KSBi-BIML 2023

Bioinformatics & Machine Learning(BIML)
Workshop for Life Scientists, Data Scientists,
and Bioinformatians

생물정보학&머신러닝 워크샵(온라인)

Bioinformatics for Cancer Immunotherapy

김상우 _ 연세대학교

SBi 한국생명정보학회
Korean Society for Bioinformatics

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2023 워크샵 온라인 수업을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고**합니다.

KSBi-BIML 2023

Bioinformatics & Machine Learning (BIML)
Workshop for Life Scientists, Data Scientists, and Bioinformatians

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2023에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 9차를 맞이하게 되었습니다. 지난 2년간은 심각한 코로나 대유행으로 인해 아쉽게도 모든 강의가 온라인으로 진행되어 현장 강의에서만 가능한 강의자와 수강생 사이에 다양한 소통의 기회가 없음에 대한 아쉬움이 있었습니다. 다행히도 최근 사회적 거리두기 완화로 현장 강의가 가능해져 올해는 현장 강의를 재개함으로써 온라인과 현장 강의의 장점을 모두 갖춘 프로그램을 구성할 수 있게 되었습니다.

BIML 워크샵은 전통적으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되었습니다. 올해 AI 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 심층학습(Deep learning) 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체 분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 싱글셀오믹스 분석과 메타유전체분석 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다. 이외에 다양한 생명정보학 분야에 대하여 30개 이상의 온라인 강좌가 개설되어 제공되며 온라인 강의의 한계를 극복하기 위해서 실시간 Q&A 세션 또한 마련했습니다. 특히 BIML은 각 분야 국내 최고 전문가들의 강의로 구성되어 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것입니다.

이번 BIML-2023을 준비하기까지 너무나 많은 수고를 해주신 BIML-2023 운영위원회의 남진우, 우현구, 백대현, 정성원, 정인경, 장혜식, 박종은 교수님과 KOBIC 이병욱 박사님께 커다란 감사를 드립니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 현장 강의와 온라인 강의를 준비하시는데 노고를 아끼지 않으신 모든 연사분께 깊은 감사를 드립니다.

2023년 2월

한국생명정보학회장 이 인 석

강의개요

Bioinformatics for Cancer Immunotherapy

본 강의에서는 차세대 암 치료 기법으로 떠오르고 있는 면역항암 치료의 원리와, 이를 수행하는 데에 필요한 다양한 생물정보학 분석 기법을 설명한다. 효율적인 암 면역치료의 기반이 되는 암 면역 특성과 환경에 대한 분석, 특히 암 유전체 분석, 변이 분석, 항원 분석의 원리를 이해하고 나아가 이를 다양한 데이터에 활용할 수 있는 기초지식을 다시는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- 암 면역치료의 역사와 개요
- 암 면역치료의 방법 및 연계된 유전체 분석 기법
- 효과적인 암 정밀 면역치료를 위한 암 면역특성 및 면역환경 분석 기법
- 생물정보학 분석 도구 소개

* 참고강의교재:
  - 유인물 배포 예정

* 강의 난이도: 초급

* 강의: 김상우 교수 (연세대학교 의과대학)

Curriculum Vitae

Speaker Name: Sangwoo Kim, Ph.D.

▶ **Personal Info**

Name	Sangwoo Kim
Title	Associate Professor
Affiliation	Yonsei Univ. College of Medicine

▶ **Contact Information**

Address	50-1, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Email	swkim@yuhs.ac
Phone Number	010-3407-9861

Research Interest

Genomic analysis of human disease, Algorithm development

Educational Experience

2002	B.S. in Computer Science, KAIST
2004	M.S. in Bioinformatics, KAIST
2010	Ph.D. in Bioinformatics, KAIST

Professional Experience

2010-2013	Post-doc research fellow, UC San Diego
2014-2019	Assistant Professor, Yonsei University College of Medicine
2020-	Associate Professor, Yonsei University College of Medicine

Selected Publications (5 maximum)

1. Kim TM, Yang IS, Seung BJ, Lee S, Kim D, Ha YJ, Seo MK, Kim KK, Kim HS, Cheong JH, Sur JH, Nam H and **Kim S***, Cross-species Oncogenic Signatures of Breast Cancer in Canine Mammary Tumors, Nature Communications, 2020 11, article number 3616
2. Jo S-Y, Kim E, and **Kim S***, Impact of mouse contamination in genomic profiling of patient-derived models and best practice for robust analysis, Genome Biology 2019, (20):231
3. Kim J, Kim D, Lim JS, Maeng JH, Son H, Kang H-C, Nam H, Lee JH* and **Kim S***, The use of technical replication for detection of low-level somatic mutations in next-generation sequencing, Nature Communications 2019, article 1047
4. Lee G, Ryu HJ, Choi JW, Kang H, Yang WI, Yang IS, Seo M-K, **Kim S*** and Yoon SO*, Characteristic gene alterations in primary gastrointestinal T and NK cell lymphomas, Leukemia 2019 33:1797-1832
5. Kim S, Kim HS, Kim E, Lee MG, Shin E-C, Paik S, and **Kim S***, Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information, Annals of Oncology 2018, 29(4):1030-1036

KSBi-BIML

Bioinformatics for Cancer Immunotherapy

Sangwoo Kim, Yonsei University

Cancer Immunotherapy:

Exploit host's immune system to treat cancer

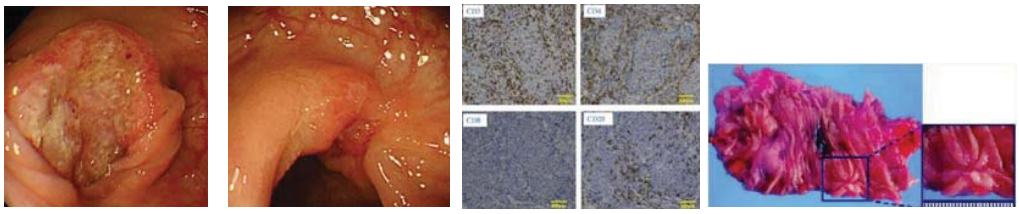
- Generate or augment an immune response against cancer

SBi 한국생명정보학회 Korean Society for Bioinformatics

Immune and cancer

- Immunosuppressed patients have a higher risk for cancer

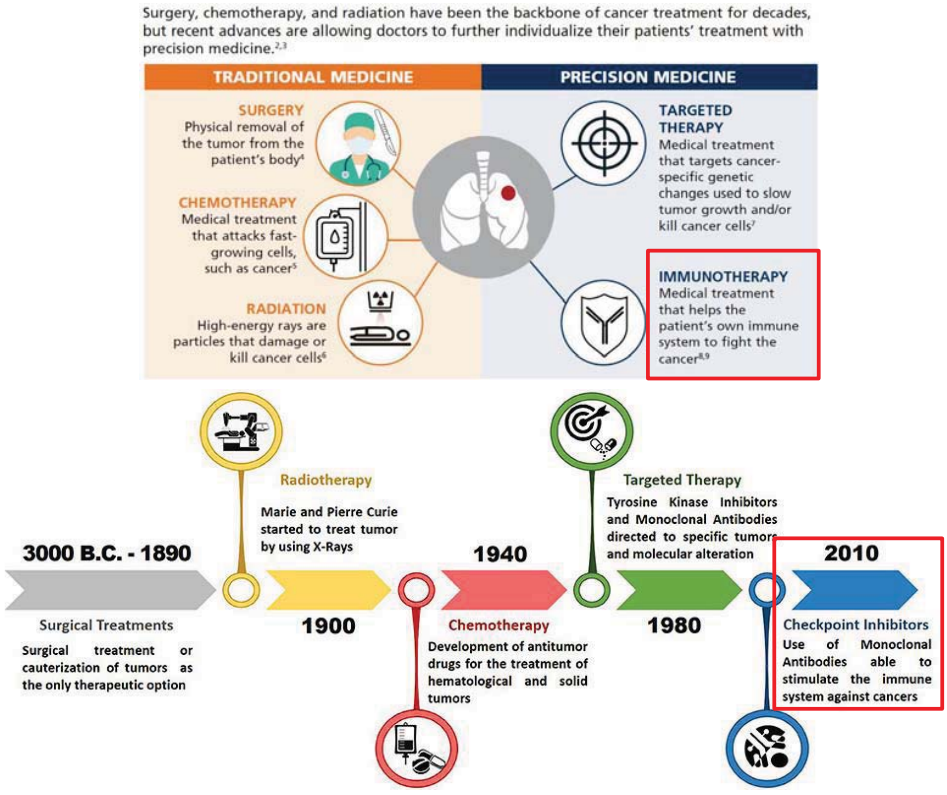
- Spontaneous regression occurs one in every 60,000 to 100,000 ca ncer cases



First visit | 1 week later | CD3+CD4+ T-cell | No cancer

Chida *et al, Surg Case Rep 2017*

Cancer Immunotherapy as a new hope



The history of immunotherapy

New York Times - July 29, 1908

ERYSIPELAS GERMS AS CURE FOR CANCER

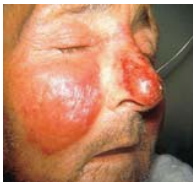
Dr. Coley's Remedy of Mixed Toxins Makes One Disease Cast Out the Other.

MANY CASES CURED HERE

Physician Has Used the Cure for 15 Years and Treated 430 Cases— Probably 150 Sure Cures.

Following news from St. Louis that two men have been cured of cancer in the City Hospital there by the use of a fluid discovered by Dr. William B. Coley of New York, it came out yesterday that nearly 100 cases of that supposely incurable disease have been cured in this city during the last few years, all through the use of the fluid discovered by Dr. Coley.


William Bradley Coley


erysipelas

CONTRIBUTION TO THE KNOWLEDGE OF SARCOMA.[1]

By WILLIAM B. COLEY, M.D.,

OF NEW YORK.

I. A Case of Periosteal Round-Celled Sarcoma of the Metacarpal Bone; Amputation of the Forearm; General Dissemination in Four Weeks; Death Six Weeks Later.

II. The General Course and Prognosis of Sarcoma, Based Upon an Analysis of Ninety Unpublished Cases.

III. The Treatment of Sarcoma by Inoculation With Erysipelas, With a Report of Three Recent (Original) Cases.

I. THE patient a young lady, æt. 18, had been in perfect health from earliest childhood. The family history was likewise good with the exception of a remote tubercular tendency, and the fact that an ancestor, three generations before, had died of "cancer" of the lip, presumably epithelioma.

In the early part of July, 1890, she received a slight blow upon the back of the right hand. The hand became a little swollen and somewhat painful the first night. The next few days the pain became a trifle less and the swelling subsided, but did not entirely disappear. About a week later the swelling again began to increase very slowly, and the pain became more severe. She consulted a physician at the time of the injury, but there being no evidence of anything more than an ordinary bruise the usual local applications were applied.

August 12. The pain and swelling continuing, she again sought

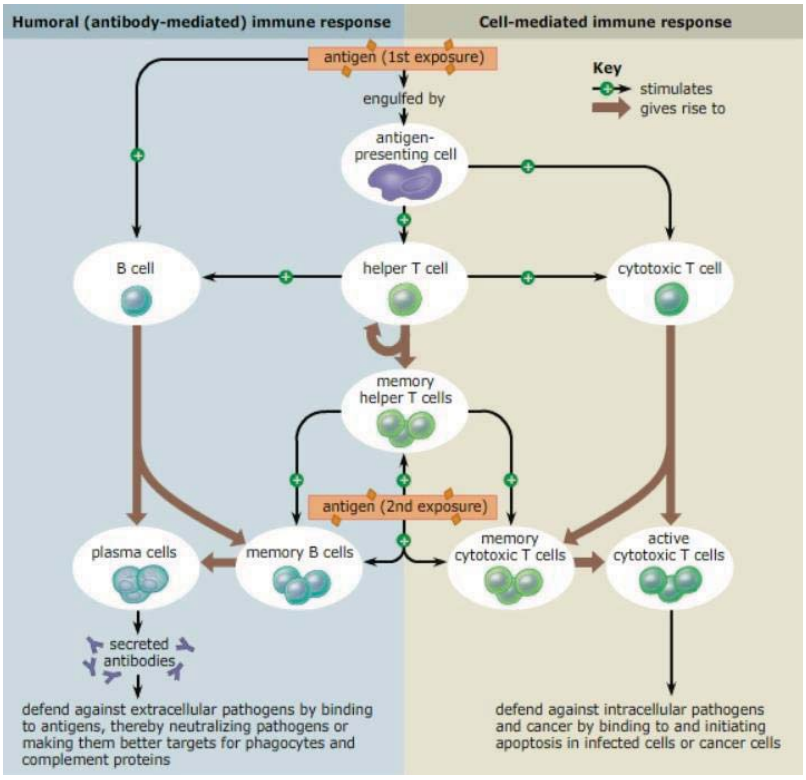
[1]Read before the Surgical Section of the New York Academy of Medicine, April 27, 1891. (With a report of three cases treated since).

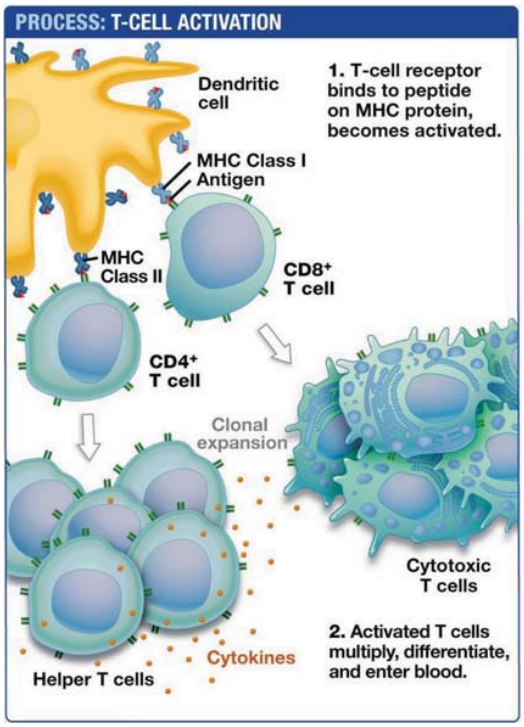
(199)

Coley, Annals of Surgery, 1981

SBi 한국생명정보학회 Korean Society for Bioinformatics

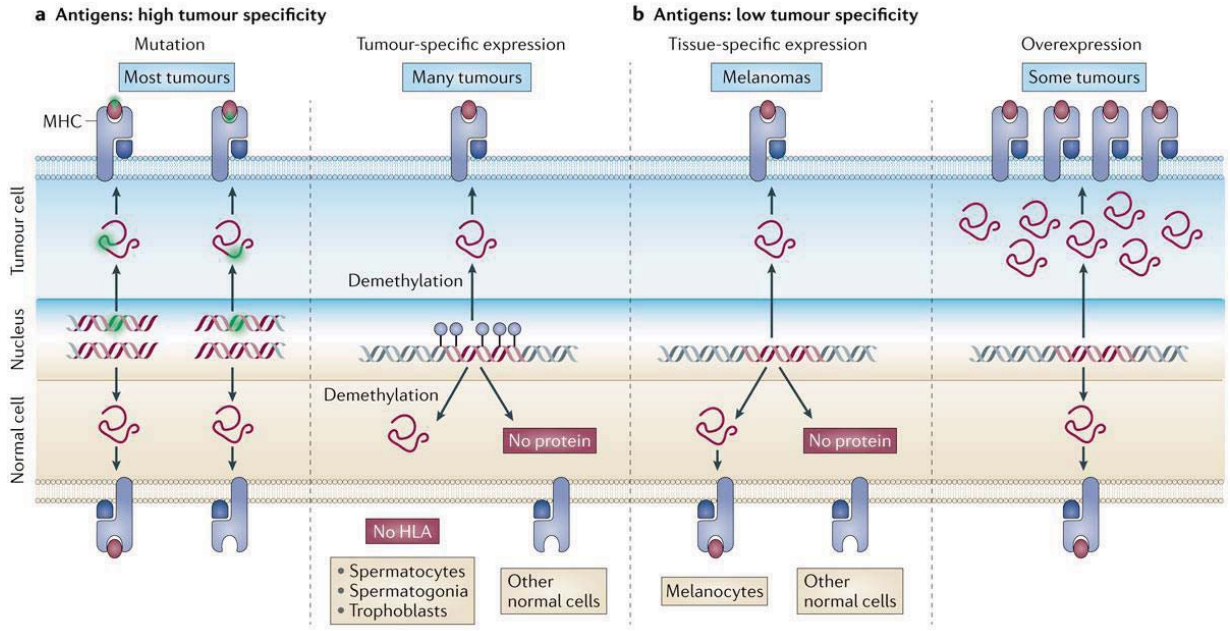
Adaptive Immunity / T-cell activation





SBi 한국생명정보학회 Korean Society for Bioinformatics

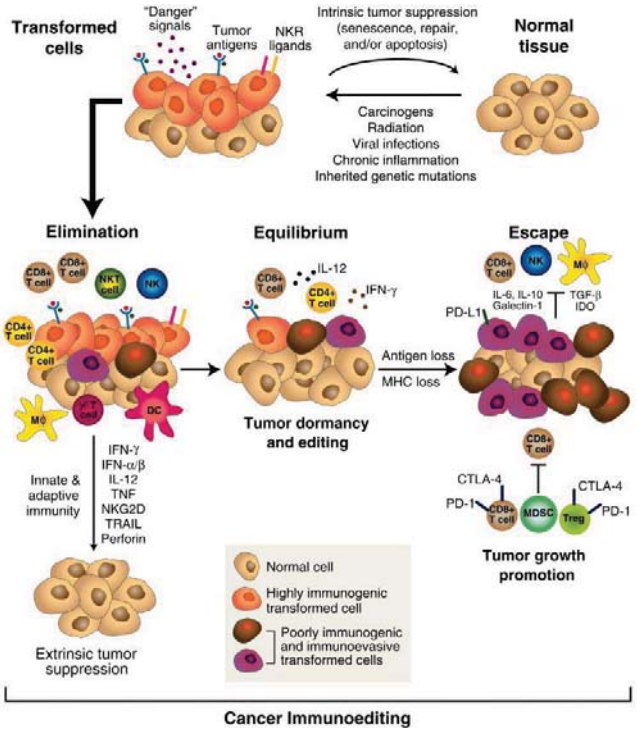
Tumor Antigens



TAA (Tumor Associated Antigen): presented in tumor cells + (some normal cells)
TSA (Tumor Specific Antigen): presented only in tumor cells

SBi 한국생명정보학회 Korean Society for Bioinformatics

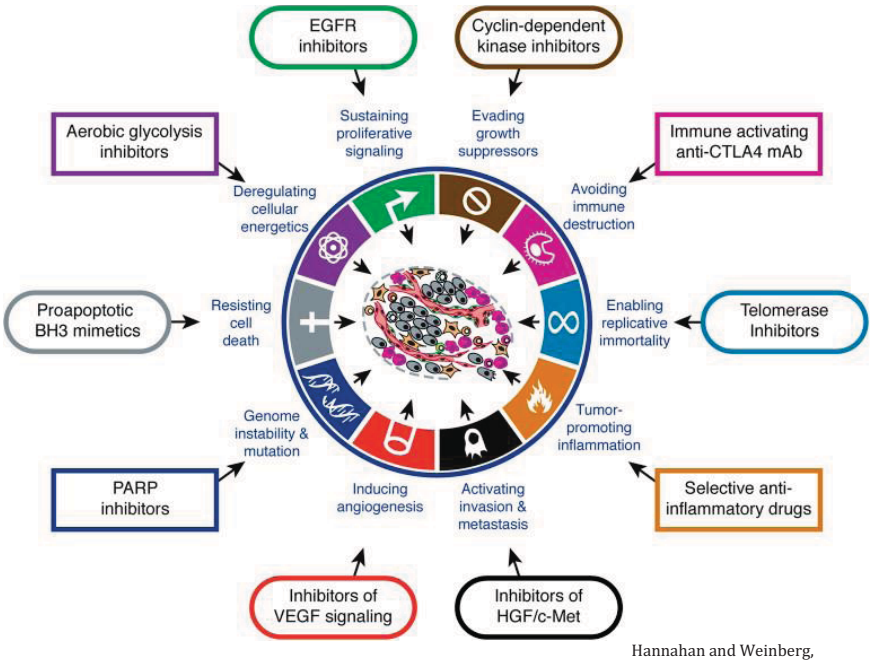
Immunoediting of cancer



- **Elimination (immunosurveillance):**
  - Initial damage (possible destruction) of tumor cells by innate immune system
  - Tumor antigen presentation and attacked by CD4+, CD8+ T-cells
- **Equilibrium:**
  - Survived tumor cells do not progress and remain dormant
- **Escape:**
  - Cancer cells grow and metastasize due to the loss of control by the immune system

