zygotic transcription coincides with the formation of an extraembryonic tissue, the yolk syncytial layer (YSL), from which Mxtx2 induces the expression of the Nodal genes *cyclops* and *squint* (Hong et al., 2011; Xu et al., 2012). The roles of *cyclops* and *squint* in zebrafish mesendoderm development were elucidated using single and double mutants (Feldman et al., 1998; Hatta et al., 1991; Heisenberg and Nüsslein-Volhard, 1997). *cyclops* and *squint* single mutants exhibit mild embryonic phenotypes, with defects in prechordal plate development and cyclopia (Figure 1-7) (Hatta et al., 1991; Heisenberg and Nüsslein-Volhard, 1997). By comparison, *cyclops;squint* double mutants are much more severe: the embryos fail to generate the derivatives of the mesendoderm, including the notochord, somites, blood, heart and gut (Feldman et al., 1998) (Figure 1-8). This phenotype is shared with embryos lacking the Nodal coreceptor Oep (Gritsman et al., 1999) or the Nodal signal transducer Smad2 (Dubrulle et al., 2015). The single and double mutant phenotypes reveal that *cyclops* and *squint* possess partially overlapping roles in patterning the mesendoderm.

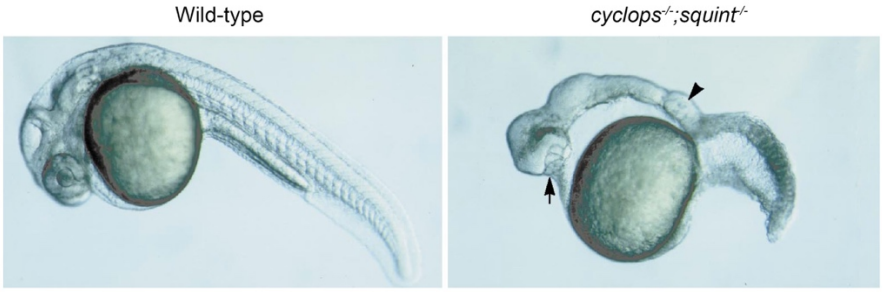


Figure 1-8. *cyclops*$^{-/-}$*; squint*$^{-/-}$ double mutants lack derivatives of the mesendoderm.
Adapted from (Feldman et al., 1998)

After the induction of *cyclops* and *squint* expression, the proteins diffuse through the blastula, inducing the expression of many target genes that pattern the mesendoderm tissues, including transcription factors and cytoskeletal components (Bennett et al., 2007; Fodor et al., 2013; Liu

et al., 2011). Embryological experiments suggest that Cyclops acts over a short range, Squint acts over a medium range, and the inhibitors Lefty1 and 2 act over a long range (Chen and Schier, 2001; 2002) (Figure 1-9). This has led to the suggestion that Nodal acts as a morphogen and operates according to a reaction-diffusion system. To test this model, members of the Schier lab measured the distribution, clearance and diffusivity properties of the zebrafish Nodals and Leftys *in vivo*, and found them to be consistent with such a system (Muller et al., 2012). In order to understand how Nodal can induce a concentration-dependent effect on gene expression, Julien Dubrulle and colleagues in the Schier lab measured the effect of Nodal concentration on gene expression. They found that the concentration of Nodal alone was unable to predict where and when target genes are induced: instead, each target gene integrates its rate of transcription with the onset of expression to determine its spatial expression range (Dubrulle et al., 2015). Current models of zebrafish mesendoderm patterning have assumed that Cyclops and Squint signal as homodimers (Shen, 2007).

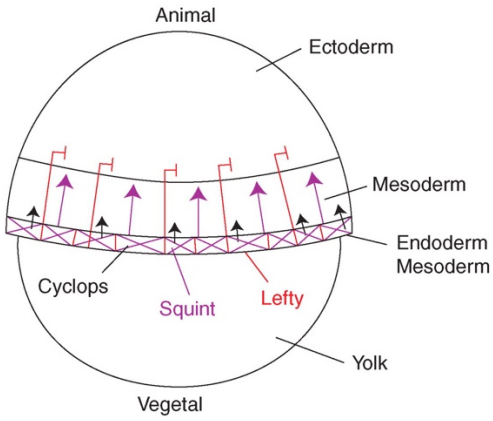


Figure 1-9. Nodal signaling in the zebrafish embryo.
Nodal ligands Cyclops and Squint are secreted from the margin of the blastula embryo and diffuse through the embryo, inducing mesendoderm. Lefty, an inhibitor of Nodal, has a higher diffusivity than Cyclops and Squint. From (Schier, 2009).

Aims and Approaches

Over 30 years ago the localized, maternal mRNA *vg1* was identified in *Xenopus*, and it was predicted to be an essential inducer of vertebrate mesoderm. However, though the *Xenopus,* zebrafish, chick and mouse mature domains of Vg1 were shown to possess mesoderm-inducing properties when fused to heterologous prodomains, an endogenous function for Vg1 in mesoderm induction was not identified. Nodal activity and *Nodal* mutants could account for the mouse and zebrafish mesodermal patterning systems, and the field was missing the genetic interrogation of Vg1 function.

In 2012, at the beginning of my graduate work, the standard method to test zebrafish gene function was with morpholinos [8] (MOs) (Nasevicius and Ekker, 2000), which are DNA oligonucleotides that sterically hinder translation or splicing. However, they have critical shortcomings, for instance they can induce off-target effects (Eisen and Smith, 2008), and the quality of knockdown cannot be quantified. Since the development of the CRISPR/Cas9 system, a number of published MO phenotypes have been shown to be artifacts (Kok et al., 2014; Schulte-Merker and Stainier, 2014; van Impel et al., 2014). Before CRISPR/Cas9 methods became available in zebrafish, we developed a method to knock down RNA in zebrafish using antisense oligonucleotides (ASOs), which cause RNase H-mediated degradation of target mRNAs, and therefore allow the effectiveness of knockdown to be quantified (see Chapter 2). Following the development of CRISPR/Cas9 methods, we developed a web tool to design guide RNAs (gRNAs), as it became clear that the effectiveness of a knockout was predominantly

[8] There was also 'TILLING' (Moens et al., 2008), a strategy that generates random mutants, but it requires luck that a mutation is generated in your gene-of-interest and that it's a loss-of-function allele.

determined by the quality of the gRNA (Gagnon et al., 2014; Wang et al., 2014) (see Chapter 3). Subsequent developments in the CRISPR/Cas9 technology, combined with large-scale studies of gRNA efficiency, illuminated additional rules for gRNA design that we incorporated into a second version of our web tool. (see Chapter 4). Finally, with the availability of efficient CRISPR/Cas9 knockout strategies in zebrafish, we knocked out *vg1* in zebrafish to determine the gene's function (Chapter 5).

Chapter 2: Antisense Oligonucleotide-Mediated Transcript Knockdown in Zebrafish

Preface

This chapter was previously published in PLOS ONE on October 5 2015 (Pauli et al., 2015). Andrea Pauli and Alexander F. Schier conceived the project; Andrea Pauli and I performed the experiments; Andrea Pauli, Alexander F. Schier and I designed the experiments and wrote the paper; and Kim A. Lennox and Mark A. Behlke contributed reagents.

Pauli, A.*, Montague, T.G.*, Lennox, K.A., Behlke, M.A., and Schier, A.F. (2015). Antisense Oligonucleotide-Mediated Transcript Knockdown in Zebrafish. PLoS ONE *10*, e0139504.

*equal contribution

Abstract

Antisense oligonucleotides (ASOs) are synthetic, single-strand RNA-DNA hybrids that induce catalytic degradation of complementary cellular RNAs via RNase H. ASOs are widely used as gene knockdown reagents in tissue culture and in *Xenopus* and mouse model systems. To test their effectiveness in zebrafish, we targeted 20 developmental genes and compared the morphological changes with mutant and morpholino (MO)-induced phenotypes. ASO-mediated transcript knockdown reproduced the published loss- of-function phenotypes for *oep, chordin, dnd, ctnnb2, bmp7a, alk8, smad2* and *smad5* in a dosage-sensitive manner. ASOs knocked down both maternal and zygotic transcripts, as well as the long noncoding RNA (lncRNA) *MALAT1*. ASOs were only effective within a narrow concentration range and were toxic at higher concentrations. Despite this drawback, quantitation of knockdown efficiency and the ability to degrade lncRNAs make ASOs a useful knockdown reagent in zebrafish.

Introduction

One effective strategy for interrogating gene function is to disrupt the generation of a gene product by knockdown or knockout. Knockout technologies, such as CRISPR/Cas9 and homologous recombination, alter the DNA locus of the gene by either introducing a premature stop codon or removing the entire locus (Figure 2-1A) (Cong et al., 2013; Mali et al., 2013c). Knockdown methods, on the other hand, such as RNAi, siRNAs and modified antisense oligonucleotides (Kurreck, 2003; Prakash, 2011), target the mRNA rather than alter the DNA. While it is most reliable to infer gene function by generating a mutant organism, knockdown reagents can provide a more immediate assessment of gene function and can be used to target gene products without disrupting regulatory DNA elements.

Because siRNAs have been used with limited success in zebrafish (Kelly and Hurlstone, 2011), an alternative knockdown reagent, morpholino oligonucleotides (MOs), has been the preferred zebrafish knockdown reagent (Nasevicius and Ekker, 2000) (Figure 2-1A). MOs are synthetic oligonucleotides composed of around 25 nucleotides that possess a morpholine ring instead of the ribose ring found in DNA and RNA, with non-ionic phosphorodiamidate linkages instead of a phosphodiester backbone. This provides nuclease resistance and allows them to bind RNA molecules through conventional Watson-Crick base pairing. MOs interfere with gene function by either sterically hindering translation (Summerton, 1999) or splicing (Draper et al., 2001) or functioning as target protectors (Choi et al., 2007) (Figure 2-1A). However, they do not degrade the target RNA, impeding quantitation of knockdown efficiency. This is particularly an issue for translation-blocking RNAs, which do not create changes to the splicing pattern of the target RNA. MOs can also cause significant off-target activity, resulting in misleading phenotypic artifacts (Eisen and Smith, 2008). An increasing number of mutants that have been generated

subsequent to MO-based studies are either non-phenotypic or produce alternative phenotypes from the published MO-mediated knockdowns (Kok et al., 2014; Rossi et al., 2015; Schulte-Merker and Stainier, 2014; van Impel et al., 2014). For this reason, the availability of an alternative knockdown reagent would be of use to complement MO studies.

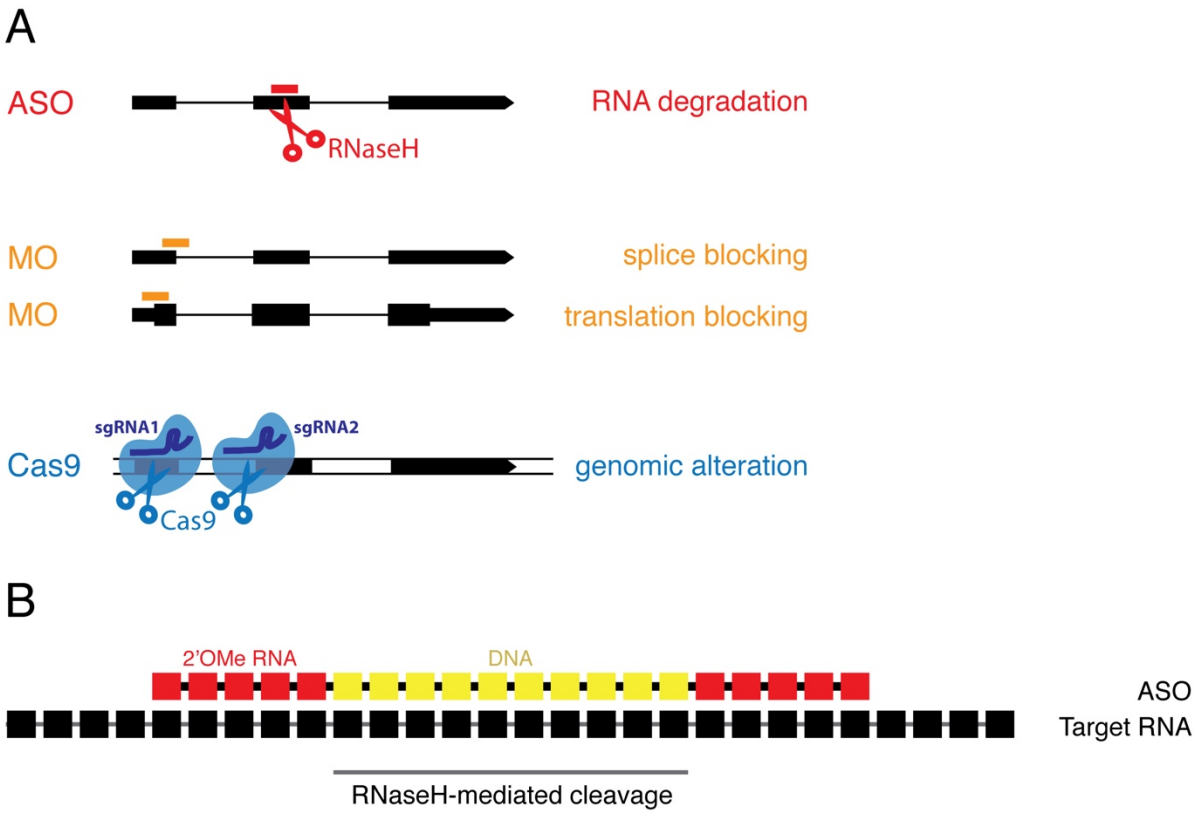


Figure 2-1. Overview of knockdown and loss-of-function technologies in zebrafish.
A) Antisense oligos (ASOs, red) degrade target RNA, morpholinos (MOs, orange) either block splicing or inhibit translation, and Cas9-sgRNA complexes (blue) create double-strand breaks in DNA leading to genomic alterations. B) ASOs are RNA-DNA hybrid oligonucleotides containing 10 central DNA nucleotides flanked by 5 2'O-Methyl (2'OMe) modified RNA nucleotides on either side (5-10-5 arrangement). Individual nucleotides in the ASO are linked by phosphorothioate bonds to increase stability.

The present study employs RNA-DNA hybrid ASOs (also known as gapmers), which are chimeric oligonucleotides containing 10 DNA nucleotides flanked by 5 2'O-Methyl (2'OMe) RNA residues (5-10-5 arrangement) (Ideue et al., 2009; Vickers et al., 2003). The phosphate backbone in the

DNA and RNA is replaced with phosphorothioate bonds to increase ASO stability, and addition of 2'OMe RNA modifications increases nuclease resistance (Figure 2-1B). The RNA and DNA portions in the hybrid molecule serve distinct functions: the RNA nucleotides increase affinity to complementary RNAs, while the central DNA stretch serves as a guide for RNase H-mediated degradation of the complementary RNA. This strategy has been widely used as a knockdown approach in Xenopus (Heasman et al., 1994; Mir and Heasman, 2008; Torpey et al., 1992; Zhang et al., 1998a), tissue culture (Dimitrova et al., 2014), mouse models (where gapmer ASOs have reversed disease phenotypes (Kordasiewicz et al., 2012; Wheeler et al., 2012)), and it has shown promise in gene therapy (Bennett and Swayze, 2010; Meng et al., 2014). RNase H degradation is catalytic, and the ASO itself is recycled, meaning that a single ASO can direct degradation of multiple copies of the target RNA. In contrast, a single steric-blocking MO can only bind and inactivate a single target RNA molecule.

ASOs offer a number of advantages over MOs. First, they cause degradation of the transcript via RNase H. Thus, the efficiency of the knockdown can be quantified. Second, due to degradation of the RNA, rather than prevention of splicing or translation, they can be used to eliminate spliced maternal RNAs. Third, they can target both protein-coding and noncoding RNAs due to activity in the nucleus: ASOs have been shown to shuttle between the cytoplasm and nucleoplasm (Lorenz et al., 2000), and can efficiently target nuclear-retained long noncoding RNAs (lncRNAs) (Prasanth et al., 2005; Wheeler et al., 2012) and nascent RNAs (Vickers et al., 2003). Finally, ASOs are significantly cheaper than MOs, with an average current cost (as of July 2015) of ~$200 (rather than ~$400). Additionally, only 1/10-1/100 of the MO concentration is required for ASO experiments. Therefore, ASOs combine several properties (quantifiable knockdown rates, specificity, efficiency, nuclear activity and persistence *in vivo* (Wheeler et al., 2012)) that highlight their potential as alternatives to MO-mediated knockdown.

To test the feasibility of using ASOs as an alternative knockdown reagent in zebrafish, we targeted 18 genes with known embryonic loss-of-function phenotypes. ASO-mediated knockdown reproduced the published phenotypes for 8 developmental protein-coding genes (*one- eyed pinhead (oep), chordin, dead-end (dnd), beta-catenin 2 (ctnnb2), bmp7a, alk8, smad2* and *smad5*). In addition, ASOs substantially knocked down a lncRNA, *MALAT1*. These results establish ASOs as useful knockdown reagents in zebrafish.

Results

In order to test ASO efficacy in zebrafish we designed ASOs against the Nodal coreceptor *oep*. *oep* was chosen as a test candidate because it is expressed both maternally and zygotically and has dosage-dependent phenotypes. The complete phenotype only becomes apparent when both maternal and zygotic *oep* (MZ*oep*) are inactivated (Gritsman et al., 1999). 5 ASOs were designed against different regions of the *oep* mRNA using *in silico* RNA-folding predictions (see Materials and Methods) (Hofacker, 2003). Each ASO was injected at multiple concentrations (1 to 500 pg/embryo) into single-cell zygotes. Two ASOs caused *oep*-specific phenotypes when injected between 30 and 150 pg (Figure 2-2A, 2-2D and 2-2F). We found that all ASOs, regardless of their nucleotide sequence, were toxic to embryos when injected above 200 pg, causing deformation and death. This toxicity was not substantially ameliorated by co-injecting a *p53* MO (Figure 2-2E) (Langheinrich et al., 2002). The ASO targeting a 20 nt region close to the 3'end of the *oep* ORF was most effective: injection of 30–60 pg of this ASO resulted in partial loss-of-function phenotypes, resembling partial *oep* loss-of-function mutants, and injection of 100–150 pg of the ASO caused phenotypes indistinguishable from complete loss-of-function MZ*oep* mutants (Figure 2-2A; quantitation of phenotypes in Figure 2-2F).

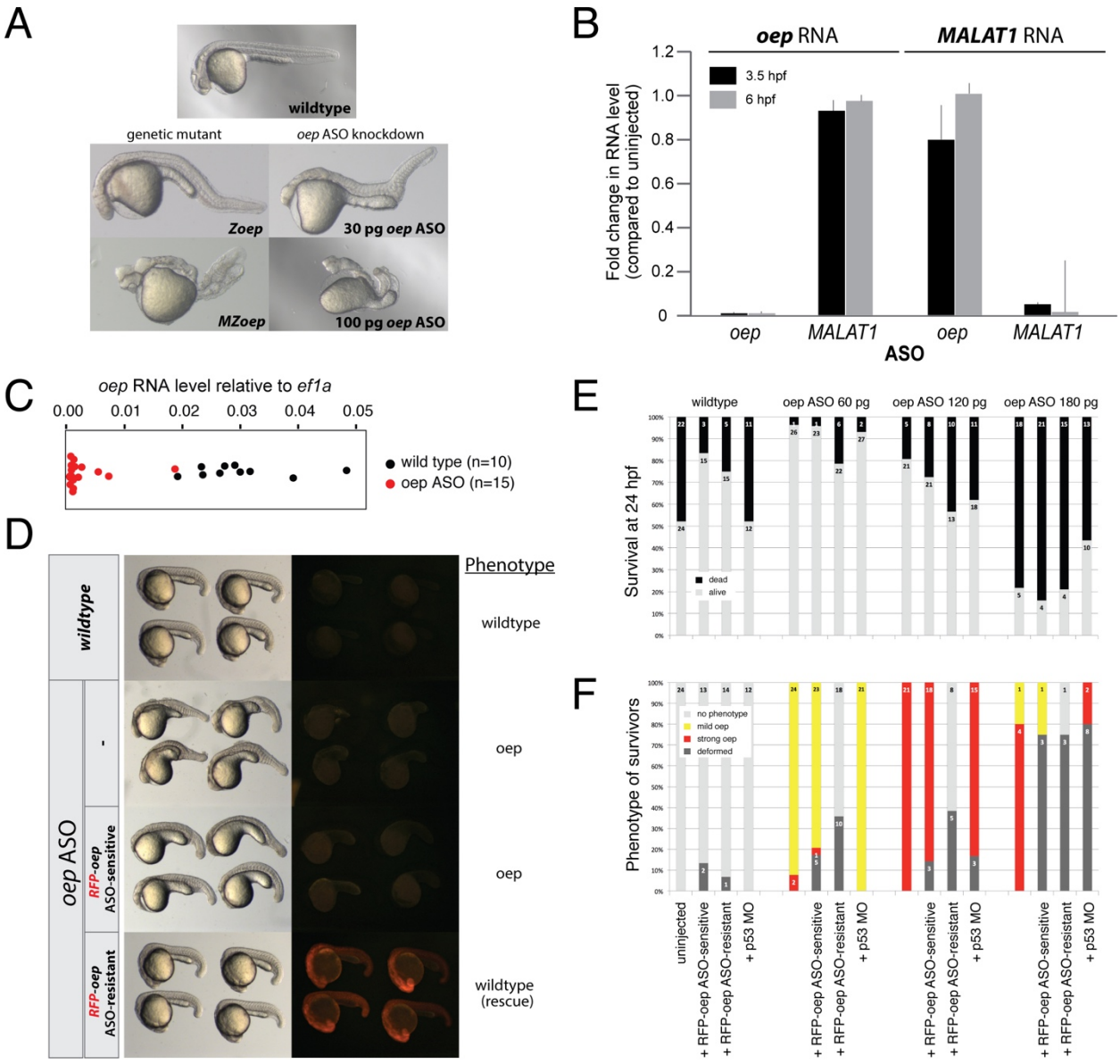


Figure 2-2. Efficiency and specificity of oep ASO-induced oep (one-eyed-pinhead) mutant phenotypes.

A) Injection of oep ASO induces dosage- dependent oep phenotypes that resemble zygotic (Zoep) and maternal-zygotic (MZoep) oep genetic mutants. B) oep ASO and MALAT1 ASO knockdowns are specific. The RNA levels of oep and MALAT1 were measured by qPCR in oep ASO (100 pg) and MALAT1 ASO (80 pg)-injected embryos. Shown is the fold change in RNA level compared to WT (wildtype), normalized to ef1a (error bars: standard deviation of the mean of 3 independent experiments). C) qPCR-based measurement of oep RNA levels in individual oep ASO (100 pg)-injected (red) or uninjected (black) embryos at shield stage (6 hpf). D) Rescue of oep ASO-induced oep phenotypes by coinjection of an oep ASO-resistant RFP-oep fusion mRNA. Note that the oep ASO-sensitive RFP-oep fusion mRNA is efficiently knocked down (no red fluorescence) and does not rescue. E) Quantitation of survival at 24 hpf and F) quantitation of phenotypic strength in survivors at 24 hpf in the presence versus absence of p53 (p53 MO-injected embryos) or RFP-oep fusion mRNA rescue construct. The number of embryos in each category is indicated.

Quantitative real-time PCR (qPCR) confirmed the efficient and concentration-dependent knockdown of *oep* mRNA: 1–3% of *oep* mRNA remained at 3.5 hours post fertilization (hpf) and shield stage (6 hpf) (Figure 2-2B). Because a small number of *oep* ASO-injected embryos did not show a specific phenotype at 24 hpf, we tested knockdown efficiencies in individual embryos to correlate variability in phenotype with variability in knockdown levels. We found that the level of *oep* mRNA knockdown across individual embryos at shield stage was in line with the variability in phenotypes at 24 hpf (7/21 strong *oep* phenotype, 11/21 dead, 2/21 partial *oep* phenotype, 1/21 deformed, versus 13/15 ASO-injected embryos with a >3-fold reduction in *oep* mRNA levels) (Figure 2-2C). The observed phenotype was specific to the knockdown of *oep* mRNA as injection of an *oep* mRNA containing 7 nucleotide changes within the ASO recognition site was able to rescue the ASO-induced phenotype (Figure 2-2D and 2-2F). Moreover, quantitation of the levels of oep and MALAT1 RNA in oep ASO-injected and *MALAT1* ASO-injected embryos (see below) revealed that each ASO was specifically knocking down the target RNA and not the unrelated RNA (Figure 2-2B).

To assess the perdurance of ASO-mediated transcript knockdown in zebrafish and to test whether ASOs could be used to knock down non-coding RNAs in zebrafish, we chose to target *MALAT1*. This lncRNA is one of the most highly expressed transcripts during development, it localizes to the nucleus, and has been efficiently targeted with ASOs in mouse models (Wheeler et al., 2012). Notably, 2/3 ASOs against *MALAT1* reduced *MALAT1* transcript levels to 1–10% of wild-type levels (Figure 2-2B and data not shown), and substantial knockdown persisted for at least 5 days after injection of *MALAT1* ASO but not after injection of an unrelated ASO (*dnd* ASO) (Figure 2-3). Consistent with the lack of detectable phenotypes in *MALAT1* knockout mice (Eißmann et al., 2014; Nakagawa et al., 2012; Zhang et al., 2012), development proceeded normally in zebrafish embryos depleted of *MALAT1* RNA. Nevertheless, the perdurance of

MALAT1 RNA knockdown shows that ASOs can be useful reagents to cause sustained knockdown of zebrafish mRNAs and lncRNAs for several days post injection.

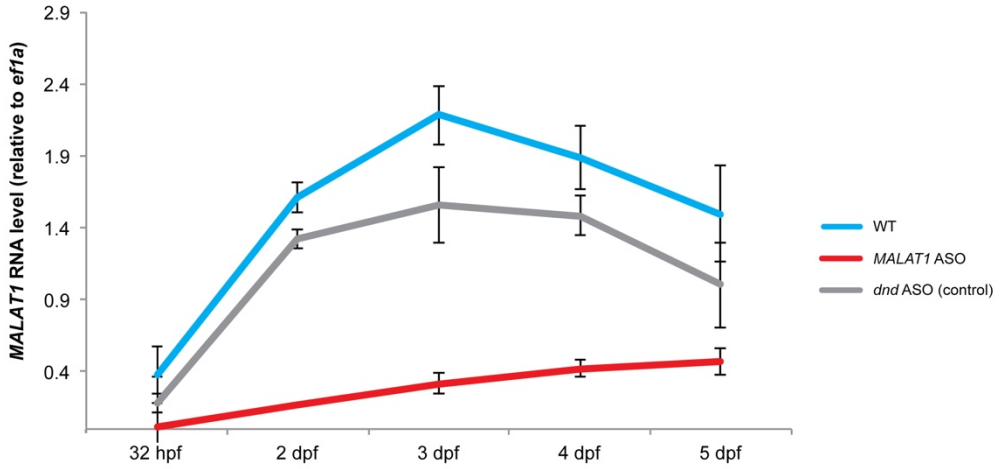


Figure 2-3. ASO-mediated RNA knockdown persists for several days.
MALAT1 and *ef1a* RNA levels were measured for 5 days post injection by qPCR in uninjected (= wildtype), *MALAT1* ASO (100 pg) and *dnd* ASO (25 pg)-injected embryos. Shown are *MALAT1* RNA levels normalized to *ef1a*. Error bars show standard deviation of the mean of 2 independent experiments (10 embryos each).

To test whether ASOs could be used as a general knockdown reagent in zebrafish, we selected an additional 17 embryonically expressed genes with known mutant phenotypes (see Table 1). For 10 of the targeted genes, we assessed whether ASO injection leads to degradation of the target RNA by qPCR. Each ASO was injected at 4 concentrations, and samples were collected at the peak times of the corresponding gene's expression during the first 36 hours of development to assess target mRNA levels. After monitoring development of the remaining embryos for 24 hours to determine the highest concentration of ASO injection that produced minimal embryonic death, we processed the corresponding samples for qPCR to determine if there was knockdown of the target mRNA. We observed a strong knockdown for *alk8, smad5, smad2, chordin* and *bmp7a*, a partial knockdown for *ntla* and *wnt11* and no knockdown for *nacre,*

tolloid and *wnt5b* (Figure 2-4A). Notably, knockdown efficiency correlated with phenotype. First, ASOs that caused efficient knockdown of their target mRNAs (e.g. *alk8, smad5, smad2, chordin* and *bmp7a* ASOs) reproduced published mutant and knockdown phenotypes in surviving embryos (Bauer et al., 2001; Dick et al., 2000; Halpern et al., 1993; Hammerschmidt et al., 1996; Hild et al., 1999; Mintzer et al., 2001; Schmid et al., 2000; Schulte-Merker et al., 1997) (Figure 2-4B). Second, 2 of the 3 ASOs that failed to knock down their cognate mRNAs (*nacre* and *tolloid* ASOs) did not produce a specific phenotype: they either caused no phenotype or resulted in embryonic deformation and death. Results for *wnt5b* ASO, the third ASO that failed to knock down its target gene based on qPCR, were inconclusive due to high variability and high toxicity (data not shown). Third, *ntla* and *wnt11* ASO injection lead to partial knockdown and reproduced the published mutant phenotype (Halpern et al., 1993) in a smaller proportion of embryos. Although we found that in most cases a successful knockdown (measured by qPCR) predicted a loss-of-function phenotype, there was a single case in which ASO knockdown achieved up to 90% reduction in the target mRNA, and yet the injected embryos had few gene-specific phenotypes (*oep* ASO#1). To ensure the knockdown was specific to the ASO that was injected, we assessed the level of *smad5*, *bmp7a* and *alk8* RNA in uninjected embryos and those injected with either a *smad5*, *bmp7a* or *alk8* ASO. Indeed, only the RNA corresponding to the injected ASO was reduced (Figure 2-4C). Together, these results suggest that qPCR can be a useful assay to pre-screen ASOs for their ability to knock down target mRNAs. To test whether ASOs could be an efficient knockdown reagent for transcripts required in only a small subset of cells, we chose to target *dnd*, a germ plasm component required for germ cell migration and survival (Weidinger et al., 2003). Loss of *dnd* functionality leads to loss of germ cells — a phenotype that can be monitored by labeling germ cells with eGFP (Ciruna et al., 2002). Injection of 25 pg of *dnd* ASO resulted in complete germ cell loss at 24 hpf (Figure 2-4D), reproducing the phenotype seen with MO-mediated translational inhibition (Ciruna et al., 2002).

In summary, after targeting 20 genes with 50 individual ASOs (see Table 1), our results reveal that ASOs can be an effective knockdown reagent for protein-coding and noncoding transcripts.

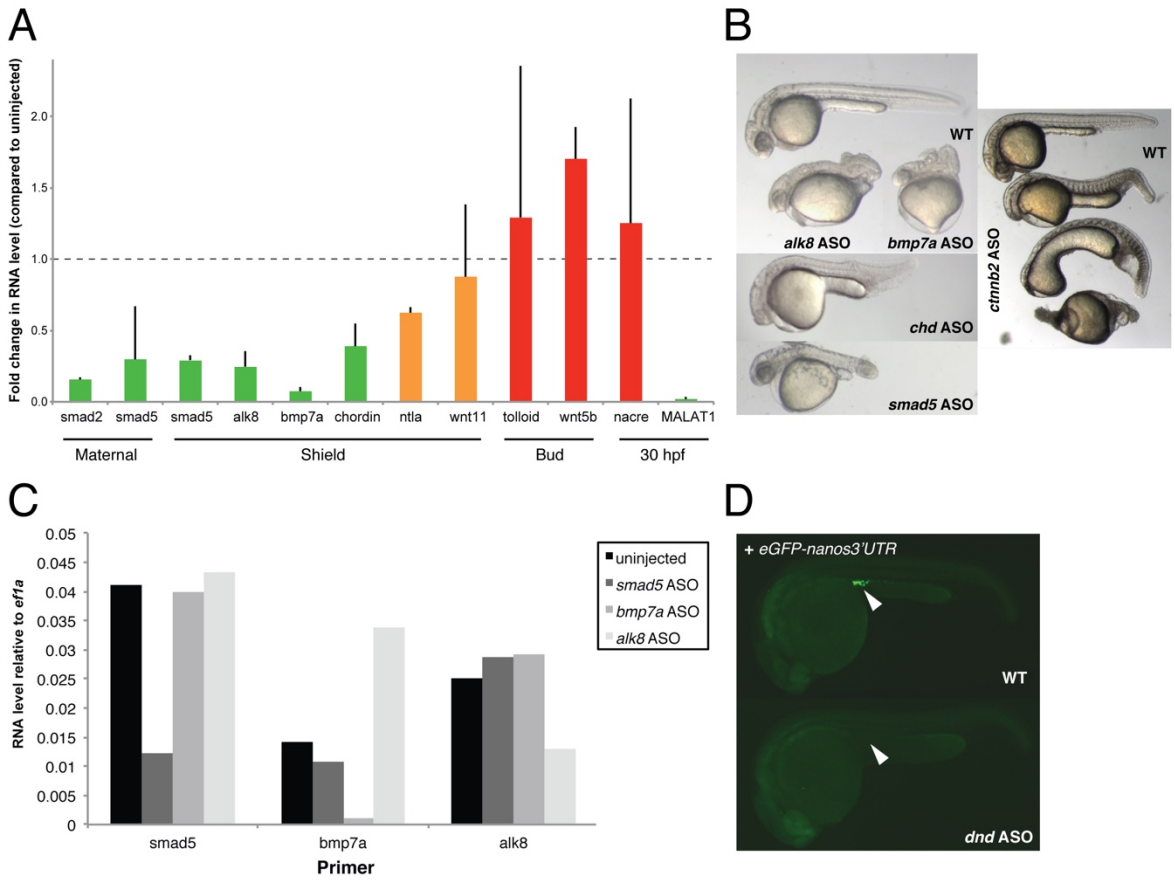


Figure 2-4. ASO-mediated RNA knockdown correlates with phenotype.
A) The RNA level of the gene corresponding to each ASO was measured (compared to WT, normalized to *ef1a*) and correlated to the presence of a phenotype: green = reproduced published phenotype (in case of *MALAT1*: no mutant phenotype), orange = reproduced published phenotype in a smaller percentage of embryos, red = did not produce a phenotype. Injected amount of ASO per embryo: 50 pg *smad2* ASO, 50 pg *smad5* ASO, 50 pg *alk8* ASO, 50 pg *bmp7a* ASO, 150 pg *chordin* ASO, 100 pg *ntla* ASO, 50 pg *wnt11* ASO, 150 pg *tolloid* ASO, 100 pg *wnt5b* ASO, 100 pg *nacre* ASO, 150 pg *MALAT1* ASO. Error bars show standard deviation of the mean of 2 independent experiments (10 embryos each). B) ASO-generated phenotypes for *alk8, bmp7a, chordin, smad5* and *ctnnb2* (shown are 3 embryos representative of the different severities of *ctnnb2* ASO-induced phenotypes). C) ASOs only target the cognate RNA, and not unrelated RNAs. qPCR-based assessment of ASO specificity to their cognate target genes (10 embryos each). D) *dnd* ASOs block germ cell formation. Germ cells were labeled by injection of 80 pg *eGFP-nanos3'UTR* mRNA. Coinjection of 25 pg *dnd* ASO caused complete loss of green germ cells (white arrow).

