OREGON HEALTH & SCIENCES UNIVERSITY
SCHOOL OF MEDICINE – GRADUATE STUDIES
Guidelines and Regulations for Completion of Master's and Ph.D. Degrees

IN SILICO ASSESSMENT OF SOMATIC MUTATION GENERATED T CELL EPITOPES IN VARIOUS SUBSETS OF ACUTE MYELOID LEUKEMIA

By

Christopher P. Loo, Ph.D.

A MASTER'S NON-THESIS MANUSCRIPT

Presented to the Department of Molecular Informatics and Clinical Epidemiology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Master of Science,
Bioinformatics and Computational Biomedicine

September 2021

OREGON HEALTH & SCIENCES UNIVERSITY
SCHOOL OF MEDICINE – GRADUATE STUDIES
Guidelines and Regulations for Completion of Master's and Ph.D. Degrees

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

_____

This is to certify that the Master's non-thesis manuscript of

Christopher P. Loo

has been approved

Reid Thompson

_____
Mentor/Advisor

_____
Member

_____
Member

ARTICLE TITLE

In silico assessment of somatic mutation generated T cell epitopes in various subsets of acute myeloid leukemia

SHORT TITLE

Generation of T cell epitopes for AML

AUTHORS

Christopher P. Loo[1], Mary A. Wood[2], Ben Weeder[1,3], Reid F. Thompson[1,2,3,4,5]

AFFILIATIONS

1. Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University
2. Phase Genomics, Seattle, Washington
3. Computational Biology Program, Oregon Health & Science University
4. Portland VA Research Foundation
5. Department of Radiation Medicine, Oregon Health & Science University
6. Division of Hospital and Specialty Medicine, VA Portland Healthcare System

Corresponding author

Reid F. Thompson: KR-COM, 2720 S. Moody Avenue, Portland, Oregon 97201, thompsonre@ohsu.edu, 503-494-6594

Funding: National Library of Medicine T15 grant (5 T15 LM 7088)

Word count (text): 2280
Word count (abstract): 234
Figure/table count: 2
Reference count: 23

Christopher P. Loo, Ph.D.
Postdoctoral Researcher
Department of Molecular Informatics and Clinical Epidemiology
Oregon Health & Science University
3181 SW Sam Jackson Park Rd, Mail Code: BICC
Portland, Oregon 97239
lab) 503-494-4502
email: looc@ohsu.edu

Title: In silico assessment of somatic mutation generated T cell epitopes in various subsets of acute myeloid leukemia

Abstract:

Despite decades of conventional chemotherapy, the prognosis for patients with acute myeloid leukemia (AML) remains dismal, with 75% of patients succumbing to their disease within 5 years of diagnosis. Recent advances in cancer therapy have turned to the immune system for specific targeting and clearance of tumors. However, to date, there is not a comprehensive understanding of the neoepitope landscape in AML that could be used in the next generation of immunotherapies. In this study, genomic data from 562 patients (cohort: Beat AML program, Oregon Health & Science University) were analyzed computationally to identify tumor variants, altered mRNA sequences of variants, and HLA-type. Using a computational pipeline and algorithm (neoepiscope), we were able to predict 8-11 amino acid peptide sequences (aka: epitopes) from DNA-seq of complementary tumor and normal patient samples which consider germline context and the potential for co-occurrence of two or more somatic variants on the same mRNA transcript. Without consideration of these phenomena, existing approaches are likely to produce both false positive and false negative results, resulting in an inaccurate and incomplete picture of the cancer neoepitope landscape. In this study we show that in *silico* neoepitope prediction will accurately predict viable novel peptides that can be tested for target immunotherapy in AML. Our future plan is to synthesize predictions and use them to stimulate banked patient samples and identify bioactive peptides that can be used as the basis of targeted immunotherapy.

Introduction:

Acute myeloid leukemia (AML) is the most common type of leukemia in adults, as incidence

increases with age. AML represents 1.2% of all cancer cases [1]; however, the US anticipates that the number of new AML cases will rise significantly as the post-World War II "baby boomers" will increase the population of citizens 65 years of age or older. AML disease increases the number of immature myeloid derived white blood cells (blasts) that fill the bone marrow (BM) and circulate in the blood. These immature myeloid blast cells have mutations that manifest in excessive cellular proliferation and cell survival, which in turn, causes an overrepresentation of blast cells and an underrepresentation of healthy white blood cells, red blood cells, and platelets (in bone marrow and blood). As a result, AML can lead to infection, anemia, fatigue, bruising, and excessive bleeding.[2]

