

# KSBI-BIML 2026

Bioinformatics & Machine Learning(BIML)  
Workshop for Life Scientists

생명정보학 & 머신러닝 워크샵(온라인)



## Introduction to adaptive immune repertoire sequencing

박대찬 \_ 아주대학교



**KSBI**  
KOREAN SOCIETY FOR  
BIOINFORMATICS

한국생명정보학회



본 강의 자료는 한국생명정보학회가 주관하는 BIML 2026 워크샵을 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 행위자 본인에게 있음**을 알립니다.

# KSBI-BIML 2026

## Bioinformatics & Machine Learning (BIML) Workshop for Life Scientists

한국생명정보학회가 주최하는 BIML-2026 동계 Bioinformatics & Machine Learning 교육 워크숍에 여러분을 초대합니다.

BIML 워크숍은 생명정보학 연구자들이 최신 AI바이오 분야의 인공지능 기반 분석 기술과 바이오 데이터 분석 기법을 이론과 실습을 통해 체계적으로 배울 수 있는 전문 교육 프로그램입니다. 2015년에 시작된 BIML 워크숍은 올해로 12년 차를 맞이하며, 국내 생명정보학 분야의 최초이자 최고 수준의 교육 프로그램으로 자리 잡았습니다. 이번 워크숍은 크게 인공지능바이오(AI바이오) 분야와 디지털바이오 분야, 두 분야로 구성됩니다.

AI바이오 분야에서는 생명정보 분석에 폭넓게 응용되고 있는 다양한 인공지능 기반 자료 모델링 기법을 다룰 예정입니다. 특히, 인공지능 심층학습을 활용한 단백질 구조 예측, 유전체 분석, 신약 개발에 대한 이론 및 실습 강의를 진행됩니다.

또한 디지털바이오 분야에서는 단일세포오믹스, 공간오믹스, 멀티오믹스, 메타오믹스에 대한 강의도 마련되어 있어, 연구자들의 분석 역량 강화에 실질적인 도움을 줄 것으로 기대됩니다.

또한 2024년부터 추가된 의료정보 자료 분석을 다루는 강의를 올해도 지속해서 운영하고자 합니다. 이는 최근 의료정보 자료 분석에 관한 연구 수요 증가를 반영한 것으로, 관련 연구를 수행하는 의과학자 및 의료정보 연구자들에게 유용한 지침을 제공할 것입니다.

또한, 올해도 생명정보학 기술의 다양화에 발맞춰 온라인 강좌를 대폭 확대했습니다. 올해는 무료 강좌 10개를 포함한 총 40개 이상의 강좌가 개설되며, 연구 주제에 맞는 강좌 추천과 강연료 할인 혜택도 제공합니다.

BIML-2026는 국내 주요 연구 중심 대학의 전임 교수 및 각 분야 최고 전문가들의 강의로 구성되어 있으며, 기초 이론부터 최신 연구 동향까지 아우르는 심도 있는 교육의 장이 될 것으로 확신합니다.

여러분의 많은 관심과 참여를 기대합니다!

2026년 2월

한국생명정보학회장 류 성 호

# Introduction to adaptive immune repertoire sequencing

암 면역 환경, 감염병 반응 면역 등에 대한 관심이 증가함에 따라 적응면역의 다양성 분석에 대한 기술이 성장하고 있다. 특히, B cell receptor (=antibody, BCR)와 T cell receptor (TCR) 서열 분석은 면역항암제와 감염병 치료제 발굴 및 반응성 이해에 핵심적인 요소이다. BCR과 TCR 서열의 천문학적 다양성과 세포별 서열 특이성으로 인해 일반적인 RNA-seq, scRNA-seq, WES 등은 adaptive immune repertoire를 포괄적으로 이해하는 데 한계가 있다. 따라서, high throughput으로 BCR과 TCR full-length 서열을 얻는 방법, 단일세포 BCR의 heavy/light chain (TCR은 alpha/beta chain)을 동시에 시퀀싱하는 기법들이 개발되고 있다.

본 강의에서는 NGS를 이용한 BCR/TCR 시퀀싱 데이터 이론을 학습하고 일련의 생명정보학적 분석과정을 이해하는 것이 목표이다. 이론으로, BCR 유전자의 구조, BCR-seq library를 만드는 최신 기법을 학습한다. 기본 분석에 사용되는 IMGT와 MIXCR 활용 예시를 통해 BCR의 V gene usage와 complementarity-determining regions (CDR) 서열을 동정하는 법을 배운다. 동정된 항체 서열의 특징을 해석하는 downstream 분석 과정을 학습한다.

강의는 다음의 내용을 포함한다:

- Immune repertoire 개요
- BCR sequencing을 위한 실험 기법
- BCR sequencing 데이터 분석 및 해석

\* 고강의교재: 강의자료

\* 육생준비물: 노트북 (메모리 8GB 이상, 디스크 여유공간 30GB 이상), RStudio 설치

\* 강의 난이도: 초급

\* 강의: 박대찬 교수 (아주대학교 생명과학과)

# Curriculum Vitae

Speaker Name: Daechan Park, Ph.D.



## ► Personal Info

Name Daechan Park  
Title Associate Professor  
Affiliation Ajou University

## ► Contact Information

Address 206 Worldcup-ro, Yeongtong-gu, Suwon 16499, Korea  
Email dpark@ajou.ac.kr

---

## Research Interest

bioinformatics, immune repertoire, epigenomics, synthetic biology

## Educational Experience

2014 Ph.D. in Genomics, The University of Texas at Austin, Austin, TX, USA  
2008 B.S. in Biological Sciences, Seoul National University, Seoul, Korea

## Professional Experience

2018-present Assistant / Associate Professor, Ajou University  
2016-2018 Senior Research Scientist, Korea Institute of Science and Technology  
2014-2016 Postdoctoral Fellow, The University of Texas at Austin, Austin, TX, USA

## Selected Publications (5 maximum)

1. Jae Yun Moon, Jina Seo, Jaewoo Lee#, and **Daechan Park#**, Assessment of attenuation of varicella-oster virus vaccines based on genomic comparison (2023) *Journal of Medical Virology* 95(3), e28590 #co-corresponding (IF=12.7)
2. Do Young Hyeon\*, Dowoon Nam\*, Youngmin Han\*, Duk Ki Kim\*, Gibeom Kim\*, Daeun Kim\*, Jingi Bae Seunghoon Back, Dong-Gi Mun, Inamul Hasan Madar, Hangyeore Lee, Su-Jin Kim, Hokeun Kim, Sangyeop Hyun, Chang Rok Kim, Juhee Jeong, Suwan Jeon, Yeon Woong Choo, Kyung Bun Lee, Wooil Kwon, Seunghyuk Choi, Taewan Goo, Taesung Park, Young Ah Suh, Hongbeom Kim, Ja-Lok Ku, Min-Sik Kim, Eunok Paek, **Daechan Park#**, Keehoon Jung#, Sung Hee Baek#, Jin-Young Jang#, Daehee Hwang# and Sang-Won Lee#, Proteogenomic landscape of human pancreatic ductal adenocarcinoma in an Asian population reveals tumor cell-enriched and immune-rich subtypes (2023) *Nature Cancer* 4(2):290-307. #co-corresponding (IF=22.7)
3. Se Won Park\*, Jaehoon Kim\*, Sungryong Oh\*, Jeongyoon Lee, Joowon Cha, Hyun Sik Lee, Keun Il Kim, **Daechan Park#** and Sung Hee Baek#, PHF20 is crucial for epigenetic control of starvation-induced autophagy through enhancer activation (2022) *Nucleic Acids Research* 50(14):7856-7872. #co-corresponding (IF=14.9)
4. Chuna Kim\*, Sanghyun Sung\*, Jong-Seo Kim\*, Hyunji Lee, Yoonseok Jung, Sanghee Shin, Eunkyeong Kim, Jenny J. Seo, Jun Kim, Daeun Kim, Hiroyuki Niida, Narry V. Kim, **Daechan Park#** and Junho Lee#, Telomeres reformed with non-telomeric sequences in mouse embryonic stem cells (2021) *Nature Communications* 12(1):1097. #co-corresponding (IF=16.6)
5. Sungryong Oh\*, Kyungjin Boo\*, Jaebeom Kim, Seon Ah Baek, Yoon Jeon, Junghyun You, Ho Lee, Hee-Jung Choi, **Daechan Park#**, Ji Min Lee# and Sung Hee Baek#, The chromatin-binding protein PHF6 functions as an E3 ubiquitin ligase of H2BK120 via H2BK12Ac recognition for activation of trophodermal genes (2020) *Nucleic Acids Research* 48(16):9037-9052. #co-corresponding (IF=14.9)

# KSBi-BIML 2024

## Introduction to adaptive immune repertoire sequencing

아주대학교  
박대찬

## 강의 순서

1

Adaptive Immune Receptor Repertoire (AIRR) 의 정의와 다양성

2

AIRR 연구를 위한 오믹스 접근법

3

AIRR 생명정보 분석법

4

AIRR 연구 사례 및 분석 예시

## 레퍼토리 (repertoire)는 무엇인가?



About 2,200,000 results (0.26 seconds)

공연예술, 같은 내용의 이야기 등을 말하는 단어인 레퍼토리를 이야기꾼으로 순화시켜 보았습니다. 하나의 흐름(끈)을 가진 이야기라는 의미에서 이야기+끈 이라고 순화시켰습니다.



 쉬운 우리말을 쓰자  
<https://www.plainkorean.kr> > part > change > title=레퍼...  
레퍼토리 - 쉬운 우리말을 쓰자

우리는 다양한 레퍼토리를 기대한다.