Discussion

This study reveals that ASOs can be an effective RNA knockdown reagent for zebrafish. Three observations establish the suitability of ASO use in zebrafish. First, ASOs are specific, because (i) we can rescue the phenotype caused by ASO-mediated knockdown of *oep* with an ASO-resistant *oep* mRNA, and (ii) ASOs cause substantial reduction of the target mRNA without corresponding reductions in unrelated mRNAs. Second, ASOs can target maternal, zygotic, nuclear and cytoplasmic coding and noncoding RNAs, since we were able to target maternal *oep, smad5* and *dnd*, zygotic *oep, chordin, bmp7a* and *alk8*, and the nuclear noncoding RNA *MALAT1* (Bernard et al., 2010). Third, unlike MOs, ASO-mediated knockdown can be quantified by qPCR, which generally allows phenotypes to be correlated with a reduction in expression of the target RNA.

Although our experiments highlight the power of ASOs to efficiently knock down target RNAs, there are three limitations associated with ASO use. First, ASOs are toxic to the embryo when injected above 200 pg, irrespective of the nucleotide sequence. The toxicity is usually manifested as cell death followed by embryonic lethality. Thus, high levels of ASOs should be used with caution if the resulting phenotype could be linked to general toxicity in the embryo (Levin, 1999). Second, ASOs appear to only be effective in a narrow concentration range. We found that some ASOs produced no knockdown at moderate levels (e.g. 100 pg) but were toxic and induced lethality at higher levels (200 pg). Therefore, it is important to inject at least 2–3 different concentrations to capture the ideal concentration range and produce efficient target knockdown. Third, not all ASO designs produced knockdown phenotypes. We designed 2–5 ASOs for each gene but, with a few exceptions such as *oep* and *dnd*, we generally found that either all or no ASOs were effective for a particular gene. Of the 50 ASOs we designed for 20 genes, we found

that 21 (42%) effectively targeted 11 genes, of which 16 (32%) caused loss-of-function-like phenotypes and 5 (10%) caused partial knockdown phenotypes (see Table 1). Steric-blocking MOs are only effective when targeting sequences around the translational start, limiting design options. By contrast, ASOs designed against any region of the target RNA can lead to target degradation. One current limitation of ASO design is the lack of an effective target prediction algorithm. Currently, ASOs are designed using antisense reagent target prediction strategies (Pan and Clawson, 2006) that have limited predictive power (MAB, personal communication), necessitating the need to empirically test each ASO for knockdown efficiency. However, as ASO use becomes more widespread, target prediction tools will likely become available. A good example of such a development comes from the RNAi field, where machine learning algorithms have been trained to improve siRNA knockdown rates (Peek and Behlke, 2007).

Due to these limitations, ASOs — like all knockdown reagents — should be used with caution and with appropriate controls. ASOs allow fast, cost-effective and preliminary assessment of gene function, but we do not recommend that ASOs be used to firmly establish the function of previously uncharacterized genes, unless phenotypes are confirmed with genetic mutants (Gagnon et al., 2014; Hwang et al., 2013). Where mutant phenotypes are known, ASOs and other knockdown reagents that recapitulate the mutant phenotypes can be valuable, for instance to knock down maternal transcripts or non-coding RNAs, or create clutches of embryos that all have the same or similar phenotypes. Thus, ASOs are a useful addition to the zebrafish knockdown reagent toolkit.

Table 2-1. Overview of ASO experiments.

Gene name (Mutant name)	Expression pattern	No. of effective ASOs/designed	Assay of ASO knockdown efficiency	Reference for published morphology
PROTEIN-CODING				
alk8 (lost-a-fin)	maternal + zygotic	**2/2**	morphology; RT-PCR	(Bauer et al., 2001; Mintzer et al., 2001)
bmp2b (swirl)	zygotic	0/2	morphology; RT-PCR	(Kishimoto et al., 1997; Lele et al., 2001; Martínez-Barberá et al., 1997; Nguyen et al., 1998)
bmp7a (snailhouse)	zygotic	**2/2**	morphology; RT-PCR	(Dick et al., 2000; Lele et al., 2001; Schmid et al., 2000)
chordin (dino)	zygotic	**2/2**	morphology; RT-PCR	(Fisher and Halpern, 1999; Nasevicius and Ekker, 2000; Schulte-Merker et al., 1997)
ctnnb2 (ichabod)	maternal	**1/2**	morphology (ventralization)	(Bellipanni et al., 2006)
cx41.8 (leopard)	expressed in late larvae + adults	0/3	morphology (spotted pigment pattern)	(Watanabe et al., 2006)
dnd/dead-end	germ-cell specific	**1/4; 1*/4**	morphology (ablation of germ cells at 24 hpf)	(Ciruna et al., 2002; Weidinger et al., 2003)
hcrt/hypocretin	expressed from 1dpf	0/3	in situ (hcrt expressing cells)	(Prober et al., 2006)
mitfa (nacre)	expressed from 1dpf	0/2	morphology; RT-PCR	(Lister et al., 1999; Nasevicius and Ekker, 2000)
oep/one-eyed pinhead	maternal + zygotic	**2/5**	morphology; RT-PCR	(Gritsman et al., 1999; Schier et al., 1997; Zhang et al., 1998b)
slc24a5 (golden)	expressed from 1dpf	0/3	morphology (loss of pigment)	(Lamason et al., 2005)
smad2	maternal + zygotic	**2/3**	morphology; RT-PCR	(Dubrulle et al., 2015)
smad5 (somitabun)	maternal + zygotic	**2/2**	morphology; RT-PCR	(Hild et al., 1999; Lele et al., 2001)
ta-T/ntla/no-tail a	zygotic	**2*/2**	morphology; RT-PCR	(Halpern et al., 1993; Nasevicius and Ekker, 2000)
toddler/apela/ELABELA	zygotic	0/2	morphology (gastrulation + heart defect)	(Chng et al., 2013; Pauli et al., 2014)

tll1/tolloid/mini fin	zygotic	0/2	morphology; RT-PCR	(Connors et al., 1999; Lele et al., 2001)
wnt11 (silberblick)	zygotic	**2*/2**	morphology; RT-PCR	(Heisenberg et al., 2000; Lele et al., 2001)
wnt5b (pipetail)	zygotic	0/2	morphology; RT-PCR	(Lele et al., 2001; Rauch et al., 1997)
NON-CODING				
MALAT1	highly expressed from 1 dpf	**2/3**	RT-PCR	
miR-126	expressed from 1 dpf; endothelia	0/2	morphology (blood formation)	(Fish et al., 2008; Wang et al., 2008)

Materials and Methods

Ethics statement

All vertebrate animal work was performed at the facilities of Harvard University, Faculty of Arts & Sciences (HU/FAS). The HU/FAS animal care and use program maintains full AAALAC accreditation, is assured with OLAW (A3593-01), and is currently registered with the USDA. This study was approved by the Harvard University/Faculty of Arts & Sciences Standing Committee on the Use of Animals in Research & Teaching under Protocol No. 25–08.

ASO design

ASOs were manually designed to target regions of the RNA predicted by *in silico* methods (Hofacker, 2003) to have no substantial secondary structure. For ASO DNA domains (central 10 bases of an ASO) with CpG, 5'Me-dC was used instead of standard deoxy-Cytosines (dCs) to protect the ASO from potential methylation by TLR9. 2–5 ASOs were designed for each target gene. For genes with published, successful MO-induced phenotypes, one ASO was designed to target a region overlapping the MO binding site, with the exception of *oep, MALAT* and *mir-126.*

For the majority of genes, further target regions were chosen within the 5'UTR or very close to the ATG translational start codon.

Microinjection, RNA purification and qPCR

Zebrafish TLAB strain zygotes were collected and injected through the chorion with 25–200 pg of an ASO. ASO injections above 200 pg resulted in general toxicity. Co-injection experiments with *p53* MO included 2.6 ng of *tp53* MO (5'-GCGCCATTGCTTTGCAAGAATTG-3') (Langheinrich et al., 2002). Each batch of ASO-injected embryos was assessed individually for knockdown/loss-of-function phenotypes of the cognate gene by scoring embryonic morphology, performing qPCR, reporter gene expression (*GFP-nanos-3'UTR* for *dnd* ASO) or *in situ* hybridization of marker genes (for *toddler* ASO and *hypocretin* ASO; data not shown). For morphological assessment, embryos were raised to 24–30 hpf and imaged. For qPCR-based assessment of knockdown efficiencies, total RNA was isolated from 5–10 embryos of the appropriate developmental stage using the standard TRIzol (Invitrogen) protocol. Genomic DNA was removed by TURBO-DNase treatment. For reverse transcription (iScript, BioRad), equal amounts of total RNA per sample were used as input (100–500 pg of total RNA, depending on the experiment). 1 μl of a 20 μl cDNA reaction (equivalent to 0.05–0.25 pg of total RNA) was used as template for quantitative real-time PCR (qPCR). qPCR reactions were run on a Stratagene MX3000p using GoTaq (Promega) and 0.25 μM of gene-specific forward and reverse primers (see primer list below). qPCR cycling conditions: 10 min at 95 degrees Celsius, followed by 45 cycles of 30 sec at 95 degrees Celsius, 30 sec at 55 degrees Celsius, and 20 sec at 72 degrees Celsius. qPCR reactions were performed in triplicate and averaged. For each gene, gene expression levels were calculated relative to a reference gene, *ef1a*. Knockdown efficiencies were calculated as the ratio of normalized gene expression in ASO-injected versus uninjected

(or non-cognate ASO-injected) sample. Each experiment was performed at least in duplicate, using independent biological samples.

oep ASO rescue

To generate an *oep* ASO-resistant *oep* mRNA rescue construct, 7 nucleotides within the targeting site of the most efficient *oep*-targeting ASO (ASO#2 sequence: mG* mG* mC* mG* mA* A* C* A* T* G* A* C* A* A* T* mU* mG* mU* mA* mG (* denotes phosphorothioate bonds; 'm' denotes 2'O-Methyl RNA nucleotides)) were mutated by standard PCR-based site-directed mutagenesis. In brief, overlapping fragments encoding the 5' portion and the 3' portion of a fusion between *oep* and *RFP* (*oep signal peptide-RFP-oepORF*) were amplified and fused together by PCR, using standard methods. mRNAs of ASO-sensitive and ASO-resistant *oep-RFP* fusion constructs were synthesized using SP6 mMessage Machine (Ambion) and injected either with or without 100 pg of *oep* ASO#2. Rescue ability was assessed by 1) strength of *oep* mutant phenotype; and 2) persistence of red fluorescence in the presence of *oep* ASO#2.

Imaging

Fluorescently labeled embryos (mRNA injection of *eGFP-nanos3'UTR* (Ciruna et al., 2002) or *RFP-oep* (oep ASO-sensitive or -resistant constructs)) were imaged on the Zeiss Discovery Scope V12, and brightfield images were captured using the Leica MZ16F.

qPCR primers

ef1a was used as reference gene (ef1a_F agaaggaagccgctgagatgg, ef1a_R tccgttcttggagataccagcc). The following primers were used to amplify specific target genes:

alk8_F cgttatcattagcaatgatgtgacg
alk8_R tcttcttttcttctggacttgtgag
bmp2b_F agttttcatcacgaagaggctt
ntl_R agttgtccatgtagttattggtgg
oep_F gaatgacgagtcaactgttcgggttc
oep_R tcttgcagcaggtacggctttgtt

bmp2b_R ttaattctgtggaagccactcg
bmp7a_F agctttgcgaatacagtggatc
bmp7a_R ctgacatggaaggtctcgttttc
chordin_F gttcctctggccggtgttctggt
chordin_R ctcctctggggttcatcttggtgct
MALAT-1_F aaggggatctgcacttttctctctcttct
MALAT-1_R cacacaaacacttccaccacacacc
nacre_RT_F ctcaactgtgagaaagagatggac
nacre_RT_R gttactgatggaaactccagctg
ntl_F aatctggatattcacaactcggtg
smad2_F Aagcg gagcaggaggtggtggag
smad2_R gtccccaaatttcagagcaattgctgg
smad5_F gtagggtgagtttggagagatg
smad5_R gtagggtgagtttggagagatg
tolloid_F aaatggtccccaggcaatatc
tolloid_R agttatactcctgacctggctg
wnt11_F gacctcaagtctaaatacctgtcg
wnt11_R gtcttgttacactgcctgtctg
wnt5b_F cgtcatgcatataggcagcc
wnt5b_R cgaagcggtagccatagttg

Chapter 3: CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing

Preface

This chapter was previously published in Nucleic Acids Research on May 26th 2014 (Montague et al., 2014). José M. Cruz and I conceived the project and generated the first version of CHOPCHOP with input from Eivind Valen. I generated the published version of CHOPCHOP with input from José M. Cruz and substantial input and guidance from Eivind Valen. I wrote the manuscript with input from José M. Cruz, Eivind Valen and James Gagnon. James Gagnon and George Church provided CRISPR/Cas9 advice.

Montague, T.G.*, Cruz, J.M.*, Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, W401–W407.

*equal contribution

Abstract

Major advances in genome editing have recently been made possible with the development of the TALEN and CRISPR/Cas9 methods. The speed and ease of implementing these technologies has led to an explosion of mutant and transgenic organisms. A rate-limiting step in efficiently applying TALEN and CRISPR/Cas9 methods is the selection and design of targeting constructs. We have developed an online tool, CHOPCHOP (https://chopchop.rc.fas.harvard.edu), to expedite the design process. CHOPCHOP accepts a wide range of inputs (gene identifiers, genomic regions or pasted sequences) and provides an array of advanced options for target selection. It uses efficient sequence alignment algorithms to minimize search times, and rigorously predicts off-target binding of single-guide RNAs (sgRNAs) and TALENs. Each query

produces an interactive visualization of the gene with candidate target sites displayed at their genomic positions and color-coded according to quality scores. In addition, for each possible target site, restriction sites and primer candidates are visualized, facilitating a streamlined pipeline of mutant generation and validation. The ease-of-use and speed of CHOPCHOP make it a valuable tool for genome engineering.

Introduction

The discovery of numerous bacterial nucleic acid modification systems has led to the recent development of two modular, precise genome editing tools (Gaj et al., 2013; Mali et al., 2013b). The TALE (transcription activator-like effector) and CRISPR/Cas (clustered regularly interspaced short palindromic repeats) systems have recently been optimized for research use to site-specifically introduce mutations, large translocations and inversions, and manipulate transcriptional activation and repression in a variety of organisms (Bikard et al., 2013; Boch et al., 2009; Cong et al., 2013; Mali et al., 2013c; 2013a; Miller et al., 2011).

TALENs are a genome editing method derived from plant pathogenic bacteria (Gaj et al., 2013). TALE architecture is composed of 3 parts: an N-terminal domain, TALE repeat domains, and a C-terminal domain. The TALE repeat domains typically consist of 34 amino acid residues, where the twelfth and thirteenth repeat variable di-residues (RVDs) determine DNA nucleotide binding specificity (Moscou and Bogdanove, 2009; Streubel et al., 2012). Each RVD is recognized by a specific nucleotide, leading to a simple code for DNA recognition: NI for adenine, HD for cytosine, NG for thymine, and NH or NN for guanine (Cong et al., 2012; Mali et al., 2013c; Moscou and Bogdanove, 2009; Streubel et al., 2012). Importantly, the RVDs can be assembled sequentially

to bind any given target sequence. For genome editing purposes, TALEs are fused to the *FokI* nuclease domain to create TALE nucleases (TALENs). Because *FokI* only cleaves as a dimer, sites must be targeted by a pair of TALENs binding on opposite faces of the DNA strand, spaced approximately 14-20 bp apart. The *FokI* nuclease domains dimerize across the spacer sequence and create a double-strand break (DSB). The DSB can be repaired through error-prone non-homologous end-joining (NHEJ), which often results in indels and potentially frameshift mutations. For efficient binding, TALEN target sequences require a thymine at the 5' end for recognition by the TALE N-terminus (Miller et al., 2011; Moscou and Bogdanove, 2009; Streubel et al., 2012).

The CRISPR/Cas9 system originates from a bacterial immune system that has been adopted for use as a programmable genome editing tool. *S. pyogenes* Cas9 nuclease is directed to target sites in the genome by a single-guide RNA (sgRNA) (Bikard et al., 2013; Cong et al., 2013; Jinek et al., 2012). The Cas9/sgRNA complex binds a 20 bp target sequence followed by a 3 bp protospacer adjacent motif (PAM) –NGG (two invariable Gs preceded by a variable base), and it creates a DSB that is repaired in a seemingly identical manner to TALEN-induced DSBs. While the presence of an -NGG PAM motif is one of the few requirements for binding, the methods used to generate sgRNAs for targeting often impose additional restrictions. Depending on the polymerase being used for sgRNA synthesis, the 5' end dinucleotides may be limited to, for example, 5' GN- for the commonly used U6 promoter, or 5' GG- for T7 polymerase (Bikard et al., 2013; Hwang et al., 2013). In addition, certain criteria such as GC-content appear to influence binding efficiency (Gagnon et al., 2014; Wang et al., 2014). These, along with other guidelines to ensure target suitability, have been used to mostly manually design sgRNAs to generate mutations and knockouts in a variety of organisms including bacteria, yeast, zebrafish, *Xenopus,* nematodes, fruit flies, mice, and human cells (Bikard et al., 2013; Cong et al., 2013; DiCarlo et

al., 2013; Friedland et al., 2013; Gagnon et al., 2014; Gratz et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Nakayama et al., 2013; Wang et al., 2013).

TALEN and sgRNA design requires identification of target sites that fulfill certain sequence requirements while simultaneously avoiding off-targets elsewhere in the genome. Several studies have demonstrated the limited specificity of TALEN- and particularly Cas9-based genome editing strategies, highlighting the importance of determining the uniqueness of each candidate target site (Cradick et al., 2013; Fu et al., 2013; Guilinger et al., 2014; Hsu et al., 2013; Karolchik et al., 2004; Pattanayak et al., 2013). Existing tools for identifying TALEN or sgRNA target sites (Doyle et al., 2012; Gratz et al., 2014; Heigwer et al., 2014; 2013; Hsu et al., 2013; Neff et al., 2013; Sander et al., 2007; Xiao et al., 2014) have limitations, including acceptance of few input formats, slow search times, restriction to either TALEN or CRISPR/Cas9 target design, minimal or no visualization of the target locus and/or limited information about potential off-target sites.

We have developed CHOPCHOP, a web-based tool that allows users to easily and rapidly select the optimal TALEN or CRISPR/Cas9 target sequences in genes from a variety of organisms. To overcome limitations of previous tools, CHOPCHOP accepts a wide range of inputs, employs rigorous off-target search algorithms to predict the specificity of each target site in the genome (35), and displays all options in an interactive graphical interface. In addition, to expedite the validation process, CHOPCHOP designs target site-specific primers for PCR and displays them together with restriction sites in the gene context.

Materials and Methods

Target sequence

CHOPCHOP accepts three forms of input: gene name, genomic coordinates or DNA sequence. If the user provides a gene name, CHOPCHOP converts it to genomic coordinates in the relevant organism by consulting gene tables from a variety of sources (e.g. University of California Santa Cruz (UCSC) Genome Browser (Karolchik et al., 2004)). If the user provides genomic coordinates, for instance to target an intron, these coordinates (or coordinates from the gene table) are parsed by TwoBitToFa (Karolchik et al., 2004), which retrieves the DNA sequence corresponding to the genomic region. If the user provides direct DNA sequence, this sequence (or sequence from TwoBitToFa) is scanned for all potential target sites fulfilling the sequence requirements for the current search (as decided by the user).

Search for off-targets

The candidate target sites are mapped by Bowtie (Langmead et al., 2009) with the appropriate number of mismatches ('-v' mode according to the user-specified options) in a sub-region of the target site where appropriate ('-L' seed mode). In TALEN mode, two target sites are paired if they are within a specified range determined by the user. Each sgRNA or TALEN pair is then ranked according to: (i) the number of off-targets in the genome (TALEN mode considers individual off-targets and paired off-targets), and (ii) how many mismatches lie within the off-targets. In addition, for CRISPR/Cas9 mode, the results are ranked by: (iii) GC-content, and (iv) the presence of a guanine at position 20 in the sgRNA target site (14,15). Any target sites with the same score are then sorted by their position in the gene (with preference to 5′ positions). The specific metrics employed by CHOPCHOP are listed on the site under 'Scoring'. These are updated with new findings from the literature. TALEN results are clustered and suppressed to avoid the display of multiple equivalent TALENs on the results page (e.g. differing by only the size of the spacer sequence). The TALEN pair with the highest ranking in each cluster is displayed on the results page.

Visualization

Interactive visualization is produced by the D3 JavaScript library (Bostock et al., 2011). The targeted gene or locus is displayed in a zoomable interface, with each sgRNA or TALEN pair displayed at its appropriate location. Clicking on any individual sgRNA/TALEN target site results in a detailed view displaying candidate primer pairs ranking the selected target region and restriction sites.

Primer design

Primer pairs spanning the target site are designed by the batch version of Primer3 (Untergasser et al., 2012) using user-specified options. The default parameters are primers of size 18–25 bp (optimum: 22 bp), a product size of 150–290 bp, and a primer *T*m of 57–63°C (optimum: 60°C). The primers are then mapped to the rest of the genome by Bowtie (Langmead et al., 2009) (options '-v 0 –best –k 10'), and subsequently ranked according to their specificity.

Results

CHOPCHOP web tool

CHOPCHOP is an easy-to-use web tool that maximizes user flexibility while maintaining a simple and interactive interface. CHOPCHOP can be run in either CRISPR/Cas9 mode or TALEN mode. It runs with default parameters, but accepts a range of advanced options for more refined searches. CHOPCHOP employs a powerful system for finding off-target sites, and displays the output in an interactive table and within the gene architecture. CHOPCHOP also carries out automated primer design to aid with downstream genotyping steps.

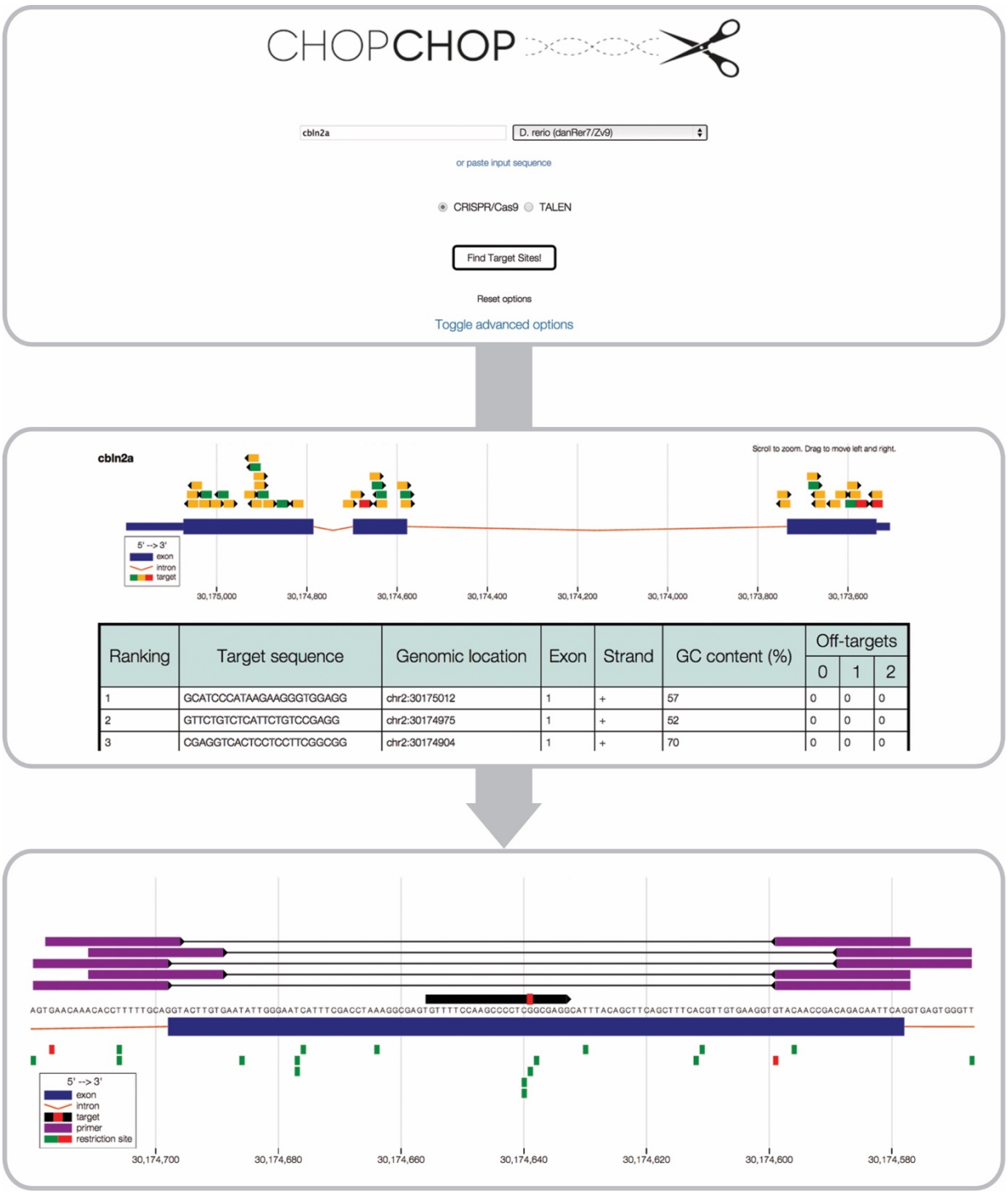


Figure 3-1. Workflow of a CHOPCHOP CRISPR/Cas9 query.
The home page of CHOPCHOP allows users to enter a gene name, genomic coordinates or a DNA sequence, and select an organism and TALEN or CRISPR/Cas9 mode. Advanced options can be toggled. The main results page presents the sgRNA or TALEN target sites within the gene architecture (exon, blue; intron, red), with each option color-coded according to ranking. Hovering over an entry in the table highlights the corresponding graphical sgRNA/TALEN and vice versa. Clicking on a specific result takes the user to a page containing the zoomed in locus with the predicted cut site highlighted in red, primer options in purple and restriction sites color-coded according to whether they are unique in the region.

Implementation

Input. CHOPCHOP can be run with as few as three basic input options, or with additional advanced parameters. The basic input comprises: (i) a gene name (accepting RefSeq, ENSEMBL, FlyBase and WormBase gene IDs), genomic coordinates or a pasted sequence; (ii) a growing list of organisms (*Homo sapiens*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Xenopus tropicalis, Rattus norvegicus, Gallus gallus, Oryzias latipes*, *Gasterosteus aculeatus* and *Anopheles gambiae*); and (iii) the choice between CRISPR/Cas9 or TALEN mode (Figure 3-1). The advanced options allow the user to target a sub-region of the gene, such as the 5′ UTR, 3′ UTR, splice sites, full exons (including UTRs), or a specified subset of exons.

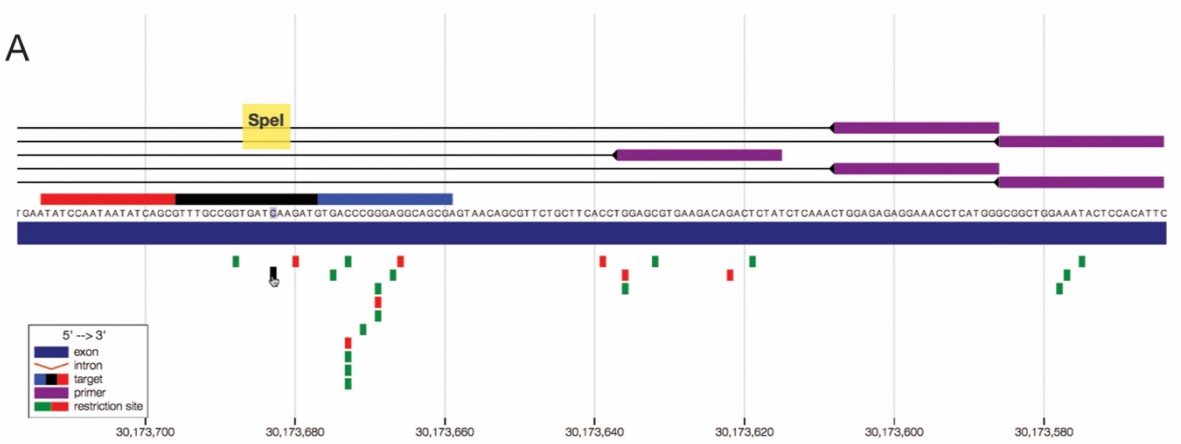
The CRISPR/Cas9 search mode locates 23 bp target sites including the PAM motif. The user may restrict this search to only target sites suited for synthesis using a particular polymerase, e.g. GG- or GN-/NG- at the 5′ end of the sgRNA. Recent reports have shown that Cas9 can have substantial off-target activity in the genome (Cradick et al., 2013; Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013) and that tolerance to mismatches shows significant variance depending on the position within the sgRNA (Cong et al., 2013; Hsu et al., 2013). Another study demonstrated that the number of mismatches tolerated is dependent upon the specific sgRNA (Fu et al., 2013), suggesting there is no universal rule for CRISPR/Cas9 sgRNA off-target prediction. For this reason, CHOPCHOP offers the choice between a variety of published methods for off-target prediction. (i) One study found mismatches were tolerated at any position except within the PAM motif (Hsu et al., 2013). CHOPCHOP provides a search mode reflecting this rule, searching for mismatches across all bases upstream of the PAM. This is the default mode. (ii) An alternative study found that single-base mismatches up to 11 bp 5′ of the PAM completely abolished cleavage by Cas9 (Cong et al., 2013). In contrast, mutations further

upstream of the PAM retained cleavage activity. CHOPCHOP therefore provides an alternative search mode that locates off-targets with mismatches only in the region where a mismatch would still induce cleavage. (iii) Finally, CHOPCHOP provides a fast mode that only searches for perfect matches of the sgRNA target sequence across the genome.

The TALEN search mode locates pairs of target sites on opposite strands, separated by a 14–20 bp spacer sequence, with the requirement that both sites have a T at the 5′ end. The TALEN-specific options allow the user to cater the target search to a particular TALEN architecture by changing the length of the spacer sequence and the length of the target sites. In addition, depending upon the assembly kit being used, the user can choose to use either the RVD 'NN' for guanines, or 'NH', which has been shown to bind guanines more specifically than 'NN' (Cong et al., 2012; Streubel et al., 2012). TALEN off-target binding does not appear to have the same position-specific complexity as CRISPR/Cas9 sgRNAs, therefore the TALEN off-target method searches for off-targets with 0, 1 or 2 mismatches across each site. The default method searches for two mismatches.

In order to analyze whether Cas9 or TALENs have successfully cleaved the target locus, users may need to amplify the region of interest for further analysis by methods such as deep sequencing or a T7E1 assay (Kim et al., 2009). CHOPCHOP therefore integrates primer design with sgRNA/TALEN target site design using Primer3 (Untergasser et al., 2012). Primers are designed to amplify the region surrounding the cut site, and mapped against the genome to avoid off-targets producing amplicons of similar length. In the advanced options, the user can adjust the primer specifications, including amplicon size, primer *T*m, primer length and the minimum distance between each primer and the target site. In addition, some users might prefer to assess successful mutagenesis using restriction enzyme digestion. CHOPCHOP

A



B

TALE sequences		
TALE 1 TGCAGAAAATGCCACTTTAGCACCAGCAGCTCGGACGGATATCCCAAACGACGA TALE 2 Generate plasmid sequences here		
There are no predicted off-target TALE pairs.		
TALE 1 off-targets		
Genomic location	Number of mismatches	Sequence (including mismatches)
>100 across the genome	0-2	n/a
TALE 2 off-targets		
Genomic location	Number of mismatches	Sequence (including mismatches)
chr23:28429363	2	GGATATaCCAAAtGACGA
chr4:17743523	2	TCGTtGTTTGGGATATCg

C

Ranking	Left primer coordinates	Left primer	Left primer Tm	Left primer off-targets	Right primer coordinates	Right primer	Right primer Tm	Right primer off-targets	Pair off-targets	Product size
1	chr2:30174775-30174797	CGTCCCAGACTGGACTAGTTTC	60.2	1	chr2:30174968-30174990	TGATGCCCAAGACAGAGTAAGA	59.9	1	0	215
2	chr2:30174775-30174797	CGTCCCAGACTGGACTAGTTTC	60.2	1	chr2:30175013-30175035	GTAGGGTATTCTTCCCACCTCC	60.1	1	0	260
3	chr2:30174717-30174739	AGTGAATCAATTGCACCTCTGA	59.7	1	chr2:30174968-30174990	TGATGCCCAAGACAGAGTAAGA	59.9	1	0	273
4	chr2:30174701-30174723	CGTTTTTCCACAAACAAGTGAA	60.0	1	chr2:30174965-30174987	TTTTGATGCCCAAGACAGAGTA	59.7	1	0	286
5	chr2:30174701-30174723	CGTTTTTCCACAAACAAGTGAA	60.0	1	chr2:30174952-30174974	TGTAAAGGGAAGGTTTTGATGC	60.3	1	0	273

Figure 3-2. CHOPCHOP provides detailed information about each CRISPR/Cas9 and TALEN target site.

(**A**) The detailed information page provides a zoomed in view of the target locus with visible DNA sequence, primer options (above the gene, purple) and restriction sites (below the gene; green if unique in the region, red if not). In TALEN mode the target site is color-coded; TALEN 1 is blue, the spacer is black and TALEN 2 is red. (**B**) Information is provided about predicted off-targets: the genomic location, number of mismatches and location of mismatches within the sequence. (**C**) Information is provided about the primer designs, including the primer sequence, *T*m and product size.

allows the user to select restriction sites from a particular restriction enzyme company, and it allows the user to specify the minimum size of the restriction site.

Output. The majority of CHOPCHOP queries are executed within a matter of seconds, and the results displayed in an interactive table and interactive gene model. CHOPCHOP ranks the search results according to a number of criteria. Both the TALEN and CRISPR/Cas9 modes are ranked by: (i) the number of off-targets, (ii) whether the off-targets contain mismatches or are perfect hits, and (iii) where the target site lies within the architecture of the gene (many users wish to create a frameshift/null mutation and therefore will prefer a mutation at the 5′ end of the gene). Additionally, for CRISPR/Cas9 mode the results are ranked by (iv) GC-content. Recent reports suggest that sgRNAs are most effective with a GC-content between 45 and 80%, and (v) a guanine at position 20 in the target site, which is associated with improved activity (Gagnon et al., 2014; Wang et al., 2014). For TALEN mode, off-targets are specifically scored by whether an individual TALEN target site occurs elsewhere in the genome, or whether both members of a pair lie within cutting distance of one another at an off-target location. For both CRISPR/Cas9 and TALEN mode, the results table provides the sequence of the target site, its ranking, genomic location (including exons and orientation) and the number of potential off-targets with 0, 1 or 2 mismatches (Figure 3-1). The CRISPR/Cas9 mode also provides the GC-content of the sgRNA target site, and the TALEN mode provides restriction sites that lie in the spacer between two TALENs, as well as the RVDs that should be synthesized for the target site. CHOPCHOP also provides an interactive graphical representation of the gene, with each sgRNA or TALEN target site color-coded according to ranking (Figure 3-1). This allows users to inspect candidate targets by their location within the gene as well as their specificity within the genome as a whole. The graphical output is generated using the D3 JavaScript library (Bostock et al., 2011) and enables the user to zoom and scroll across the gene. Finally, users can download a text file containing

the search results, as well as a GenBank file of the DNA sequence annotated with the target sites, either with or without introns.

