

Induction of Interferon-Stimulated Genes and Cellular Stress Pathways by Morpholinos.

Jason K.H. Lai^{1,¶,i}, Kristina Gagalova^{1,¶,ii}, Didier Y.R. Stainier^{1,*}.

¹ Max Planck Institute for Heart and Lung Research, Department of Developmental Genetics, Bad Nauheim 61231, Germany.

* Corresponding Author.

Email: didier.stainier@mpi-bn.mpg.de

¶ These authors contributed equally to this work.

ⁱ Present Address: Mechanobiology Institute Singapore, National University of Singapore, 117411 Singapore.

ⁱⁱ Present Address: Canada Michael Smith Genome Sciences Center, Vancouver BC V5Z 4S6, Canada.

Abstract

The phenotypes caused by morpholino-mediated interference of gene function in zebrafish are often not observed in the corresponding mutant(s). We took advantage of the availability of a relatively large collection of transcriptomic datasets to identify common signatures that characterize morpholino-injected animals (morphants). In addition to the previously reported activation of *tp53* expression, we observed increased expression of the interferon-stimulated genes (ISGs), *isg15* and *isg20*, an extrinsic cell death pathway gene, *casp8*, and other cellular stress response genes including *phlda3*, *mdm2* and *gadd45aa*. Studies involving segmentation stage embryos were more likely to show upregulation of these genes. We also found that these genes could be upregulated by an increasing dose of *egfl7* morpholino, or high doses of standard control morpholino. Thus, these data suggest that morpholinos can stimulate the innate immune response in zebrafish, as was recently reported in *Xenopus*.

Keywords: morpholino, zebrafish, ISG, p53, knockdown, mutant.

Introduction

Morpholinos are anti-sense reagents capable of knocking down gene function by inhibiting post-transcriptional regulation of RNAs [1]. They are synthetic nucleic acid analogs which can heterodimerize with complementary RNA molecules to inhibit their processing.

Morpholinos inhibit mRNAs by blocking their splicing or translation [2-3], and they can inhibit miRNA function as well [4].

The ease of morpholino use in zebrafish and its effectiveness in gene knockdown enabled scientists to transiently inhibit gene function to study various biological processes [2]. Initial testing of morpholinos indicated that morphants could largely phenocopy known mutants [2, 5]. However, the examples were limited, and restricted to genes with well characterized mutant phenotype(s). With the recent advent of gene editing, scientists have now been able to test more extensively whether reported morphant phenotypes are similar to those observed in mutants, and they have found substantial discrepancy [6]. For example, a study by Kok et al. generated mutations in more than 20 genes and observed that most mutants appear to develop normally while the corresponding morphants for ten of these genes display a developmental phenotype, suggesting additional effects introduced by morpholinos [7]. This idea was supported by the observation that the hydrocephaly phenotype in *megamind* morphants could be reproduced by injecting a *megamind* morpholino into animals devoid of the *megamind* locus, albeit at a very high dose (20 ng) [7]. In addition, Rossi, Kontarakis et al. proposed that genetic compensation in mutants but not in morphants could explain their phenotypic discrepancy [8].

One well characterized morpholino off-target effect is the activation of the p53 signaling pathway [9]. Additionally, a recent study in *Xenopus* found that morpholinos could activate

the innate immune pathway, as well as induce off-target mis-splicing of pre-mRNA transcripts [10]. These studies raise questions about the effects of morpholinos in a cell, and the need to fully understand them for future applications. In this study, we sought to further characterize the effects of morpholinos by studying various microarray gene expression datasets from morpholino studies in zebrafish obtained from the Gene Expression Omnibus (GEO) database. We found that the expression of interferon-stimulated genes, ISGs (*isg15* and *isg20*), and other cellular stress pathway genes (*tp53*, *phlda3*, *casp8*, *gadd45aa* and *mdm2*) was upregulated in several morpholino studies, especially in segmentation stage embryos. Furthermore, these genes were upregulated in response to increasing doses of an *egfl7* morpholino, or high doses of a standard control morpholino. In contrast, a *vegfaa* morpholino did not induce the expression of these genes. Our study suggests that morpholinos can induce the innate immune response and furthers our understanding of morpholino effects in developing zebrafish embryos.

Materials and Methods

Microarray data collection and curation

Microarray gene expression arrays of various zebrafish morpholino studies were obtained from the GEO database. As of the 19th of March 2015, keyword search for “zebrafish morpholino” or “zebrafish knockdown” yielded 63 unique *Danio rerio* data series (GSE). The embedded metadata and raw intensities for all studies were retrieved with the R package GEOquery [11]. We manually curated all datasets utilizing GPL1319, a platform manufactured by Affymetrix, to annotate the publication, morpholino sequence, morpholino dose, sample source and developmental stage. Datasets from Agilent arrays were set aside for meta-analysis. All annotated information can be found in the File S1 – Section 1.

