Data for this project will consist of a large number of sequences and microarray results. Raw sequences will be submitted to NCBI using GEO, which is specifically for high throughput sequence data generated by a variety of platforms including the 454. Microarray, data such as probe oligomer sequences and annotations, experimental design and results will also be uploaded to GEO. The platforms in GEO are designed to include metadata on the biological samples and protocol descriptions.

Some samples may be obtained from collections by existing sampling programs (e.g., GLOBEC), which are listed in databases such as BCO-DMO. In this case we will provide a cross-reference in this database to NCBI, and vice-versa.