Individual target sites can be inspected in a separate detailed view, displaying additional information about the genomic location of off-targets, and the location of the mismatches within the off-target sites (Figure 3-2). Upon zooming within this specific region, the DNA sequence becomes visible. For TALENs, the gene view suppresses the visibility of substantially overlapping TALEN pairs to avoid redundancy. In the detailed view, however, all of the clustered targets are listed, should the user prefer a different target sequence in the same approximate location. The detailed view also presents the user with all the restriction sites in the surrounding region that can be used for testing cleavage activity. Restriction sites are color-coded according to whether they are unique within the region. Finally, the detailed view displays primer pairs that ank the target site, and a downloadable GenBank le of the targeted region is available, containing annotations of the target site and primer designs.

Conclusion

CHOPCHOP is a user-friendly web tool that locates optimal CRISPR/Cas9 and TALEN target sites for any genomic region, and presents the information in an interactive and intuitive manner. CHOPCHOP expedites the design process for CRISPR/Cas9- or TALEN-based mutations with fast run times, powerful off-target prediction and integrated primer design.

CHOPCHOP has a number of features that separate it from the other CRISPR/Cas9 or TALEN tools currently available (Doyle et al., 2012; Gratz et al., 2014; Heigwer et al., 2013; 2014; Hsu et

al., 2013; Neff et al., 2013; Sander et al., 2007; Xiao et al., 2014). First, CHOPCHOP accepts a wide range of inputs – gene identifiers, genomic regions or pasted sequences – making it suitable for a broad range of uses. Second, CHOPCHOP provides a dynamic graphical output display that includes an interactive visualization of the gene, with each Cas9/TALEN target site displayed at its genomic position and color-coded according to its quality. The visualization of all possible target sites in the gene model makes the selection of the optimal candidate easy, and is an ideal system for the design of two sgRNAs, as used in the increasingly popular dual nickase approach (Ran et al., 2013). Third, unlike most tools, CHOPCHOP integrates TALEN and CRISPR/Cas9 target design into a single tool. Fourth, CHOPCHOP provides automatic primer generation and restriction site visualization for genotyping. Finally, CHOPCHOP provides downloadable results, including a GenBank file with annotations of the gene's exons, introns and target sites, and a GenBank file of the specific target region with primer designs. CHOPCHOP creates a streamlined process from start to end of mutant design, and is a valuable new resource for genome editing technologies.

Chapter 4: CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering

Preface

This chapter was previously published in Nucleic Acids Research on May 16 2016 (Labun et al., 2016). Kornel Labun added new features to CHOPCHOP with supervision from Eivind Valen. I maintained CHOPCHOP and added new organisms. James A. Gagnon provided CRISPR/Cas9 advice. Summer B. Thyme provided unpublished CRISPR/Cas9 results. I wrote the manuscript with input from Eivind Valen and Kornel Labun.

Labun, K.*, Montague, T.G.*, Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. *44*, W272–W276.

*equal contribution

Abstract

In just 3 years CRISPR genome editing has transformed biology, and its popularity and potency continue to grow. New CRISPR effectors and rules for locating optimum targets continue to be reported, highlighting the need for computational CRISPR targeting tools to compile these rules and facilitate target selection and design. CHOPCHOP is one of the most widely used web tools for CRISPR- and TALEN-based genome editing. Its overarching principle is to provide an intuitive and powerful tool that can serve both novice and experienced users. In this major update we introduce tools for the next generation of CRISPR advances, including Cpf1 and Cas9 nickases. We support a number of new features that improve the targeting power, usability and efficiency of CHOPCHOP. To increase targeting range and specificity we provide support for custom length

sgRNAs, and we evaluate the sequence composition of the whole sgRNA and its surrounding region using models compiled from multiple large-scale studies. These and other new features, coupled with an updated interface for increased usability and support for a continually growing list of organisms, maintain CHOPCHOP as one of the leading tools for CRISPR genome editing. CHOPCHOP v2 can be found at http://chopchop.cbu.uib.no

Introduction

The discovery and adoption of the CRISPR bacterial system for genome editing has led to a revolution in biology: targeted mutations are now possible in a multitude of organisms, including many not previously amenable to genetic manipulation. This has both transformed our approach to answering biological questions and unlocked the possibility of correcting human genetic diseases.

Originally harnessed from the *Streptococcus pyogenes* type II system (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013c), CRISPR genome editing is based on a two-component system: a Cas9 nuclease and a single guide RNA (sgRNA), which directs the nuclease to a specific site in the genome. In the presence of the sgRNA, Cas9 locates the target site and makes a double-strand break (DSB). The DSB is repaired by the host non-homologous end-joining pathway, but often the repair is imperfect, creating indels and in many cases frameshift mutations. Since the technology's inception, research to improve the technology has focused on two main challenges: optimization of cutting efficiency and specificity of cutting. A substantial portion of sgRNAs designed for a given gene will produce a low or zero cutting rate, and many sgRNAs have the capacity to bind promiscuously in the genome, which can lead to off-target mutagenesis

(Cradick et al., 2013; Fu et al., 2013; Hsu et al., 2013; Kim et al., 2015; Pattanayak et al., 2013; Tsai et al., 2015; Wang et al., 2015b). To address these issues, research has focused on identifying the sequence features that contribute to effective (and ineffective) sgRNAs (Chari et al., 2015; Doench et al., 2016; 2014; Gagnon et al., 2014; Moreno-Mateos et al., 2015; Xu et al., 2015), as well as the development of new CRISPR variants that expand the targeting range and specificity of the nuclease (Fu et al., 2014; Kleinstiver et al., 2016; 2015; Zetsche et al., 2015). With the contribution of so many factors to optimum sgRNA target selection, it has become necessary to use software to aid selection of CRISPR target sites for experiments. CHOPCHOP (Montague et al., 2014) provides an intuitive online environment for target selection that optimizes efficiency and specificity according to the latest large-scale studies, as well as performing primer design and restriction site identification, all in a user-friendly, graphical interface (Figure 4-1). This new update of CHOPCHOP provides additional flexibility by offering new options for sgRNA design, as well as additional metrics by which sgRNA targets are scored and ranked.

Improvements in the 2016 release

CHOPCHOP accepts multiple input formats (gene identifiers, genomic coordinates and pasted sequences) for a wide range of organisms, and provides instant, visual output as well as downloadable data (GenBank, text tables and FASTA les). In this new version users can also view the output data in the UCSC browser (Kent et al., 2002) with a single click, enabling results to be viewed in the context of annotated genomic features, such as transcription factor binding sites and chromatin architecture and accessibility (Figure 4-2).

CHOPCHOP offers flexible targeting to sub-regions of protein-coding and non-coding genes, including coding regions, UTRs, splice sites and individual exons. In this new version we also offer a promoter-targeting mode (Figure 4-2) for experiments such as down- or upregulating gene expression using catalytically dead Cas9 (dCas9) or transcriptionally active dCas9 (e.g. dCas9-VP64), respectively (Bikard et al., 2013; Hilton et al., 2015; Qi et al., 2013). CHOPCHOP determines potential off-target sites for all sgRNAs using Bowtie (Langmead et al., 2009) and automatically generates primers for target sites using Primer3 (Untergasser et al., 2012). The length and annealing temperature of the primers, as well as the size of the amplicon, can be specified. CHOPCHOP visualizes all elements in a dynamic visual interface that includes information about restriction sites, which can be used for downstream validation. In addition to these improvements, the new iteration of CHOPCHOP introduces the following major new features.

Support for a new generation of CRISPR effectors

The most widely used CRISPR effector is Cas9, derived from the type II *S. pyogenes* system. While the RNA-mediated targeting of Cas9 offers great versatility in selecting a target site, a limiting factor is the requirement for an NGG protospacer adjacent motif (PAM) motif adjacent to the target. The occurrence of this motif is not rare in most genomes, but it imposes a restriction that can be inimical to achieving the high genomic precision required for certain experiments, or for targeting small genes. The new generation of CRISPR effectors vastly expands the universe of viable targets by offering alternative PAM motifs. CHOPCHOP now provides support for alternative CRISPR effectors, including Cpf1 from *Acidaminococcus*, which utilizes an AT-rich PAM (Zetsche et al., 2015) and Cas9 homologs from *S. pyogenes, Streptococcus thermophilus*, *Staphylococcus aureus* and *Neisseria meningitidis* (Fonfara et al., 2014). In addition, CHOPCHOP also accepts user-defined custom PAMs that can be anchored to the 5′ (Cpf1) or 3′ (Cas9) end

of the sgRNA. This field accepts the standard IUPAC nucleotide alphabet (Cornish-Bowden, 1985), including ambiguity codes. CHOPCHOP therefore provides support for the sequence requirements of any currently known CRISPR effector and enables immediate adoption of any new CRISPR effectors. This greatly increases the targeting range of CRISPR experiments that can be designed with CHOPCHOP, including improved targeting of AT-rich genomes such as *Plasmodium falciparum*.

New rules for optimizing cutting efficiency

CRISPR sgRNAs can be ranked by 2 criteria: (i) efficiency – the likelihood that the particular sgRNA facilitates cutting, and (ii) specificity – the likelihood that the sgRNA binds off-target sites.

The initial release of CHOPCHOP provided two simple metrics for efficiency based on experimental studies. First, the GC-content of the sgRNA – ideally between 40 and 80% – and second, whether the sgRNA contains a G at position 20 (Gagnon et al., 2014; Wang et al., 2014). Since the initial release of CHOPCHOP, several refinements have been proposed. A study from Doench *et al*. produced a large dataset to calculate efficiencies across a wide range of sgRNAs (Doench et al., 2014), and the rules for computationally-aided sgRNA design were recently further refined by the same group (Doench et al., 2016). Moreno-Mateos *et al*. conducted similar screens and found that sgRNA stability, which depends on guanine enrichment and adenine depletion, was a major determinant of sgRNA efficiency (12). Chari *et al*. conducted a study exploiting the bias of lentiviral integration into transcriptionally active regions, which: (i) revealed that accessible DNA is more amenable to cutting with Cas9; (ii) separated the influence of DNA accessibility and sequence composition on sgRNA efficiency (Chari et al., 2015). CHOPCHOP users can now view results in the UCSC browser (Kent et al., 2002) in the context of DNase I hypersensitivity sites to predict accessible DNA regions (Figure 4-2). Finally, a meta study by Xu *et al*. compiled the

sequence specificities across multiple datasets to build an aggregate model (Xu et al., 2015). We have implemented all of these metrics in the new release to give the user a broad selection of metrics to choose from (the default is the Xu *et al*. metric). Using these methods, CHOPCHOP can now score every sgRNA using position-specific scoring matrices or support vector machines that consider each individual position of the sgRNA as well as the sequence downstream of the PAM and upstream of the binding site. In the results table this score is reported as the 'efficiency score'.

Other factors also play a role in whether an sgRNA is likely to cut at its intended target. Recently, we and others showed that self-complementarity of the sgRNA can inhibit its efficient incorporation into the effector complex (Moreno-Mateos et al., 2015; Thyme et al., 2016). CHOPCHOP now includes the basic self-complementarity score of the Thyme *et al*. study (Thyme et al., 2016), which computes the number of potential 4 bp stems within the sgRNA and between the sgRNA and the backbone. The user can therefore opt to avoid sgRNAs with self-complementarity using this option.

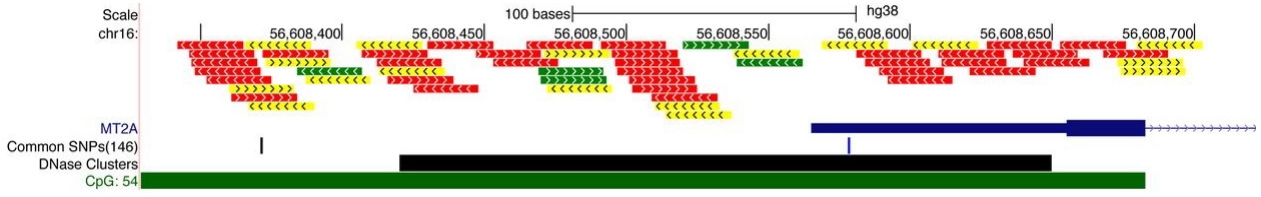


Figure 4-1. CHOPCHOP results can be exported to the UCSC browser with a single click.
Here, the sgRNAs (in this example in promoter-targeting mode) are viewed in the context of the genome. The tracks displayed in this example are DNase sensitive regions, common SNPs and CpG islands.

Strategies to increase specificity

A significant challenge in CRISPR experiments is the possibility of inducing cleavage at sites other than the intended target. An emerging tool to alleviate this problem is the paired nickase approach (Ran et al., 2013). Unlike natural CRISPR effectors, nickases have been modified to cut only one DNA strand. In order to create a DSB, a pair of nickases must be targeted to opposite strands and bind within 10–31 bp of each other (Ran et al., 2013). These requirements vastly reduce the likelihood of creating off-target DSBs, and CHOPCHOP has now added support for paired nickase experiments. In this mode, sites on opposite strands within a specified distance (either default or user-defined) are paired as potential nickase sites. For these sites, in addition to the default off-target search, each pair of sites is evaluated for off-targets where binding and cutting would result in a DSB. Nickase sites are visualized with two CRISPR targets surrounding a 'break' region (Figure 4-1).

Recent studies have highlighted the need to search for more than two mismatches when identifying off-targets (Tsai et al., 2015) so CHOPCHOP now counts off-targets with up to three mismatches. While off-targets with more than three mismatches have been reported (Tsai et al., 2015), evidence suggests that almost all predicted sites of four mismatches or more are not cleaved (Tsai et al., 2015) and therefore the vast majority of such predicted sites would be misleading and unnecessarily time-consuming to search for during sgRNA selection. Another strategy that has been shown to decrease off-target cleavage is the use of truncated sgRNAs (Fu et al., 2014; Tsai et al., 2015). Besides increasing specificity, 5′ shortening of the customary 20 bp also increases the targeting range. The new version of CHOPCHOP therefore provides support for sgRNAs of user-defined lengths.

Thus, this version of CHOPCHOP supports a number of new features that: (i) improve the ability to target a broader range of sequences, and (ii) more thoroughly predict potential off-target sites in the genome.

New genomes

In addition to a new range of features, CHOPCHOP strives to accommodate all requests for new genomes and gene annotation sets. So far we have incorporated all inquiries received, and CHOPCHOP now supports a total of 32 organisms. Furthermore, all genomes have been updated to their most recent assemblies and suggestions for new species can easily be submitted through a link on the main page.

Discussion and future developments

The overarching principle of CHOPCHOP is to provide an intuitive and powerful tool that can serve first time as well as experienced users. The basic mode offers optimized defaults for the basic user, while more advanced users can select from a wide range of options curated from the literature by their relevance and utility. All options are presented in a tabulated and organized manner to help users quickly visualize and evaluate options when designing CRISPR experiments.

This release retains the general layout of the previous release but updates the visual profile to a modern look and to accommodate new features. The site is now mobile and tablet friendly, and to streamline the user's experience we use cookies to remember the selection of species and targeting options for subsequent searches. All reported bugs have been fixed, and the

implementation is now optimized for future development to facilitate both rapid adoption of any future effectors and new targeting data from large-scale studies. This major update maintains CHOPCHOP as one of the most easy-to-use, versatile and powerful CRISPR targeting tools available.

Chapter 5: Vg1/Nodal heterodimers are the endogenous inducers of mesendoderm

Preface

This chapter was previously published in eLIFE on November 15 2017. Alexander F. Schier and I conceived the experiments, analyzed the data and wrote the manuscript. I performed the experiments. Nate Lord performed the modeling.

Montague, T.G., and Schier, A.F. (2017). Vg1-Nodal heterodimers are the endogenous inducers of mesendoderm. Elife *6*, 178.

Abstract

Nodal is considered the key inducer of mesendoderm in vertebrate embryos and embryonic stem cells. Other TGF-beta-related signals, such as Vg1/Dvr1/Gdf3, have also been implicated in this process but their roles have been unclear or controversial. Here we report that zebrafish embryos without maternally provided *vg1* fail to form endoderm and head and trunk mesoderm, and closely resemble *nodal* loss-of-function mutants. Although Nodal is processed and secreted without Vg1, it requires Vg1 for its endogenous activity. Conversely, Vg1 is unprocessed and resides in the endoplasmic reticulum without Nodal, and is only secreted, processed and active in the presence of Nodal. Co-expression of Nodal and Vg1 results in heterodimer formation and mesendoderm induction. Thus, mesendoderm induction relies on the combination of two TGF-beta-related signals: maternal and ubiquitous Vg1, and zygotic and localized Nodal. Modeling reveals that the pool of maternal Vg1 enables rapid signaling at low concentrations of zygotic Nodal.

Introduction

The induction of mesoderm and endoderm (mesendoderm) during embryogenesis and embryonic stem cell differentiation generates the precursors of the heart, liver, gut, pancreas, kidney and other internal organs. Nodal, a ligand in the TGF-beta protein family, is the key inducer of vertebrate mesendoderm (Schier and Shen, 2000; Schier, 2009; Shen, 2007), ranging from zebrafish and mouse embryos to human embryonic stem cells. Nodal mutants fail to form mesendodermal cell lineages in zebrafish and mouse (Conlon et al., 1991; 1994; Feldman et al., 1998; Zhou et al., 1993), and activation of the Nodal signaling pathway drives the *in vitro* differentiation of embryonic stem cells into mesendodermal progenitors (Brandenberger et al., 2004; Camus et al., 2006; D'Amour et al., 2005; Hoveizi et al., 2014; Kubo, 2004; Parisi et al., 2003; Schier and Shen, 2000; Shen, 2007; Smith et al., 2008; Takenaga et al., 2007; Vallier et al., 2004; Yasunaga et al., 2005). Following its role in mesendoderm induction, Nodal activity also patterns the left-right axis. Nodal ligands are expressed in the left lateral plate mesoderm (Collignon et al., 1996; Levin et al., 1995; Long et al., 2003; Lowe et al., 1996; Pagán-Westphal and Tabin, 1998), and mutants that lack left-sided Nodal signaling exhibit multiple left-right defects (Brennan et al., 2002; Kumar et al., 2008; Long et al., 2003; Noël et al., 2013; Saijoh et al., 2003; Yan et al., 1999).

Nodal is not the only TGF-beta-related signal implicated in mesendoderm induction and left-right patterning. Members of the Vg1/GDF1/GDF3 TGF-beta subfamily have been assigned various roles in these processes, although there are puzzling contradictions from the level of gene expression to the loss-of-function and gain-of-function phenotypes. The role of GDF1 in left-right patterning is well established. *Gdf1* mutant mice exhibit left-right asymmetry defects (Rankin et al., 2000) and morpholino studies indicate that zebrafish *vg1* (*dvr1*/*gdf3*) is required

for left-right patterning (Peterson et al., 2013). GDF1/Vg1 alone is unable to activate the Nodal signaling pathway, but it increases the activity and range of mouse and zebrafish Nodal ligands in *Xenopus* assays (Peterson et al., 2013; Tanaka et al., 2007) and the activity of mouse Nodal in tissue culture cells (Andersson et al., 2007; Fuerer et al., 2014). Thus, Nodal and Vg1/GDF1 family members cooperate to pattern the left-right axis.

The role of the Vg1/GDF1/GDF3 TGF-beta subfamily in mesendoderm formation is less clear. In *Xenopus* – where Vg1 was first discovered – *vg1* mRNA is localized to a vegetal crescent in the oocyte and in the vegetal hemisphere of the early embryo (Rebagliati et al., 1985; Weeks and Melton, 1987). By contrast, zebrafish *vg1* mRNA is localized to the animal pole of late stage oocytes (Marlow and Mullins, 2008), and it is present ubiquitously in the early embryo (Dohrmann et al., 1996; Helde and Grunwald, 1993; Peterson et al., 2013). *Gdf1* and *Gdf3*, which are considered to be the mammalian Vg1 orthologs (Andersson et al., 2007; Chen et al., 2006; Rankin et al., 2000; Wall et al., 2000), are expressed in the 16-cell morula (*Gdf3*) and epiblast prior to gastrulation (*Gdf3 and Gdf1*) (Chen et al., 2006; Wall et al., 2000). Some *Gdf3* mutants lack a subset of endodermal and mesodermal markers, while others grow to fertile adults (Andersson et al., 2007; Chen et al., 2006), conversely *Gdf1* mutants only exhibit left-right asymmetry defects (Rankin et al., 2000). Some *Gdf1;Gdf3* double mutants exhibit more severe defects in endoderm and mesoderm formation than *Gdf3* single mutants (Andersson et al., 2007) and *Gdf1-/-;Nodal+/-* mutants resemble hypomorphic *nodal* mutants (Andersson et al., 2006; Lowe et al., 2001), suggesting some synergy between GDF1 and Nodal functions. Experiments in the chick and mouse indicate that Vg1/GDF1/GDF3 may act upstream of Nodal (Andersson et al., 2007; Chen et al., 2006; Rankin et al., 2000; Shah et al., 1997; Skromne and Stern, 2001; Tanaka et al., 2007). Thus, these analyses suggest that mouse Nodal, GDF1 and GDF3 may cooperate during early amniote development, but their regulatory and molecular relationships have remained

unclear. Morphant studies in zebrafish suggest a function for Vg1 in left-right axis formation but not in mesendoderm induction (Peterson et al., 2013). Antisense oligonucleotide-mediated knockdown of *Xenopus vg1* leads to defects in dorsal mesoderm induction (Birsoy et al., 2006), but most mesendodermal derivatives still form. Taken together, loss-of-function studies establish crucial roles for both Nodal and Vg1/GDF1 in left-right development and for Nodal in mesendoderm induction, but the roles of GDF1/GDF3/Vg1 in mesendoderm induction remain poorly understood.

Another puzzling aspect of Vg1's function is its apparent inability to be processed and secreted. This is in stark contrast to other members of the TGF-beta superfamily, which are generated as pro-proteins that dimerize and are cleaved to generate a secreted, mature dimer that binds receptors (Constam, 2014; Dutko and Mullins, 2011). Neither cleavage nor secretion of Vg1 have been detected in *Xenopus* and zebrafish, and correspondingly, overexpression does not yield a phenotype (Dale et al., 1989; 1993; Dohrmann et al., 1996; Tannahill and Melton, 1989; Thomsen and Melton, 1993). Conflicting results have been reported for GDF1 processing in heterologous systems, ranging from cleavage but inactivity in *Xenopus* oocytes (Tanaka et al., 2007) to no detectable cleavage in *Xenopus* embryos (Wall et al., 2000). Mouse Nodal-GDF1 heterodimers, but not zebrafish Nodal-Vg1 heterodimers, have been detected in a heterologous *Xenopus* system (Peterson et al., 2013; Tanaka et al., 2007). Upon fusion of the Vg1, GDF1 or GDF3 mature domain to the Activin or BMP prodomain, Vg1 is processed and induces mesoderm formation (Chen et al., 2006; Dale et al., 1993; Dohrmann et al., 1996; Kessler and Melton, 1995; Thomsen and Melton, 1993; Wall et al., 2000). However, it is unclear if these constructs reveal the true nature of Vg1, or whether the fused prodomains generate ectopic functions. Thus, it remains to be resolved how Vg1 processing, secretion, dimerization and activity are regulated.

In this study we address the long-standing question of Vg1's role in vertebrate mesendoderm induction and its relationship to Nodal, using zebrafish as a model system. Current models of zebrafish mesendoderm induction have focused entirely on the roles of the two zebrafish Nodal genes, *cyclops* (*cyc*) and *squint* (*sqt*), with no consideration of *vg1* (Bodenstine et al., 2016; Cartwright et al., 2008; Chea et al., 2005; Constam, 2009; Hirokawa et al., 2006; Juan and Hamada, 2001; Liang and Rubinstein, 2003; Papanayotou and Collignon, 2014; Pauklin and Vallier, 2015; Quail et al., 2013; Robertson, 2014; Schier and Shen, 2000; Shen, 2007; Signore et al., 2016; Strizzi et al., 2012; 2009; Tian and Meng, 2006; Wang and Tsang, 2007; Whitman, 2001). *cyc* and *sqt* are zygotically-expressed at the embryonic margin and act as concentration-dependent inducers of mesendoderm (Schier, 2009). *cyc;sqt* double mutants (Feldman et al., 1998) and other zebrafish Nodal signaling mutants (Dubrulle et al., 2015; Gritsman et al., 1999) fail to form endoderm and head and trunk mesoderm. Conversely, ectopic expression of *cyc* or *sqt* induces mesendoderm formation (Bisgrove et al., 1999; Feldman et al., 1998; Gritsman et al., 1999; 2000; Meno et al., 1999; Sampath et al., 1998). These results, and the lack of a *vg1* morphant mesendoderm phenotype (Peterson et al., 2013), have been interpreted to mean that Cyc and Sqt are the sole inducers of mesendoderm, without a requirement for Vg1 or other TGF-beta family members. Contrary to these models, we now report that *vg1* is absolutely essential for mesendoderm induction. Vg1 is only secreted, processed and active in the presence of Nodal, while Nodal requires Vg1 for activity. Co-expression of Nodal and Vg1 results in heterodimer formation and mesendoderm induction.

Results

Maternal *vg1* is required for mesendoderm formation

To determine the function of zebrafish Vg1, we generated *vg1* mutants using CRISPR/Cas9 (Appendix Figure 3A). We recovered 8 bp and 29 bp deletion alleles that cause frameshifts, truncating Vg1 from a 355 amino acid protein to predicted 18 and 11 amino acid peptides, respectively (Appendix Figure 3B). Zygotic homozygous *vg1* (Z*vg1)* mutants were viable, with no strong left-right asymmetry defects (Appendix Figure 4), allowing the generation of maternal *vg1* (M*vg1*) mutants from homozygous females crossed to wild-type males (Figure 5-1A). M*vg1* embryos lacked the derivatives of the mesendoderm, including heart, blood, pronephros, notochord, gut and trunk somites (Figure 5-1A). To test whether the phenotype is caused by the loss of *vg1*, we performed rescue experiments by injecting 5 concentrations of *vg1* mRNA, spanning a 1600-fold range. 0.5-100 pg of *vg1* rescued the phenotype, revealing that the embryo can tolerate a large range of *vg1* concentrations (Figure 5-1B). 50pg of a *vg1* mRNA containing the 8 bp deletion found in the genetic mutant was unable to rescue the phenotype (Figure 5-1C). In contrast to previous *vg1* morpholino experiments (Peterson et al., 2013), these results reveal that *vg1* is essential for mesendoderm formation.

Endogenous Nodal signaling requires Vg1

The phenotype of M*vg1* and maternal-zygotic *vg1* (MZ*vg1)* embryos closely resembles that of embryos that lack Nodal (Feldman et al., 1998), its coreceptor Oep (Gritsman et al., 1999), or its signal transducer Smad2 (Dubrulle et al., 2015) (Figure 5-2A). To determine whether M*vg1* embryos are defective in Nodal signaling, we analyzed the expression of a selection of Nodal target genes. The expression of these mesendoderm genes showed the same defects in M*vg1* mutants as in Nodal signaling mutants, indicating that Nodal signaling is not functional in the absence of Vg1 (Figure 5-2B).

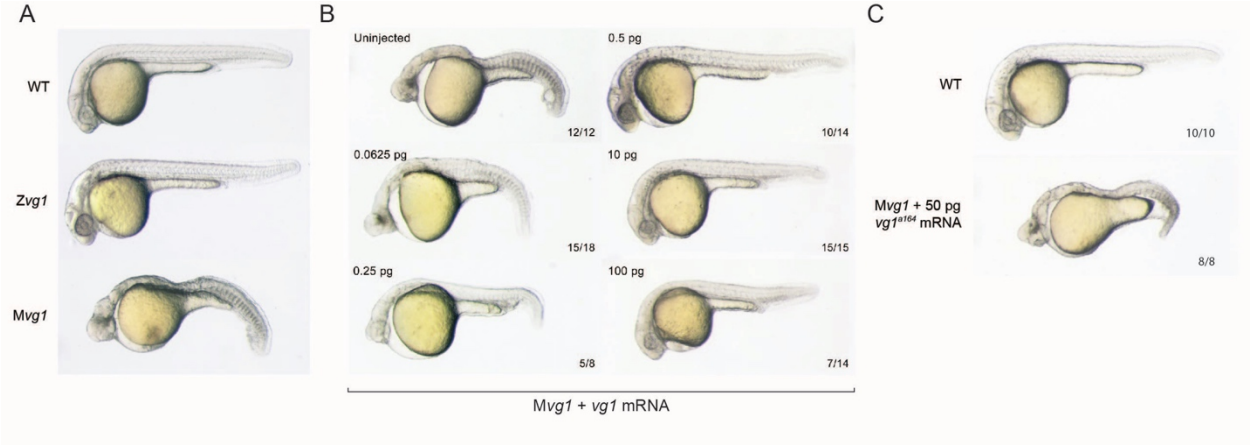


Figure 5-1. Maternal *vg1* is required for mesendoderm formation.
(A) Zygotic and maternal *vg1* (Z*vg1* and M*vg1*) mutants and wild-type (WT) embryo at 28 hours post-fertlization (hpf). See Appendix Figure 3 for information about the *vg1* mutant alleles, and Appendix Figure 4 for analysis of left-right asymmetry in WT and Z*vg1* embryos.
(B) M*vg1* embryos injected with 0.0625-100 pg of *vg1* mRNA.
(C) M*vg1* embryo injected with 50 pg of *vg1* mRNA containing the 8 bp deletion allele found in the genetic mutants (*vg1*^{sa164}).

One way Nodal signaling might be disrupted in M*vg1* embryos is through loss of Nodal gene expression. We analyzed *cyc* and *sqt* expression in wild-type and M*vg1* embryos. *cyc* and *sqt* were initially expressed at comparable levels across both genotypes, but mRNA levels subsequently increased in wild-type embryos by autoregulation (Meno et al., 1999) while they generally remained low in M*vg1* embryos (Figure 5-2C). These results suggest that Vg1 is required for the auto-induction but not initiation of Nodal gene expression, and that the remaining endogenous levels of Nodal are not able to induce mesendoderm in the absence of Vg1.

To test whether the Nodal ligands might be inactive in the absence of Vg1, we overexpressed *cyc* or *sqt* in M*vg1* embryos and analyzed Nodal target gene expression. High levels (50 pg of mRNA) of *cyc* failed to induce target gene expression in M*vg1* embryos (Figure 5-2D), whereas *sqt* at low (0.2 pg) but not high (2-50 pg) levels of overexpression failed to induce target gene expression (Figures 2D and E). Co-expression of 20 pg of *cyc* mRNA in M*vg1* embryos with increasing concentrations of *vg1* mRNA caused an increase in induction of Nodal target gene

expression with co-expression of 5 pg of *vg1* (Figure 5-2F). These results indicate that Vg1 is necessary for Cyc activity and partially needed for Sqt activity.

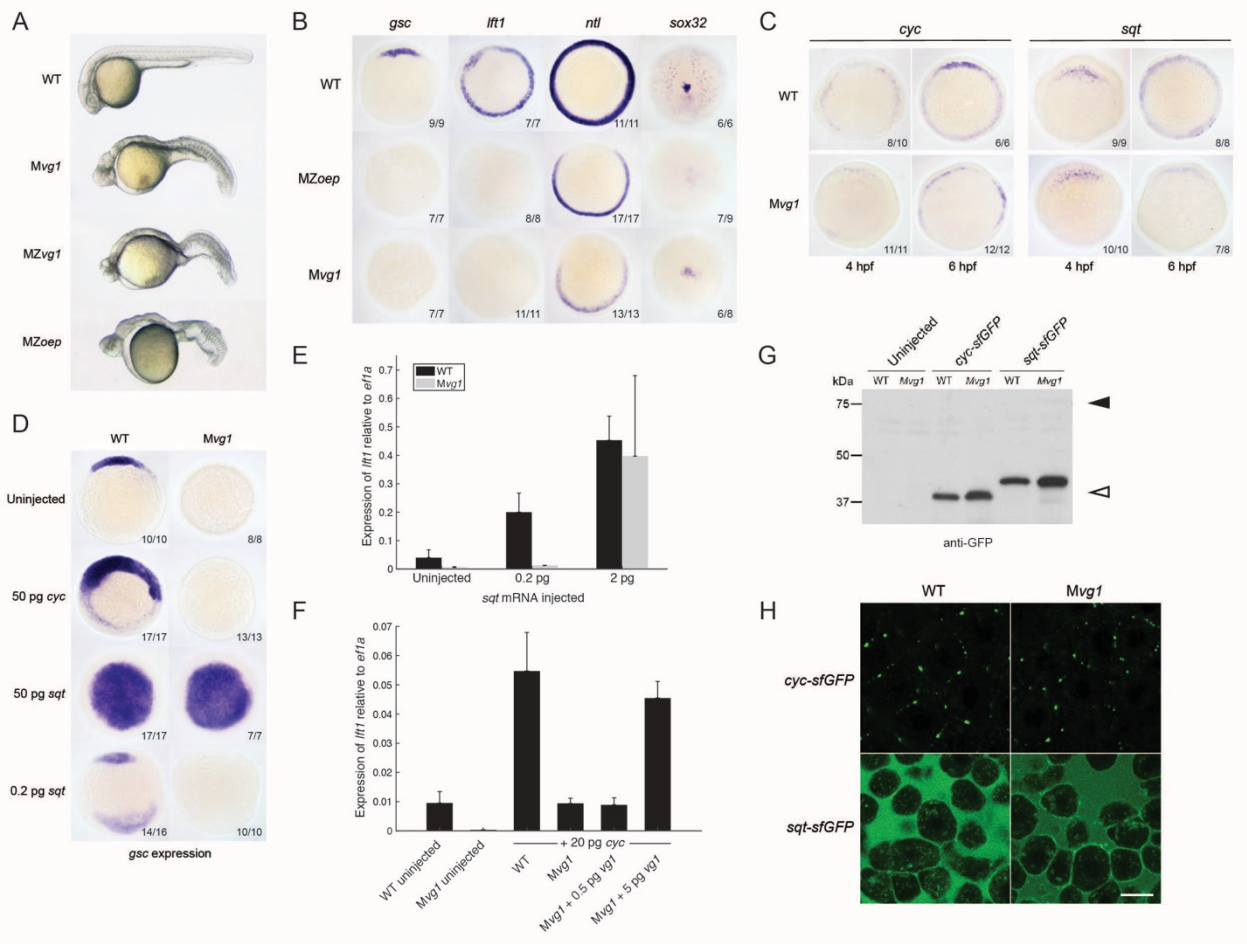


Figure 5-2. Endogenous Nodal signaling requires Vg1.
(A) M*vg1*, maternal-zygotic *vg1* (MZ*vg1*) and maternal-zygotic *oep* (MZ*oep*) mutants at 28 hpf.
(B) Expression of Nodal target genes *gsc*, *lft1* and *ntl* at 50% epiboly and *sox32* at 90% epiboly in WT, M*vg1* and MZ*oep* embryos.
(C) *cyc* and *sqt* expression at 4 and 6 hpf in WT and M*vg1* embryos.
(D) *gsc* expression at 50% epiboly in WT and M*vg1* embryos injected with 50 pg of *cyc* mRNA, 50 pg or 0.2 pg of *sqt* mRNA.
(E) qPCR of *lft1* expression at 50% epiboly relative to *ef1a* in WT and M*vg1* embryos injected with 0.2 pg or 2 pg of *sqt* mRNA. The mean and standard error of the mean (SEM) was plotted.
(F) qPCR of *lft1* expression at 50% epiboly relative to *ef1a* in embryos injected with 20 pg of *cyc* mRNA in combination with 0.5 pg or 5 pg of *vg1* mRNA. The mean and SEM was plotted.
(G) Anti-GFP reducing immunoblot of WT and M*vg1* embryos injected with 50 pg of *cyc-sfGFP* or *sqt-sfGFP* mRNA. Black arrowhead indicates the position of full-length protein, open arrowhead indicates processed protein. 8 embryos at 50% epiboly were loaded per well.
(H) Live imaging of the animal cap of sphere-stage WT and M*vg1* embryos injected with 50 pg of *cyc-sfGFP* or *sqt-sfGFP* mRNA. Scale bar, 17 um.