Common chemotherapy targets highly replicating cells, a hallmark of tumor cells. Due to the side effects of chemotherapy patients can experience loss of hair, changes in skin pigmentation, energy weakness, and sometimes lethargy. [3] AML blast cells deviate from the normal cellular lifespan to a state of prolonged proliferation and survival. AML can originate and transform from another blood malignancy or there could be a de novo AML disease. Despite these differences, treatment against AML has stayed relatively the same for decades, standard induction chemotherapy consisting of a cytarabine and idarubicin regimen and hypomethylating drugs. [4] These therapies cull over proliferation of WBCs both peripherally and in the bone marrow. Despite treatment with chemotherapy and BM transplantation, prognosis to clear the AML remains poor; patients are able to reach remission but not a cure status. It has become evident that the immune system plays a direct role in helping the tumor seed into a tissue by increasing oncogenic growth factors as well as creating an inflammatory environment that dampens the immune system's own defense mechanisms against neoplasms. However, immunotherapy has promise as in targeting the activation of the adaptive immune response and suppressing

inhibitory immune factors. [5–7] Focus on new immune therapies is thought to help with long-term treatment leading to a potential cure. [8] The optimism of immunotherapy is the specificity of T cells and antibodies from B cells to target foreign proteins or peptides. The specificity of these immune factors as well as the long-term durability of immune responses is a key mechanism of successful vaccinations. In particular, cytotoxic T cells have the ability to kill foreign or diseased cells. Cytotoxic T cells particularly bind to short peptides bound by an MHC-complex on a target cell. Recognition of both the peptide and the MHC complex defines whether the target cell is killed or accepted by the cytotoxic T cell. Any mismatch of the MHC complex defines the target cell as foreign (i.e., mismatch graft tissue) and allows the target cell to be eliminated. If cognate MHC binds with a foreign peptide (i.e., infected cell producing pathogenic peptides), the cytotoxic T cell immune response is to eliminate the target cell. Therefore, a major obstacle of new immunotherapies is tuning the immune system to be solely specific for a target cell, overcoming immune tolerance, and licensing adaptive immune cells to attack tumors – allowing controlled autoimmunity to occur. [9]

Breakthroughs in whole-genome and exome sequencing have provided valuable information on the tumors from AML patients. Common mutations associated with the disease have classified subtypes of AML disease that provide insight to target cancer therapy. Furthermore, genetic sequencing has provided insight to the low mutational burden of AML, which on average has 13 somatic mutations per patient.[10,11] This mutational burden of AML is in contrast to tumors that contain a higher number of mutations, such as, melanoma with greater than 400 mutations per patient. [12] The low mutational burden of AML emphasizes the difficulty to find new biomarkers that can define or mark AML tumors in vivo.

Neoepitope prediction is a process that can help with adding specificity to the next generation

of immunotherapies.[13] The mutational burden in tumors has the potential to produce new proteins, or neoantigens, that are recognized as foreign to the immune system. With the use of DNA and RNA sequencing (-seq) data from patient tumors, it is possible to computationally identify the tumor variants of a patient, the altered mRNA sequences of the variants, and predict the 8-11 amino acid peptide sequence (aka: epitopes) of the larger protein that can possibly be presented by the tumor and recognized by the adaptive immune system (B cell antibodies or cytotoxic T cells). Standard methods of identifying neoepitopes is a laborious experimental task of isolating target cells, cleaving off MHC complexes from the cell surface, extracting the peptide attached from the MHC complex, purifying individual peptides, and performing mass-spec to elucidate the amino-acid peptide sequence. [14] Mining for neoepitopes computationally allows for analyzing all mutations and predicting new antigenic epitopes that can be tested in vitro for immunogenicity. Current tools for neoepitope prediction from genomic sequencing of complementary tumor and normal patient samples generally do not consider germline context or the potential for co-occurrence of two or more somatic variants on the same mRNA transcript (e.g., Epi-Seq) or the cleavage peptide sites from the proteasome. [15,16] These approaches produce both false positive and false negative results, decreasing the likelihood of identifying antigenic neoepitopes.

To predict neoepitopes in various subsets of AML, we applied Neoepiscope to the Beat AML dataset of ~900 AML patient samples (consisting of blood and BM aspirates). The Beat AML cohort data set was generated over the last 5 years at our institute as part of a national collaborative network aimed at defining the mutational landscape of this disease.[10] More significantly, it is the largest comprehensive biorepository of AML samples for the computational screening of genomic data, the in vitro screening of small molecule therapeutics

on blood and bone marrow samples, and immunological phenotyping and functional analysis. Discovery of neoepitopes in the Beat AML cohort will have significant importance in the route to provide precision immunotherapy to future AML patients. In this project, we predict a series of immunogenic neoepitopes associated with HLA-A, HLA-B, and HLA-C with a binding affinity less than 500 nm and specific to a patient's subtype of AML.

