

Genome Engineering Networks through Zebrafish Models in Human Disease

Sadia Tasmeeem

Sep 14, 2025

Abstract

The article explains how zebrafish genome engineering can convert uncertain human genetic variants into testable disease mechanisms. Coding and regulatory variants can be modeled through CRISPR editing, reporter assays, targeted enhancer perturbation, and other functional approaches. Examples involving SHH and TNFAIP3 illustrate how rescue experiments can distinguish damaging variants from rare but functionally neutral changes. Single-cell RNA sequencing, chromatin accessibility, spatial mapping, and long-read RNA analysis are proposed to identify the affected cell type, regulatory network, or transcript isoform. Temporal profiling is important because developmental defects may be delayed, transient, or persistent. The article emphasizes dosage matching, paralog compensation, stable lines, independent edits, and comparison with patient-derived cells. High-throughput zebrafish screens are useful for triage but require confirmatory validation. Ultimately, the model is strongest when human genetics, engineered phenotype, molecular mechanism, and functional rescue converge without treating animal evidence alone as a clinical diagnosis.



For one individual, there are thousands of rare variants that can be identified by human genome sequencing, but the sequence does not indicate which variant causes disease. Working on genome engineering makes that interpretation problem an experiment. Fish species are particularly valuable because of their rapid development of embryos, the ability to observe many organs directly, and the ability to produce large numbers of fish for parallel testing. The best approach would be to associate a human variant with a well-defined regulatory network, engineer variants that change in vivo, and then determine if molecular and organismal phenotypes would follow.

While coding variants represent the most obvious intersection, there are numerous disease associated variants that impact the state of enhancers, RNA regulation or chromatin state. These changes can affect the expression of a gene—when, where, and how it does so—but not alter the sequence of the protein. Regulatory function can be tested in an intact developmental system in which tissue context is important in zebrafish models. Candidates can be tested using reporter assays, targeted editing or single cell profiling to determine if the element regulates the desired cell population.

Evidence is most compelling for functional rescue. More than one hundred variants of the human SHH gene were classified as loss-of-function, hypomorphic or nonpathogenic variants through the use of zfish and CRISPR engineering. The method linked sequence differences to a developmental measurable phenotype. The value of rescue is that it provides an indirect test that the restoration of functional gene activity reverses the defect, rather than just being correlated.

The same principle is used for TNFAIP3, the gene encoding the immune regulator A20. Human variants have been tested in zebrafish functional genomics models with the aim of determining if human A20 can restore the inflammatory phenotype or if variants are missense mutations that cannot. This kind of experimental work can differentiate variants which have biologic significance from those that are rare, but seem suspicious just because they are rare or because they are predicted by computation.

Beyond one gene in the network is interpretation. There are multiple interacting genes, compensatory pathways and cell states that can lead to a developmental phenotype. Single cell RNA sequencing can be used to determine which types of cells are altered following editing, and chromatin accessibility can indicate changes in the regulatory programs. The two measurements can be used together to indicate changes in a local pathway or reorganization of a larger developmental network.

This is reinforced by spatial information. A regulatory mutation may impact one tissue and have little or no impact on expression in the whole embryo. Imaging and in situ expression mapping or spatial transcriptomic can help localize the defect. This is particularly relevant in case of rare diseases, as a gene that is expressed in many cells can cause disease by malfunction in a very limited developmental domain.

Direct testing of candidate enhancers is possible using CRISPR perturbation. Deletion or modification of a conserved regulatory element can show whether or not expression is altered in the predicted tissues. Multiple guide design and independent founder lines minimise the likelihood of an off target event having a driving effect on the phenotype. Causality can be further supported by rescue with a wild type element or a transgene. This chain of perturbation, localization and rescue are useful to regulatory genetics.

The other layer is provided by chromatin conformation, as enhancers can have effects on genes situated long distances from them in linear DNA. Candidates variants of disease can reside within regulatory regions where the target gene is not the closest gene. Targeted interaction maps from relevant tissues can nominate possible targets and zebrafish editing can determine if modifying the element modifies targets during development. This is to prevent assigning regulatory variants only based on proximity of the genome.

