Unraveling Immunogenomic Diversity in Single-Cell Data

Ahmad Al Ajami[1,2,3], Jonas Schuck[1,2,3], Federico Marini[4], Katharina Imkeller[1,2,3]

[1]Goethe University, University Hospital Frankfurt, Neurological Institute (Edinger Institute), Frankfurt/Main, Germany; [2]University Cancer Center, Frankfurt/Main, Germany; [3]Frankfurt Cancer Institute, Frankfurt/Main, Germany; [4]Johannes Gutenberg University, University Medical Center Mainz, IMBEI, Mainz, Germany
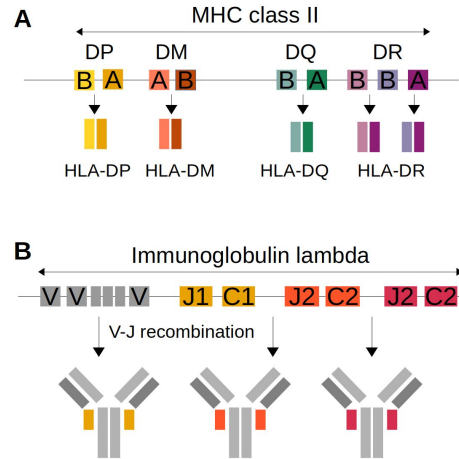
Grand Rapids
25 July 2024

Structural diversity of receptors

- Receptors that coordinate antigen specificity in immune responses
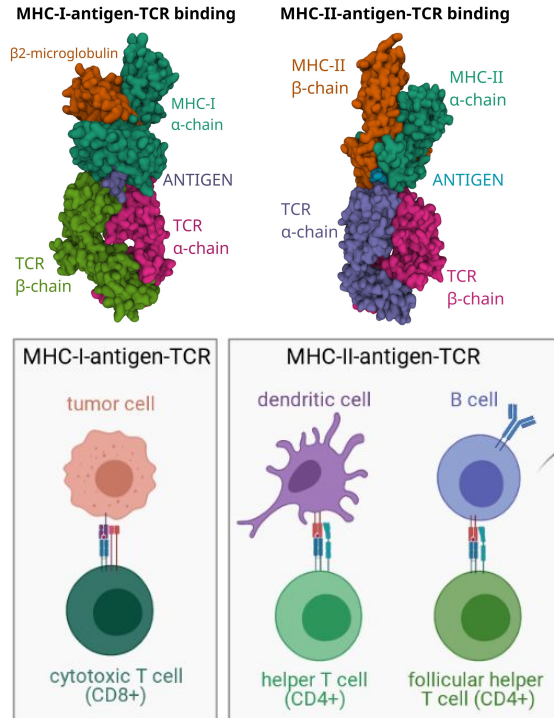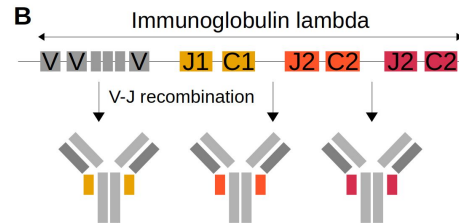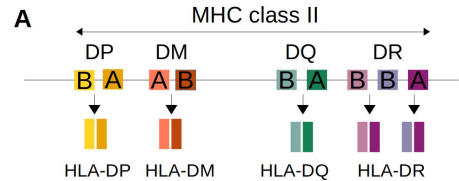
Genetic mechanisms of structural diversity

