Bayesian Network Models of Multiple Myeloma: Drivers of High Risk and Durable Response

Fred Gruber[1], Boris Hayete[1], Jonathan Keats[2], Kyle McBride[3], Karl Runge[1], Mary DeRome[4], Sagar Lonial[5], Iya Khalil, Daniel Auclair[4]

[1]GNS Healthcare, [2]Translational Genomics Research Institute, [3]Instat, [4]Multiple Myeloma Research Foundation, [5]Department of Hematology & Medical Oncology, Emory University School of Medicine.

Introduction

In this study, we use the Multiple Myeloma Research Foundation (MMRF) CoMMpass trial (NCT0145429) Interim Analysis 9 (IA9) dataset of newly-diagnosed multiple myeloma (MM) patients to learn an ensemble of Bayesian networks in order to elucidate important biological mechanism in MM.

We demonstrate that our approach finds a number of known drug targets and identifies potentially novel ones. These targets, in our simulations, affect a number of treatment efficacy outcomes.

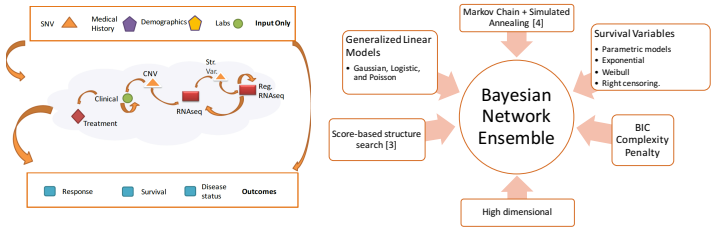
Data Collection

The CoMMpass trial (NCT0145429) started in July 2011 and includes patients from the United States, Canada, and the European Union. For each patient, tumor and matched constitutional samples are analyzed with Long-insert Whole Genome Sequencing, Whole Exomme Sequencing, and RNAseq. Clinical parameters are colected at study entry and every three months for a minimum of 5 years.

After integrating the different data modalities (somatic single nucleotide variants (SNV), structural variants, somatic copy numbers (SCNV), and RNAseq gene expression) we obtained a table with 645 patients samples, and 30426 variables that completely characterize the molecular and clinical traits of each patient.

Methodology For Analysis

We have applied REFS™ [2], our state-of-the-art Bayesian causal inference engine, to reverse-engineer the molecular pathways that most likely affect treatment.



- Network structures are constrained to a minimal set of biological relations, but are otherwise *de novo*.
- To account for intrinsic uncertainty of the structure, we used REFS to learn an ensemble of 256 networks representing a statistical sample of the most probable structures that explain the data.

"Treatment" Effects

- The resulting Bayesian network model is interrogated by exhaustively enumerating the effects of perturbations of the variables in the model upon outcomes for each of the networks in the ensemble [5].
- This approach could simulate, for example, a gene knockdown experiment, a change in drug treatment, or an imposition of an enrollment criterion upon a clinical study.