Affymetrix microarray data normalization and processing

8 GSE data series were rejected because they were not morpholino studies, were a duplicate, or were not linked to a publication. The remaining data series were pooled together for normalization. The raw data were read with the ReadAffy function from affy package in R [12], followed by the implementation of Single Channel Array Normalization (SCAN) [13] to normalize the pooled dataset. Default parameters were used. Integrating microarray datasets from various studies will inevitably introduce batch effects. Thus, we corrected the normalized data for batch effects using ComBat [14], and defined the unique GSE ID as the primary source of batch effects.

The data were then filtered by removing control probes as well as averaging all probesets that interrogate the same gene. We also filtered genes that had little variation in gene expression across all pooled samples using the varFilter from geneFilter package. These data were used to generate the Principal Component Analysis (PCA) plots.

Finally, we grouped genes that show high correlation in the microarray dataset using a 0.95 score cutoff. This step was carried out with the findCorrelation from caret R package [15]. Details and quality control plots can be found in the File S1 – Section 2.

Feature Selection

We tested various models with built-in feature selection on our dataset, and we also tested the final predictive accuracy. These models include: “glmnet”, “pam”, “gaussprLinear”, “svmLinear”, “svmPoly”, “svmRadial”, “nb”, “rf”, “knn” and “J48”. We optimized the parameters for each method through the train function of caret R package [15], using Leave One Out Cross Validation (LOOCV) and the accuracy as the readout score. The models were further tested against scrambled-labels datasets. Relevant genes were ranked with t-test statistic as the linear models performed best. We analyzed this set of genes by Gene Set Enrichment Analysis (GSEA) with GOrilla [16] as well as pathway analysis with Ingenuity Pathway Analysis (Qiagen).

The features were shortlisted through recursive feature elimination and backward selection with rfe caret function. The models were evaluated with 4-folds cross-validation, iterated 100 times. The model accuracy increases logarithmically with the number of features. To avoid overfitting the model, we selected the number of features at the point of diminishing returns, which is 5% below the maximum accuracy score.

Two-way ANOVA analysis

We conducted a standard two-way ANOVA analysis stratifying for developmental stage and treatment (i.e. morphants and controls). Both variables were converted into categorical variables. To determine developmental stage, we followed ZFIN's guidelines.

Agilent microarray processing and analysis

A total of 34 microarray studies were done on a variety of Agilent microarray platforms. Among these studies, 12 were rejected due to the same reasons mentioned for the rejected Affymetrix studies. A further 5 raw datasets were not usable for a standard pipeline because they were uploaded to GEO in a wrong file format or were not readable. 1 other dataset could not be analyzed as the experimental design is insufficient to run statistical tests. The R package limma [17] was used to process the raw intensity files as well as downstream analyses.

Briefly, the raw files were read and compiled into a standard “eset” object for a standard normalization pipeline and statistical analyses [18]. Both dual- and single-channel arrays were first background corrected using the “normexp” method [19], followed by within-array normalization by the “loess” method to correct for dye biases in dual-channel arrays [18]. Finally, between-array normalization was performed for both microarray configurations by the “quantile” method [20]. Control probes were then removed prior to applying the empirical Bayesian statistical analyses [21]. We extracted the average log₂-transformed fold-change (logFC) and standard errors for all probes interrogating *isg15*, *isg20*, *casp8*, *mdm2*, *gadd45aa* and *tp53*. As some genes were interrogated by multiple probes, we took the probe that detected the maximum absolute logFC.

Ethics Statement

All zebrafish husbandry was performed under standard conditions in accordance with institutional (MPG) and national ethical and animal welfare guidelines.

Zebrafish

Transgenic line used in this study:

Tg(kdrl:EGFP)^{s843Tg} [22]

Zebrafish micro-injection and Real-time PCR (qPCR) assay

The three morpholinos used in this study are *egfl7* (5'-CAGGTGTGTCTGACAGCAGAAAGAG-3') [23], *vegfaa* (5'-GTATCAAATAAACAACCAAGTTCAT-3') [24] and standard control morpholino (5'-CCTCTTACCTCAGTTACAATTTATA-3') (Gene Tools). Each morpholino was injected into 1-cell stage *Tg(kdrl:EGFP)^{s843Tg}* zebrafish embryos (1 nL volume). The morpholino doses used are as follows: *egfl7* - 0.0625, 0.125, 0.25, 0.5, 1 and 2 ng; *vegfaa* - 0.5, 1, 2, 4 and 8 ng; control morpholino - 0.5, 1, 2, 8 and 16 ng. We also injected the buffer alone to serve as a blank control. Each sample was collected at 32 hours post-fertilization (hpf). 5 embryos were pooled into one biological replicate, and a total of 4 biological replicates was obtained for each sample. We extracted RNA with TRIzol® (Ambion). The aqueous phase was passed through a RNA purification column (Zymogen) and treated with RQ1 DNase (Promega). Samples were re-extracted by acid phenol:chloroform (Ambion) followed by chloroform back-extraction. Biological replicates with poor RNA quality or insufficient RNA amount were discarded. Nevertheless, each group had at least 3 biological replicates. 150 ng of RNA from each sample was converted to cDNA by High Capacity RNA-to-cDNA Kit (Applied Biosystems). Primer sequences and optimized cycling conditions for *rpl13*, *gapdh*, *isg20*, *isg15*, *casp8*, *phlda3* and *tp53/113* can be found in Table S1.

