Developmental Biology 425 (2017) 8–20

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Resource

Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in *Ciona*

Shashank Gandhi[a,1], Maximilian Haeussler[b], Florian Razy-Krajka[a], Lionel Christiaen[a,*], Alberto Stolfi[a,*]

[a] Department of Biology, New York University, New York, USA
[b] Santa Cruz Genomics Institute, University of California, MS CBSE, Santa Cruz, USA

ARTICLE INFO

Keywords:
sgRNA
Cardiopharyngeal mesoderm
Tunicate

ABSTRACT

The CRISPR/Cas9 system has emerged as an important tool for various genome engineering applications. A current obstacle to high throughput applications of CRISPR/Cas9 is the imprecise prediction of highly active single guide RNAs (sgRNAs). We previously implemented the CRISPR/Cas9 system to induce tissue-specific mutations in the tunicate *Ciona*. In the present study, we designed and tested 83 single guide RNA (sgRNA) vectors targeting 23 genes expressed in the cardiopharyngeal progenitors and surrounding tissues of *Ciona* embryo. Using high-throughput sequencing of mutagenized alleles, we identified guide sequences that correlate with sgRNA mutagenesis activity and used this information for the rational design of all possible sgRNAs targeting the *Ciona* transcriptome. We also describe a one-step cloning-free protocol for the assembly of sgRNA expression cassettes. These cassettes can be directly electroporated as unpurified PCR products into *Ciona* embryos for sgRNA expression *in vivo*, resulting in high frequency of CRISPR/Cas9-mediated mutagenesis in somatic cells of electroporated embryos. We found a strong correlation between the frequency of an *Ebf* loss-of-function phenotype and the mutagenesis efficacies of individual *Ebf*-targeting sgRNAs tested using this method. We anticipate that our approach can be scaled up to systematically design and deliver highly efficient sgRNAs for the tissue-specific investigation of gene functions in *Ciona*.

1. Introduction

A platform for targeted mutagenesis has been recently developed based on the prokaryotic immune response system known as CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated) (Barrangou et al., 2007). In its most common derivation for genome engineering applications, the system makes use of a short RNA sequence, known as a single guide RNA (sgRNA) to direct the Cas9 nuclease of *Streptococcus pyogenes* to a specific target DNA sequence. (Jinek et al., 2012, 2013; Cong et al., 2013; Mali et al., 2013). Although initial Cas9 binding requires a Protospacer Adjacent Motif (PAM) sequence, most commonly "NGG", the high specificity of this system is accounted for by Watson-Crick base pairing between the 5′ end of the sgRNA and a 17–20 bp "protospacer" sequence immediately adjacent to the PAM (Fu et al., 2014). Upon sgRNA-guided binding to the intended target, Cas9 generates a double stranded break (DSB) within the protospacer sequence. Imperfect repair of these DSBs by non-homologous end joining (NHEJ) often results in short insertions or deletions (indels) that may disrupt the function of the targeted sequence. Numerous reports have confirmed the high efficiency of CRISPR/Cas9 for genome editing purposes (Dickinson et al., 2013; Hwang et al., 2013; H. Wang et al., 2013; Koike-Yusa et al., 2014; Paix et al., 2014; Shalem et al., 2014; Wang et al. 2014; Gantz and Bier, 2015; Sanjana et al., 2016).

The tunicate *Ciona* is a model organism for the study of chordate-specific developmental processes (Satoh, 2013). The CRISPR/Cas9 system was adapted to induce site-specific DSBs in the *Ciona* genome (Sasaki et al., 2014; Stolfi et al., 2014). Using electroporation to transiently transfect *Ciona* embryos with plasmids encoding CRISPR/Cas9 components, we were able to generate clonal populations of somatic cells carrying loss-of-function mutations of *Ebf*, a transcription-factor-coding gene required for muscle and neuron development, in F0-generation embryos (Stolfi et al., 2014). By using developmentally regulated *cis*-regulatory elements to drive Cas9 expression in specific cell lineages or tissue types, we were thus able to control the disruption of *Ebf* function with spatiotemporal precision. Following

* Corresponding authors.
E-mail addresses: lc121@nyu.edu (L. Christiaen), stolfidobranchia@gmail.com (A. Stolfi).
[1] Current address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA.

http://dx.doi.org/10.1016/j.ydbio.2017.03.003
Received 6 October 2016; Received in revised form 6 February 2017; Accepted 5 March 2017
Available online 22 March 2017
0012-1606/

this proof-of-principle, tissue-specific CRISPR/Cas9 has rapidly propagated as a simple yet powerful tool to elucidate gene function in the *Ciona* embryo (Abdul-Wajid et al., 2015; Cota and Davidson, 2015; Segade et al., 2016; Tolkin and Christiaen, 2016).