To determine whether the Nodal ligands are processed and secreted in M*vg1* embryos, we expressed superfolderGFP (sfGFP)-tagged derivatives of Cyc and Sqt (Muller et al., 2012; Pédelacq et al., 2006). No differences in cleavage or localization of Cyc and Sqt were detected in the presence or absence of Vg1 (Figures 2G and H). Taken together, these results suggest that Vg1 is necessary for the endogenous activities, but not the processing and secretion, of Cyc and Sqt.

Vg1 processing requires Nodal

Previous studies did not detect Vg1 processing in early embryos (Dale et al., 1989; Dohrmann et al., 1996; Tannahill and Melton, 1989; Thomsen and Melton, 1993). To examine the relationship of Vg1 processing to presence or absence of Nodal proteins, we first inserted sfGFP downstream of the Vg1 cleavage site (Figure 5-3A). *vg1-sfGFP* rescued M*vg1* mutants (Figure 5-3B) but cleavage of Vg1 protein was undetectable (Figure 5-3C). To test whether Vg1 needs to be processed to be functional, we mutated the basic residues in the Vg1 cleavage site to non-basic residues (Vg1-NC, "Non-Cleavable" (Figure 5-3E)). This abolished Vg1 rescuing activity (Figure 5-3D), suggesting that endogenous Vg1 cleavage is not detectable but is required for Vg1 function.

Given that the M*vg1* phenotype resembles Nodal loss-of-function phenotypes, and Vg1 requires its cleavage site, we asked whether Nodal might induce Vg1 cleavage. We co-expressed *vg1-sfGFP* with *cyc* or *sqt* and discovered that Vg1-sfGFP was cleaved to its mature form in the presence of Nodal (Figure 5-3E, Appendix Figure 5). By contrast, Vg1-sfGFP was not cleaved upon co-expression with an alternative TGF-beta-related ligand, *bmp7a* (Appendix Figure 5B),

and non-cleavable Vg1-sfGFP (Vg1-NC-sfGFP) was not cleaved in the presence of Cyc or Sqt (Figure 5-3E, Appendix Figure 5A). These data reveal that Nodal induces Vg1 processing.

Vg1 secretion requires Nodal

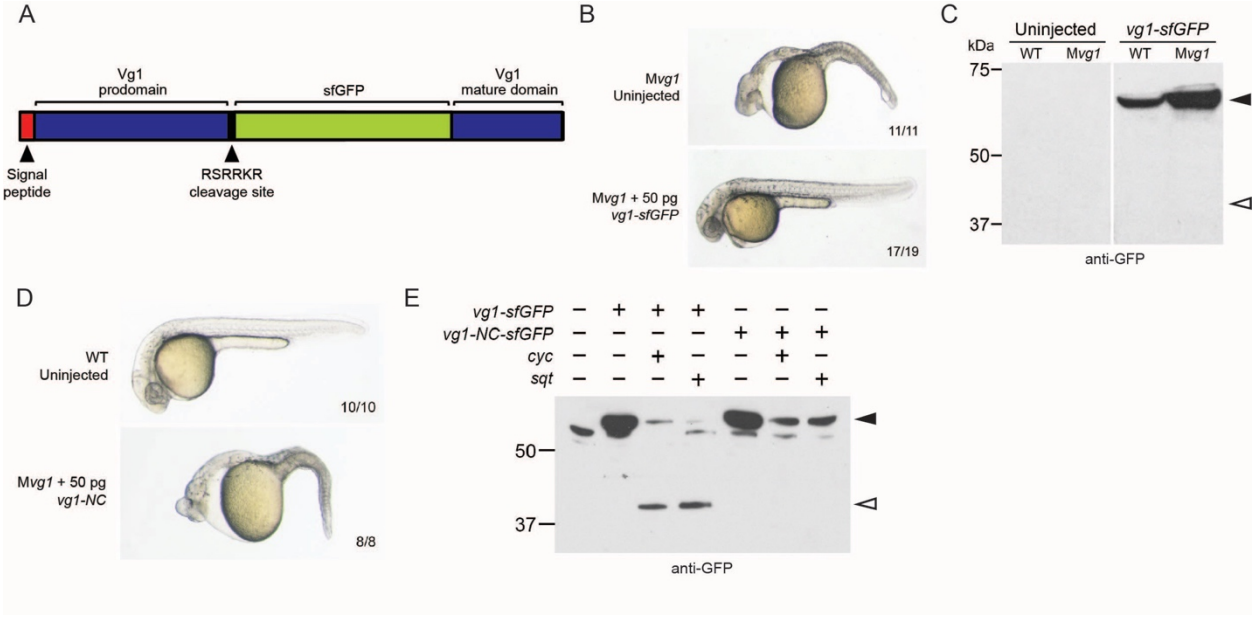


Figure 5-3. Vg1 processing requires Nodal.
(A) *superfolderGFP* (*sfGFP*) was inserted into *vg1* downstream of the predicted basic cleavage site.
(B) M*vg1* embryo injected with 50 pg of *vg1-sfGFP* mRNA, shown at 28 hpf.
(C) Anti-GFP reducing immunoblot of WT and M*vg1* embryos injected with 50 pg of *vg1-sfGFP* mRNA. Black arrowhead indicates full-length Vg1-sfGFP; open arrowhead indicates the predicted size of cleaved Vg1-sfGFP. 8 embryos at 50% epiboly were loaded per well.
(D) M*vg1* embryo injected with 50 pg of non-cleavable *vg1* mRNA (*vg1-NC,* RSRRKR->SQNTSN), shown at 28 hpf. Embryos were injected with up to 200 pg of *vg1-NC* mRNA with no rescue.
(E) Anti-GFP reducing immunoblot of M*vg1* embryos injected with 10 pg of *vg1-sfGFP* or *vg1-NC-sfGFP* mRNA and 10 pg of *cyc* or *sqt* mRNA. Black arrowhead indicates full-length Vg1-sfGFP, open arrowhead indicates cleaved Vg1-sfGFP. Molecular weights in kDa. 8 embryos at 50% epiboly were loaded per well. See also Appendix Figure 5.

To examine the secretion and localization of Vg1, we expressed *vg1-sfGFP* in wild-type or M*vg1* embryos for *in vivo* imaging. In contrast to the extracellular localization of Cyc and Sqt (Figure 5-2H), Vg1 was only detected intracellularly, predominantly in the endoplasmic reticulum (ER) (Figure 5-4A) (Fodero-Tavoletti et al., 2005; Southall et al., 2006; Szul and Sztul, 2011; Tu et al., 2002).

To determine whether Nodal can induce not only Vg1 processing but also secretion, we co-expressed *vg1-sfGFP* with *cyc* or *sqt*. Notably, Vg1-sfGFP formed extracellular puncta and/or diffuse extracellular signal upon co-expression with *cyc* or *sqt* (Figure 5-4B, Appendix Figure 6A, Table 1). By contrast, Vg1-sfGFP was not secreted upon co-expression with *bmp7a* (Figure 5-4B).

To directly test whether Vg1 is secreted in the presence of Nodal, we tagged Vg1 with the pH-sensitive fluorescent protein pHluorin2 (Mahon, 2011). pHluorin2 is non-fluorescent at acidic pH, as found in intracellular vesicles, but it fluoresces in the neutral pH of the extracellular space. Vg1-pHluorin2 fluorescent puncta were only visible upon co-expression with *cyc* or *sqt*, indicating that Vg1 is secreted in the presence of Nodal (Figure 5-4C, Appendix Figure 6B).

To independently test if Vg1 is only secreted in the presence of Nodal, we expressed *vg1-sfGFP* in single-cell embryos and co-expressed *cyc-RFP* in 1 cell at the 16-cell stage. At sphere stage, Vg1-sfGFP was only secreted in the cells that also expressed Cyc-RFP (Figure 5-4D).

To determine whether Vg1 and Nodal co-localize, we co-expressed *vg1-sfGFP* with *cyc-RFP* or *sqt-RFP.* Vg1-sfGFP displayed extensive extracellular co-localization with Cyc-RFP and Sqt-RFP (Figure 5-4E, Appendix Figure 6C). Taken together, these results reveal that Nodal induces the secretion of Vg1, and that Vg1 and Nodal co-localize.

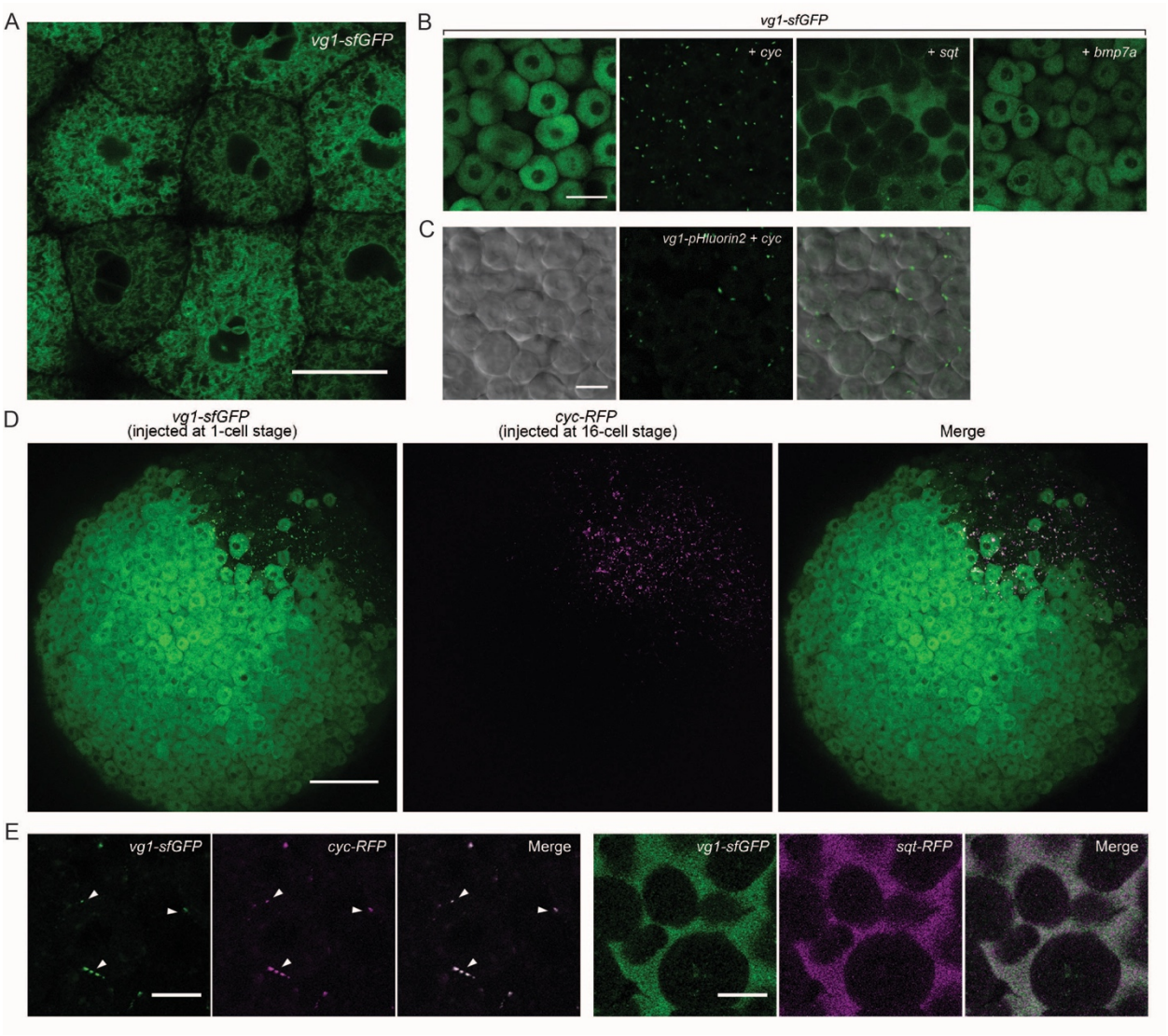


Figure 5-4. Vg1 secretion requires Nodal.
(A) Live imaging of M*vg1* embryo injected with 50 pg of *vg1-sfGFP* mRNA. Scale bar, 17 um.
(B) M*vg1* embryos co-injected with 50 pg of *vg1-sfGFP* mRNA and 50 pg of *cyc, sqt* or *bmp7a* mRNA. Scale bar, 17 um. See also Appendix Figure 6A.
(C) M*vg1* embryo co-injected with 50 pg of pH-sensitive fluorescent *vg1* (*vg1-pHluorin2)* and 50 pg of *cyc* mRNA. Scale bar, 17 um. See also Appendix Figure 6B.
(D) M*vg1* embryo co-injected with 50 pg of *vg1-sfGFP* mRNA at the 1-cell stage and 10 pg of *cyc-RFP* mRNA into 1 cell at the 16-cell stage. Scale bar, 100 um.
(E) M*vg1* embryos co-injected with 50 pg of *vg1-sfGFP* mRNA and 50 pg of *cyc-* or *sqt-RFP* mRNA. Arrowheads indicate examples of co-localization. Scale bar, 17 um. See also Appendix Figure 6C.

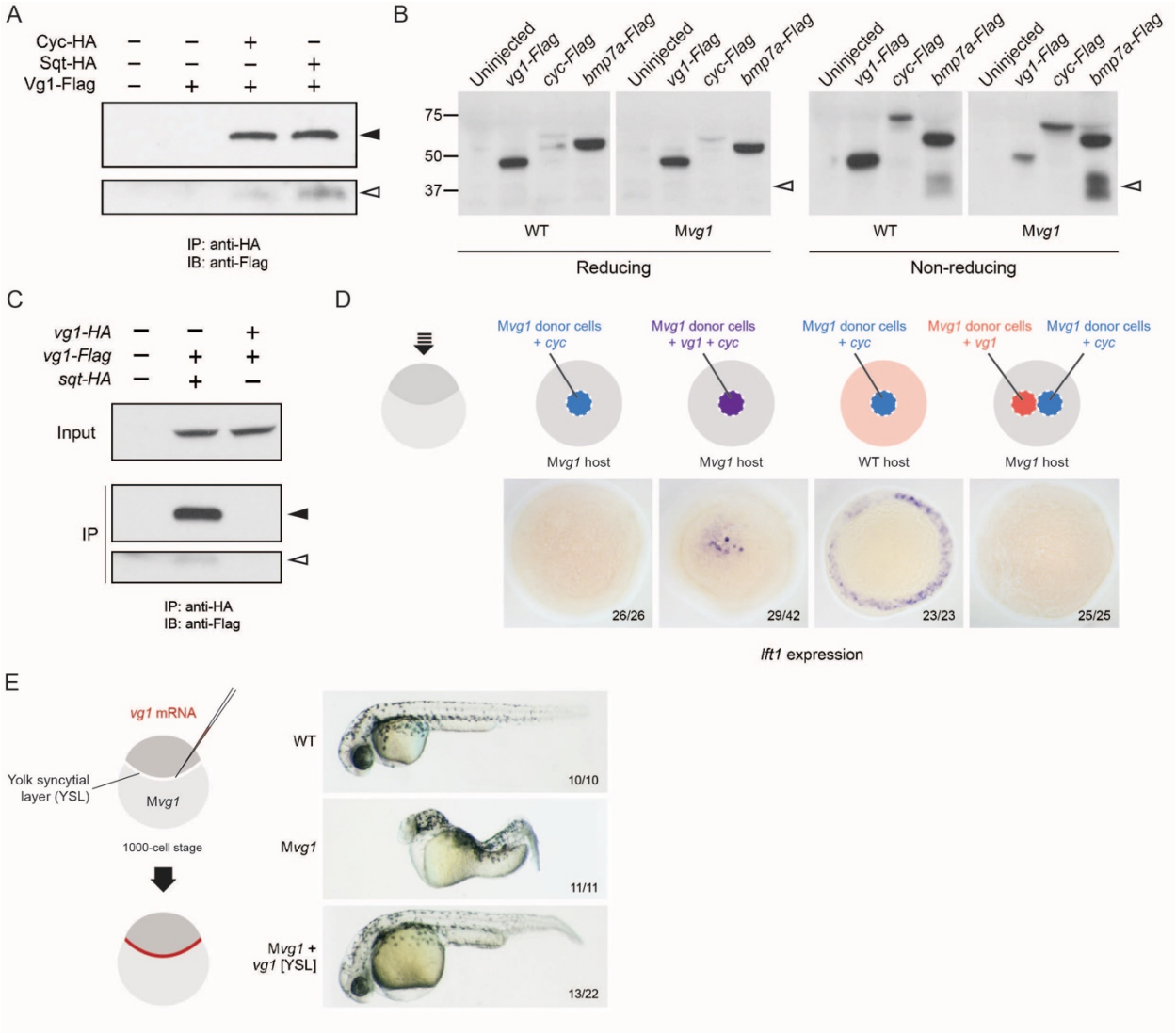


Figure 5-5. Vg1 and Nodal form heterodimers, and are only active when co-expressed.

(A) Anti-Flag reducing immunoblot (IB) of anti-HA immunoprecipitates (IP) from lysates of M*vg1* embryos injected with 50 pg of *cyc-HA*, *sqt-HA* and/or 50 pg of *vg1-Flag* mRNA. Black arrowhead indicates full-length Vg1-Flag; open arrowhead indicates cleaved Vg1-Flag. See also Appendix Figures 7A and B.

(B) Anti-Flag reducing and non-reducing immunoblots of WT and M*vg1* embryos injected with 50 pg of *vg1-Flag*, *cyc-Flag* or *bmp7a-Flag,* collected at 50% epiboly. Under reducing conditions the proteins migrated at sizes consistent with the theoretical molecular weights for full-length monomers: Vg1-Flag – 42 kDa; Cyc-Flag – 58 kDa; Bmp7a-Flag – 50 kDa. Open arrowhead indicates where mature Bmp7a-Flag homodimers are expected to migrate as 2 species under non-reducing conditions (Little and Mullins, 2009). For an annotated gel, see Appendix Figure 7C.

(C) Anti-Flag reducing immunoblot of anti-HA IP from lysates of M*vg1* embryos injected with 50 pg of *sqt-HA* and *vg1-Flag* mRNAs or *vg1-HA* and *vg1-Flag* mRNAs. Black arrowhead indicates full-length Vg1-Flag; open arrowhead indicates cleaved Vg1-Flag.

(D)Transplantation of cells from donor embryos injected with *cyc*, *vg1* or *cyc* and *vg1* mRNAs into host embryos for analysis of Nodal target gene *lft1* expression at 50% epiboly. mRNAs were co-injected with *sfGFP* mRNA to verify successful transplantation using DAB staining (Figure 5-Figure Supplement 2A).

(E) Injection of *vg1* mRNA into the yolk syncytial layer (YSL) of M*vg1* mutants, shown at 32 hpf. For Nodal target gene expression in *vg1* mRNA YSL-injected embryos see Figure 5-Figure Supplement 2B. *vg1* was co-injected with a fluorescent dextran to verify YSL localization (Figure 5-Figure Supplement 2C).

Vg1 and Nodal form heterodimers

The co-localization of Vg1 and Nodal suggested that these secreted ligands might form heterodimers, as detected for GDF1 and Nodal, and some other TGF-beta-related signals (Aono et al., 1995; Dutko and Mullins, 2011; Eimon and Harland, 2002; Fuerer et al., 2014; Guo and Wu, 2012; Hazama et al., 1995; Israel et al., 1996; Little and Mullins, 2009; Nishimatsu and Thomsen, 1998; Schmid et al., 2000; Shimmi et al., 2005; Suzuki et al., 1997; Tanaka et al., 2007). To test this hypothesis, we performed co-immunoprecipitation experiments by co-expressing 50 pg of *vg1-Flag* with 50 pg of *cyc-HA* or *sqt-HA.* Vg1-Flag co-immunoprecipitated with Cyc-HA or Sqt-HA (Figure 5-5A, Appendix Figure 7A). To test the specificity of this interaction we used two different concentrations of *sqt-HA* mRNA in combination with three different concentrations of *vg1-Flag* or *bmp7a-Flag* mRNA. We detected an interaction between Sqt-HA and Vg1-Flag at all 6 concentrations tested, whereas an interaction between Sqt-HA and Bmp7a-Flag was only detected at the highest concentration of each mRNA (Appendix Figure 7B). Thus, Vg1 specifically interacts with Nodal to form heterodimers.

The heterodimerization of Vg1 and Nodal raises the possibility that Vg1 is maintained in a monomeric state in the absence of Nodal. Indeed, a previous study found that Vg1 does not form homodimers, and that endogenous Vg1 is predominantly monomeric (Dale et al., 1993). To test the monomeric or dimeric states of Vg1 in the absence of Nodal, we performed reducing and non-reducing immunoblots of wild-type and M*vg1* embryos expressing *vg1-Flag, cyc-Flag* or *bmp7a-Flag* mRNA. TGF-beta family members are disulfide-linked dimers: under reducing conditions the disulfide bonds are broken, while under non-reducing conditions the bonds are maintained, allowing the detection of dimers. Bmp7a-Flag mature homodimers were visible under non-reducing conditions, whereas Vg1-Flag homodimers were not detected (Figure 5-5B,

Appendix Figure 7C). Using a complementary approach, we tested whether Vg1 forms homodimers by co-immunoprecipitation. While Sqt-HA and Vg1-Flag co-precipitated, Vg1-HA and Vg1-Flag did not (Figure 5-5C). These results indicate that Vg1 does not form homodimers and might be present as monomers in the absence of Nodal.

Vg1 protein is synthesized before Nodal transcription and translation begin, raising the possibility that newly synthesized Nodal monomers bind to preexisting Vg1 monomers. Alternatively, Nodal might only heterodimerize with Vg1 protein that is co-translated with Nodal. To distinguish between these possibilities, we generated a Vg1-Dendra2 photoconvertible fusion protein and injected it at the 1-cell stage. At the 64-cell stage we photoconverted Vg1-Dendra2 from green to red, and co-injected the embryos with 5 pg of *cyc* mRNA. Imaging revealed the production of red puncta, indicating that Vg1 protein synthesized prior to Nodal synthesis was able to heterodimerize and be secreted with Nodal (Appendix Figure 7D). This data suggests that Nodal can heterodimerize with pre-existing Vg1.

Vg1 and Nodal are only active when co-expressed

To determine if co-expression of Vg1 and Nodal in the same cells is required for activity, we used transplantation assays to compare target gene induction in cells co-expressing *vg1* and *cyc* versus neighboring cells expressing either *vg1* or *cyc.* Nodal target gene induction only occurred when *vg1* and *cyc* were co-expressed in the same cells (Figure 5-5D and Appendix Figure 8A). Analogously, deposition of *vg1* mRNA to the yolk syncytial layer (YSL) of M*vg1* mutants, where *cyc* and *sqt* are expressed endogenously, was sufficient to rescue M*vg1* mutants by morphology and gene expression (Figure 5-5E and Appendix Figure 8B). Confocal imaging of embryos injected with *vg1* mRNA and a fluorescent dextran into the YSL indicated that the majority of

fluorescence was localized to the YSL, but a few cells in the margin also inherited the fluorescent dextran (Appendix Figure 8C). Thus, although *vg1* is ubiquitously expressed in the early embryo (Helde and Grunwald, 1993; Peterson et al., 2013) (Appendix Figure 8D), its co-localization with *cyc* and *sqt* is sufficient for its role in mesendoderm formation. Taken together, these results suggest that Vg1 and Nodal are active when expressed in the same cells, where they form heterodimers.

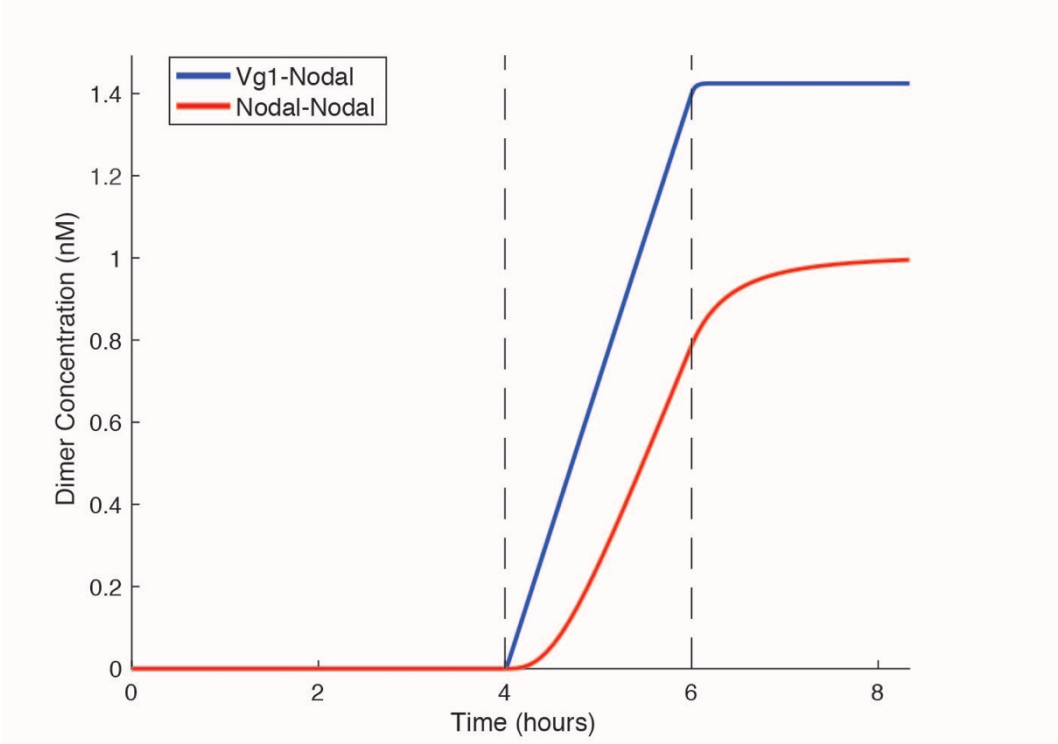


Figure 5-6. Vg1 can enable rapid response to low concentrations of Nodal.
Kinetic model comparing Nodal homodimer formation in the absence of Vg1 (red line) and Vg1-Nodal heterodimer formation in the presence of a maternal Vg1 pool (blue line). For both conditions Nodal monomer production begins at 4 hpf (the onset of zygotic transcription and translation, first dotted line) and concludes after mesendodermal patterning (6 hpf, second dotted line). For the heterodimer simulation an excess of Vg1 is provided in the initial conditions.

Vg1 can enable rapid response to low concentrations of Nodal

The requirement for Vg1-Nodal heterodimers for mesendoderm induction raises the question of why the embryo relies on both a ubiquitous ligand, Vg1, and localized ligands, Cyc and Sqt. We developed a basic kinetic model to test the rate of Nodal homodimer formation versus Nodal-Vg1 heterodimer formation in the presence of a maternal Vg1 pool. Simulating these two conditions revealed that the preloading of inactive Vg1 monomers in the cell allows Nodal to immediately form heterodimers whereas dimer formation is delayed when Nodal must form homodimers (Figure 5-6). Thus, Vg1-Nodal heterodimers can initiate signaling more quickly than Nodal homodimers, and already at low Nodal levels. Even if Nodal homodimers were as active as Vg1-Nodal heterodimers, Nodal target gene induction would still be slower in the absence of maternal Vg1, because the association of two Nodal monomers is less likely at low Nodal concentrations than Vg1-Nodal dimerization (Figure 5-6). These simulations reveal that low concentrations of zygotic Nodal can be directly transformed into pathway activation via association with maternal Vg1.

Discussion

The results in this study indicate that mesendoderm induction depends on the co-expression and heterodimer formation of Nodal and Vg1. This conclusion is based on five new findings: Vg1 is essential for mesendoderm induction; Vg1 is only processed, secreted and active in the presence of Nodal; Nodal activity, but not processing and secretion, depends on Vg1; Vg1 and Nodal form heterodimers; and Vg1-Nodal heterodimers are more active than Nodal alone. Together with previous studies, our findings suggest a unifying 5-step model for mesendoderm

induction in zebrafish: (1) *vg1* mRNA is inherited from the mother and is ubiquitous in the early embryo; (2) Vg1 protein is synthesized ubiquitously and retained predominantly in the ER; (3) *cyc* and *sqt* are transcribed and translated in the YSL; (4) Cyc and Sqt form heterodimers with pre-existing Vg1, resulting in Vg1 secretion and cleavage; (5) Cyc-Vg1 and Sqt-Vg1 heterodimers activate the Nodal signaling pathway to induce mesendoderm (Figure 5-7).

Vg1 is as essential as Nodal for mesendoderm induction

Knockdown studies in zebrafish suggested no requirement for Vg1 in mesendoderm induction (Peterson et al., 2013), but the loss-of-function mutants reported here reveal that Vg1 is absolutely required for the induction of head and trunk mesoderm and endoderm. Strikingly, *vg1* mutants strongly resemble Nodal signaling mutants, showing that zebrafish Vg1 has as essential a function as Nodal, which has been considered the sole mesendoderm inducer.

The results in zebrafish warrant a re-analysis of the requirements for Vg1 orthologs in other systems. For example, mouse *Gdf1;Gdf3* double mutants have incompletely penetrant mesendodermal phenotypes (Andersson et al., 2007). A closer comparison to *Nodal* mutants might reveal functions of GDF1 and GDF3 that are equivalent to Vg1. It is also possible that mouse Nodal is expressed at sufficiently high levels to be less dependent on GDF1/GDF3, akin to the overexpression of zebrafish *sqt*. Similarly, zebrafish *southpaw* might be expressed at high enough levels to act independently of zygotic *vg1* during left-right development. Knockdown studies in *Xenopus* have suggested that Vg1 is mainly involved in inducing notochord precursors, but not other mesendodermal progenitors (Birsoy et al., 2006). Mutant studies in *Xenopus* might reveal broader roles for Vg1, or alternatively additional TGF-beta-related signals such as Derrière

(Sun et al., 1999), which has been shown to interact with Nodal (Eimon and Harland, 2002), might have complementary or overlapping functions with Vg1.

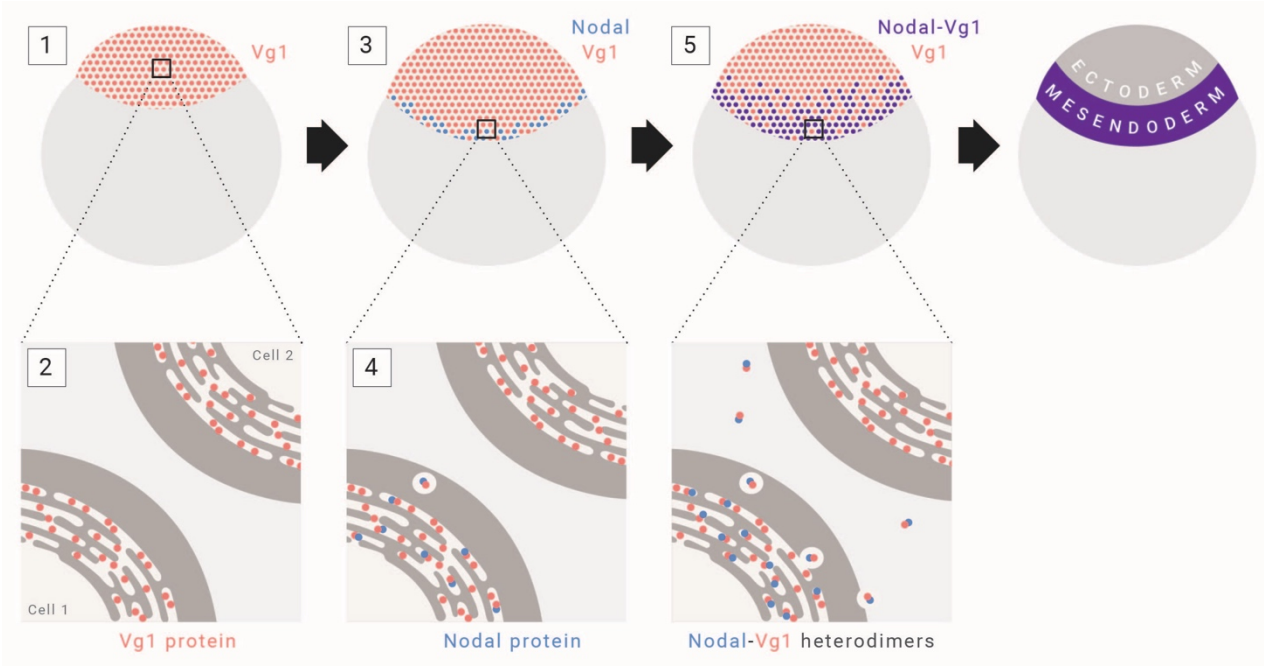


Figure 5-7. Model for mesendoderm induction in zebrafish.
As described in the main text: (1) *vg1* mRNA is inherited from the mother, and is ubiquitous in the early embryo; (2) Vg1 protein is synthesized ubiquitously and retained predominantly in the ER; (3) *cyc* and *sqt* are transcribed and translated in the YSL; (4) Cyc and Sqt form heterodimers with Vg1, resulting in Vg1 secretion and cleavage; (5) Cyc-Vg1 and Sqt-Vg1 heterodimers activate the Nodal signaling pathway to induce mesendoderm.

More generally, it is conceivable that the activities of the Nodal and Vg1/GDF1/GDF3 subfamilies are co-dependent in all contexts. This idea is not only supported by the co-dependence of Nodal and Vg1 in zebrafish mesendoderm induction reported here, but also the observation that Nodal expression coincides with the expression of Vg1 family members in numerous contexts (Agius et al., 2000; Levin et al., 1995; Onai et al., 2010; Range and Lepage, 2011; Range et al., 2007; Seleiro et al., 1996). Moreover, mouse *Gdf1* mutants display very similar left-right defects as mutants with impaired Nodal signaling (Andersson et al., 2006; Cheng et al., 2003; Lowe et al.,

2001; Yan et al., 1999). It is therefore tempting to speculate that wherever and whenever Nodal subfamily members are expressed and active, they are accompanied by Vg1 subfamily members. In this scenario, Vg1/GDF1/GDF3 act in parallel with Nodal even when their expression is precedent, and any apparent upstream functions of Vg1 (Andersson et al., 2007; Chen et al., 2006; Rankin et al., 2000; Shah et al., 1997; Skromne and Stern, 2001; Tanaka et al., 2007) are actually the result of Nodal autoinduction: Vg1 is required together with Nodal to fully activate Nodal gene expression but it is not needed for the initiation of Nodal expression, as shown in the M*vg1* mutants. Finally, Nodal and Vg1 act through the same coreceptors (Andersson et al., 2007; Cheng et al., 2003; Fuerer et al., 2014; Tanaka et al., 2007) and are both inhibited by Lefty ligands (Agathon et al., 2001; Bisgrove et al., 1999; Chen and Shen, 2004; Chen and Schier, 2002; Cheng et al., 2004; Meno et al., 1999; 1996; Thisse et al., 2000; Thisse and Thisse, 1999). These observations suggest that Nodal signaling should henceforth be considered Nodal/Vg1 signaling.