Statistical Analyses of qPCR assays

All statistical analyses were performed with R statistical software. All C_T values were normalized to *gapdh* and compared to uninjected samples ($-\Delta\Delta C_T$). Morpholino dose is transformed to a log-scale (base 2). A dose-response curve test was applied using the *LL.4* function from drc package in R to calculate the upper limit, lower limit and the slope. Otherwise, linear plots were tested using linear regression to estimate the slope. For plotting of graphs, both the in-built plot function of R as well as ggplot2 package [25] were utilized.

Whole mount *in situ* hybridization (WISH)

We injected 0.5 ng of *egfl7* or control morpholino into 1-cell WT embryos (ABs). Embryos were collected at 32 hpf and fixed with 4% PFA in 1X PBST (1X PBS, 0.1% Tween-20), followed by standard WISH protocol [26]. Primers used to clone *isg20* for synthesizing the *in situ* probe are: 5'-GGAGAATCATGGGAAGTGGA-3'; and 5'-GGCATTGAGGTTGGCAGTAT-3'.

Results

Integration of Affymetrix microarray gene expression datasets

We retrieved 63 microarray gene expression datasets from Gene Expression Omnibus (GEO) by keyword search. The majority of microarrays used were manufactured by Agilent (n = 34) or Affymetrix (n = 19). Duplicated datasets, datasets with no accompanying publications and datasets in incorrect file formats (Agilent, n = 12; Affymetrix, n = 8) were removed from further analyses. We applied algorithms to eliminate batch effects and other technical noise in order to integrate datasets generated using the Affymetrix microarray platform (n = 11) into one large dataset. Principal component analysis (PCA) of this dataset revealed that developmental stage and tissue source were the main factors driving the clustering in the principal components (PC) 1 and 2, respectively (Fig 1A). In addition, we observed an occasional segregation of morphants (MO) from controls (CO or WT) for a given cluster group (Fig 1B).

Feature selection uncovers activation of ISGs and the p53 signaling pathway

To identify genes differentially regulated between uninjected controls or buffer injected controls (WT) and morphants, we fed the entire dataset into a supervised machine learning algorithm. The model assigns a higher score to a gene if it performs better at predicting morphants. Among the various models of feature selection, we found that the linear feature selection model tended to perform best (Table S2). We then tested the accuracy of the model and short-listed the optimal number of genes that successfully distinguish the morphants from controls (see Materials and Methods). We refer to this set of genes as “top performing genes” (Table S3).

We detected *tp53* as one of these top performing genes upregulated in morphants compared to controls, corroborating previous findings [9]. Surprisingly, we found another gene, *isg20*, which scored higher than *tp53*, raising the possibility that interferon signaling was also activated in morphants. Furthermore, we found biological processes such as apoptosis and various cell stress responses enriched when further analyzing the top performing genes (Fig 2).

ISG and p53 signaling activation in morphants are most pronounced at the segmentation stage

As mentioned earlier, the samples can be separated by developmental stage along the first PC, as expected by the dynamic nature of gene expression in the developing zebrafish embryo [27-28]. Thus, we proceeded to analyze our dataset by stratifying based on developmental stage. Using a two-way ANOVA approach, we identified 157 genes that were significantly ($\alpha = 0.001$) differentially regulated between morphants and controls, depending on developmental stage (S4 Table). Interestingly, we found that *isg15*, *isg20*, *casp8*, *mdm2*, *gadd45aa*, *phlda3* and *tp53* were profoundly upregulated during segmentation stages in morphants compared to controls (Fig 3). Interestingly, Robu and colleagues similarly reported the activation of *isg20l2*, *phlda3* and *casp8* in their microarray assay (Fig 9C in [9]). Genes that are differentially regulated at other developmental stages can be found in the File S1 – Section 3.

Analysis of Agilent microarray datasets confirms the activation of ISGs and the p53 signaling pathway in morphants

The Agilent microarray datasets were composed of different platform configurations which makes data integration challenging and could potentially introduce artefacts into the post-

processed data. Therefore, we individually normalized each microarray dataset to investigate the expression levels of *isg15*, *isg20*, *casp8*, *mdm2*, *gadd45aa* and *tp53*. Indeed, these genes are similarly upregulated in some morphants compared to control embryos, and many of these morphants were studied at segmentation and pharyngula stages (Fig 4).

