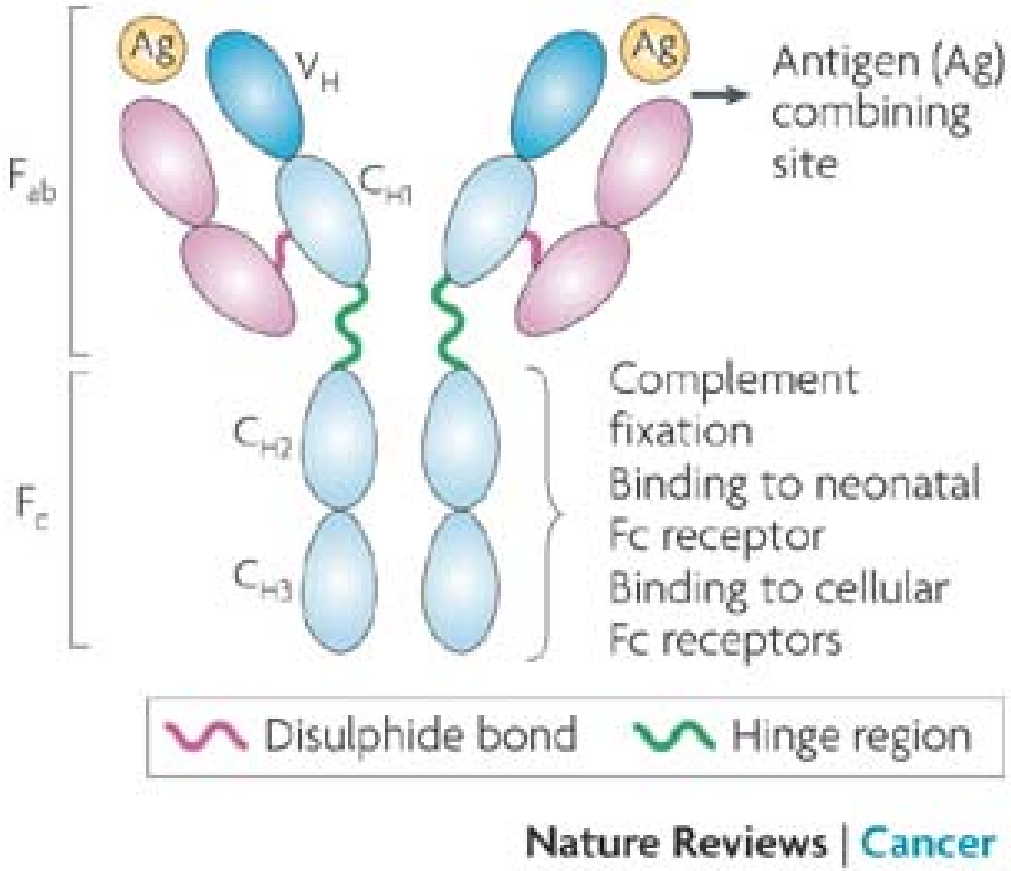
Next-generation sequencing and proteomics to study the antibody repertoire and generate monoclonal antibodies