Vg1 is only processed, secreted and active in the presence of Nodal

Our study clarifies previously puzzling observations on the activity and processing of Vg1 that contrast with the properties of other TGF-beta-related signals: overexpression of wild-type *vg1* does not cause a phenotype (Dale et al., 1993; Dohrmann et al., 1996; Tannahill and Melton, 1989; Thomsen and Melton, 1993), neither secreted (Dale et al., 1993; Tannahill and Melton, 1989) nor processed Vg1 has been reliably detected (Dale et al., 1989; Dohrmann et al., 1996; Tannahill and Melton, 1989; Thomsen and Melton, 1993), but fusion of the Vg1 mature domain to the Activin or BMP prodomains results in processed and active Vg1 (Dale et al., 1993; Dohrmann et al., 1996; Kessler and Melton, 1995; Thomsen and Melton, 1993; Wall et al., 2000). Our study explains these conundrums by revealing that Vg1 is only processed, secreted and

active in the presence of Nodal. Without Nodal, Vg1 is unprocessed and predominantly resides in the ER. Upon overexpression, Vg1 contributes to this inert pool, and only fusion to heterologous prodomains allows secretion, cleavage and activation in the absence of Nodal. Thus, the dependence of Vg1 processing, secretion and activity on Nodal accounts for many of the previously confusing observations. One conundrum remains: we and others have not been able to detect the processing (Figure 5-3C) or secretion (Figure 5-4A) of Vg1 at endogenous levels of Nodal. We speculate that endogenous Nodal is expressed at very low levels and in few cells, resulting in cleavage and secretion of only a small (and undetectable) fraction of the total pool of Vg1. Only upon ectopic Nodal expression, sufficiently high levels of Vg1 are processed to become detectable. The development of more sensitive detection methods is needed to directly demonstrate the cleavage, processing and secretion of endogenous Vg1.

Vg1 and Nodal form heterodimers

Our study reveals that Nodal and Vg1 form heterodimers, and that Vg1 exists in a monomeric state prior to heterodimerization with Nodal. The initial localization of Vg1 to the ER suggests that this is the site of heterodimerization, which is consistent with previous studies of other heterodimers (Duitman et al., 2008; Hurtley and Helenius, 1989; Jalah et al., 2013; Lorenz et al., 2002; Persson and Pettersson, 1991; Tu et al., 2002). For example, in the case of the uroplakin proteins UPIb and UPIII, UPIb can autonomously exit the ER and translocate to the plasma membrane. By contrast, UPIII must heterodimerize with UPIb in the ER in order to exit and move to the plasma membrane (Tu et al., 2002). Although we currently favor a model in which monomeric Vg1 meets Nodal in the ER, more complex scenarios are conceivable. For example, it is unclear whether pre-existing Vg1 might associate with other ER-resident proteins to maintain or prepare it in a state that allows dimerization with newly synthesized Nodal.

Our results suggest that Nodal-Vg1 heterodimers are more potent than Nodal alone: in the case of Cyc, such heterodimers seem to be required for Cyc to activate signaling, whereas Sqt-Vg1 heterodimers appear to be more active than Sqt alone (Figures 5-2D, E and F). The molecular basis of the increased activity is unknown, but based on previous studies in the BMP system, heterodimers might be necessary to assemble heteromeric combinations of two types of class I or II receptors (Little and Mullins, 2009).

Our results also extend and generalize the previous observation that mouse GDF1 and Nodal form heterodimers (Fuerer et al., 2014; Tanaka et al., 2007), although those studies did not address the requirement, localization, or processing of Vg1/GDF1/GDF3 during mesendoderm formation, and instead proposed that heterodimer formation might increase the potency and/or range of Nodal. Our results uncover the alternative or additional mechanism that heterodimer formation triggers processing and secretion of Vg1 and allows Nodal to be active at physiological concentrations.

_vg1_ mRNA and protein do not need to be localized in the embryo

Our results demonstrate a novel mode to restrict TGF-beta-related protein activity through heterodimer formation. *vg1* mRNA and protein do not need to be localized in the embryo to restrict Vg1 activity: instead, it is the absence of Nodal that blocks Vg1 processing, secretion and activity. Indeed, zebrafish Vg1 is present ubiquitously in early embryos and *vg1* is expressed in broader domains than Nodal in all systems analyzed. The question therefore arises whether the exquisite vegetal localization of *Xenopus vg1* is important for development (Weeks and Melton, 1987). The localized activation of *Xenopus* Nodal genes might be sufficient to restrict mesendoderm formation to vegetal and marginal regions, but it is also possible that localized

Vg1 provides an additional safeguard to spatially restrict pathway activation. Rescue experiments similar to those reported here could address this question.

Vg1 can enable rapid response to low concentrations of Nodal

Modeling of hetero- and homodimerization kinetics reveals that the maternal pool of Vg1 accelerates the onset of ligand dimerization relative to a system that relies on Nodal-Nodal dimerization alone. This could be advantageous in the embryo, where mesendoderm induction cannot initiate until after the maternal-to-zygotic transition. Although it may be counterintuitive for a spatially localized signal to rely on a ubiquitous signal for pathway activation, the preloading of Vg1 in the ER could be instrumental for ensuring Nodal signaling initiates in a rapid and temporally reliable manner. Thus, the requirement for Vg1 in the zebrafish embryo can ensure rapid and sensitive response to the low concentrations of Nodal that initiate mesendoderm induction.

Concluding remarks

The finding that Vg1 – together with Nodal – is an endogenous mesoderm inducer resolves some of the historical controversies in the field. Vg1 was described in 1987 as a TGF-beta-related signal present at the right time and place to be a mesoderm inducer (Weeks and Melton, 1987), but the lack of a functional requirement raised doubts about its importance. Conversely, Activin was reported in 1990 as a TGF-beta-related signal that can induce mesoderm (Smith et al., 1990; van den Eijnden-van Raaij et al., 1990), but its absence during early embryogenesis (Dohrmann et al., 1993; Thomsen et al., 1990), and the lack of loss-of-function phenotypes (Hawley et al., 1995; Kessler and Melton, 1995; Matzuk et al., 1995; Schulte-Merker et al., 1994; Sun et al.,

1999), raised doubts about its importance (Schier and Shen, 2000). With the discovery of the essential roles of mouse Nodal (Conlon et al., 1994; Zhou et al., 1993) and zebrafish Nodal (Feldman et al., 1998) in mesoderm induction, the field converged to the view that Nodal is the key inducer. Our study indicates instead that Nodal-Vg1 heterodimers are the essential endogenous inducers of mesendoderm, while Activin serves as a powerful reagent to induce mesendoderm from embryonic stem cells.

Materials and Methods

Ethics statement

All vertebrate animal work was performed at the facilities of Harvard University, Faculty of Arts & Sciences (HU/FAS). The HU/FAS animal care and use program maintains full AAALAC accreditation, is assured with OLAW (A3593-01), and is currently registered with the USDA. This study was approved by the Harvard University/Faculty of Arts & Sciences Standing Committee on the Use of Animals in Research & Teaching under Protocol No. 25–08.

CRISPR/Cas9-mediated mutagenesis of *vg1*

sgRNAs targeting the *vg1*/*dvr1*/*gdf3* gene were designed using CHOPCHOP (Labun et al., 2016; Montague et al., 2014) and synthesized as previously described (Gagnon et al., 2014) (See also Appendix Figure 3). *vg1* sgRNAs were co-injected with ~0.5 nL of 50 uM Cas9 protein into TLAB wild-type embryos. Injected embryos were raised to adulthood and outcrossed to TLAB adults. Clutches of embryos with potential heterozygous individuals were used to identify founders with germline mutations in *vg1* by extracting DNA from 10 embryos and genotyping by MiSeq sequencing. The offspring of confirmed founders were raised to adulthood and genotyped to

identify heterozygous *vg1* adults. Heterozygous *vg1* mutants were intercrossed to generate zygotic homozygous (*Zvg1*) fish. For maintaining the *vg1* mutant line, homozygous *Zvg1* male fish were crossed to heterozygous female fish, and the resulting progeny were genotyped to identify *Zvg1* adults. To generate maternal *vg1* (M*vg1*) mutants, TLAB wild-type male fish were crossed to homozygous *Zvg1* female fish.

Genotyping of *vg1* mutants

Two deletion alleles of 8 bp and 29 bp (*vg1*^{a164} and *vg1*^{a165} respectively) were recovered in the first exon of *vg1* from the sgRNA targeting the sequence GGGTCAGAAGACAGGCTCTGAGG. Genomic DNA was extracted using the HotSHOT method (Meeker et al., 2007) and PCR was performed using standard conditions (see primer sequences below), followed by Sanger sequencing or MiSeq sequencing for the 8 bp allele (Gagnon et al., 2014) or 2% gel electrophoresis for the 29 bp allele.

Cloning of expression constructs

The *vg1* CDS sequence was PCR amplified from a high-stage cDNA library and cloned into the pSC vector (Agilent) with a beta-globin 5'UTR and an SV40 3'UTR using Gibson assembly (Gibson et al., 2009) to generate pSC-*vg1*. To generate pCS2(+)-*cyc* and pCS2(+)-*sqt,* the *cyc* and *sqt* CDS sequences were PCR amplified from a high-stage cDNA library and cloned into the pCS2(+) vector using Gibson assembly. To generate non-cleavable forms of *vg1* and *vg1-sfGFP*, site-directed mutagenesis was used to replace the RSRRKR cleavage site with SQNTSN using a Q5 Site-Directed Mutagenesis Kit (NEB).

Cloning of fusion and epitope tag constructs

All superfolder GFP (sfGFP) (Pédelacq et al., 2006), RFP, Dendra2 and pHluorin2 (Mahon, 2011) fusion constructs were generated by PCR-based methods and cloned into the pCS2(+) vector using Gibson assembly. Flag (DYKDDDDK) and HA tag (YPYDVPDYA) sequences were inserted by site-directed mutagenesis of pSC-*vg1*, pCS2(+)-*cyc* and pCS2(+)-*sqt*. For Vg1 fusions, sequences encoding the fluorescent protein or Flag tag were inserted downstream of the cleavage site (RSRRKR) with a GSTGTT linker separating the prodomain and fluorescent protein, and a GS linker separating the fluorescent protein and the Vg1 mature domain. For Cyc fusions, sequences encoding the fluorescent proteins or HA tag were inserted 2 amino acids downstream of the cleavage site (RRGRR) (Muller et al., 2012). For Sqt fusions, fluorescent protein and HA tag sequences were inserted 10 amino acids downstream of the cleavage site (RRHRR) with a GSTGTT linker separating the prodomain and fluorescent protein, and a GS linker separating the fluorescent protein and the mature domain (Muller et al., 2012).

mRNA synthesis and microinjection

Vectors were linearized by digestion with NotI (pCS2(+) vectors) or XhoI (pSC vectors). Capped mRNAs were synthesized using the SP6 or T7 mMessage Machine Kits (ThermoFisher), respectively. For *in situ* hybridization, immunoblot, imaging and qPCR experiments, embryos were dechorionated using 1mg/ml Pronase (Protease type XIV from Streptomyces griseus, Sigma) prior to injection, and subsequently cultured in agarose-coated dishes. Embryos were injected at the one-cell stage unless otherwise stated.

Zebrafish husbandry

Zebrafish embryos were grown at 28°C and staged according to (Kimmel et al., 1995). Embryos were cultured in blue water (250 mg/L Instant Ocean salt, 1 mg/L methylene blue in reverse osmosis water adjusted to pH 7 with NaHCO3).

Morphological analysis of *vg1* mutant phenotypes

Embryos were analyzed for mutant phenotypes at 28-32 hpf. For imaging, embryos were anesthetized in Tricaine (Sigma) and mounted in 2% methylcellulose then imaged using a Zeiss SteREO Discovery.V12 microscope.

Live imaging

Embryos were raised to sphere stage and mounted in 1% low gelling temperature agarose (Sigma) on glass-bottomed dishes (MatTek) with the animal pole facing the glass. Imaging was performed on Zeiss LSM 700 and LSM 880 inverted confocal microscopes.

Photoconversion

Embryos were injected at the 1-cell stage with 100 pg of *vg1-Dendra2* mRNA then grown at 28°C to the 64-cell stage and injected with 5 pg of *cyc* mRNA into 6 locations in the embryo. Embryos were mounted in 1% low gelling temperature agarose and photoconverted with 2 minutes of UV light at 10x magnification on the Zeiss LSM 700 inverted confocal microscope. The embryos were incubated at 28°C for 30 minutes before imaging on the LSM 700 microscope over a period of 2 hours.

Image adjustments

Images were processed in FIJI/ImageJ (Schindelin et al., 2012). Brightness, contrast and color balance was applied uniformly to images.

***in situ* hybridization and DAB staining**

Embryos were fixed in 4% formaldehyde overnight at room temperature. Whole mount *in situ* hybridizations were performed according to standard protocols (Thisse and Thisse, 2008). DIG-labeled antisense RNA probes against *cmlc2, spaw, gsc, lft1, ntl, sox32, cyc, sqt* and *vg1* were synthesized using a DIG Probe Synthesis Kit (Roche). NBT/BCIP/Alkaline phosphatase-stained embryos were dehydrated in methanol and imaged in benzyl benzoate:benzyl alcohol (BBBA) using a Zeiss Axio Imager.Z1 microscope. For DAB staining, embryos were rehydrated in PBST after completing the *in situ* protocol, and blocked in 10% normal goat serum/1% DMSO before incubation in primary antibody overnight (1:400 rabbit anti-GFP-HRP, ThermoFisher A10260, RRID:AB_2534022). Embryos were washed multiple times in PBST, incubated in DAB solution (KPL #71-00-48), and dehydrated before imaging in BBBA.

Transplantation and YSL injection

For transplantation experiments, donor embryos were injected with 50 pg of *cyc* mRNA and/or *vg1* mRNA and 50 pg of GFP mRNA and grown to sphere stage (4 hpf). At sphere stage, cells were transplanted from donor embryos to host embryos, and host embryos were grown to shield stage before fixation for *in situ* hybridization. For YSL injections, 1000-cell stage embryos were injected through the chorion into the YSL with approximately 100 pg of *vg1* mRNA and 500 pg of Alexa Fluor 488 dextran (ThermoFisher) .

Immunoblotting

Embryos were injected at the one-cell stage with 50 pg of each mRNA and grown to early gastrulation (50% epiboly). 8 embryos per sample were manually deyolked with forceps and frozen in liquid nitrogen. The samples were boiled at 95°C for 5 minutes with 2x SDS loading buffer (10 uL) and DTT (reducing gels only, 150 mM final concentration) then loaded onto Any kD protein gels (Bio-Rad). Samples were transferred to polyvinylidene fluoride (PVDF) membranes

(GE Healthcare). Membranes were blocked in 5% non-fat milk (Bio-Rad) in TBST and incubated overnight at 4°C in primary antibodies (1:5000 rabbit anti-GFP, ThermoFisher A11122, RRID:AB_221569; 1:2000 rabbit anti-Flag, Sigma F7425, RRID:AB_439687). Proteins were detected using HRP-coupled secondary antibody (1:15,000 goat anti-rabbit, Jackson ImmunoResearch Labs 111-035-144, RRID:AB_2307391). Chemiluminescence was detected using Amersham ECL reagent (GE Healthcare).

Co-immunoprecipitation

Dechorionated embryos were injected at the one-cell stage with 5, 20 or 50 pg of mRNA encoding epitope-tagged constructs and grown to 50% epiboly. 50-100 embryos were transferred to 400 uL of cold lysis buffer (50 mM Tris at pH 7.5, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 1% Triton X-100 and protease inhibitors, Sigma 11836170001) and crushed using a homogenizer and disposable pestles before incubation on ice for 30 minutes with vortexing every 5 minutes. Samples were spun at maximum speed at 4°C for 30 minutes and the supernatant was transferred to tubes containing 50 uL of anti-HA affinity matrix (Roche 11815016001, RRID:AB_390914) that was pre-washed twice in lysis buffer. Samples were placed on a rotating platform at 4°C overnight. The matrix was spun down for 2 minutes at 3000 rcf and washed in 600 uL of cold wash buffer (50 mM Tris at pH 7.5, 150 mM NaCl, 1% Triton X-100 and protease inhibitors) 5 times. 2x SDS loading buffer and DTT (150 mM final concentration) was added to the matrix in 10 uL of wash buffer. Immunoblots were performed as above.

qPCR

Embryos were injected at the one-cell stage with *sqt*, *cyc* and/or *vg1* mRNAs and grown to 50% epiboly. For the *sqt* experiment (Figure 5-2E) 2 sets of 10 embryos were collected per condition; for the *cyc* experiment (Figure 5-2F) 2 sets of 12 embryos were collected per condition. Embryos

were flash frozen in liquid nitrogen and RNA was extracted using an E.Z.N.A. Total RNA Kit (Omega) and reverse transcription was carried out using an iScript cDNA Synthesis Kit (Bio-Rad). qPCR reactions were run on a CFX96 machine (Bio-Rad) using iTaq Universal SYBR Green Supermix (Bio-Rad) and 0.25 uM of primers (see primer sequences below). Gene expression levels were calculated relative to a reference gene, *ef1a.* The mean and standard error of the mean was plotted for each condition. Two technical replicates in addition to biological replicates were used per condition. Both experiments were performed multiple times.

Primer sequences

vg1_genotype_F	CCTGTGTGTGTTCTTTGCTCTG
vg1_genotype_R	CTGTTTAAAGATTTTCCACATCTGTG
ef1a_qPCR_F	AGAAGGAAGCCGCTGAGATGG
ef1a_qPCR_R	TCCGTTCTTGGAGATACCAGCC
lft1_qPCR_F	GAGATGGCCAAGTGTGTCCA
lft1_qPCR_R	CTGCAGCACATTTCACGGTC

Modeling

In the 'primed model', Vg1 is already present in excess when Nodal production begins. This model describes the dynamics of Nodal monomers (*N*), Vg1 monomers (*V*) and Nodal-Vg1 dimers (*D*).

$$\frac{dN}{dt} = \lambda_N - \beta_N N - \lambda_D NV$$

$$\frac{dV}{dt} = -\beta_V V - \lambda_D NV$$

$$\frac{dD}{dt} = \lambda_D NV$$

Assumptions: constitutive production of *N* (rate λ_N), first-order degradation (component half-lives of $ln2/\beta_N$ and $ln2/\beta_V$, respectively) and bimolecular heterodimerization with rate $\lambda_D NV$. Vg1 is assumed to be maternally deposited, and is thus provided to the system via the initial conditions.

In the 'cold-start' model, Nodal monomers accumulate and dimerize after the onset of Nodal production. This model describes the dynamics of Nodal monomers (*N*), Vg1 monomers (*V*) and Nodal-Nodal dimers (*D*).

$$\frac{dN}{dt} = \lambda_N - \beta_N N - \lambda_D N^2$$

$$\frac{dD}{dt} = \lambda_D N^2$$

Assumptions: constitutive production of *N* (rate λ_N), first-order degradation (component half-life of $ln2/\beta_N$) and bimolecular homodimerization with rate $\lambda_D N^2$.

Simulation

The system begins with an excess of *V,* representing maternally deposited Vg1. All other component concentrations begin at 0, and Nodal monomer production is assumed to be off. Nodal monomer production begins at 4 hpf (the onset of zygotic transcription and translation) and concludes after mesendodermal patterning (6 hpf). The concentration of dimer (either Vg1-Nodal heterodimer or Nodal-Nodal homodimer) was plotted.

Parameter	Description	Value	Units
λ_N	Nodal synthesis rate	0, t < 4h, t > 6h $2x10^{-13}$, 4h < t < 6h (Muller et al., 2012)	$M\ s^{-1}$
β_N	Nodal degradation rate	$1.16x10^{-4}$ (Muller et al., 2012)	s^{-1}
β_V	Vg1 degradation rate	$1.16x10^{-4}$ (Muller et al., 2012)	s^{-1}
λ_D	Dimerization rate	$1x10^{6}$ (Gerstle and Fried, 1993; Kohler and Schepartz, 2001; Northrup and Erickson, 1992)	$M^{-1}\ s^{-1}$
V	Vg1 initial concentration	100	nM

Table 5-1. Quantification of Vg1-sfGFP localization in M*vg1* embryos co-injected with 20 pg of *vg1-sfGFP* mRNA and 0.5-20 pg of *cyc* or *sqt* mRNA.
See Appendix Figure 6A for examples of Vg1-sfGFP extracellular localization.

mRNA co-injected with 20 pg *vg1-sfGFP*		Intracellular	Extracellular puncta	Extracellular diffuse	Extracellular puncta + diffuse
	0.5	2	3	0	0
	1	5	5	0	0
	2	0	8	0	0
	5	0	7	0	0

cyc (pg)	10	0	6	0	0
	20	0	9	0	0
sqt (pg)	0.5	5	0	0	0
	1	4	4	1	0
	2	0	3	2	4
	5	0	6	0	0
	10	2	0	3	5
	20	0	2	5	3

Chapter 6: Conclusions and prospects

Through my graduate work we have uncovered that Vg1 is essential for mesendoderm patterning in zebrafish. We propose the following mechanism: (1) *vg1* is inherited ubiquitously from the mother; (2) Vg1 protein localizes to the ER where it remains in an inactive state; (3) After the initiation of zygotic transcription, Nodal is synthesized in the cells at the margin of the blastula-stage embryo; (4) Nodal enters the ER and binds to Vg1; (5) the Vg1-Nodal heterodimer is secreted from the cell and cleaved to form a mature, heterodimeric ligand; (6) Activation of receptors by Vg1-Nodal induces gene expression and mesendoderm formation (Montague and Schier, 2017) (Chapter 5). This work demonstrates that *vg1* has an endogenous function in development, and furthermore, is as essential as Nodal for the induction of mesendoderm.

Our findings shed light on many of the confusing experimental results from the last 30 years: most prominently, the apparent inability for endogenous Vg1 to be secreted, cleaved or induce mesendoderm.

Historic problems solved

After discovering the exquisite localization of *vg1*, predictions were made that Vg1 is an endogenous inducer of mesoderm (Rebagliati et al., 1985; Weeks and Melton, 1987). However, although Vg1 was predicted to belong to the TGF-β family of growth and differentiation factors, Vg1 cleavage and secretion were not detected (Dale et al., 1989; Tannahill and Melton, 1989). Our data provides an explanation for these results. Vg1 is trapped in the ER in the absence of

Nodal, where it is neither cleaved nor secreted. Thus, it is likely that only the small fraction of endogenous Vg1 that encounters Nodal is cleaved and secreted in the embryo. This is consistent with previous arguments (Dale et al., 1989; Tannahill and Melton, 1989). However, while our experiments revealed robust cleavage of Vg1 when it was co-expressed with Nodal, we were also unable to detect the endogenous cleavage of Vg1. The development of more sensitive biochemical or imaging methods should allow the visualization of endogenous Vg1 cleavage. Overexpression of Vg1 also failed to produce cleaved Vg1, probably because the embryo has only a limited supply of endogenous Nodal to facilitate cleavage. Secretion of Vg1 by oocytes likely failed because *Nodal* is not expressed in oocytes. Our model does not explain why expression of zebrafish *vg1* in *Xenopus* embryos and oocytes resulted in cleavage and activity (Dohrmann et al., 1996). It is possible that zVg1 was able to heterodimerize with a different *Xenopus* protein in the embryo, or a proprotein convertase that recognizes zVg1 and not Vg1 was present in the oocyte.

The strongest evidence against Vg1 being a mesoderm inducer came from over-expression studies: injection of *vg1* into *Xenopus* or zebrafish embryos resulted in no mesoderm induction or phenotype (Dale et al., 1993; Dohrmann et al., 1993; Thomsen and Melton, 1993). In hindsight this makes sense: if embryos already have an excess of Vg1 that is dormant in the ER, any additional Vg1 will only add to that inert pool. To achieve ectopic mesoderm induction, co-expression of *vg1* and *Nodal* would have been necessary. Chimeric proteins were subsequently developed with the prodomains of BMP2, BMP4 or Activin fused to the mature domain of Vg1 (Dale et al., 1993; Kessler and Melton, 1995; Thomsen and Melton, 1993). These constructs solved the problem that *vg1* injections were ineffectual – now, mature Vg1 could be generated, it was secreted, and it appeared to induce mesoderm (Dale et al., 1993; Kessler and Melton, 1995; Thomsen and Melton, 1993). Given that the prodomains of TGF-β family members can

alter the folding, disulfide bond formation, export, stability and signaling range of dimers (Degnin et al., 2004; Le Good et al., 2005; Massagué, 1987) as well as affect a monomer's ability to form heterodimers (Neugebauer et al., 2015), it is quite possible that the heterologous prodomains conferred ectopic function on Vg1. Nonetheless, in light of our work, it is most likely that the fused prodomains permitted the formation of active Vg1 homodimers. This is in contrast to the endogenous context, where Vg1 appears to primarily exist in a monomeric state (Dale et al., 1993) (Chapter 5).

Our data indicates that Vg1 remains in the ER in the absence of Nodal (Montague and Schier, 2017) (Chapter 5). This is consistent with the observation that the Vg1 signal sequence is not cleaved upon entry into the ER (Dale et al., 1993), unlike other members of the TGF-β family (Massagué, 1987). So, what retains Vg1 in the ER? One possibility is that Vg1 is bound by an ER-resident protein, which is only displaced by the arrival of Nodal. Alternatively, Vg1's hydrophobic signal peptide could maintain Vg1 an incorrectly folded state until Nodal heterodimerization allows correct folding, peptide cleavage and ER exit. Either way, retention in the ER is likely a function of the Vg1 prodomain: thus, replacing it with a BMP or Activin prodomain facilitates secretion and cleavage.

Experiments using a truncated version of the Activin type II receptor inhibited both mesoderm formation (Hemmati-Brivanlou and Melton, 1992) and ectopic BMP-Vg1 activity (Schulte-Merker et al., 1994), suggesting that Vg1 activates the type II Activin receptor. However, receptor binding assays revealed that mature Vg1 does not directly bind the type II Activin receptor, suggesting Vg1 might activate the receptors through an indirect mechanism (Kessler and Melton, 1995). This result is consistent with our model: a Vg1/Nodal heterodimer could assemble a tetramer of two type I and type II Activin receptors; Nodal might bind the type II receptors and Vg1 might bind

the type I receptors, akin to the BMP2-7 heterodimer system (Little and Mullins, 2009). If this is the case, Vg1 homodimers produced from BMP- or Activin-Vg1 fusion proteins may assemble and signal via a different array of receptors, or they may more weakly bind the type I/type II heteromeric complex than a Nodal/Vg1 heterodimer. Since all BMP-Vg1 experiments were performed under overexpression conditions, even a relatively weaker activation of the receptors could result in robust gene expression. Indeed, we find that Vg1/Nodal heterodimers are more potent than Nodal homodimers (Montague and Schier, 2017) (Chapter 5).

A Vg1/Nodal heterodimer in *Xenopus?*

The current view of *Xenopus* mesoderm development is as follows: *VegT*, a maternal mRNA encoding a transcription factor, is localized to the vegetal cortex of the egg, and upon fertilization, diffuses through the embryo. Cleavage events passing through the equator of the embryo restrict VegT's localization to the vegetal half of the embryo. At the 256-cell stage, β-catenin initiates expression of *Xnr5* and 6 on the dorsal side (Yang et al., 2002), and after zygotic transcription begins at the 4000-cell stage, VegT activates expression of the *Xnr1, 2 and 4* genes in the vegetal hemisphere. The *Xnr* genes activate multiple target genes to initiate mesoderm formation (Clements et al., 1999; Hyde and Old, 2000; Kofron et al., 1999; Xanthos et al., 2001; Zhang et al., 1998a).

Might Vg1 form obligate heterodimers with the Xnr proteins? Like *VegT*, *Vg1* is localized to the vegetal cortex of the egg, and then diffuses through the vegetal hemisphere after fertilization (Thomas and Moos, 2007). Thus, the embryonic domains of Vg1 and the Xnr proteins must largely overlap. If Vg1 and Xnr1-5 form obligate heterodimers, expression of either mRNA alone

should fail to activate gene expression. Injection of *Xnr1* into *Xenopus* animal caps results in ectopic mesoderm formation (Agius et al., 2000), suggesting that Xnr1 can act independently of Vg1. However, quantification of *vg1* expression in different regions of the embryo suggested there is some *vg1* present in the animal cap (Agius et al., 2000), which could be enough to permit heterodimer formation. Alternatively, similar to zebrafish Nodal, Xnr1 may form active homodimers at artificially high concentrations (Chapter 5). In support of Xnr/Vg1 heterodimers, depletion of *Vg1* and *Xnr5* in combination reduces pSmad2 signaling more than either mRNA alone (Tadjuidje et al., 2016).

If Xnr and Vg1 form obligate heterodimers, why does *vg1* localize to the vegetal cortex? The elegance of the zebrafish Vg1/Nodal patterning system is that even if Vg1 is ubiquitous, it cannot induce ectopic signaling since it remains in an inactive state in the ER in the absence of Nodal. *Xenopus Vg1* might not need its precise localization, or perhaps it presents an additional layer of regulation, restricting the generation of active Nodal/Vg1 signals to vegetal regions.

Vg1 in left-right asymmetry

One of the mammalian Vg1 orthologs, GDF1, forms heterodimers with Nodal to pattern the left-right axis (Tanaka et al., 2007). A morpholino-based study in zebrafish suggested a similar model for *vg1* in zebrafish left-right axis patterning (Peterson et al., 2013). Our results contradict this study and show instead that zebrafish Vg1 has an early role in mesendoderm patterning (Chapter 5). However, while our zygotic *vg1* mutants do not exhibit a phenotype, it's possible a left-right patterning phenotype is masked. There are two periods of *vg1* expression: maternal *vg1* mRNA is present until mid-late gastrulation, and zygotic *vg1* transcription resumes 3-4 hours later

(Montague and Schier, 2017) (Appendix Figure 8). Although there is a 3 to 4-hour period when *vg1* isn't expressed, there is some data to suggest that maternal Vg1 protein remains in the embryo. *Xenopus* Vg1 protein measurements suggest it has high stability (Dale et al., 1989), and uncleaved mouse Nodal appears to be stable, while cleaving it causes rapid internalization and degradation (Le Good et al., 2005). Since the majority of Vg1 in the embryo is never cleaved, it could be stable for hours. Thus, even in a zygotic *vg1* mutant, there could be Vg1 protein residing in the ER of every cell, which could form heterodimers with Southpaw[9]. To address this issue, a rapidly degrading version of *vg1* could be injected into MZ*vg1* embryos to facilitate rescue of mesendoderm but not left-right patterning.

Future directions

Our work has solved some of the pressing Vg1 mysteries but leaves other questions open. For instance, we present a convincing mechanism for the regulation of Vg1 processing and secretion, but we remain unable to detect the cleavage and secretion of endogenous Vg1. It will be important to develop methods to detect endogenous proteins in zebrafish, for instance with new antibody strategies and by fluorescently tagging endogenous loci.

The majority of work interrogating the function of Vg1 was carried out in *Xenopus* where, until recently, generating mutants was not feasible. With the development of CRISPR/Cas9 methods in *Xenopus* (Wang et al., 2015a), it will be interesting to find out if Vg1 plays a similar role in the

[9] In a paper co-published with ours, the authors claim that Vg1 is required for left-right asymmetry by presenting phenotypic evidence after injecting multiple MOs into the *vg1* mutants. Their conclusion may be correct, but the experimental design is flawed and inconsistent with their model (Pelliccia et al., 2017).

frog as it does in the zebrafish. All 5 Xnr proteins might require heterodimerization with Vg1 for activity; alternatively, other TGF-β family members such as Derrière (Eimon and Harland, 2002) might replace the function of Vg1.

This work, and the work of others, suggests that heterodimer-mediated signaling can confer a number of advantages on a system, for instance higher potency, temporal and spatial control of signaling, faster signaling dynamics and modulation of signaling range. Future work might reveal that heterodimers are more widespread in signaling pathways than previously anticipated.

Appendix

Appendix 1: Reverse transcription, semi-quantitative PCR and DNA extraction on board the International Space Station

Preface

This chapter is currently under review. The project initiated from the United Arab Emirates Genes in Space competition. I mentored 14-year-old Alia Almansoori in preparation for the competition, which she subsequently won. Her prize was to send her experiment to space. Alia Almansoori conceived the initial idea for this project and developed the experiment with input and supervision from me. I designed the experiment that was sent to space, with input from Alia Almansoori, Ezequiel Alvarez Saavedra and Emily Gleason. I performed the on-Earth experiments and prepared the samples that were sent to space with Ezequiel Alvarez Saavedra. Astronaut Peggy Whitson performed the experiments in space. I performed the final experiments on Earth upon return of the samples from space and analyzed the data with Alia Almansoori and Ezequiel Alvarez Saavedra.

Abstract

The distance and duration of human spaceflight missions is set to markedly increase over the coming decade as we prepare to send astronauts to Mars. However, the health impact of long-term exposure to cosmic radiation and microgravity is just beginning to be understood. In order to identify the molecular mechanisms underpinning the effects of space travel on human health, we must develop the capacity to monitor changes in gene expression and DNA integrity in space. Here we report successful implementation of three molecular biology procedures on board the International Space Station (ISS) using a miniaturized thermal cycler system and *C. elegans* as a model organism: first, DNA extraction, which is the initial step for any type of DNA analysis;

second, reverse transcription of RNA to generate complementary DNA (cDNA); and third, the subsequent semi-quantitative PCR amplification of cDNA to analyze gene expression changes in space. These molecular procedures represent a significant expansion of the budding molecular biology capabilities of the ISS, and will permit more complex analyses of space-induced genetic changes during spaceflight missions aboard the ISS and beyond.