ISGs are upregulated by increasing doses of some morpholinos

We injected increasing doses of *egfl7*, *vegfaa* and standard control morpholinos, followed by qPCR assays of *isg20*, *isg15*, *tp53Δ113* [9], *phlda3* and *casp8*. We found that these genes were upregulated by increasing doses of *egfl7* morpholino, and in some assayed genes, a sigmoid curve was observed (Fig 5). In contrast, these genes were not upregulated by increasing doses of *vegfaa* or standard control morpholinos (Fig 5). However, when we injected 8 or 16 ng of standard control morpholino, not only did the injected embryos exhibit upregulation of the aforementioned genes (Fig 5), but these embryos also exhibited a delay in angiogenesis (Fig S1). We performed whole-mount *in situ* hybridization to investigate the *isg20* expression pattern in *egfl7* morphants, and found that *isg20* was upregulated throughout the embryo (Fig S2), but not restricted to the vascular endothelium which expresses *egfl7* specifically [23].

Discussion

Morpholinos are widely used as a tool to inactivate gene and non-coding RNA function. However, we do not yet fully understand the responses of a cell to morpholinos. Indeed, several reported side-effects of morpholinos, including *tp53* upregulation [9], mis-splicing of transcripts and induction of the innate immune response [10], could complicate conclusions arising from morpholino studies. Similarly, our current study identified genes related to the innate immune response which are upregulated across multiple zebrafish morphants as well as zebrafish embryos injected with an excess of standard control morpholino.

Apart from the p53 signaling pathway [9], we further identified the activation of ISGs. *isg20* was previously shown to be induced by the presence of dsRNA [29-30]. dsRNA can be recognized by receptors including Toll-Like Receptor 3 (TLR3), RIG1 and MDA5, which trigger downstream interferon type-I (IFN-I) signaling [31-33]. Interestingly, we detected activation of the IL-8 signaling pathway in our top upregulated genes (data not shown), suggesting the activation of TLRs and IFN-I, possibly by an RNA-morpholino heteroduplex.

Large-scale gene expression analyses have been used to identify gene expression signatures that correlate to tumor progression, and with thorough validation, have become diagnostic markers and/or drug targets [34-35]. Moreover, public databases, such as Oncomine [36], have been established to mine out markers across multiple studies. We applied a similar strategy to multiple gene expression datasets from zebrafish morpholino studies in order to identify gene expression signatures that are common to morpholino-injected zebrafish, and found that many genes in the innate immunity pathway are upregulated in morphants.

Additional work will be needed to understand the underlying biology of these genes and how they can affect morpholino experiments. Clearer understanding of morpholino effects, should help scientists make informed decisions when conducting morpholino experiments (see also latest guidelines in ref. 37), and to facilitate interpretation of data arising from these experiments.

Acknowledgements

We would like to thank Sven Reischauer, Oliver Stone, and Hans-Martin Maischein for their insights regarding morpholinos, Mario Looso and his team for feedback and suggestions on bioinformatics, Pedro Moura for insightful discussions regarding dsRNA induction of IFN-I, and Andrea Rossi, Zacharias Kontarakis and Albert Wang for comments on the manuscript.

Funding

The Stainier lab is supported in part by the Max Planck Society, the European Union (ERC), and the German Research Foundation (DFG).

References

1. Summerton J. Morpholino antisense oligomers: The case for an RNase H-independent structural type. *Biochim Biophys Acta - Gene Struct Expr.* 1999;1489(1):141–58.
2. Nasevicius a, Ekker SC. Effective targeted gene “knockdown” in zebrafish. *Nat Genet.* 2000;26(2):216–20.
3. Draper BW, Morcos PA, Kimmel CB. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: A quantifiable method for gene knockdown. *Genesis.* 2001;30: 154–156.
4. Choi W-Y, Giraldez AJ, Schier AF. Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science.* 2007;318: 271–4.
5. Ekker SC. Morphants: a new systematic vertebrate functional genomics approach. *Yeast.* 2000;17(4):302–6.
6. Stainier DYR, Kontarakis Z, Rossi A. Making sense of anti-sense data. *Dev Cell.* Elsevier Inc.; 2015;32: 7–8.
7. Kok FO, Shin M, Ni C-W, Gupta A, Grosse AS, van Impel A, et al. Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. *Dev Cell.* Elsevier Inc.; 2015;32(1):97–108.
8. Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature.* 2015
9. Robu ME, Larson JD, Nasevicius A, Beiraghi S, Brenner C, Farber SA, Ekker SC. P53 Activation By Knockdown Technologies. *PLoS Genet.* 2007;3: 787–801.

10. Gentsch GE, Spruce T, Monteiro RS, Owens NDL, Martin SR, Smith JC. Innate Immune Response and Off-Target Mis-splicing Are Common Morpholino-Induced Side Effects in *Xenopus*. *Dev. Cell*. 2018;44(5):597-610.
11. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–7.
12. Gautier L, Cope L, Bolstad BM, Irizarry R a. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307–15.
13. Piccolo SR, Sun Y, Campbell JD, Lenburg ME, Bild AH, Johnson WE. A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics*. Elsevier Inc.; 2012 Dec;100(6):337–44.
14. Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–27.
15. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw*. 2008;28(5):1–26.
16. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10(1):48.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47.
18. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods*. 2003;31(4):265–73.
19. Ritchie ME, Silver J, Oshlack a., Holmes M, Diyagama D, Holloway a., et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics*. 2007;23(20):2700–7.