Mario Nuvolone

Technical Journal Club

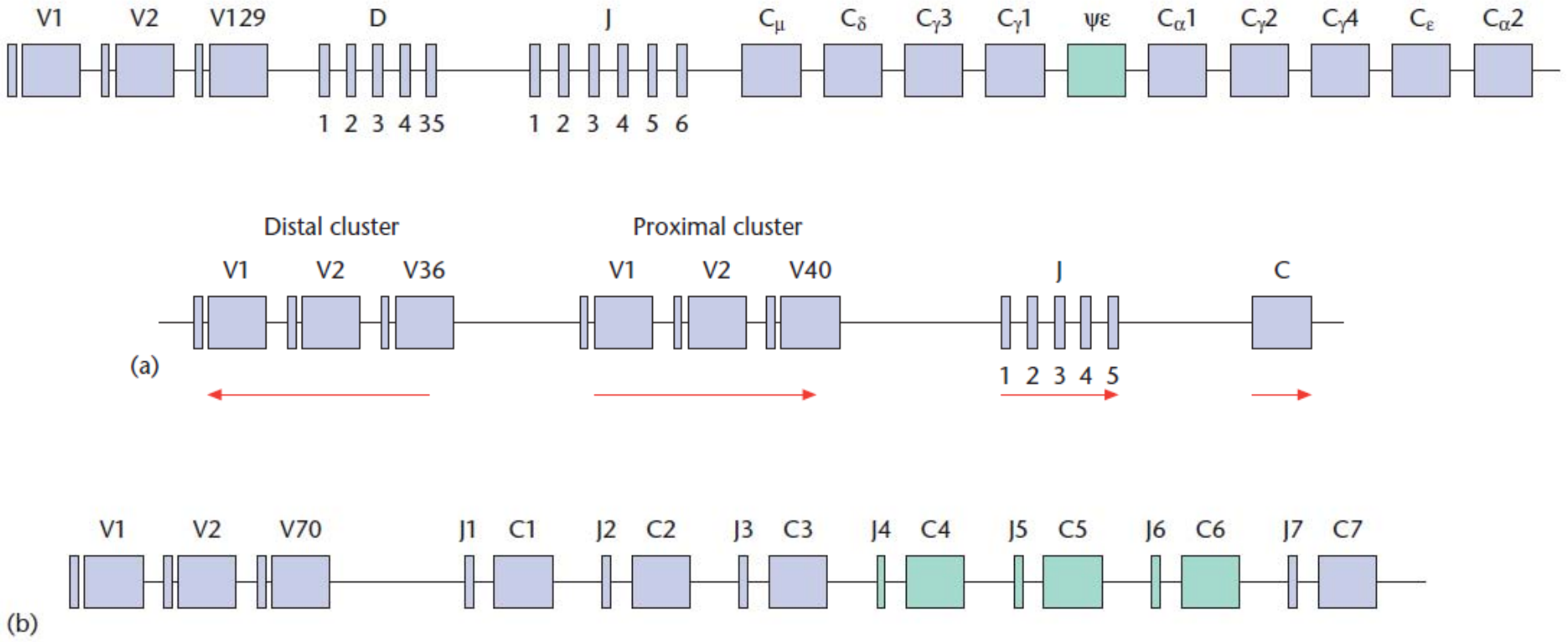
7th May 2013

Antibodies



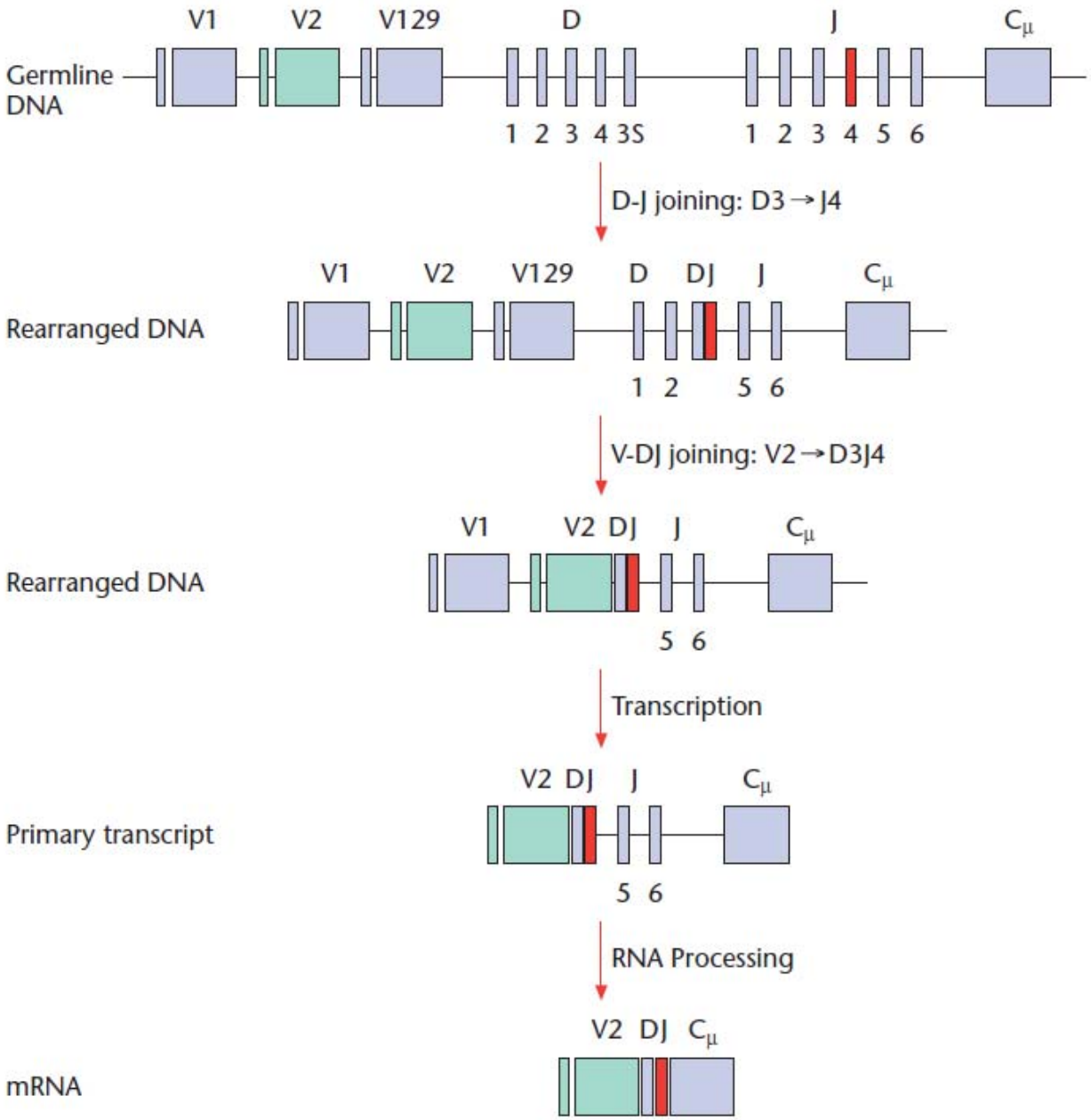
