CAUSAL AI *IN SILICO* PATIENT MODEL IDENTIFIES MINICHROMOSOME (MCM) FAMILY GENES AS NOVEL PREDICTORS FOR OVERALL SURVIVAL IN MULTIPLE MYELOMA

D. D. VAGIE[1], D. M. WALKAMA[1], S. REDDY[1], L. M. MAYHEW[1], T. E. OAKLAND[1], and B. W. CHURCH[1]

1. GNS Healthcare, Inc., Somerville, MA, USA

INTRODUCTION

The minichromosome maintenance (MCM) complex is a well-known protein family that is involved in eukaryotic replication initiation. This conserved protein complex is assembled as a double-hexamer ring structure by six MCM proteins (MCM[2-7]) that work to unwind DNA double strands and facilitate replication fork progression[1].

Overexpression of proteins of the MCM complex has been demonstrated as a biomarker for rapid cell proliferation and tumor progression in multiple cancer types, including prostate, breast, lung cancers, among others. However, there is little scientific research regarding the dysregulation of the MCM complex in the context of hematological oncology, particularly in Multiple Myeloma[1] (MM).

Here, we demonstrate the first causal link between members of the MCM complex, specifically MCM2 and MCM4, and overall survival (OS) of patients with Multiple Myeloma.

METHODS

In this study, an *in-silico* patient model of Multiple Myeloma Research Foundation (MMRF) IA19 CoMMpass Study was developed using GNS's Causal AI technology, REFS™. REFS™ has been previously used to analyze MMRF data to identify the epigenetic regulator PHF19 as a marker of aggressive disease[2].

REFS™ learns an ensemble of the most probable models (n=128) given the data, where the ensemble of networks is constructed via Metropolis-Hastings Markov Chain Monte Carlo algorithm by the addition and subtraction of scored network fragments based on a maximum entropy structural prior[3] [Figure 1].

Simulated genetic perturbations were performed on the in-silico model to determine causal effects of changes in the expression of MCM family genes on OS in MM patients. These targets were validated using a Cox proportional-hazards model on an out-of-cohort Dana-Farber Cancer Institute MM dataset (DFCI).

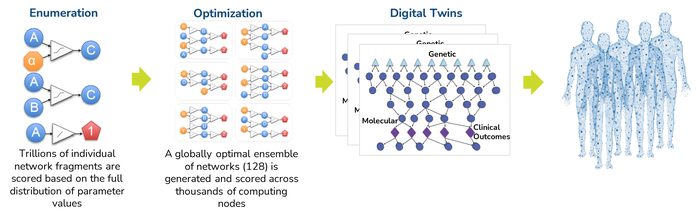

Figure 1: Machine Learning Methods (REFS™). Visualization of REFS™ enumeration of model fragments and reverse-engineering of model ensemble to create *in silico* digital twins.

RESULTS

Leveraging causal AI, simulations on REFS™ models elucidate a causal influence between genes in the MCM complex and overall survival (OS) in patients with Multiple Myeloma (MM). The prognostic role of MCM genes on OS in MM has been previously reported[4], however little to no work has investigated the causal role of the MCM complex on OS.

We first validated several MCM genes as prognostic biomarkers by simulating overall survival and conditioning on sampled values of gene expression for MCM2 and MCM4 [Figure 2 (a) and (b), respectively]. Prognostic simulations on the learned causal model line up well with Kaplan – Meier estimates of OS [Figure 2].

Furthermore, our model demonstrates that MCM genes are causal drivers of OS [Table 1]. Simulated interventions (upregulation) are applied to MCM[2-4] which show significant causal effects on OS for MCM2 and MCM4, but not for MCM3. These results were validated using Cox proportional hazards estimates on an out-of-cohort dataset from DFCI.

Preliminary investigation of the causal network paths from MCM4 to OS [Figure 3] show that MCM4 drives many genes related to DNA replication (POLE, DNMT1, HELLS, CHEK1, etc) on its path to OS. Further studies are required to understand these causal mechanisms and their biological context.

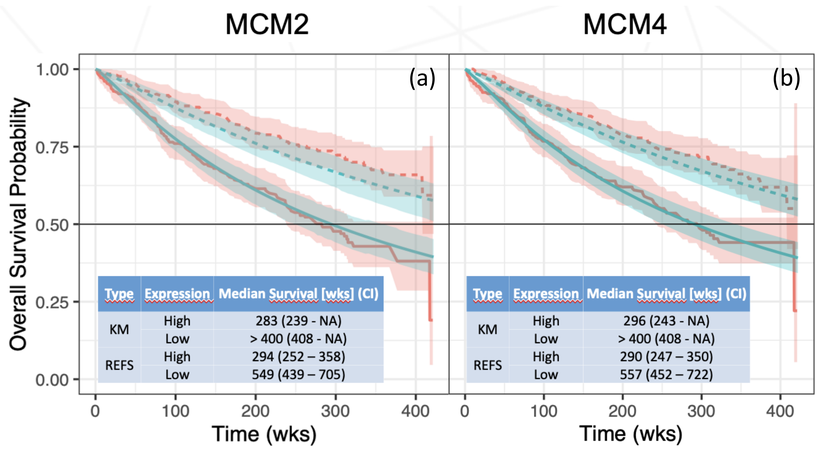

Figure 2: Analysis of MCM expression on OS. (a) - (b) *Overall survival probability is estimated by sampling the random variables in the causal networks learned from REFS™ and conditioning the simulation results on sampled values of MCM2 and MCM4, respectively. These simulated survival curves are compared to Kaplan – Meier (KM) estimates based on stratification of OS data on MCM2 and MCM4 expression level, respectively. Median survivals and the 5th – 95th percentiles (confidence intervals) are reported for REFS™ simulated survival and KM estimates for MCM2 and MCM4, respectively.*