Introduction

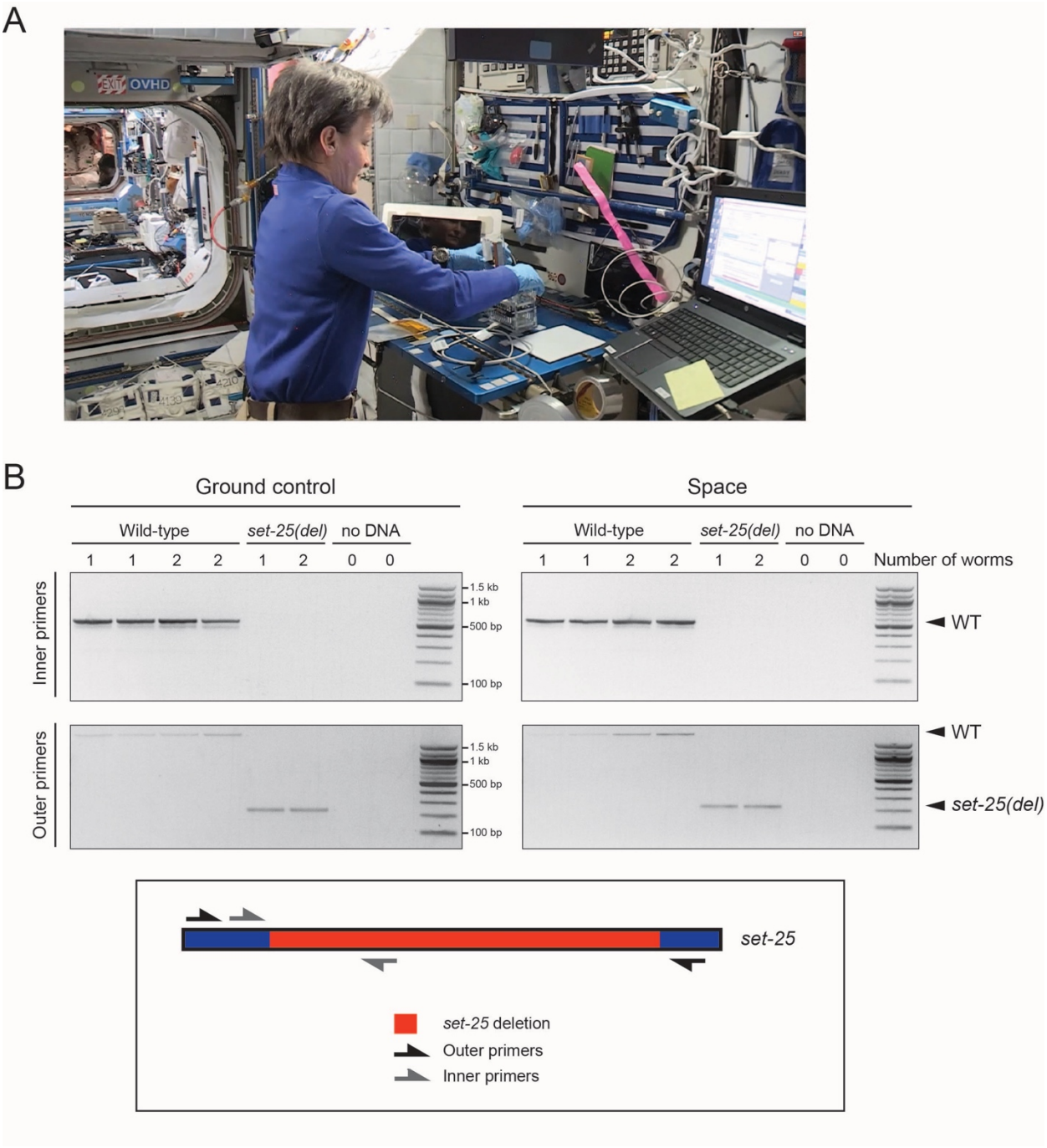
Human spaceflight is about to undergo a major shift. While the last few decades have seen short- or medium-term (<1 year) orbital flights within Earth's protective magnetic field, in the coming years astronauts will transition to longer-term explorations of deep space. There are many challenges associated with deep space missions, but the predominant health risks that must be mitigated before humans can safely undergo interplanetary travel are the consequences of exposure to cosmic radiation and microgravity. The high-energy protons and high-charge and high-energy nuclei that comprise cosmic rays can have devastating effects on human health. One of the most vulnerable sites for cosmic radiation damage is DNA, where mutations can lead to the development of cancer (Durante and Cucinotta, 2008; Moreno-Villanueva et al., 2017). In addition, the combination of microgravity and cosmic radiation can negatively impact many normal biological processes in the skeletal, immune and nervous systems (White and Averner, 2001). In many cases, alterations in gene expression precede and accompany the physiological changes (Barrila et al., 2016; Crucian et al., 2015; Cucinotta et al., 2014; Semov, 2002; Terada et al., 2016), providing a means to predict and understand the nature of the damage. Thus, during long-term space missions, detecting molecular changes in the DNA and RNA of astronauts will be useful to monitor their health and inform possible treatments.

Recent developments in molecular biology tools on the ISS have resulted in the first reports of DNA amplification and sequencing in space (Boguraev et al., 2017; McIntyre et al., 2016). However, a larger repertoire of nucleic acid techniques will be needed to render the ISS capable of complete molecular biology procedures. In this Genes in Space-4 study we tested three of the major techniques used on Earth to monitor gene expression and detect DNA sequence changes on board the ISS. We used the nematode worm *Caenorhabditis elegans* (*C. elegans)* as a model organism, which has already been successfully flown to space for scientific study (Szewczyk et al., 2008). We performed a DNA extraction experiment, reverse transcribed RNA into cDNA, and performed semi-quantitative PCR on board the ISS. These experiments provide a foundation for more complex experimental procedures to monitor molecular changes to human health in space.

Results

DNA extraction from *C. elegans* on board the ISS

To detect changes in an organism's DNA, including mutations induced by cosmic radiation, it is first necessary to extract DNA from biological tissues. To test the feasibility of DNA extraction in space we used *C. elegans* from two different genotypes: wild-type and a *set-25(n5021)* deletion strain (Andersen and Horvitz, 2007). Individual worms were mixed with lysis buffer on Earth then frozen, and on board the ISS the samples were incubated in a miniPCR thermal cycler (Appendix Figure 1A) at 60°C for 60 minutes followed by 95°C for 15 minutes to inactivate the proteinase K present in the buffer. The samples were refrozen and sent back to Earth, where the genomic DNA was tested as a template for amplification using the polymerase chain reaction (PCR). Two sets of primers were used: one flanking the *set-25(n5021)* deletion and one with a primer internal to the deletion. In combination, these amplification reactions permit the identification of the molecular



Appendix Figure 1. DNA extraction from *C. elegans* on board the ISS.
(A) The miniPCR system was operated on the International Space Station (ISS) by astronaut Peggy Whitson.
(B) Wild-type and *set-25(del) C. elegans* were digested in lysis buffer and proteinase K on board the ISS in a miniPCR thermal cycler. The resulting DNA was used as input in a PCR amplification reaction on Earth that tested for the presence of the deletion found in the *set-25* mutants. The *set-25* inner primers amplify a 573 bp product with wild-type DNA, but don't produce a product with *set-25* mutant DNA due to the deletion. The *set-25* outer primers amplify different sized products depending on the genotype: wild-type DNA produces a 2220 bp band, whereas *set-25* mutant DNA produces a 242 bp band.

lesion in the *set-25* mutants. Amplification from the extracted genomic DNA was successful, and the molecular lesion was identifiable in both the ground controls and space samples (Appendix Figure 1B), indicating that DNA can be successfully extracted in space with the same protocol as on Earth. Furthermore, our experiments reveal that DNA extracted on board the ISS can be frozen, stowed, and later utilized in molecular analyses back on Earth. Given the small volume used for DNA extraction, in future experiments, PCR reagents could be added directly to the DNA extraction reaction tubes to efficiently perform the complete DNA extraction and PCR in space.

Semi-quantitative PCR in space

Due to limitations on the collection of live biological samples on board the ISS, we simulated space-induced changes in gene expression using the heat shock system of proteins in *C. elegans.* The heat shock proteins are part of a stress response system that is highly conserved from bacteria to humans (Schlesinger, 1990). Stresses such as heat or microgravity (Cubano and Lewis, 2001; Zupanska et al., 2013) cause up-regulation of the heat shock genes and synthesis of the heat shock proteins, which act as molecular chaperones to aid the correct folding of proteins (Schlesinger, 1990). We simulated spaceflight-induced expression of the *hsp-70* gene by subjecting *C. elegans* to a 30-minute heat shock on Earth, followed by purification of RNA from heat shock-treated and control animals after a 30-minute recovery period (Appendix Figure 2A).

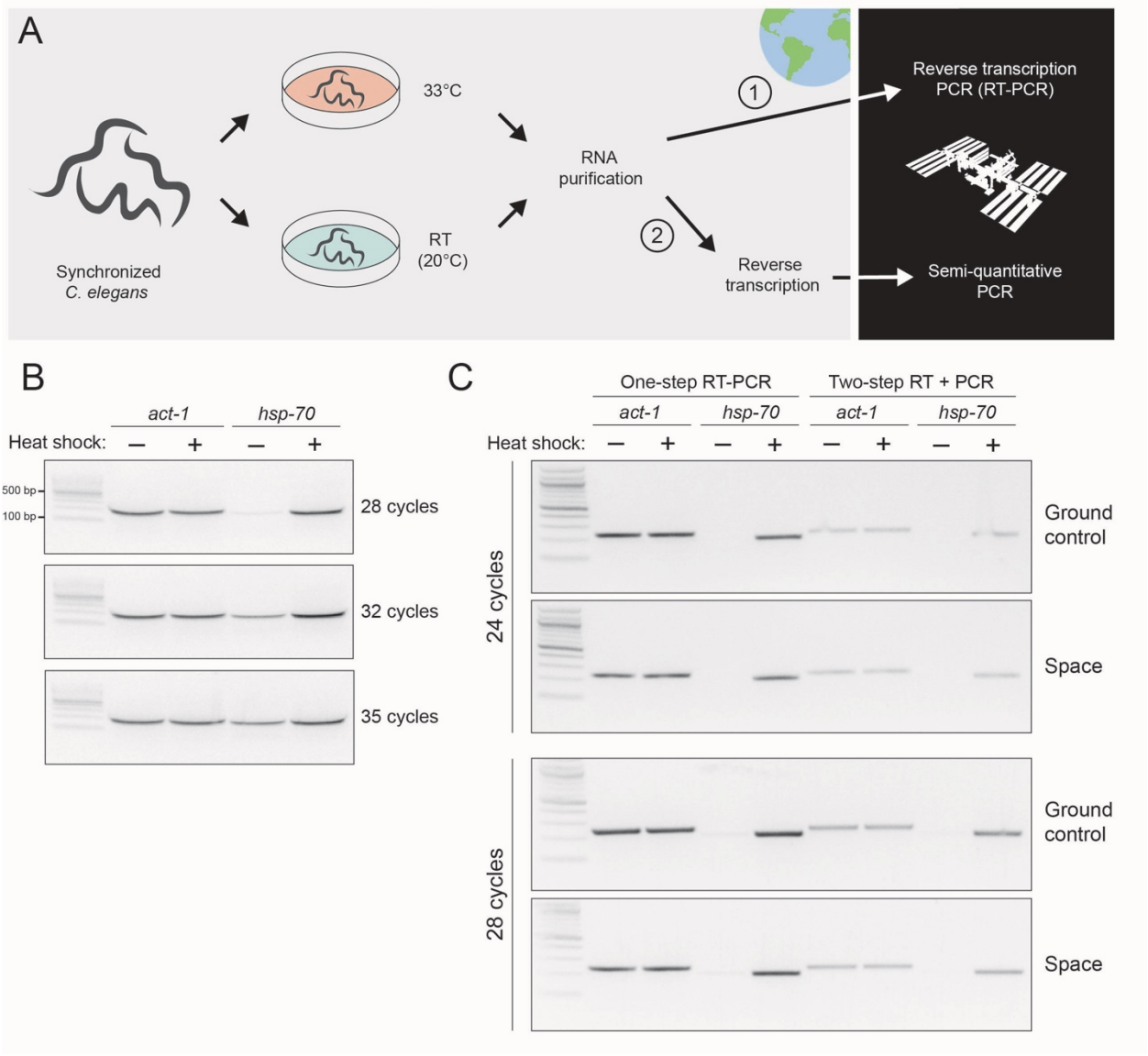
To measure gene expression, we reverse transcribed RNA to generate cDNA, and then performed semi-quantitative DNA amplification. Semi-quantitative PCR uses a cycle number that lies within the exponential phase of DNA amplification so that the amount of product synthesized between different conditions is proportional to the amount of starting template. Primers are selected that target the gene of interest as well as a housekeeping (control) gene

that is not expected to change between experimental conditions. We used primers targeting the heat shock gene *hsp-70* and the housekeeping gene *actin (act-1).* We converted total RNA to cDNA on Earth, then set up reaction tubes containing cDNA and the PCR reagents and sent them frozen to the ISS. Astronaut Peggy Whitson performed the semi-quantitative PCR in space following a protocol that was optimized on ground samples. The samples were then refrozen and returned to ground where we analyzed the results by gel electrophoresis. During optimization of the procedure, we determined the range of PCR cycles within which *hsp-70* amplification remained in the exponential phase by running parallel PCR experiments with different cycle numbers (Appendix Figure 2B). This allowed for maximum distinction between the experimental conditions tested.

In space, the *act-1* product was amplified to equivalent levels under heat shock and no-heat shock conditions, similar to the ground controls (Appendix Figure 2C). The *hsp-70* product was undetectable, or almost not detectable, in the no-heat shock condition, indicating that basal expression of *hsp-70* is low under normal physiological conditions. By contrast, under heat shock conditions, the *hsp-70* PCR product was amplified to high levels, indicating robust induction of *hsp-70* gene expression (Appendix Figure 2C). The results were indistinguishable between the ground controls and the ISS samples. Thus, semi-quantitative PCR can be carried out in space to measure changes in gene expression, such as up-regulation of *hsp-70* triggered by stress conditions.

One-step reverse transcription and semi-quantitative PCR in space

For molecular assays to be performed successfully by astronauts with limited resources and potentially limited laboratory training, the workflow of sample collection, RNA extraction, reverse transcription and quantitative or semi-quantitative PCR should be as streamlined as possible. As a step towards a simplified procedure, we tested if one-step reverse transcription and semi-



Appendix Figure 2. One-step reverse transcription and semi-quantitative DNA in space.
(A) *C. elegans* were heat shocked for 30 minutes at 33°C to simulate space-induced changes in the expression of heat shock gene *hsp-70*. Control worms were maintained at 20°C. After a 30-minute recovery period, RNA was purified from both populations of worms, and the RNA was subjected to DNase digestion. One set of samples was combined with reagents for one-step reverse transcription-PCR (RT-PCR) and frozen (1), while a second set of RNA samples was used as the input for a reverse transcription experiment on Earth, and the resulting cDNA was combined with PCR reagents and frozen (2). Half of each sample set was sent to the ISS while the other half remained on Earth for ground controls. Semi-quantitative PCR reactions were run on Earth and in space. All samples were analyzed on Earth using gel electrophoresis.
(B) To determine the range of exponential DNA amplification of the *hsp-70* gene amplicon during PCR, cDNA was amplified using 28, 32 or 35 cycles of PCR.
(C) One-step reverse transcription-PCR (RT-PCR) and two-step reverse transcription and then PCR were run on earth and in space using 24 and 28 cycles of amplification. cDNA was amplified using primers targeting *hsp-70* to semi-quantitatively determine the amount of *hsp-70* gene expression in samples subjected to heat shock or no-heat shock. Amplification of *actin* (*act-1*) was used as a control.

quantitative PCR in the same reaction tube, and without astronaut intervention, was feasible in space. We combined the RNA extracted on Earth with a complete enzyme mix, containing all of the reagents for reverse transcription and PCR in one tube. The samples were frozen and sent to space where a combined reverse transcription-PCR (RT-PCR) protocol was performed on a miniPCR thermal cycler (Appendix Figure 2A). The RT-PCR successfully amplified the *act-1* and *hsp-70* amplicons, producing similar results as the reaction on Earth and as the two-step reverse transcription and PCR (Appendix Figure 2C). Thus, using RNA as starting material, gene expression can be assessed with minimal astronaut intervention.

Discussion

In this study we describe the successful execution of DNA extraction, reverse transcription and semi-quantitative PCR in space. Current ISS-based experiments generally rely on in-space sample collection followed by on-Earth processing and analysis. Our experiments reveal that three additional molecular techniques can be performed in space, thus expanding the molecular capabilities of the ISS. Soon, astronauts will be able to generate and analyze data on their health and the molecular status of the living
environment entirely in space.

A round-trip mission to Mars is expected to take approximately 3 years (Cucinotta et al., 2013). The astronauts embarking on this journey will need to be entirely self-sufficient to survive the hostile environment of deep space. Not only will they be solely responsible for their own health, but upon exiting Earth's magnetosphere, the astronauts will be subject to approximately three times the cosmic radiation as spacecraft within Earth's orbit (Cucinotta et al., 2013). In order to closely monitor their health, crewmembers will need to track not only their physiology, but also

their DNA integrity. By successfully demonstrating DNA extraction in space we are one step closer to astronauts acquiring their own biological samples, extracting the DNA and sequencing it, for example using the MinION nanopore sequencer that has been successfully operated in microgravity (McIntyre et al., 2016). In addition, if a crewmember acquires an infection, astronauts could potentially use DNA analysis to identify the organism and any genetic changes it has undergone to inform their decisions for treatment.

The human body has been shown to undergo gene expression changes in space, which could reveal underlying health issues or changes to physiology, such as cellular stress or inflammation. By demonstrating successful reverse transcription of RNA and semi-quantitative PCR in space, we provide evidence that two of the critical steps of gene expression analysis can be carried out in space. Future developments should take us closer to the ultimate goal of allowing astronauts to carry out full genetic and gene expression analyses in space using human samples.

Methods

Strains

C. elegans were maintained at 20°C. N2 Bristol was used as wild-type. MT17463 [*set-25(n5021)*] animals (Andersen and Horvitz, 2007) contain a 1978 bp deletion in the *set-25* gene. *set-25* mutants were genotyped using two primer pairs. The outer primers (see Primer Sequences) produce a 2220 bp band with wild-type DNA, and a 242 bp band with *set-25* DNA; the inner primers produce a 573 bp band with wild-type DNA, and no PCR product with mutant DNA.

DNA extraction

Either 1 or 2 animals were placed in tubes with 5 µL of lysis buffer (50 mM KCl,10 mM Tris pH 8.3, 2.5 mM $MgCl_2$, 0.45% IGEPAL (Sigma, I8896), 0.45% Tween-20 and 0.01% Gelatin) supplemented with 100 µg/ml proteinase K and the tubes were frozen at -80°C. The samples remained frozen until operations on board the ISS. The lysis protocol on the miniPCR machine was: 60°C/60 min, 95°C/15 min. However, due to input error, the heat inactivation step in space was carried out for 1.5 minutes instead of 15 minutes. Thus, upon their return to Earth, the samples were subjected to the remaining 13.5-minute heat inactivation. On Earth, 1 µL of the extracted DNA was placed in tubes with OneTaq Hot Start Taq kit (NEB, M0484S) and 0.4 µM of the *set-25* primers for DNA amplification using the following protocol: 95°C/2 min, [95°C/30 sec, 55°C/30 sec, 72°C/3 min] x35, 72°C/5 min. The completed PCR reactions were run on a 2% agarose gel.

Heat shock

Worms were synchronized and grown to 4 hours post-L4 stage on NGM agar plates and submerged in a 33°C water bath for 30 minutes. The control “no heat shock” plates were maintained at room temperature. The heat shocked plates were allowed to recover for 30 minutes at 20°C before washing the worms off the plates, and freezing them in liquid nitrogen.

RNA extraction and DNase digestion

RNA was extracted from the worms using TRIzol (ThermoFisher, 15596026). Briefly, frozen worms were resuspended in TRIzol, flash frozen in liquid nitrogen, then thawed at 37°C and vortexed for 1 minute. The freeze-thaw-vortex cycle was repeated twice, and the samples were combined with chloroform and centrifuged at 4°C for 15 minutes before the aqueous layer was removed and combined with isopropanol. The RNA was pelleted by spinning at 4°C for 15 minutes, the pellets were washed in ethanol and then resuspended in RNase-free water and combined with Turbo DNase and buffer (ThermoFisher, AM2239). DNase digestion was carried

out at 37°C for 30 minutes and then an additional 1 µL of Turbo DNase was added to the reactions and they were incubated for a further 30 minutes. RNA was purified from the samples by ethanol precipitation.

cDNA synthesis

For the two-step reverse transcription and PCR, 1 µg of RNA from each condition was combined with 1 µL of iScript reverse transcriptase (Bio-Rad, 1708890) and 4 µL of iScript buffer in a total reaction volume of 20 µL. The cDNA synthesis program was as follows: 25°C/5 min, 46°C/20 min, 95°C/1 min.

Semi-quantitative PCR

For the combined one-step RT-PCR, 100 ng of RNA from each condition was added to a tube and combined with OneTaq One-Step RT-PCR mix (NEB, E5315S) and 0.4 µM of each primer (see Primer Sequences). The reactions were frozen at -80°C and maintained in a frozen state until operations on board the ISS. The RT-PCR protocol was as follows: 48°C/20 min, [94°C/20 sec, 56°C/30 sec, 68°C/60 sec] x24/28, 68°C/10 sec. The two-step semi-quantitative PCR combined 1 µL of cDNA as the template, OneTaq Hot Start master mix (NEB, M0484S) and 0.4 µM of each primer. The samples were frozen at -80°C and remained frozen until operations on board the ISS. The PCR protocol was identical for the one-step and two-step protocols. Upon returning to earth, the samples were run on a 2% agarose gel for analysis.

Thermal cycler system

All experiments were carried out in a mini8 miniPCR thermal cycler (www.minipcr.com).

ISS operations

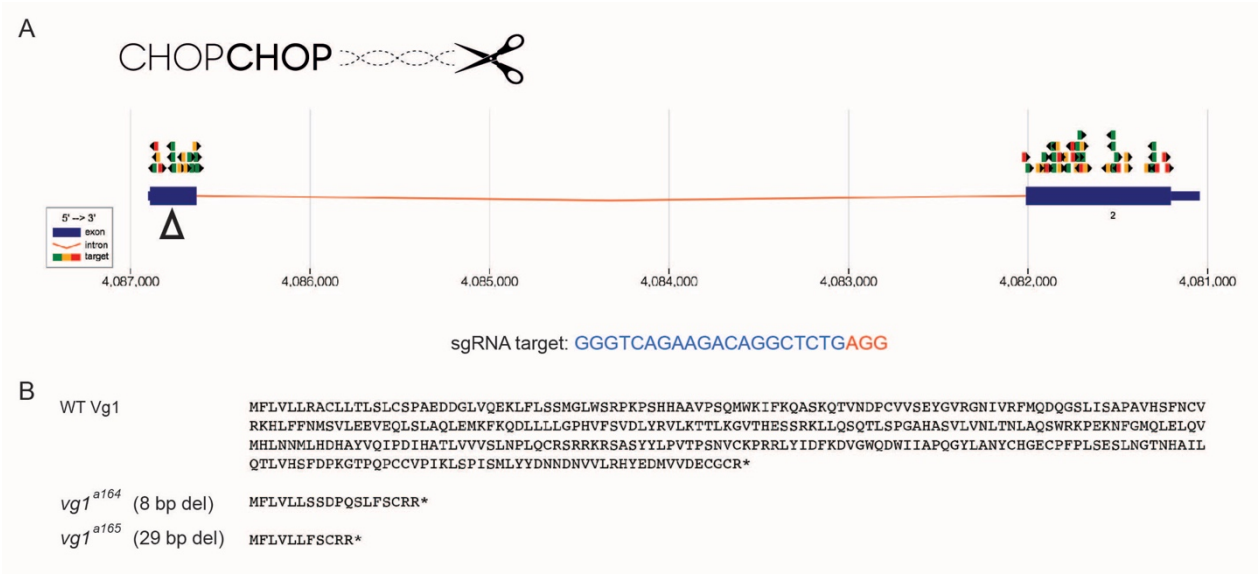
The Genes in Space samples were prepared on Earth, frozen, and sent on dry ice to the Kennedy Space Center, where they arrived 6 days before the launch. The samples were stored in a POLAR freezer until 24 hours prior to launch when it was moved to the SpaceX Dragon vehicle, kept at -95°C, and launched on a SpaceX Falcon rocket to the ISS. Upon docking with the ISS, the samples were transferred to the GLACIER -95°C freezer on orbit until further processing. During experimental operations, ISS crew removed the samples from the freezer, allowed them to thaw, and took preliminary photos to show the sample strip conditions prior to being removed from the flight bag. A visual inspection of the samples strips was also performed at this time to verify all sample tube caps were sealed and that there was no visible leakage from the strip. The MWA (maintenance work area) was prepped by setting up the miniPCR, ISS Station Support Computer (SSC) laptop, inverter (to change power from 120VDC to 120VAC), along with all cables being connected. The sample and thermal cycling protocols were loaded onto the miniPCR machine. At this point the sample run was started and the astronaut allowed miniPCR operations to complete while working on other tasks. The ground team monitored the miniPCR process via live video stream. Upon completion of the miniPCR run, the samples were allowed to cool for a minimum of 30 minutes before being removed from the miniPCR. Photos were taken again to show the strip configuration prior to being placed back into the flight bag in the MELFI or GLACIER freezer, where they remained until loaded into the SpaceX Dragon inside a POLAR freezer set to -95°C. Upon splashing down to Earth in the Pacific Ocean off the coast of California, the samples were unloaded from the Dragon vehicle and placed in a freezer to maintain temperature on a fast boat that took them to land. The samples were then placed inside a freezer for return to the Johnson Space Center, where the Boeing team took delivery and packaged for return to Harvard University for analysis.

Primer sequences

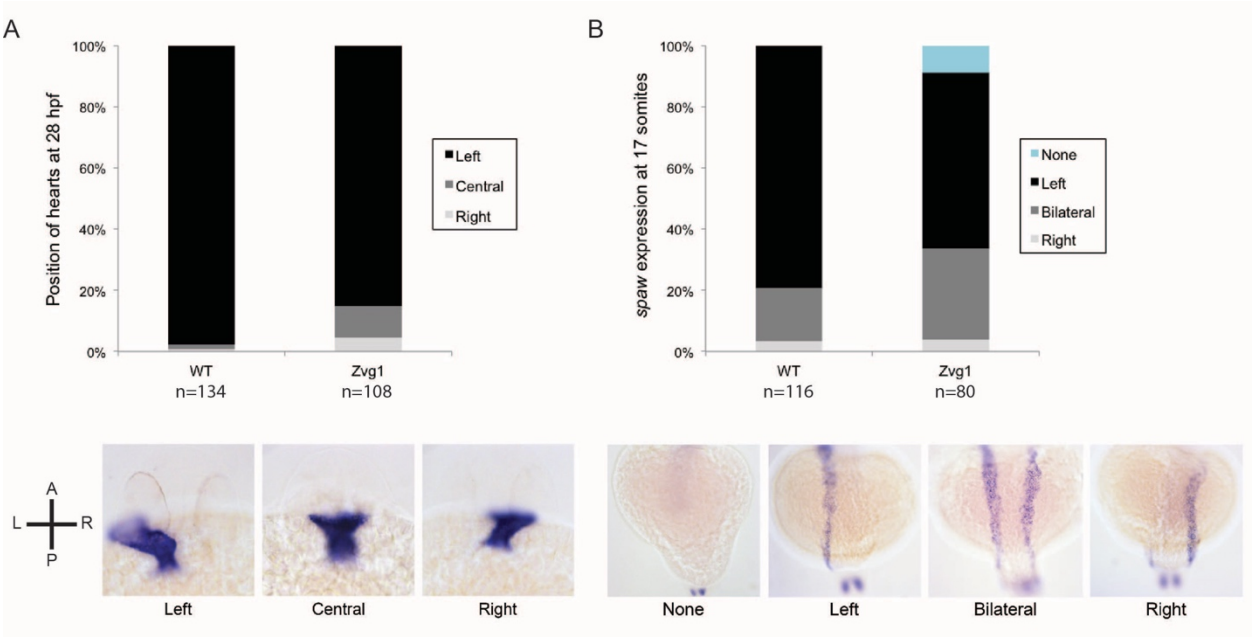
All semi-quantitative PCR primers were designed with one primer across a splice junction to avoid amplification of contaminating genomic DNA.

Set-25_outer_F	TTCGTTGGTTTTTCGGACA
Set-25_outer_R	CCGTGCTACGCGGTAAGTAT
Set-25_inner_F	AGACTTCGACGAACACCGAG
Set-25_inner_R	CGCGTGGAAAAGTTTCGTGT
Act-1_F (Gaiser et al., 2011)	AATCCAAGAGAGGTATCCTTA
Act-1_R (Gaiser et al., 2011)	GATGGCGACATACATGGCT
Hsp-70_F	GAAAATCACACGTGCAAGATTCG
Hsp-70_R	GAGCAGTTGAGGTCCTTCCC

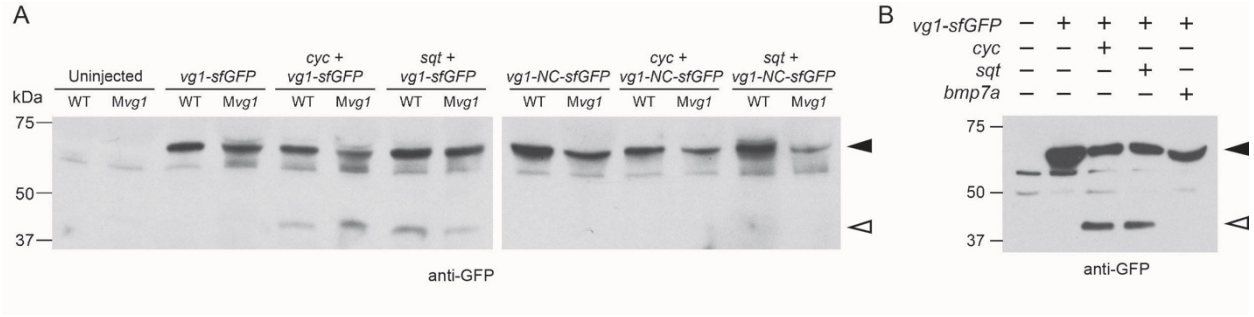
Appendix 2: Supplementary figures



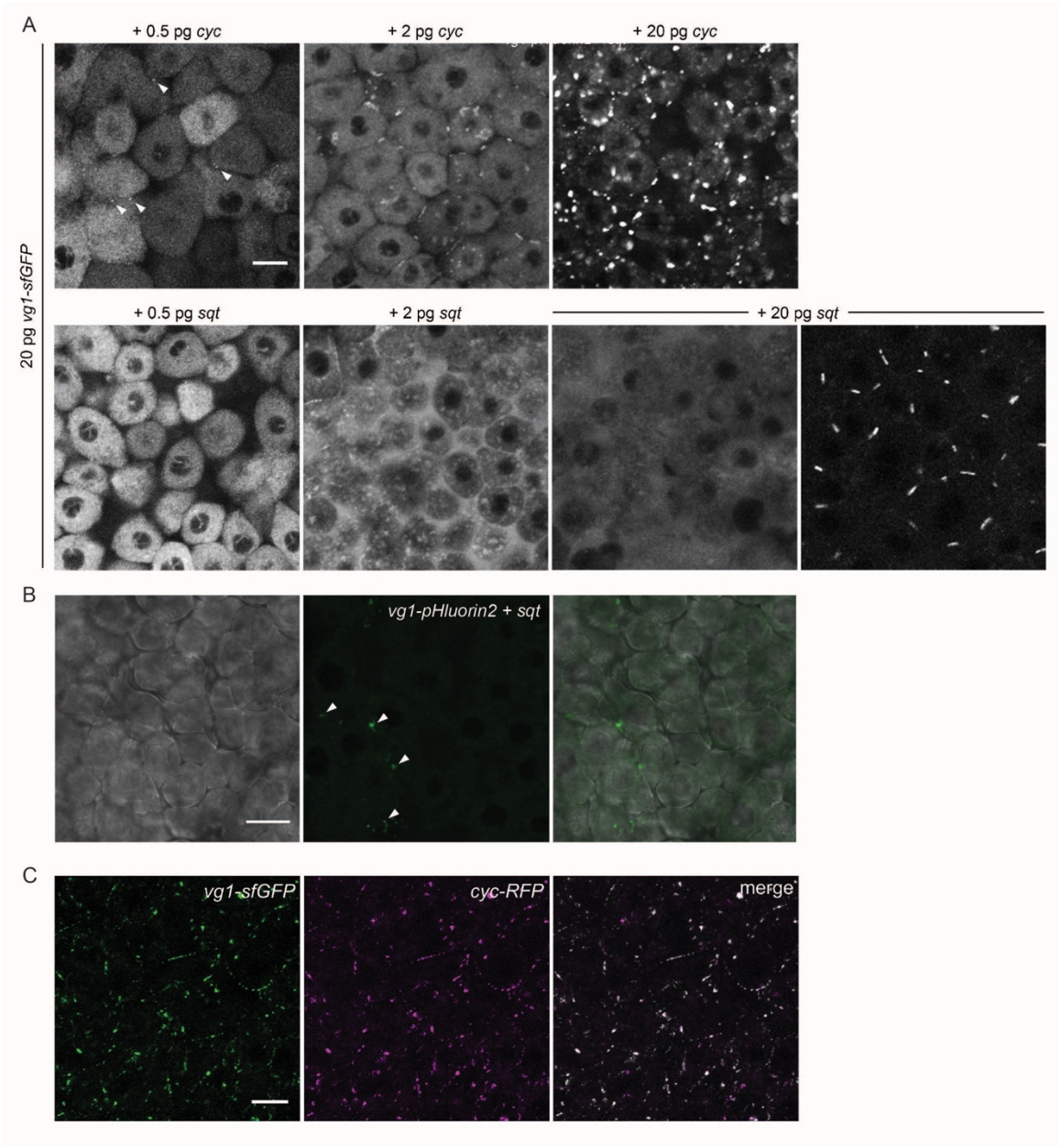
Appendix Figure 3. *vg1* mutant alleles.
(A) sgRNAs targeting the *vg1/dvr1/gdf3* gene were designed using CHOPCHOP (Labun et al., 2016; Montague et al., 2014). Colored arrows indicate possible sgRNA targets, color-coded according to predicted efficiency (green: good to red: bad). Open arrowhead indicates the location of the sgRNA used in this study. The sgRNA sequence is provided in blue, with the PAM motif in red.
(B) 8 bp (*vg1*^{a164}) and 29 bp (*vg1*^{a165}) deletion alleles cause frameshifts in Vg1, truncating it from a 355 amino acid protein to predicted 18 and 11 amino acid peptides, respectively.