20. Bolstad BM, Irizarry R., Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–93.
21. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat Appl Genet Mol Biol*. 2004;3(1):1–25.
22. Jin S-WS-W, Beis D, Mitchell T, Chen J-N, Stainier DYR. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development*. The Company of Biologists Ltd; 2005;132: 5199–5209.
23. Parker LH, Schmidt M, Jin S-W, Gray AM, Beis D, Pham T, et al. The endothelial-cell-derived secreted factor Egfl7 regulates vascular tube formation. *Nature*. 2004;428(6984):754–8.
24. Nasevicius a, Larson J, Ekker SC. Distinct requirements for zebrafish angiogenesis revealed by a VEGF-A morphant. *Yeast*. 2000;17(4):294–301.
25. Wickham H. *ggplot2: elegant graphics for data analysis*. Springer New York; 2009.
26. Thisse C, Thisse B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc*. 2008 Jan;3(1):59–69.
27. Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J. The zebrafish transcriptome during early development. *BMC Dev Biol*. BioMed Central Ltd; 2011;11(1):30.
28. Mathavan S, Lee SGP, Mak A, Miller LD, Murthy KRK, Govindarajan KR, et al. Transcriptome Analysis of Zebrafish Embryogenesis Using Microarrays. *PLoS Genet*. 2005;1(2):e29.
29. Espert L, Rey C, Gonzalez L, Degols G, Chelbi-Alix MK, Mechti N, et al. The exonuclease ISG20 is directly induced by synthetic dsRNA via NF-kappaB and IRF1 activation. *Oncogene*. 2004;23(26):4636–40.

30. Horio T, Murai M, Inoue T, Hamasaki T, Tanaka T, Ohgi T. Crystal structure of human ISG20, an interferon-induced antiviral ribonuclease. *FEBS Lett.* 2004;577(1-2):111–6.
31. Alexopoulou L, Holt AC, Medzhitov R, Flavell RA. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature.* Macmillan Magazines Ltd.; 2001;413: 732.
32. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol.* 2004;5(7):730–7.
33. Kang D, Gopalkrishnan R V, Wu Q, Jankowsky E, Pyle AM, Fisher PB. mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A.* 2002;99(2):637–42.
34. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer.* 2007;7: 545–53.
35. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer.* 2006;6: 857–66.
36. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia.* 2004;6: 1–6.
37. Stainier DYS, Raz E, Lawson ND, Ekker SC, Burdine RD, Eisen JS, et al. Guidelines for morpholino use in zebrafish. *PLoS Genet.* 2017;13: 6–10.

Figure Legends

Fig 1. Principal component analysis (PCA) of integrated gene expression data compiled from morpholino studies in zebrafish. (A) Gene expression patterns are clustered by developmental stage along PC1 and clustered by tissue source along PC2. (B) Some morpholino experiments show morphants clustering marginally away from controls (demarcated in red).

Fig 2. Gene set enrichment analysis. Biological processes such as apoptosis and the p53 signaling pathway are enriched in the top performing genes (S3 Table).

Fig 3. Relative fold change of gene expression plotted against developmental stage. Genes from the p53 signaling pathway (*tp53*, *mdm2*, *gadd45aa*, *phlda3*), the caspase cascade (*casp8*), as well as interferon stimulated genes (*isg15*, *isg20*) are upregulated in morphants at the segmentation stage. Morphants (MO) were compared to uninjected, buffer-injected or control morpholino (CTRL), depending on the design of the experiment from the microarray datasets.

Fig 4. Analysis of *isg20*, *isg15*, *tp53*, *casp8*, *phlda3*, *gadd45aa* and *mdm2* expression from Agilent microarray datasets. Differential expression of *isg20*, *isg15*, *tp53*, *casp8*, *phlda3*, *gadd45aa* and *mdm2* between morphants and controls. Red indicates significant upregulation ($P < 0.05$); green indicates significant downregulation ($P < 0.05$). The darker shade of red/green indicates $\log_{2}FC > 1$.

Fig 5. ISGs and stress response genes are upregulated in a morpholino dose-dependent manner. qPCR assays for *isg20*, *isg15*, *tp53*, *phlda3* and *casp8* expression in 32 hpf

zebrafish embryos injected with increasing doses of standard control morpholino, *egfl7* morpholino and *vegfaa* morpholino. N = 4 per morpholino per dose. m - slope; LL - lower limit of sigmoid curve ; UL - upper limit of sigmoid curve. *P*-values for individual parameters in parantheses. Grey shaded region reflects the 95% confidence band of the curve.

Supporting Information

Fig S1. 36 hpf zebrafish embryo trunks. Micrographs of zebrafish trunks marked by an endothelial transgenic marker, and overlays on the bright field channel. Embryos injected with 8 ng control morpholino exhibit a mild delay in intersegmental vessel sprouting, and this delay is more pronounced in embryos injected with 16 ng control morpholino.

Fig S2. Whole mount *in situ* for *isg20* expression in *egfl7* morphants and controls.