SBi 한국생명정보학회 Korean Society for Bioinformatics

Immune evasion

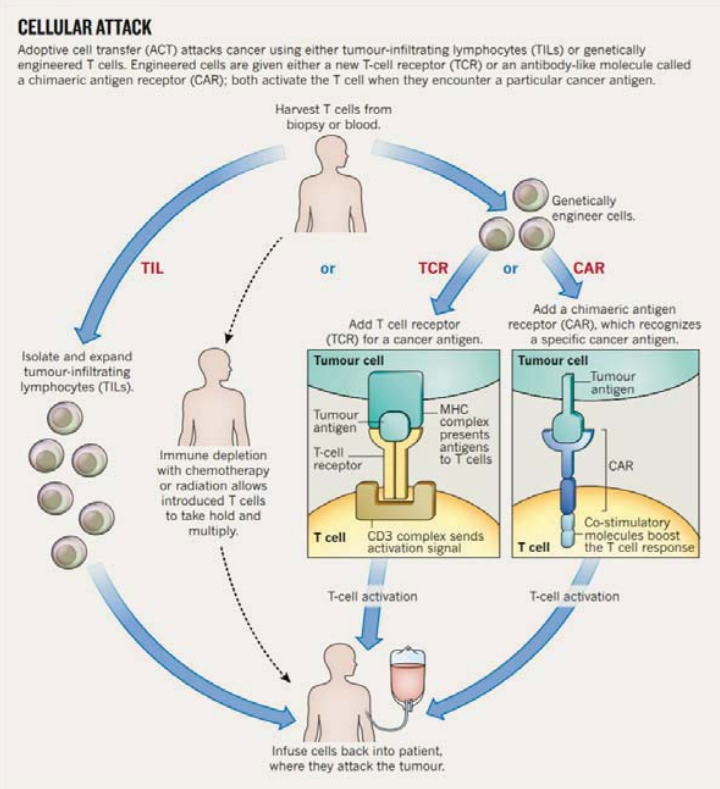


Hannahan and Weinberg, Hallmarks of cancer: The Next Generation, Cell 2011

- Paralyze CTLs and NK cells by secreting TGF-β or immunosuppressive factors
- Recruitment of regulatory T-cell (Tregs) and myeloid-derived suppressor cells (MDSCs)
- Loss of MHC class I expresssion

CURRENT APPROACHES

1. Adoptive Cell Transfer



- TILs (tumor-infiltrating lymphocytes) - metastatic melanoma
  - tissue surrounding tumor may contain immune cells and antitumor activity
  - culture TILs and re-infuse
  - deplete endogenous immune cells
- TCR (T-cell receptor)
  - give cells new receptor
  - viral vector in patient's T-cell
  - T-cell receptor must be genetically match to the patient's immune type
- CAR (chimeric antigen receptor)
  - artificial, antibody-like protein
  - antibody (binding to cancer antigen)
  - cell activating receptor
  - stimulatory molecule

Courtney Humpreies, Nature 504, S13-15, 2013

SBi 한국생명정보학회 Korean Society for Bioinformatics

Adverse effects and personalization

Table 1

Select examples of adverse events resulting from clinical application of immunotherapies targeting public antigens

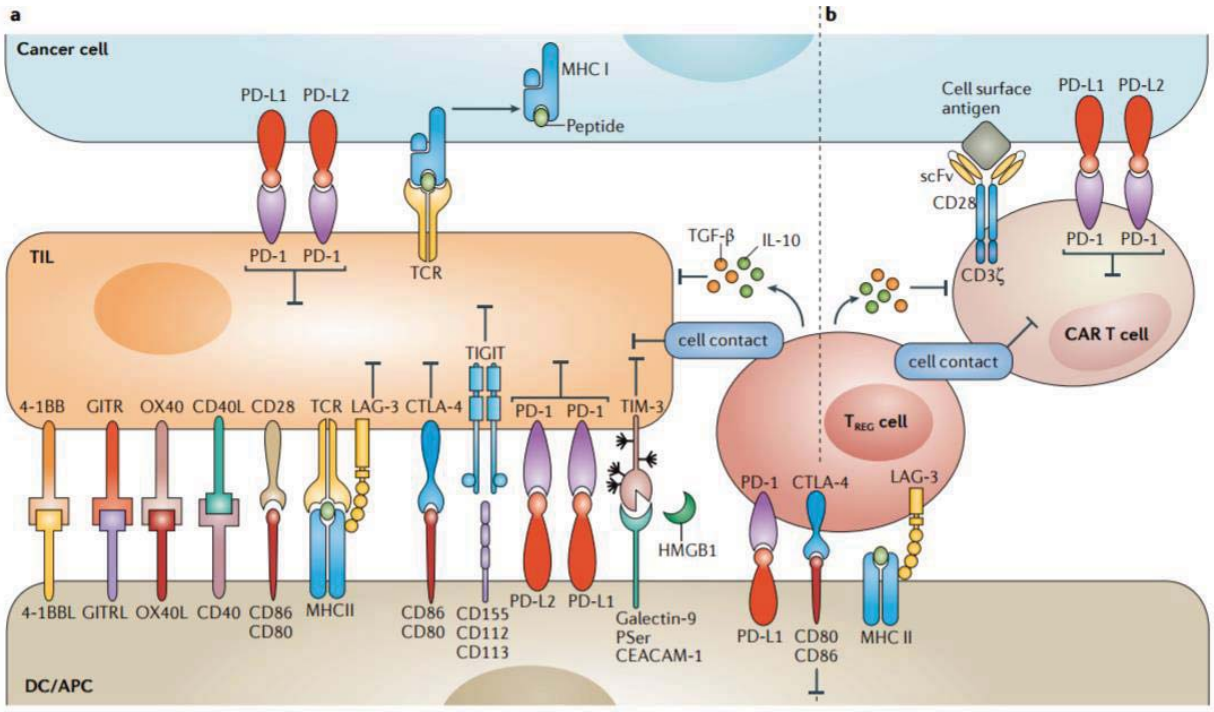
Antigen	Immunotherapy	Adverse event	Cause	Ref.
MART-1/MelanA	TCR	Fatal neural and cardiac toxicity	High levels of inflammatory cytokines alone or in combination with semi-acute heart failure and epileptic seizure	[30]
		Uveitis, Hearing loss, Loss of pigmentation	On-target activity of TCR-engineered T cells targeting normal cells expressing the cognate epitope	[24•]
	TCR + DC vaccination	Acute respiratory distress	High levels of inflammatory cytokines	[31]
NY-ESO-1	TCR (Affinity enhanced)	Skin rash with lymphocytosis, diarrheal syndrome	Autologous GVHD-like syndrome possibly due to loss of self-tolerance	[32]
MAGE-A3	TCR (Affinity enhanced)	Fatal cardiogenic shock	Cross-reactivity with an unrelated epitope from the Titin protein presented on cardiac tissue	[28]
	TCR (Affinity enhanced)	Mental status changes, comas, necrotizing leukoencephalopathy with extensive white matter defects	Reactivity to similar MAGE-A12-derived epitope presented on neural cells	[33]

- Adverse effects in ACT
  - cytokine storm
- Need to target "tumor-specific" antigen
  - Neoantigen?

Courtney Humpreies, Nature 504, S13-15, 2013

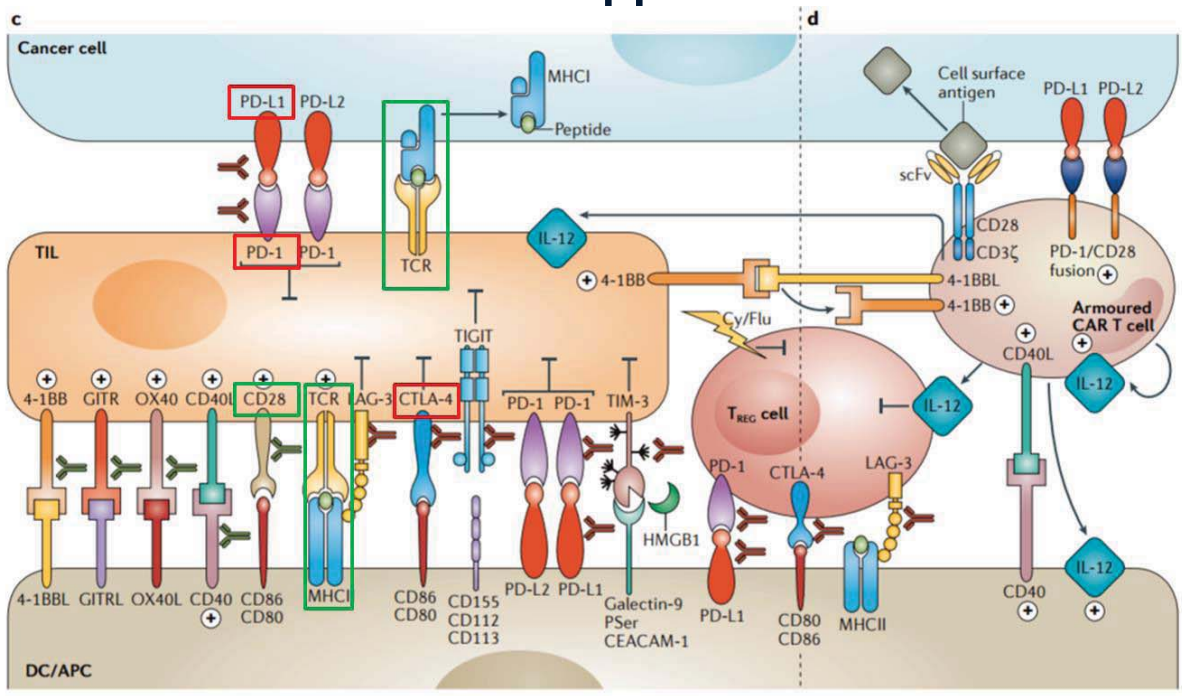
SBi 한국생명정보학회 Korean Society for Bioinformatics

2. Checkpoint inhibitors



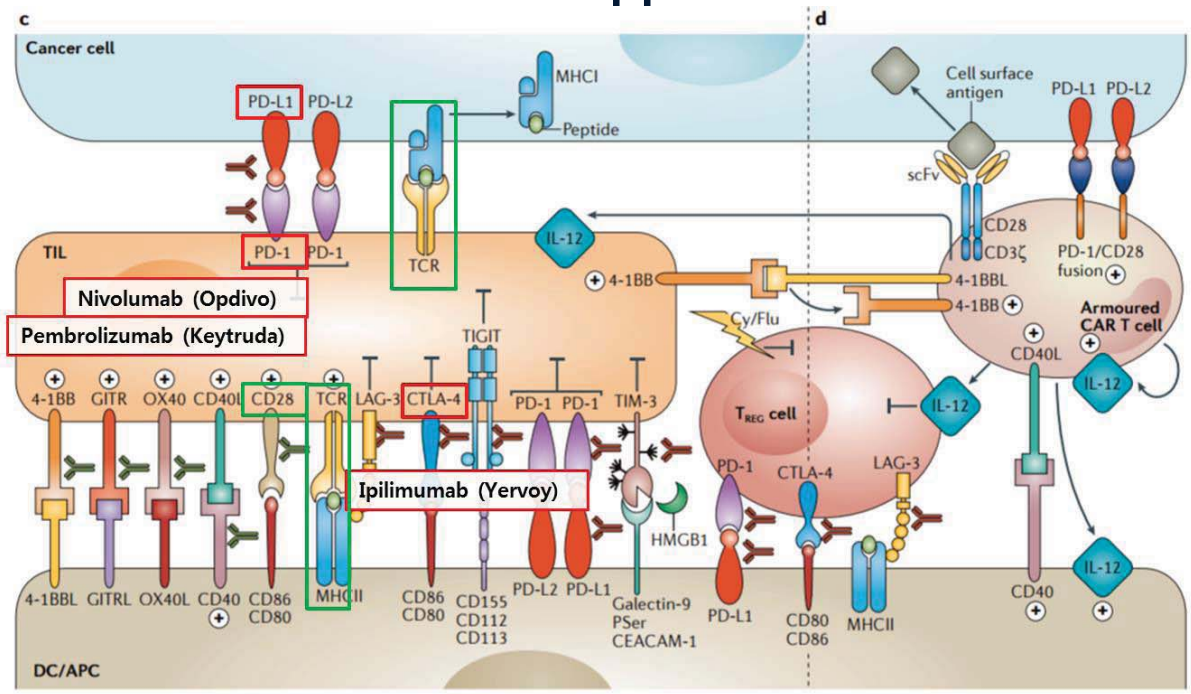
SBi 한국생명정보학회 Korean Society for Bioinformatics

Immunomodulatory mABs to overcome immunosuppression



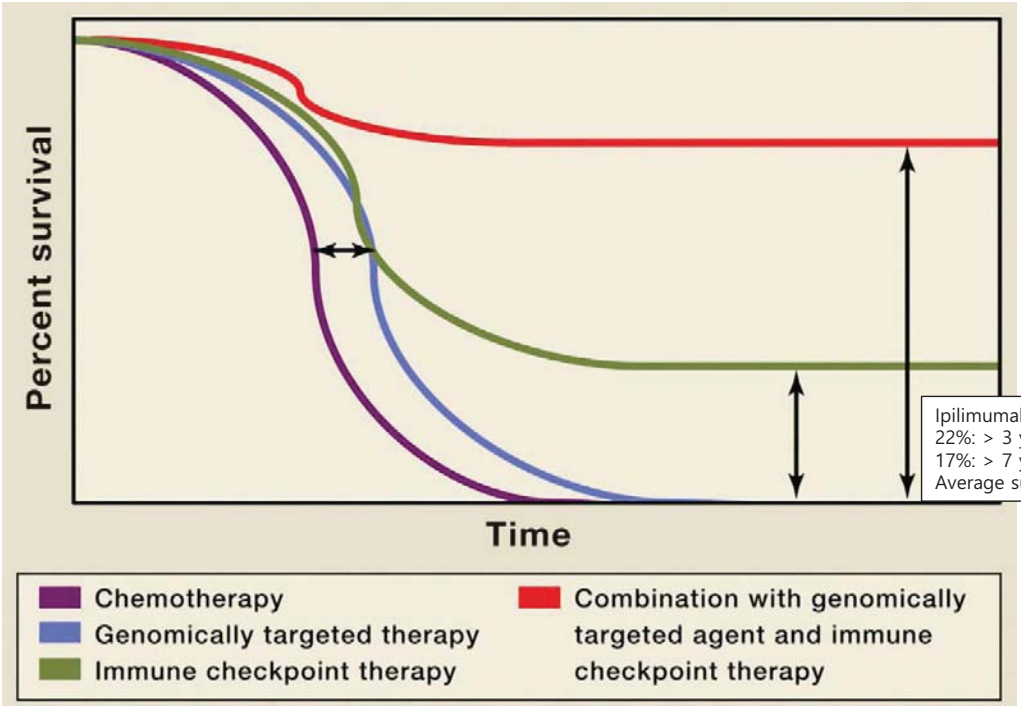
SBi 한국생명정보학회 Korean Society for Bioinformatics

Immunomodulatory mABs to overcome immunosuppression



SBi 한국생명정보학회 Korean Society for Bioinformatics

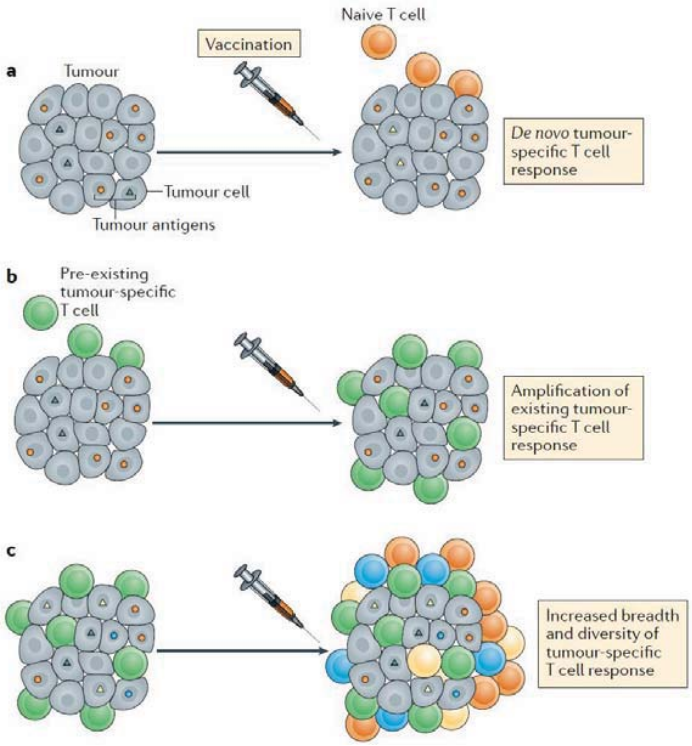
The benefits from cancer immunotherapy



Ipilimumab in melanoma (1861 patients)
22%: > 3 years
17%: > 7 years
Average survival (6-9 month to >1 yrs)

SBi 한국생명정보학회 Korean Society for Bioinformatics

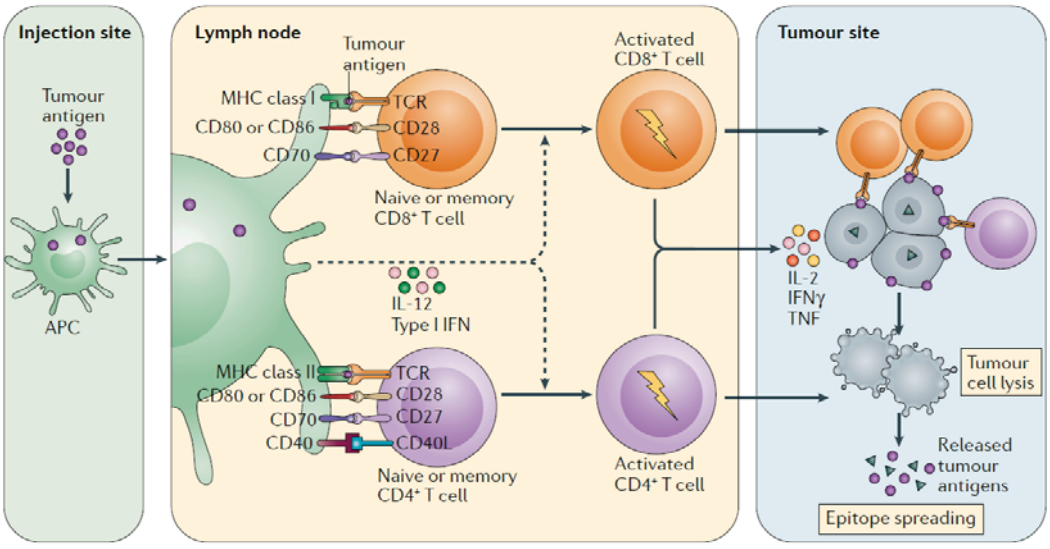
3. Cancer Vaccine



Hu *et al, Nat. Rev. Immunol 2018*

Cancer vaccines:
- Injection of tumor antigens
- generate new antigen-specific T-cell response
- amplification of existing T-cell response
- increase breadth and diversity of T-cell response

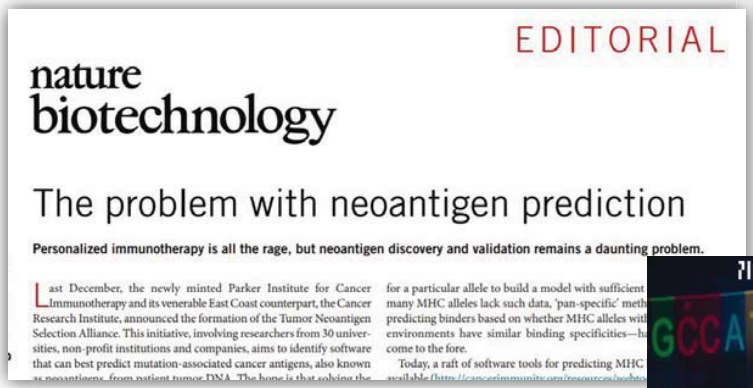
How cancer vaccine works



Hu *et al, Nat. Rev. Immunol 2018*

- Antigen injection (or DC vaccine):
- Migration of APC to present antigens to T-cells (signal 1)
- Co-stimulatory signals (signal 2)
- Migration of T-cells to tumor site
- Kill tumor cells (cytotoxicity, IFNγ, TNF..)

Neoantigen prediction is a key challenge


EDITORIAL

nature biotechnology

The problem with neoantigen prediction

Personalized immunotherapy is all the rage, but neoantigen discovery and validation remains a daunting problem.

Last December, the newly minted Parker Institute for Cancer Immunotherapy and its venerable East Coast counterpart, the Cancer Research Institute, announced the formation of the Tumor Neoantigen Selection Alliance. This initiative, involving researchers from 30 universities, non-profit institutions and companies, aims to identify software that can best predict mutation-associated cancer antigens, also known

for a particular allele to build a model with sufficient

many MHC alleles lack such data, 'pan-specific' meth

predicting binders based on whether MHC alleles wil

environments have similar binding specificities—h

come to the fore.

Today, a raft of software tools for predicting MHC

Editorial, Nat. Biotech. 2017 35(2)


RESEARCH PROJECT

Tumor Neoantigen Selection Alliance

OVERVIEW

Penn

Stanford MEDICINE

- Neoantigen prediction for markers of checkpoint inhibitor
- Neoantigen prediction for finding tumor-specific (non-self) antigens for ACT

TUMOR MUTATION BURDEN (TMB)

Who can benefit from checkpoint inhibitor?

The NEW ENGLAND JOURNAL of MEDICINE

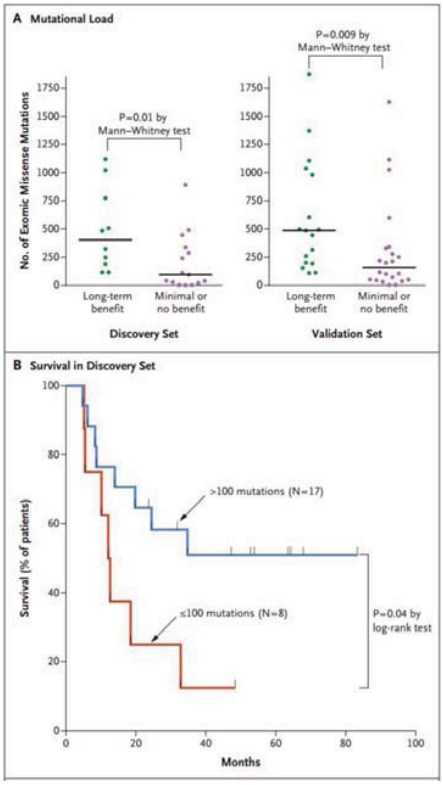
ORIGINAL ARTICLE

Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma

Alexandra Snyder, M.D., Vladimir Makarov, M.D., Taha Merghoub, Ph.D., Jianda Yuan, M.D., Ph.D., Jesse M. Zaretsky, B.S., Alexis Desrichard, Ph.D., Logan A. Walsh, Ph.D., Michael A. Postow, M.D., Phillip Wong, Ph.D., Teresa S. Ho, B.S., Travis J. Hollmann, M.D., Ph.D., Cameron Bruggeman, M.A., Kasthuri Kannan, Ph.D., Yanyun Li, M.D., Ph.D., Ceyhan Elipenahli, B.S., Cailian Liu, M.D., Christopher T. Harbison, Ph.D., Lisu Wang, M.D., Antoni Ribas, M.D., Ph.D., Jedd D. Wolchok, M.D., Ph.D., and Timothy A. Chan, M.D., Ph.D.