Weiner Nat Rev Cancer 2007

Genomic organization of human Ig genes



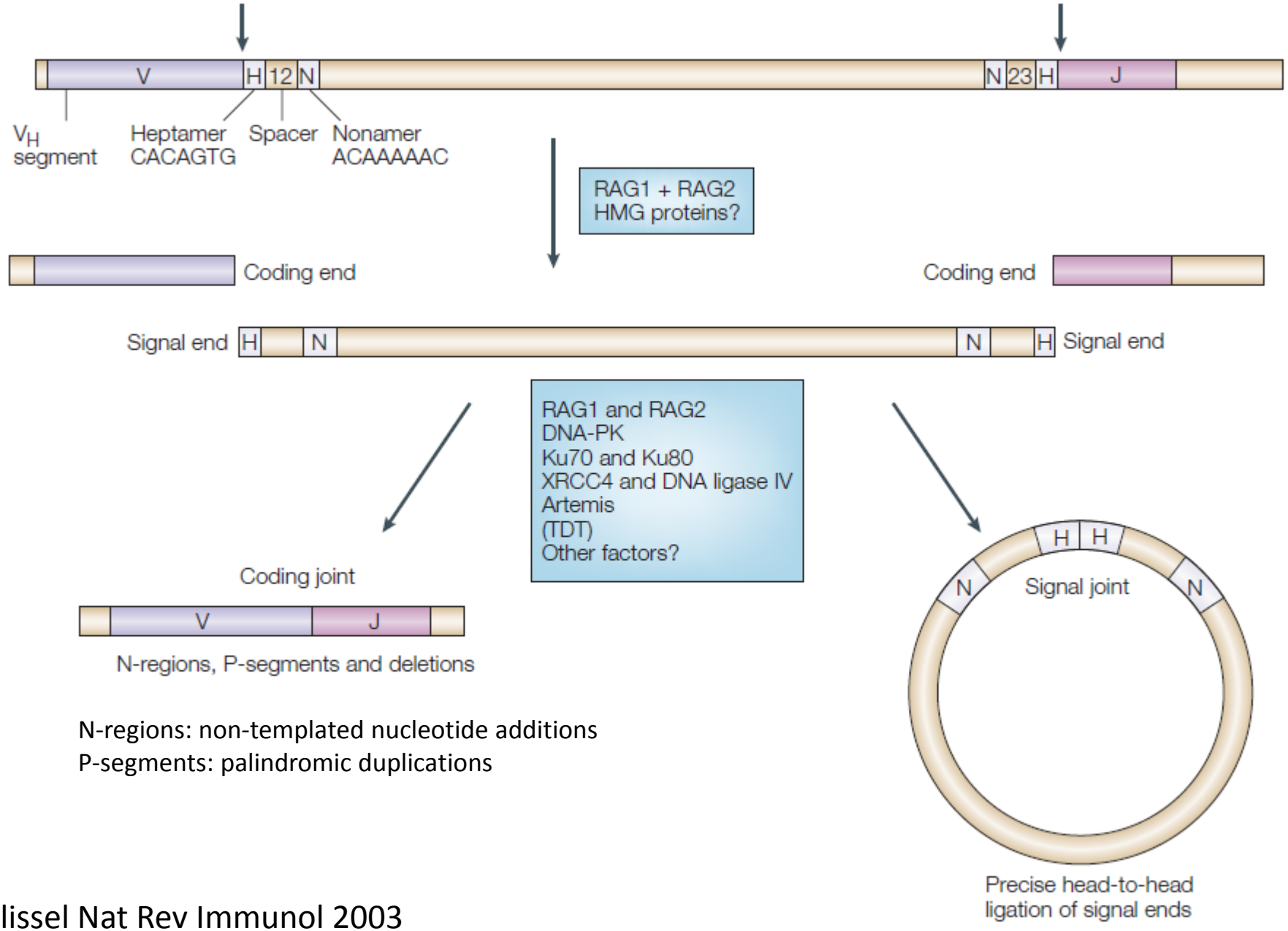
Lucas Encyclopedia of life sciences 2003