Polygeny:
Multiple similar genes encode same function

A — MHC class II

DP DM DQ DR

HLA-DP HLA-DM HLA-DQ HLA-DR

B — Immunoglobulin lambda

V-J recombination

Gene group	Number of **similar genes** (incl. pseudogenes)
IGHG	5
HLA-DRB	8
…	…

Genetic mechanisms of structural diversity

Polygeny:
Multiple similar genes
encode same function

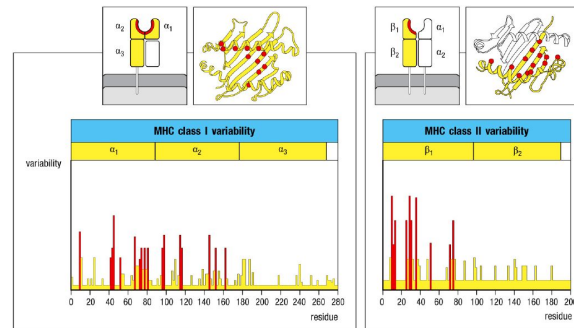**Hyperpolymorphism:**
Multiple alleles
for one gene

A — MHC class II: DP, DM, DQ, DR → HLA-DP, HLA-DM, HLA-DQ, HLA-DR

B — Immunoglobulin lambda: V V V ... V, J1 C1, J2 C2, J2 C2; V-J recombination

MHC class I variability — α_1, α_2, α_3

MHC class II variability — β_1, β_2

Figure 6.21 Janeway's Immunobiology, 9th ed. (© Garland Science 2017)

Gene group	Number of **similar genes** (incl. pseudogenes)
IGHG	5
HLA-DRB	8
…	…

| Gene group	Number of **alleles** (incl. pseudogenes)
IGHG	71
HLA-DRB	~4200
…	…

Source: IMGT and IPD-IMGT/HLA

Typical transcriptomic analysis workflow

1. For each sequencing read:
   a. identify cellular barcode
   b. identify unique molecular identifier (UMI)
   c. map read to reference genome
2. Count unique transcripts per gene and barcode using UMI
3. Generate feature count matrix (genes x cells)
4. Downstream analysis
   a. dimension reduction
   b. clustering
   c. differential gene expression
   d. …

Typical transcriptomic analysis workflow

1. For each sequencing read:
   a. identify cellular barcode
   b. identify unique molecular identifier (UMI)
   c. map read to reference genome
2. Count unique transcripts per gene and barcode using UMI
3. Generate feature count matrix (genes x cells)
4. Downstream analysis
   a. dimension reduction
   b. clustering
   c. differential gene expression
   d. …

SingleCellExperiment
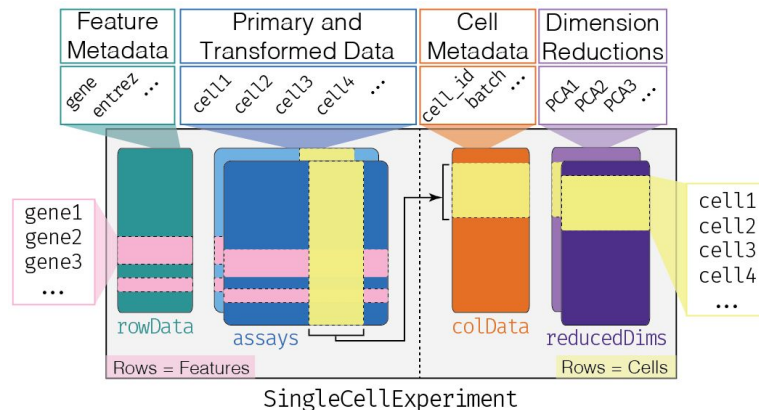
Allelic diversity and polygeny pose a bioinformatic challenge