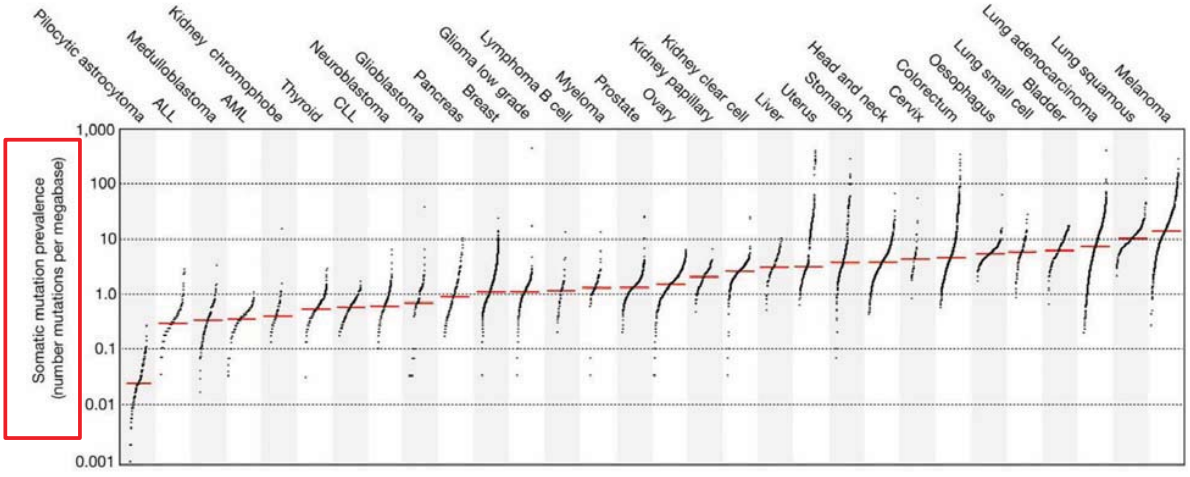
64 melanoma patients (25 discovery set, 39 validation set) treated with Ipilimumab .

Patients with high mutation burden: good survival, long-term benefit



SBi 한국생명정보학회 Korean Society for Bioinformatics

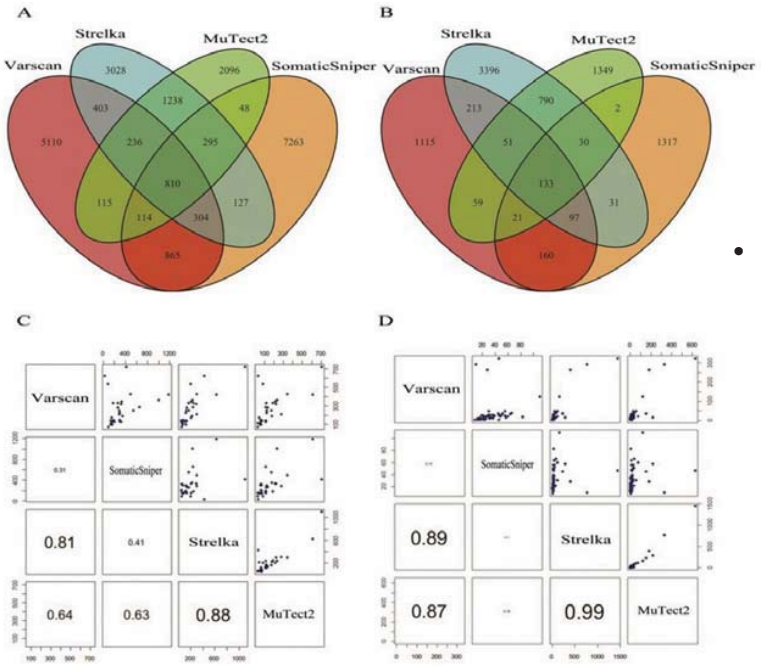
Tumor mutation burden



- **Tumor Mutation Burden (TMB)** $= \frac{\#total\_somatic\_mutation}{total\_targeted\_genome\_size(Mb)}$

SBi 한국생명정보학회 Korean Society for Bioinformatics

Inconsistence of somatic mutation calls



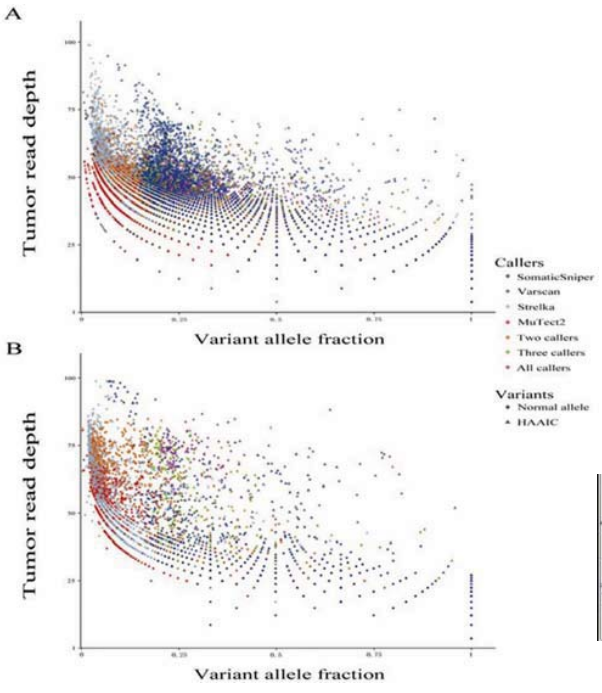
Low depth (~50x) Targeted, High depth (~370x)

- The number of somatic mutations are largely dependent on the variant caller used

Cai *et al, Sci Rep. 2016*

SBi 한국생명정보학회 Korean Society for Bioinformatics

Tumor mutation burden



- The number of somatic mut ations are largely dependent on the read depth

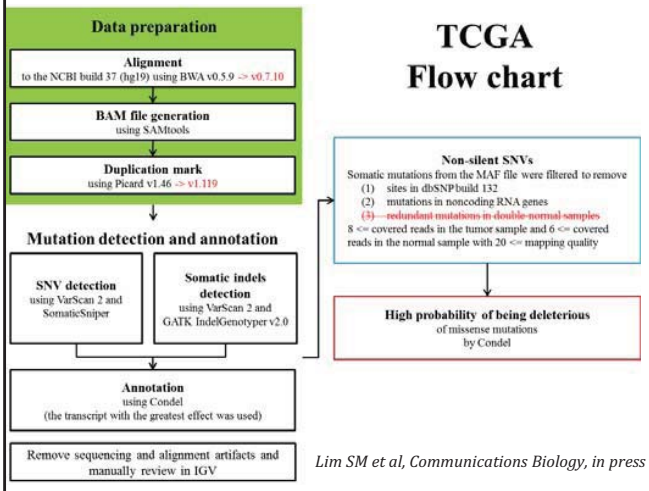
- And the read depth is simply not uniform

Cai *et al, Sci Rep. 2016*

SBi 한국생명정보학회 Korean Society for Bioinformatics

Fixing pipeline

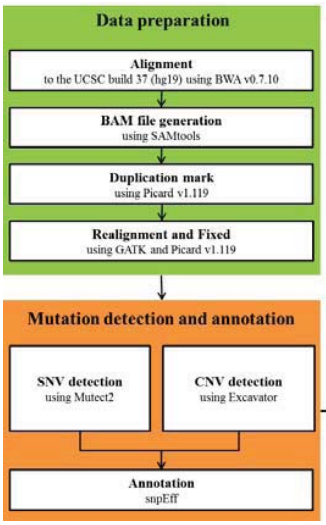
mut/MB (SNV)
5.533993
5.178398
3.056166
1.459616
1.471475
1.453428
1.706356



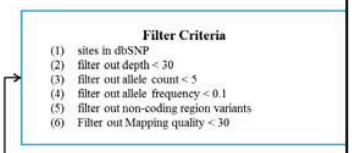
TCGA Flow chart

Lim SM et al, Communications Biology, in press

mut/MB
2.991871
2.4023
1.641857
1.27743
1.820113
1.108711
1.003712

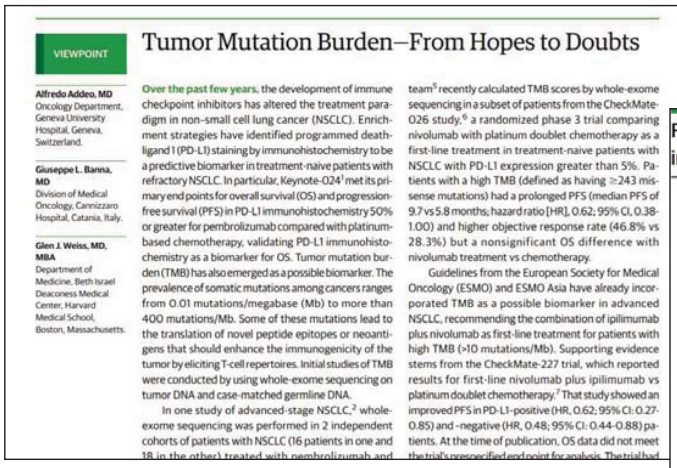


In-house Flow chart



SBi 한국생명정보학회 Korean Society for Bioinformatics

Potential pitfalls (use with care)



VIEWPOINT

Tumor Mutation Burden—From Hopes to Doubts

Alfredo Addeo, MD
Oncology Department, Geneva University Hospital, Geneva, Switzerland.

Giuseppe L. Banna, MD
Division of Medical Oncology, Cannizzaro Hospital, Catania, Italy.

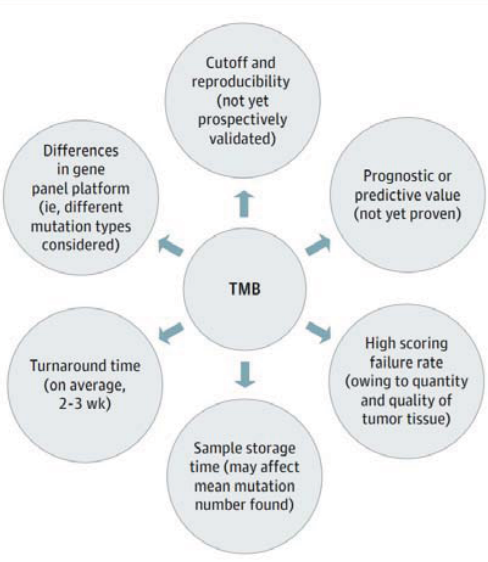
Glen J. Weiss, MD, MBA
Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts.

Over the past few years, the development of immune checkpoint inhibitors has altered the treatment paradigm in non-small cell lung cancer (NSCLC). Enrichment strategies have identified programmed death-ligand 1 (PD-L1) staining by immunohistochemistry to be a predictive biomarker in treatment-naive patients with refractory NSCLC. In particular, Keynote-024[1] met its primary end points for overall survival (OS) and progression-free survival (PFS) in PD-L1 immunohistochemistry 50% or greater for pembrolizumab compared with platinum-based chemotherapy, validating PD-L1 immunohistochemistry as a biomarker for OS. Tumor mutation burden (TMB) has also emerged as a possible biomarker. The prevalence of somatic mutations among cancers ranges from 0.01 mutations/megabase (Mb) to more than 400 mutations/Mb. Some of these mutations lead to the translation of novel peptide epitopes or neoantigens that should enhance the immunogenicity of the tumor by eliciting T-cell repertoires. Initial studies of TMB were conducted by using whole-exome sequencing on tumor DNA and case-matched germline DNA.

In one study of advanced-stage NSCLC,[2] whole-exome sequencing was performed in 2 independent cohorts of patients with NSCLC (16 patients in one and 18 in the other) treated with pembrolizumab and team[5] recently calculated TMB scores by whole-exome sequencing in a subset of patients from the CheckMate-026 study,[6] a randomized phase 3 trial comparing nivolumab with platinum doublet chemotherapy as a first-line treatment in treatment-naive patients with NSCLC with PD-L1 expression greater than 5%. Patients with a high TMB (defined as having ≥243 missense mutations) had a prolonged PFS (median PFS of 9.7 vs 5.8 months; hazard ratio [HR], 0.62; 95% CI, 0.38-1.00) and higher objective response rate (46.8% vs 28.3%) but a nonsignificant OS difference with nivolumab treatment vs chemotherapy.

Guidelines from the European Society for Medical Oncology (ESMO) and ESMO Asia have already incorporated TMB as a possible biomarker in advanced NSCLC, recommending the combination of ipilimumab plus nivolumab as first-line treatment for patients with high TMB (>10 mutations/Mb). Supporting evidence stems from the CheckMate-227 trial, which reported results for first-line nivolumab plus ipilimumab vs platinum doublet chemotherapy.[7] That study showed an improved PFS in PD-L1-positive (HR, 0.62; 95% CI: 0.27-0.85) and -negative (HR, 0.48; 95% CI: 0.44-0.88) patients. At the time of publication, OS data did not meet the trial's prespecified end point for analysis. The trial had

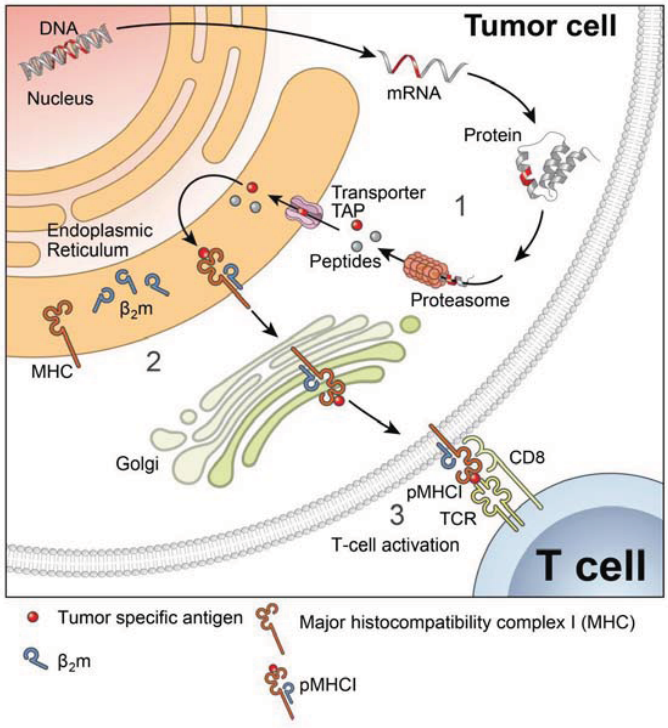
Figure. Pitfalls of Tumor Mutation Burden (TMB) for Clinical Application in Non-Small Cell Lung Cancer



SBi 한국생명정보학회 Korean Society for Bioinformatics

HLA TYPING IN THE ANTIGEN PROCESSING

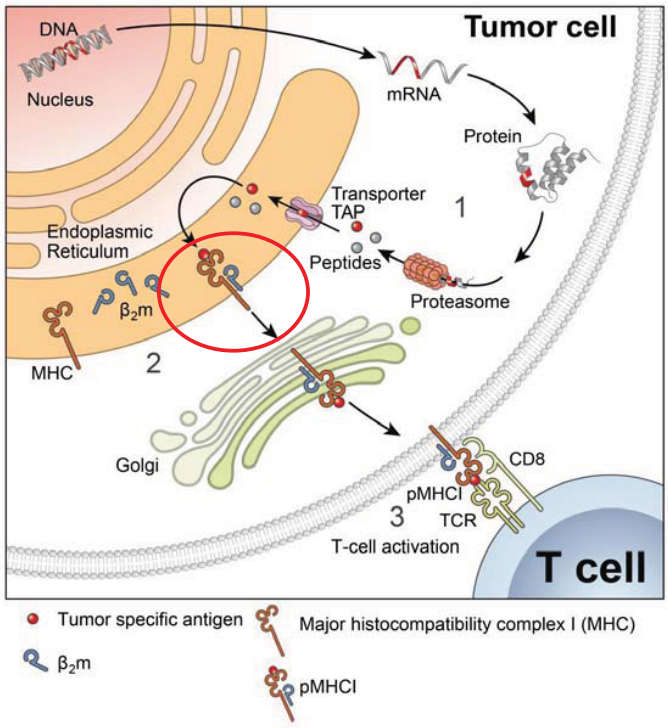
Neoantigen processing



somatic (passenger) mutations
↓
amino acid alternation
↓
protein degradation
↓
neopeptides
↓
binding to MHC molecule
↓
peptide-MHC presentation
↓
recognition by T-cell
↓
immune response

SBi 한국생명정보학회 Korean Society for Bioinformatics

Neoantigen processing



somatic (passenger) mutations

↓

amino acid alternation

↓

protein degradation

↓

neopeptides

↓

binding to MHC molecule

↓

peptide-MHC presentation

↓

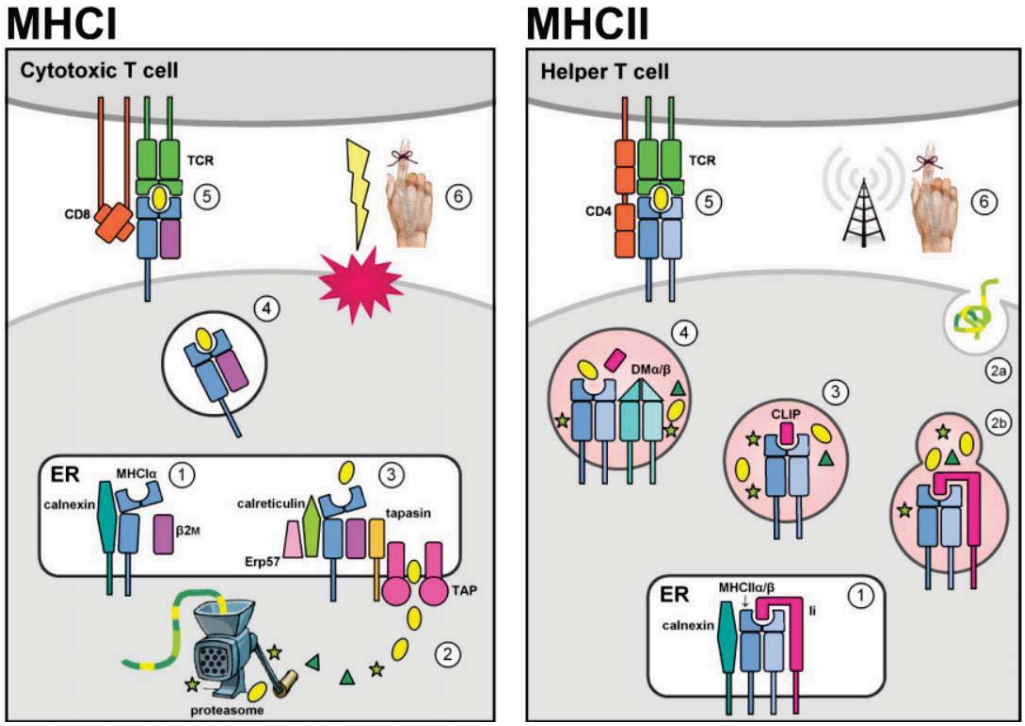
recognition by T-cell

↓

immune response

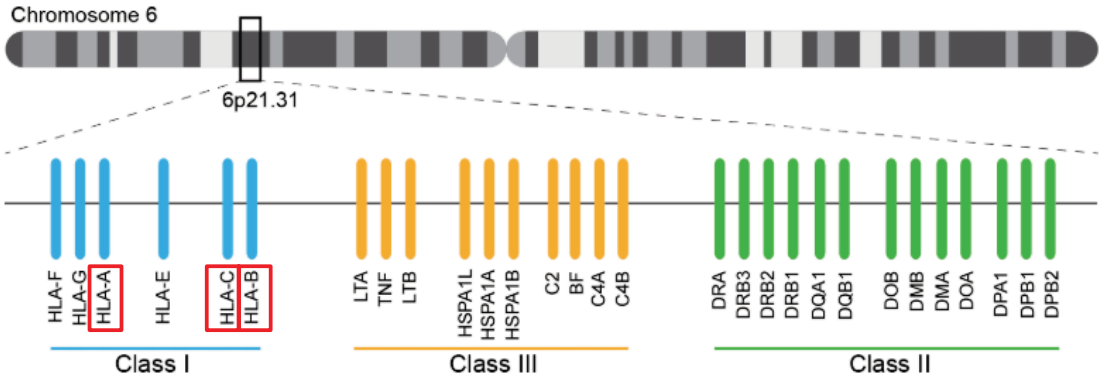
SBi 한국생명정보학회 Korean Society for Bioinformatics

