Who pairs with whom?

High-throughput sequencing of the human paired heavy and light chain repertoire

Technical Journal Club September 15th

Christina Müller

nature biotechnology

High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire

Brandon J DeKosky[1], Gregory C Ippolito[2], Ryan P Deschner[1], Jason J Lavinder[3], Yariv Wine[1], Brandon M Rawlings[1], Navin Varadarajan[4], Claudia Giesecke[5,6], Thomas Dörner[5,6], Sarah F Andrews[7], Patrick C Wilson[7], Scott P Hunicke-Smith[3], C Grant Willson[1,8], Andrew D Ellington[3,8] & George Georgiou[1–3,9]

TECHNICAL REPORTS

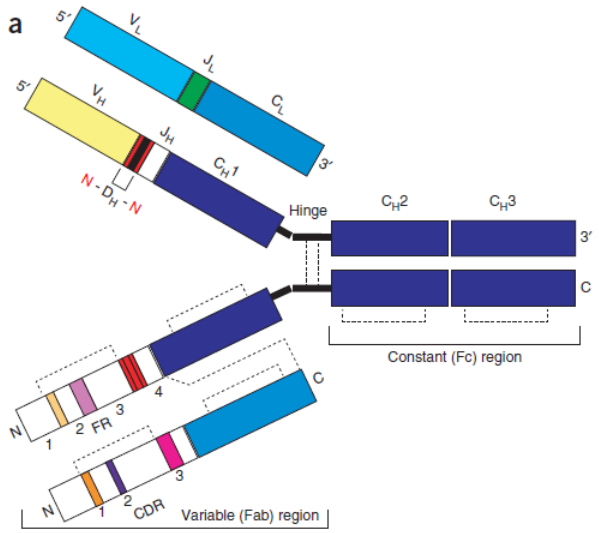
nature medicine

In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire

Brandon J DeKosky[1], Takaaki Kojima[1,2], Alexa Rodin[1], Wissam Charab[1], Gregory C Ippolito[3], Andrew D Ellington[4] & George Georgiou[1,3,5,6]

Background

- antibody repertoire is the sum of all circulating antibodies produced by the B cells
- total number of B lymphocytes ~ 1-2 x 10^{11}

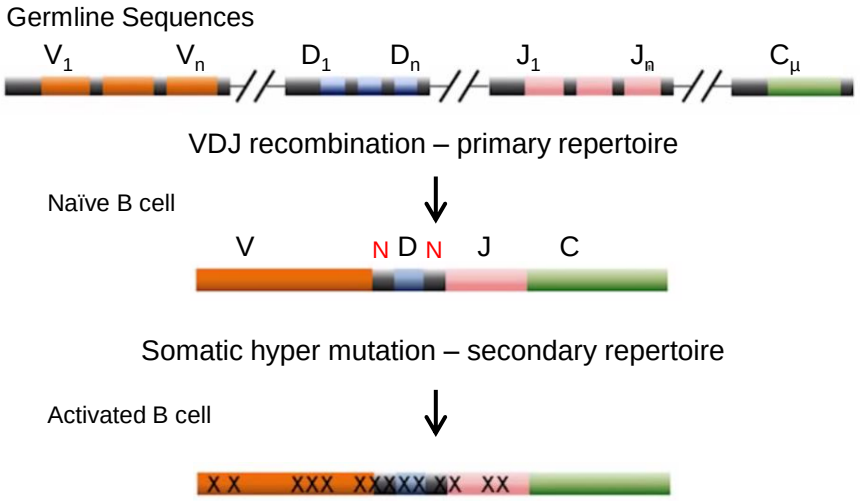


Georgiou et al. 2014

Generation of the antibody repertoire:

- V(D)J recombination
- addition or deletion of nucleotides in regions where junctions occur

Functional gene segments in human IgG loci			
Segment	Light chains		Heavy chain
	κ	λ	H
Variable (V)	34-38	29-33	38-46
Diversity (D)	-	-	27
Joining (J)	5	4-5	6
Constant (C)	1	4-5	9



Background

Why is it important to determine the antibody repertoire?

- provides important information on protective and pathogenic immunity
- capturing the nature of a successful antibody response
- Ig-sequencing combined with other techniques such as the expression and isolation of antigen-specific antibodies, sequencing of multiple RNAs from single cells or proteomic analysis help to identify antibody properties mediating protection against infectious diseases or autoimmune response

Background

Why is it important to determine the antibody repertoire?

- provides important information on protective and pathogenic immunity
- capturing the nature of a successful antibody response
- Ig-sequencing combined with other techniques such as the expression and isolation of antigen-specific antibodies, sequencing of multiple RNAs from single cells or proteomic analysis help to identify antibody properties mediating protection against infectious diseases or autoimmune response

First steps – Low-throughput analysis of the antibody repertoire:

- determination of IgH and IgL V(D)J recombinants in a few hundred B cells per experiment based on Sanger sequencing (1990s)

Background

Why is it important to determine the antibody repertoire?

- provides important information on protective and pathogenic immunity
- capturing the nature of a successful antibody response
- Ig-sequencing combined with other techniques such as the expression and isolation of antigen-specific antibodies, sequencing of multiple RNAs from single cells or proteomic analysis help to identify antibody properties mediating protection against infectious diseases or autoimmune response

First steps – Low-throughput analysis of the antibody repertoire:

- determination of IgH and IgL V(D)J recombinants in a few hundred B cells per experiment based on Sanger sequencing (1990s)

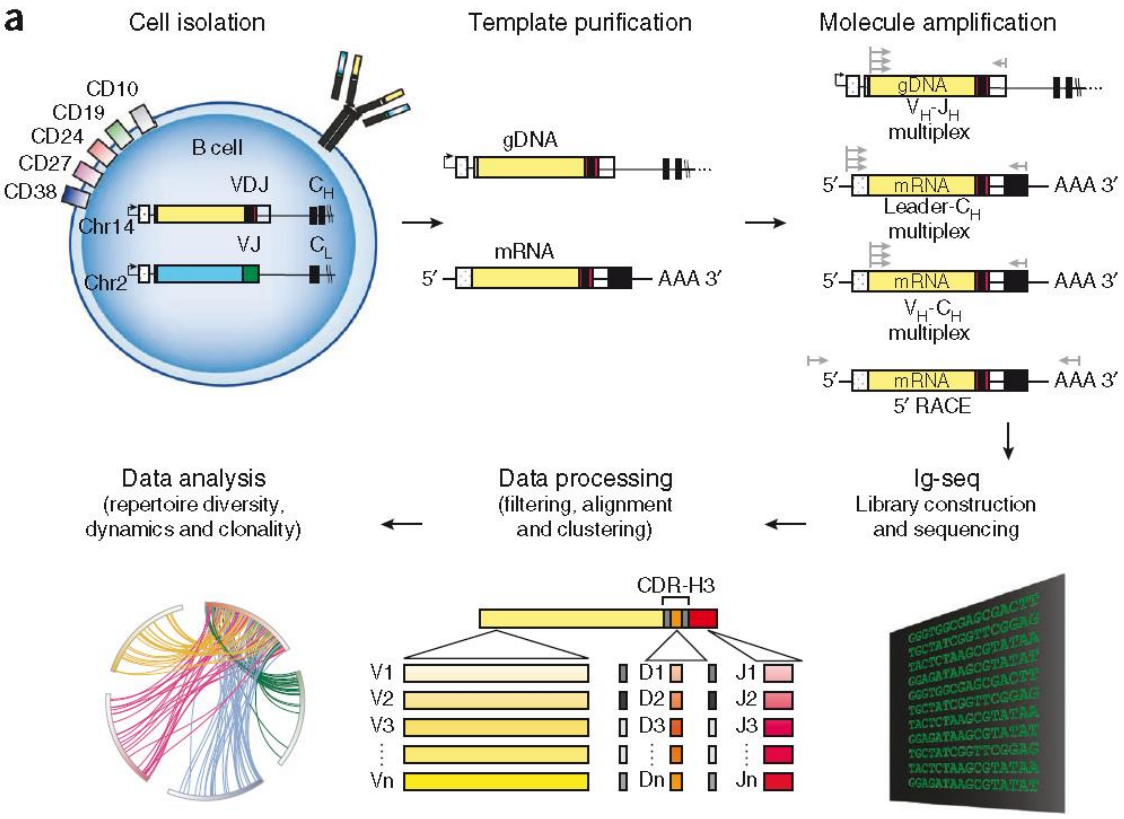
BUT

Low-throughput analysis provides only a small amount of information about the entire antibody repertoire & too labor intensive