We sought to expand the strategy to target more genes, with the ultimate goal of building a genome-wide library of sgRNAs for systematic genetic loss-of-function assays in *Ciona* embryos. However, not all sgRNAs drive robust CRISPR/Cas9-induced mutagenesis, and few guidelines exist for the rational design of highly active sgRNAs, which are critical for rapid gene disruption in F0. This variability and unpredictable efficacy demands experimental validation of each sgRNA tested. Individual studies have revealed certain nucleotide sequence features that correlate with high sgRNA expression and/or activity in CRISPR/Cas9-mediated DNA cleavage (Doench et al., 2014, 2016; Gagnon et al., 2014; Ren et al., 2014; Wang et al., 2014; Chari et al., 2015; Fusi et al., 2015; Housden et al., 2015; Moreno-Mateos et al., 2015; Wong et al., 2015; Xu et al., 2015). These studies have been performed in different organisms using a variety of sgRNA and Cas9 delivery methods and show varying ability to predict sgRNA activities across platforms (Haeussler et al., 2016).

Given the uncertainty regarding how sgRNA design principles gleaned from experiments in other species might be applicable to *Ciona*, we tested a collection of sgRNAs using our own modified tools for tissue-specific CRISPR/Cas9-mediated mutagenesis in *Ciona* embryos. We describe here the construction and validation of this collection using high-throughput sequencing of PCR-amplified target sequences. This dataset allowed us to develop a practical pipeline for optimal design and efficient assembly of sgRNA expression constructs for use in *Ciona*.

Table 1
Genes targeted for CRISPR/Cas9-mediated mutagenesis.

#	Gene Symbol	Aliases	2012 KyotoHoya ID
1	*Bmp2/4*	*Bone morphogenetic protein 2/4*	KH.C4.125
2	*Ddr*	*Discoidin Domain Receptor Tyrosine Kinase 1/2*	KH.C9.371
3	*Ebf*	*Collier/Olf/EBF; COE*	KH.L24.10
4	*Eph.a*	*Ephrin type-A receptor.a; Eph1*	KH.C1.404
5	*Ets.b*	*Ets/Pointed2*	KH.C11.10
6	*Fgf4/6*	*Fibroblast growth factor 4/6; FGF, unassigned 1*	KH.C1.697
7	*Fgf8/17/18*	*Fibroblast growth factor 8/17/18*	KH.C5.5
8	*Fgfr*	*Fibroblast growth factor receptor*	KH.S742.2
9	*Foxf*	*FoxF*	KH.C3.170
10	*Foxg-r*	*Foxg-related; Orphan Fox-4; Ci-ZF248*	KH.C5.74
11	*Fzd5/8*	*Frizzled5/8*	KH.C9.260
12	*Gata4/5/6*	*GATA-a*	KH.L20.1
13	*Hand*	*Heart And Neural Crest Derivatives Expressed 1/2*	KH.C14.604
14	*Hand-r*	*Hand-related; Hand-like; NoTrlc*	KH.C1.1116
15	*Htr7-r*	*5-Hydroxytryptamine Receptor 7-related*	KH.S555.1
16	*Isl*	*Islet1/2*	KH.L152.2
17	*Lef1*	*Lef/TCF*	KH.C6.71
18	*Mrf*	*Muscle regulatory factor; MyoD*	KH.C14.307
19	*Neurog*	*Neurogenin; Ngn*	KH.C6.129
20	*Nk4*	*Nkx2-5; Tinman*	KH.C8.482
21	*Rhod/f*	*RhoD/F; Rif*	KH.C1.129
22	*Tle.b*	*Groucho2*	KH.L96.50
23	*Tll*	*Tolloid-like; Tolloid*	KH.C12.156

The 23 genes targeted in the initial screen, each identified here by official gene symbol, aliases, and KyotoHoya identifier.