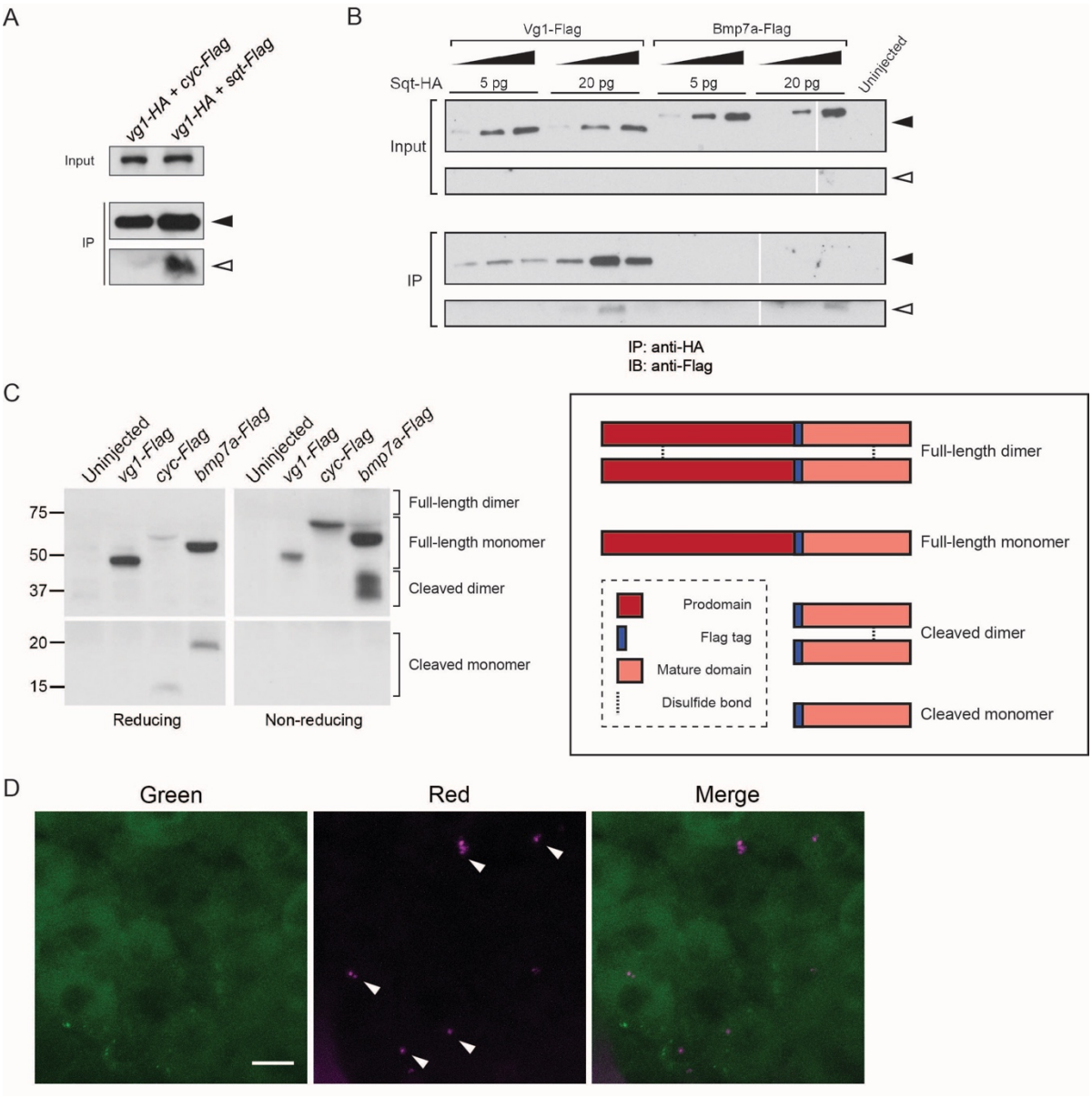
Appendix Figure 4. Left-right asymmetry in *Zvg1* mutants.
(A) Heart positioning at 28 hours post-fertilization (hpf) in WT and *Zvg1* embryos, detected by expression of *cmlc2* by *in situ* hybridization.
(B) Expression of *southpaw* (*spaw)*, a Nodal ligand involved in left-right asymmetry, at the 17-somite stage in WT and *Zvg1* embryos. *spaw* is expressed in the lateral plate mesoderm.



Appendix Figure 5. Vg1 processing requires Nodal.
(A) Anti-GFP reducing immunoblot of WT and M*vg1* embryos injected with 50 pg of *vg1-sfGFP* or *vg1-NC-sfGFP* mRNA and 50 pg of *cyc* or *sqt* mRNA. Black arrowhead indicates full-length Vg1-sfGFP, open arrowhead indicates cleaved Vg1-sfGFP. 8 embryos at 50% epiboly were loaded per well.
(B) Anti-GFP reducing immunoblot of M*vg1* embryos injected with 50 pg of *vg1-sfGFP* mRNA and 50 pg of *cyc, sqt* or *bmp7a* mRNA. Black arrowhead indicates full-length Vg1-sfGFP, open arrowhead indicates processed Vg1-sfGFP. Molecular weights in kDa. 8 embryos at 50% epiboly were loaded per well.

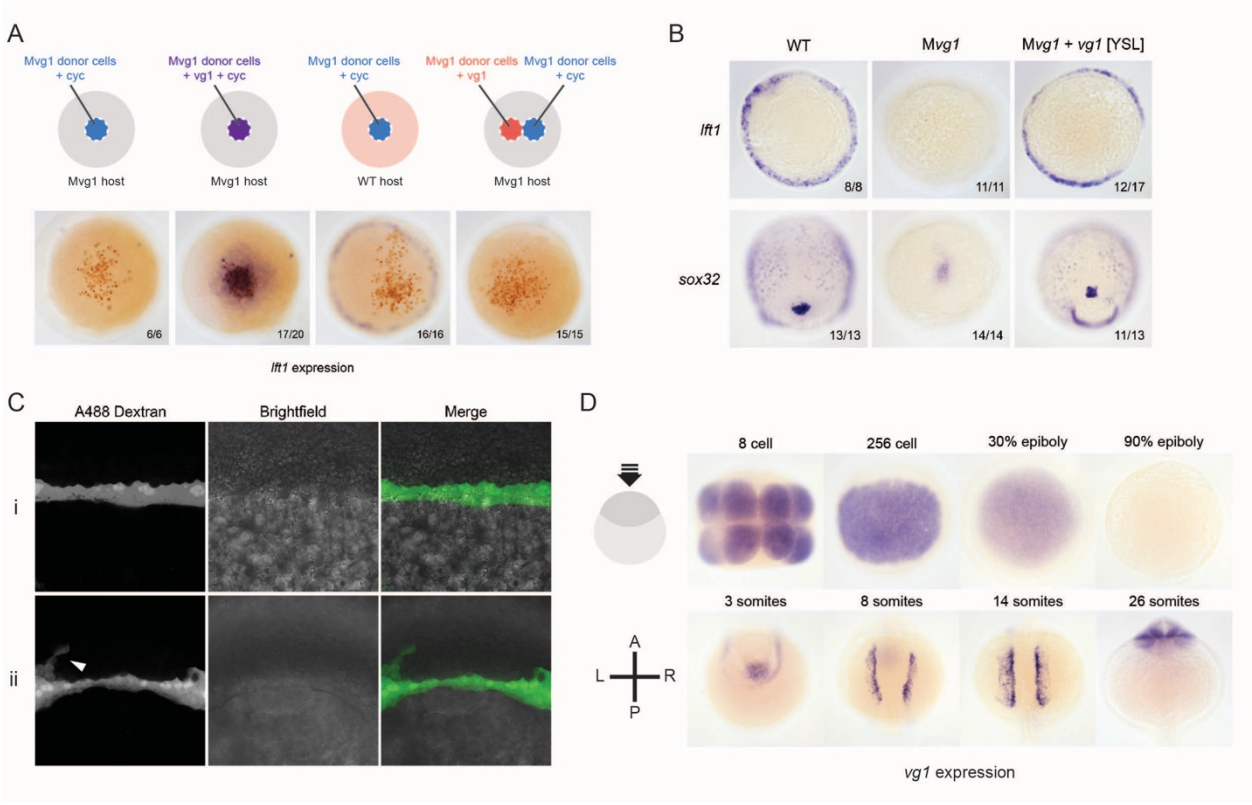


Appendix Figure 6. Vg1 secretion requires Nodal.
(A) M*vg1* embryos co-injected with 20 pg of *vg1-sfGFP* mRNA and 0.5 pg, 2 pg or 20 pg of *cyc* or *sqt* mRNA. These images were captured under the same microscope settings. Arrowheads indicate small extracellular puncta. Note that the puncta increase in size with increasing concentrations of *cyc* mRNA. Co-expression of *vg1-sfGFP* mRNA and *sqt* mRNA can result in extracellular diffuse signal and/or extracellular puncta. Scale bar, 17 um. See Table 1 for quantification.
(B) M*vg1* embryo co-injected with pH-sensitive fluorescent *vg1* (*vg1-pHluorin2)* and *sqt.* Arrowheads indicate fluorescent puncta. Scale bar, 17 um.
(C) Z-stack of M*vg1* embryo co-injected with 50 pg of *vg1-sfGFP* and 50 pg of *cyc-sfGFP*. Scale bar, 17 um.



Appendix Figure 7. Vg1 and Nodal form heterodimers.

(A) Anti-Flag reducing IB of anti-HA IP from lysates of WT embryos injected with 50 pg of *vg1-HA* mRNA and 50 pg of *cyc-Flag* or *sqt-Flag* mRNA. Black arrowhead indicates full-length protein; open arrowhead indicates cleaved protein.
(B) Anti-Flag reducing IB of anti-HA IP from lysates of M*vg1* embryos injected with either 5 pg or 20 pg of *sqt-HA* mRNA combined with increasing concentrations (5 pg, 20 pg or 50 pg) of *vg1-Flag* or *bmp7a-Flag* mRNA. Black arrowhead indicates full-length protein; open arrowhead indicates cleaved protein. In contrast to Vg1-Flag, the mature domain of Bmp7a-Flag is more readily detected than the full-length protein (Little and Mullins, 2009). The samples were loaded onto 3 gels run under the same conditions at the same time (gel boundaries are marked by white vertical lines).
(C) Anti-Flag reducing and non-reducing immunoblots of M*vg1* embryos injected with *vg1-Flag*, *cyc-Flag* or *bmp7a-Flag* mRNA (from Figure 5-5B). Approximate locations of expected monomers and dimers are indicated. TGF-beta family protein dimers are connected by disulfide bonds, which are broken under reducing conditions.
(D) Z-stack of ~256-cell stage M*vg1* embryo co-injected with 100 pg of *vg1-Dendra2* mRNA at the 1-cell stage and 5 pg of *cyc* mRNA at the 64-cell stage immediately followed by photoconversion. The red puncta represent Vg1-Dendra2 protein synthesized before the 64-cell stage that has heterodimerized and been secreted with Cyc protein.



Appendix Figure 8. Vg1 and Nodal are only active when co-expressed.
(A) Transplantation experiments with DAB staining to show the location of the transplanted cells (orange) and *lft1* expression (purple).
(B) *in situ* hybridizations for a mesoderm marker (*lft1*) and an endoderm marker (*sox32*) in *vg1* mRNA YSL-injected embryos.
(C) Confocal imaging of M*vg1* embryos injected with 100 pg of *vg1* and 500 pg of Alexa 488 dextran into the embryonic YSL at the 1000-cell stage. The majority of fluorescent dextran localized to the YSL (i), but in some cases dextran was detectable in a few margin cells (ii, white arrowhead).
(D) Endogenous *vg1* expression in WT embryos.

References

Agathon, A., Thisse, B., and Thisse, C. (2001). Morpholino knock□down of antivin1 and antivin2 upregulates nodal signaling. Genesis *30*, 178–182.

Agius, E., Oelgeschläger, M., Wessely, O., Kemp, C., and De Robertis, E.M. (2000). Endodermal Nodal-related signals and mesoderm induction in Xenopus. Development (Cambridge, England) *127*, 1173–1183.

Andersen, E.C., and Horvitz, H.R. (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. Development (Cambridge, England) *134*, 2991–2999.

Andersson, O., Bertolino, P., and Ibáñez, C.F. (2007). Distinct and cooperative roles of mammalian Vg1 homologs GDF1 and GDF3 during early embryonic development. Dev. Biol. *311*, 500–511.

Andersson, O., Reissmann, E., Jörnvall, H., and Ibáñez, C.F. (2006). Synergistic interaction between Gdf1 and Nodal during anterior axis development. Dev. Biol. *293*, 370–381.

Aono, A., Hazama, M., Notoya, K., Taketomi, S., Yamasaki, H., Tsukuda, R., Sasaki, S., and Fujisawa, Y. (1995). Potent ectopic bone-inducing activity of bone morphogenetic protein-4/7 heterodimer. Biochem. Biophys. Res. Commun. *210*, 670–677.

Attisano, L. (1993). Identification of human activin and TGF□ type I receptors that form heteromeric kinase complexes with type II receptors. Cell *75*, 671–680.

Attisano, L., Wrana, J.L., Cheifetz, S., and Massagué, J. (1992). Novel activin receptors: Distinct genes and alternative mRNA splicing generate a repertoire of serine/threonine kinase receptors. Cell *68*, 97–108.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature *226*, 1209–1211.

Barrila, J., Ott, C.M., LeBlanc, C., Mehta, S.K., Crabbé, A., Stafford, P., Pierson, D.L., and Nickerson, C.A. (2016). Spaceflight modulates gene expression in the whole blood of astronauts. NPJ Microgravity *2*, 16039.

Bauer, H., Lele, Z., Rauch, G.J., Geisler, R., and Hammerschmidt, M. (2001). The type I serine/threonine kinase receptor Alk8/Lost-a-fin is required for Bmp2b/7 signal transduction during dorsoventral patterning of the zebrafish embryo. Development (Cambridge, England) *128*, 849–858.

Beck, S., Le Good, J.A., Guzman, M., Ben Haim, N., Roy, K., Beermann, F., and Constam, D.B. (2002). Extraembryonic proteases regulate Nodal signalling during gastrulation. Nat. Cell Biol. *4*, 981–985.

Bellipanni, G., Varga, M., Maegawa, S., Imai, Y., Kelly, C., Myers, A.P., Chu, F., Talbot, W.S., and Weinberg, E.S. (2006). Essential and opposing roles of zebrafish beta-catenins in the formation of dorsal axial structures and neurectoderm. Development (Cambridge, England) *133*, 1299–1309.

Ben Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E.J., and Constam, D.B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. Dev. Cell *11*, 313–323.

Bennett, C.F., and Swayze, E.E. (2010). RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. Annu. Rev. Pharmacol. Toxicol. *50*, 259–293.

Bennett, J.T., Joubin, K., Cheng, S., Aanstad, P., Herwig, R., Clark, M., Lehrach, H., and Schier, A.F. (2007). Nodal signaling activates differentiation genes during zebrafish gastrulation. Dev. Biol. *304*, 525–540.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. Embo J. *7*, 1749–1756.

Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. Embo J. *29*, 3082–3093.

Bertocchini, F., and Stern, C.D. (2002). The hypoblast of the chick embryo positions the primitive streak by antagonizing nodal signaling. Dev. Cell *3*, 735–744.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. *41*, 7429–7437.

Birsoy, B. (2005). XPACE4 is a localized pro-protein convertase required for mesoderm induction and the cleavage of specific TGF proteins in Xenopus development. Development (Cambridge, England) *132*, 591–602.

Birsoy, B., Kofron, M., Schaible, K., Wylie, C., and Heasman, J. (2006). Vg1 is an essential signaling molecule in Xenopus development. Development (Cambridge, England) *133*, 15–20.

Bisgrove, B.W., Essner, J.J., and Yost, H.J. (1999). Regulation of midline development by antagonism of lefty and nodal signaling. Development (Cambridge, England) *126*, 3253–3262.

Blanchet, M.-H., Le Good, J.A., Mesnard, D., Oorschot, V., Baflast, S., Minchiotti, G., Klumperman, J., and Constam, D.B. (2008). Cripto recruits Furin and PACE4 and controls Nodal trafficking during proteolytic maturation. Embo J. *27*, 2580–2591.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509–1512.

Bodenstine, T.M., Chandler, G.S., Seftor, R.E.B., Seftor, E.A., and Hendrix, M.J.C. (2016). Plasticity underlies tumor progression: role of Nodal signaling. Cancer Metastasis Rev. *35*, 21–39.

Boguraev, A.-S., Christensen, H.C., Bonneau, A.R., Pezza, J.A., Nichols, N.M., Giraldez, A.J., Gray, M.M., Wagner, B.M., Aken, J.T., Foley, K.D., et al. (2017). Successful amplification of DNA aboard the International Space Station. NPJ Microgravity *3*, 26.

Bostock, M., Ogievetsky, V., and Heer, J. (2011). D^3: Data-Driven Documents. IEEE Trans Vis Comput Graph *17*, 2301–2309.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., et al. (2004). Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat. Biotechnol. *22*, 707–716.

Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S., and Robertson, E.J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. Nature *411*, 965–969.

Brennan, J., Norris, D.P., and Robertson, E.J. (2002). Nodal activity in the node governs left-right asymmetry. Genes Dev. *16*, 2339–2344.

Camus, A., Perea-Gomez, A., Moreau, A., and Collignon, J. (2006). Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. Dev. Biol. *295*, 743–755.

Cartwright, J.H.E., Piro, N., Piro, O., and Tuval, I. (2008). Fluid dynamics of nodal flow and left-right patterning in development. Dev. Dyn. *237*, 3477–3490.

Chari, R., Mali, P., Moosburner, M., and Church, G.M. (2015). Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Meth.

Chea, H.K., Wright, C.V., and Swalla, B.J. (2005). Nodal signaling and the evolution of deuterostome gastrulation. Dev. Dyn. *234*, 269–278.

Chen, C., and Shen, M.M. (2004). Two modes by which Lefty proteins inhibit nodal signaling. Curr. Biol. *14*, 618–624.

Chen, C., Ware, S.M., Sato, A., Houston-Hawkins, D.E., Habas, R., Matzuk, M.M., Shen, M.M., and Brown, C.W. (2006). The Vg1-related protein Gdf3 acts in a Nodal signaling pathway in the pre-gastrulation mouse embryo. Development (Cambridge, England) *133*, 319–329.

Chen, Y., and Schier, A.F. (2001). erratum: The zebrafish Nodal signal Squint functions as a morphogen. Nature *412*, 566–566.

Chen, Y., and Schier, A.F. (2002). Lefty proteins are long-range inhibitors of squint-mediated nodal signaling. Curr. Biol. *12*, 2124–2128.

Cheng, S.K., Olale, F., Bennett, J.T., Brivanlou, A.H., and Schier, A.F. (2003). EGF-CFC proteins are essential coreceptors for the TGF-beta signals Vg1 and GDF1. Genes Dev. *17*, 31–36.

Cheng, S.K., Olale, F., Brivanlou, A.H., and Schier, A.F. (2004). Lefty blocks a subset of TGFbeta signals by antagonizing EGF-CFC coreceptors. PLoS Biol. *2*, E30.

Chng, S.C., Ho, L., Tian, J., and Reversade, B. (2013). ELABELA: A Hormone Essential for Heart Development Signals via the Apelin Receptor. Dev. Cell *27*, 672–680.

Choi, W.-Y., Giraldez, A.J., and Schier, A.F. (2007). Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. Science *318*, 271–274.

Ciruna, B., Weidinger, G., Knaut, H., Thisse, B., Thisse, C., Raz, E., and Schier, A.F. (2002). Production of maternal-zygotic mutant zebrafish by germ-line replacement. Proc. Natl. Acad. Sci. U.S.a. *99*, 14919–14924.

Clements, D., Friday, R.V., and Woodland, H.R. (1999). Mode of action of VegT in mesoderm and endoderm formation. Development (Cambridge, England) *126*, 4903–4911.

Collignon, J., Varlet, I., and Robertson, E.J. (1996). Relationship between asymmetric nodal expression and the direction of embryonic turning. Nature *381*, 155–158.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cong, L., Zhou, R., Kuo, Y.-C., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun *3*, 968.

Conklin, E.G. (1905). Mosaic development in ascidian eggs. Journal of Experimental Zoology *2*, 145–223.

Conlon, F.L., Barth, K.S., and Robertson, E.J. (1991). A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the developmental fate of embryonic stem cells homozygous for the 413.d proviral integration. Development (Cambridge, England) *111*, 969–981.

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development (Cambridge, England) *120*, 1919–1928.

Connors, S.A., Trout, J., Ekker, M., and Mullins, M.C. (1999). The role of tolloid/mini fin in dorsoventral pattern formation of the zebrafish embryo. Development (Cambridge, England) *126*, 3119–3130.

Constam, D.B. (2009). Running the gauntlet: an overview of the modalities of travel employed by the putative morphogen Nodal. Curr. Opin. Genet. Dev. *19*, 302–307.

Constam, D.B. (2014). Regulation of TGF and related signals by precursor processing. Semin. Cell Dev. Biol. *32*, 85–97.

Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: rcommendations 1984.

Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. (2013). CRISPR/Cas9 systems targeting □-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. *41*, 9584–9592.

Crucian, B., Stowe, R.P., Mehta, S., Quiriarte, H., Pierson, D., and Sams, C. (2015). Alterations in adaptive immunity persist during long-duration spaceflight. NPJ Microgravity *1*, 15013.

Cubano, L.A., and Lewis, M.L. (2001). Effect of vibrational stress and spaceflight on regulation of heat shock proteins hsp70 and hsp27 in human lymphocytes (Jurkat). J. Leukoc. Biol. *69*, 755–761.

Cucinotta, F.A., Alp, M., Sulzman, F.M., and Wang, M. (2014). Space radiation risks to the central nervous system. Life Sciences in Space Research *2*, 54–69.

Cucinotta, F.A., Kim, M.-H.Y., Chappell, L.J., and Huff, J.L. (2013). How safe is safe enough? Radiation risk for a human mission to Mars. PLoS ONE *8*, e74988.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. *23*, 1534–1541.

Dale, L., Matthews, G., and Colman, A. (1993). Secretion and mesoderm-inducing activity of the TGF-beta-related domain of Xenopus Vg1. Embo J. *12*, 4471–4480.

Dale, L., Matthews, G., Tabe, L., and Colman, A. (1989). Developmental expression of the protein product of Vg1, a localized maternal mRNA in the frog Xenopus laevis. Embo J. *8*, 1057.

De Robertis, E.M. (2006). Spemann's organizer and self-regulation in amphibian embryos. Nat. Rev. Mol. Cell Biol. *7*, 296–302.

Degnin, C., Jean, F., Thomas, G., and Christian, J.L. (2004). Cleavages within the prodomain direct intracellular trafficking and degradation of mature bone morphogenetic protein-4. Mol. Biol. Cell *15*, 5012–5020.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. *41*, 4336–4343.

Dick, A., Hild, M., Bauer, H., Imai, Y., Maifeld, H., Schier, A.F., Talbot, W.S., Bouwmeester, T., and Hammerschmidt, M. (2000). Essential role of Bmp7 (snailhouse) and its prodomain in dorsoventral patterning of the zebrafish embryo. Development (Cambridge, England) *127*, 343–354.

Dillon, S.L., Williamson, D.M., Elferich, J., Radler, D., Joshi, R., Thomas, G., and Shinde, U. (2012). Propeptides Are Sufficient to Regulate Organelle-Specific pH-Dependent Activation of Furin and Proprotein Convertase 1/3. J. Mol. Biol. *423*, 47–62.

Dimitrova, N., Zamudio, J.R., Jong, R.M., Soukup, D., Resnick, R., Sarma, K., Ward, A.J., Raj, A., Lee, J.T., Sharp, P.A., et al. (2014). LincRNA-p21 Activates p21 In cis to Promote Polycomb Target Gene Expression and to Enforce the G1/S Checkpoint. Mol. Cell *54*, 777–790.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. *32*, 1262–1267.

Dohrmann, C.E., Hemmati-Brivanlou, A., Thomsen, G.H., Fields, A., Woolf, T.M., and Melton, D.A. (1993). Expression of activin mRNA during early development in Xenopus laevis. Dev. Biol. *157*, 474–483.

Dohrmann, C.E., Kessler, D.S., and Melton, D.A. (1996). Induction of axial mesoderm by zDVR-1, the zebrafish orthologue of Xenopus Vg1. Dev. Biol. *175*, 108–117.

Dougan, S.T., Warga, R.M., Kane, D.A., Schier, A.F., and Talbot, W.S. (2003). The role of the zebrafish nodal-related genes squint and cyclops in patterning of mesendoderm. Development (Cambridge, England) *130*, 1837–1851.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K., and Bogdanove, A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res. *40*, W117–W122.

Draper, B.W., Morcos, P.A., and Kimmel, C.B. (2001). Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. Genesis *30*, 154–156.

Dubrulle, J., Jordan, B.M., Akhmetova, L., Farrell, J.A., Kim, S.-H., Solnica-Krezel, L., and Schier, A.F. (2015). Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. Elife *4*.

Duitman, E.H., Orinska, Z., Bulanova, E., Paus, R., and Bulfone-Paus, S. (2008). How a cytokine is chaperoned through the secretory pathway by complexing with its own receptor: lessons from interleukin-15 (IL-15)/IL-15 receptor alpha. Mol. Cell. Biol. *28*, 4851–4861.

Durante, M., and Cucinotta, F.A. (2008). Heavy ion carcinogenesis and human space exploration. Nat. Rev. Cancer *8*, 465–472.

Dutko, J.A., and Mullins, M.C. (2011). SnapShot: BMP signaling in development. Cell *145*, 636–636.e1–2.

Dyson, S., and Gurdon, J.B. (1997). Activin signalling has a necessary function in Xenopus early development. Curr. Biol. *7*, 81–84.

Efstratiadis, A., Maniatis, T., Kafatos, F.C., Jeffrey, A., and Vournakis, J.N. (1975). Full length and discrete partial reverse transcripts of globin and chorion mRNAs. Cell *4*, 367–378.

Efstratiadis, A., Kafatos, F.C., Maxam, A.M., and Maniatis, T. (1976). Enzymatic in vitro synthesis of globin genes. Cell *7*, 279–288.

Eimon, P.M., and Harland, R.M. (2002). Effects of heterodimerization and proteolytic processing on Derrière and Nodal activity: implications for mesoderm induction in Xenopus. Development (Cambridge, England) *129*, 3089–3103.

Eisen, J.S., and Smith, J.C. (2008). Controlling morpholino experiments: don't stop making antisense. Development (Cambridge, England) *135*, 1735–1743.

Eißmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Groß, M., Schirmacher, P., Rippe, K., Braun, T., Diederichs, S., et al. (2014). Loss of the abundant nuclear non-coding RNA MALAT1is compatible with life and development. RNA Biology *9*, 1076–1087.

Erter, C.E., Solnica-Krezel, L., and Wright, C.V. (1998). Zebrafish nodal-related 2 encodes an early mesendodermal inducer signaling from the extraembryonic yolk syncytial layer. Dev. Biol. *204*, 361–372.

Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F., and Talbot, W.S. (1998). Zebrafish organizer development and germ-layer formation require nodal-related signals. Nature *395*, 181–185.

Fish, J.E., Santoro, M.M., Morton, S.U., Yu, S., Yeh, R.-F., Wythe, J.D., Ivey, K.N., Bruneau, B.G., Stainier, D.Y.R., and Srivastava, D. (2008). miR-126 Regulates Angiogenic Signaling and Vascular Integrity. Dev. Cell *15*, 272–284.

Fisher, S., and Halpern, M.E. (1999). Patterning the zebrafish axial skeleton requires early chordin

function. Nat. Genet. *23*, 442–446.

Fodero-Tavoletti, M.T., Hardy, M.P., Cornell, B., Katsis, F., Sadek, C.M., Mitchell, C.A., Kemp, B.E., and Tiganis, T. (2005). Protein tyrosine phosphatase hPTPN20a is targeted to sites of actin polymerization. Biochem. J. *389*, 343–354.

Fodor, E., Zsigmond, Á., Horváth, B., Molnár, J., Nagy, I., Tóth, G., Wilson, S.W., and Varga, M. (2013). Full Transcriptome Analysis of Early Dorsoventral Patterning in Zebrafish. PLoS ONE *8*, e70053.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.-L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. *42*, 2577–2590.

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J.A. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Meth *10*, 741–743.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. Cell *47*, 735–746.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. *32*, 279–284.

Fuerer, C., Nostro, M.C., and Constam, D.B. (2014). Nodal·Gdf1 Heterodimers with Bound Prodomains Enable Serum-independent Nodal Signaling and Endoderm Differentiation. Journal of Biological Chemistry *289*, 17854–17871.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. PLoS ONE *9*, e98186.

Gaiser, A.M., Kaiser, C.J.O., Haslbeck, V., and Richter, K. (2011). Downregulation of the Hsp90 system causes defects in muscle cells of Caenorhabditis elegans. PLoS ONE *6*, e25485.

Gaj, T., Gersbach, C.A., and Barbas, C.F., III (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. *31*, 397–405.

Gerstle, J.T., and Fried, M.G. (1993). Measurement of binding kinetics using the gel electrophoresis mobility shift assay. Electrophoresis *14*, 725–731.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth *6*, 343–345.

Gilbert, S.F. (2016). Developmental Biology (Sinauer Associates IncorporatedU.S.).

Gimlich, R.L., and Gerhart, J.C. (1984). Early cellular interactions promote embryonic axis formation in Xenopus laevis. Dev. Biol. *104*, 117–130.

Graff, J.M., Bansal, A., and Melton, D.A. (1996). Xenopus Mad Proteins Transduce Distinct Subsets of Signals for the TGF Superfamily. Cell *85*, 479–487.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. Genetics *196*, 961–971.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics *194*, 1029–1035.

Gray, A.M., and Mason, A.J. (1990). Requirement for Activin-a and Transforming Growth Factor-Beta-1 Pro-Regions in Homodimer Assembly. Science *247*, 1328–1330.

Gritsman, K., Talbot, W.S., and Schier, A.F. (2000). Nodal signaling patterns the organizer. Development (Cambridge, England) *127*, 921–932.

Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., and Schier, A.F. (1999). The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell *97*, 121–132.

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. Gene *25*, 263–269.

Guilinger, J.P., Pattanayak, V., Reyon, D., Tsai, S.Q., Sander, J.D., Joung, J.K., and Liu, D.R. (2014). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. Nat Meth *11*, 429–435.

Guo, J., and Wu, G. (2012). The signaling and functions of heterodimeric bone morphogenetic proteins. Cytokine Growth Factor Rev. *23*, 61–67.

Gurdon, J.B., Mohun, T.J., Fairman, S., and Brennan, S. (1985). All components required for the eventual activation of muscle-specific actin genes are localized in the subequatorial region of an uncleaved amphibian egg. Pnas *82*, 139–143.

Halpern, M.E., Ho, R.K., Walker, C., and Kimmel, C.B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. Cell *75*, 99–111.

Hamada, H., Meno, C., Watanabe, D., and Saijoh, Y. (2002). Establishment of vertebrate left-right asymmetry. Nat. Rev. Genet. *3*, 103–113.

Hammerschmidt, M., Pelegri, F., Mullins, M.C., Kane, D.A., van Eeden, F.J., Granato, M., Brand, M., Furutani-Seiki, M., Haffter, P., Heisenberg, C.P., et al. (1996). dino and mercedes, two genes regulating dorsal development in the zebrafish embryo. Development (Cambridge, England) *123*, 95–102.

Hatta, K., Kimmel, C.B., Ho, R.K., and Walker, C. (1991). The cyclops mutation blocks specification of the floor plate of the zebrafish central nervous system. Nature *350*, 339–341.

Hawley, S.H., Wünnenberg-Stapleton, K., Hashimoto, C., Laurent, M.N., Watabe, T., Blumberg, B.W., and Cho, K.W. (1995). Disruption of BMP signals in embryonic Xenopus ectoderm leads to direct neural induction. Genes Dev. *9*, 2923–2935.

Hazama, M., Aono, A., Ueno, N., and Fujisawa, Y. (1995). Efficient expression of a heterodimer of bone morphogenetic protein subunits using a baculovirus expression system. Biochem. Biophys. Res. Commun. *209*, 859–866.

Heasman, J., Ginsberg, D., Geiger, B., Goldstone, K., Pratt, T., Yoshida-Noro, C., and Wylie, C. (1994). A functional test for maternally inherited cadherin in Xenopus shows its importance in cell adhesion at the blastula stage. Development (Cambridge, England) *120*, 49–57.

Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. Nat Meth *11*, 122–123.

Heigwer, F., Kerr, G., Walther, N., Glaeser, K., Pelz, O., Breinig, M., and Boutros, M. (2013). E-TALEN: a web tool to design TALENs for genome engineering. Nucleic Acids Res. *41*, e190.

Heijne, von, G. (1990). The signal peptide. J. Membrain Biol. *115*, 195–201.

Heisenberg, C.P., Tada, M., Rauch, G.J., Saúde, L., Concha, M.L., Geisler, R., Stemple, D.L., Smith, J.C., and Wilson, S.W. (2000). Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. Nature *405*, 76–81.

Heisenberg, C.-P., and Nüsslein-Volhard, C. (1997). The Function of silberblick in the Positioning of the Eye Anlage in the Zebrafish Embryo. Dev. Biol. *184*, 85–94.

Helde, K.A., and Grunwald, D.J. (1993). The DVR-1 (Vg1) transcript of zebrafish is maternally supplied and distributed throughout the embryo. Dev. Biol. *159*, 418–426.

Hemmati-Brivanlou, A., and Melton, D.A. (1992). A truncated activin receptor inhibits mesoderm induction and formation of axial structures in Xenopus embryos. Nature *359*, 609–614.

Hild, M., Dick, A., Rauch, G.J., Meier, A., Bouwmeester, T., Haffter, P., and Hammerschmidt, M. (1999). The smad5 mutation somitabun blocks Bmp2b signaling during early dorsoventral patterning of the zebrafish embryo. Development (Cambridge, England) *126*, 2149–2159.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hirokawa, N., Tanaka, Y., Okada, Y., and Takeda, S. (2006). Nodal flow and the generation of left-right asymmetry. Cell *125*, 33–45.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429–3431.

Hong, S.-K., Jang, M.K., Brown, J.L., McBride, A.A., and Feldman, B. (2011). Embryonic mesoderm and endoderm induction requires the actions of non-embryonic Nodal-related ligands and Mxtx2. Development (Cambridge, England) *138*, 787–795.

Hoveizi, E., Nabiuni, M., Parivar, K., Ai, J., and Massumi, M. (2014). Definitive endoderm differentiation of human-induced pluripotent stem cells using signaling molecules and IDE1 in three-dimensional polymer scaffold. J Biomed Mater Res A *102*, 4027–4036.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hurtley, S.M., and Helenius, A. (1989). Protein oligomerization in the endoplasmic reticulum. Annu. Rev. Cell Biol. *5*, 277–307.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Hyatt, B.A., and Yost, H.J. (1998). The left-right coordinator: the role of Vg1 in organizing left-right axis formation. Cell *93*, 37–46.

Hyatt, B.A., Lohr, J.L., and Yost, H.J. (1996). Initiation of vertebrate left-right axis formation by maternal Vg1. Nature *384*, 62–65.

Hyde, C.E., and Old, R.W. (2000). Regulation of the early expression of the Xenopus nodal-related 1 gene, Xnr1. Development (Cambridge, England) *127*, 1221–1229.

Iannaccone, P.M., Zhou, X., Khokha, M., Boucher, D., and Kuehn, M.R. (1992). Insertional mutation of a gene involved in growth regulation of the early mouse embryo. Dev. Dyn. *194*, 198–208.

Ideue, T., Hino, K., Kitao, S., Yokoi, T., and Hirose, T. (2009). Efficient oligonucleotide-mediated degradation of nuclear noncoding RNAs in mammalian cultured cells. Rna *15*, 1578–1587.

Israel, D.I., Nove, J., Kerns, K.M., Kaufman, R.J., Rosen, V., Cox, K.A., and Wozney, J.M. (1996). Heterodimeric bone morphogenetic proteins show enhanced activity in vitro and in vivo. Growth Factors *13*, 291–300.

Jalah, R., Rosati, M., Ganneru, B., Pilkington, G.R., Valentin, A., Kulkarni, V., Bergamaschi, C., Chowdhury, B., Zhang, G.-M., Beach, R.K., et al. (2013). The p40 subunit of interleukin (IL)-12 promotes stabilization and export of the p35 subunit: implications for improved IL-12 cytokine production. Journal of Biological Chemistry *288*, 6763–6776.

Jeffery, W.R., Tomlinson, C.R., and Brodeur, R.D. (1983). Localization of actin messenger RNA during early ascidian development. Dev. Biol. *99*, 408–417.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Jones, C.M., Kuehn, M.R., Hogan, B.L., Smith, J.C., and Wright, C.V. (1995). Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. Development (Cambridge, England) *121*, 3651–3662.

Jones, C.M., Simon-Chazottes, D., Guenet, J.L., and Hogan, B.L. (1992). Isolation of Vgr-2, a novel member of the transforming growth factor-beta-related gene family. Mol. Endocrinol. *6*, 1961–1968.

Joseph, E.M., and Melton, D.A. (1997). Xnr4: a Xenopus nodal-related gene expressed in the Spemann organizer. Dev. Biol. *184*, 367–372.

Joseph, E.M., and Melton, D.A. (1998). Mutant Vg1 ligands disrupt endoderm and mesoderm formation in Xenopus embryos. Development (Cambridge, England) *125*, 2677–2685.