Background

High-throughput sequencing of the antibody repertoire

- NGS allows in-depth antibody repertoire studies

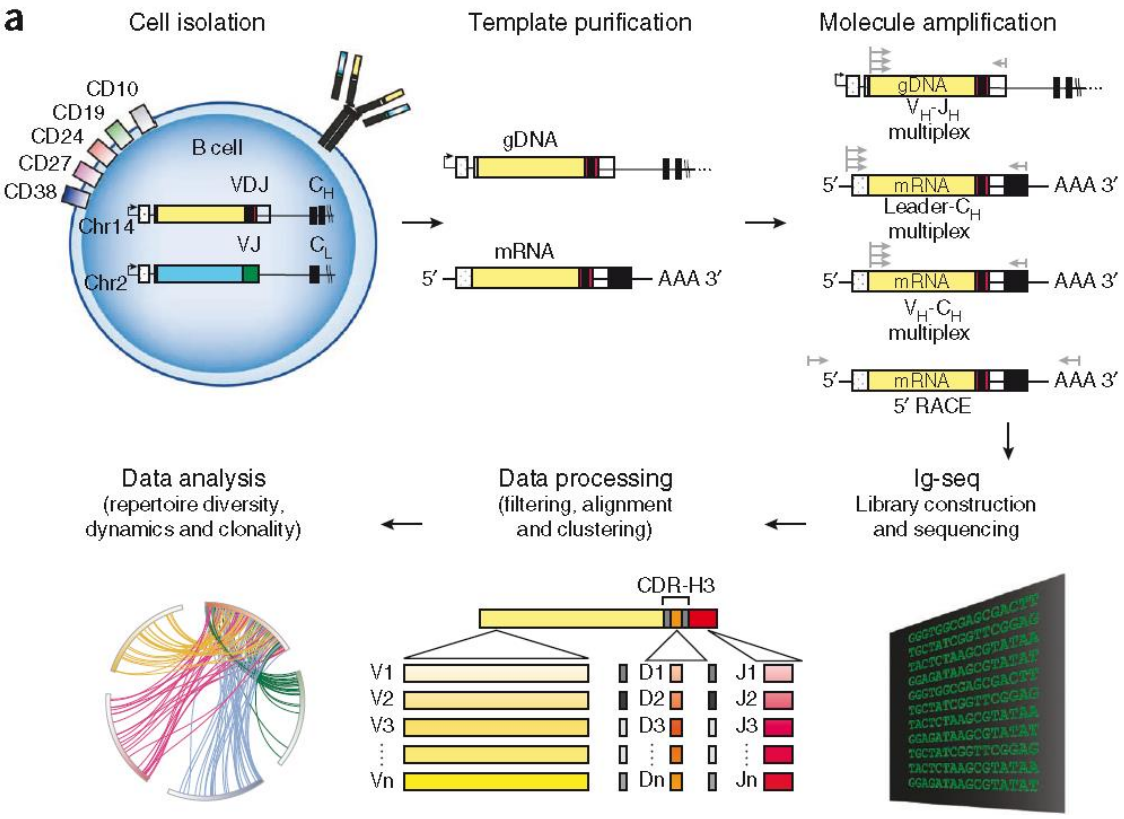


Georgiou et al. 2014

Background

High-throughput sequencing of the antibody repertoire

- NGS allows in-depth antibody repertoire studies

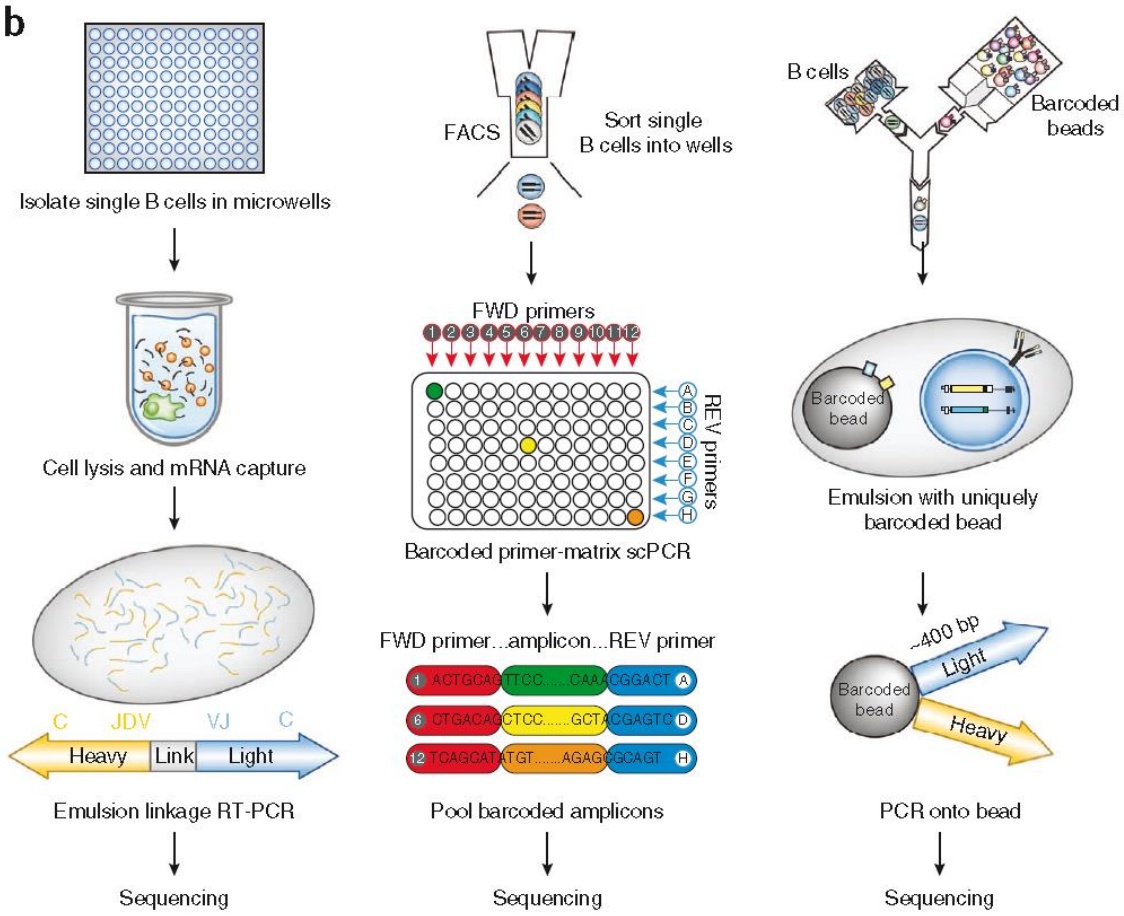


Georgiou et al. 2014

Information about the endogenous pairing of the V heavy (V_H) and V light (V_L) chain is lacking

Background

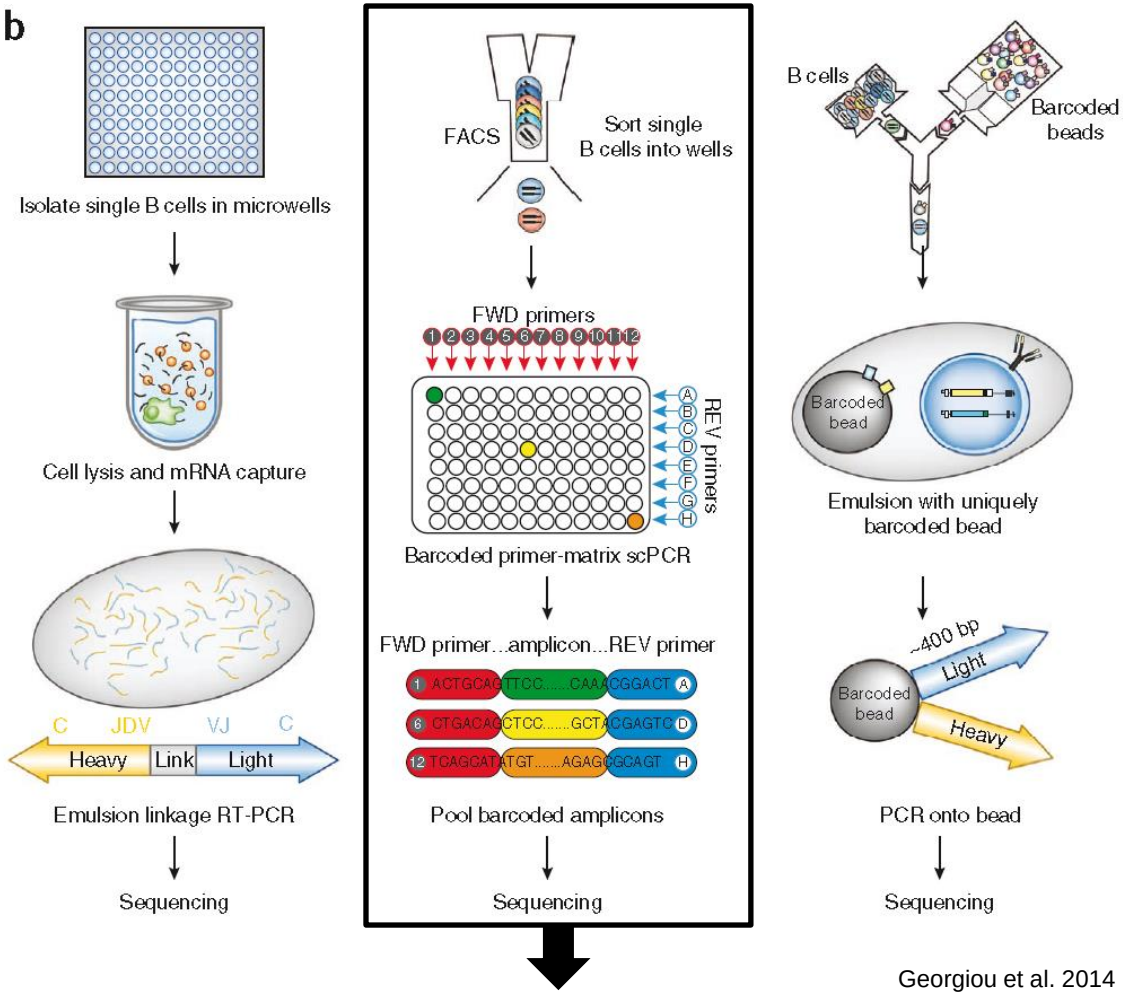
High-throughput sequencing of the antibody repertoire



Georgiou et al. 2014

Background

High-throughput sequencing of the antibody repertoire

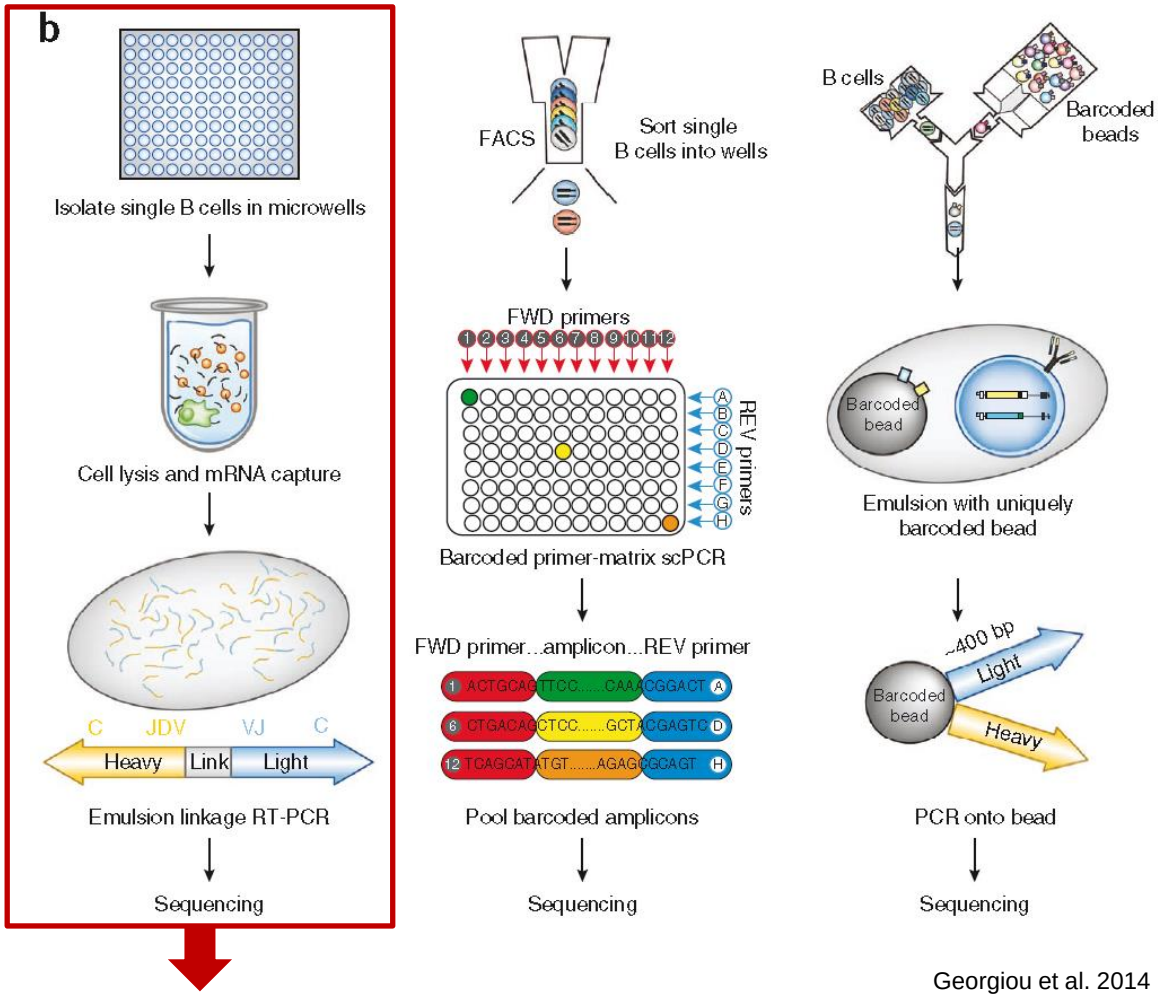


Georgiou et al. 2014

~46 000 individual B cells
(Busse et al. 2013)

Background

High-throughput sequencing of the antibody repertoire



~50 000 individual B cells
(DeKosky et al. 2013)

Georgiou et al. 2014

LETTERS

nature biotechnology

High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire

Brandon J DeKosky[1], Gregory C Ippolito[2], Ryan P Deschner[1], Jason J Lavinder[3], Yariv Wine[1], Brandon M Rawlings[1], Navin Varadarajan[4], Claudia Giesecke[5,6], Thomas Dörner[5,6], Sarah F Andrews[7], Patrick C Wilson[7], Scott P Hunicke-Smith[3], C Grant Willson[1,8], Andrew D Ellington[3,8] & George Georgiou[1–3,9]