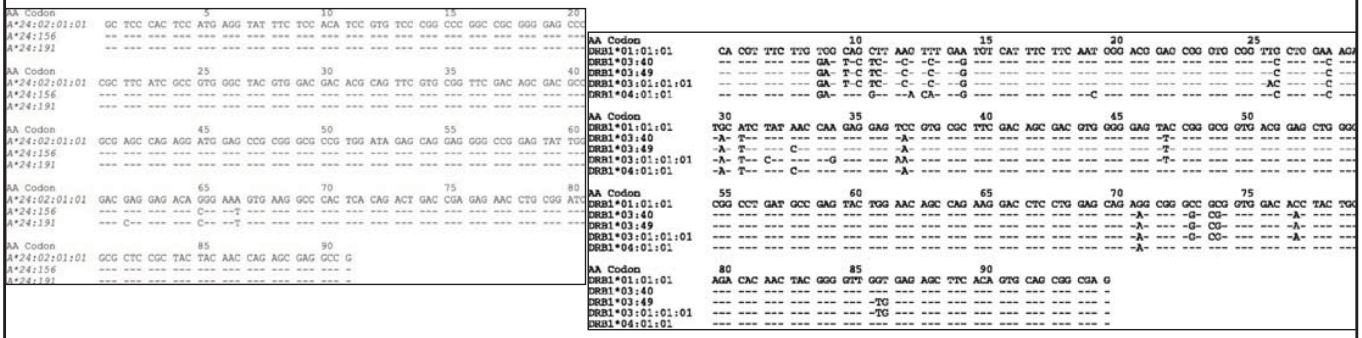
MHC (Major Histocompatibility Complex)



SBi 한국생명정보학회 Korean Society for Bioinformatics

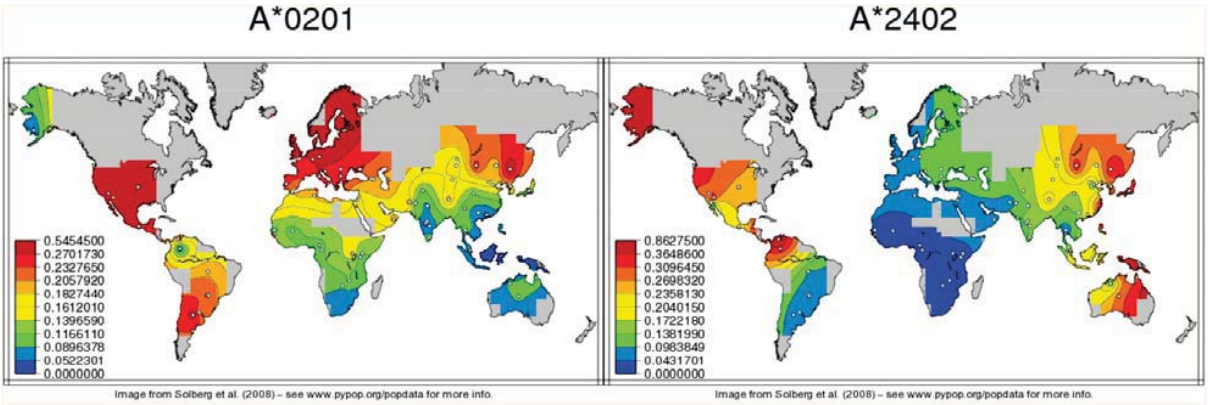
HLA (Human Leukocyte Antigen)

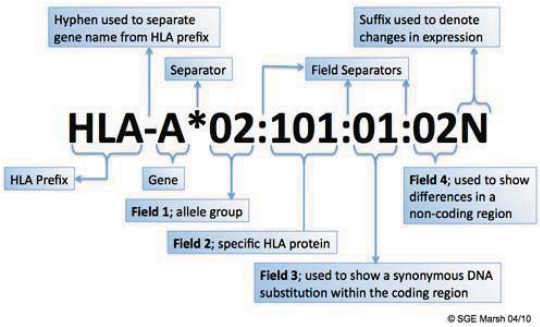




SBi 한국생명정보학회 Korean Society for Bioinformatics

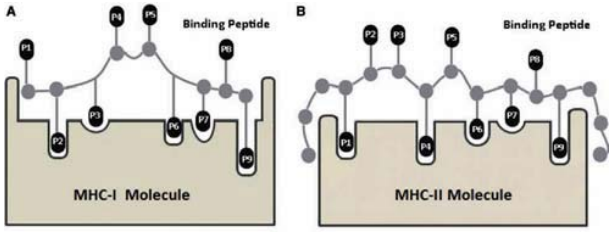
HLA alleles are ethnic specific





SBi 한국생명정보학회 Korean Society for Bioinformatics

MHC-peptide binding



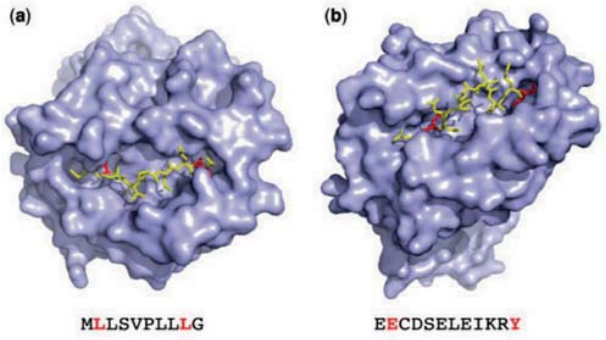
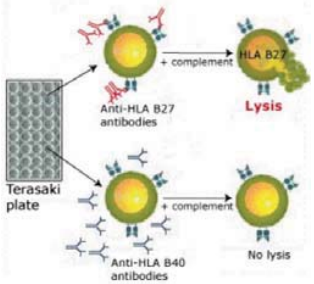


Fig. 5. 3D structures for two MHC class I molecules with bound peptides longer than 9 amino acids (PDB references 2CLR and 4JQX). (**a**) The 10mer peptide MLLSVPLLLG bound to HLA-A*02:01 extends at the C terminus with a glycine (G) amino acid. The residues at the anchor positions P2 (L) and P9 (L) are highlighted. (**b**) The 12mer EECDSELEIKRY bound to HLA-B*44:03 has anchors at its second (E) and last (Y) positions and bulges out from the middle of the MHC binding groove

But it is highly dependent on the HLA alleles
- That's why we need to know HLA allele (of the patient)

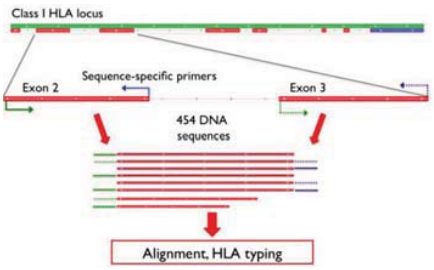
HLA typing methods

1. Serology-based typing

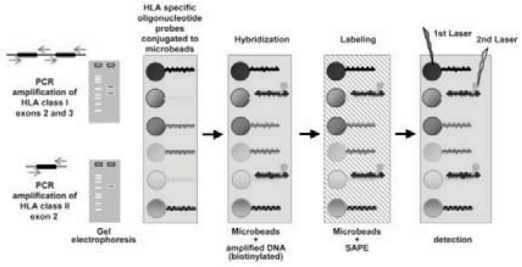


- Use of microcytotoxicity
  - complement mediated lysis
- Simple and low-cost
- Mostly used in HLA-A and HLA-B
- Can type allele groups an d alleles only

2. Sanger sequencing



3. Sequence-specific Oligonucleotide Hybridization (SSO)



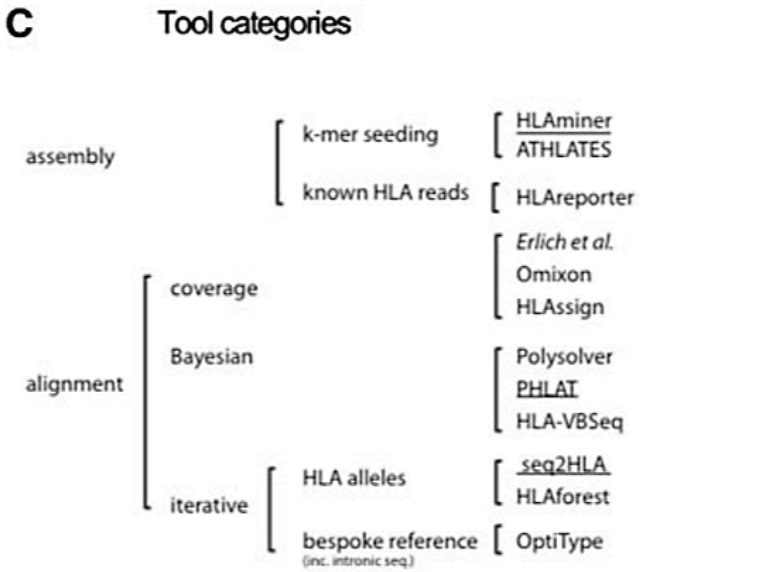
- Amplify targeted regions with biotin-labeled primers
- Hybridized sequences emit fluorescence

NGS-based HLA typing

- PROS
- Use of (already) produced NGS-data
- No extra-cost
- Fast

- Threat
- Short-read
- HLA genes are GC-rich: lower-sequencing coverage

NGS-based HLA typing



Bauer *et al, Briefings in Bioinformatics. 2018*

SBi 한국생명정보학회 Korean Society for Bioinformatics

Assembly-based HLA typing

Warren et al. Genome Medicine 2012, 4:95
http://genomemedicine.com/content/4/12/95

Genome Medicine

METHOD Open Access

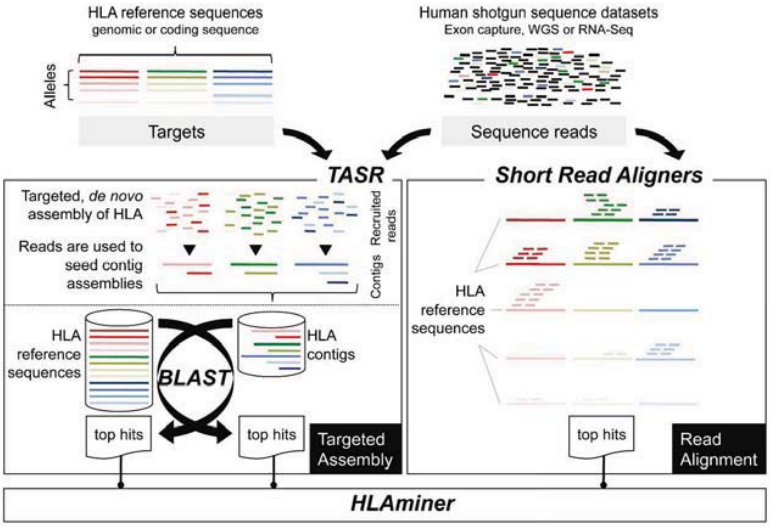
Derivation of HLA types from shotgun sequence datasets

René L Warren[1], Gina Choe[1], Douglas J Freeman[1], Mauro Castellarin[1], Sarah Munro[1], Richard Moore[1] and Robert A Holt[1,2*]

Abstract

The human leukocyte antigen (HLA) is key to many aspects of human physiology and medicine. All current sequence-based HLA typing methodologies are targeted approaches requiring the amplification of specific HLA gene segments. Whole genome, exome and transcriptome shotgun sequencing can generate prodigious data but due to the complexity of HLA loci these data have not been immediately informative regarding HLA genotype. We describe HLAminer, a computational method for identifying HLA alleles directly from shotgun sequence datasets (http://www.bcgsc.ca/platform/bioinfo/software/hlaminer). This approach circumvents the additional time and cost of generating HLA-specific data and capitalizes on the increasing accessibility and affordability of massively parallel sequencing.

HLAminer



Alignment-based HLA typing

ANALYSIS

computational BIOLOGY

nature biotechnology

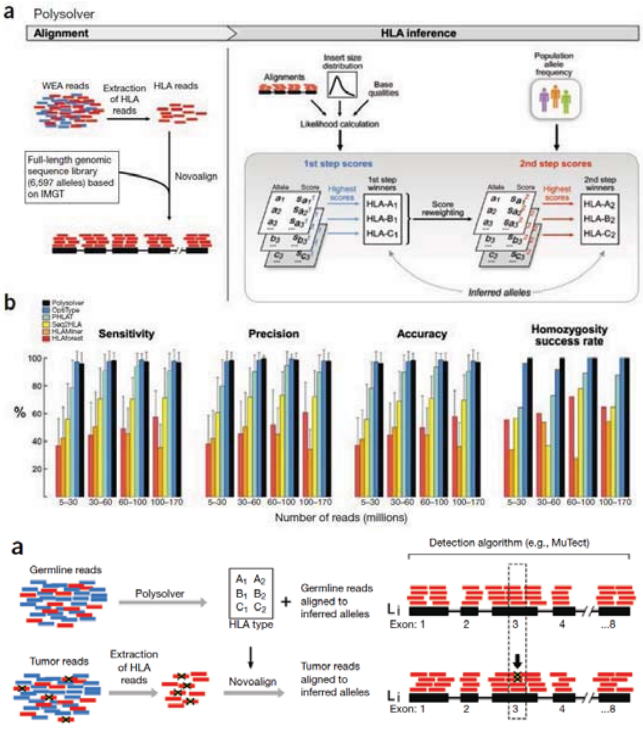
Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes

Sachet A Shukla[1–3], Michael S Rooney[2,4], Mohini Rajasagi[1,5], Grace Tiao[2], Philip M Dixon[3], Michael S Lawrence[2], Jonathan Stevens[6], William J Lane[6,7], Jamie L Dellagatta[6], Scott Steelman[3], Carrie Sougnez[2], Kristian Cibulskis[2], Adam Kiezun[2], Nir Hacohen[2,8,9], Vladimir Brusic[1,5], Catherine J Wu[1,2,5,8,11] & Gad Getz[2,10,11]

Detection of somatic mutations in human leukocyte antigen (HLA) genes using whole-exome sequencing

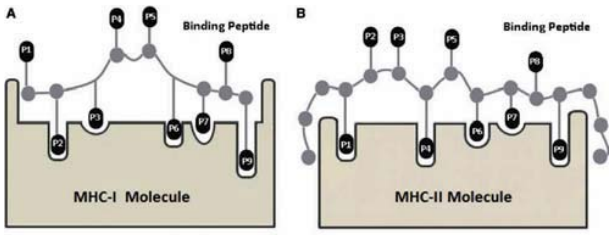
adenocarcinoma and diffuse large B-cell lymphoma[1–5]. The HLA locus, located on chromosome 6, is among the most polymorphic

Polysolver



MHC BINDING PREDICTION

MHC-peptide binding



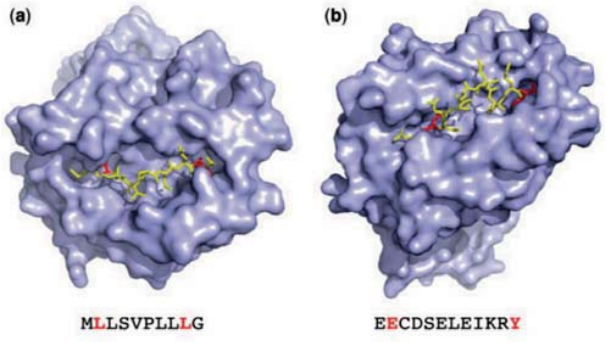


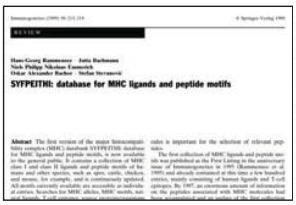
Fig. 5. 3D structures for two MHC class I molecules with bound peptides longer than 9 amino acids (PDB references 2CLR and 4JQX). (**a**) The 10mer peptide MLLSVPLLLG bound to HLA-A*02:01 extends at the C terminus with a glycine (G) amino acid. The residues at the anchor positions P2 (L) and P9 (L) are highlighted. (**b**) The 12mer EECDSELEIKRY bound to HLA-B*44:03 has anchors at its second (E) and last (Y) positions and bulges out from the middle of the MHC binding groove

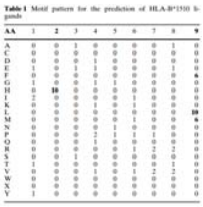
Can we predict if a given peptide will bind to MHC?

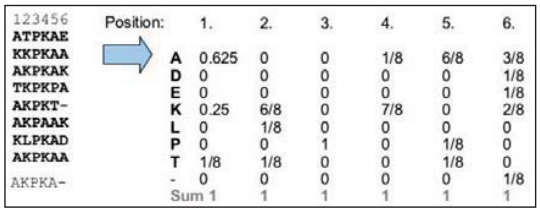
SBi 한국생명정보학회 Korean Society for Bioinformatics

Prediction algorithms

- SYFPEITHI: using PSSM

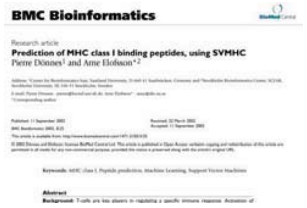

SYFPEITHI: database for MHC ligands and peptide motifs



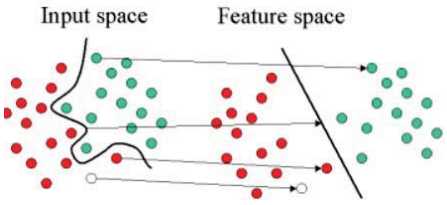


123456	Position:	1.	2.	3.	4.	5.	6.
ATPKAE							
KKPKAA	A	0.625	0	0	1/8	6/8	3/8
AKPKAK	D	0	0	0	0	0	1/8
TKPKPA	E	0	0	0	0	0	1/8
AKPKT-	K	0.25	6/8	0	7/8	0	2/8
AKPAAK	L	0	1/8	0	0	0	0
KLPKAD	P	0	0	1	0	1/8	0
AKPKAA	T	1/8	1/8	0	0	1/8	0
AKPKA-	-	0	0	0	0	0	1/8
	Sum 1		1	1	1	1	1

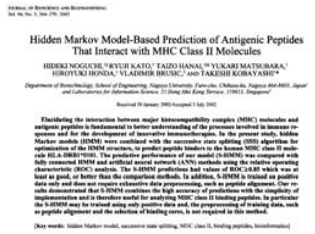
- SVMHC: using Support Vector Machine

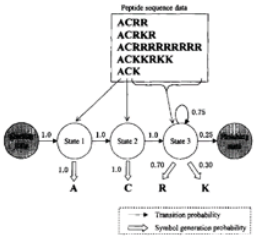

BMC Bioinformatics

Prediction of MHC class I binding peptides, using SVMHC



- S-HMM: using Hidden Markov Model

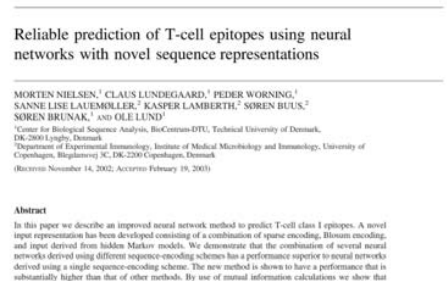

Hidden Markov Model-Based Prediction of Antigenic Peptides That Interact with MHC Class II Molecules

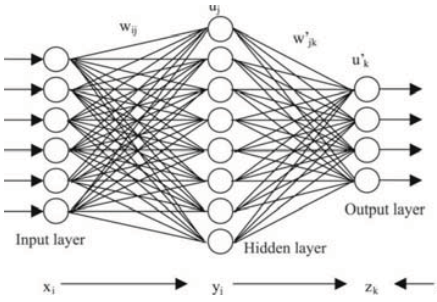


SBi 한국생명정보학회 Korean Society for Bioinformatics

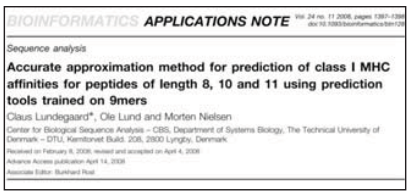
ANN based algorithms

NetMHC: Classification of MHC-I binding peptides using ANN


Reliable prediction of T-cell epitopes using neural networks with novel sequence representations

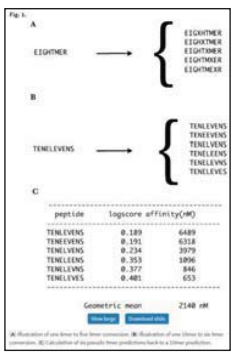


NetMHC-3.0

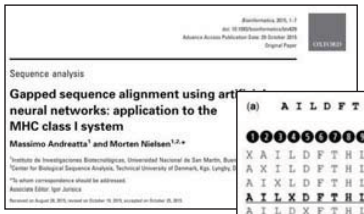

BIOINFORMATICS APPLICATIONS NOTE

Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers

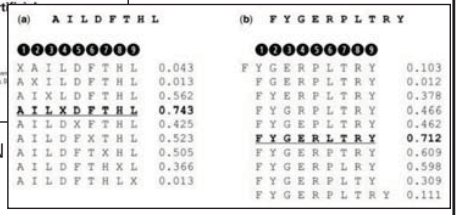
Approximation of 8, 10, 11 from 9 mer model



NetMHC-4.0


Gapped sequence alignment using artificial neural networks: application to the MHC class I system

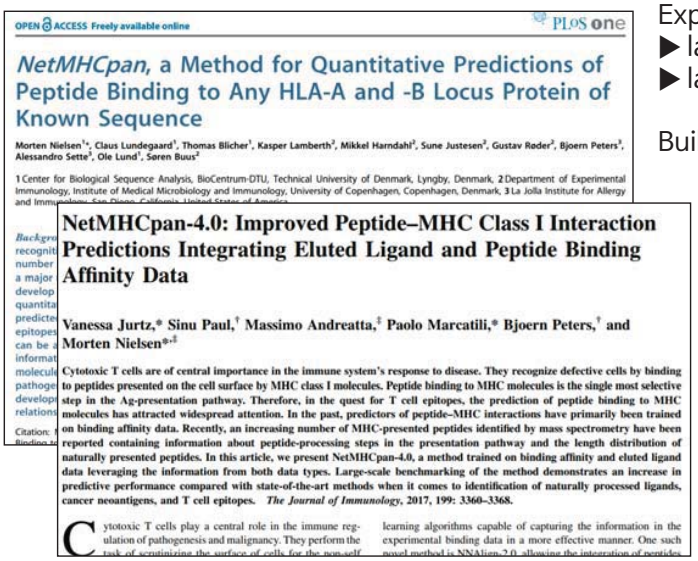
Gapped alignment to ANN : 9 to 8~11 mer



SBi 한국생명정보학회 Korean Society for Bioinformatics

Regarding all HLA-types at once

NetMHCpan: Prediction on all HLA-A/B alleles, simultaneously



OPEN ACCESS Freely available online — PLoS one

NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence

Morten Nielsen[1]*, Claus Lundegaard[1], Thomas Blicher[1], Kasper Lamberth[2], Mikkel Harndahl[2], Sune Justesen[2], Gustav Røder[2], Bjoern Peters[3], Alessandro Sette[3], Ole Lund[1], Søren Buus[2]

1 Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, Lyngby, Denmark, 2 Department of Experimental Immunology, Institute of Medical Microbiology and Immunology, University of Copenhagen, Copenhagen, Denmark, 3 La Jolla Institute for Allergy and Immunology, San Diego, California, United States of America

Background — recognit… number … a major … develop … quantita… predicte… epitopes … can be … informat… molecul… pathoge… develop… relations…

Citation: …

NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data

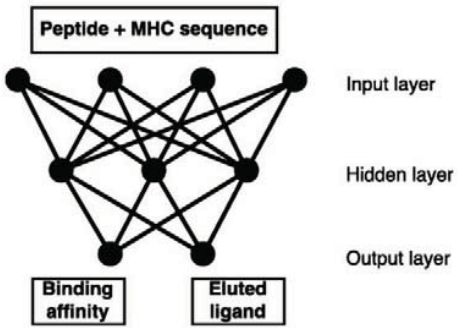
Vanessa Jurtz,* Sinu Paul,† Massimo Andreatta,‡ Paolo Marcatili,* Bjoern Peters,† and Morten Nielsen*,‡

Cytotoxic T cells are of central importance in the immune system's response to disease. They recognize defective cells by binding to peptides presented on the cell surface by MHC class I molecules. Peptide binding to MHC molecules is the single most selective step in the Ag-presentation pathway. Therefore, in the quest for T cell epitopes, the prediction of peptide binding to MHC molecules has attracted widespread attention. In the past, predictors of peptide–MHC interactions have primarily been trained on binding affinity data. Recently, an increasing number of MHC-presented peptides identified by mass spectrometry have been reported containing information about peptide-processing steps in the presentation pathway and the length distribution of naturally presented peptides. In this article, we present NetMHCpan-4.0, a method trained on binding affinity and eluted ligand data leveraging the information from both data types. Large-scale benchmarking of the method demonstrates an increase in predictive performance compared with state-of-the-art methods when it comes to identification of naturally processed ligands, cancer neoantigens, and T cell epitopes. *The Journal of Immunology*, 2017, 199: 3360–3368.

