Erythroid Cell Lineage Genomic Anatomy Project (ELGAP) Workshop Summary and Recommendations

An NIDDK-sponsored workshop to discuss erythroid cell functional genomics was held December 19-20, 2001. The overall goals of erythroid functional genomics are to understand normal erythropoiesis and the diseases associated with it and to develop novel therapeutic approaches to these diseases. In summary, the workshop highlighted the current state of the field and pointed to specific research areas that need to be addressed.

Topics discussed are as follows:

- **Erythroid Cell Biology**
  - It is now possible to purify progenitor cells that give rise to erythroid cells and megakaryocytes. Likewise, stromal cells can be purified to obtain the "womb" of the emerging, committed erythroid cells. Gene expression profiles have been generated by hybridization to microarrays.
  - It is also possible to identify, purify, and isolate erythroid progenitor cells from the developing mouse and zebrafish embryos. The primitive erythrocytes have unique morphology, physiology, and molecular biology associated with them.
  - Many knockout mouse and zebrafish mutants have been generated, and they have failed to either develop or differentiate normal erythrocyte.

- **Emerging Genomics Technologies**
  - Gene expression profiles of single hematopoietic cells can be determined by hybridization to microarrays. Thus expression profiles can be developed for each stage of hematopoiesis, despite the small numbers of cells that can be purified.
  - Rapid computational analysis is currently available to evaluate changing expression levels during different stages of development or differentiation.
  - Databases are being developed.

- **Emerging Proteomics Technology**
  - Cellular proteins from small numbers of cells may now be analyzed without the use of polyacrylamide gel electrophoresis.
  - An example of a complete proteome atlas was discussed. Currently, this atlas is being developed for *C. elegans.*

- **Bioinformatics and Computational Biology**
  - Databases of gene expression profiles are being generated from hematopoietic cells that precede the committed erythroid cell during differentiation.
  - Databases of gene expression profiles have been generated from erythropoietin-dependent erythroid cells.

- Integration of both genomic and cDNA data sets discovers functionally important biological themes.
- In some well-studied model systems, such as changes in sugar metabolism in yeast or developmentally specific gene expression in sea urchins, mathematical, logical models can be built that describe the changes in terms of regulatory sites in DNA and proteins, as well as physical constants for their interactions. Such models are refined computationally and can be tested experimentally to define regulatory modules.

■ **Evolutionary Biology**

- Comparisons of genomes or sections of genomes have proved to be useful for identifying and characterizing biologically relevant motifs in erythroid cell development and differentiation.

Areas that need further research were identified as

■ **Global, molecular, genetic descriptions of erythroid cells at various stages of development and differentiation.** These descriptions of various stages of development (i.e., embryonic, fetal liver and adult) and differentiation (i.e., committed CFU-E, pronormoblast, orthochromatophilic erythroblast, polychromatophilic erythroblast, reticulocyte and red blood cell) are needed as standards for the entire scientific community. In other words, scientists need to determine (1) which sets of genes are expressed at the various stages of erythropoiesis and (2) the function of the red blood cell.

- Gene expression profiles may be standardized by use of a common set of microarrays. In addition, expression profiles need to be developed in cells with specific gene knockouts or knockdowns.
- Proteomic profiles are also needed, since some proteins are present but not active because of post-translational modification. Also needed are (1) complete analysis of the sub-cellular localization of all proteins in the erythroid proteome, (2) a protein-protein interaction map of the erythroid proteome, and (3) a comprehensive phenotypic map of loss-of-function of genes encoding the erythroid proteome. These results might be stored as maps widely available on the Internet.

■ **Applications of molecular knowledge to human diseases.** Among inherited disorders of the red blood cell, sickle cell anemia and the thalassemias are the most prevalent worldwide. The molecular basis of these diseases is known, but scientists need to learn more about the contribution of other genes to the clinical course of these illnesses and their overall severity.

■ **Computational biologic programs.** Computational biologic programs need to be developed to integrate data from all levels of biology, namely DNA, mRNA, protein-protein interactions, and others. The goal is to develop mathematical models and graphical displays of the red blood cell.