Table S1. Primers for qPCR assays of *isg20*, *isg15*, *tp53A113*, *casp8*, *phlda3* and *gapdh* expression levels.

Table S2. Prediction accuracy of various machine learning methods.

Table S3. Top performing genes ranked by t-test statistics and linear models.

Table S4. Genes that are significantly differentially regulated in a two-way ANOVA analysis.

File S1. Supplementary Information.

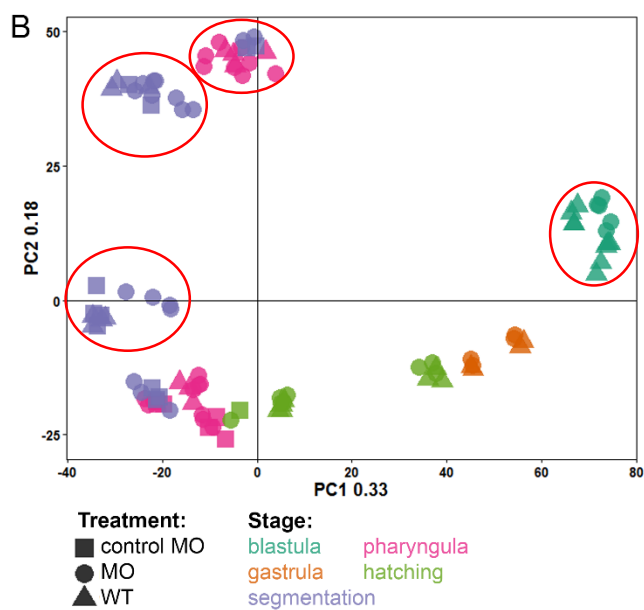
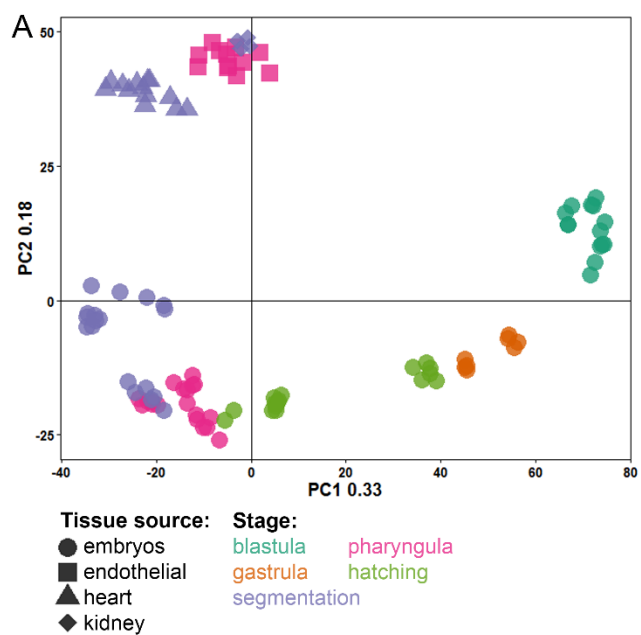