## Different People, Different Repertoire, Different B/T cells



## A pool of Adaptive Immune Receptor is a Repertoire

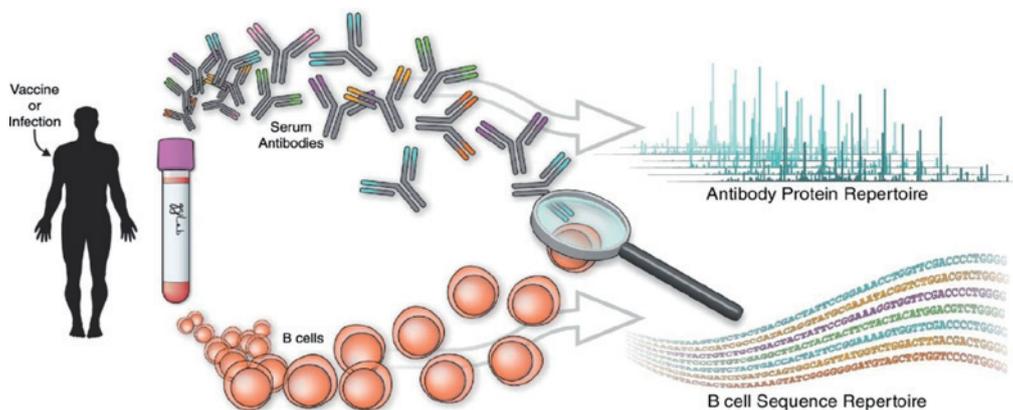
# Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data

Florian Rubelt<sup>1,21</sup>, Christian E Busse<sup>2,21</sup>, Syed Ahmad Chan Bukhari<sup>3,21</sup>, Jean-Philippe Bürckert<sup>4</sup>, Encarnita Mariotti-Ferrandiz<sup>5</sup>, Lindsay G Cowell<sup>6</sup>, Corey T Watson<sup>7</sup>, Nishanth Marthandan<sup>8</sup>, William J Faison<sup>9</sup>, Uri Hershberg<sup>10</sup>, Uri Laserson<sup>11</sup>, Brian D Corrie<sup>12,13</sup>, Mark M Davis<sup>1,14</sup>, Bjoern Peters<sup>15</sup>, Marie-Paule Lefranc<sup>16</sup>, Jamie K Scott<sup>8,12,17</sup>, Felix Breden<sup>12,13</sup>, The AIRR Community<sup>18</sup>, Eline T Luning Prak<sup>19,22</sup> & Steven H Kleinstein<sup>3,20,22</sup>

High-throughput sequencing of B and T cell receptors is routinely being applied in studies of adaptive immunity. The Adaptive Immune Receptor Repertoire (AIRR) Community was formed in 2015 to address issues in AIRR sequencing studies, including the development of reporting standards for the sharing of data sets.

VOLUME 18 NUMBER 12 DECEMBER 2017 NATURE IMMUNOLOGY

## B cell receptor (BCR) and T cell receptor (TCR) are extremely diverse and personalized due to clonal evolution



### By individuals

- MHC
- Genotype
- Age
- Race

### By disease

- Cancer
- Infectious disease
- Autoimmune disease

Lavinder et al., (2015) *Current Opinion in Chemical Biology*

# B cell receptor (BCR) and T cell receptor (TCR) are extremely diverse

## TCR

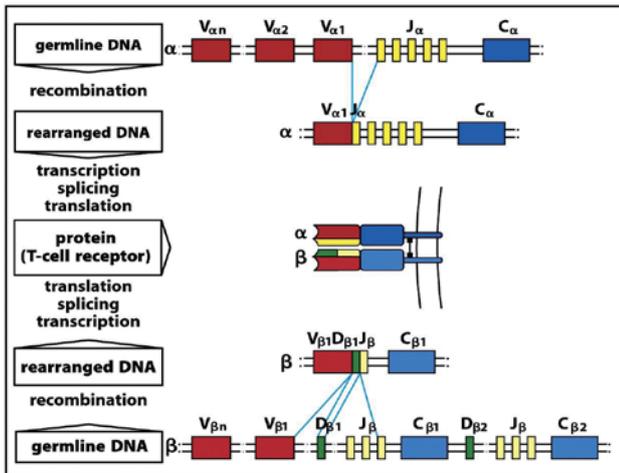


Figure 4-10 Immunobiology, 7ed. (© Garland Science 2008)

~10<sup>18</sup> receptors

## BCR

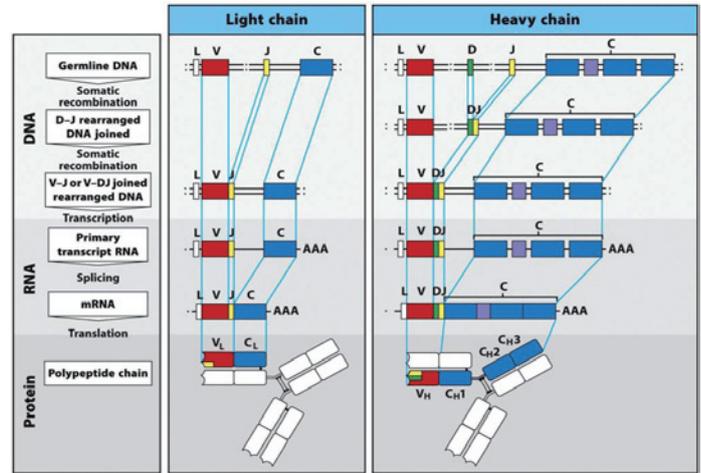
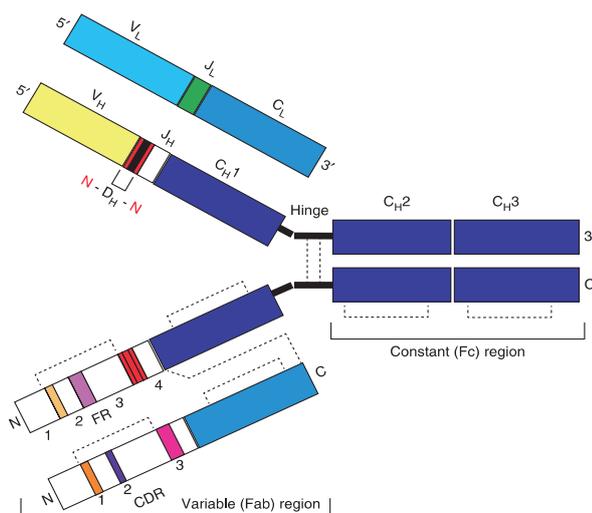


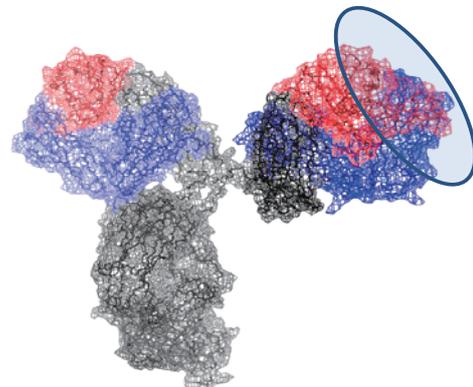
Figure 4-2 Immunobiology, 7ed. (© Garland Science 2008)