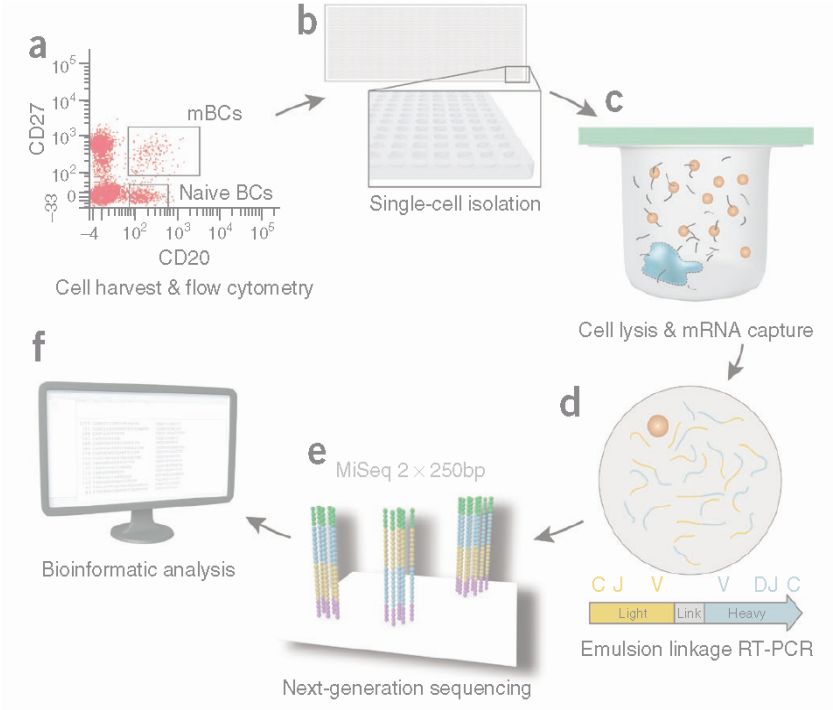
Experimental design

Source of B cells

- Lymph nodes (28%), Spleen and mucosal surface (23%), Red bone marrow (17%)
- Peripheral blood (only 2% of the 1-2 x 10^{11} B cells in the human body)

! Ig transcription varies up to 100 fold between naive B cells and plasma cells **!**

→ using unsorted bulk B cells will make it difficult to deduce cellular clonal frequencies



DeKosky et al. 2013

Experimental design

Source of B cells

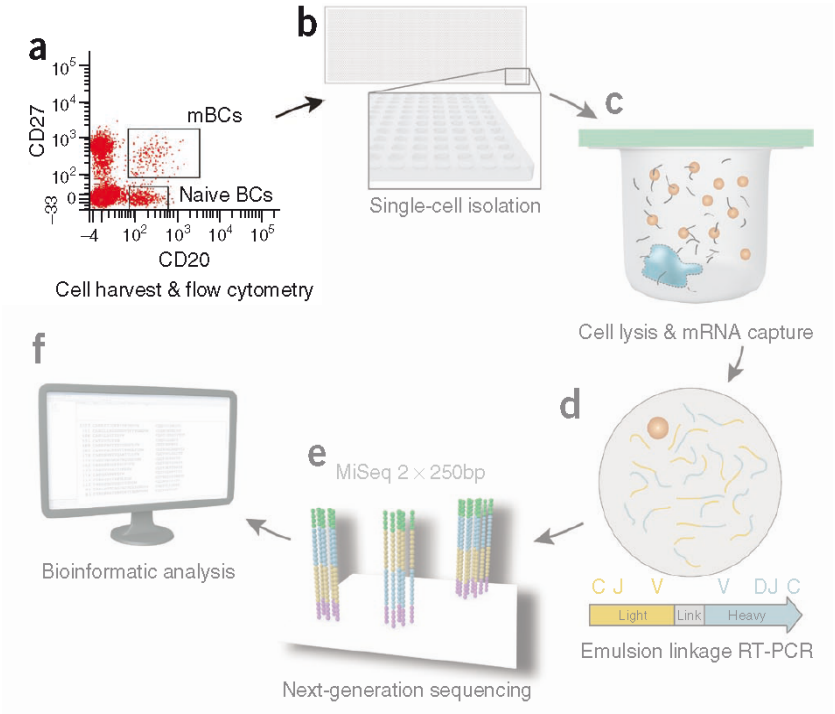
- Lymph nodes (28%), Spleen and mucosal surface (23%), Red bone marrow (17%)
- Peripheral blood (only 2% of the 1-2 x 10^{11} B cells in the human body)

! Ig transcription varies up to 100 fold between naive B cells and plasma cells **!**

→ using unsorted bulk B cells will make it difficult to deduce cellular clonal frequencies

a. FACS sorting

- PBMCs sorted for $CD19^{+}CD3^{-}CD27^{+}CD38^{int}$ memory B cells



DeKosky et al. 2013

Experimental design

Source of B cells

- Lymph nodes (28%), Spleen and mucosal surface (23%), Red bone marrow (17%)
- Peripheral blood (only 2% of the 1-2 x 10^{11} B cells in the human body)

! Ig transcription varies up to 100 fold between naive B cells and plasma cells **!**

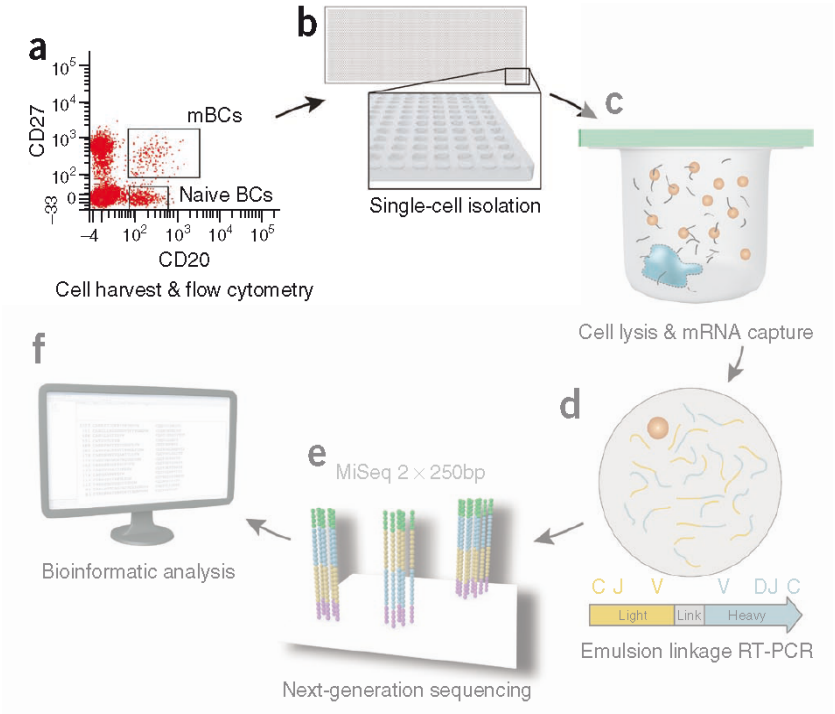
→ using unsorted bulk B cells will make it difficult to deduce cellular clonal frequencies

a. FACS sorting

- PBMCs sorted for $CD19^{+}CD3^{-}CD27^{+}CD38^{int}$ memory B cells

b. Single-cell isolation

- 125-pl wells molded in polydimethylsildoxan (PDMS) slides
- 1.7 x 10^{5} wells per slide



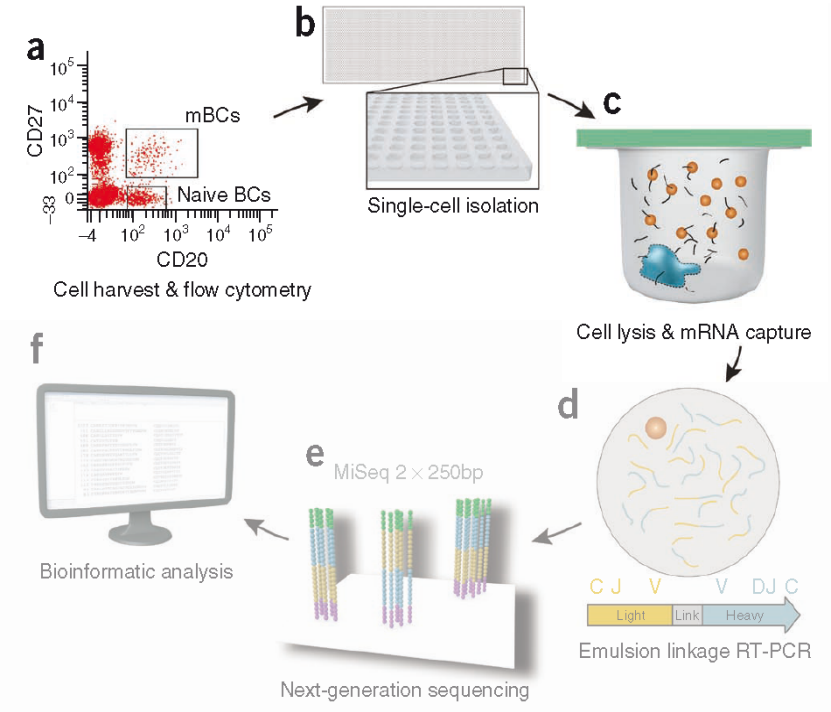
DeKosky et al. 2013

Experimental design

c. Cell lysis & mRNA capture

- poly(dT) magnetic beads added at an average 55 beads/well
- slides were incubated with optimized cell lysis solution (1% lithium dodecyl sulfate)

  → complete cell lysis within < 1min



DeKosky et al. 2013

Experimental design

c. Cell lysis & mRNA capture

- poly(dT) magnetic beads added at an average 55 beads/well
- slides were incubated with optimized cell lysis solution (1% lithium dodecyl sulfate)

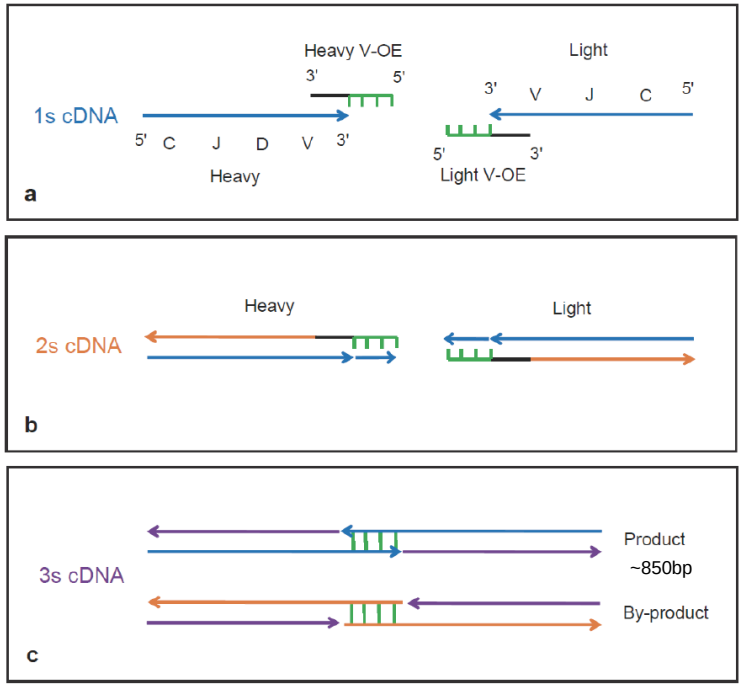
  → complete cell lysis within < 1min

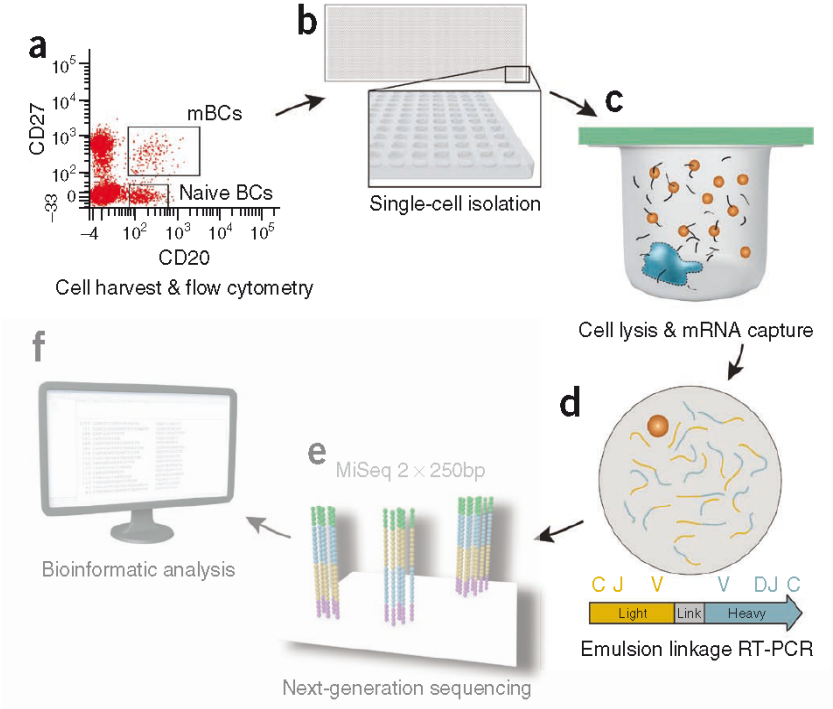
d. Emulsion linkage RT-PCR

- captured mRNA was emulsified with

  primers, RTase and thermostable DNA polymerase

  → RT-PCR & linkage PCR (Meijer et al. 2006)