Fig 1

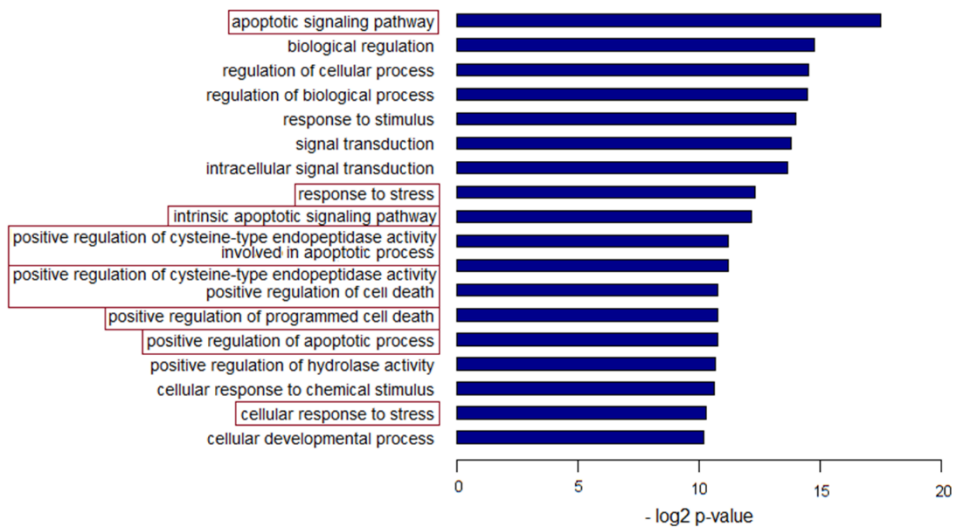


Fig 2

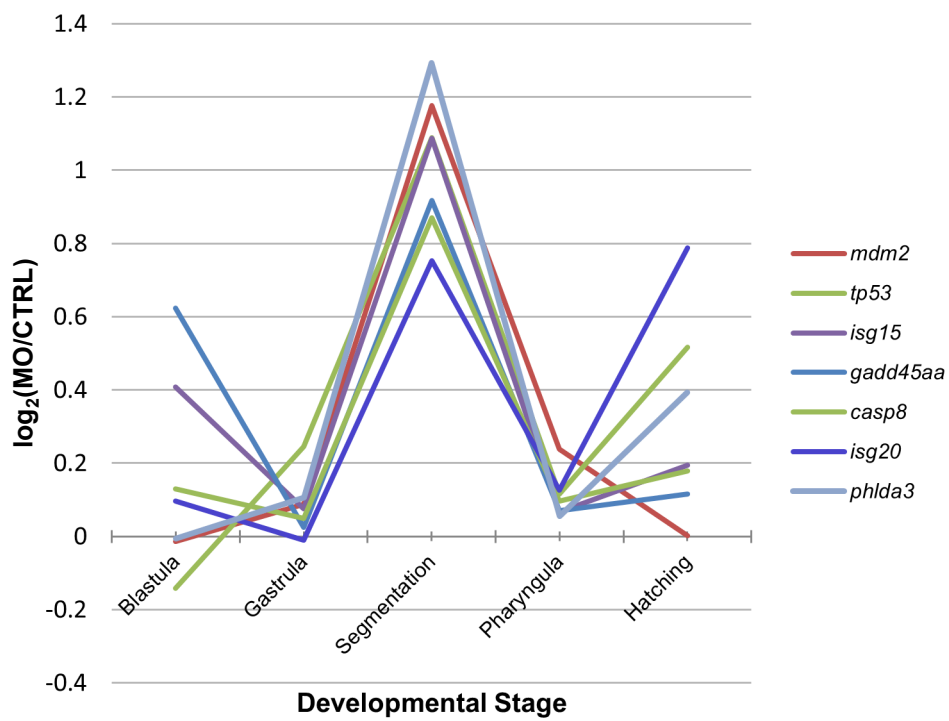


Fig 3

GSE	Morpholino	Control	Developmental Stage	<i>isg20</i>	<i>isg15</i>	<i>tp53</i>	<i>casp8</i>	<i>phlda3</i>	<i>gadd45</i> <i>aa</i>	<i>mdm2</i>
GSE21539	<i>ovol1</i>	WT	Segmentation							
GSE27569	<i>esco2</i>	WT	Segmentation							
GSE27569	<i>esco2</i>	WT	Pharyngula							
GSE34930	<i>ptpn6</i>	WT	Pharyngula							
GSE37332	<i>mitfa</i>	WT	Hatching							
GSE50376	<i>nr3c1</i>	controlMO	Segmentation							
GSE50376	<i>nr3c1</i>	controlMO	Pharyngula							
GSE13157	<i>esr2a</i>	WT	Hatching							
GSE13158	<i>esr2a</i>	WT	Hatching							
GSE38441	<i>grna</i>	WT	Segmentation							
GSE38441	<i>grna</i>	WT	Segmentation							
GSE38441	<i>grna</i>	WT	Pharyngula							
GSE38441	<i>grna</i>	WT	Hatching							
GSE42070	<i>pcsk7+tp53</i>	controlMO	Gastrula							
GSE42070	<i>pcsk7+tp53</i>	controlMO	Segmentation							
GSE64426	<i>tbx5a</i>	controlMO	Pharyngula							
GSE64426	<i>tbx5a</i>	controlMO	Segmentation							
GSE61625c	<i>dio3b</i>	controlMO	Hatching							
GSE61625a	<i>dio3b</i>	controlMO	Hatching							
GSE61625b	<i>dio3b</i>	controlMO	Hatching							
GSE61625a	<i>dio1/2</i>	controlMO	Hatching							
GSE61625b	<i>dio1/2</i>	controlMO	Hatching							
GSE61625c	<i>dio1/2</i>	controlMO	Hatching							
GSE17773	<i>arrb1/2</i>	controlMO	Segmentation							
GSE57836	<i>slc2a2</i>	controlMO	Hatching							
GSE16740	<i>tnnt2</i>	controlMO	Pharyngula							
GSE16740	<i>tnnt2</i>	controlMO	Pharyngula							
GSE16740	<i>tnnt2</i>	controlMO	Hatching							
GSE54754	<i>gba</i>	controlMO	Pharyngula							
GSE51541	<i>atg5</i>	WT	Hatching							
GSE51541	<i>becn1</i>	WT	Hatching							
GSE51541	<i>atg7</i>	WT	Hatching							
GSE18830	<i>sox2/3/19a/19b</i>	WT	Blastula							
GSE18830	<i>sox2/3/19a/19b</i>	WT	Gastrula							
GSE18830	<i>sox2/3/19a/19b</i>	WT	Gastrula							
GSE32594	<i>nfe2l2a</i>	controlMO	Hatching							
GSE32594	<i>nfe2l2b</i>	controlMO	Hatching							
GSE32594	<i>nfe2l2a+nfe2l2b</i>	controlMO	Hatching							
GSE19206	<i>spi1b</i>	WT	Pharyngula							
GSE45011	<i>klf2a</i>	controlMO	Blastula							
GSE45011	<i>klf2b</i>	controlMO	Blastula							
GSE52229	<i>ahrra</i>	WT	Hatching							
GSE58683	<i>aplnra/b</i>	WT	Blastula							
GSE52229	<i>ahrrb</i>	WT	Hatching							
GSE12012	<i>miR-126_MO1</i>	WT	Pharyngula							
GSE12012	<i>miR-126_MO2</i>	WT	Pharyngula							
GSE46844	<i>miR-34b</i>	controlMO	Hatching							
GSE34508	<i>slc2a10</i>	controlMO	Pharyngula							
GSE32914	<i>sox31</i>	WT	Blastula							
GSE20179	<i>tnnt2</i>	controlMO	Larval							
GSE63360	<i>nr3c1_MO1</i>	controlMO	Pharyngula							
GSE63360	<i>nr3c1_MO2</i>	controlMO	Pharyngula							