~10<sup>11</sup> receptors

# Complementarity-determining regions (CDRs)



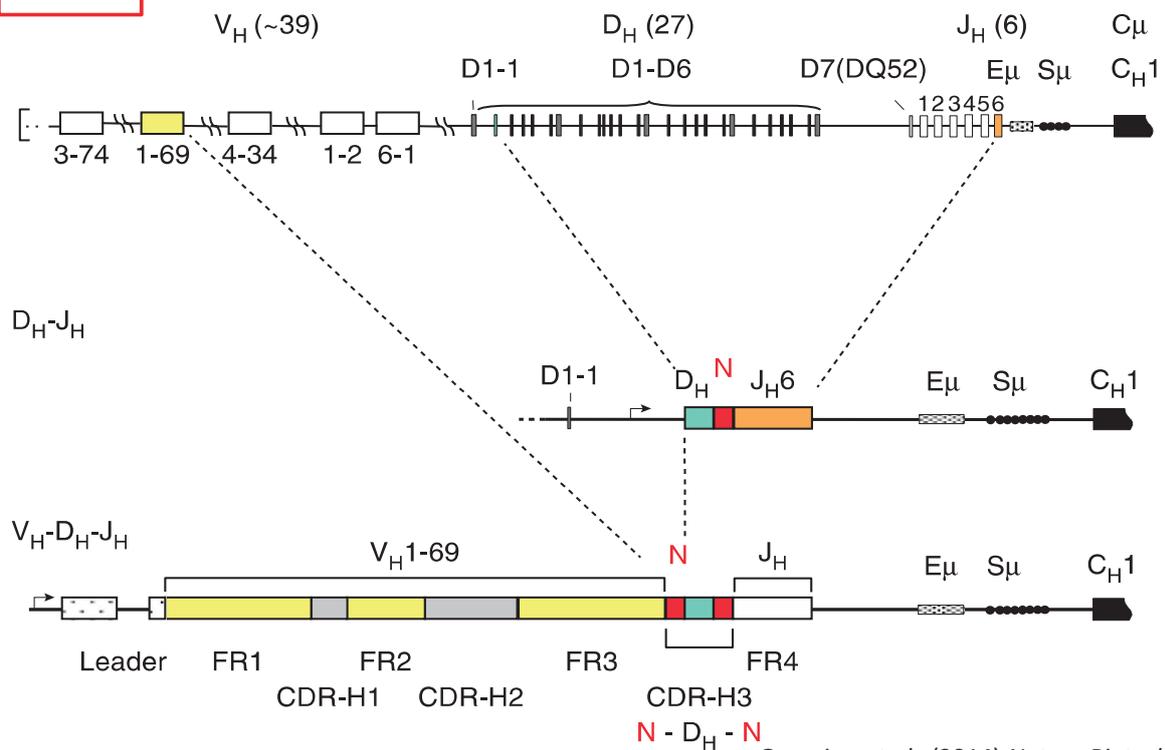
VH-VL domain: Binding to antigen



Georgiou et al., (2014) Nature Biotechnology

# Antibody transcripts are highly diverse

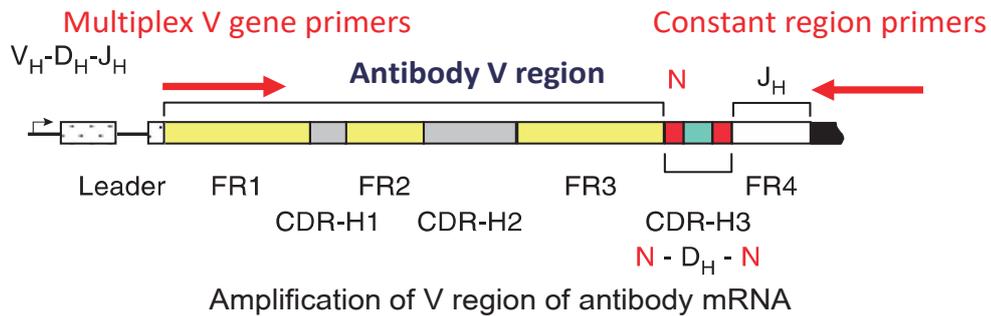
Germline



Georgiou et al., (2014) Nature Biotechnology

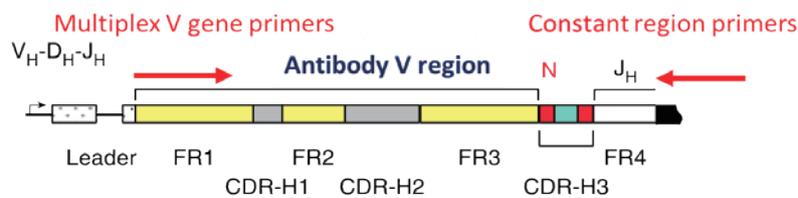
- 1 Adaptive Immune Receptor Repertoire (AIRR)의 정의와 다양성
- 2 AIRR 연구를 위한 오믹스 접근법
- 3 AIRR 생명정보 분석법
- 4 AIRR 연구 사례 및 분석 예시

## B cell Receptor sequencing (BCR-seq)



Next Generation Sequencing  
MiSeq 2X300 Platform

## B cell Receptor sequencing (BCR-seq)



### Important points

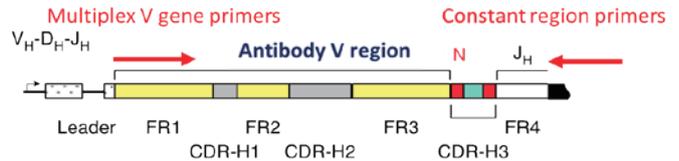
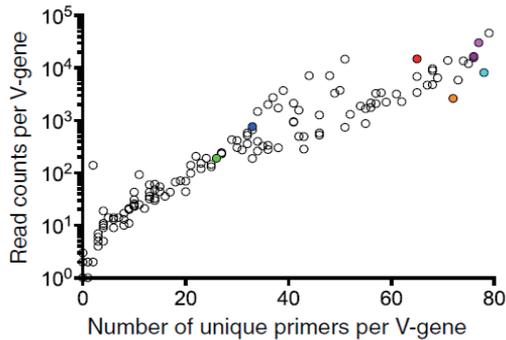
- Clone수를 많이 보는 BCR-seq은 주로 amplicon 으로 시퀀싱 한다.
- PCR을 할 때 forward primer는 주로 multiplexing primer가 사용된다.
- V region을 다 cover하려면 NGS read 길이가 길어야 한다.
- Error rate 낮은 NGS platform을 사용해야 한다 .

# BCR-seq is erroneous

## IMMUNOLOGY

### Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting

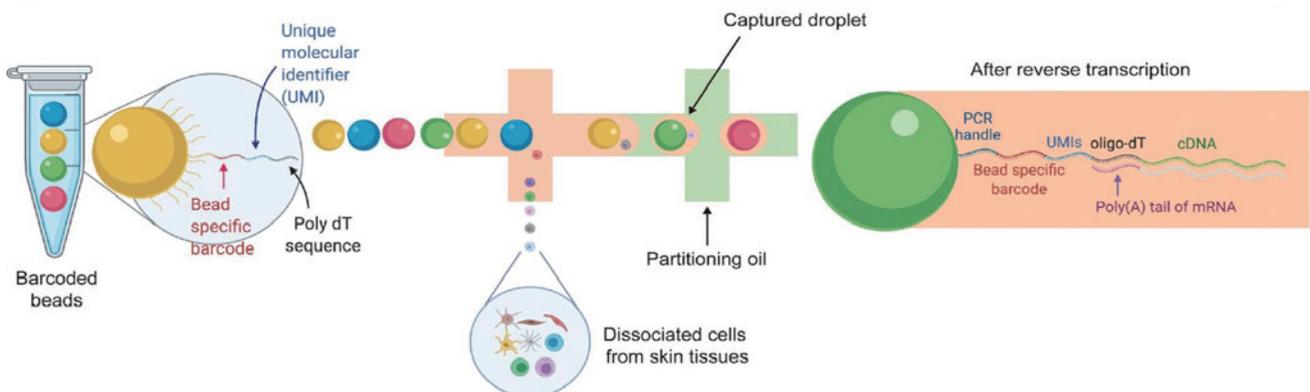
Tarik A. Khan,<sup>1</sup> Simon Friedensohn,<sup>1</sup> Arthur R. Gorter de Vries,<sup>1</sup> Jakub Straszewski,<sup>1,2</sup> Hans-Joachim Ruscheweyh,<sup>1,2,3</sup> Sai T. Reddy<sup>1\*</sup>



Unique Molecular Identifier를 사용하면 문제를 (일부) 해결할 수 있음

Khan et al., (2016) Science Advances

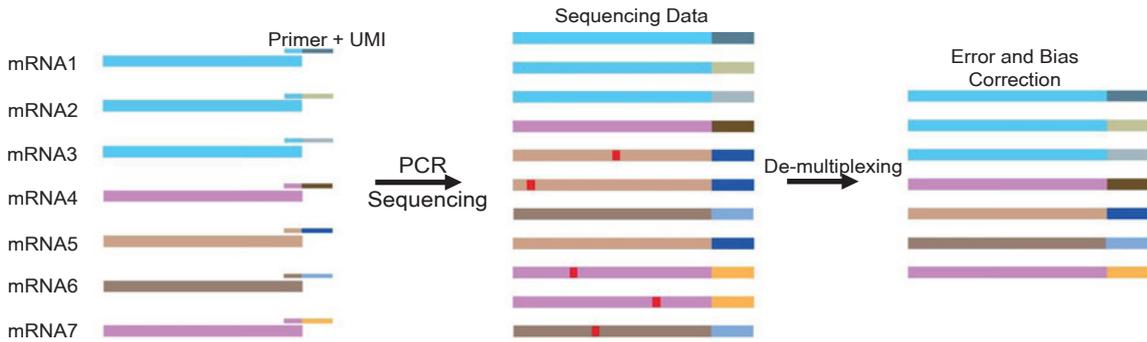
# Single cell RNA-seq의 UMI와 사용 목적이 조금 다름



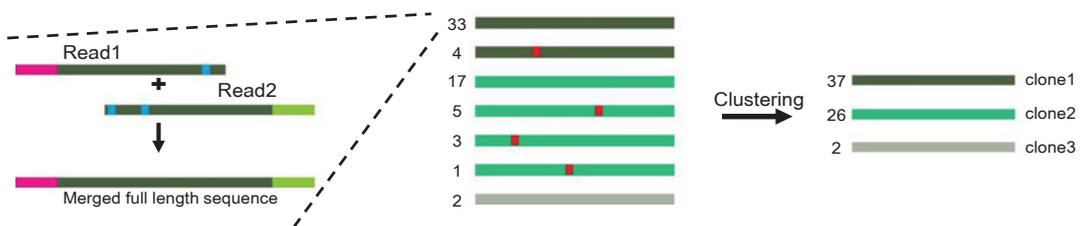
UMI counts represent gene expression levels

# Error correction for detection of true diversity

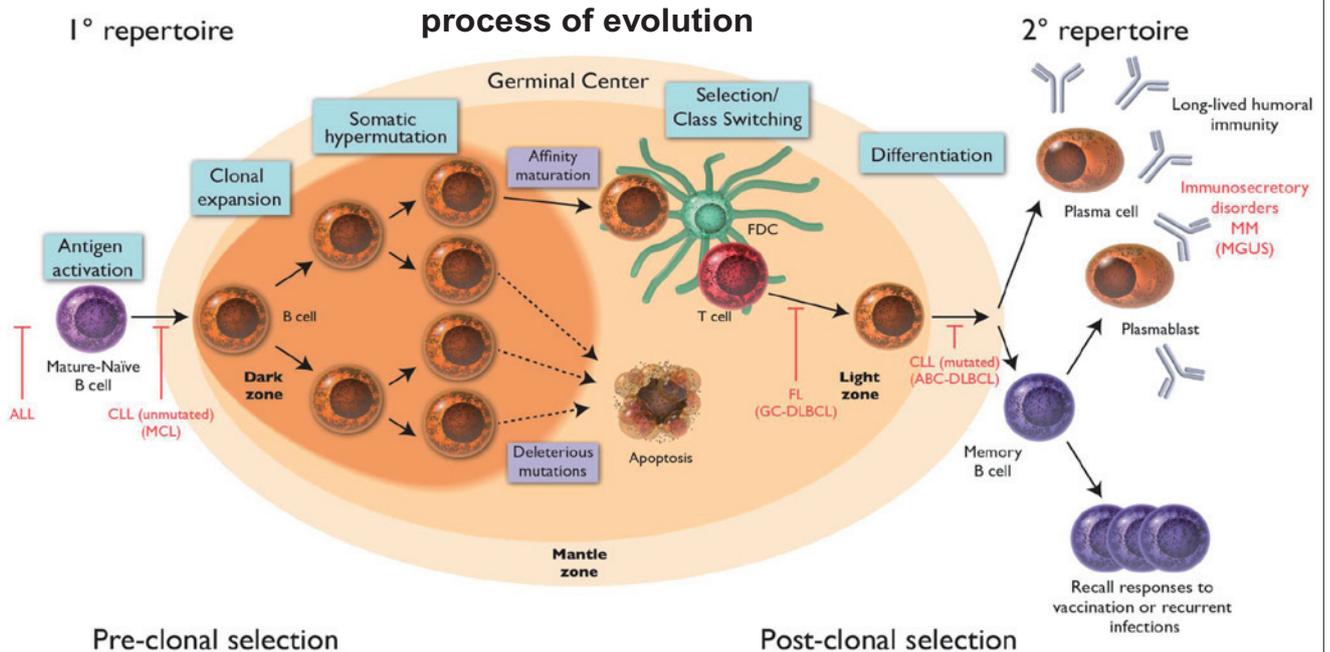
(A) Correction PCR and Sequencing Errors