DeKosky et al. 2013

Experimental design

c. Cell lysis & mRNA capture

- poly(dT) magnetic beads added at an average 55 beads/well
- slides were incubated with optimized cell lysis solution (1% lithium dodecyl sulfate)

→ complete cell lysis within < 1min

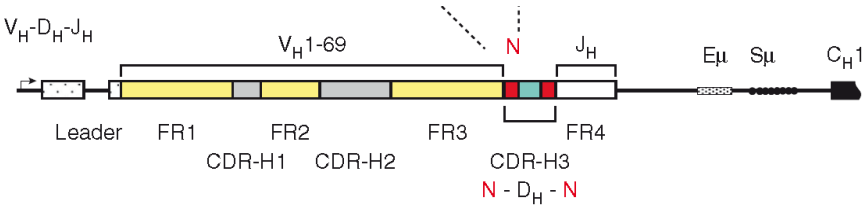
d. Emulsion linkage RT-PCR

- captured mRNA was emulsified with primers, RTase and thermostable DNA polymerase

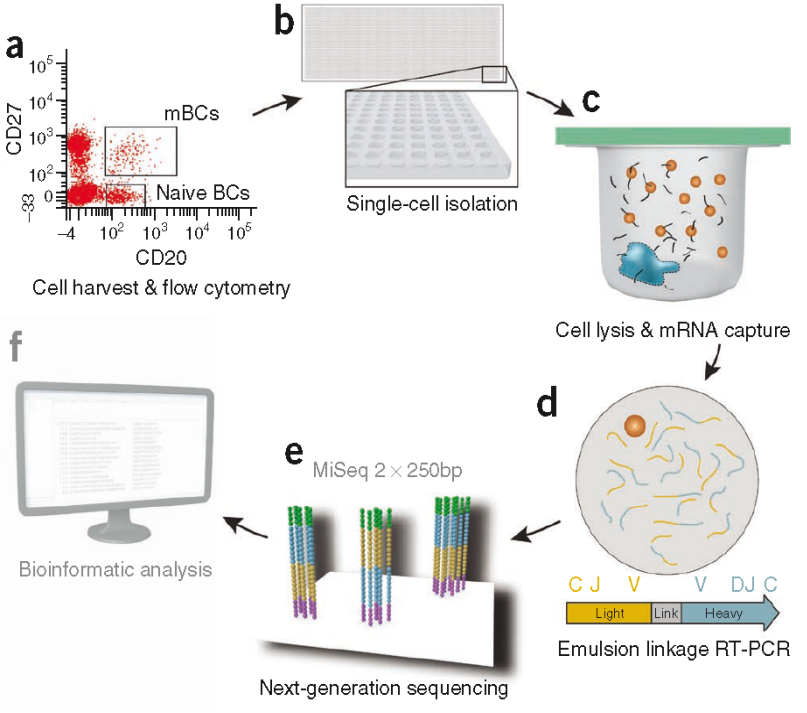
→ RT-PCR & linkage PCR (Meijer et al. 2006)

e. NGS using Illumina MiSeq 2 x250pb

- sequencing of CDR-H3 and CDR-L3



Georgiou et al. 2014



DeKosky et al. 2013

Experimental design

c. Cell lysis & mRNA capture

- poly(dT) magnetic beads added at an average 55 beads/well
- slides were incubated with optimized cell lysis solution (1% lithium dodecyl sulfate)

→ complete cell lysis within < 1min

d. Emulsion linkage RT-PCR

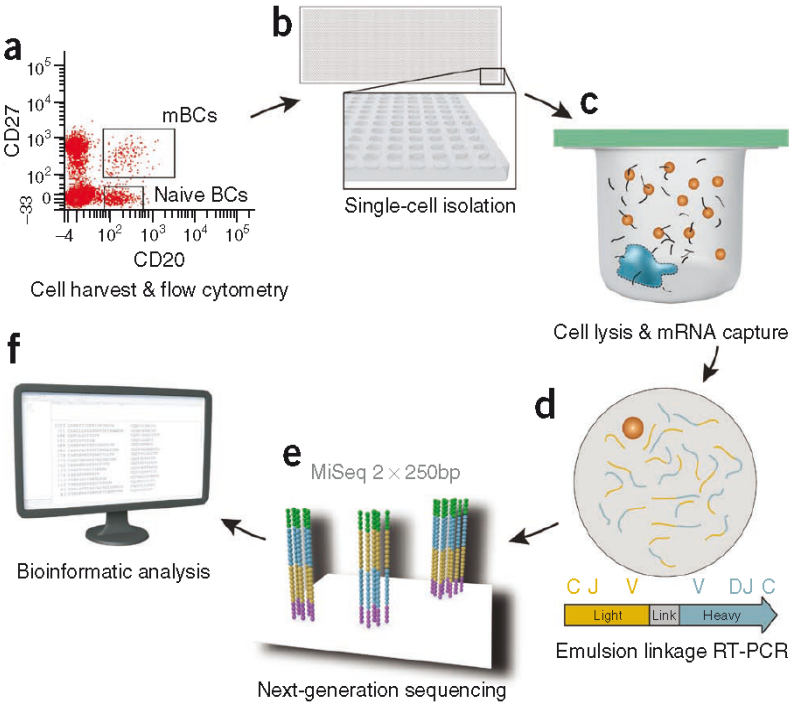
- captured mRNA was emulsified with primers, RTase and thermostable DNA polymerase

→ RT-PCR & linkage PCR

e. NGS using Illumina MiSeq 2 x250pb

- sequencing of CDR-H3 and CDR-L3

f. Bioinformatic analysis



DeKosky et al. 2013

Evaluation of Methodology

IgG$^+$ B cells from two healthy individuals

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

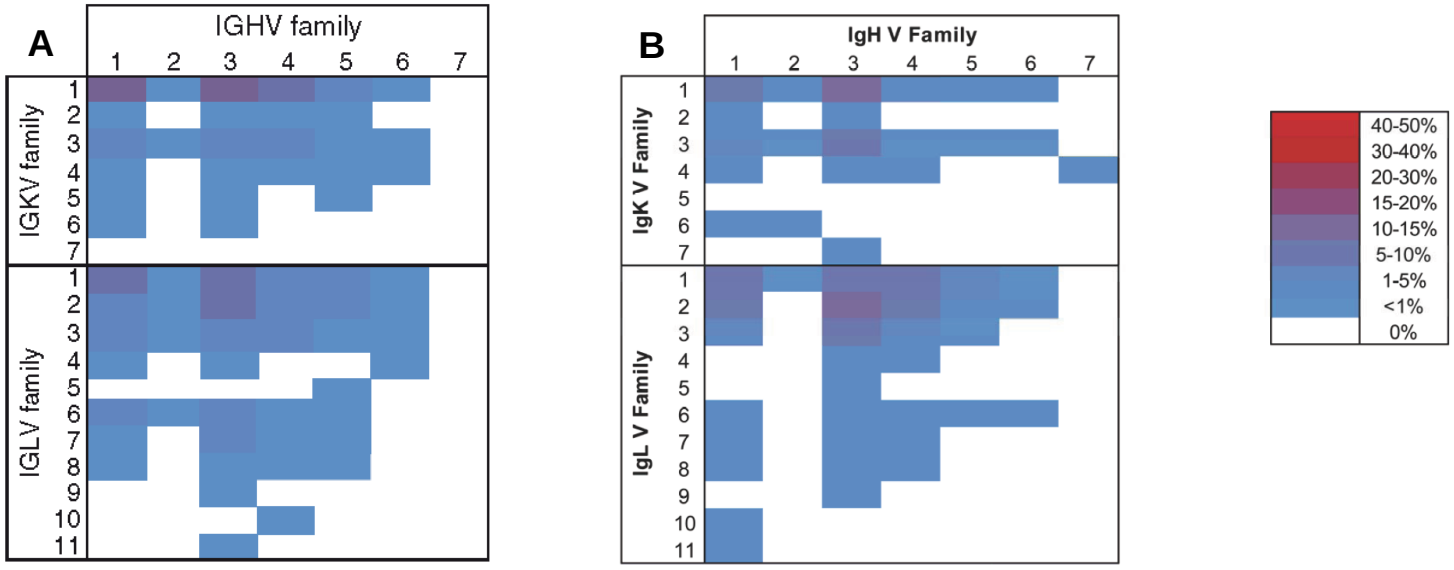
Evaluation of Methodology

IgG⁺ B cells from two healthy individuals

B cells were spiked with immortalized IM-9 lymphoblasts (~4% of total mixture)

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination



- 61,000 B cells
- 2,716 unique pairs
- correct pairing of IM-9 V_H and V_L 78 fold above background

- 47,000 B cells
- 2,248 unique pairs
- correct pairing of IM-9 V_H and V_L 125 fold above background

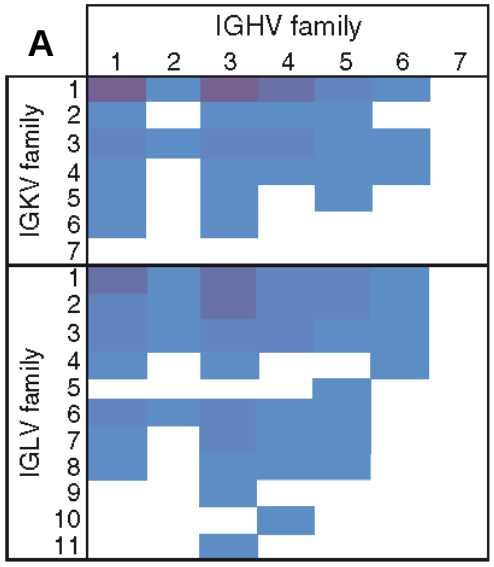
Evaluation of Methodology

IgG$^+$ B cells from two healthy individuals

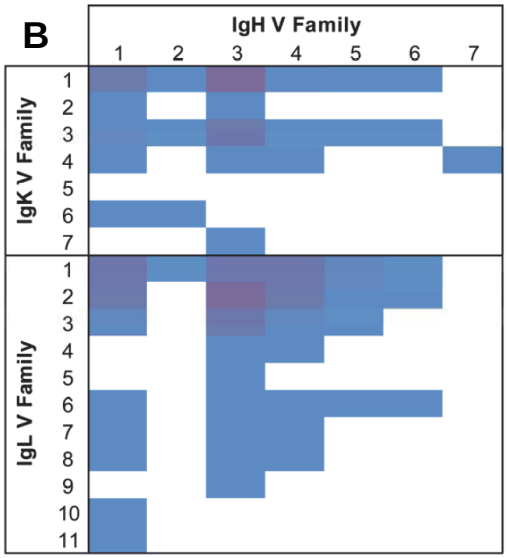
B cells were spiked with immortalized IM-9 lymphoblasts (~4% of total mixture)

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

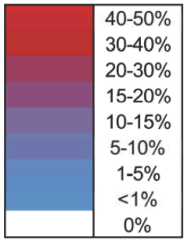
Memory B cells from a healthy individual 14d after influenza vaccination



- 61,000 B cells
- 2,716 unique pairs
- correct pairing of IM-9 V_H and V_L 78 fold above background



- 47,000 B cells
- 2,248 unique pairs
- correct pairing of IM-9 V_H and V_L 125 fold above background



