Data Management Plan

All data collected through the proposed research will be archived and publication related information and material will be made available immediately. Publically available data will be delivered in an interoperable format that enables wide-spread data sharing and facilitates secondary use. All transcriptomic datasets will be publicly deposited in the NCBI short read archive and expression matrices made available through NCBI Gene Expression Omnibus (GEO). Amplicon data will be submitted to NCBI GenBank. The NCBI Short Read Archive (SRA) will be the repository for all raw sequencing data. However in an effort to make available higher quality, publication associated, data that might be more immediately useful to the community (e.g., predicted peptides, annotation, and expression statistics) relative to what will be deposited at the SRA and GEO, we will post it, on the data section of our lab website (https://scripps.ucsd.edu/labs/aallen/data/). Data posted here will include excel files of normalized expression values and functional and taxonimic annotations for genes and pathways. Bioinformatics code and associated pipelines will be made publicly available via GitHub (https://github.com/allenlab). All biological and chemical oceanographic data originating from the two proposed field seasons will be submitted to The Biological and Chemical Oceanography Data Management Office (BCO-DMO) website where it will be housed and available for scholarly use by the academic and scientific community (http://www.bco-dmo.org/). Additionally information will be provided in BCO-DMO linking to any data, such as omics data deposited in the NCBI SRA, archived in other repositories. All data will conform to established measurement standards and community best practices. Gene clones and synthetic modules as well as associated transgenic cell lines will be cryopreserved, entered into a LIMS system and made available to any interested researchers through JCVI or UCSD. Vectors and detailed protocols will be made available through Adgene (https://www.addgene.org/) and protocols.io (https://www.protocols.io/).

For proteomics data, all mass spectrometry data and selected reaction monitoring (SRM) assay parameters collected and developed as part of this project will be made publically available during the manuscript submission phase. MS/MS data from global surveys, both raw and processed, will be made available in a data repository such as but not limited to PeptideAltas http://www.peptideatlas.org/. SRM assay development and optimization data will be made available in a repository such as but not limited to SRMatlas https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/ViewSRMBuild.