(B) Clustering after Assembly of Paired-end Reads



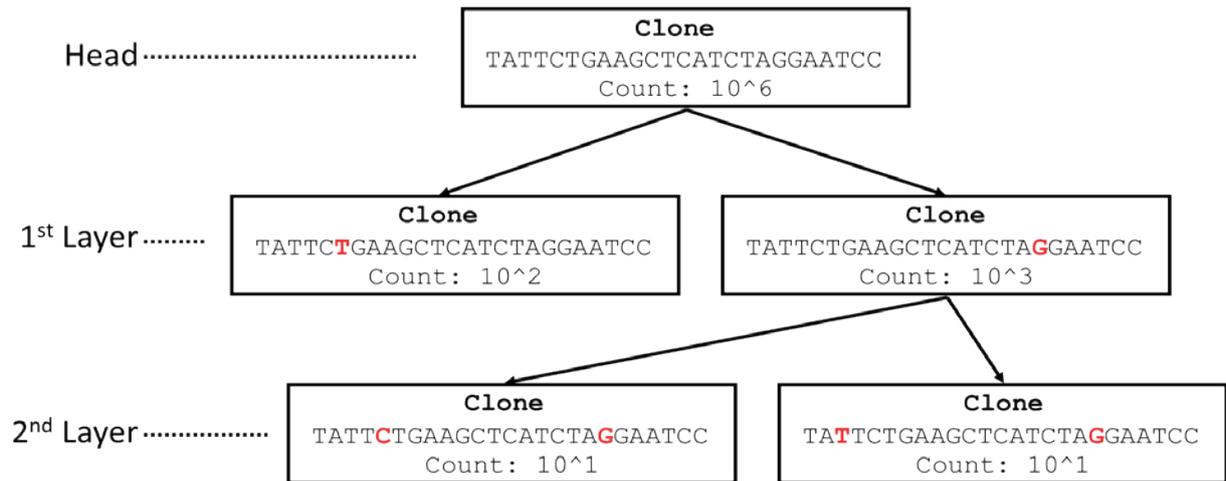
# Germinal Center (GC) Reaction



ALL = Acute Lymphoblastic Leukemia  
 CLL = Chronic Lymphocytic Leukemia  
 FL = Follicular Lymphoma  
 DLBCL = Diffuse Large B Cell Lymphoma  
 MM = Multiple Myeloma

Georgiou et al., *Nature Biotechnology*

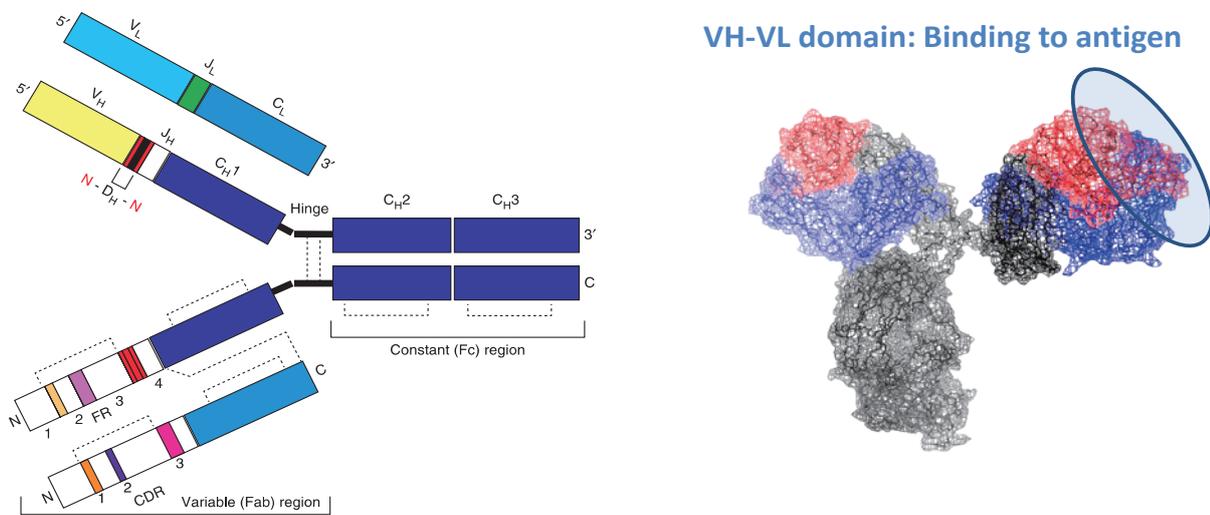
## Neighbor joining algorithm to define clones



## AIRR 연구를 위한 오믹스 접근법

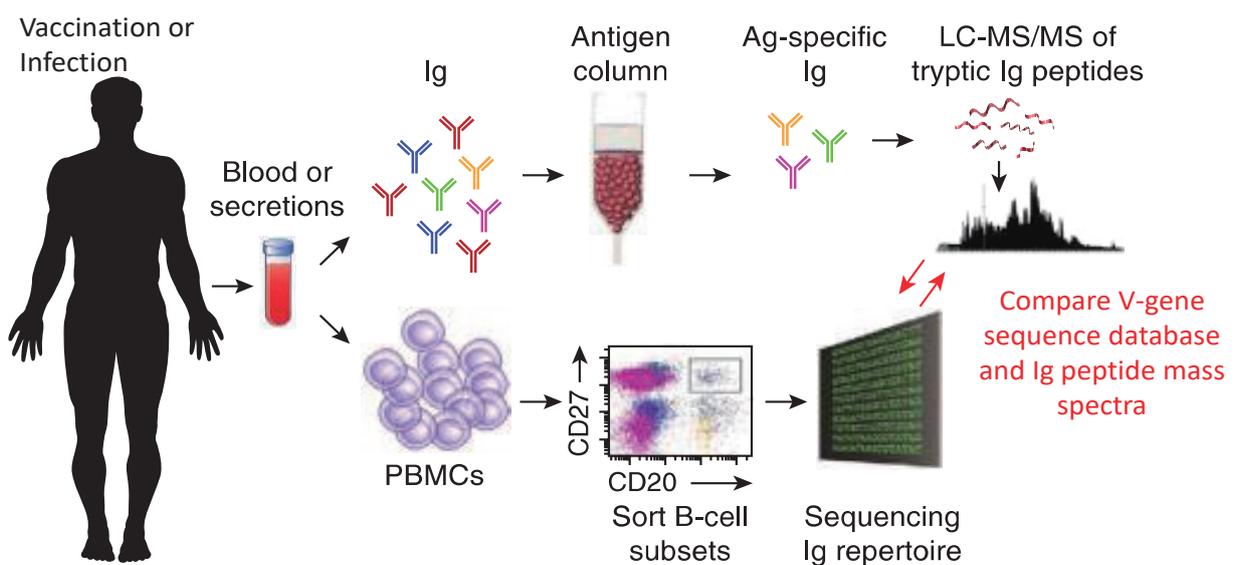
- BCR-seq 데이터 분석에서 UMI의 사용과 DNA 서열 clustering 이 중요도
- proteomics + BCR-seq 통합 분석
- Heavy / light chain 을 동시에 시퀀싱 하기
- Short read로 assembly하기

# Ig-seq is proteomics analysis of immunoglobulin



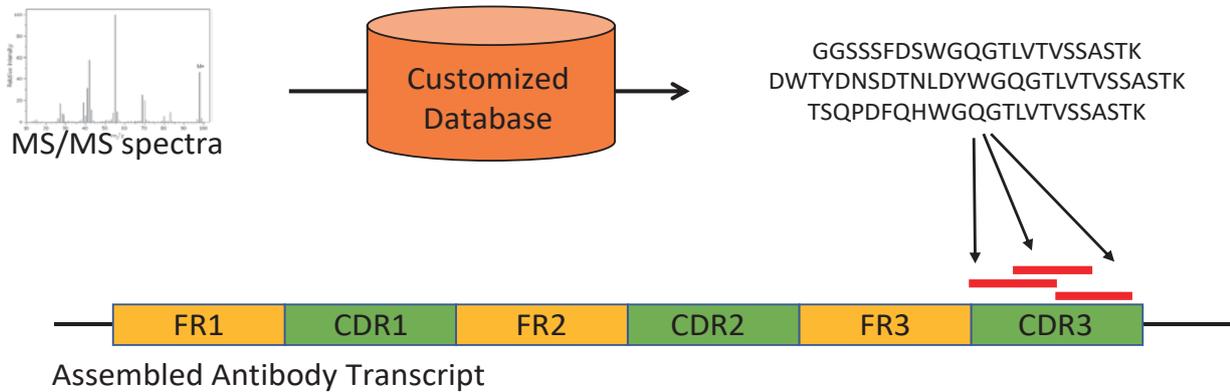
Georgiou et al., (2014) *Nature Biotechnology*