Polygeny:
Multiple similar genes
encode same function
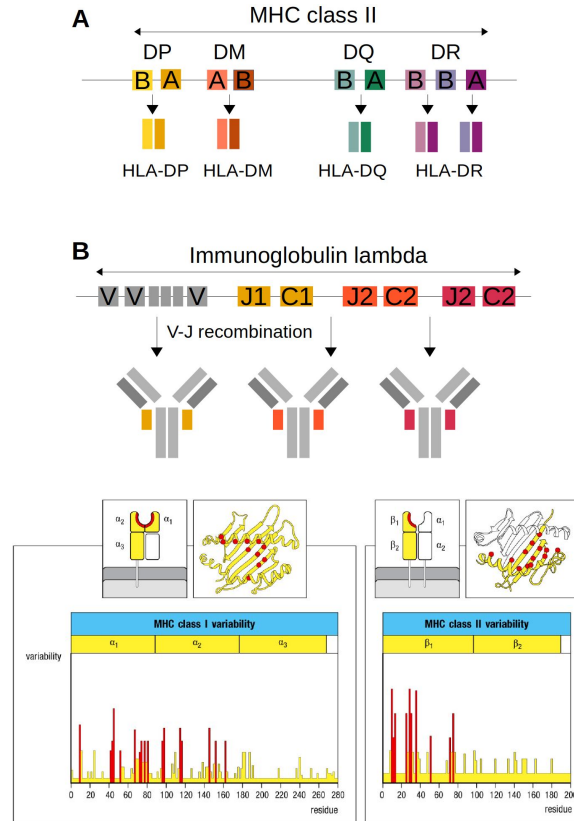
Bioinformatic challenge

Cross-mapping: uniquely
mapping to the wrong gene

Hyperpolymorphism:
Multiple alleles
for one gene

Figure 6.21 Janeway's Immunobiology, 9th ed. (© Garland Science 2017)

Allelic diversity and polygeny pose a bioinformatic challenge

Polygeny:
Multiple similar genes
encode same function

Hyperpolymorphism:
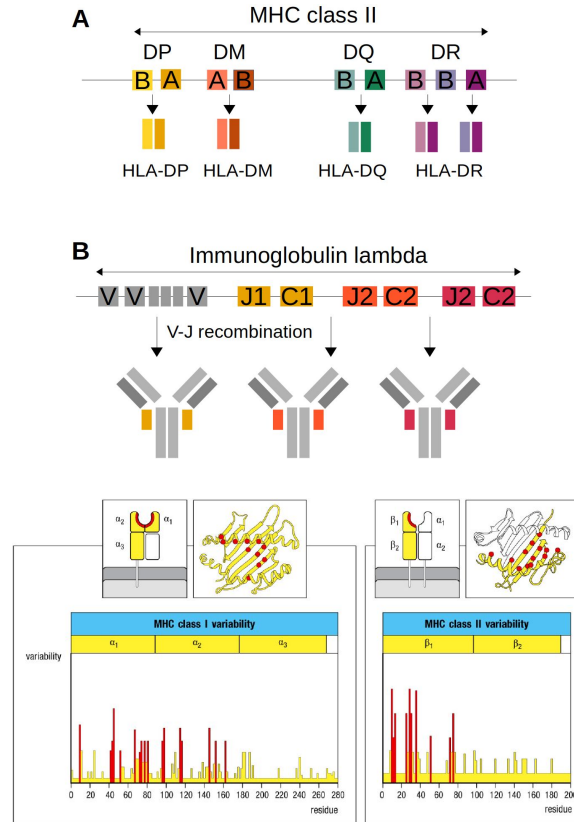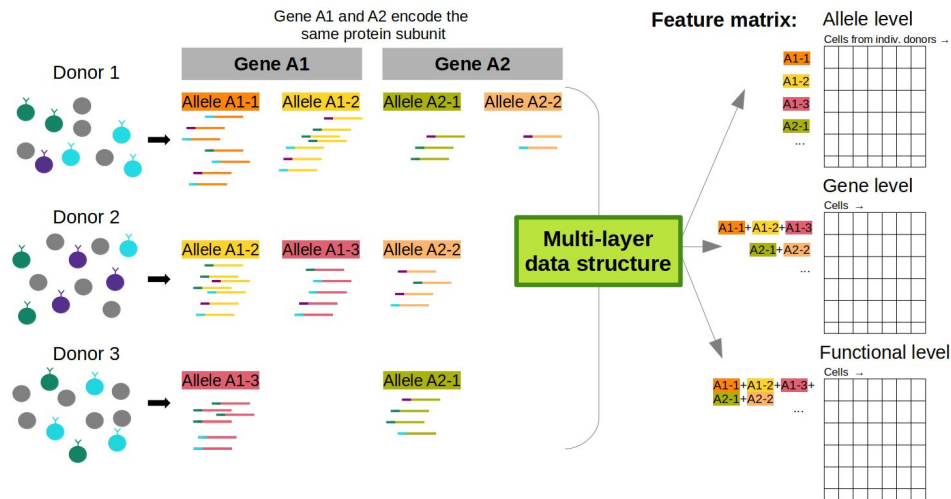Multiple alleles
for one gene