Results:

HLA haplotype prediction was performed on 864 samples using WES derived from the BeatAML database[10]. Data fasta.gz format files using read 1 and read 2 were utilized in the HLA prediction through the use of Optitype software. The data shows HLA prediction of each allele based on the highest prediction probability. Forty eight distinct HLA-A alleles were detected, the greatest count within the population of patients were A*02:01, A*02:02, and A*03:01 (Figure 2A). Within the HLA-B subtype there were 89 distinct alleles predicted. HLA-B*07:02, B*08:01, and B*44:02 were the three alleles with the highest frequency in the cohort (figure 2B). Within the HLA-C subtype there were 36 distinct alleles with C*07:01, C*07:02, and C*04:01 (figure 1C). This data shows HLA-B with the greatest distribution of alleles within this cohort and is consistent with the reports that HLA-B has the greatest breadth of alleles within the MHC-class I subtypes.

Neoepitope prediction was a culmination of using both normal and tumor DNA-seq from the BeatAML cohort. Bam files were used for variant calling and consensus sequences were made from the output from mutect and varsan callers. Variant phasing was done using HapCUT2 and

with the input HLA prediction and MHC binding affinity, we used Neoepiscope software to predict $4.6*10^7$ neoepitopes across 864 patients (figure 1A-B). The distribution of predicted neoepitopes for each patient is shown in the boxplot in figure 2A. There is a wide distribution of predicted epitopes in the cohort with the median number of patients having approximately 2000 predicted epitopes but in the upper quartile, the 75th percentile is approximately 70,000 predicted epitopes. The max number of predicted epitopes in a patient is approximately 345,000. With the large distribution of predicted epitopes we further categorized the numbers of neoepitopes to the specific diseases within the BeatAML cohort as it could help designate the enrichment of predicted epitopes to subtypes of AML or coordinating disease. The barplot in figure 2B shows the number of predicted epitopes to specific diagnosis. The barplot in figure 2B shows the number of predicted epitopes associated with specific disease from the BeatAML cohort. The number of epitopes associated with the subtype of AML is greatest in number within the group of AML mutated NPM1, AML with myelodysplasia-related changes, and AML/NOS. These numbers are closely related to the number of patients with the respective diagnosis. AML mutated NPM1, AML with myelodysplasia-related changes, and AML/NOS are also the top three diagnoses.

Discussion:

Immune based therapies against solid and blood tumors are an actively growing field, especially as monoclonal antibodies have shown success against melanoma and many other cancer types. However, monoclonal antibody immune-checkpoint (i.e., aPD-1, aCTLA-4) therapy has specificity for T cells and helps to release the T cell suppression from the inflammatory tumor microenvironment. In this study, we elucidated neoepitopes from AML patients using whole

genome sequencing data. These neoepitopes are specific tumor proteins that can be used to prime naive T cells to become memory T cells specific for the patient's type of AML disease. Once activated, CD8 T cells clonally expand to form an army and efficiently kill unwanted cells by the release of toxic proteins onto target cells. However, CD8 T cell immunity is dependent on identifying what is foreign and what is self and, in this study, we have identified specific tumor epitopes that can be used to condition cytotoxic T cells to eradicate AML tumor cells.

We used DNA sequencing (DNA-seq) data from the largest AML cohort assembled to date (Beat AML, OHSU) to computationally identify the variants, the altered mRNA sequences of the variants, HLA-type, and used the computational algorithm Neoepiscope to predict the peptide sequences that have strong binding affinity to each patient's specific HLA class-I haplotype. The mechanisms of targeted-immunotherapy are still at the beginning stages for blood tumors. However, our results add important HLA patient information to the BeatAML dataset. Identification of specific risk or preventative HLA alleles associated with particular subtypes of AML is added information in the efforts for early AML detection.