# Application of BCR-seq and Ig-seq



Georgiou et al., (2014) *Nature Biotechnology*

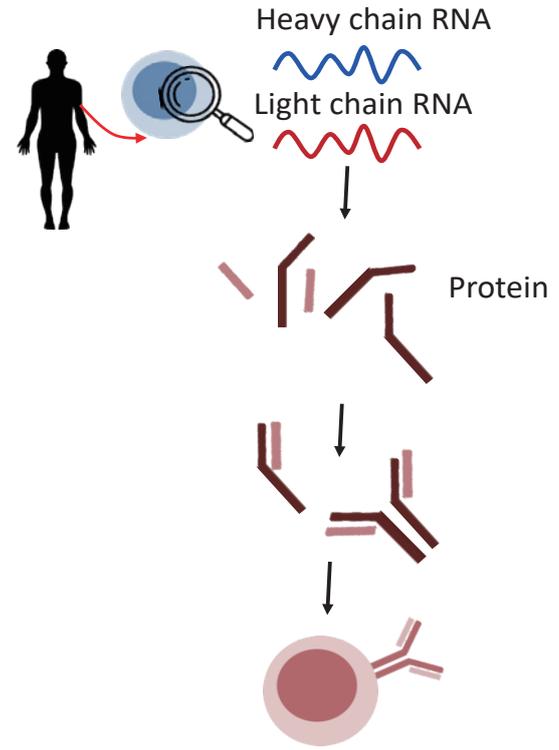
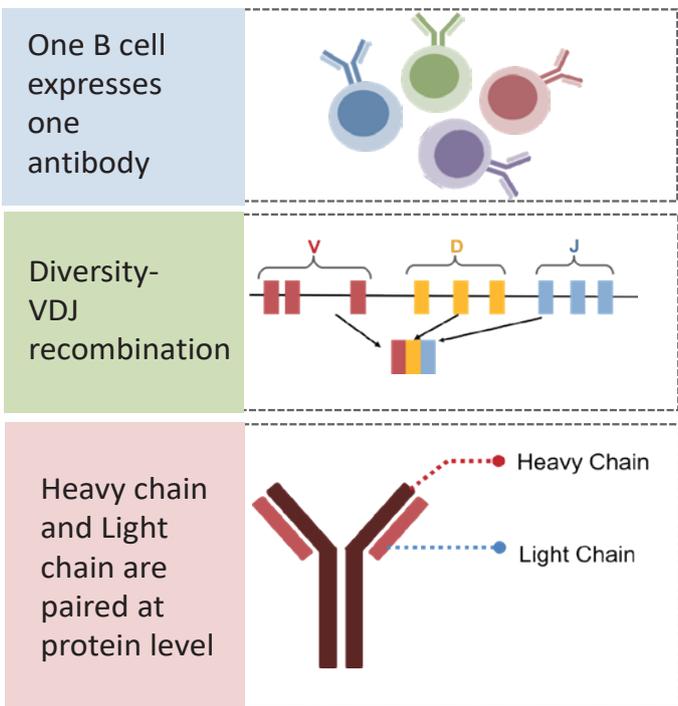
# Serological Antibodies Identification by CDR3 sequences in LC-MS/MS



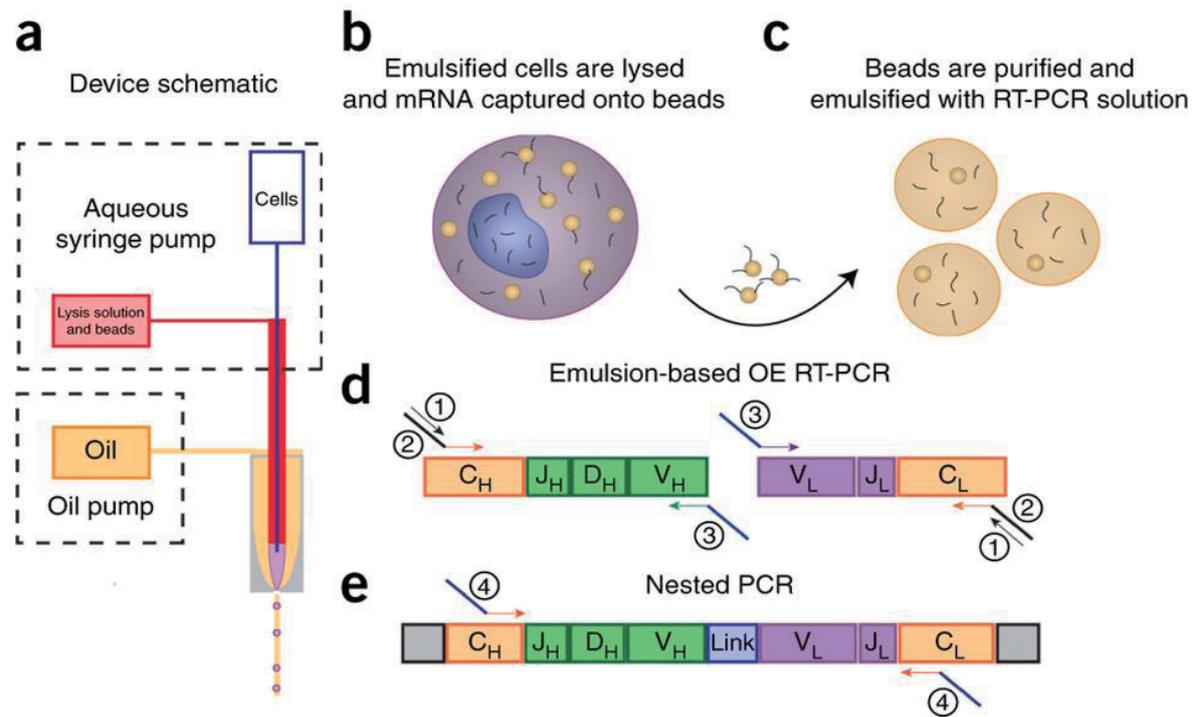
## AIRR 연구를 위한 오믹스 접근법

- BCR-seq 데이터 분석에서 UMI의 사용과 DNA 서열 clustering 이 중요도
- proteomics + BCR-seq 통합 분석
- Heavy / light chain 을 동시에 시퀀싱 하기
- Short read로 assembly하기

# Difficulties with antibody sequencing



# Pairing sequencing technology

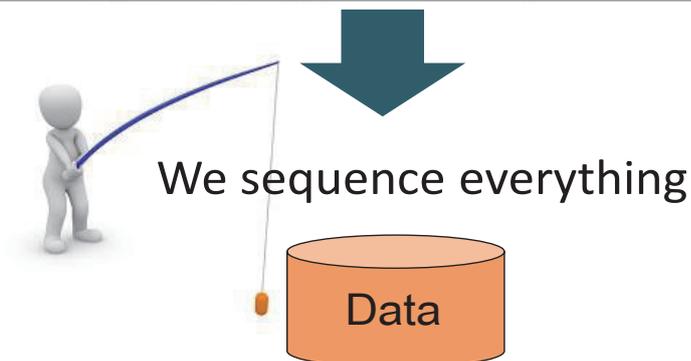


DeKosky et al., (2015) *Nature Medicine*  
 McDaniel et al., (2016) *Nature Protocols*

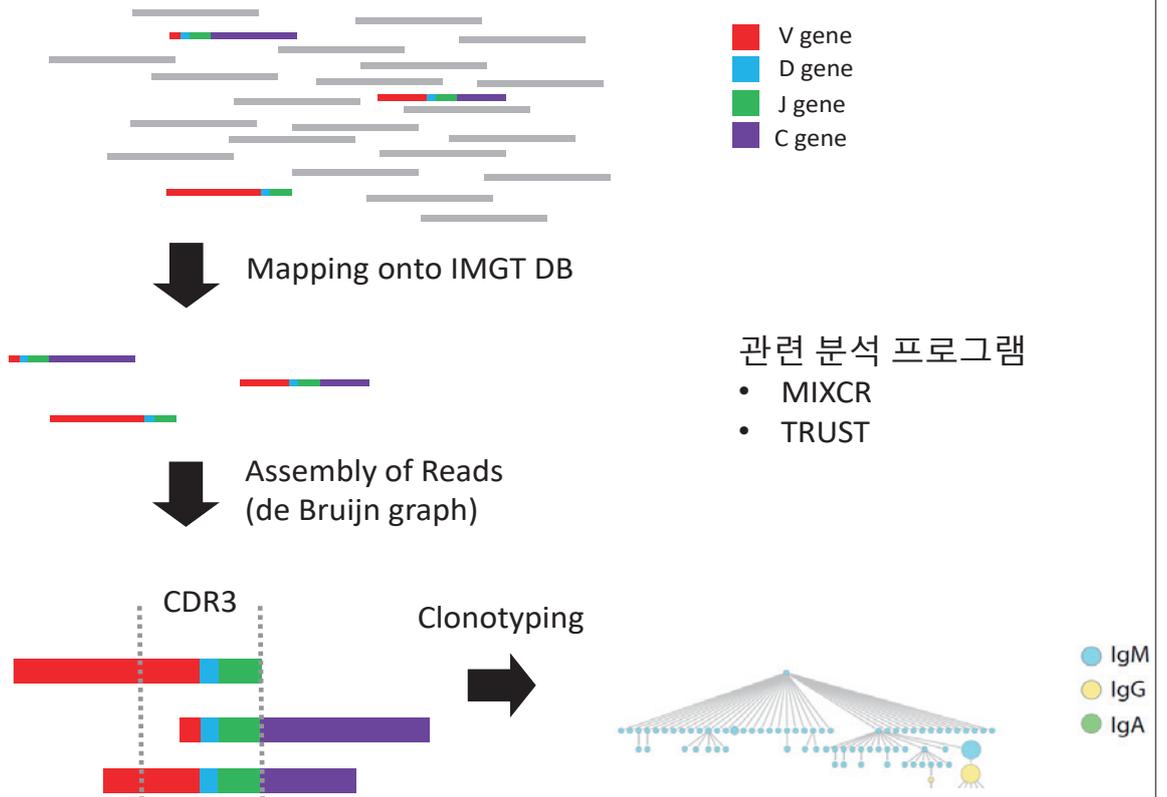
## AIRR 연구를 위한 오믹스 접근법

- BCR-seq 데이터 분석에서 UMI의 사용과 DNA 서열 clustering 이 중요도
- proteomics + BCR-seq 통합 분석
- Heavy / light chain 을 동시에 시퀀싱 하기
- Short read로 assembly하기