Fig 4

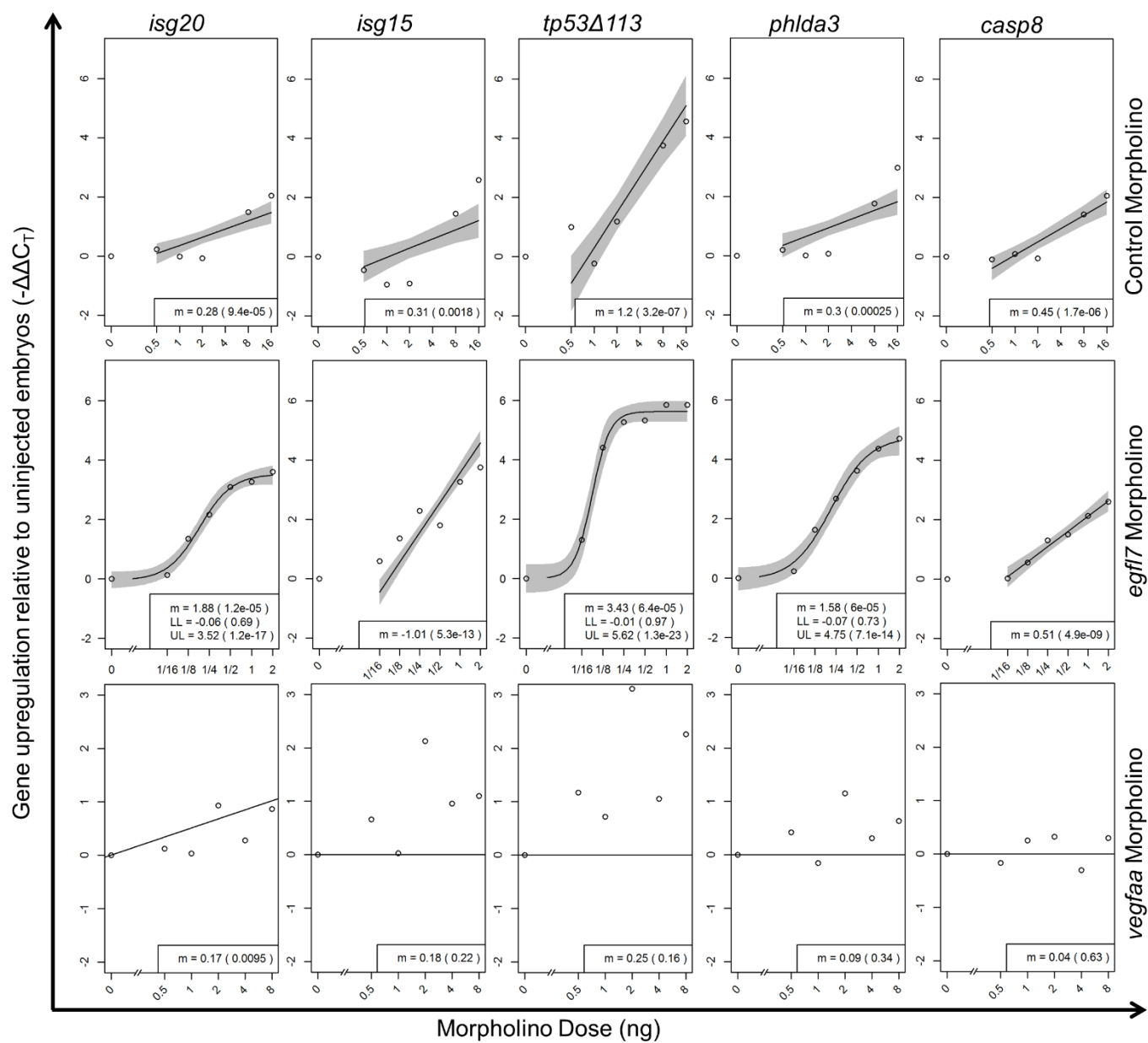


Fig 5