Existing tools for neoepitope prediction from DNA-seq of complementary tumor and normal patient samples do not always consider germline context or the potential for co-occurrence of two or more somatic variants on the same mRNA transcript. Without this consideration, existing approaches are likely to produce both false positive and false negative results, resulting in an inaccurate and incomplete picture of the cancer neoepitope landscape. We believe that using Neoepitope has produced a thorough estimation of neoepitopes within the Beat AML cohort of patient samples that could be directly used for screening immunogenicity against AML samples and utilized for the development of patient-specific immunotherapy. Our results show that it is possible to generate predictive epitopes in AML despite it being classified as a type of cancer

with low antigenic burden in regards to the number of somatic mutations that have been measured. While our data is not at the stage of being tested, we have further interests in continuing to prune the neoepitope prediction by adding a proteasome cleavage filter[10], especially in regards to MHC class I restricted HLA where presented peptides go through a proteome for trimming off of the C-terminal and N-terminals before the peptide coordinates with the TAP protein for the binding to the MHC class I. Neoepitope clustering is another area where we are investigating, finding like neoepitopes that have similar sequences (i.e. 1 hamming distance) to form in clusters (i.e. 3-5 peptides sequences). This would provide more confidence in a viable prediction if particular neoepitopes were predicted in groups versus isolated predicted peptide sequences. Furthermore it is in our interest to test neoepitopes from this project in vitro by synthesizing peptides, making peptide pools, and adding the peptides to AML blood samples for an in vitro functional T cell assay for cytokine production or cellular proliferation. The strength and range of the responses would support the prediction methods presented in this project.

Methods:

SOFTWARE

Neoepitope prediction started from unaligned DNA-seq reads of tumor samples and matched normal samples. FASTQ files are used for prediction of HLA genotype for each sample using OptiType software.[17] In our study, BAM files of sequence alignment were performed by the Beat AML project. [10] Germline and somatic variant calling was performed using VarScan and Mutect software producing VCF files. [18,19] Variant phasing was performed using HapCUT2 software and binding affinity of peptide to HLA using prediction tools including MHCflurry,

MHCnuggets, and NetMHCpan software.[20–22] We used Neoepiscope software (https://github.com/pdxgx/neoepiscope) to consider germline context and the potential for co-occurrence of two or more somatic variants on the same mRNA transcript. Neoepiscope is an open-source software licensed under the MIT license and is available at https://github.com/ohsu-comp-bio/neoepiscope.

Data processing was performed using the Advanced Computing Center (Oregon Health and Science University) and Exacloud Cluster (https://www.ohsu.edu/advanced-computing-center).

PATIENT SAMPLES

Patient samples were generated in a collaboration with Evan Lind, PhD, a collaborator in the Beat AML cohort at Oregon Health & Science University. All participating patients submitted informed consent to participate in the Beat AML consortium and volunteer their samples for research study. The Beat AML consortium consisted of ten research institutions: Oregon Health & Science University, University of Utah, University of Texas Medical Center (UT Southwestern), Stanford University, University of Miami, University of Colorado, University of Florida, National Institutes of Health (NIH), Fox Chase Cancer Center and University of Kansas (KUMC). [10]

DATA

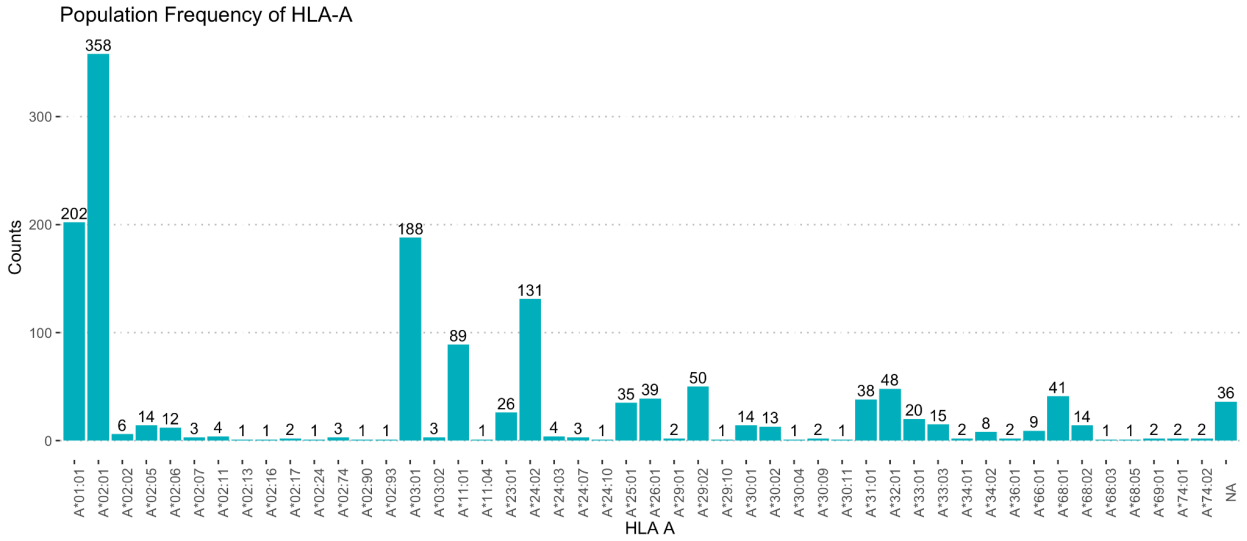
Publicly available genomic data from the Beat AML consortium can be found at