H chain production



Lucas Encyclopedia of life sciences 2003

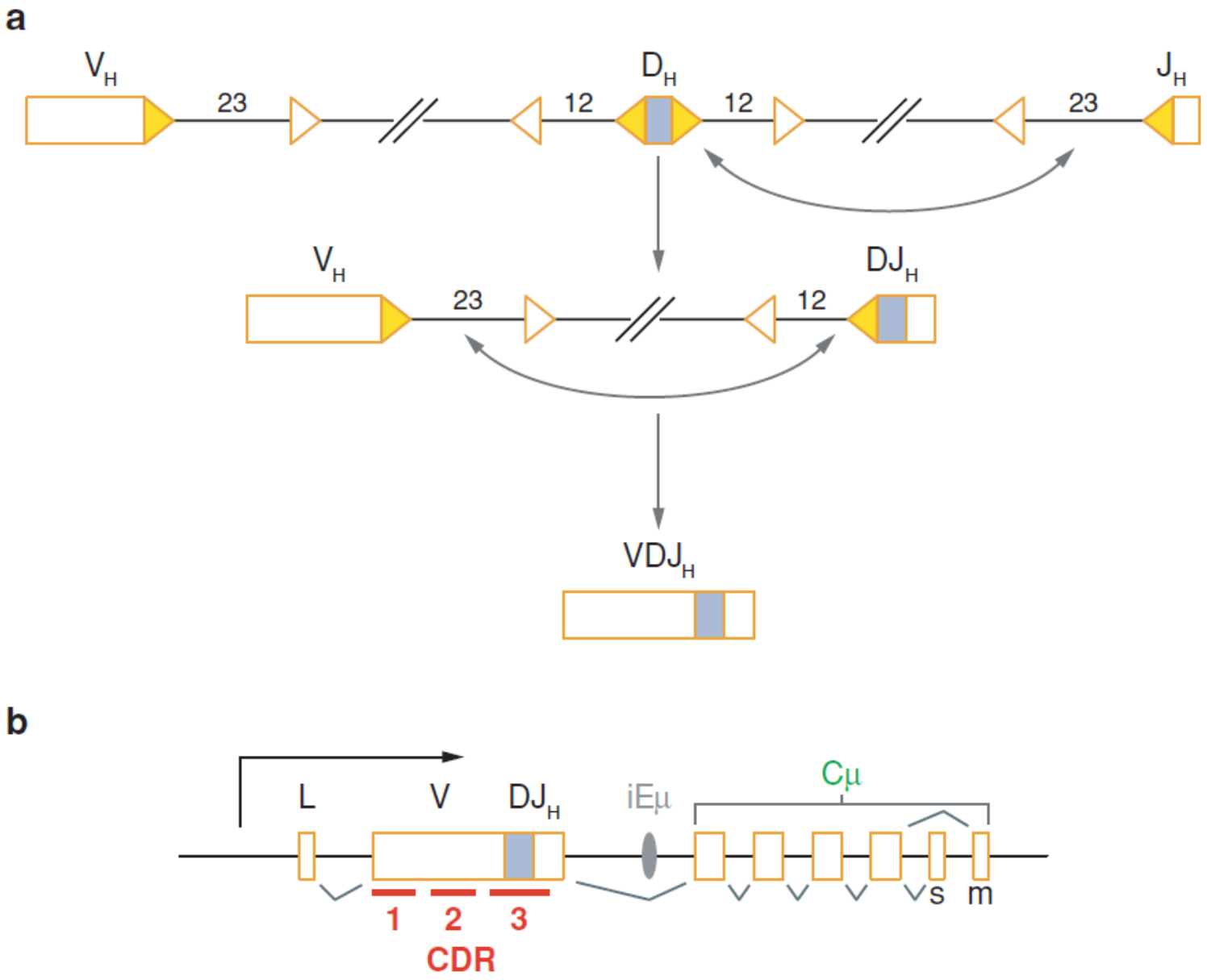
Mechanisms of V(D)J recombination



N-regions: non-templated nucleotide additions
P-segments: palindromic duplications