Spearman rank correlation coefficient= 0.804; $P < 10^{-29}$

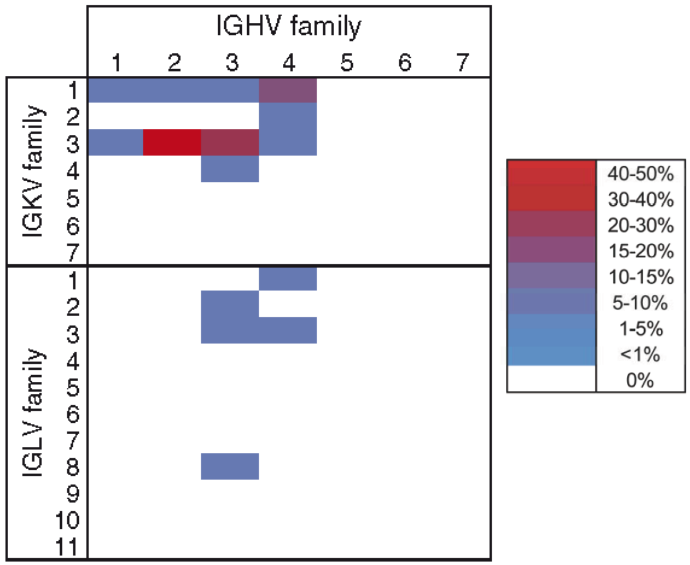
Evaluation of Methodology

IgG⁺ B cells from two healthy individuals

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

B cells were spiked with immortalized ARH-77



- 400 recovered B cells
- 86 unique pairs

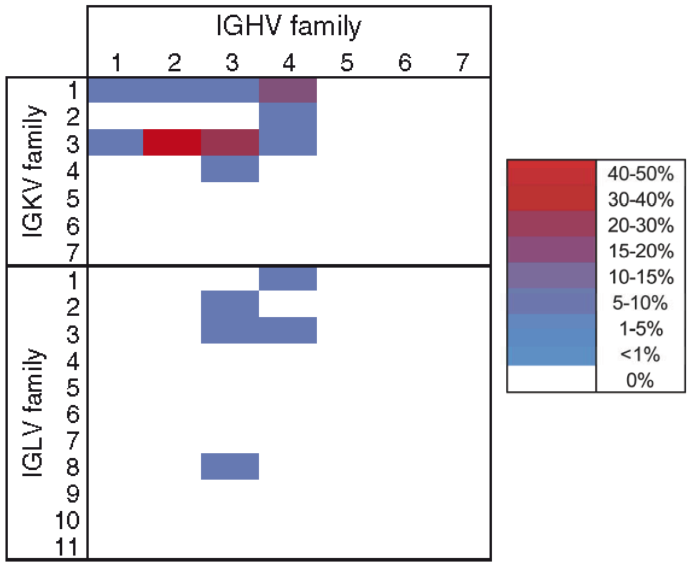
Evaluation of Methodology

IgG⁺ B cells from two healthy individuals

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

B cells were spiked with immortalized ARH-77



- 400 recovered B cells
- 86 unique pairs

⇩

expression of ten of the identified V_H:V_L pairs as IgG proteins in HEK293 cells

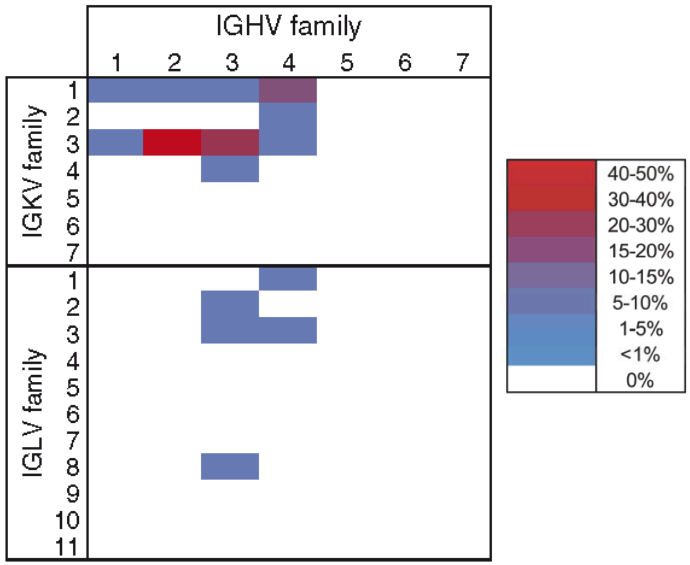
Evaluation of Methodology

IgG⁺ B cells from two healthy individuals

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

B cells were spiked with immortalized ARH-77



- 400 recovered B cells
- 86 unique pairs

expression of ten of the identified V_H:V_L pairs as IgG proteins in HEK293 cells

ELISA

Table 1 TT-binding affinities of IgG antibodies sequenced from TT⁺ peripheral plasmablasts

Antibody ID	Gene family assignment[a]	Affinity (K_D)
TT1	HV3-HD1-HJ6: KV3-KJ5	1.6 ± 0.1 nM
TT2	HV3-HD3-HJ4: LV3-LJ1	14 ± 3 nM
TT3	HV1-HD2-HJ4: KV3-KJ5	3.6 ± 1.8 nM
TT4	HV2-HD2-HJ4: KV1-KJ1	2.7 ± 0.3 nM
TT5	HV4-HD2-HJ6: KV2-KJ3	18 ± 4 nM
TT6	HV1-HD3-HJ4: KV1-KJ2	0.57 ± 0.03 nM
TT7	HV4-HD3-HJ4: KV1-KJ2	0.46 ± 0.01 nM
TT8	HV3-HD3-HJ4: LV8-LJ3	2.8 ± 0.3 nM
TT9	HV4-HD2-HJ4: KV1-KJ1	0.10 ± 0.01 nM
TT10	HV1-HD3-HJ5: KV3-KJ5	1.6 ± 0.1 nM

Evaluation of Methodology

IgG$^+$ B cells from two healthy individuals

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

B cells were spiked with immortalized IM-9

Evaluation of Methodology

IgG⁺ B cells from two healthy individuals

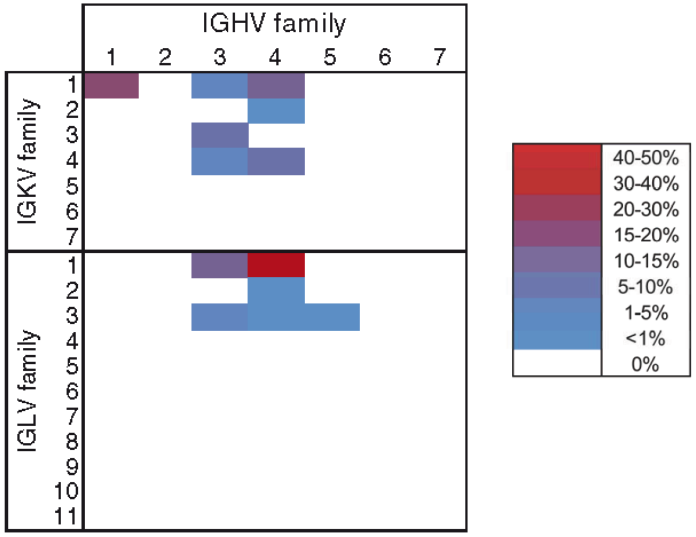
Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

B cells were spiked with immortalized IM-9

Identification of V_H:V_L pairs by high-throughput approach vs scRT-PCR:

Sanger scRT-PCR	High-throughput approach
168 single memory B cells	8,000 single memory B cells
168 RT- & 504 nested PCR reactions → Sanger Sequencing	Workflow as described → MiSeq
50 unique V_H:V_L pairs	240 unique V_H:V_L pairs



- 8000 recovered B cells
- 240 unique pairs

Evaluation of Methodology

IgG$^+$ B cells from two healthy individuals

Plasmablasts from a healthy individual 7d after tetanus toxin immunization

Memory B cells from a healthy individual 14d after influenza vaccination

B cells were spiked with immortalized IM-9

Identification of V_H:V_L pairs by high-throughput approach vs scRT-PCR:

Sanger scRT-PCR	High-throughput approach
168 single memory B cells	8,000 single memory B cells
168 RT- & 504 nested PCR reactions → Sanger Sequencing	Workflow as described → MiSeq
50 unique V_H:V_L pairs	240 unique V_H:V_L pairs

Seq ID	Isotype	CDR-H3	Paired CDR-L3[1]	Source
2D02	IgM	gcgagaggcggaaatgggcgacccttgacaac	gcagcatgggatgacagcctgaatggttgggtg	Sanger scRT-PCR
2D02	IgM	gcgagaggcggaaatgggcgacccttgacaac	gcagcatgggatgacagcctgaatggttgggtg	MiSeq VH:VL
3D05	IgM	gcgagaaggtactttgactac	gnagcatgggatgacagcctgaatgtttggntg	Sanger scRT-PCR
3D05	IgM	gcgagaaggtactttgactac	gcagcatgggatgacagcctgaatgtttggctg	MiSeq VH:VL
1E02	IgG1	gcgcgacatggccctgcgggaaaaagcgcgtatggttttgatatc	cagtcctatgacagcggactgaatggttatgtggtc	Sanger scRT-PCR
1E02	IgG	gcgcgacatggccctgcgggaaaaagcgcgtatggttttgatatc	cagtcctatgacaacagactgaatggttatgtggtg	MiSeq VH:VL
3A01	IgG3	gcgagagtaatagcagctcgcgaccgccggatcactcctaactactaccgccctatggacgtc	caggtgtgggatagtagtagtgaccatcaggtg	Sanger scRT-PCR
3A01	IgG	gcgagagtaatagcagctcgcgaccgccggatcactcctaattactaccgccctatggacgtc	caggtgtgggacagtagtagtgatcatcaggtg	MiSeq VH:VL

[1] The 2D02 and 3D05 CDR-L3 sequences are highly similar but differ by two bases

Conclusion Paper 1

Workflow for high-throughput sequencing of the paired heavy and light chain repertoire

- entire process can be completed by a single investigator in 10 working hours over 4days

  → identification of 2,716 unique V_H:V_L to a cost of $550

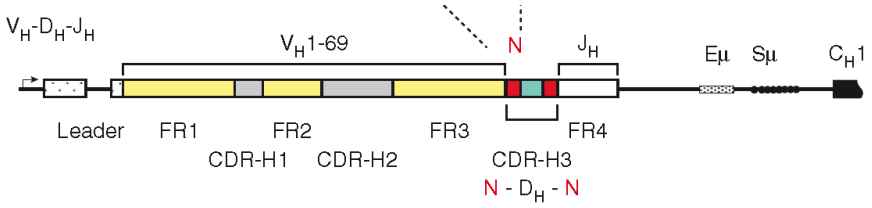
  vs. > $25,000 using scRT-PCR protocol

- identification of TT specific antibodies with high affinity

  → can be applied to investigate vaccine efficacy

- high CDR-H3:CDR-L3 pairing accuracy

  → longer sequencing reads are needed to distinguish somatic variants based on mutations between the FR1 and CDR2 region



Georgiou et al. 2014

- capacity > 5 x 10^4 single cells per experiment

  → greater depth still needed (10ml blood draw contains ~0.7 x 10^6 to 4 x10^6 B cells)

