Systems Model for Prognostic Cancer Biology

Xuefei Wang[1], Bai-Lian Li[2], Cheryl L. Willman[3], and Susan R. Atlas[4]

***Short Abstract* — Human cancers display profound biological heterogeneities that belie their strict characterization as organ-based diseases. Identifying underlying disregulated genetic pathways may reveal novel cancer subtypes and patterns of disease evolution, ultimately leading to a unified cancer systems biology transcending tumor site of origin. Toward this goal, we describe a neuro-fuzzy systems engineering model developed for outcome prediction in acute leukemia. We show how gene-patient expression biclusters, identified as fuzzy membership functions in the fuzzy inference model, can be reinterpreted as novel cancer subtypes associated with distinct, competing disease pathways.**

***Keywords* — Biological pathways, machine learning, biclustering, fuzzy rules, neuro-fuzzy inference, gene expression, cancer outcome prediction, acute leukemia.**

I. Introduction

A major challenge in contemporary cancer diagnosis and treatment is the development of systematic methodologies for identifying patterns of disease specific to distinct subgroups of patients, in order to develop tailored therapies that neither undertreat nor induce excessive toxicity [1]. Through genomic profiling in conjunction with clinical data, *molecular taxonomies* have been proposed for various cancers, such as leukemia [2]. In parallel with these efforts, cell line and mouse model experiments have sought to isolate critical mutated pathways characteristic of human cancers in general. These studies are aimed at understanding cancer natural history—from oncogenesis and neoplastic transformation through metastasis—in terms of fundamental cell biological processes and pathways [3-4]. Here we describe an approach that enables these distinct perspectives to coexist within a unified mathematical model.

Acknowledgements: SRA thanks D Dai for many inspiring discussions relating to cancer biology. This work was supported by NIH/NCI grants CA88361 and CA32102, the W. M. Keck Foundation, the Leukemia and Lymphoma Society, and funds from the Dedicated Health Research Fund of the State of New Mexico. We thank the University of New Mexico Center for High Performance Computing (CHPC) and UNM Cancer Center Shared Resource for Bioinformatics and Computational Biology for computational resources, the Children's Oncology Group for clinical data, and the Keck-UNM Genomics Resource for expert microarray support.

[1]UNM Cancer Center and CHPC, Univ. New Mexico, Albuquerque, NM 87131. E-mail: xuefei@hpc.unm.edu

[2]Dept. Botany and Plant Sciences, Univ. California, Riverside, CA 92521. E-mail: bai-lian.li@ucr.edu

[3]UNM Cancer Center and Department of Pathology, Univ. New Mexico, Albuquerque, NM 87131. E-mail: cwillman@unm.edu

[4]Department of Physics & Astronomy and Center for Advanced Studies, UNM, Albuquerque, NM 87131. E-mail: susie@sapphire.phys.unm.edu

II. Methodology

The language of systems engineering—specifically, fuzzy logic—provides a natural framework for developing cancer outcome prediction models that merge clinical data 'expert knowledge' with gene expression-based molecular subtypes. Using data from leukemia patient cohorts, three classes of fuzzy IF-THEN rules have been generated: clinical rules (via survival analysis); global gene expression rules (fuzzy-c-means clustering); and gene expression subtype rules (biclustering algorithm, [5]). These may be used to build and train a neuro-fuzzy classifier for prognostic inference [6].

In the fuzzy classifier, the biclustering-based rules play an *engineering* role, serving as mathematical classes in which patients exhibit varying degrees of fuzzy membership. However, they also possess a complementary *systems biology* interpretation: by correlating coherent gene over- and under-expression patterns with patient subsets and clinical outcomes, they define potentially novel disease subtypes, pathways, and therapeutic targets.

III. Results and Conclusion

We describe the biclustering algorithm and fuzzy rule translation procedure, and highlight two distinct classes of biclusters and corresponding pathways, using data for acute leukemia. The classes correspond to 'hallmark' oncogenic features [3] and phenotypes such as therapeutic response [7].

This work illustrates how mathematically-derived constructs such as gene-patient biclusters can play multiple roles in biological modeling. Designed to improve numerical outcome prediction models, they can also provide valuable insights into the underlying cancer cell biology.

References

[1] Couzin J (2005) To what extent are genetic variation and personal health linked? *Science* **309**, 81.

[2] Willman CL (2004) Discovery of novel molecular classification schemes and genes predictive of outcome in leukemia. *Hematol. J.* **5**, 138–143.

[3] Hahn WC and Weinberg R (2002) Modelling the molecular circuitry of cancer. *Nat. Rev. Cancer* **2**, 331-441.

[4] Atlas SR, Luo W *et al.* (2008) Deconstructing a tumor, reconstructing a history [abstract]. In: *Proceedings of the American Association for Cancer Research Annual Meeting (San Diego, CA),* **49**, 188.

[5] Wang X, Willman CL, and Atlas SR (2008) Robust semi-supervised biclustering for discovery of novel phenotypes in cancer expression data. Ms., to be submitted.

[6] Wang X (2007) *Hybrid neuro-fuzzy inference models for outcome prediction in acute leukemia using gene expression and covariate data*. Ph.D. dissertation, University of New Mexico.

[7] Atlas SR, Wang X, and Willman CL "Identifying novel prognostic subtypes in cancer," in preparation.