2. Results

2.1. High-Throughput sequencing to estimate sgRNA-specific mutagenesis rates

Previous studies using CRISPR/Cas9-based mutagenesis in *Ciona* revealed that different sgRNAs have varying ability to induce mutations (Sasaki et al., 2014; Stolfi et al., 2014). In order to test a larger number of sgRNAs and identify parameters that may influence mutagenesis efficacy, we constructed a library of 83 sgRNA expression plasmids targeting a set of 23 genes (Table 1). The majority of these genes are transcription factors and signaling molecules of potential interest in the study of cardiopharyngeal mesoderm development. The cardiopharyngeal mesoderm of *Ciona*, also known as the Trunk Ventral Cells (TVCs), are multipotent cells that invariantly give rise to the heart and pharyngeal muscles of the adult (Hirano and Nishida, 1997; Stolfi et al., 2010; Razy-Krajka et al., 2014), thus sharing a common ontogenetic motif with the cardiopharyngeal mesoderm of vertebrates (W. Wang et al., 2013; Diogo et al., 2015; Kaplan et al., 2015).

We followed a high-throughput-sequencing-based approach to quantify the efficacy of each sgRNA, i.e. its ability to cause CRISPR/Cas9-induced mutations in targeted sequences in the genome (Fig. 1a–c). The 83 sgRNA plasmids were co-electroporated with *Eef1a1 > nls::Cas9::nls* plasmid. The ubiquitous *Eef1a1* promoter is active in all cell lineages of the embryo and has been used to express a variety of site-specific nucleases for targeted somatic knockouts in *Ciona* (Sasakura et al., 2010; Kawai et al., 2012; Sasaki et al., 2014; Stolfi et al., 2014; Treen et al., 2014). Each individual sgRNA + Cas9 vector combination was electroporated into pooled *Ciona* zygotes, which were then grown at 18 °C for 16 hours post-fertilization (hpf; embryonic stage 25) (Hotta et al., 2007). Targeted sequences were individually PCR-amplified from each pool of embryos. Each target was also amplified from "negative control" embryos grown in parallel and electroporated with *Eef1a1 > nls::Cas9::nls* and "*U6 > Negative Control*" sgRNA vector. Agarose gel-selected and purified amplicons (varying from 108 to 350 bp in length) were pooled in a series of barcoded "targeted" and "negative control" Illumina sequencing libraries and sequenced using the Illumina MiSeq platform.

Alignment of the resulting reads to the reference genome sequence (Satou et al., 2008) revealed that targeted sites were represented on average by 16,204 reads, with a median of 3899 reads each (Supplementary Table 1). The ability of each sgRNA to guide Cas9 to induce DSBs at its intended target was detected by the presence of insertions and deletions (indels) within the targeted protospacer + PAM. The ratio of [indels]/[total reads] represents our estimation of the mutagenesis efficacy of the sgRNA (Fig. 1b–d, Supplementary Table 1). For each mutation, an unknown number of cell divisions occur between the moment Cas9 induces a DSB and the time of sample collection and genomic DNA extraction. This prevents us from quantifying the true mutagenesis rates. However, we can surmise that this applies comparably to all sgRNAs, such that our values still represent an accurate ranking of mutagenesis efficacy.

For simplicity, we did not count single nucleotide point mutations, even though a fraction of them may result from NHEJ-repair of a DSB event. Our data indicated that all sgRNAs (with the exception of Neurog.2) were able to induce DSBs, with estimated efficacies varying from 0.05% (Ebf.4) to 59.63% (Htr7-r.2). Although each sgRNA was tested only once, we did not observe any evidence of electroporation variability or batch effects that may have confounded our estimates (Supplementary Figure 1).

This conservative approach most likely underestimates the actual mutagenesis rates. First, we excluded point mutations potentially resulting from imperfect DSB repair. Second, but more importantly, amplicons from transfected cells are always diluted by wild-type sequences from untransfected cells in the same sample, due to mosaic incorporation of sgRNA and Cas9 plasmids. Indeed, we previously observed an enrichment of mutated sequences amplified from reporter transgene-expressing cells isolated by magnetic-activated cell sorting (representing the transfected population of cells) relative to unsorted