TECHNICAL REPORTS

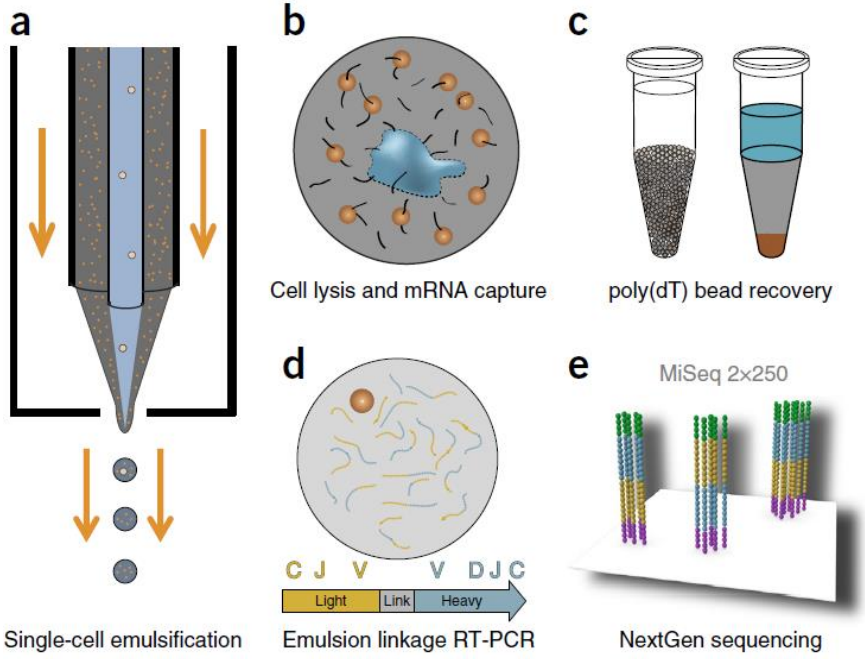
nature medicine

In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire

Brandon J DeKosky[1], Takaaki Kojima[1,2], Alexa Rodin[1], Wissam Charab[1], Gregory C Ippolito[3], Andrew D Ellington[4] & George Georgiou[1,3,5,6]

Experimental design

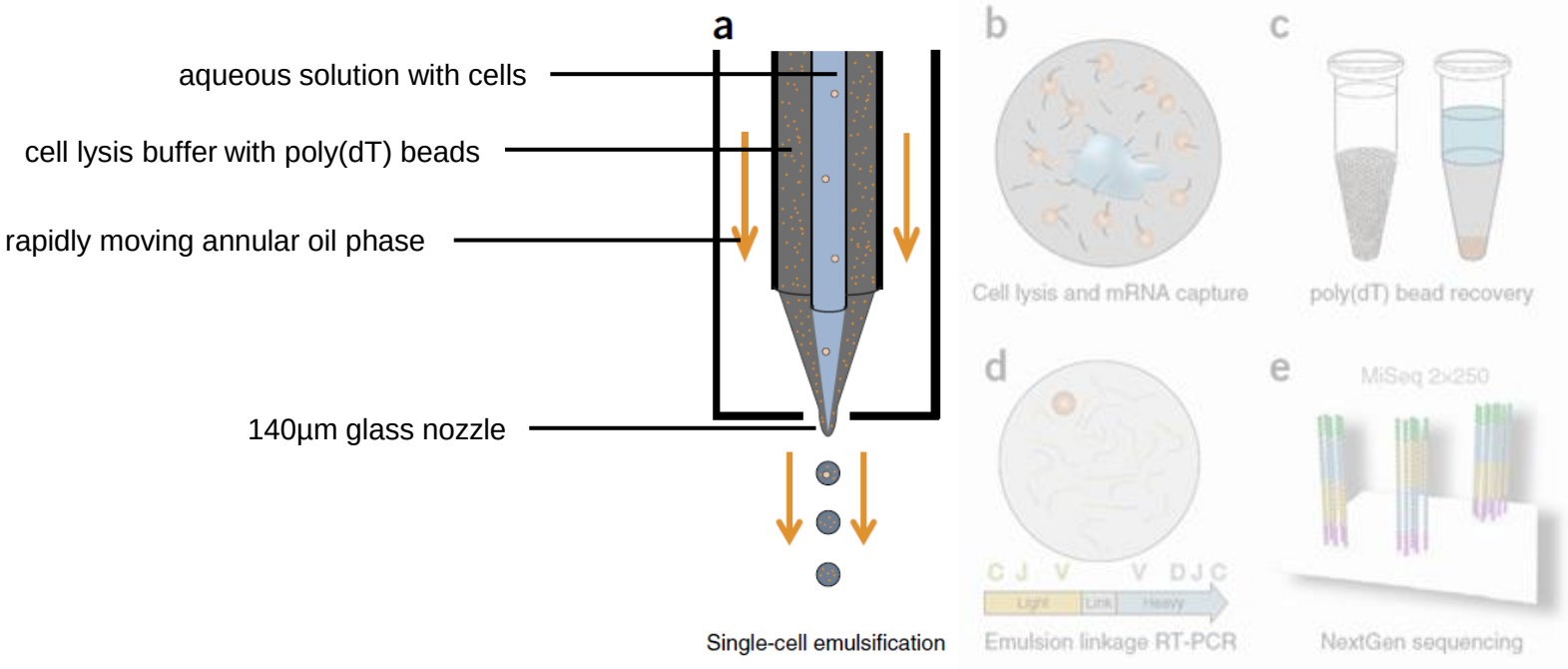
Greater depth by using an axissymmetric flow-focusing devices



DeKosky et al. 2015

Experimental design

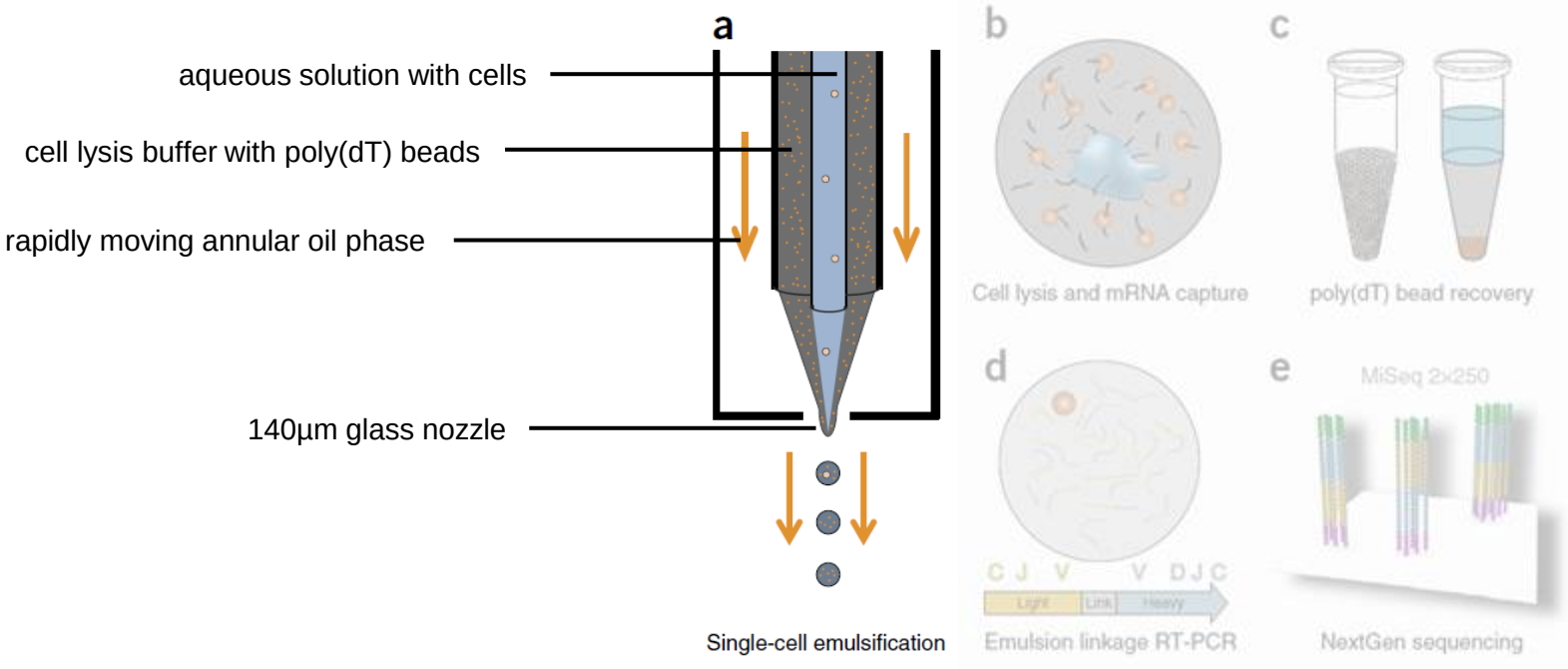
Greater depth by using an axissymmetric flow-focusing devices



DeKosky et al. 2015

Experimental design

Greater depth by using an axissymmetric flow-focusing devices



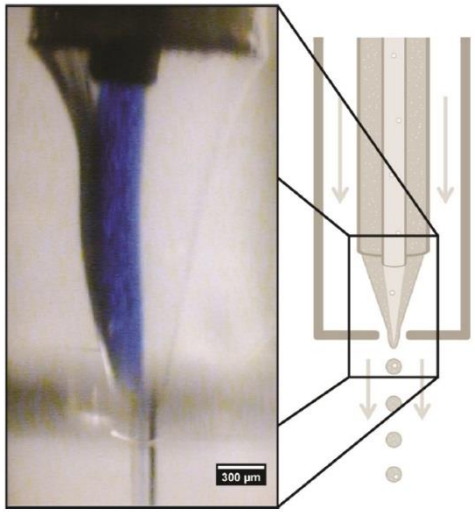
DeKosky et al. 2015

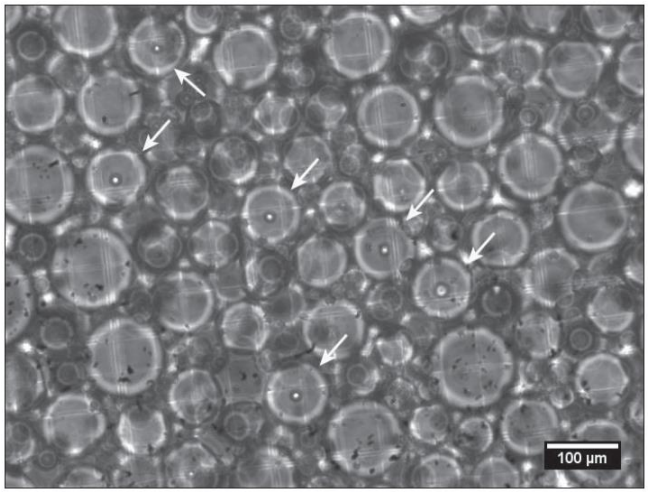
droplet formation with predictable size distribution

Evaluation of Methodology

Evaluation of encapsulation and droplet size distribution

- middle tubing containing PBS and Trypan blue (0,4% v/v)
- 250,000 cells per minute

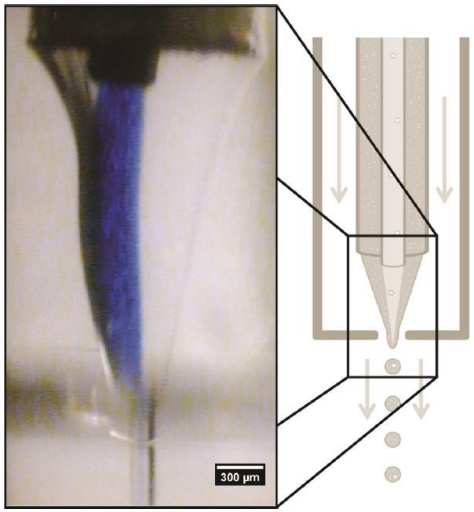


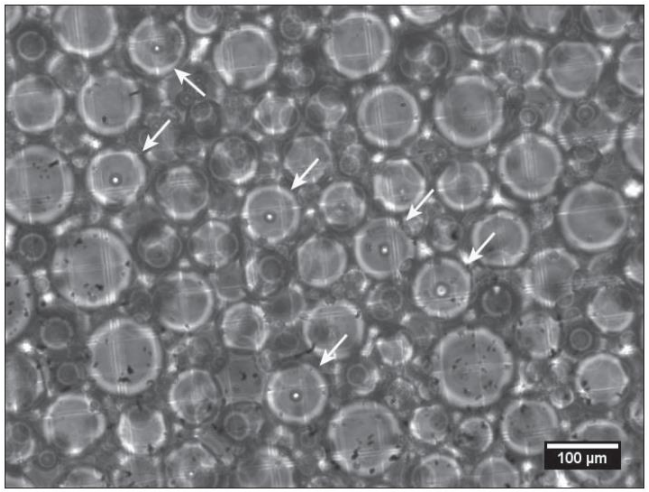


Evaluation of Methodology

Evaluation of encapsulation and droplet size distribution

- middle tubing containing PBS and Trypan blue (0,4% v/v)
- 250,000 cells per minute