http://vizome.org. Code and analysis can be found at

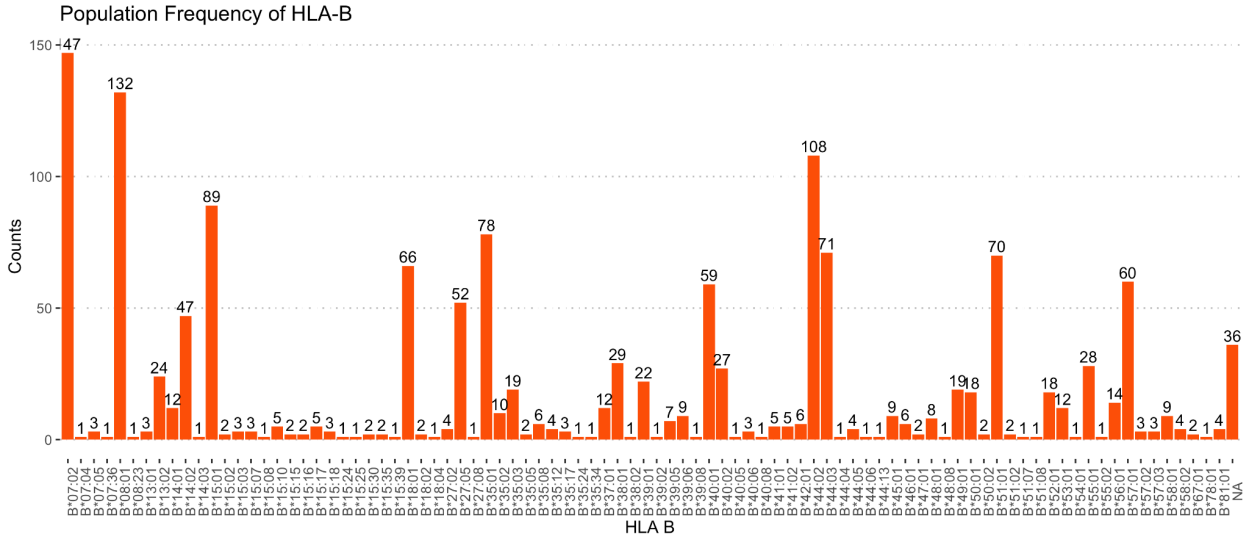
https://github.com/looc27/AML_neoepitope_prediction.

Figures:

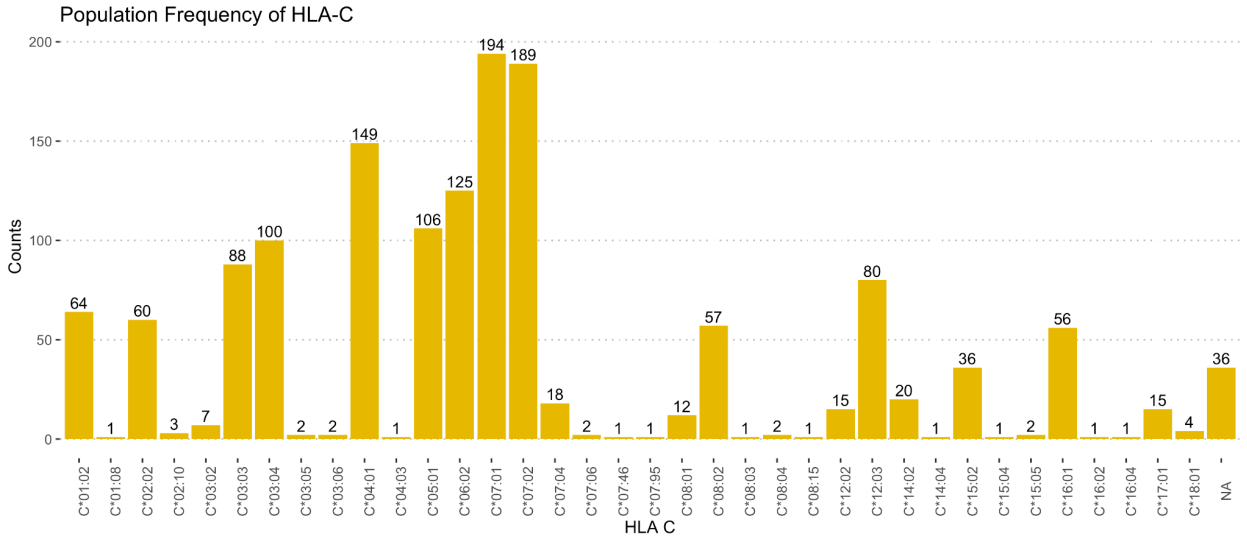
1A)



