S2

# Functional Genomic Analysis of *C. elegans* Development by Systematic RNAi

Asako Sugimoto

*RIKEN Center for Developmental Biology, Kobe, Japan*

The genome of each organism contains the developmental program by which a fertilized egg develops into an individual animal with a complex body structure. The nematode *Caenorhabditis elegans* is an excellent model organism to study how the information encoded in the genome regulates animal development, because of its complete genome sequence information as well as the availability of powerful genetic techniques and its fully described cell lineage. The genome size of *C. elegans* is 100 Mb and ~20,000 genes are predicted, more than half of which are conserved in metazoans. To understand how these genes in the *C. elegans* genome orchestrate the developmental program, we have been systematically characterizing the function of each gene by double-stranded RNA (dsRNA)-mediated interference (RNAi). For the high-throughput RNAi analysis, we optimized a "soaking method", in which worms are simply soaked in dsRNA solutions and their loss-of-function phenotypes in the next generation are examined. As templates for *in vitro* dsRNA synthesis, a set of tag-sequenced non-redundant cDNAs has been used. Among ~6000 genes thus far examined, ~25% showed detectable phenotypes under a dissection microscope. For the genes resulted in embryonic lethality, the detailed terminal phenotypes have been characterized using differential interference contrast (DIC) microscopy. In addition, potential additional post-embryonic functions of these embryonic lethal genes have been examined by a stage-specific RNAi methodology. More than half of the embryonic lethal genes caused defects in post-embryonic development as well, thus, many essential genes function at multiple developmental stages and/or in multiple tissues. The RNAi phenotypes are recorded using multiple independent criteria (e.g., cell number, cellular morphology, tissue differentiation, body shape...) as well as DIC images, and archived in a database that allows the phenotype search with any combination of phenotypic criteria. Furthermore, this RNAi phenotype dataset can be converted into binary dataset that can then be used for computer-assisted analyses. As one of such computational approaches, we utilized these digitized RNAi phenotype dataset for gene classification, with the aim to predict functions of uncharacterized genes, and to identify genetic networks. Gene clustering based on RNAi phenotypes was performed as follows. First, the RNAi phenotypes described using multiple criteria are converted into binary (0 or 1) data of 40 dimensions. Next, using principal components analysis, dominating features from the original binary data set were separated and the dimensionality of the data were reduced into 14 dimensions. Finally, these condensed dataset are used for a hierarchical cluster analysis and a dendrogram was created. As a pilot case, the terminal RNAi phenotype dataset of 254 embryonic lethal genes was subjected for this analysis. In the resulting dendrogram, genes with similar phenotypes were clustered in an intuitively understandable manner: The genes were classified into three major clusters (Cell division, Cell fate determination, and Morphogenesis), and each cluster was further classified into several sub-clusters. (For example, the "Cell division" cluster consists of the "1-cell arrest" sub-cluster and the "cell size variable"sub-cluster.) Furthermore, genes known to function in the same pathway are closely clustered, confirming the validity of this classification. Thus, the gene clustering based on loss-of-function phenotypes is a powerful method to identify genes involved in the same developmental processes, and will aid the understanding of the genetic program that control development.