- Trypan blue exclusion → cells remained viable
- droplet diameter 73 ± 20µm

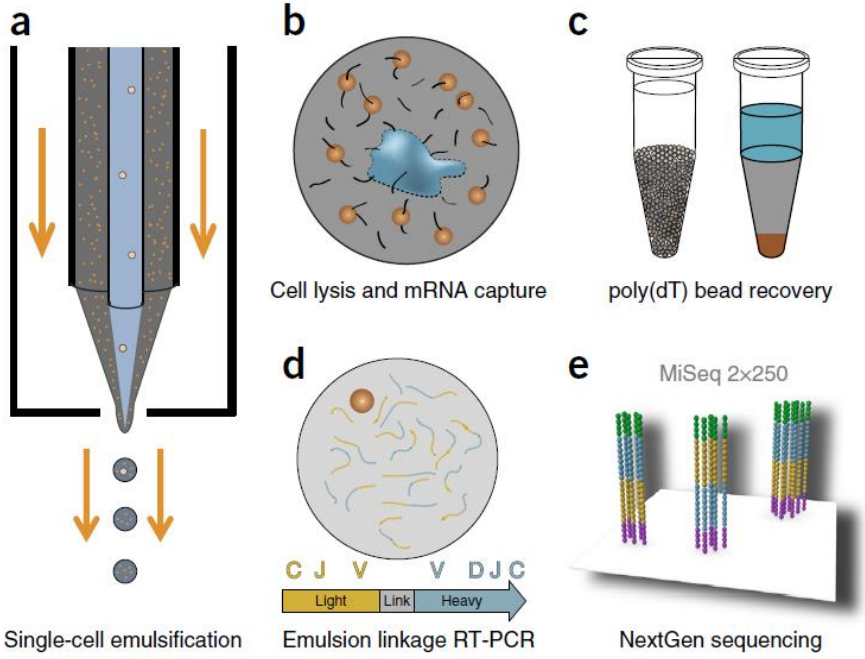
Experimental design

b. **Cell lysis & mRNA capture**
   - emulsion maintained for 3min

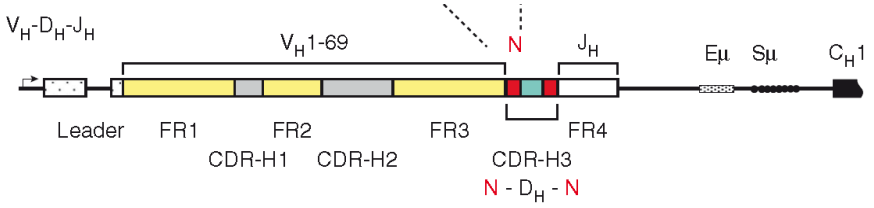
c. **poly(dT) bead recovery**
   - chemical breakage of emulsion

d. **Emulsion linkage RT-PCR**

e. **NGS using Illumina MiSeq 2 x 250pb**
   - sequencing of CDR-H3 and CDR-L3



DeKosky et al. 2015



Georgiou et al. 2014

Evaluation of Methodology

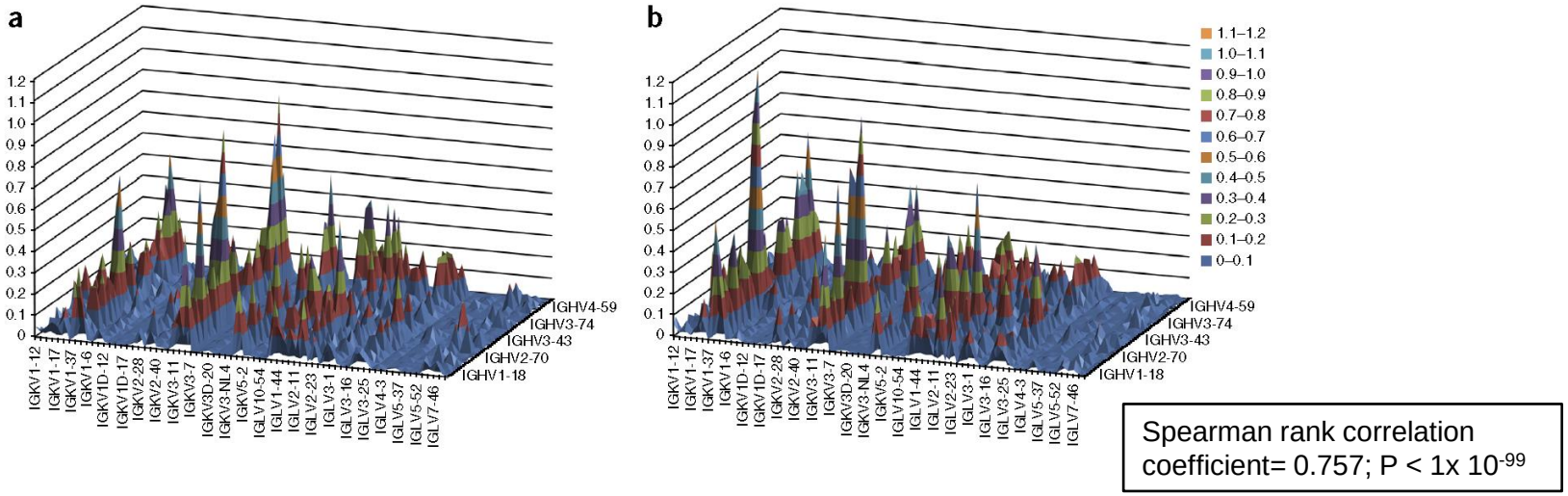
Memory B cells from healthy individuals in two technical replicates

B cells were expanded for 4 days

Evaluation of Methodology

Memory B cells from healthy individuals in two technical replicates

B cells were expanded for 4 days



Spearman rank correlation coefficient= 0.757; $P < 1x\ 10^{-99}$

Table 1 High-throughput VH-VL sequence analysis of CD3⁻CD19⁺CD20⁺CD27⁺ *in vitro*–expanded human B cells

Human donor	V-region primer set	No. cells analyzed	Emulsification rate (cells per minute)	Observed VH-VL clusters	CDR-H3 detected in both replicates	CDR-H3–CDR-L3 clusters detected in both replicates	VH-VL pairing precision
Donor 1	Framework 1	1,600,000	50,000	129,097	37,995	36,468	98.0%
Donor 2	Framework 1	810,000	50,000	53,679	19,096	18,115	97.4%
Donor 3	Leader peptide	210,000	33,000	15,372	4,267	4,170	98.9%

Evaluation of Methodology

Identification and characterization of promiscuous V_L junctions

- light chains have a much lower theoretical diversity than heavy chains

→ light chain sequences pair with multiple heavy chains

(promiscuous light chains)

Functional gene segments in human IgG loci

Segment	Light chains		Heavy chain
	κ	λ	H
Variable (V)	34-38	29-33	38-46
Diversity (D)	-	-	27
Joining (J)	5	4-5	6
Constant (C)	1	4-5	9

Evaluation of Methodology

Identification and characterization of promiscuous V_L junctions

- light chains have a much lower theoretical diversity than heavy chains

  → light chain sequences pair with multiple heavy chains

  (promiscuous light chains)

Functional gene segments in human IgG loci			
Segment	Light chains		Heavy chain
	κ	λ	H
Variable (V)	34-38	29-33	38-46
Diversity (D)	-	-	27
Joining (J)	5	4-5	6
Constant (C)	1	4-5	9

- Are the 50 most frequent promiscuous V_L junctions of donor 1 also present in donor 2 and 3?

  49/50 were present

  → promiscuous V_L are shared across individuals **("public" V_L)**

Evaluation of Methodology

Identification and characterization of promiscuous V_L junctions

- light chains have a much lower theoretical diversity than heavy chains

  → light chain sequences pair with multiple heavy chains

  (promiscuous light chains)

Functional gene segments in human IgG loci			
Segment	Light chains		Heavy chain
	κ	λ	H
Variable (V)	34-38	29-33	38-46
Diversity (D)	-	-	27
Joining (J)	5	4-5	6
Constant (C)	1	4-5	9

- Are the 50 most frequent promiscuous V_L junctions of donor 1 also present in donor 2 and 3?

  49/50 were present

  → promiscuous V_L are shared across individuals **("public" V_L)**

- Average of nontemplated bases in the VJ junction?

Evaluation of Methodology

Identification and characterization of promiscuous V_L junctions

- light chains have a much lower theoretical diversity than heavy chains

  → light chain sequences pair with multiple heavy chains

  (promiscuous light chains)

Functional gene segments in human IgG loci			
Segment	Light chains		Heavy chain
	κ	λ	H
Variable (V)	34-38	29-33	38-46
Diversity (D)	-	-	27
Joining (J)	5	4-5	6
Constant (C)	1	4-5	9

- Are the 50 most frequent promiscuous V_L junctions of donor 1 also present in donor 2 and 3?

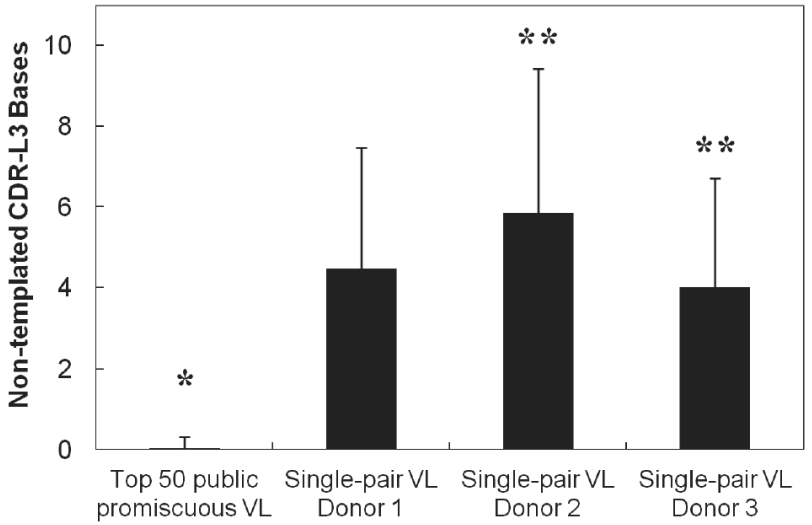
  49/50 were present

  → promiscuous V_L are shared across individuals **("public" V_L)**

- Average of nontemplated bases in the VJ junction?