Cytotoxic T cells play a central role in the immune regulation of pathogenesis and malignancy. They perform the task of scrutinizing the surface of cells for the non-self… learning algorithms capable of capturing the information in the experimental binding data in a more effective manner. One such novel method is NNAlign_2.0, allowing the integration of peptides…

Experimental data are biased to major HLA alleles
▶ lack of training data in rare alleles
▶ lack of accuracy

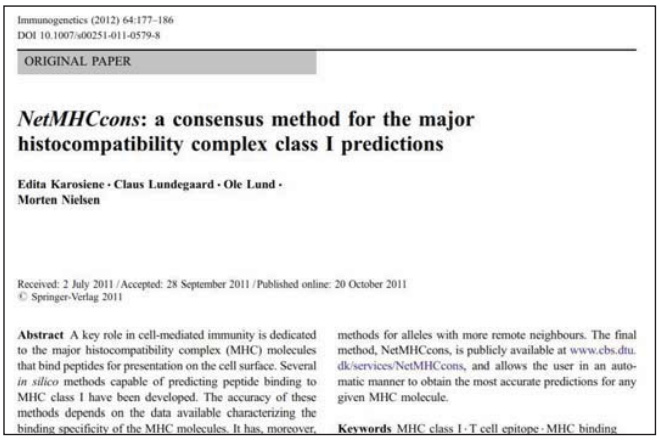
Build a classifier that work on HLA-peptide pair



SBi 한국생명정보학회 Korean Society for Bioinformatics

Too many methods. Need a consensus

NetMHCcons: Prediction on all HLA-A/B alleles, simultaneously



Immunogenetics (2012) 64:177–186
DOI 10.1007/s00251-011-0579-8

ORIGINAL PAPER

***NetMHCcons*: a consensus method for the major histocompatibility complex class I predictions**

Edita Karosiene · Claus Lundegaard · Ole Lund · Morten Nielsen

Received: 2 July 2011 / Accepted: 28 September 2011 / Published online: 20 October 2011
© Springer-Verlag 2011

Abstract A key role in cell-mediated immunity is dedicated to the major histocompatibility complex (MHC) molecules that bind peptides for presentation on the cell surface. Several *in silico* methods capable of predicting peptide binding to MHC class I have been developed. The accuracy of these methods depends on the data available characterizing the binding specificity of the MHC molecules. It has, moreover, methods for alleles with more remote neighbours. The final method, NetMHCcons, is publicly available at www.cbs.dtu.dk/services/NetMHCcons, and allows the user in an automatic manner to obtain the most accurate predictions for any given MHC molecule.

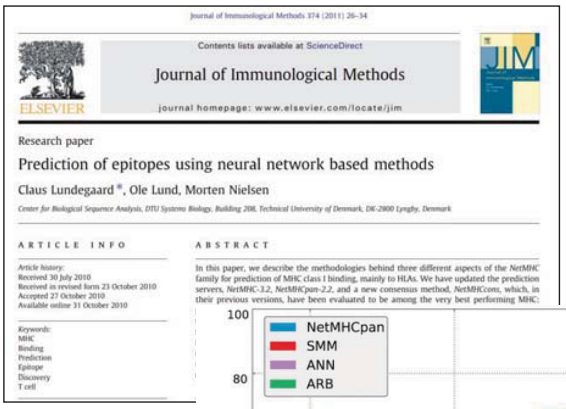
Keywords MHC class I · T cell epitope · MHC binding

$$\text{NetMHCcons} = \begin{cases} \text{NetMHCpan} & \text{for } N_p < 50 \text{ and } N_b < 10 \\ \text{NetMHC} + \text{NetMHCpan} & \text{otherwise} \end{cases}$$

We demonstrate that a simple combination of NetMHC and NetMHCpan gives the highest performance when the allele in question is included in the training and is characterized by at least 50 data points with at least ten binders. Otherwise, NetMHCpan is the best predictor.

SBi 한국생명정보학회 Korean Society for Bioinformatics

Benchmarks and competitions



Journal of Immunological Methods 374 (2011) 26–34

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Research paper

Prediction of epitopes using neural network based methods

Claus Lundegaard *, Ole Lund, Morten Nielsen

Center for Biological Sequence Analysis, DTU Systems Biology, Building 208, Technical University of Denmark, DK-2800 Lyngby, Denmark

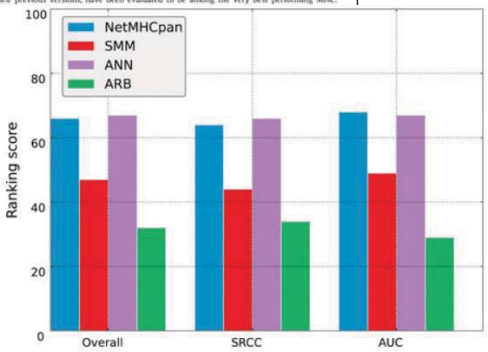
ARTICLE INFO

Article history:
Received 30 July 2010
Received in revised form 23 October 2010
Accepted 27 October 2010
Available online 31 October 2010

Keywords:
MHC
Binding
Prediction
Epitope
Discovery
T cell

ABSTRACT

In this paper, we describe the methodologies behind three different aspects of the *NetMHC* family for prediction of MHC class I binding, mainly to HLAs. We have updated the prediction servers, *NetMHC-3.2*, *NetMHCpan-2.2*, and a new consensus method, *NetMHCcons*, which, in their previous versions, have been evaluated to be among the very best performing MHC:



2nd Machine Learning Competition in Immunology 2012

Sponsors: InCoB 2012 and ICIW 2012

Prediction task:
Predict peptides naturally processed by MHC Class I pathway ("eluted peptides") for each target MHC molecule. For e target molecule, the competitors are asked to submit a set of predicted eluted peptides from the test set.

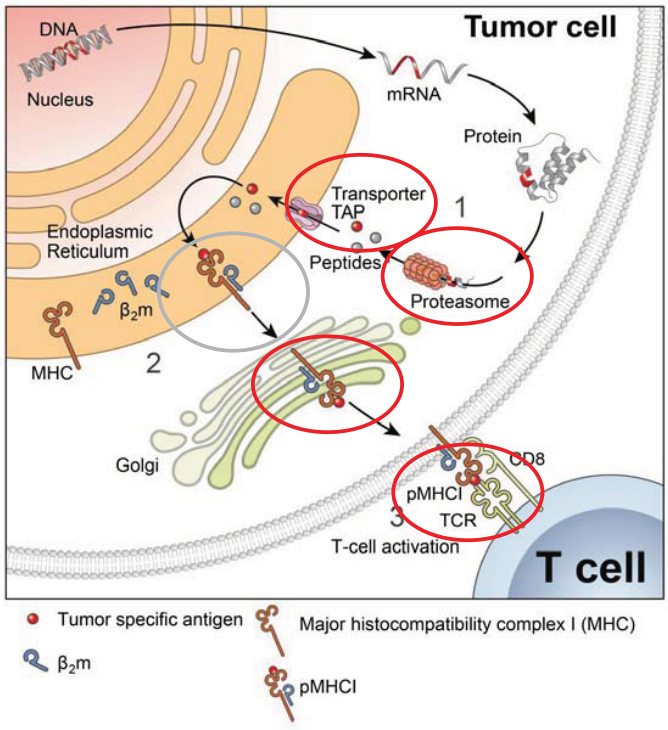
Winners MLCI 2012

A total of 32 submissions were submitted for the competition. Of these, 24 submissions (Group 1) provided a set of thresholds (elution score based predictors) for each peptide and each MHC molecule. Another 8 submissions (Group 2) provided lists of peptides that were predicted as eluted from specific MHC molecules (eluted peptide list based predictors) for each of 8 studied MHC alleles. The NetMHC 3.2 server (1D-BENCH) results were used as a benchmark method.

Winning Team	Predictor No.	Prediction Method	Winning Category
Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M, Technical University of Denmark	1D-BENCH	NetMHC 3.2 (Reference)	Group 1: A*0201
Giguere S, Drouin A, Lacoste A, Laval University, Canada	2F	A Bayesian model averaging method over several SVMs using the GS kernel.	Group 1: B*0702, H-2D^{b} and H-2K^{b}
Nielsen M, et al., Technical University of Denmark	9D	A combination of NetMHC, NetMHCpan and MHCkernel predictions.	Group 1: B*3501 and B*4403
Giguere S, Drouin A, Lacoste A, Laval University, Canada	2D	A SVM classifier and a novel string kernel (GS kernel)	Group 1: B*5301
Xiang Z, He Y, University of Michigan Medical School, Ann Arbor, MI, USA.	20D	A position-specific scoring matrix (PSSM) with statistical P-value as the cutoff.	Group 1: B*5701
Yu Ting Wei, Department of Probability and Statistics, School of Mathematical Sciences, Peking University; Wen Jun Shen and Hau-San Wong, Department of Computer Science, City University of Hong Kong	14A	ConsMHC: a consensus program incorporating the results of kernelRLSpan-I, NetMHC, NetMHCpan and PickPocket by SVM	Group 2

ANTIGEN PROCESSING STEPS

Neoantigen processing revisited



somatic (passenger) mutations

↓

amino acid alternation

↓

protein degradation

↓

neopeptides

↓

binding to MHC molecule

↓

peptide-MHC presentation

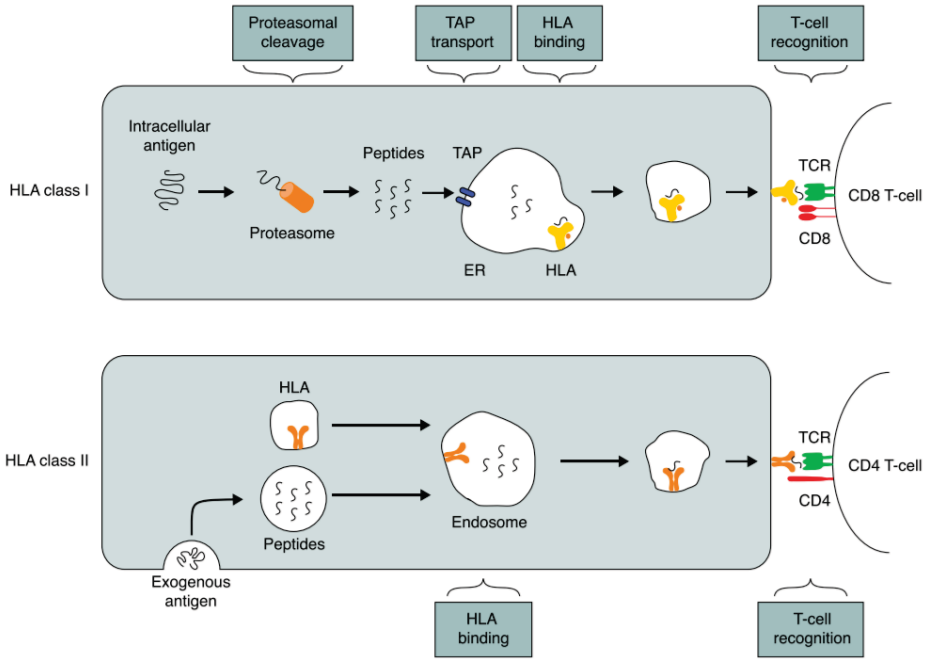
↓

recognition by T-cell

↓

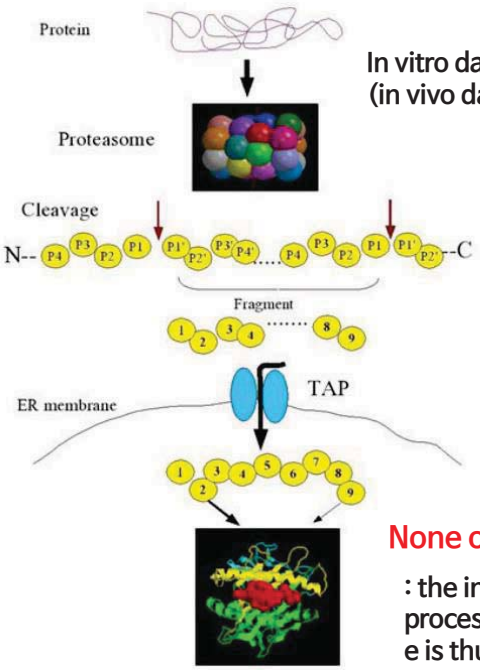
immune response

Antigen Processing Pathways for MHC class I/II



Backert and Kohlbacher, *Genome Medicine, 2015*

Proteasomal cleavage



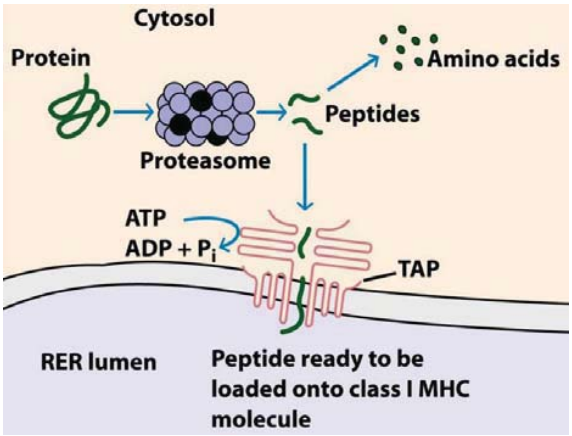
In vitro data created with purified proteasomes in the laboratory
(in vivo data are harder to collect)

C-terminus: commonly determined by proteasomal cleavage
N-terminus: can undergo further trimming by proteases located in the cytosol or ER

None of the predictors achieved an MCC above 0.3

: the in vitro data do not capture the full complexity of proteasomal processing in vivo. The value of predictions of proteasomal cleavage is thus rather limited

TAP transport prediction



- Primarily owing to the scarcity of data, there are few published methods on TAP transport prediction.
- No unbiased blind benchmarks for TAP transport methods have been published so far, and a comparative assessment of the various methods is thus currently difficult

Considering MHC-binding stability, not affinity

European Journal of Immunology

Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity

Mikkel Harndahl[1], Michael Rasmussen[1], Gustav Roder[1], Ida Dalgaard Pedersen[1], Mikael Sørensen[2], Morten Nielsen[2] and Søren Buus[1]

[1] Laboratory of Experimental Immunology, Faculty of Health Sciences, University of Copenhagen, Denmark
[2] Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Denmark

Efficient presentation of peptide-MHC class I (pMHC-I) complexes to immune T cells should benefit from a stable peptide-MHC-I interaction. However, it has been difficult to distinguish stability from other requirements for MHC-I binding, for example, affinity. We have recently established a high-throughput assay for pMHC-I stability. Here, we have generated a large database containing stability measurements of pMHC-I complexes, and re-examined a previously reported unbiased analysis of the relative contributions of antigen processing and presentation in defining cytotoxic T lymphocyte (CTL) immunogenicity [Assarsson et al., J. Immunol. 2007. 178: 7890–7901]. Using an affinity-balanced approach, we demonstrated that immunogenic peptides tend to be more stably bound to MHC-I molecules compared with nonimmunogenic peptides. We also developed a bioinformatics method to predict pMHC-I stability, which suggested that 30% of the nonimmunogenic binders hitherto classified as "holes in the T-cell repertoire" can be explained as being unstably bound to MHC-I. Finally, we suggest that nonoptimal anchor

Binding (kinetic) stability

We also developed a bioinformatics method to predict pMHC-I stab ility, which suggested that 30% of the nonimmunogenic binders hith erto classified as "holes in the T-cell repertoire" can be explained as being unstably bound to MHC-I.

