



The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

624.HODGKIN LYMPHOMAS AND T/NK CELL LYMPHOMAS: CLINICAL AND EPIDEMIOLOGICAL

Machine Learning-Based Survival Analysis Reveals Prognostic Clinical and Genetic Insights for Patients with Cutaneous T-Cell Lymphoma

Celine M. Schreidah¹, David M. DeStefano, MPH², Samuel S. Pan, MSc², Shikun Wang, PhD², Haoyang Shen, MS³, Casey N. Ta, PhD⁴, George Bingham Reynolds, MS³, Lauren M. Fahmy, BS¹, Emily R. Gordon, BA¹, Oluwaseyi Adeuyan, BS¹, Bradley D. Kwinta, MD¹, Connor J. Stonesifer, MD⁵, Warren H. Chan, MD MS⁶, Jaehyuk Choi, MD PhD⁷, Madeleine Duvic, MD⁸, Fernando Gallardo, MD PhD⁹, Michael Girardi, MD¹⁰, Joan Guitart, MD⁷, Youn H. Kim, MD¹¹, Michael S. Khodadoust, MD PhD¹¹, Safa Najidh, MD¹², Xiao Ni, MD PhD⁸, Ramon M. Pujol, MD PhD⁹, Cornelis P. Tensen, PhD¹², Maarten H. Vermeer, MD PhD¹², Sean Whittaker, MD¹³, Nicholas P. Tatonetti, PhD^{14,3,4}, Herbert S. Chase, MD MA⁴, Itzik Pe'er, PhD^{3,15,16}, Larisa J. Geskin, MD^{17,2}

¹ Columbia University Vagelos College of Physicians and Surgeons, New York, NY

² Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY

³ The Data Science Institute, Columbia University, New York, NY

⁴ Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY

⁵ Dr. Phillip Frost Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, FL

⁶ Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, NY

⁷ Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago, IL

⁸ MD Anderson Cancer Center, The University of Texas, Houston, TX

⁹ Department of Dermatology, Hospital del Mar-Parc de Salut Mar, Barcelona, Spain

¹⁰ Department of Dermatology, Yale School of Medicine, New Haven, CT

¹¹ Departments of Dermatology and Medicine - Oncology, Stanford University School of Medicine, Stanford, CA

¹² Department of Dermatology, Leiden University Medical Centre, Leiden, Netherlands

¹³ St John's Institute of Dermatology, Guy's and St Thomas' NHS Foundation Trust, London, GBR

¹⁴ Department of Computational Biomedicine and Cedars-Sinai Cancer Center, Cedars-Sinai Medical Center, Los Angeles, CA

¹⁵ Department of Computer Science, Columbia University, New York, NY

¹⁶ Department of Systems Biology, Columbia University, New York, NY

¹⁷ Department of Dermatology, Columbia University Irving Medical Center, New York, NY

Introduction: Cutaneous T-cell lymphomas (CTCL) are heterogeneous lymphoproliferative disorders on a spectrum of disease presentation and severity. Around two-thirds of cutaneous T-cell lymphomas can be classified as mycosis fungoides (MF) or Sézary syndrome (SS). While advanced stages of MF and SS are associated with decreased survival and worse outcomes, even early-stage patients can possess a variable course. Numerous deep sequencing studies have fallen short in identifying genetic abnormalities that drive disease pathogenesis and predict prognosis. Large-cell transformation and elevated lactate dehydrogenase levels are associated with worse prognosis in SS; however, such features cannot accurately prognosticate patient survival. There is a need for investigation that may assist in the prognostication of survival in patients with CTCL. Machine learning methods may help to elucidate correlations between clinical and genetic factors to predict disease progression and outcomes.

Objective: Using integrated clinical and genomic data from six international sequencing studies, this investigation aimed to perform survival analysis to identify both clinical and genetic features of survival outcomes using artificial intelligence/machine learning methods.

Methods: A total of 126 eligible patients were identified, of which 99 had sufficient clinical data and 88 had sufficient clinical and genetic data. Poisson distribution was used to assess genomic data and significant genetic abnormalities for each individual patient were linked with their corresponding clinical outcomes. Genetic inputs included mutational data at a frequency of greater than or equal to four dataset occurrences. Multiple imputation using a random forest model was applied to all

included variables (rate of missingness <10%). Overall survival was assessed using three separate cox models fit with patient clinical, laboratory, or treatment covariates.

Ten-fold cross validated Least Absolute Shrinkage and Selection Operator (LASSO) was applied using adaptive regularization, and an iterated approach using 100 ten-fold cross validation repeats (resampling validation) to select the lambda value with the highest average C-index. The one-standard-error rule was used for the adaptive LASSO, while the iterated LASSO penalty was relaxed by 0.1. Prior to adjusting the lambdas, the best performing C-index was 0.73 and 0.64 for the adaptive LASSO and iterated LASSO, respectively. Following adjustment, the C-indexes were 0.71 and 0.60. To elucidate genetic candidates, we performed genome-wide association studies (GWAS) which used the false discovery rate (FDR) multiple comparisons correction and adjusted for the first three principal components in a principal component analysis (PCA) that included the imputed non-mutational covariates.

Results: We have used standard statistical and machine learning methodologies to elucidate prognostic factors in MF and SS. Using standard cox regression analysis and in agreement with prior investigations, our investigation showed significant associations for age at diagnosis (hazard ratio = 1.06, $P < 0.001$), stage at sampling (hazard ratio = 1.99, $P = 0.007$), and lymph node involvement at diagnosis (hazard ratio = 4.59, $P < 0.001$).

For the first time, using PCA-adjusted GWAS and iterated and adaptive LASSO, we demonstrate the association of mutated genes with survival in patients with MF and SS (e.g. most significant GWAS hazard ratio of 0.17, $P = 0.007$, FDR=0.35). Moreover, several mutated genes were associated with particularly poor outcomes and high mortality (e.g. most significant GWAS hazard ratio of 5.26, $P = 0.003$, FDR=0.63). Mutated genes with the highest or lowest magnitude using LASSO effect estimates in GWAS (lowest P values) were highly associated with survival outcomes. When present, the mutated genes carried significant prognostic implications for these patients.

Conclusions: Taken together, we demonstrate the potential of machine learning and artificial intelligence methodologies for investigation of novel genetic associations with survival prognostication. Future investigations are needed to validate our findings in prospective studies.

Disclosures Choi: *Moonlight Bio*: Current equity holder in private company, Membership on an entity's Board of Directors or advisory committees, Other: Co-founder, Patents & Royalties. **Kim:** *Eisai*: Research Funding; *Citius*: Research Funding; *Kyowa Kirin*: Research Funding; *Innate*: Research Funding; *Corvus*: Research Funding; *Trillium*: Research Funding; *Elorac*: Research Funding; *CRISPR Therapeutics*: Research Funding; *Takeda*: Research Funding; *Drenbio*: Research Funding. **Khodadoust:** *CRISPR Therapeutics*: Research Funding; *Daiichi Sankyo*: Membership on an entity's Board of Directors or advisory committees; *Nutcracker Therapeutics*: Research Funding.

<https://doi.org/10.1182/blood-2023-190775>