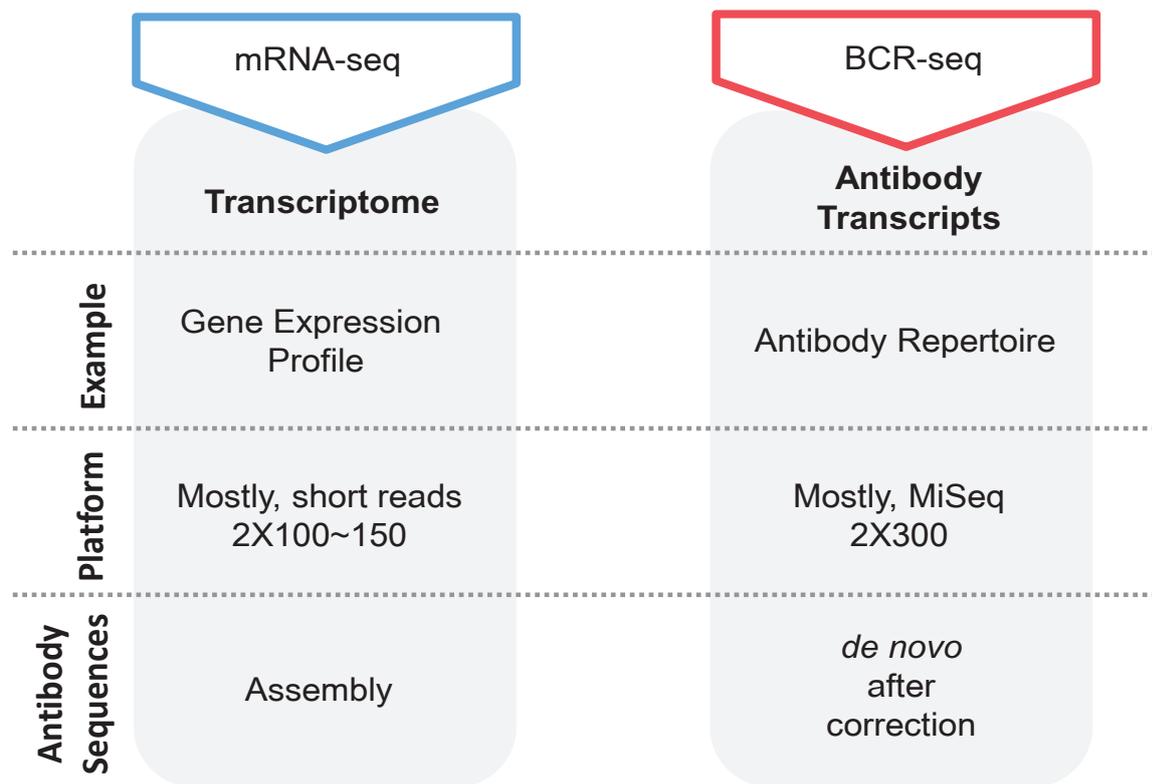
## Tumor infiltrating lymphocytes from mRNA-seq



# Reconstruction of Antibody from mRNA-seq



# Choice for TIL repertoire: mRNA-seq and BCR-seq



1

Adaptive Immune Receptor Repertoire (AIRR) 의 정의와 다양성

2

AIRR 연구를 위한 오믹스 접근법

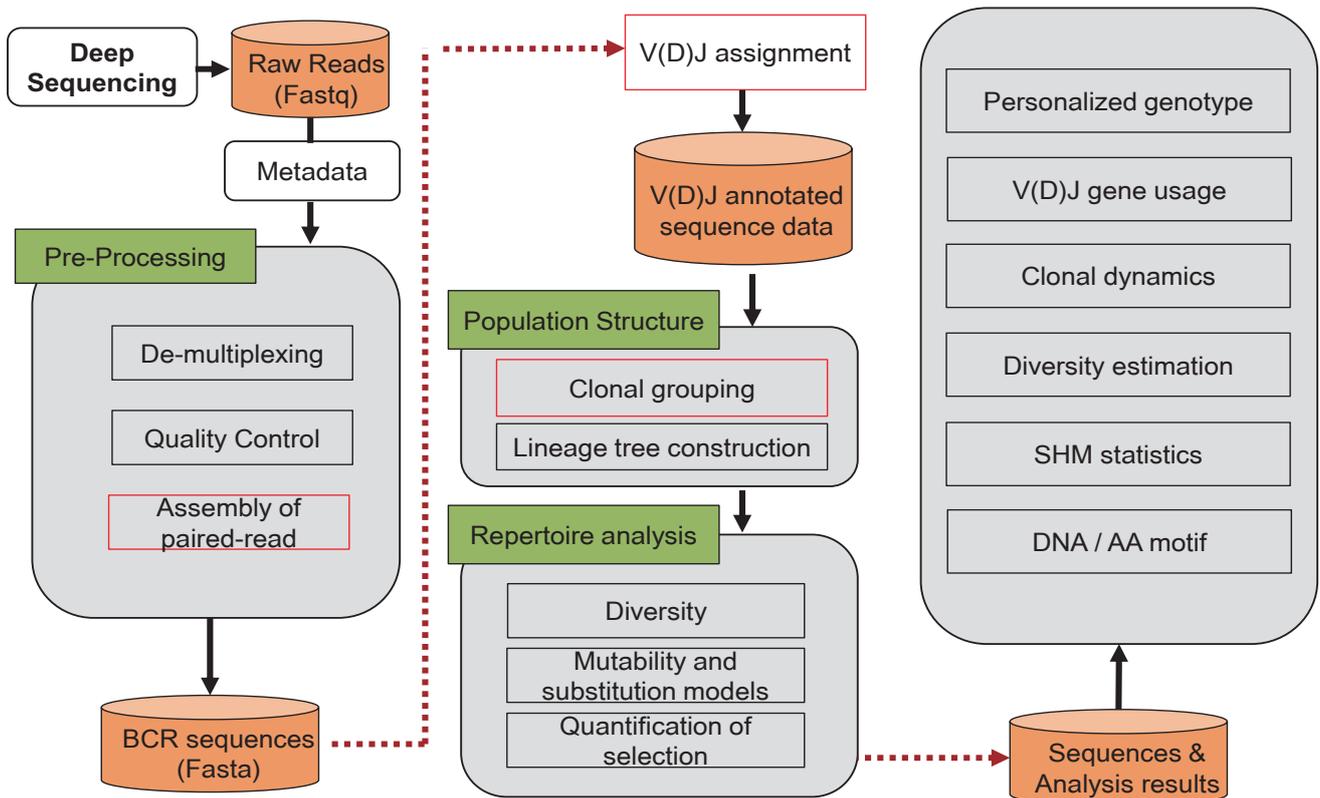
3

AIRR 생명정보 분석법

4

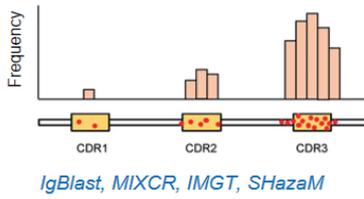
AIRR 연구 사례 및 분석 예시

## Bioinformatics Pipeline for BCR-seq

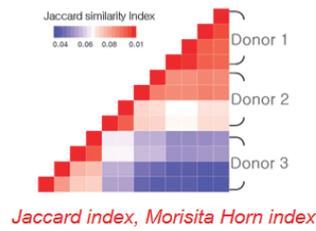


# Examples of BCR-seq downstream analysis

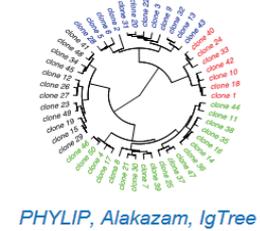
(A) Quantification of Somatic Hyper Mutation