Prediction on the stability

NetMHCstab: predicting stability of pMHC-I complexes

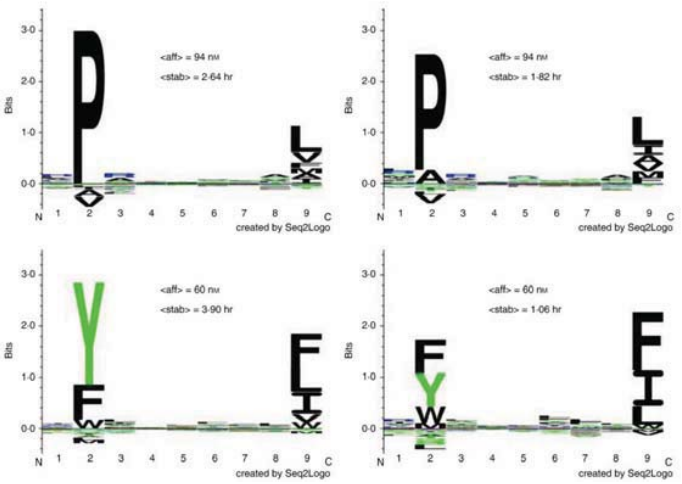
Immunology
The Journal of cells, molecules, systems and technologies
British Society for immunology
IMMUNOLOGY ORIGINAL ARTICLE

NetMHCstab – predicting stability of peptide–MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery

Kasper W. Jørgensen,[1,*] Michael Rasmussen,[2,*] Søren Buus[2] and Morten Nielsen[1,3]
[1]Department of Systems Biology, Centre for Biological Sequence Analysis, Technical University of Denmark, Lyngby, [2]Laboratory of Experimental Immunology, University of Copenhagen, Copenhagen N, Denmark, and [3]Instituto de Investigaciones Biotecnologicas, Universidad Nacional de San Martin, San Martin, Buenos Aires, Argentina

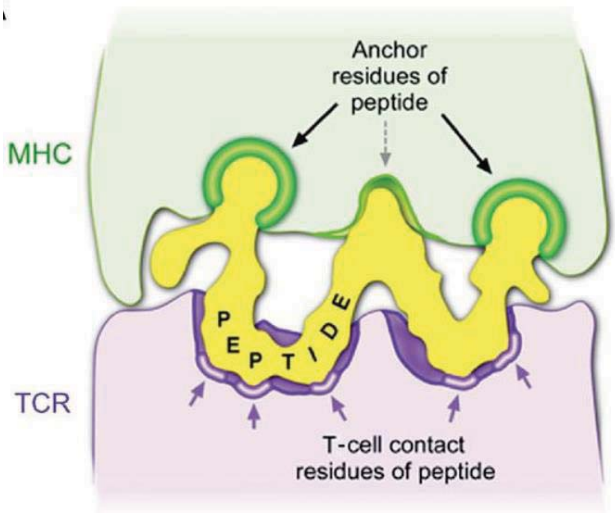
Summary

Major histocompatibility complex class I (MHC-I) molecules play an essential role in the cellular immune response, presenting peptides to cytotoxic T lymphocytes (CTLs) allowing the immune system to scrutinize ongoing intracellular production of proteins. In the early 1990s, immunogenicity and stability of the peptide–MHC-I (pMHC-I) complex were shown to be correlated. At that time, measuring stability was cumbersome and time consuming and only small data sets were analysed. Here, we investigate this fairly unexplored area on a large scale compared with earlier studies. A recent small-scale study demonstrated that pMHC-I complex stability was a better correlate of CTL immunogenicity than peptide–MHC-I affinity. We here extended this study and analysed a total of 5509 distinct peptide stability measurements covering 10 different HLA class I molecules. Artificial neural networks were used to construct stability predictors capable of predicting the half-life of the pMHC-I complex. These



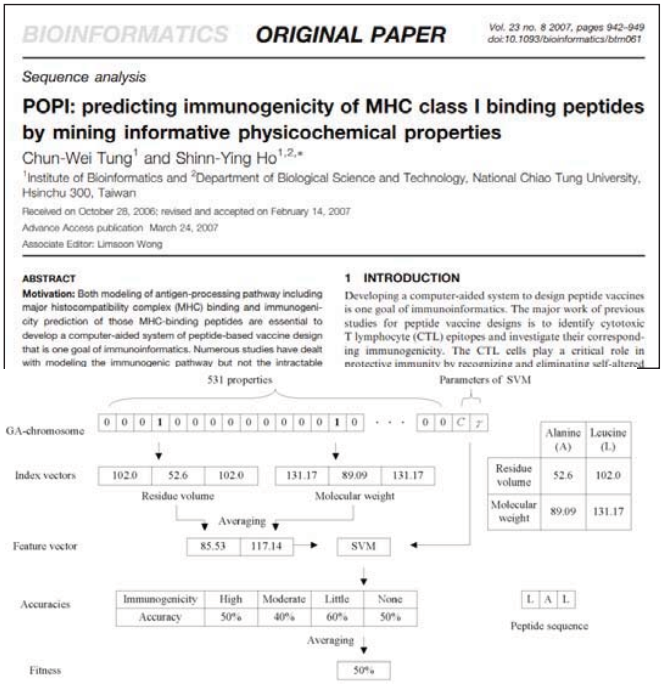
stable

Prediction on pMHC-TCR binding



Fritsch *et al, Cancer Immunology Research. 2014*

TCR immunogenicity prediction



BIOINFORMATICS **ORIGINAL PAPER** *Vol. 23 no. 8 2007, pages 942–949 doi:10.1093/bioinformatics/btm061*

Sequence analysis

POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties

Chun-Wei Tung[1] and Shinn-Ying Ho[1,2,*]

[1]Institute of Bioinformatics and [2]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 300, Taiwan

Received on October 28, 2006; revised and accepted on February 14, 2007

Advance Access publication March 24, 2007

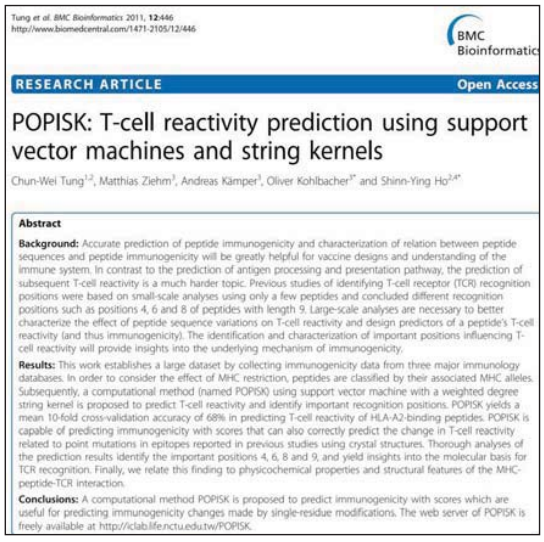
Associate Editor: Limsoon Wong

ABSTRACT

Motivation: Both modeling of antigen-processing pathway including major histocompatibility complex (MHC) binding and immunogenicity prediction of those MHC-binding peptides are essential to develop a computer-aided system of peptide-based vaccine design that is one goal of immunoinformatics. Numerous studies have dealt with modeling the immunogenic pathway but not the intractable

1 INTRODUCTION

Developing a computer-aided system to design peptide vaccines is one goal of immunoinformatics. The major work of previous studies for peptide vaccine designs is to identify cytotoxic T lymphocyte (CTL) epitopes and investigate their corresponding immunogenicity. The CTL cells play a critical role in protective immunity by recognizing and eliminating self-altered



Tung *et al. BMC Bioinformatics* 2011, **12**:446
http://www.biomedcentral.com/1471-2105/12/446

BMC Bioinformatics

RESEARCH ARTICLE **Open Access**

POPISK: T-cell reactivity prediction using support vector machines and string kernels

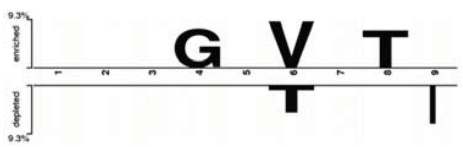
Chun-Wei Tung[1,2], Matthias Ziehm[3], Andreas Kämper[3], Oliver Kohlbacher[3*] and Shinn-Ying Ho[2,4*]

Abstract

Background: Accurate prediction of peptide immunogenicity and characterization of relation between peptide sequences and peptide immunogenicity will be greatly helpful for vaccine designs and understanding of the immune system. In contrast to the prediction of antigen processing and presentation pathway, the prediction of subsequent T-cell reactivity is a much harder topic. Previous studies of identifying T-cell receptor (TCR) recognition positions were based on small-scale analyses using only a few peptides and concluded different recognition positions such as positions 4, 6 and 8 of peptides with length 9. Large-scale analyses are necessary to better characterize the effect of peptide sequence variations on T-cell reactivity and design predictors of a peptide's T-cell reactivity (and thus immunogenicity). The identification and characterization of important positions influencing T-cell reactivity will provide insights into the underlying mechanism of immunogenicity.

Results: This work establishes a large dataset by collecting immunogenicity data from three major immunology databases. In order to consider the effect of MHC restriction, peptides are classified by their associated MHC alleles. Subsequently, a computational method (named POPISK) using support vector machine with a weighted degree string kernel is proposed to predict T-cell reactivity and identify important recognition positions. POPISK yields a mean 10-fold cross-validation accuracy of 68% in predicting T-cell reactivity of HLA-A2-binding peptides. POPISK is capable of predicting immunogenicity with scores that can also correctly predict the change in T-cell reactivity related to point mutations in epitopes reported in previous studies using crystal structures. Thorough analyses of the prediction results identify the important positions 4, 6, 8 and 9, and yield insights into the molecular basis for TCR recognition. Finally, we relate this finding to physicochemical properties and structural features of the MHC-peptide-TCR interaction.

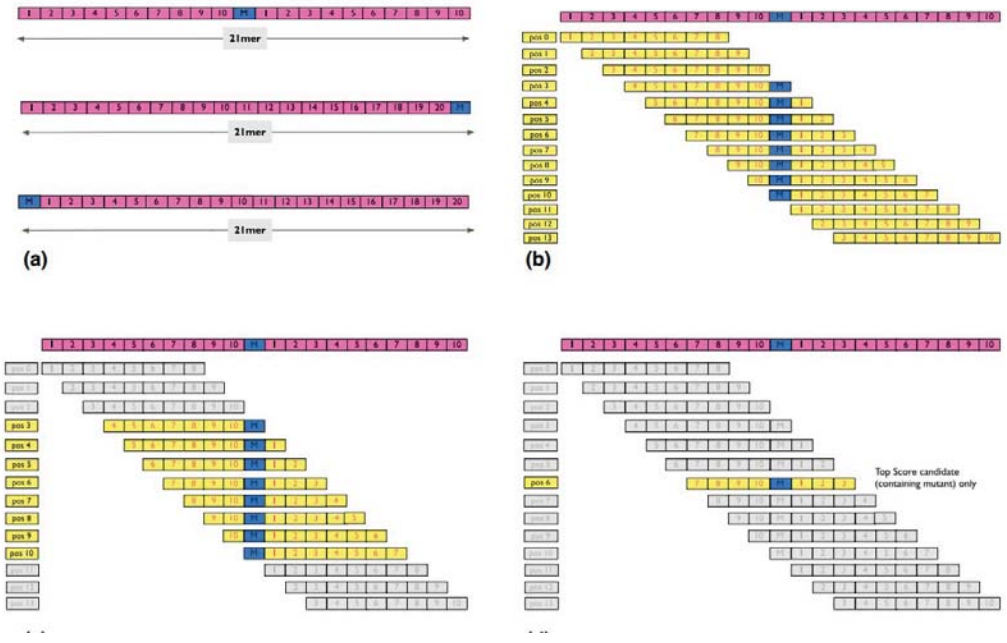
Conclusions: A computational method POPISK is proposed to predict immunogenicity with scores which are useful for predicting immunogenicity changes made by single-residue modifications. The web server of POPISK is freely available at http://iclab.life.nctu.edu.tw/POPISK.



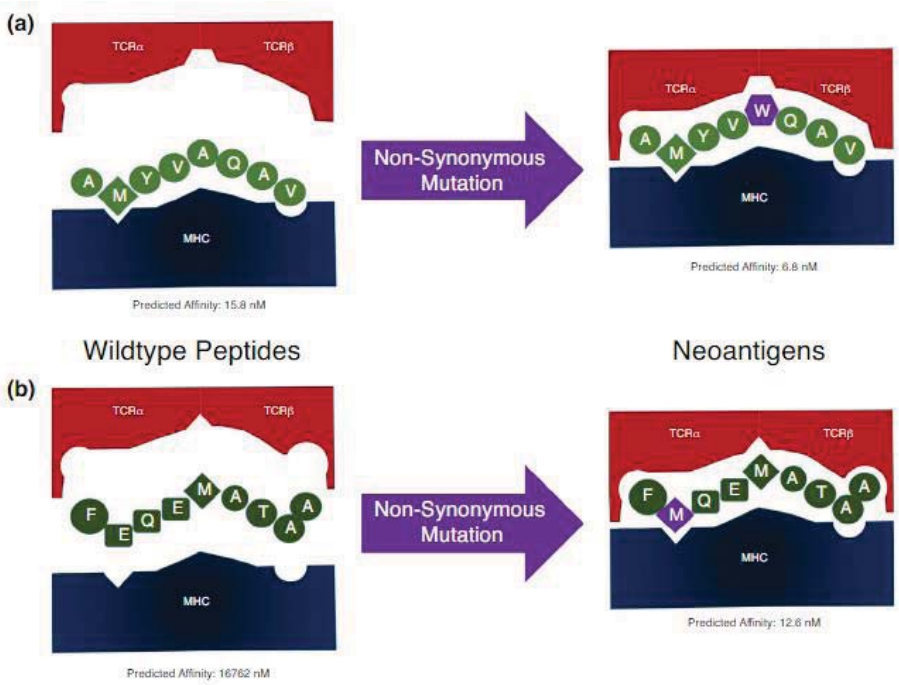
The current performance of immunogenicity predictors is certainly not satisfying.
The amount and reliability of experimental data on T-cell reactivity is certainly one reason for this. But clearly our lack of understanding of the details of the processes leading to central and peripheral tolerance hamper the development of more predictive methods too (Toussant *et al*, BCB11, 2011)

NEOANTIGEN ANALYSIS & INTEGRATED PIPELINES

Somatic mutation derived neopeptide



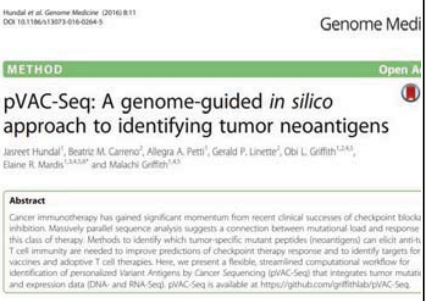
And Neoantigens



Oiseth et al, *J Cancer Metastasis and Treatment, 2017*

SBi 한국생명정보학회 Korean Society for Bioinformatics

Overall Pipeline



Genome Medi

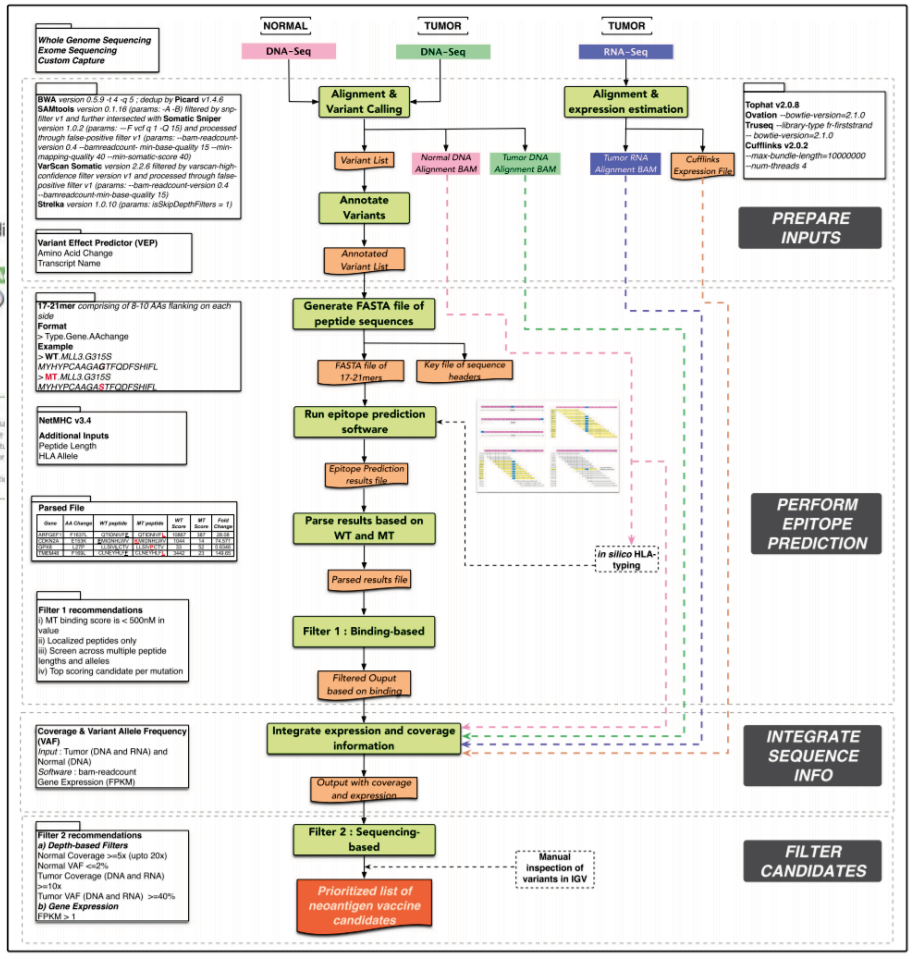
METHOD

pVAC-Seq: A genome-guided *in silico* approach to identifying tumor neoantigens

Jasreet Hundal[1], Beatriz M. Carreno[2], Allegra A. Petti[1], Gerald P. Linette[2], Obi L. Griffith[1,2,4,5], Elaine R. Mardis[1,3,4,5,6*] and Malachi Griffith[1,4,5]

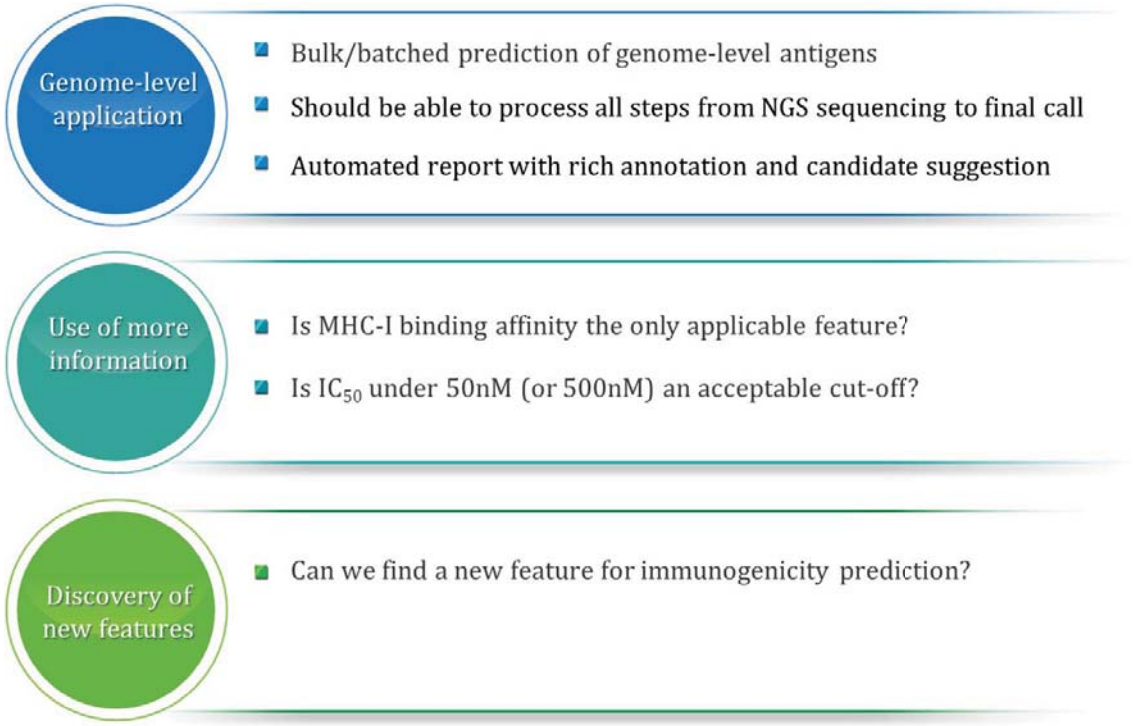
Abstract

Cancer immunotherapy has gained significant momentum from recent clinical successes of checkpoint blocka inhibition. Massively parallel sequence analysis suggests a connection between mutational load and response this class of therapy. Methods to identify which tumor-specific mutant peptides (neoantigens) can elicit anti-t T cell immunity are needed to improve predictions of checkpoint therapy response and to identify targets for vaccines and adoptive T cell therapies. Here, we present a flexible, streamlined computational workflow for identification of personalized Variant Antigens by Cancer Sequencing (pVAC-Seq) that integrates tumor mutati and expression data (DNA- and RNA-Seq). pVAC-Seq is available at https://github.com/griffithlab/pVAC-Seq.



SBi 한국생명정보학회 Korean Society for Bioinformatics

Things need to be resolved for practical application



SBi 한국생명정보학회 Korean Society for Bioinformatics

NGS based Genome-level application

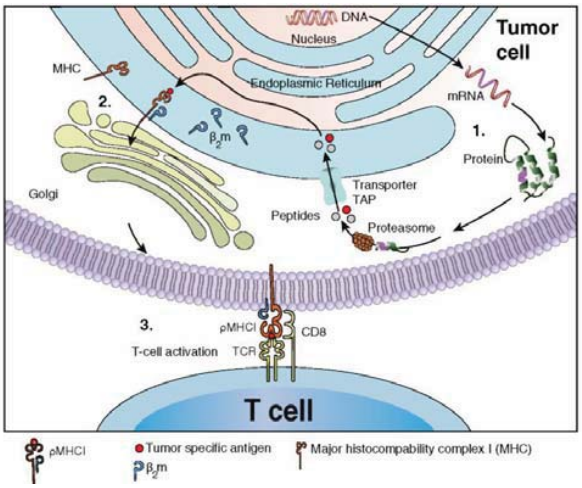
For genome-level application, the followings should be automated or properly handled:

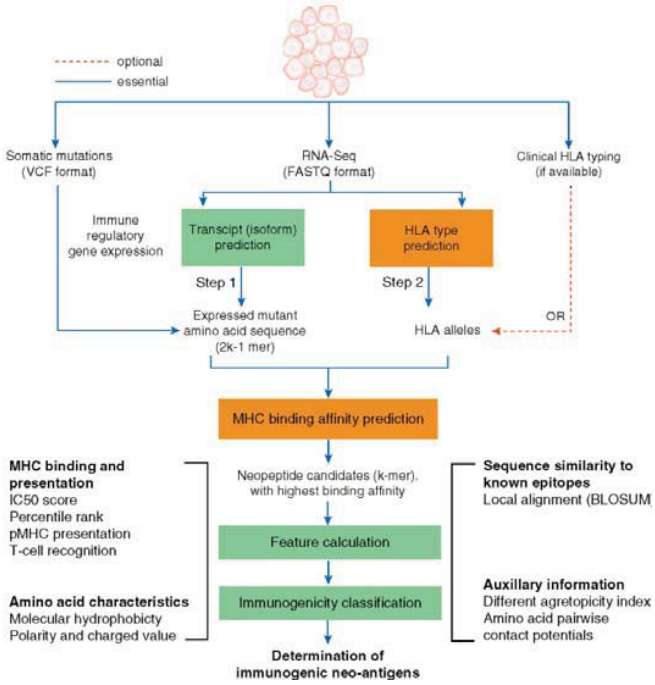
1. **Accurate calling of somatic mutations from NGS data**
2. **Conversion of genetic variants to protein sequence alteration**
   1. must consider transcript structures, or which to use for backbone
   2. need to cut into shorter peptides (e.g. 9-mer)
3. **Inference of HLA alleles**
4. **Expression level analysis of:**
   1. immune-regulatory genes
   2. genes containing candidate neopeptides
5. **Calculation of immunogenicity features including:**
   1. MHC-binding affinity (IC50)
   2. And other information (as much as possible)
6. **Effective binding of information sources and determination of final call**

SBi 한국생명정보학회 Korean Society for Bioinformatics

Neopepsee

Neopepsee: accurate genome-level prediction of neoantigens

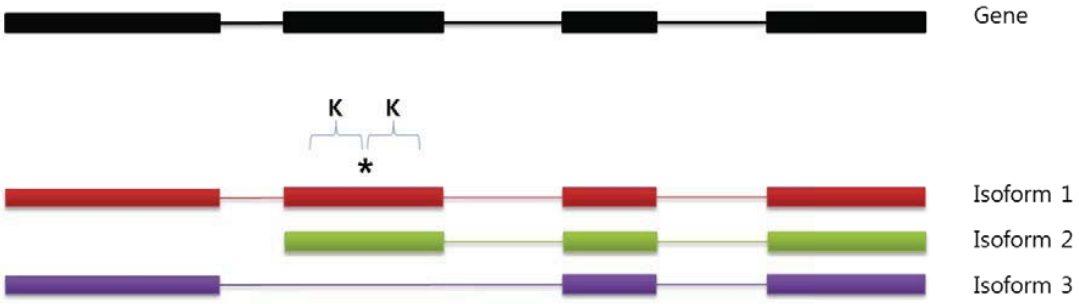




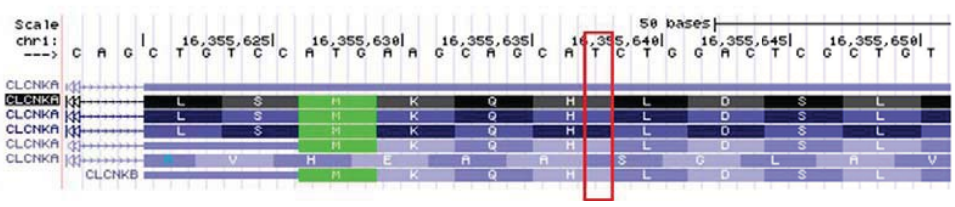
Sora Kim *et al, Annals of Oncology, 2018*

SBi 한국생명정보학회 Korean Society for Bioinformatics

Regarding Transcript-specific peptides



dbSNP ID	Transcript ID (by knwongenes)	Ref	Alt	WT	MT
	hg19_knownGene_uc001axv.3	T	A	ASRLSMKQ**H**LDSLFDNH	ASRLSMKQ**Q**LDSLFDNH
rs79751787	hg19_knownGene_uc010obx.1	T	A	MKQ**H**LDSLFDNH	MKQ**Q**LDSLFDNH
	hg19_knownGene_uc010oby.1	T	A	FSAVHEAA**S**GLAVRQPL	FSAVHEAA**T**GLAVRQPL



Neopepsee determines the sequence of neopeptides regarding the most expressed transcript isoform.