1B)



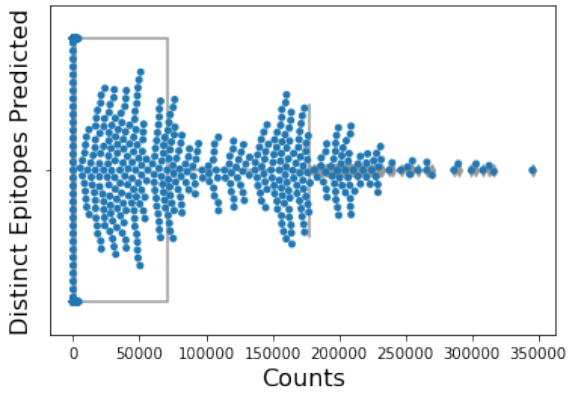
1C)



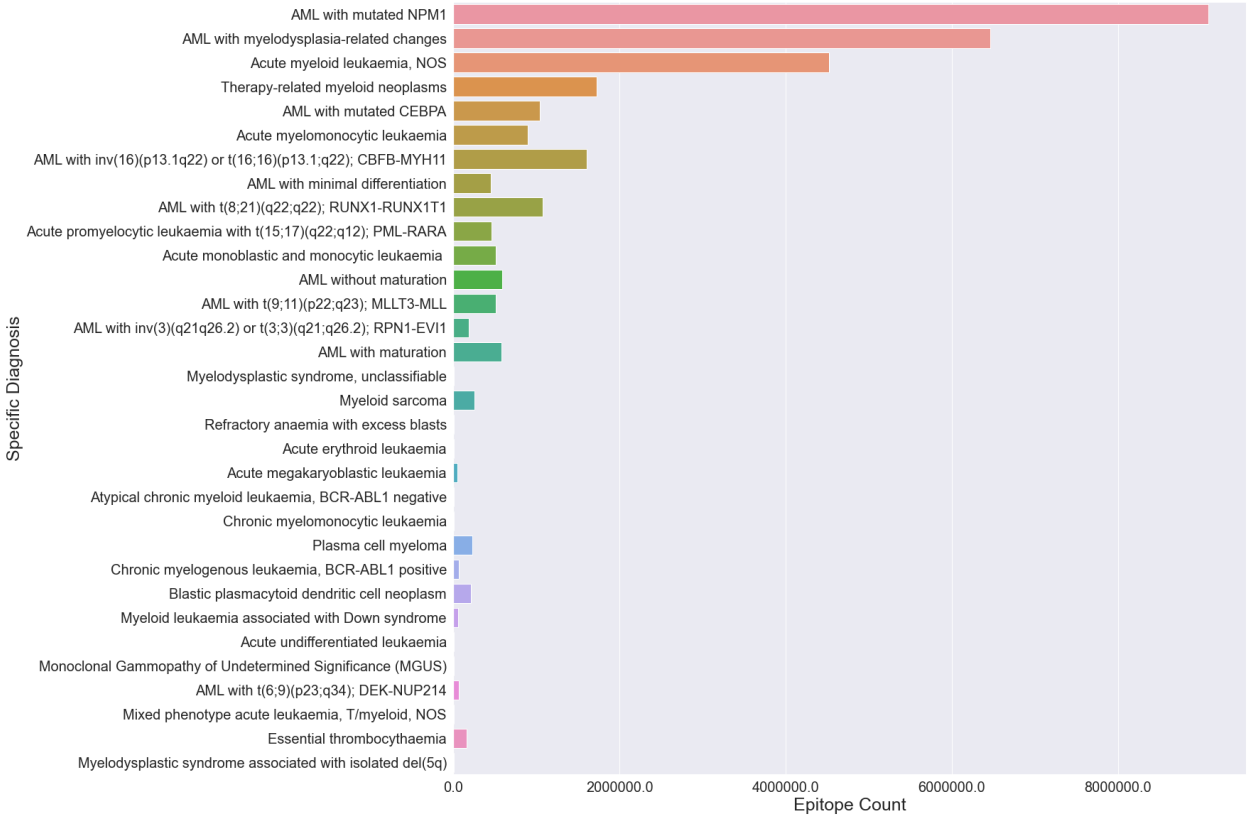
Distribution of HLA types of the frequency of HLA types in AML patients.