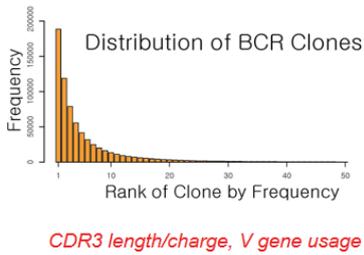
(B) Pairwise Overlap Analysis



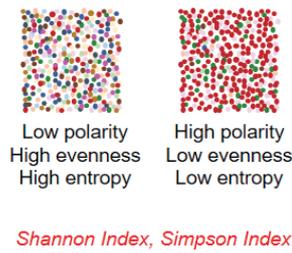
(C) Lineage Construction



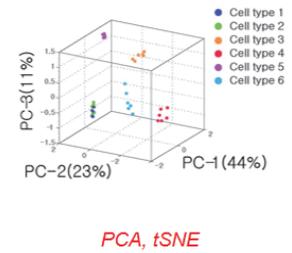
(D) Statistics and Distribution



(E) Ecological Diversity



(F) Dimension Reduction



1

Adaptive Immune Receptor Repertoire (AIRR) 의 정의와 다양성

2

AIRR 연구를 위한 오믹스 접근법

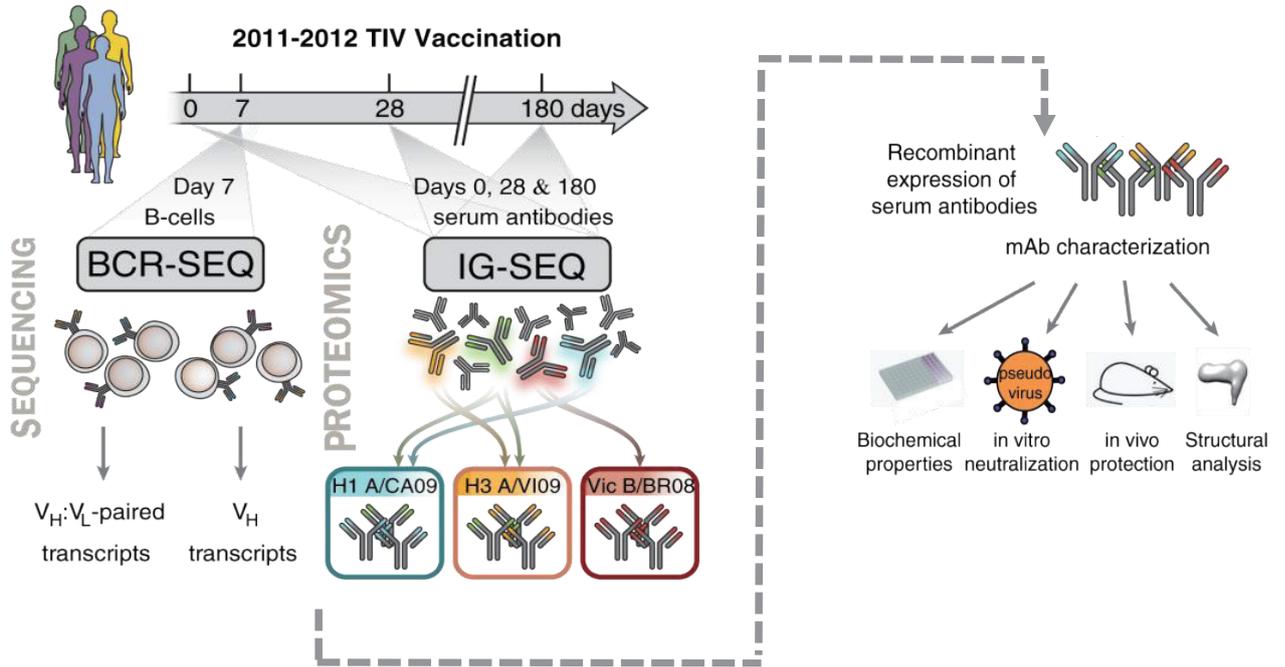
3

AIRR 생명정보 분석법

4

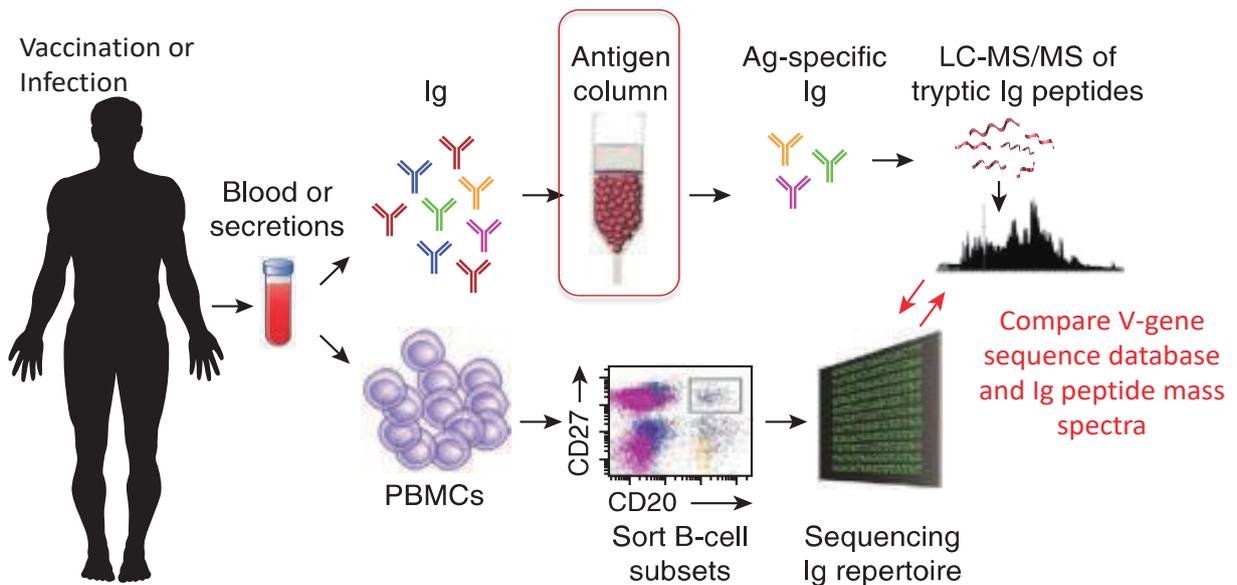
AIRR 연구 사례 및 분석 예시

# Example of Immunoproteogenomics: Flu Repertoire



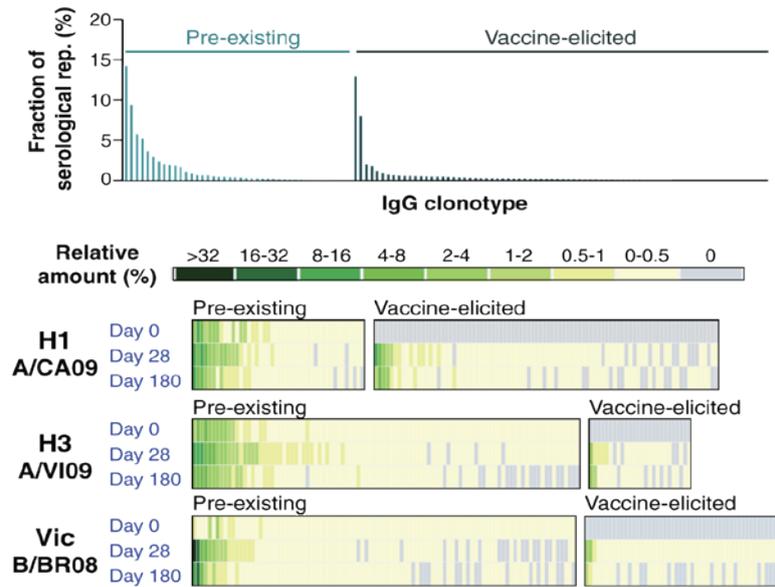
Georgiou and colleagues (2016) *Nature Medicine*

# Application of BCR-seq and Ig-seq



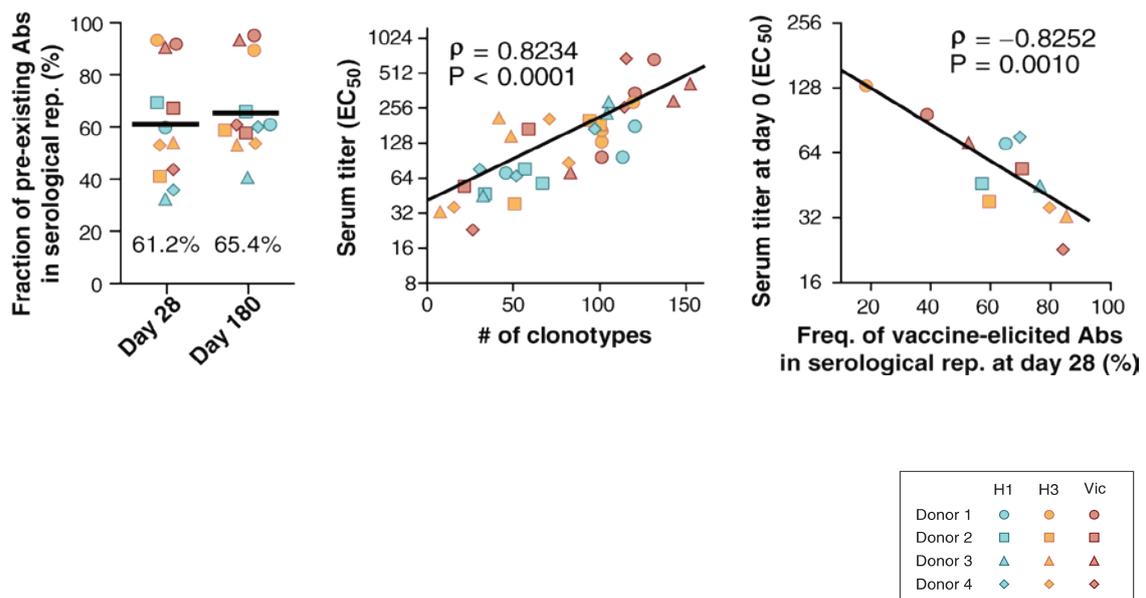
Georgiou et al., (2014) *Nature Biotechnology*

## Delineation of the serological repertoire



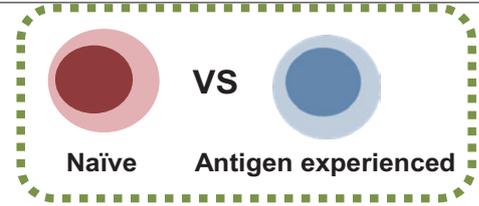
Georgiou and colleagues (2016) *Nature Medicine*

## Delineation of the serological repertoire



Georgiou and colleagues (2016) *Nature Medicine*

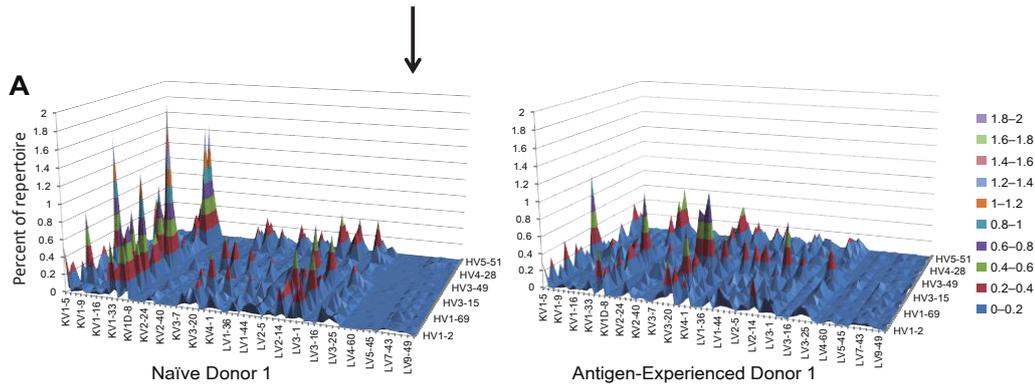
# VH:VL Pairing Experiments



70 ml Blood Draw  
 ↓  
 BCR pairing sequencing

< Unique Antibody Sequences >

Donor	CD3 <sup>-</sup> CD19 <sup>+</sup> CD20 <sup>+</sup> CD27 <sup>-</sup> Naïve	CD3 <sup>-</sup> CD19 <sup>+</sup> CD20 <sup>+</sup> CD27 <sup>+</sup> Ag-Exp
1	13,780	34,692
2	26,372	89,249
3	15,203	NA
Total	55,355	123,941