  Average of 0.04 (promiscuous)

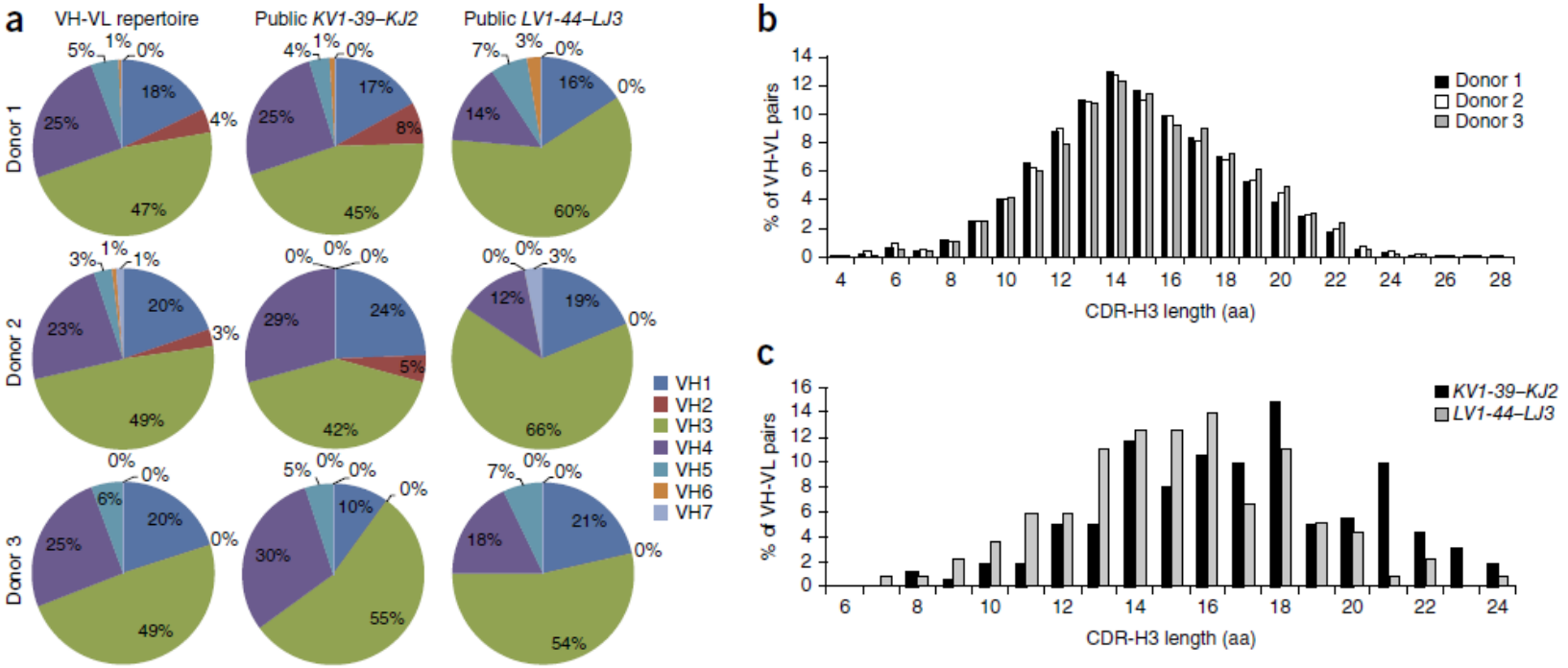
  vs. 5 (non-promiscuous)



Evaluation of Methodology

Identification and characterization of promiscuous V_L junctions

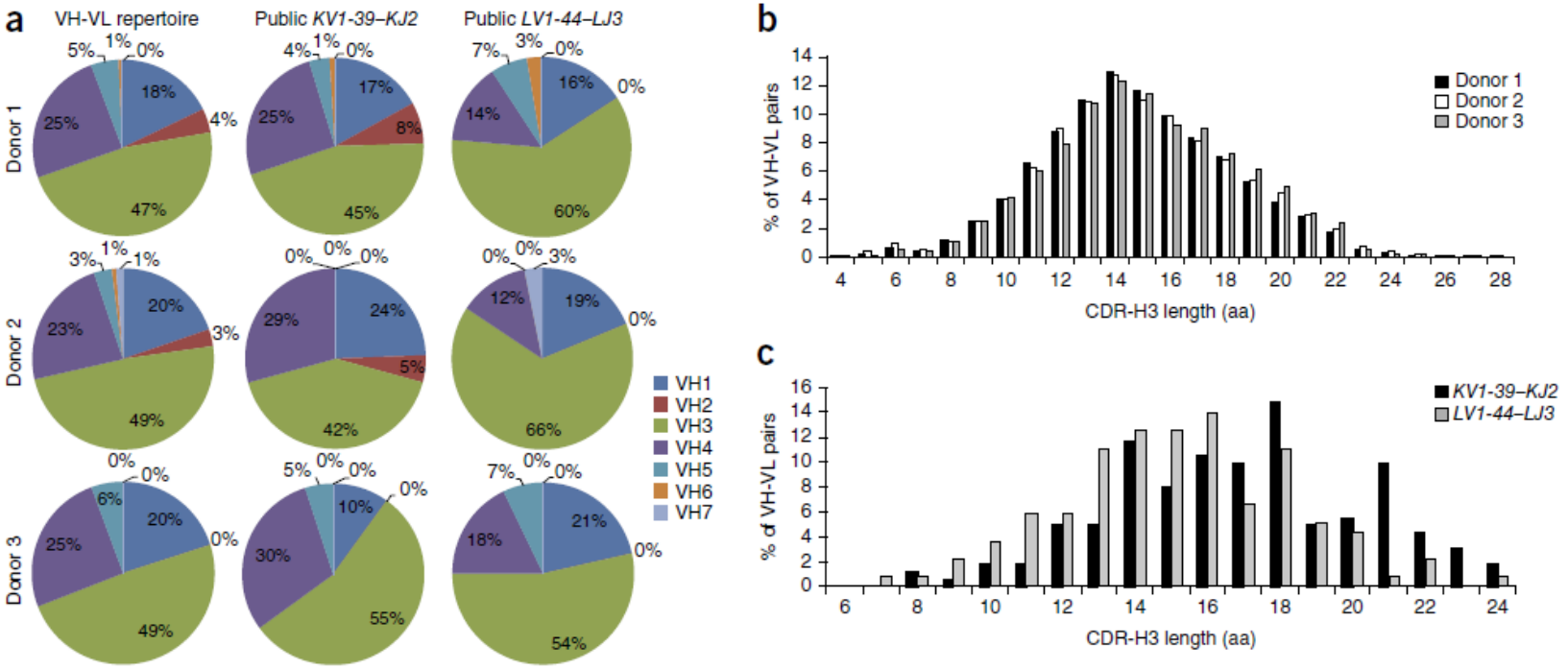
- ***KV1-39-KJ2*** (9aa CDR-L3) and ***LV1-44-LJ3*** (11aa CDR-L3) were examined in more detail (containing V and G genes with high prevalence in steady-state human immune repertoires)



Evaluation of Methodology

Identification and characterization of promiscuous V_L junctions

- ***KV1-39-KJ2*** (9aa CDR-L3) and ***LV1-44-LJ3*** (11aa CDR-L3) were examined in more detail (containing V and G genes with high prevalence in steady-state human immune repertoires)



→ V_L nucleotide promiscuity due to VL recombination rather than due to B cell activation and clonal expansion

Evaluation of Methodology

Quantification of allelic inclusion

- one B cell expressing two B cell receptors
- involved in autoimmunity
- detected in 0.2-0.5% of human memory B cells (Giachino et al. 1995)

Evaluation of Methodology

Quantification of allelic inclusion

- one B cell expressing two B cell receptors
- involved in autoimmunity
- detected in 0.2-0.5% of human memory B cells (Giachino et al. 1995)

➢ Can the event of allelic inclusion be detected?

– Identification of heavy chain clusters paired with multiple light chains which were encoded by distinct Vκλ and J κλ genes using an iterative loop

– cross-confirmation in both replicates

Evaluation of Methodology

Quantification of allelic inclusion

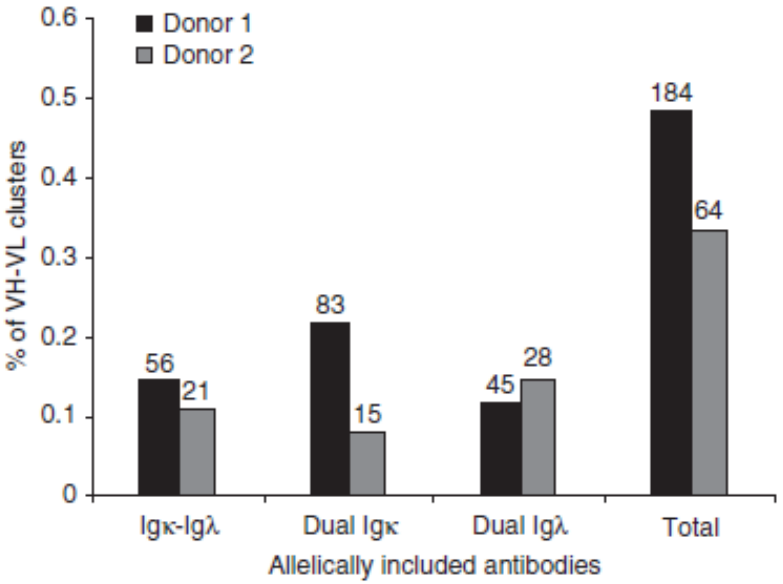
- one B cell expressing two B cell receptors
- involved in autoimmunity
- detected in 0.2-0.5% of human memory B cells (Giachino et al. 1995)

➢ Can the event of allelic inclusion be detected?

– Identification of heavy chain clusters paired with multiple light chains which were encoded by distinct Vκλ and J κλ genes using an iterative loop

– cross-confirmation in both replicates

detection in 0.4% of V_H clusters of donor 1 and 2



Evaluation of Methodology

Quantification of allelic inclusion

- one B cell expressing two B cell receptors
- involved in autoimmunity
- detected in 0.2-0.5% of human memory B cells (Giachino et al. 1995)

➢ Can the event of allelic inclusion be detected?

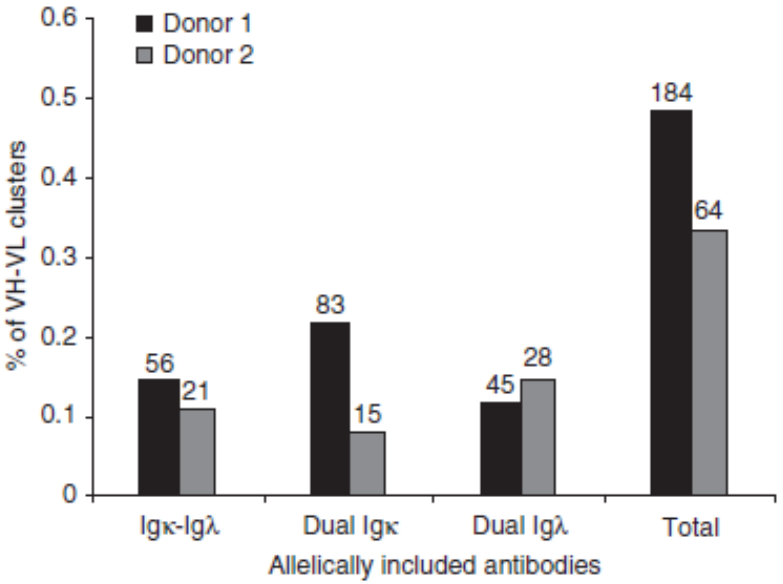
– Identification of heavy chain clusters paired with multiple light chains which were encoded by distinct Vκλ and J κλ genes using an iterative loop

– cross-confirmation in both replicates

detection in 0.4% of V_H clusters of donor 1 and 2

→ consistent with the study by Giachino

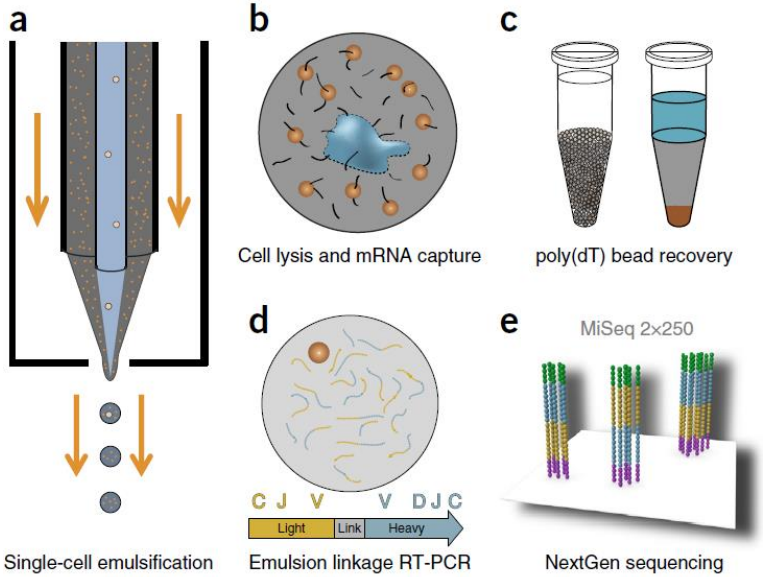
Allelic inclusion can be studied and quantified using this approach



Conclusion

High-throughput sequencing of the paired heavy and light chain repertoire

- great depth >2 x 10^6 per experiment (vs. capacity > 5 x 10^4 Paper 1)
- high accuracy of CDR-H3:CDR-L3 pairing
- fast and low costs
- can be used as tool to study allelic inclusion or vaccine efficacy



DeKosky et al. 2015

Thank you for your attention!!!