SBi 한국생명정보학회 Korean Society for Bioinformatics

Considering multiple features at once

1. **MHC binding and presentation**
    1. predicted IC_{50} value
    2. percentile rank
    3. protein cleavage
    4. TAP (transporter associated with antigen processing) efficiency
    5. T-cell recognition

2. **Amino-acid characteristics**
    1. amino acid hydrophobicity
    2. amino acid polarity and charge

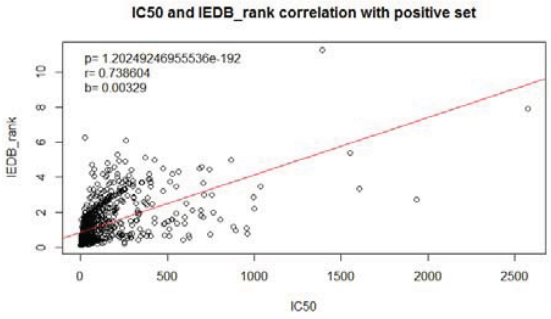
3. **Auxiliary features**
    1. DAI: differential agretopicity
    2. AAPP: amino acid pairwise contact potential

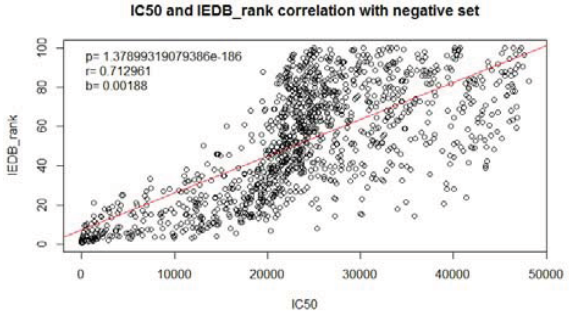
4. **Sequence similarity to known epitopes**

SBi 한국생명정보학회 Korean Society for Bioinformatics

On rare HLA alleles

MHC binding affinity in IC50 score is not available for rare HLA alleles.





***Percentile rank*:**
***rank* of the predicted affinity of the given peptide sequence among ~400,000 random natural peptides**

SBi 한국생명정보학회 Korean Society for Bioinformatics

Automatic calculation of multiple features

Immunogenetics (2010) 62:357–368
DOI 10.1007/s00251-010-0441-4

ORIGINAL PAPER

NetCTLpan: pan-specific MHC class I pathway epitope predictions

Thomas Stranzl · Mette Voldby Larsen · Claus Lundegaard · Morten Nielsen

Received: 31 October 2009 / Accepted: 16 March 2010 / Published online: 9 April 2010

- **MHC score**
- **TAP score**
- **Cleavage score**
- **Combined score**

DESCRIPTION

The prediction output consists of 11 columns.

- Prediction number
- Protein identifier
- HLA Allele
- Peptide sequence
- MHC Prediction score (in 1-log50K(aff) uniqs)
- TAP Prediction score
- Cleavage Prediction score
- Combined Prediction score
- %Random - %Rank of prediction score to a set of 1000.000 random natural 9mer peptides
- Epitope assignment

EXAMPLE OUTPUT

```
# NetCTLpan version 1.1

# Peptide length 9
# NetCTLpan predictions for HLA-A01:01 allele.

#  N   Sequence Name     Allele      Peptide     MHC     TAP      Cle     Comb  %Rank
   0 143B_BOVIN_P29    HLA-A01:01   TMDKSELVQ  0.10500 -0.18300  0.16188  0.13685  50.00
   1 143B_BOVIN_P29    HLA-A01:01   MDKSELVQK  0.02300  0.21200  0.53837  0.14943  50.00
   2 143B_BOVIN_P29    HLA-A01:01   DKSELVQKA  0.01200 -0.77000  0.78670  0.16976  50.00
   3 143B_BOVIN_P29    HLA-A01:01   KSELVQKAK  0.07600  0.32900  0.45985  0.18769  32.00
   4 143B_BOVIN_P29    HLA-A01:01   SELVQKAKL  0.01400  0.99100  0.91927  0.24561  32.00
...
 235 143B_BOVIN_P29    HLA-A01:01   DEGDAGEGE  0.00300 -2.21000  0.04146 -0.04292  50.00
 236 143B_BOVIN_P29    HLA-A01:01   EGDAGEGEN  0.02000 -2.10100  0.05666 -0.01978  50.00
-----------------------------------------------------------------------------

Number of MHC ligands 4 identified. Number of peptides 237. Allele HLA-A0101. Protein name 143B_BOVIN_P29
```

SBi 한국생명정보학회 Korean Society for Bioinformatics

pMHC-I presentation for recognition by TCR

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

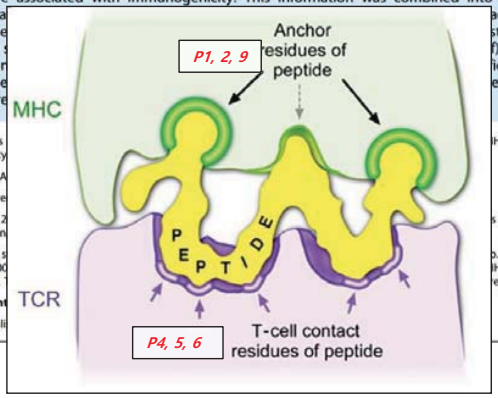
Properties of MHC Class I Presented Peptides Enhance Immunogenicity

Jorg J. A. Calis[1]*, Matt Maybeno[2], Jason A. Greenbaum[2], Daniela Weiskopf[2], Aruna D
Alessandro Sette[2], Can Keşmir[1], Bjoern Peters[2]

1 Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, The Netherlands, 2 Division of Vaccine Discovery, La Jolla Institute for Alle
California, United States of America, 3 Genetech Research Institute, Colombo, Sri Lanka

Abstract

T-cells have to recognize peptides presented on MHC molecules to be activated and elicit their effector
studies demonstrate that some peptides are more immunogenic than others and therefore more likely to
We set out to determine which properties cause such differences in immunogenicity. To this end, we colle
a large set of data describing the immunogenicity of peptides presented on various MHC-I molecules. Tw
could be drawn from this analysis: First, in line with previous observations, we showed that positions P
peptide are more important for immunogenicity. Second, some amino acids, especially those with large
chains, are associated with immunogenicity. This information was combined into a simple model
demonstra
immunoge
we could
elucidation
immunoge
immune re



	IEDB	Vaccinia	Arena	Coxiella
T-cell recognized				
HLA restricted	1492	63	116	-
H-2 restricted	537	-	-	11
Unrecognized				
HLA restricted	99	33	159	-
H-2 restricted	53	-	-	16

Combine all data sets
Select 9mers
Exclude redundant pMHCs

	Mouse	Human
T-cell recognized		
HLA restricted	308	602
H-2 restricted	292	-
Unrecognized		
HLA restricted	135	1
H-2 restricted	46	-

Figure 1. Data acquisition and handling oversight. Data was collected from four different sources (see Methods). The first panel shows how many pMHCs were derived from each data set and their respective MHC restrictions and immunogenicity status. Data from all sets was combined, the number of non-redundant 9mers with respect to the host in which the data was obtained is shown in the second panel.
doi:10.1371/journal.pcbi.1003266.g001

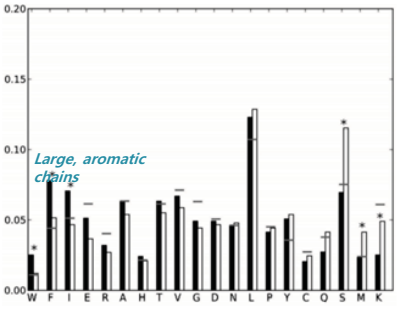


Figure 2. T-cell preferences for different amino acids in HLA class I presented peptides. The fraction of an amino acid in immunogenic (left bar, filled) and non-immunogenic (right bar, unfilled) peptides presented on HLA class I molecules is shown. Significantly different distributions are indicated with a star (Permutation test, see Methods: p<0.05; False discovery rate (FDR) for multiple testing determined as in [46]: q<0.05). The background frequency for each amino acid in the protein sequences that were a source of the immunogenic or non-immunogenic peptides is shown by a grey line.
doi:10.1371/journal.pcbi.1003266.g002

SBi 한국생명정보학회 Korean Society for Bioinformatics

Amino acids features – hydrophobicity

PNAS

TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes

Diego Chowell[a,b,1], Sri Krishna[b,c,1], Pablo D. Becker[d], Clément Cocita[d], Jack Shu[e], Xuefang Tan[e], Philip D. Greenberg[e], Linda S. Klavinskis[d,2], Joseph N. Blattman[f,2], and Karen S. Anderson[b,2]

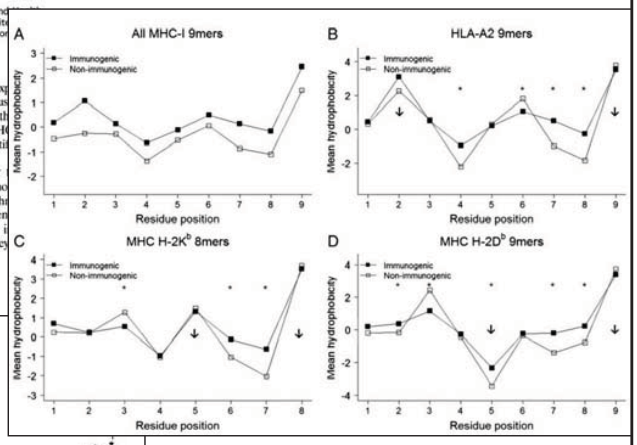
[a]Simon A. Levin Mathematical, Computational, and Modeling Sciences Center, [b]Center for Personalized Diagnostics, and [c]School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287; [d]Department of Immunobiology, King's College London, London SE1 9RT, United Kingdom; [e]Department of Immunology, University of Washington, Seattle, WA 98195; and [f]Center for Infectious Diseases and Vaccinology, Arizona State University, Tempe, AZ 85287

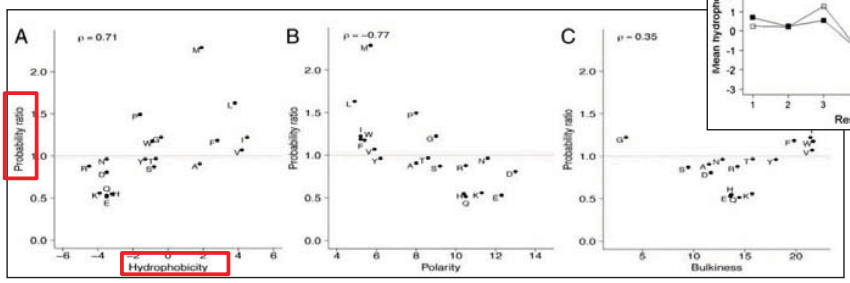
Edited by Ira Mellman, Genentech, Inc., South San Francisco, CA, and approved March 2, 2015 (received for review January 21, 2015)

Despite the availability of major histocompatibility complex (MHC)-binding peptide prediction algorithms, the development of T-cell vaccines against pathogen and tumor antigens remains challenged by inefficient identification of immunogenic epitopes. CD8+ T cells must distinguish immunogenic epitopes from nonimmunogenic self peptides to respond effectively against an antigen without endangering the viability of the host. Because this discrimination is fundamental to our understanding of immune recognition and critical for rational vaccine design, we interrogated the biochemical properties of 9,888 MHC class I peptides. We identified a strong bias toward hydrophobic amino acids at T-cell receptor contact residues within immunogenic epitopes of MHC allomorphs, which permitted

confirmation of MHC-bound peptides, and scarcity of ex
tally confirmed immunogenic epitopes within the infectious
proteome (4). As a result, T-cell epitope prediction algorith
focused on amino acid binding affinity for specific MH
and the protein's proteasomal cleavage pattern to identi
date T-cell epitopes (11–14).

Although computational tools have improved over
decade, they have not been trained to predict immuno
The major limitation when using such prediction algorith
presence of a significant number of binders from a given
that will never lead to an immune response (15). Thus,
genic CTL epitopes fulfill additional criteria that go be





SBi 한국생명정보학회 Korean Society for Bioinformatics

Amino acid features – polarity and charged values

Journal of Computer-Aided Molecular Design, **15:** 573–586, 2001.
KLUWER/ESCOM

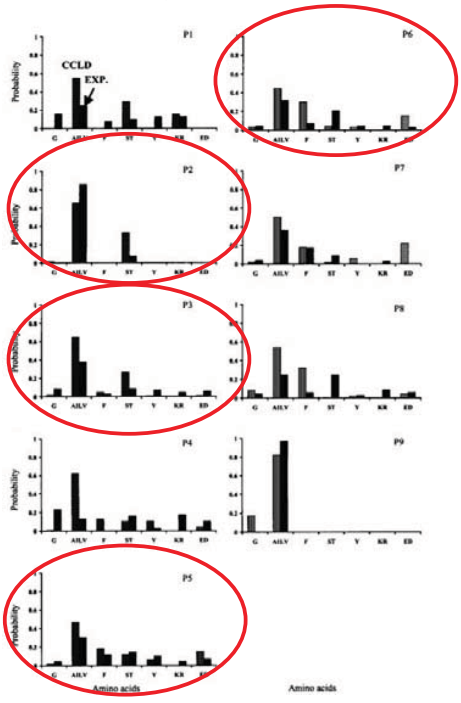
573

Predicting sequences and structures of MHC-binding peptides: a computational combinatorial approach

Jun Zeng[a,d,*], Herbert R. Treutlein[a,b,d] & George B. Rudy[c,e]
[a]*Molecular Modelling Laboratory, Ludwig Institute for Cancer Research, P.O. Box 2008, Royal Melbourne Hospital, Parkville, VIC 3050, Australia;* [b]*Cooperative Research Centre for Cellular Growth Factors, P.O. Royal Melbourne Hospital, Parkville, VIC 3050, Australia;* [c]*Genetics and Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Parkville, VIC 3050, Australia;* [d]*Current Address: Cytopia Pty Ltd, 7th Floor, Daly Wing, St Vincent's Hospital, 41 Victoria Parade, Fitzroy, VIC 3065, Australia;* [e]*Current address: GeneType Pty Ltd, P.O. Box 115, Fitzroy, VIC 3065, Australia*

Received 20 September 2000; accepted 18 April 2001

Key words: computational combinatorial chemistry, docking, major histocompatibility complex, MCSS, peptide design



SBi 한국생명정보학회 Korean Society for Bioinformatics

Entropy and molecular weight

Research article

Vertical T cell immunodominance and epitope entropy determine HI

Michael K.P. Liu,[1] Natalie Hawkins,[2] Adam J. Ritchie,[1] V
Simon Brackenridge,[1] Hui Li,[4] Jeffrey W. Pavlicek,[5] Fangp
Florette Treurnicht,[6] Peter Hraber,[7] Catherine Riou,[8] Clive
Li-Hua Ping,[9,10] Jeffrey A. Anderson,[9,10,11] Ronald Swanstro
Salim S. Abdool Karim,[14] Barton Haynes,[5] Persephon
George M. Shaw,[4] Beatrice H. Hahn,[4] Carolyn Wi
Feng Gao,[5] Steve Self,[2] Andrew McMichael,

[1]Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom. (SCHARP), University of Washington, Seattle, Washington, USA. [3]Department of Microbiol [4]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. North Carolina, USA. [9]Division of Medical Virology, Faculty of Health Sciences, University of Los Alamos National Laboratory, Los Alamos, New Mexico, USA. [13]Institute of Infectious Dis University of Cape Town, Cape Town, South Africa. [14]UNC Center for AIDS Resear [11]Division of Infectious Diseases, School of Medicine, [12]Department of Biochemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [14]Centre f University of KwaZulu-Natal, Durban, South Africa. [15]Nuffield Department of Clini University of Oxford, Oxford, United Kingdom. [16]Santa Fe Institut

Proc. Natl. Acad. Sci. USA
Vol. 73, No. 10, pp. 3671–3675, October 1976
Immunology

Molecular determinants of immunogenicity: The immunon model of immune response

(polyacrylamide/dinitrophenyl-/receptor linkage)

H. M. DINTZIS*, R. Z. DINTZIS*†, AND B. VOGELSTEIN*

* Department of Biophysics and † Department of Anatomy, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205

Communicated by Herman N. Eisen, July 9, 1976

ABSTRACT The immunological response *in vivo* to a series of size-fractionated linear polymers of acrylamide substituted with hapten has been measured in mice. A sharp threshold was observed in immunogenic response elicited by various polymer preparations. All polymers with less than 12 to 16 appropriately spaced hapten groups per molecule were nonimmunogenic, while those polymers with greater than this number were fully immunogenic. The results lead to the conclusion that the immunological response at its most elementary level is quantized, i.e., a minimum specific number of antigen receptors (approximately 12 to 16) must be connected together as a spatially

a systematic variation of molecular parameters might give clues as to the mechanisms involved.