The population frequency of HLA-A, -B, -C from AML patients. Alleles are ordered ascending and the specific count of alleles are represented on top of each bar. (HLA-A, blue; HLA-B, red; HLA-C, yellow).

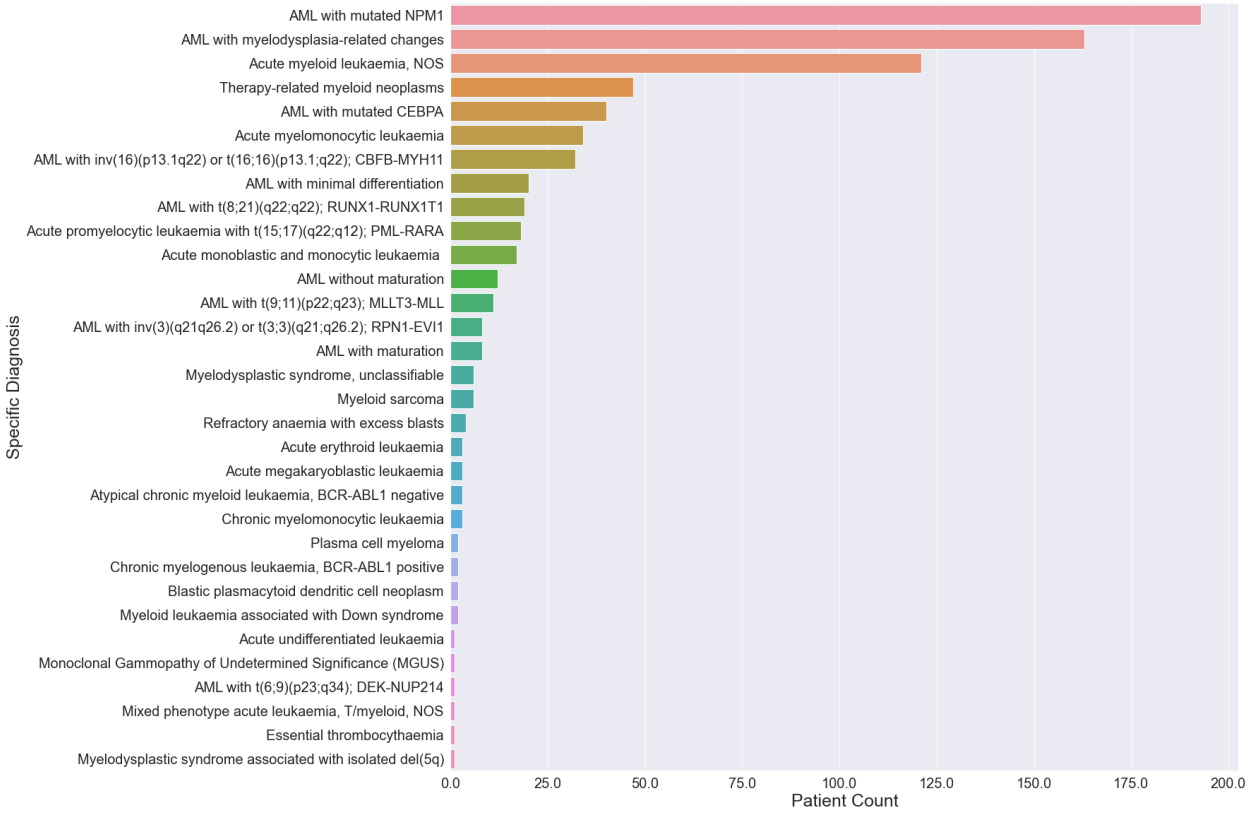
2A)


Distinct Epitopes Predicted
0
50000
100000
150000
200000
250000
300000
350000
Counts

2B)