Gene	Cox Hazard Ratio		REFS Median Effect [wks]	
	DFCI (CI)	IA19 (CI)	Causal (TP)	Prognostic (TP)
MCM4	1.31 (1.05-1.64)	1.75 (1.5-2.05)	-315.139 (0.137)	-317.217 (0.056)
MCM2	1.37 (1.07-1.75)	1.62 (1.41-1.85)	-37.857 (0.297)	-291.438 (0.069)
MCM3	1.37 (1.07-1.76)	1.72 (1.42-2.1)	-11.452 (0.416)	-228.913 (0.062)

Table 1: Cox Proportional Hazards Ratio and REFS™ Median Effect. Cox proportional hazard regression analysis is used to estimate the hazard ratio and confidence interval (CI) of three MCM genes on OS using both DFCI and MMRF IA19 datasets. For each gene, high expression is associated with increased hazard, significant at the $p \leq 0.05$ level. The last two columns detail the causal and prognostic median effect sizes on overall survival in weeks estimated from intervening or sampling the genes in the causal networks learned from REFS, respectively. The tail probability (TP) corresponds to the probability of simulating an effect with an opposite sign compared to the median effect.

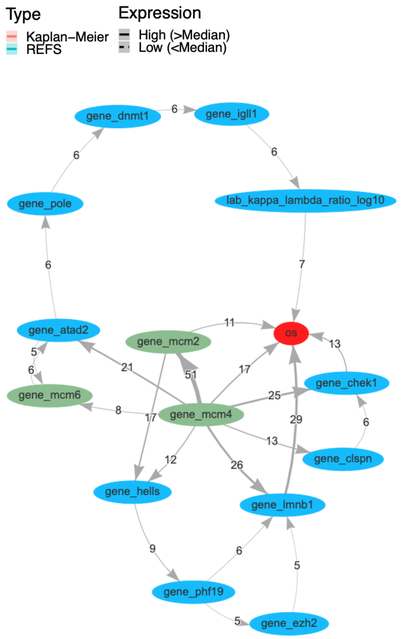

Figure 3: Consensus causal paths from MCM4 to OS. *REFS™ learns an ensemble of 128 independent Bayesian network models that detail causal relationships between (tens of) thousands of provided features. In this graph, millions of edges from these 128 networks are filtered to only include all paths between MCM4 and OS within 6 vertex-hops (distance between two vertices) and with a minimum edge count of 5 networks. Green, blue, and red vertices denote MCM genes, other model features (RNA expression, labs, clinical outcomes, DNA variants, etc.), and OS, respectively. Edge labels and thickness correspond to counts of networks (out of 128) which edge exists within the path between MCM4 and overall survival. Arrows denote paths of influence.*

CONCLUSIONS

In summary, AI causal modeling on the MMRF IA19 dataset using REFS™ has identified the first causal link between the MCM complex and OS in MM patients. We validated these findings using an out of cohort dataset to confirm the overexpression of MCM2 and MCM4 as causal predictors of increased hazard on OS. Preliminary investigation of the causal network paths between MCM4 and OS reveal interesting possible causal mechanisms for decreased OS related to the DNA replication pathway. Our findings suggest that the MCM complex is a potential novel drug target for hematological cancers such as MM.

REFERENCES

1. Yu S, Wang G, Shi Y, Xu H, Zheng Y, Chen Y. MCMS in cancer: Prognostic potential and Mechanisms. Analytical Cellular Pathology. Hindawi; 2020

2. Mason, M.J., Schinke, C., Eng, C.L.P. et al. Multiple Myeloma DREAM Challenge reveals epigenetic regulator PHF19 as marker of aggressive disease. Leukemia 34, 1866–1874 (2020).

3. Latourelle JC, Beste MT, Hadzi TC, Miller RE, Oppenheim JN, Valko MP, et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's Disease: A longitudinal cohort study and validation. The Lancet. Neurology. U.S. National Library of Medicine; 2017

4. Quan L, Qian T, Cui L, Liu Y, Fu L, Si C. Prognostic role of minichromosome maintenance family in multiple myeloma. Cancer Gene Ther. 2020 Nov;27(10-11):819-829.

ACKNOWLEDGEMENTS

We would like to acknowledge MMRF, TGen (Translational Genomics Research Institute), and DFCI for providing data for these studies. We'd also like to give a special thanks to Fred Gruber[1], Karl Runge[1], and Jonathan Keats[2] for support with this project.

1. GNS Healthcare, Inc., Somerville, MA, USA
2. Translational Genomics Research Institute, Phoenix, AZ, USA

CONTACT INFORMATION

Todd Oakland, VP Customer Value and Success
GNS Healthcare, Inc. Somerville, MA, USA
Toakland@gnshealthcare.com