Aim

To perform quantification of allele-specific expression for immunogenomic analysis

Figure 6.21 Janeway's Immunobiology, 9th ed. (© Garland Science 2017)

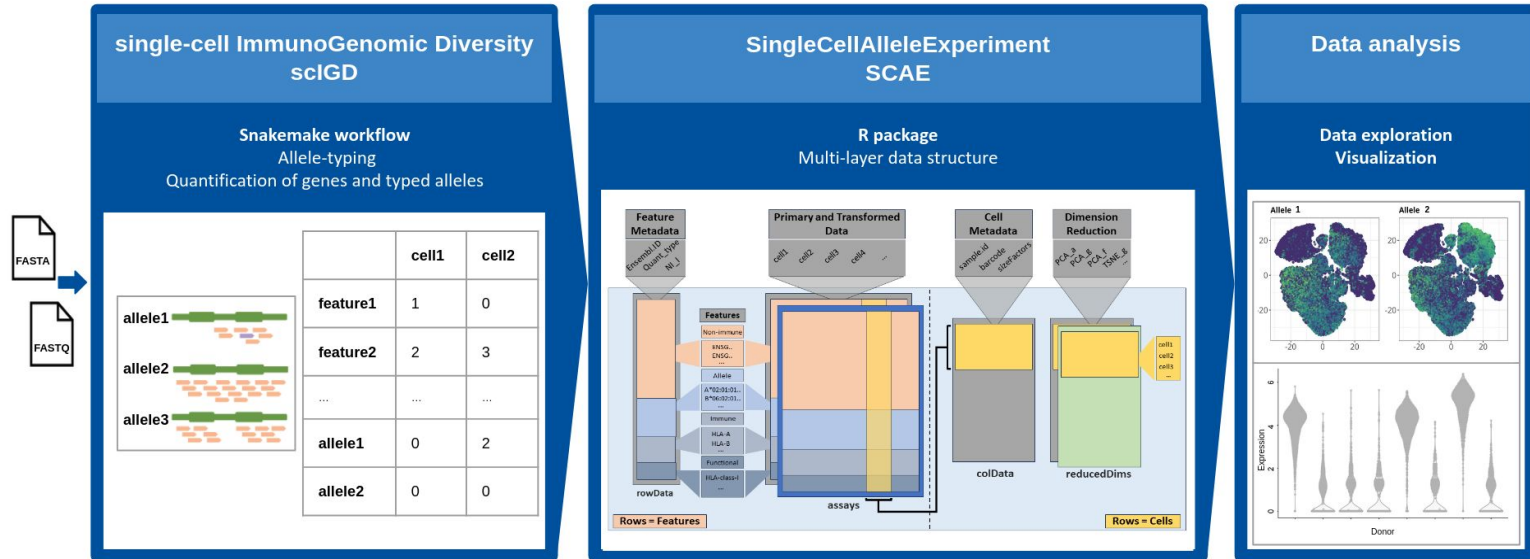# Implementation of allele and functional information
What we want

- Building a multi-layer data structure for immune gene representation in single-cell data

Workflow
How to get there

- Example datasets present in *scaeData* – R/ExperimentHub data package

Acknowledgements

Project partner

Federico Marini

Group of Computational Immunology

Katharina Imkeller

Jonas Schuck

✉ alajami@med.uni-frankfurt.de

https://agimkeller.github.io

scIGD

scaeData

SingleCellAlleleExperiment