Specific Diagnosis
AML with mutated NPM1
AML with myelodysplasia-related changes
Acute myeloid leukaemia, NOS
Therapy-related myeloid neoplasms
AML with mutated CEBPA
Acute myelomonocytic leukaemia
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11
AML with minimal differentiation
AML with t(8;21)(q22;q22); RUNX1-RUNX1T1
Acute promyelocytic leukaemia with t(15;17)(q22;q12); PML-RARA
Acute monoblastic and monocytic leukaemia
AML without maturation
AML with t(9;11)(p22;q23); MLLT3-MLL
AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); RPN1-EVI1
AML with maturation
Myelodysplastic syndrome, unclassifiable
Myeloid sarcoma
Refractory anaemia with excess blasts
Acute erythroid leukaemia
Acute megakaryoblastic leukaemia
Atypical chronic myeloid leukaemia, BCR-ABL1 negative
Chronic myelomonocytic leukaemia
Plasma cell myeloma
Chronic myelogenous leukaemia, BCR-ABL1 positive
Blastic plasmacytoid dendritic cell neoplasm
Myeloid leukaemia associated with Down syndrome
Acute undifferentiated leukaemia
Monoclonal Gammopathy of Undetermined Significance (MGUS)
AML with t(6;9)(p23;q34); DEK-NUP214
Mixed phenotype acute leukaemia, T/myeloid, NOS
Essential thrombocythaemia
Myelodysplastic syndrome associated with isolated del(5q)
0.0
2000000.0
4000000.0
6000000.0
8000000.0
Epitope Count

2C)



Counts and distribution of Neoepitopes to AML.

The number of predicted epitopes per patient is measured in figure 2A. Neoepitopes in specific diseases are plotted in figure 2B while the number of patients with specific diagnosis is shown in 2C.

References

1. Jemal A, Thomas A, Murray T, Thun M. Cancer statistics, 2002. *CA. Cancer J. Clin.* 2002;52(1):23–47.
2. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N. Engl. J. Med.* 2015;373(12):1136–1152.
3. Aslam MS, Naveed S, Ahmed A, et al. Side effects of chemotherapy in cancer patients and evaluation of patient's opinion about starvation based differential chemotherapy. *J. Cancer Ther.* 2014;2014:.
4. Murphy T, Yee KWL. Cytarabine and daunorubicin for the treatment of acute myeloid leukemia. *Expert Opin. Pharmacother.* 2017;18(16):1765–1780.
5. Gubin MM, Zhang X, Schuster H, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature*. 2014;515(7528):577–581.
6. McGranahan N, Furness AJS, Rosenthal R, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. 2016;351(6280):1463–1469.
7. Stahl M, Goldberg AD. Immune Checkpoint Inhibitors in acute myeloid leukemia: novel combinations and therapeutic targets. *Curr. Oncol. Rep.* 2019;21(4):37.
8. Acheampong DO, Adokoh CK, Asante D-B, et al. Immunotherapy for acute myeloid leukemia (AML): a potent alternative therapy. *Biomed. Pharmacother.* 2018;97:225–232.
9. Caspi RR. Immunotherapy of autoimmunity and cancer: the penalty for success. *Nat. Rev. Immunol.* 2008;8(12):970–976.
10. Tyner JW, Tognon CE, Bottomley D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526–531.
11. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–421.
12. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science*. 2015;349(6255):1483–1489.
13. Brennick CA, George MM, Srivastava PK, Karandikar SH. Prediction of cancer neoepitopes needs new rules. *Semin. Immunol.* 2020;47:101387.
14. van der Lee DI, Reijmers RM, Honders MW, et al. Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia. *J. Clin. Invest.* 2019;129(2):774–785.
15. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061–1073.
16. Duan F, Duitama J, Al Seesi S, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* 2014;211(11):2231–2248.

17. Szolek A, Schubert B, Mohr C, et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinforma. Oxf. Engl.* 2014;30(23):3310–3316.
18. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22(3):568–576.
19. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 2013;31(3):213–219.
20. O'Donnell TJ, Rubinsteyn A, Bonsack M, et al. MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst.* 2018;7(1):129-132.e4.
21. Bhattacharya R, et. al. Evaluation of machine learning methods to predict peptide binding to MHC class I proteins | bioRxiv. **doi:** doi.org/10.1101/154757 2017
22. Nielsen M, Lundegaard C, Blicher T, et al. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput. Biol.* 2008;4(7):e1000107.
23. Weeder, B, Wood MA, Li E, Nellore A, Thompson RF. Pepsickle rapidly and accurately predicts proteasomal cleavage sites for improved neoantigen identification. bioRxiv. **doi:** doi.org/10.1101/2021.05.14.444244. 2021