In this study, we sought to vary several molecular properties of our ideal antigen, looking for those responsible for the triggering of a bone-marrow-derived lymphocyte (B-cell) to differentiate and to produce specific antibody in the primary immune response. We desired as our ideal antigen a molecule with the following properties: (*i*) It should consist of a nonimmunogenic carrier or "backbone" structure made of repeating subunits and with hapten groups projecting from it. (*ii*) The molecular weight, and therefore the length, of the carrier should be manipulable. (*iii*) It should be nondegradable by the host organism. (*iv*) The molecule should be linear, flexible, uncharged, and hydrophilic so that it might interact freely with cell surface receptors in whatever geometrical arrangement

Table 1. Characteristics of preparations of fractionated Dnp-polyacrylamide

Preparation	A	B	C	D	E	F
Immunogenic ?	No	No	Yes	Yes	No	Yes
Molecular weight, × 10^{-3}	0.5	0.8	1.4	1.8	1.3	3.3
Acrylamide monomer subunits/molecule	670	1050	1850	2350	1830	4650
Extended length of polymer chain, Å	1700	2600	4600	6000	4600	11,600
Acrylamide monomer subunits/Dnp	48	42	38	36	230	270
Average distance between Dnp groups, Å	120	105	95	90	575	675
Total Dnp groups/molecule	14	25	48	66	8	17
"Effective" Dnp groups/molecule	5–7	8–12	16–24	22–33	7–8	16–17

SBi 한국생명정보학회 Korean Society for Bioinformatics

Protein sequence similarity

Known Epitope sequence
approximately 400,000

Vaccinia virus ERQDYR

patient origin sequence EKQDYR

```
7C12   AVQLVESGGGSVQAGGSLRLTCAASGRTSRSYGMGWFRQAPGKEREFVS
       .|:|.|||||.||||.|||::||||||....|.|||||||||||||||:
EG2    QVKLEESGGGLVQAGDSLRVSCAAGRDFSDYVMGWFRQAPGKEREFVA
```

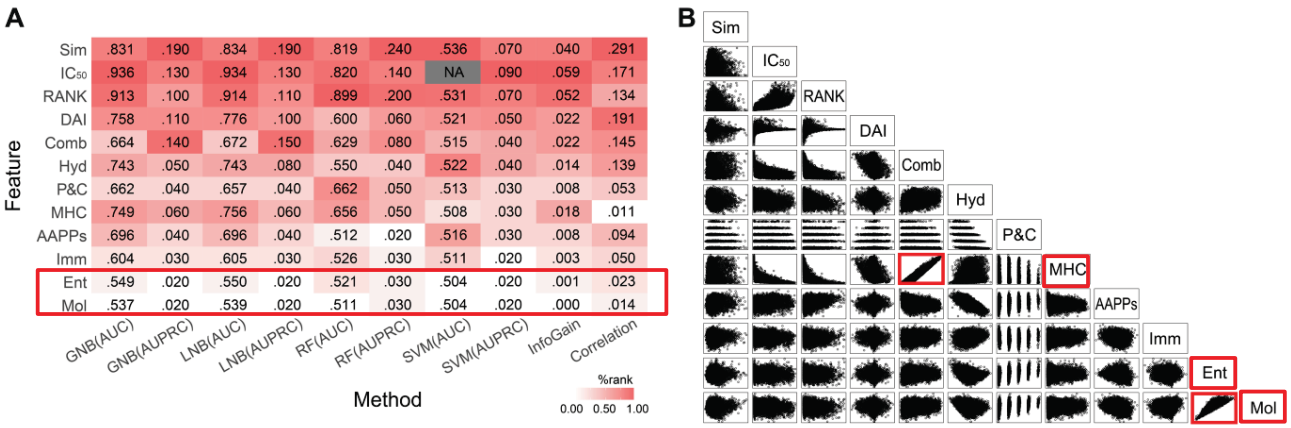
Protein sequence local alignment

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X
A	8	-3	-4	-5	-2	-2	-3	-1	-4	-4	-4	-2	-3	-5	-2	1	-1	-6	-5	-2	-4	-2	-2
R	-3	10	-2	-5	-8	0	-2	-6	-1	-7	-6	3	-4	-6	-5	-3	-3	-7	-5	-6	-4	-1	-3
N	-4	-2	11	1	-5	-1	-2	-2	0	-7	-7	-1	-5	-7	-5	0	-1	-8	-5	-7	5	-2	-3
D	-5	-5	1	10	-8	-2	2	-4	-3	-8	-8	-3	-8	-8	-5	-2	-4	-10	-7	-8	6	0	-4
C	-2	-8	-5	-8	14	-7	-9	-7	-8	-3	-5	-8	-4	-4	-8	-3	-3	-7	-6	-3	-7	-8	-5
Q	-2	0	-1	-2	-7	11	2	-5	1	-6	-5	2	-2	-6	-4	-2	-3	-5	-4	-5	-2	5	-2
E	-3	-2	-2	2	-9	2	10	-6	-2	-7	-7	0	-5	-8	-4	-2	-3	-8	-7	-5	0	7	-3
G	-1	-6	-2	-4	-7	-5	-6	9	-6	-9	-8	-5	-7	-8	-6	-2	-5	-7	-8	-8	-3	-5	-4
H	-4	-1	0	-3	-8	1	-2	-6	13	-7	-6	-3	-5	-4	-5	-3	-4	-5	1	-7	-2	-1	-4
I	-4	-7	-7	-8	-3	-6	-7	-9	-7	8	2	-6	1	-2	-7	-5	-3	-6	-4	4	-8	-7	-3
L	-4	-6	-7	-8	-5	-5	-7	-8	-6	2	8	-6	3	0	-7	-6	-4	-5	-4	0	-8	-6	-3
K	-2	3	-1	-3	-8	2	0	-5	-3	-6	-6	10	-4	-6	-3	-2	-3	-8	-5	-5	-2	0	-3
M	-3	-4	-5	-8	-4	-2	-5	-7	-5	1	3	-4	12	-1	-5	-4	-2	-4	-5	0	-7	-4	-3
F	-5	-6	-7	-8	-4	-6	-8	-8	-4	-2	0	-6	-1	11	-7	-5	-5	0	4	-3	-7	-7	-4
P	-2	-5	-5	-5	-8	-4	-4	-6	-5	-7	-7	-3	-5	-7	12	-3	-4	-8	-7	-6	-5	-4	-4
S	1	-3	0	-2	-3	-2	-2	-2	-3	-5	-6	-2	-4	-5	-3	9	2	-7	-5	-4	-1	-2	-2
T	-1	-3	-1	-4	-3	-3	-3	-5	-4	-3	-4	-3	-2	-5	-4	2	9	-7	-5	-1	-2	-3	-2
W	-6	-7	-8	-10	-7	-5	-8	-7	-5	-6	-5	-8	-4	0	-8	-7	-7	17	2	-5	-9	-7	-6
Y	-5	-5	-5	-7	-6	-4	-7	-8	1	-4	-4	-5	-5	4	-7	-5	-5	2	12	-5	-6	-6	-4
V	-2	-6	-7	-8	-3	-5	-5	-8	-7	4	0	-5	0	-3	-6	-4	-1	-5	-5	8	-7	-5	-3
B	-4	-4	5	6	-7	-2	0	-3	-2	-8	-8	-2	-7	-7	-5	-1	-2	-9	-6	-7	6	0	-4
Z	-2	-1	-2	0	-8	5	7	-5	-1	-7	-6	0	-4	-7	-4	-2	-3	-7	-6	-5	0	6	-2
X	-2	-3	-3	-4	-5	-2	-3	-4	-4	-3	-3	-3	-3	-4	-4	-2	-2	-6	-4	-3	-4	-2	-3

BLOSUM 100 Matrix

SBi 한국생명정보학회 Korean Society for Bioinformatics

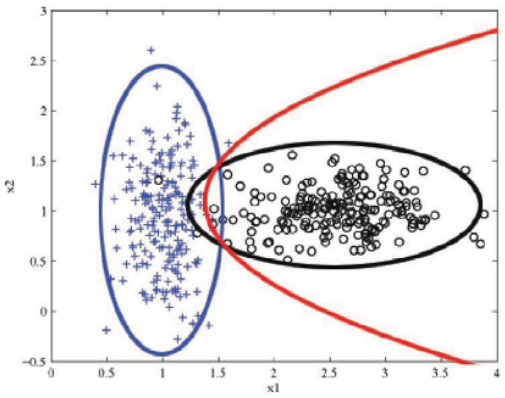
Selecting what to use (feature selection)



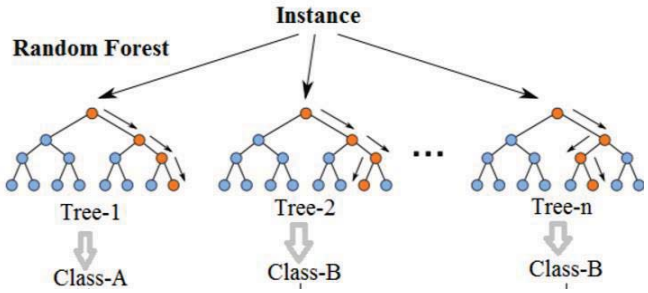
Single feature based classifier

Inter-dependency of features

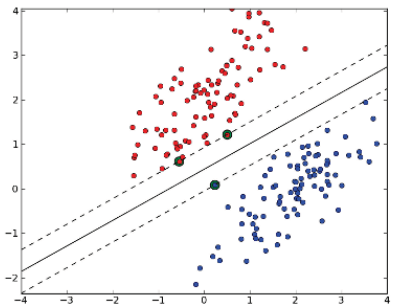
Integration by machine learning



1. Naïve Bayes and
2. locally weighted Naïve Bayes



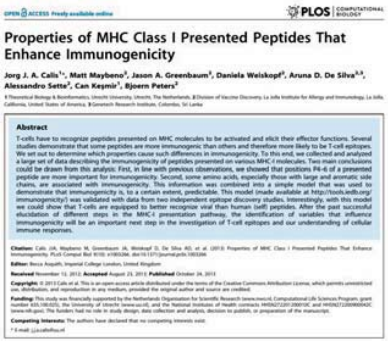
3. Random forest



4. Support vector machine

Training data set

Positive dataset (N=311)



1,113 experimentally validated epitopes

311 wild-type assignable epitopes

Negative dataset (N=14,633)

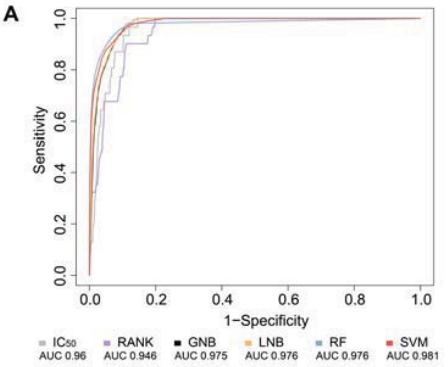
22,245 common (MAF>0.05), non-synonymous SNVs

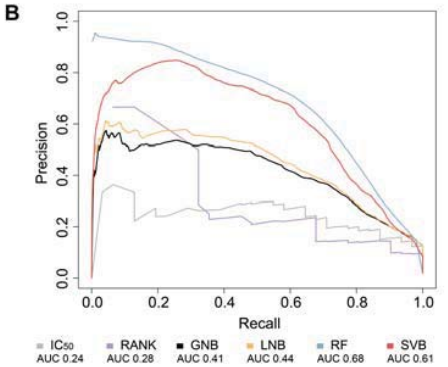
wild-type:
corresponding human reference protein

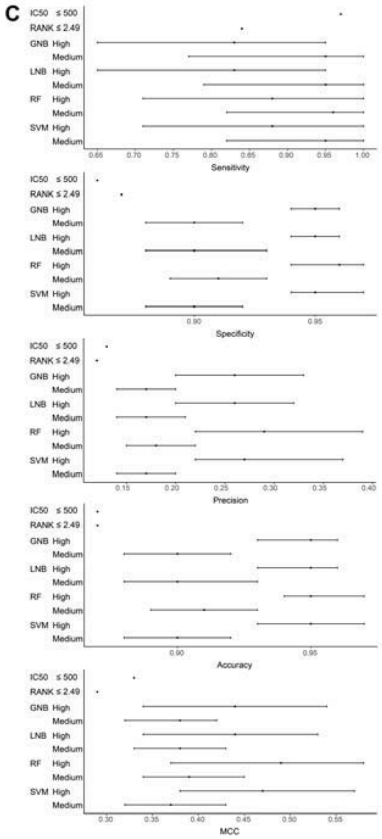
reduced to 14,633

SBi 한국생명정보학회 Korean Society for Bioinformatics

Neopepsee Accuracy





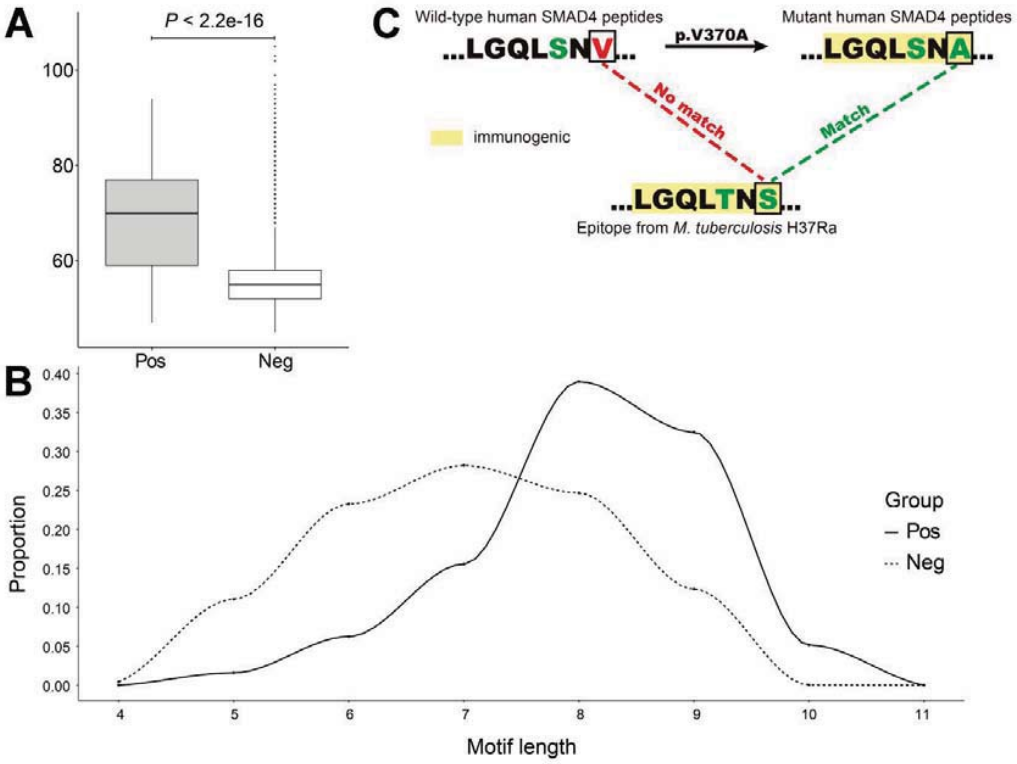


Classification power of Neopepsee vs. single IC50 threshold and single rank threshold

Sora Kim *et al*

SBi 한국생명정보학회 Korean Society for Bioinformatics

Peptide sequence similarity



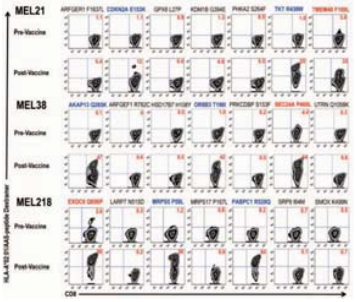
Validation of scores

CANCER IMMUNOTHERAPY

A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells

Beatriz M. Carreno,[1*] Vincent Magrini,[2] Michelle Becker-Hapak,[1] Saghar Kaabinejadian,[3] Jasreet Hundal,[2] Allegra A. Petti,[2] Amy Ly,[2] Wen-Rong Lie,[4] William H. Hildebrand,[3] Elaine R. Mardis,[2] Gerald P. Linette[1]

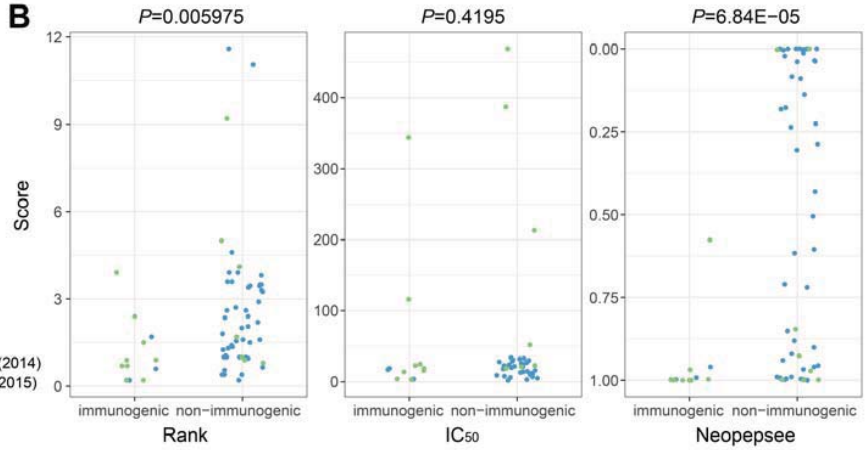
T cell immunity directed against tumor-encoded amino acid substitutions occurs in some melanoma patients. This implicates missense mutations as a source of patient-specific



A

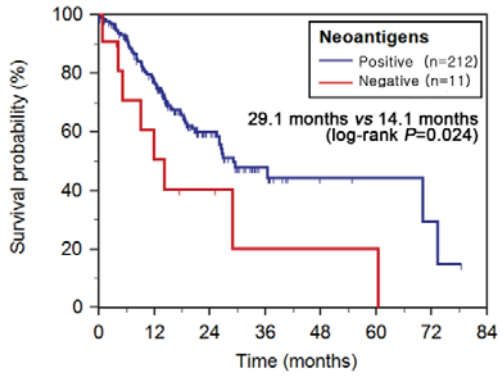
	IC50	RANK	Neopepsee	
	≤500nM	≤2.49	medium	high
# of calls	283	184	259	197
# of hits	12	11	12	10
# of FPs	29	31	26	14
Sensitivity	1.00	0.92	1.00	0.83
Specificitiy	0.45	0.42	0.51	0.74
F-score	0.45	0.41	0.48	0.56

B

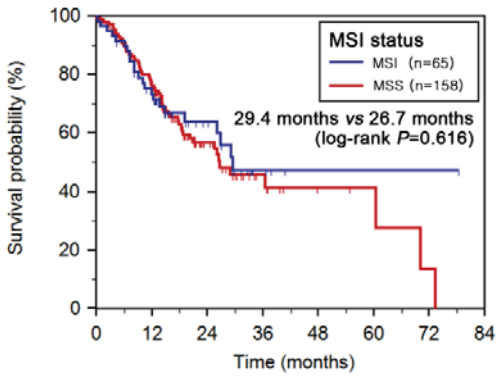


Application to TCGA data

C



D



E

Variable	Category	Univariate analysis			Multivariate analysis		
		HR	95% CI	P	HR	95% CI	P
Neoantigens	negative *v* positive (ref)	3.1	1.18 to 8.47	0.022*	2.2	1.04 to 4.82	0.040*
Stage	III, IV *v* I, II (ref)	2.4	1.14 to 5.08	0.021*	2.0	1.25 to 3.16	0.004*
Sex	female *v* male (ref)	0.9	0.44 to 2.10	0.923	1.1	0.72 to 1.86	0.545
Age	≥65 *v* <65 (ref)	1.1	0.73 to 1.76	0.571	1.0	0.66 to 1.63	0.878
Cytolytic activity (Rooney, et al)	high *v* low (ref)	0.8	0.52 to 1.30	0.398	0.8	0.49 to 1.25	0.306
Microsatellite instability (MSI)	MSI *v* MSS (ref)	0.9	0.54 to 1.43	0.617	1.0	0.61 to 1.65	0.989

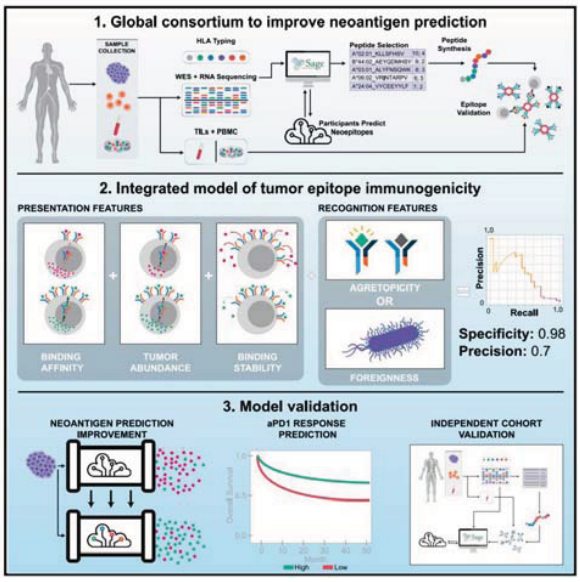
SBi 한국생명정보학회 Korean Society for Bioinformatics

Community-based Guideline for Neoantigen Prediction

Cell

Resource

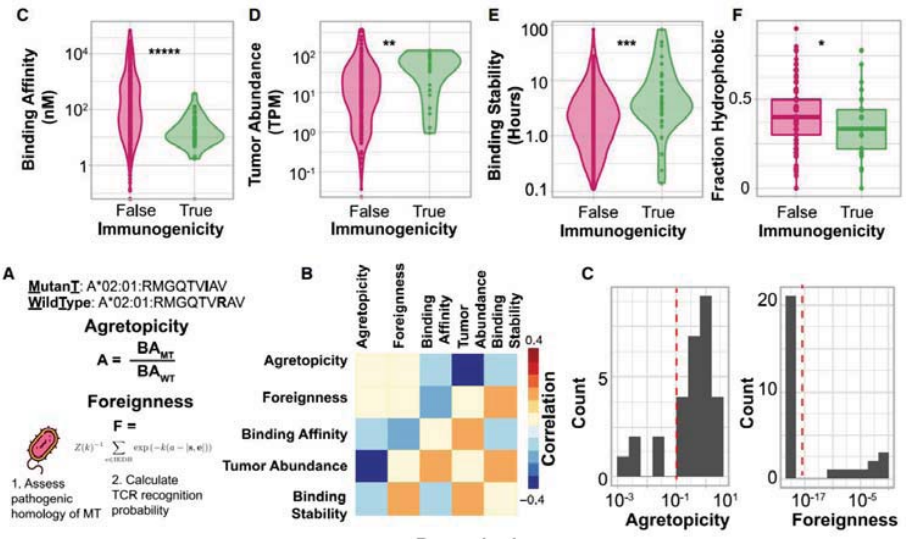
Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction



- 6 subjects
  (3 with metastatic melanoma, 3 with NSCLC)
- 25/28 teams participated
- each team reported 7 to 81,904 candidates
  - median 204
- 608 were selected and validated (multimer-based assay)
- 37/608 (6%) were immunogenic

SBi 한국생명정보학회 Korean Society for Bioinformatics

Community-based Guideline for Neoantigen Prediction



- **Derive informative features**
  - **binding affinity, Tumor abundance, Binding stability, Hydrophobicity**
  - **Agretopicity, Foreignness**

Conclusion

- 다양한 cancer immunotherapy 의 발전으로 자신의 면역 시스템을 이용한 치료가 각광받고 있음
- 더 큰 효과와 적은 부작용을 위하여 환자, 종양 특이적 antigen 발굴이 필요함
- HLA type, MHC binding, Antigen processing 등 다양한 step 단계를 예측할 수 있는 computational algorithm 이 존재하며, 발전하고 있음
- NGS 에 기반하여 면역항암치료의 반응을 예측하고, 환자 특이적 치료를 할 수 있는 분석을 진행할 수 있음

Thank you

Your success is our success. We've prescription for your business.
We are professional communication group.