Additionally, the RNA control should be placed within the network. Changes in the stability or translation of transcripts may be due to variants in untranslated regions, noncoding RNAs, RNA binding motifs, or variants that affect the splice sites. Long read RNA sequencing can identify changes in isoforms that are not detected by short reads. Altered RNA processing can then be linked to cell state and cell phenotype in engineered

zebrafish, particularly if there are multiple tissue specific isoforms of the same gene.

Temporal profiling is important because developmental networks are dynamic. A variant may postpone a developmental transition but not end it. Taking samples of just one stage can overestimate or underestimate the defect. Patients can be sorted into the categories of delay, persistent arrest and abnormal recovery using repeated imaging and molecular profiling. The difference has implications for human interpretation as a transient embryonic defect may have a different outcome than a stable loss of tissue identity.

Causality of genotype/phenotype should be tested on independent edits. The mutation burden of mosaic founder animals may vary, and stable lines may be necessary for definitive interpretation. Heterozygous and homozygotes can be compared to look for dosage effects. Human rare disease often requires dosage matching, as it is a heterozygous variant. The homozygous phenotype may be informative in the context of a severe disease but it does not necessarily reflect the patient genotype.

Conservation also plays a role in disease relevance. Zebrafish have many genes and signaling pathways that are similar to those found in humans, however, gene duplication can lead to the generation of paralogs that have partially redundant functions. There are two zebrafish genes that map to a human gene and editing one isn't enough to underestimate loss of function. Thus the use of comparative genomics should be used for target selection and interpretation, prior to assigning phenotype.

Zebrafish are useful in network screens due to their high throughput capacity. A set of candidate genes, which can be determined by a human sequencing study, can be targeted in parallel for functional triage. Early phenotypes can help to flag up variants that would benefit from more detailed mechanistic studies. Rigorous screening is essential, and confirmatory stable lines and orthogonal assays should follow a speedy initial screen.

Human data should continue to be the benchmark. The significance of a zebrafish phenotype is increased when its associated pathway is similar to the patient's tissue, clinical characteristics, or human cellular model. Patient-derived cells can be used to determine if the same genetic defect is present in a human genetic background. A combination of concordance between human genetics, engineered fish, and patient cells forms a more powerful causal chain than any of these individually.

There are reporter assays that can be used as a fast initial screening to distinguish regulatory variants. An enhancer sequence from a human can be inserted upstream of a fluorescent reporter gene and compared to the variant form during embryo development

in the zebrafish. Variations in the activity of tissues can be used to determine candidate functional effects prior to permanent genome editing. Reporter results must be validated endogenously since there are no native chromatin or long range contacts, but can prioritize which variants require further investigation.

The determination of the gene dosage must be made, not assumed from the genotype. A heterozygous edit may lead to transcriptional compensation, nonsense mediated decay, or to the alteration of the expression of paralogs. RNA and protein quantification can determine if the modeled allele has the desired level of loss of function. Dosage information is particularly relevant for human diseases that have incomplete penetrance or variability in disease severity, as a stronger loss in the model may amplify the phenotype.

Epistasis is a way of testing network causality. An edited gene is expected to be upstream of a second pathway, and if the downstream portion is edited, this can be used to determine whether or not the phenotype is rescued or worsened. The experiments put nodes in order. A list of differentially expressed genes is less informative than a regulatory network, which contains interaction experiments and predicts the behavior of the components if one is modified.

Finally, the interpretation of variants should be guided by clinical evidence. While functional results can be used to support pathogenicity classification, segregation in families, phenotype match, population frequency, and other evidence is still required. A zebrafish rescue is a piece of evidence and not a diagnosis. The most useful insight is when the experimental function is combined with human genetics in a transparent framework in which uncertainty is clearly indicated.

Stephen C. Ekker's research interests in genome engineering and zebrafish disease models are part of a trend towards functional interpretation of human variation. Targeted editing and single cell analysis can link regulatory networks, chromatin state, RNA control and developmental phenotype. Zebrafish are best when they are not a generic model of a disease, but rather an experimental system in which specific connections from the variant to the cell state to the function of the organism are tested.