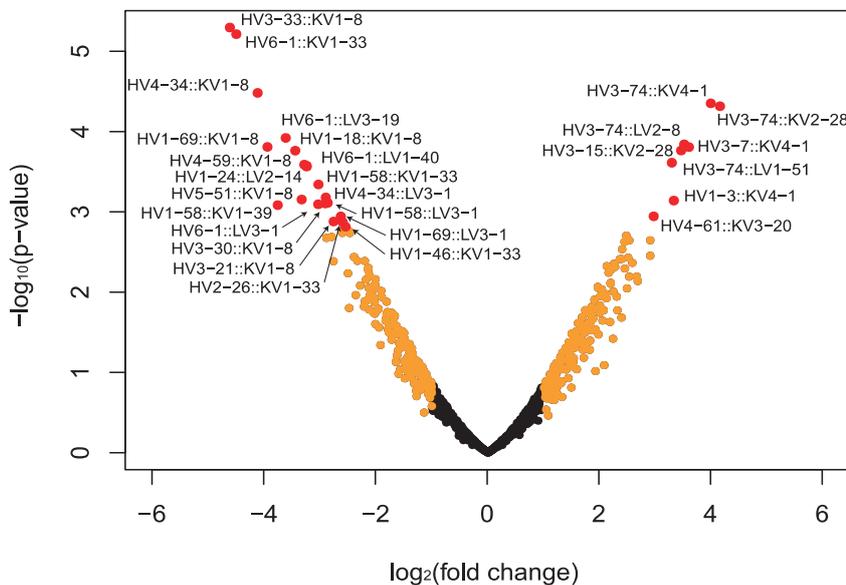


DeKosky\*, Oana\*, Park *et al.*, (2016) PNAS

# Identification of differentially paired VH:VL genes

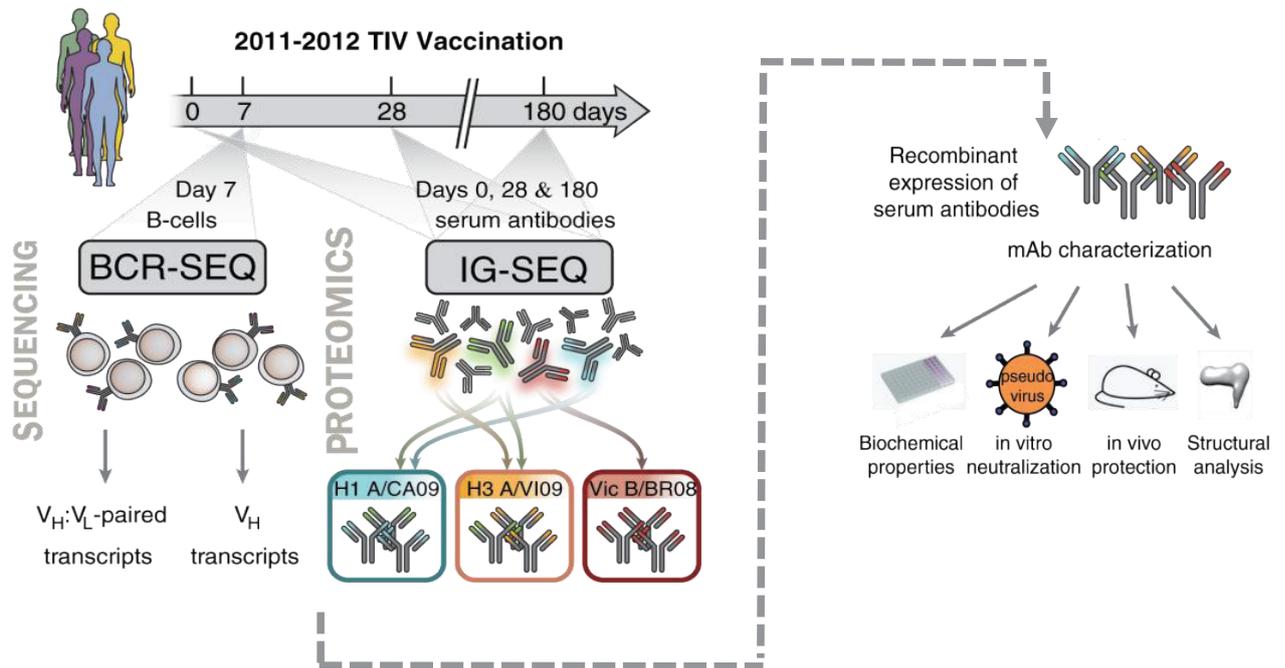


Differentially Paired Genes



DeKosky\*, Oana\*, Park *et al.*, (2016) PNAS

## Example of Immunoproteogenomics: Flu Repertoire



Georgiou and colleagues (2017) *Nature Medicine*

## Summary

1

### Adaptive Immune Receptor Repertoire (AIRR) 의 정의와 다양성

- $V_H$ 의 구조

2

### AIRR 연구를 위한 오믹스 접근법

- BCR-seq 데이터 분석에서 UMI의 사용과 DNA 서열 clustering의 중요도
- proteomics + BCR-seq 통합 분석
- Heavy / light chain 을 동시에 시퀀싱 하기
- Short read로 assembly하기

3

### AIRR 생명정보 분석법

4

### AIRR 연구 사례 및 분석 예시

1

Adaptive Immune Receptor Repertoire (AIRR) 의 정의와 다양성

2

AIRR 연구를 위한 오믹스 접근법

3

AIRR 생명정보 분석법

4

AIRR 연구 사례 및 분석 예시

## 실습순서

### 1. 데이터 다운로드하기

- Sratoolkit

### 2. 데이터 확인하기

- 프라이머 서열 확인: 리눅스 grep & awk 명령어
- 항체 서열인지 확인해보기: UCSC genome browser & IMGT

### 3. 분석하기

- R1 and R2 이어 붙이기
- MIXCR 프로그램 이용해서 V gene과 CDR3 서열 찾기

<https://www.theserverside.com/blog/Coffee-Talk-Java-News-Stories-and-Opinions/How-do-I-install-Java-on-Ubuntu>

RESEARCH ARTICLE | IMMUNOLOGY AND INFLAMMATION | ●

PNAS

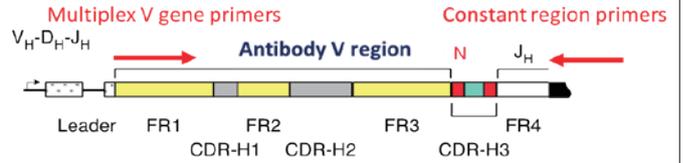
## Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires

Brandon J. DeKosky, Qana I. Lungu, Daechan Park, , and George Georgiou  [Authors Info & Affiliations](#)

Edited by James A. Wells, University of California, San Francisco, CA, and approved March 30, 2016 (received for review December 24, 2015)

April 25, 2016 | 113 (19) E2636-E2645 | <https://doi.org/10.1073/pnas.1525510113>

2x250 or 2x300 technology. Forward primers targeted the antibody Framework 1 regions (3); reverse primers targeted the IgM/IgG/IgA constant region for CD27<sup>+</sup> NBCs, and IgM/IgG/IgA/IgE/IgA reverse primers were used for CD27<sup>+</sup> AEBs. Full length VH and VL genes were generated for antigen-experienced repertoires via bioinformatic assembly of three Illumina sequencing samples (VH.VL, VH only, and VL only) as described previously (2–4). The following barcoded primers were used for VH-only amplification and sequencing (barcodes are italicized): Donor 1 Replicate 1 5'-NNNN *TGAAGG* GGCTAGCTATTCCCATCGCGG-3', Donor 1 Replicate 2 5'-NNNN *CGCGTC* GGCTAGCTATTCCCATCGCGG-3', Donor 2 Replicate 1 5'-NNNN *TAAGAA* GGCTAGCTATTCCCATCGCGG-3', Donor 2 Replicate 2 5'-NNNN *AGCGAG* GGCTAGCTATTCCCATCGCGG-3'. The following barcoded primers were used for VL-only amplification and sequencing (barcodes are italicized): Donor 1 Replicate 1 5'-NNNN *TGAAGG* GCGCCGCGATGGGAAT-3', Donor 1 Replicate 2 5'-NNNN *CGCGTC* GCGCCGCGATGGGAAT-3', Donor 2 Replicate 1 5'-NNNN *TAAGAA* GCGCCGCGATGGGAAT-3', Donor 2 Replicate 2 5'-NNNN *AGCGAG* GCGCCGCGATGGGAAT-3'.



R= A or G

	Primer Sequence (enter letters only, 5' to 3')
hVH1	<i>tattccatcgcggcgc</i> CAGGTCCAGCTKGTRCAGTCTGG
hVH157	<i>tattccatcgcggcgc</i> CAGGTGCAGCTGGTGSARTCTGG
hVH2	<i>tattccatcgcggcgc</i> CAGRTCACCTTGAAGGAGTCTG
hVH3	<i>tattccatcgcggcgc</i> GAGGTGCAGCTGKTGGAGWCY
hVH4	<i>tattccatcgcggcgc</i> CAGGTGCAGCTGCAGGAGTCSG
hVH6	<i>tattccatcgcggcgc</i> CAGGTACAGCTGCAGCAGTCA
hVH3N	<i>tattccatcgcggcgc</i> TCAACACAACGGTTCCAGTTA
hlg_MiSqRev1_4N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG NNNN ATGGGCCCTG SGATGGGCCCTTGGTGGARGC
hlgM_MiSqRev1_4N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG NNNN ATGGGCCCTG GGTGGGGCGGATGCACTCC
hlgA_MiSqRev1_4N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG NNNN ATGGGCCCTG CTTGGGGCTGGTCGGGGATG