Selected Findings

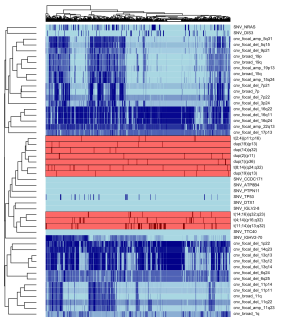
- **General Network Properties**

Important Variables According to Out-Degree

Variable	Description	Type
WHSC1	Wolf-Hirschhorn syndrome candidate 1 location: 4p16.3	Protein Coding Gene Expression
LIK3	unc-51 like kinase 3 location: 15q24.1	Protein Coding Gene Expression
KMT2D	lysine (K)-specific methyltransferase 2D location: 12q13.12	Protein Coding Gene Expression
FLI1	Fli-1 proto-oncogene, ETS transcription factor location: 11q24.3	Protein Coding Gene Expression
NFKB2	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100) location: 10q24.32	Protein Coding Gene Expression
LHX4	LIM homeobox 4 location: 1q25.2	Protein Coding Gene Expression
RELA	v-rel avian reticuloendotheliosis viral oncogene homolog A location: 11q13.1	Protein Coding Gene Expression
MIR1302-9	microRNA 1302-9 location: 9p24.3	NonCoding RNA Expression
MIR4453	microRNA 4453 location: 4q31.3	NonCoding RNA Expression
RP11-108N6.5	ENSID: ENSG00000258776 type: lincRNA location: 14q22.3	NonCoding RNA Expression
RP5-1057J7.6	ENSID: ENSG00000261326 type: lincRNA location: 1p36.12	NonCoding RNA Expression
SNV_IGHV2-70	Weighted burden score SNV in immunoglobulin heavy variable 2-70 location: 14q32.33	Somatic SNV
SNV_DIS3	Weighted burden score SNV in DIS3 mitotic control homolog (S. cerevisiae) location: 13q22.1	Somatic SNV
SNV_IGLV2-8	Weighted burden score SNV in immunoglobulin lambda variable 2-8 location: 22q11.22	Somatic SNV
SNV_NRAS	Weighted burden score SNV in neuroblastoma RAS viral (v-ras) oncogene homolog location: 1p13.2	Somatic SNV
SNV_ATP8B4	Weighted burden score SNV in ATPase, class I, type 8B, member 4 location: 15q21.2	Somatic SNV
SNV_CCDC171	Weighted burden score SNV in coiled-coil domain containing 171 location: 9p22.3	Somatic SNV
SNV_PTPN11	Weighted burden score SNV in protein tyrosine phosphatase, non-receptor type 11 location: 12q24.13	Somatic SNV
SNV_TP53	Weighted burden score SNV in tumor protein p53 location: 17p13.1	Somatic SNV
SNV_TTC40	Weighted burden score SNV in tetratricopeptide repeat domain 40 location: 10q26.3	Somatic SNV
SNV_DTX1	Weighted burden score SNV in deltex homolog 1 (Drosophila) location: 12q24.13	Somatic SNV
cnv_focal_amp_15q24	Somatic CN: Focal amplification in 15q24	Somatic CNV
cnv_broad_15q	Somatic CN: Arm level event at 15q	Somatic CNV
cnv_focal_amp_11q23	Somatic CN: Focal amplification in 11q23	Somatic CNV
cnv_focal_del_3p24	Somatic CN: Focal deletion in 3p24	Somatic CNV
t(11;14)(q13;q32)	Translocation t(11;14)(q13;q32)	Somatic Structural Variants
dup(1)(p13)	Tandem duplication dup(1)(p13)	Somatic Structural Variants
t(4;14)(p16;q32)	Translocation t(4;14)(p16;q32)	Somatic Structural Variants
dup(1)(q32)	Tandem duplication dup(1)(q32)	Somatic Structural Variants
t(14;16)(q32;q23)	Translocation t(14;16)(q32;q23)	Somatic Structural Variants

A sample of topgenes, SNV, CNV, and structural variants with large out-degree.

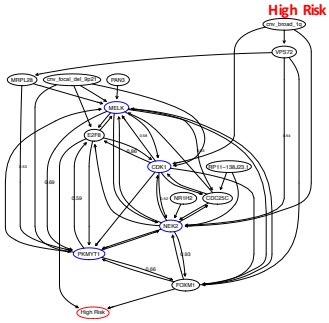
Genetic Stratification of Tumor Samples



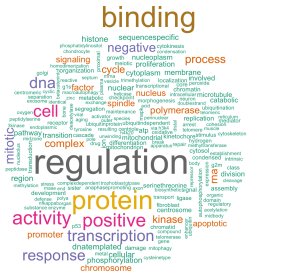
Bi-clustering of the somatic genetic variants with the largest out degree. Distance weighted by out degree.

- **Drivers of High Risk and Durable Response**

High Risk

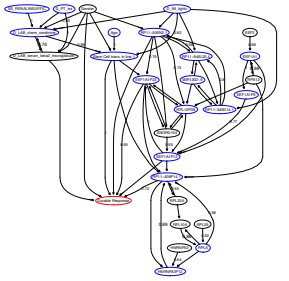


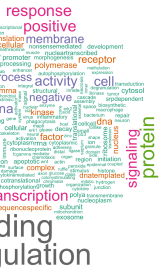
Drivers of outcomes. The blue nodes represent strong drivers of the outcome (red node). The width of edges are a function of the frequency of this edge among all networks in the ensemble and is given by the number next to the edge if it is larger than 50%.



GO terms associated with genes that are causal drivers of the outcome in at least 40% of the networks upto 4 hops away.

Durable Response





Current and Future Work

- Developing patient stratification algorithms for finding individual treatment effects and identifying possible subpopulations.
- Determining drivers of the subpopulations with different treatment effects.

References

[1] Lonial, Sagar, Venkata D Yellapantula, Winnie Liang, Ahmet Kurdoglu, Jessica Aldrich, Christophe M Legendre, Kristi Stephenson, et al. "Interim Analysis of the Mmrf Commpass Trial: Identification of Novel Rearrangements Potentially Associated with Disease Initiation and Progression." Blood 124, no. 21 (2014): 722–722.

[2] H. Xing, P. D. McDonagh, J. Bienkowska, T Cashorali, K. Runge, R. E. Miller, D. Decaprio, B. Church, R. Roubenoff, I. G. Khalil, and J. Carulli, "Causal modeling using network ensemble simulations of genetic and gene expression data predicts genes involved in rheumatoid arthritis," PLoS Comput. Biol., vol. 7, no. 3, p. e1001105, Mar. 2011

[3] D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series), 1st ed. The MIT Press, 2009.[3] N. Friedman, "Inferring Cellular Networks Using Probabilistic Graphical Models," Science, vol. 303, no. 5659, pp. 799–805, Feb. 2004.

[4] D. Madigan, J. York, and D. Allard, "Bayesian graphical models for discrete data," Int. Stat. Rev. Int. Stat., pp. 215–232, 1995..

[5] J. Pearl, Causality: Models, Reasoning and Inference, vol. 1. Cambridge University Press, 2009.

Conflict of Interest

There are no relevant conflicts of interest to disclose.