Juan, H., and Hamada, H. (2001). Roles of nodal-lefty regulatory loops in embryonic patterning of vertebrates. Genes Cells *6*, 923–930.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kelly, A., and Hurlstone, A.F. (2011). The use of RNAi technologies for gene knockdown in zebrafish. Brief Funct Genomics *10*, 189–196.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996–1006.

Kessler, D.S., and Melton, D.A. (1995). Induction of dorsal mesoderm by soluble, mature Vg1 protein. Development (Cambridge, England) *121*, 2155–2164.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.-I., and Kim, J.-S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Meth *12*, 237–43–1pfollowing243.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.-S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res *19*, 1279–1288.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.

Kishimoto, Y., Lee, K.H., Zon, L., Hammerschmidt, M., and Schulte-Merker, S. (1997). The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning. Development (Cambridge, England) *124*, 4457–4466.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P.W., Li, Z., Peterson, R.T., Yeh, J.-R.J., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kofron, M., Demel, T., Xanthos, J., Lohr, J., Sun, B., Sive, H., Osada, S., Wright, C., Wylie, C., and Heasman, J. (1999). Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors. Development (Cambridge, England) *126*, 5759–5770.

Kohler, J.J., and Schepartz, A. (2001). Kinetic Studies of Fos·Jun·DNA Complex Formation: DNA Binding Prior to Dimerization. Biochemistry *40*, 130–142.

Kok, F.O., Shin, M., Ni, C.-W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I., et al. (2014). Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. Dev. Cell.

Kordasiewicz, H.B., Stanek, L.M., Wancewicz, E.V., Mazur, C., McAlonis, M.M., Pytel, K.A., Artates, J.W., Weiss, A., Cheng, S.H., Shihabuddin, L.S., et al. (2012). Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. Neuron *74*, 1031–1044.

Kubo, A. (2004). Development of definitive endoderm from embryonic stem cells in culture. Development (Cambridge, England) *131*, 1651–1662.

Kumar, A., Lualdi, M., Lewandoski, M., and Kuehn, M.R. (2008). Broad mesodermal and endodermal deletion of Nodal at postgastrulation stages results solely in left/right axial defects. Dev. Dyn. *237*, 3591–3601.

Kumar, A., Lualdi, M., Lyozin, G.T., Sharma, P., Loncarek, J., Fu, X.-Y., and Kuehn, M.R. (2015). Nodal signaling from the visceral endoderm is required to maintain Nodal gene expression in the epiblast and drive DVE/AVE migration. Dev. Biol. *400*, 1–9.

Kurreck, J. (2003). Antisense technologies. Improvement through novel chemical modifications. Eur J Biochem *270*, 1628–1644.

Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. *44*, W272–W276.

Lamason, R.L., Mohideen, M.-A.P.K., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E., et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science *310*, 1782–1786.

Langheinrich, U., Hennen, E., Stott, G., and Vacun, G. (2002). Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling. Curr. Biol. *12*, 2023–2028.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Le Good, J.A., Joubin, K., Giraldez, A.J., Ben Haim, N., Beck, S., Chen, Y., Schier, A.F., and Constam, D.B. (2005). Nodal stability determines signaling range. Curr. Biol. *15*, 31–36.

Lee, S.-J. (1990). Identification of a Novel Member (GDF-1) of the Transforming Growth Factor- Superfamily. Mol. Endocrinol. *4*, 1034–1040.

Lele, Z., Bakkers, J., and Hammerschmidt, M. (2001). Morpholino phenocopies of the swirl, snailhouse, somitabun, minifin, silberblick, and pipetail mutations. Genesis *30*, 190–194.

Levin, A.A. (1999). A review of the issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides. Biochim. Biophys. Acta *1489*, 69–84.

Levin, M., Johnson, R.L., Sterna, C.D., Kuehn, M., and Tabin, C. (1995). A molecular pathway determining left-right asymmetry in chick embryogenesis. Cell *82*, 803–814.

Levin, M., Pagan, S., Roberts, D.J., Cooke, J., Kuehn, M.R., and Tabin, C.J. (1997). Left/right patterning signals and the independent regulation of different aspects of situs in the chick embryo. Dev. Biol. *189*, 57–67.

Liang, J.O., and Rubinstein, A.L. (2003). Patterning of the zebrafish embryo by nodal signals. Curr. Top. Dev. Biol.

Lister, J.A., Robertson, C.P., Lepage, T., Johnson, S.L., and Raible, D.W. (1999). nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. Development (Cambridge, England) *126*, 3757–3767.

Little, S.C., and Mullins, M.C. (2009). Bone morphogenetic protein heterodimers assemble heteromeric type I receptor complexes to pattern the dorsoventral axis. Nat. Cell Biol. *11*, 637–643.

Liu, Z., Lin, X., Cai, Z., Zhang, Z., Han, C., Jia, S., Meng, A., and Wang, Q. (2011). Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. Journal of Biological Chemistry *286*, 28520–28532.

Lohr, J.L., Danos, M.C., and Yost, H.J. (1997). Left-right asymmetry of a nodal-related gene is regulated by dorsoanterior midline structures during Xenopus development. Development (Cambridge, England) *124*, 1465–1472.

Long, S., Ahmad, N., and Rebagliati, M. (2003). The zebrafish nodal-related gene southpaw is required for visceral and diencephalic left-right asymmetry. Development (Cambridge, England) *130*, 2303–2316.

Lorenz, I.C., Allison, S.L., Heinz, F.X., and Helenius, A. (2002). Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. J. Virol. *76*, 5480–5491.

Lorenz, P., Misteli, T., Baker, B.F., Bennett, C.F., and Spector, D.L. (2000). Nucleocytoplasmic shuttling: a novel in vivo property of antisense phosphorothioate oligodeoxynucleotides. Nucleic Acids Res. *28*, 582–592.

Lowe, L.A., Yamada, S., and Kuehn, M.R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. Development (Cambridge, England) *128*, 1831–1843.

Lowe, L.A., Supp, D.M., Sampath, K., Yokoyama, T., Wright, C.V.E., Potter, S.S., Overbeek, P., and Kuehn, M.R. (1996). Conserved left–right asymmetry of nodal expression and alterations in murine situs inversus. Nature *381*, 158–161.

Lustig, K.D., Kroll, K., Sun, E., Ramos, R., Elmendorf, H., and Kirschner, M.W. (1996). A Xenopus nodal-related gene that acts in synergy with noggin to induce complete secondary axis and notochord formation. Development (Cambridge, England) *122*, 3275–3282.

Mahon, M.J. (2011). pHluorin2: an enhanced, ratiometric, pH-sensitive green florescent protein. Adv Biosci Biotechnol *2*, 132–137.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Mali, P., Esvelt, K.M., and Church, G.M. (2013b). Cas9 as a versatile tool for engineering biology. Nat Meth *10*, 957–963.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013c). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Marlow, F.L., and Mullins, M.C. (2008). Bucky ball functions in Balbiani body assembly and animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish. Dev. Biol. *321*, 40–50.

Martínez-Barberá, J.P., Toresson, H., Da Rocha, S., and Stefan Krauss (1997). Cloning and expression of three members of the zebrafish Bmp family: Bmp2a, Bmp2b and Bmp4. Gene *198*, 53–59.

Massagué, J. (1987). The TGF-beta family of growth and differentiation factors. Cell *49*, 437–438.

Mathews, L.S., Vale, W.W., and Kintner, C.R. (1992). Cloning of a second type of activin receptor and functional characterization in Xenopus embryos. Science *255*, 1702–1705.

Mathews, L.S., and Vale, W.W. (1991). Expression cloning of an activin receptor, a predicted transmembrane serine kinase. Cell *65*, 973–982.

Matzuk, M.M., Kumar, T.R., Vassalli, A., Bickenbach, J.R., Roop, D.R., Jaenisch, R., and Bradley, A. (1995). Functional analysis of activins during mammalian development. Nature *374*, 354–356.

McIntyre, A.B.R., Rizzardi, L., Yu, A.M., Alexander, N., Rosen, G.L., Botkin, D.J., Stahl, S.E., John, K.K., Castro-Wallace, S.L., McGrath, K., et al. (2016). Nanopore sequencing in microgravity. NPJ Microgravity *2*, 16035.

Meeker, N.D., Hutchinson, S.A., Ho, L., and Trede, N.S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. BioTechniques *43*, 610–612–614.

Melton, D.A. (1987). Translocation of a localized maternal mRNA to the vegetal pole of Xenopus oocytes. Nature *328*, 80.

Meng, L., Ward, A.J., Chun, S., Bennett, C.F., Beaudet, A.L., and Rigo, F. (2014). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. Nature.

Meno, C., Gritsman, K., Ohishi, S., Ohfuji, Y., Heckscher, E., Mochida, K., Shimono, A., Kondoh, H., Talbot, W.S., Robertson, E.J., et al. (1999). Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. Mol. Cell *4*, 287–298.

Meno, C., Saijoh, Y., Fujii, H., Ikeda, M., Yokoyama, T., Yokoyama, M., Toyoda, Y., and Hamada, H. (1996). Left–right asymmetric expression of the TGF□-family member lefty in mouse embryos. Nature *381*, 151–155.

Mi, L.-Z., Brown, C.T., Gao, Y., Tian, Y., Le, V.Q., Walz, T., and Springer, T.A. (2015). Structure of bone morphogenetic protein 9 procomplex. Proc. Natl. Acad. Sci. U.S.a. 201501303.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29*, 143–148.

Mintzer, K.A., Lee, M.A., Runke, G., Trout, J., Whitman, M., and Mullins, M.C. (2001). Lost-a-fin encodes a type I BMP receptor, Alk8, acting maternally and zygotically in dorsoventral pattern formation. Development (Cambridge, England) *128*, 859–869.

Mir, A., and Heasman, J. (2008). How the mother can help: studying maternal Wnt signaling by anti-sense-mediated depletion of maternal mRNAs and the host transfer technique. Methods Mol. Biol. *469*, 417–429.

Moens, C.B., Donn, T.M., Wolf-Saxon, E.R., and Ma, T.P. (2008). Reverse genetics in zebrafish by TILLING. Brief Funct Genomic Proteomic *7*, 454–459.

Montague, T.G., and Schier, A.F. (2017). Vg1-Nodal heterodimers are the endogenous inducers of mesendoderm. Elife *6*, 178.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, W401–W407.

Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Meth *12*, 982–988.

Moreno-Villanueva, M., Wong, M., Lu, T., Zhang, Y., and Wu, H. (2017). Interplay of space radiation and microgravity in DNA damage and DNA damage response. NPJ Microgravity *3*, 14.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Muller, P., Rogers, K.W., Jordan, B.M., Lee, J.S., Robson, D., Ramanathan, S., and Schier, A.F. (2012). Differential Diffusivity of Nodal and Lefty Underlies a Reaction-Diffusion Patterning System. Science *336*, 721–724.

Nakagawa, S., Ip, J.Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., and Prasanth, K.V. (2012). Malat1 is not an essential component of nuclear speckles in mice. Rna *18*, 1487–1499.

Nakayama, T., Fish, M.B., Fisher, M., Oomen-Hajagos, J., Thomsen, G.H., and Grainger, R.M. (2013). Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in Xenopus tropicalis. Genesis *51*, 835–843.

Nasevicius, A., and Ekker, S.C. (2000). Effective targeted gene “knockdown” in zebrafish. Nat. Genet. *26*, 216–220.

Neff, K.L., Argue, D.P., Ma, A.C., Lee, H.B., Clark, K.J., and Ekker, S.C. (2013). Mojo Hand, a TALEN design tool for genome editing applications. BMC Bioinformatics *14*, 1.

Neugebauer, J.M., Kwon, S., Kim, H.-S., Donley, N., Tilak, A., Sopory, S., and Christian, J.L. (2015). The prodomain of BMP4 is necessary and sufficient to generate stable BMP4/7 heterodimers with enhanced bioactivity in vivo. Proc. Natl. Acad. Sci. U.S.a. *112*, E2307–E2316.

Nguyen, V.H., Schmid, B., Trout, J., Connors, S.A., Ekker, M., and Mullins, M.C. (1998). Ventral and lateral regions of the zebrafish gastrula, including the neural crest progenitors, are established by a bmp2b/swirl pathway of genes. Dev. Biol. *199*, 93–110.

Nieuwkoop, P.D. (1969). The Formation of the Mesoderm in Urodelean Amphibians. W. Roux' Archiv F. Entwicklungsmechanik *163*, 298–315.

Nieuwkoop, P.D. (1977). Origin and establishment of embryonic polar axes in amphibian development. Curr. Top. Dev. Biol. *11*, 115–132.

Nieuwkoop, P.D., and Ubbels, G.A. (1972). The formation of the mesoderm in urodelean amphibians. W. Roux' Archiv F. Entwicklungsmechanik *169*, 185–199.

Nishimatsu, S., and Thomsen, G.H. (1998). Ventral mesoderm induction and patterning by bone morphogenetic protein heterodimers in Xenopus embryos. Mech. Dev. *74*, 75–88.

Noël, E.S., Verhoeven, M., Lagendijk, A.K., Tessadori, F., Smith, K., Choorapoikayil, S., Hertog, den, J., and Bakkers, J. (2013). A Nodal-independent and tissue-intrinsic mechanism controls heart-looping chirality. Nat Commun *4*, 2754.

Northrup, S.H., and Erickson, H.P. (1992). Kinetics of protein-protein association explained by Brownian dynamics computer simulation. Proc. Natl. Acad. Sci. U.S.a. *89*, 3338–3342.

Okayama, H., and Berg, P. (1982). High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. *2*, 161–170.

Onai, T., Yu, J.-K., Blitz, I.L., Cho, K.W.Y., and Holland, L.Z. (2010). Opposing Nodal/Vg1 and BMP signals mediate axial patterning in embryos of the basal chordate amphioxus. Dev. Biol. *344*, 377–389.

Pagán-Westphal, S.M., and Tabin, C.J. (1998). The transfer of left-right positional information during chick embryogenesis. Cell *93*, 25–35.

Pan, W.-H., and Clawson, G.A. (2006). Identifying accessible sites in RNA: The first step in designing antisense reagents. Curr. Med. Chem. *13*, 3083–3103.

Papanayotou, C., and Collignon, J. (2014). Activin/Nodal signalling before implantation: setting the stage for embryo patterning. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *369*.

Parisi, S., D'Andrea, D., Lago, C.T., Adamson, E.D., Persico, M.G., and Minchiotti, G. (2003). Nodal-dependent Cripto signaling promotes cardiomyogenesis and redirects the neural fate of embryonic stem cells. J. Cell Biol. *163*, 303–314.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843.

Pauklin, S., and Vallier, L. (2015). Activin/Nodal signalling in stem cells. Development (Cambridge, England) *142*, 607–619.

Pauli, A., Montague, T.G., Lennox, K.A., Behlke, M.A., and Schier, A.F. (2015). Antisense Oligonucleotide-Mediated Transcript Knockdown in Zebrafish. PLoS ONE *10*, e0139504.

Pauli, A., Norris, M.L., Valen, E., Chew, G.-L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., et al. (2014). Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. Science.

Peek, A.S., and Behlke, M.A. (2007). Design of active small interfering RNAs. Current Opinion in Molecular Therapeutics *9*, 110–118.

Pelliccia, J.L., Jindal, G.A., and Burdine, R.D. (2017). Gdf3 is required for robust Nodal signaling during germ layer formation and left-right patterning. Elife *6*, e28635.

Persson, R., and Pettersson, R.F. (1991). Formation and intracellular transport of a heterodimeric viral spike protein complex. J. Cell Biol. *112*, 257–266.

Peterson, A.G., Wang, X., and Yost, H.J. (2013). Dvr1 transfers left-right asymmetric signals from Kupffer's vesicle to lateral plate mesoderm in zebrafish. Dev. Biol. *382*, 198–208.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Prakash, T.P. (2011). An overview of sugar-modified oligonucleotides for antisense therapeutics. Chem. Biodivers. *8*, 1616–1641.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. Cell *123*, 249–263.

Prober, D.A., Rihel, J., Onah, A.A., Sung, R.-J., and Schier, A.F. (2006). Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. J. Neurosci. *26*, 13400–13410.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Quail, D.F., Siegers, G.M., Jewer, M., and Postovit, L.-M. (2013). Nodal signalling in embryogenesis and tumourigenesis. Int. J. Biochem. Cell Biol. *45*, 885–898.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Range, R., and Lepage, T. (2011). Maternal Oct1/2 is required for Nodal and Vg1/Univin expression during dorsal-ventral axis specification in the sea urchin embryo. Dev. Biol. *357*, 440–449.

Range, R., Lapraz, F., Quirin, M., Marro, S., Besnardeau, L., and Lepage, T. (2007). Cis-regulatory analysis of nodal and maternal control of dorsal-ventral axis formation by Univin, a TGF-beta related to Vg1. Development (Cambridge, England) *134*, 3649–3664.

Rankin, C.T., Bunton, T., Lawler, A.M., and Lee, S.J. (2000). Regulation of left-right patterning in mice by growth/differentiation factor-1. Nat. Genet. *24*, 262–265.

Rauch, G.J., Hammerschmidt, M., Blader, P., Schauerte, H.E., Strähle, U., Ingham, P.W., Mcmahon, A.P., and Haffter, P. (1997). WNT5 is required for tail formation in the zebrafish embryo. Cold Spring Harb. Symp. Quant. Biol. *62*, 227–234.

Rebagliati, M.R., Toyama, R., Fricke, C., and Haffter, P. (1998a). Zebrafish Nodal-Related Genes Are Implicated in Axial Patterning and Establishing Left–Right Asymmetry. Developmental ….

Rebagliati, M.R., Toyama, R., Haffter, P., and Dawid, I.B. (1998b). cyclops encodes a nodal-related factor involved in midline signaling. Proc. Natl. Acad. Sci. U.S.a. *95*, 9932–9937.

Rebagliati, M.R., Weeks, D.L., Harvey, R.P., and Melton, D.A. (1985). Identification and cloning of localized maternal RNAs from Xenopus eggs. Cell *42*, 769–777.

Robertson, E.J. (2014). Dose-dependent Nodal/Smad signals pattern the early mouse embryo. Semin. Cell Dev. Biol.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature.

Rougeon, F., and Mach, B. (1976). Stepwise biosynthesis in vitro of globin genes from globin mRNA by DNA polymerase of avian myeloblastosis virus. Pnas *73*, 3418–3422.

Rougeon, F., Kourilsky, P., and Mach, B. (1975). Insertion of a rabbit beta-globin gene sequence into an E. coli plasmid. Nucleic Acids Res. *2*, 2365–2378.

Saijoh, Y., Oki, S., Ohishi, S., and Hamada, H. (2003). Left-right patterning of the mouse lateral plate requires nodal produced in the node. Dev. Biol. *256*, 160–172.

Sampath, K., Cheng, A.M., Frisch, A., and Wright, C.V. (1997). Functional differences among Xenopus nodal-related genes in left-right axis determination. Development (Cambridge, England) *124*, 3293–3302.

Sampath, K., Rubinstein, A.L., Cheng, A.M., Liang, J.O., Fekany, K., Solnica-Krezel, L., Korzh, V., Halpern, M.E., and Wright, C.V. (1998). Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. Nature *395*, 185–189.

Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F., and Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res. *35*, W599–W605.

Schier, A.F., and Shen, M.M. (2000). Nodal signalling in vertebrate development. Nature *403*, 385–389.

Schier, A.F., Neuhauss, S.C., Helde, K.A., Talbot, W.S., and Driever, W. (1997). The one-eyed pinhead gene functions in mesoderm and endoderm formation in zebrafish and interacts with no tail. Development (Cambridge, England) *124*, 327–342.

Schier, A.F. (2003). Nodal signaling in vertebrate development. Annu. Rev. Cell Dev. Biol. *19*, 589–621.

Schier, A.F. (2009). Nodal morphogens. Cold Spring Harbor Perspectives in Biology *1*, a003459.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Meth *9*, 676–682.

Schlesinger, M.J. (1990). Heat shock proteins. J. Biol. Chem. *265*, 12111–12114.

Schmid, B., Furthauer, M., Connors, S.A., Trout, J., Thisse, B., Thisse, C., and Mullins, M.C. (2000). Equivalent genetic roles for bmp7/snailhouse and bmp2b/swirl in dorsoventral pattern formation. Development (Cambridge, England) *127*, 957–967.

Schulte-Merker, S., Lee, K.J., Mcmahon, A.P., and Hammerschmidt, M. (1997). The zebrafish organizer requires chordino. Nature *387*, 862–863.

Schulte-Merker, S., Smith, J.C., and Dale, L. (1994). Effects of truncated activin and FGF receptors and of follistatin on the inducing activities of BVg1 and activin: does activin play a role in mesoderm induction? Embo J. *13*, 3533–3541.

Schulte-Merker, S., and Stainier, D.Y.R. (2014). Out with the old, in with the new: reassessing morpholino knockdowns in light of genome editing technology. Development (Cambridge, England) *141*, 3103–3104.

Seleiro, E.A., Connolly, D.J., and Cooke, J. (1996). Early developmental expression and experimental axis determination by the chicken Vg1 gene. Curr. Biol. *6*, 1476–1486.

Semov, A. (2002). Alterations in TNF- and IL-related gene expression in space-flown WI38 human fibroblasts. The FASEB Journal.

Shah, S.B., Skromne, I., Hume, C.R., Kessler, D.S., Lee, K.J., Stern, C.D., and Dodd, J. (1997). Misexpression of chick Vg1 in the marginal zone induces primitive streak formation. Development (Cambridge, England) *124*, 5127–5138.

Shen, M.M. (2007). Nodal signaling: developmental roles and regulation. Development (Cambridge, England) *134*, 1023–1034.

Shi, M., Zhu, J., Wang, R., Chen, X., Mi, L., Walz, T., and Springer, T.A. (2011). Latent TGF- structure and activation. Nature *474*, 343–349.

Shimmi, O., Umulis, D., Othmer, H., and O'Connor, M.B. (2005). Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the Drosophila blastoderm embryo. Cell *120*, 873–886.

Signore, I.A., Palma, K., and Concha, M.L. (2016). Nodal signalling and asymmetry of the nervous system. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *371*.

Skromne, I., and Stern, C.D. (2001). Interactions between Wnt and Vg1 signalling pathways initiate primitive streak formation in the chick embryo. Development (Cambridge, England) *128*, 2915–2927.

Smith, J.C. (1987). A mesoderm-inducing factor is produced by a Xenopus cell line. Development (Cambridge, England) *99*, 3–14.

Smith, J.C., Price, B.M., Van Nimmen, K., and Huylebroeck, D. (1990). Identification of a potent Xenopus mesoderm-inducing factor as a homologue of activin A. Nature *345*, 729–731.

Smith, J.R., Vallier, L., Lupo, G., Alexander, M., Harris, W.A., and Pedersen, R.A. (2008). Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. Dev. Biol. *313*, 107–117.

Smith, W.C., McKendry, R., Ribisi, S., and Harland, R.M. (1995). A nodal-related gene defines a physical and functional domain within the Spemann organizer. Cell *82*, 37–46.

Sopory, S., Kwon, S., Wehrli, M., and Christian, J.L. (2010). Regulation of Dpp activity by tissue-specific cleavage of an upstream site within the prodomain. Dev. Biol. *346*, 102–112.

Southall, T.D., Terhzaz, S., Cabrero, P., Chintapalli, V.R., Evans, J.M., Dow, J.A.T., and Davies, S.-A. (2006). Novel subcellular locations and functions for secretory pathway Ca2+/Mn2+-ATPases. Physiol. Genomics *26*, 35–45.

Spemann, H., and Mangold, H. (1924). über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. Archiv Für Mikroskopische Anatomie Und Entwicklungsmechanik *101*, 458.

Streubel, J., Blücher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. Nat. Biotechnol. *30*, 593–595.

Strizzi, L., Hardy, K.M., Kirschmann, D.A., Ahrlund-Richter, L., and Hendrix, M.J.C. (2012). Nodal expression and detection in cancer: experience and challenges. Cancer Res. *72*, 1915–1920.

Strizzi, L., Hardy, K.M., Seftor, E.A., Costa, F.F., Kirschmann, D.A., Seftor, R.E.B., Postovit, L.-M., and Hendrix, M.J.C. (2009). Development and cancer: at the crossroads of Nodal and Notch signaling. Cancer Res. *69*, 7131–7134.

Summerton, J. (1999). Morpholino antisense oligomers: the case for an RNase H-independent structural type. Biochim. Biophys. Acta *1489*, 141–158.

Sun, B.I., Bush, S.M., Collins-Racie, L.A., LaVallie, E.R., DiBlasio-Smith, E.A., Wolfman, N.M., McCoy, J.M., and Sive, H.L. (1999). derrière: a TGF-beta family member required for posterior development in Xenopus. Development (Cambridge, England) *126*, 1467–1482.

Suzuki, A., Kaneko, E., Maeda, J., and Ueno, N. (1997). Mesoderm induction by BMP-4 and -7 heterodimers. Biochem. Biophys. Res. Commun. *232*, 153–156.

Szewczyk, N.J., Tillman, J., Conley, C.A., Granger, L., Segalat, L., Higashitani, A., Honda, S., Honda, Y., Kagawa, H., Adachi, R., et al. (2008). Description of International Caenorhabditis elegans Experiment first flight (ICE-FIRST). Adv Space Res *42*, 1072–1079.

Szul, T., and Sztul, E. (2011). COPII and COPI traffic at the ER-Golgi interface. Physiology (Bethesda) *26*, 348–364.

Tadjuidje, E., Kofron, M., Mir, A., Wylie, C., Heasman, J., and Cha, S.-W. (2016). Nodal signalling in Xenopus: the role of Xnr5 in left/right asymmetry and heart development. Open Biol *6*.

Takahashi, S., Yokota, C., Takano, K., Tanegashima, K., Onuma, Y., Goto, J., and Asashima, M. (2000). Two novel nodal-related genes initiate early inductive events in Xenopus Nieuwkoop center. Development (Cambridge, England) *127*, 5319–5329.

Takenaga, M., Fukumoto, M., and Hori, Y. (2007). Regulated Nodal signaling promotes differentiation of the definitive endoderm and mesoderm from ES cells. J. Cell. Sci. *120*, 2078–2090.

Tanaka, C., Sakuma, R., Nakamura, T., Hamada, H., and Saijoh, Y. (2007). Long-range action of Nodal requires interaction with GDF1. Genes Dev. *21*, 3272–3282.

Tannahill, D., and Melton, D.A. (1989). Localized synthesis of the Vg1 protein during early Xenopus development. Development (Cambridge, England) *106*, 775–785.

Temin, H.M., and Mizutani, S. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. Nature *226*, 1211–1213.

Terada, M., Seki, M., Takahashi, R., Yamada, S., Higashibata, A., Majima, H.J., Sudoh, M., Mukai, C., and Ishioka, N. (2016). Effects of a Closed Space Environment on Gene Expression in Hair Follicles of Astronauts in the International Space Station. PLoS ONE *11*, e0150801.

Tessadori, F., Noël, E.S., Rens, E.G., Magliozzi, R., Gogh, I.J.A.E.-V., Guardavaccaro, D., Merks, R.M.H., and Bakkers, J. (2015). Nodal Signaling Range Is Regulated by Proprotein Convertase-Mediated Maturation. Dev. Cell *32*, 631–639.

Thisse, B., Wright, C.V., and Thisse, C. (2000). Activin- and Nodal-related factors control antero-posterior patterning of the zebrafish embryo. Nature *403*, 425–428.

Thisse, C., and Thisse, B. (1999). Antivin, a novel and divergent member of the TGFbeta superfamily, negatively regulates mesoderm induction. Development (Cambridge, England) *126*, 229–240.

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc *3*, 59–69.

Thomas, J.T., and Moos, M. (2007). Vg1 has specific processing requirements that restrict its action to body axis patterning centers. Dev. Biol. *310*, 129–139.

Thompson, T.B. (2003). Structures of an ActRIIB:activin A complex reveal a novel binding mode for TGF-beta ligand:receptor interactions. Embo J. *22*, 1555–1566.

Thomsen, G.H., and Melton, D.A. (1993). Processed Vg1 protein is an axial mesoderm inducer in Xenopus. Cell *74*, 433–441.

Thomsen, G., Woolf, T., Whitman, M., Sokol, S., Vaughan, J., Vale, W., and Melton, D.A. (1990). Activins are expressed early in Xenopus embryogenesis and can induce axial mesoderm and anterior structures. Cell *63*, 485–493.

Thyme, S.B., Akhmetova, L., Montague, T.G., Valen, E., and Schier, A.F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. Nat Commun *7*, 11750–11757.

Tian, T., and Meng, A.M. (2006). Nodal signals pattern vertebrate embryos. Cell. Mol. Life Sci. *63*, 672–685.

Torpey, N., Wylie, C.C., and Heasman, J. (1992). Function of maternal cytokeratin in Xenopus development. Nature *357*, 413–415.

Toyama, R., O'Connell, M.L., Wright, C.V., Kuehn, M.R., and Dawid, I.B. (1995). Nodal induces ectopic goosecoid and lim1 expression and axis duplication in zebrafish. Development (Cambridge, England) *121*, 383–391.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197.

Tu, L., Sun, T.-T., and Kreibich, G. (2002). Specific heterodimer formation is a prerequisite for uroplakins to exit from the endoplasmic reticulum. Mol. Biol. Cell *13*, 4221–4230.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

Vallier, L., Reynolds, D., and Pedersen, R.A. (2004). Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. Dev. Biol. *275*, 403–421.

van den Eijnden-van Raaij, A.J., van Zoelent, E.J., Van Nimmen, K., Koster, C.H., Snoek, G.T., Durston, A.J., and Huylebroeck, D. (1990). Activin-like factor from a Xenopus laevis cell line responsible for mesoderm induction. Nature *345*, 732–734.

van Impel, A., Zhao, Z., Hermkens, D.M.A., Roukens, M.G., Fischer, J.C., Peterson-Maduro, J., Duckers, H., Ober, E.A., Ingham, P.W., and Schulte-Merker, S. (2014). Divergence of zebrafish and mouse lymphatic cell fate specification pathways. Development (Cambridge, England) *141*, 1228–1238.

Vickers, T.A., Koo, S., Bennett, C.F., Crooke, S.T., Dean, N.M., and Baker, B.F. (2003). Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. J. Biol. Chem. *278*, 7108–7118.

Wall, N.A., Craig, E.J., Labosky, P.A., and Kessler, D.S. (2000). Mesendoderm induction and reversal of left-right pattern by mouse Gdf1, a Vg1-related gene. Dev. Biol. *227*, 495–509.

Wang, F., Shi, Z., Cui, Y., Guo, X., Shi, Y.-B., and Chen, Y. (2015a). Targeted gene disruption in Xenopus laevis using CRISPR/Cas9. Cell & Bioscience *5*, 225.

Wang, H., and Tsang, B.K. (2007). Nodal signalling and apoptosis. Reproduction.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Wang, S., Aurora, A.B., Johnson, B.A., Qi, X., McAnally, J., Hill, J.A., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2008). The Endothelial-Specific MicroRNA miR-126 Governs Vascular Integrity and Angiogenesis. Dev. Cell *15*, 261–271.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J., and Yee, J.-K. (2015b). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat. Biotechnol. *33*, 175–178.

Wang, X., Fischer, G., and Hyvönen, M. (2016). Structure and activation of pro-activin A. Nat Commun *7*, 12052.

Watanabe, M., Iwashita, M., Ishii, M., Kurachi, Y., Kawakami, A., Kondo, S., and Okada, N. (2006). Spot pattern of leopard Danio is caused by mutation in the zebrafish connexin41.8 gene. EMBO Rep. *7*, 893–897.

Weeks, D.L., and Melton, D.A. (1987). A maternal mRNA localized to the vegetal hemisphere in Xenopus eggs codes for a growth factor related to TGF-beta. Cell *51*, 861–867.

Weidinger, G., Stebler, J., Slanchev, K., Dumstrei, K., Wise, C., Lovell-Badge, R., Thisse, C., Thisse, B., and Raz, E. (2003). dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. Curr. Biol. *13*, 1429–1434.

Wheeler, T.M., Leger, A.J., Pandey, S.K., Macleod, A.R., Nakamori, M., Cheng, S.H., Wentworth, B.M., Bennett, C.F., and Thornton, C.A. (2012). Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature *488*, 111–115.

White, R.J., and Averner, M. (2001). Humans in space. Nature *409*, 1115–1118.

Whitman, M. (2001). Nodal signaling in early vertebrate embryos: themes and variations. Dev. Cell *1*, 605–617.

Wilson, E.B. (1925). The Cell in Development and Heredity (Macmillan Company, New York).

Wittbrodt, J., and Rosa, F.M. (1994). Disruption of mesoderm and axis formation in fish by ectopic expression of activin variants: the role of maternal activin. Genes Dev. *8*, 1448–1462.

Xanthos, J.B., Kofron, M., Wylie, C., and Heasman, J. (2001). Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis. Development (Cambridge, England) *128*, 167–180.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., and Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics *30*, 1180–1182.

Xu, C., Fan, Z.P., Müller, P., Fogley, R., DiBiase, A., Trompouki, E., Unternaehrer, J., Xiong, F., Torregroza, I., Evans, T., et al. (2012). Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway. Dev. Cell *22*, 625–638.

Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C.A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J.S., et al. (2015). Sequence determinants of improved CRISPR sgRNA design. Genome Res *25*, 1147–1157.

Yan, Y.T., Gritsman, K., Ding, J., Burdine, R.D., Corrales, J.D., Price, S.M., Talbot, W.S., Schier, A.F., and Shen, M.M. (1999). Conserved requirement for EGF-CFC genes in vertebrate left-right axis formation. Genes Dev. *13*, 2527–2537.

Yang, J., Tan, C., Darken, R.S., Wilson, P.A., and Klein, P.S. (2002). □-Catenin/Tcf-regulated transcription prior to the midblastula transition. Development (Cambridge, England) *129*, 5743–5752.

Yang, T., Mendoza-Londono, R., Lu, H., Tao, J., Li, K., Keller, B., Jiang, M.M., Shah, R., Chen, Y., Bertin, T.K., et al. (2010). E-selectin ligand–1 regulates growth plate homeostasis in mice by inhibiting the intracellular processing and secretion of mature TGF-□. J. Clin. Invest. *120*, 2474–2485.

Yasunaga, M., Tada, S., Torikai-Nishikawa, S., Nakano, Y., Okada, M., Jakt, L.M., Nishikawa, S., Chiba, T., Era, T., and Nishikawa, S.-I. (2005). Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. Nat. Biotechnol. *23*, 1542–1550.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J., Zhang, C., et al. (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep *2*, 111–123.

Zhang, J., and King, M.L. (1996). Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. Development (Cambridge, England) *122*, 4119–4129.

Zhang, J., Houston, D.W., King, M.L., Payne, C., Wylie, C., and Heasman, J. (1998a). The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. Cell *94*, 515–524.

Zhang, J., Talbot, W.S., and Schier, A.F. (1998b). Positional Cloning Identifies Zebrafish one-eyed pinhead as a Permissive EGF-Related Ligand Required during Gastrulation. Cell *92*, 241–251.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L.M., and Kuehn, M.R. (1993). Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature *361*, 543–547.

Zupanska, A.K., Denison, F.C., Ferl, R.J., and Paul, A.-L. (2013). Spaceflight engages heat shock protein and other molecular chaperone genes in tissue culture cells of Arabidopsis thaliana. Am. J. Bot. *100*, 235–248.