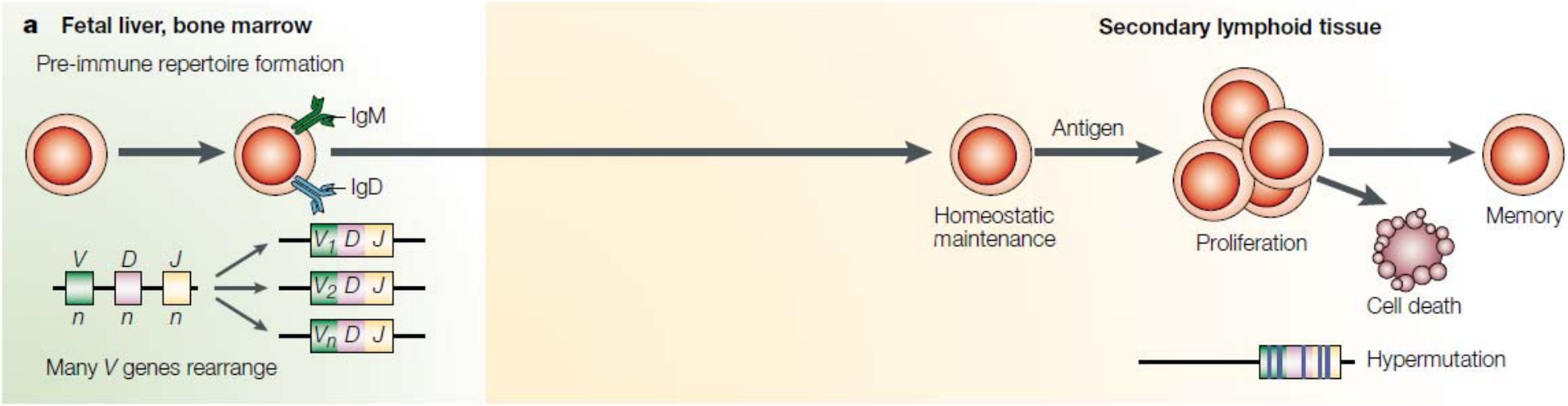
Schlissel Nat Rev Immunol 2003

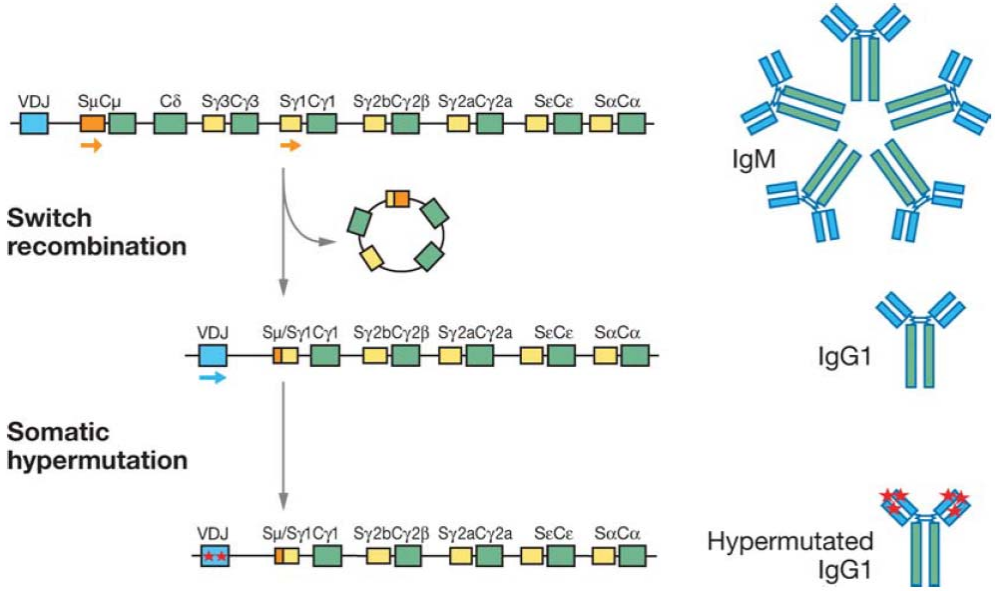
V(D)J recombination and CDR variation



Jung et al Ann Rev Immunol 2006

Somatic hypermutation

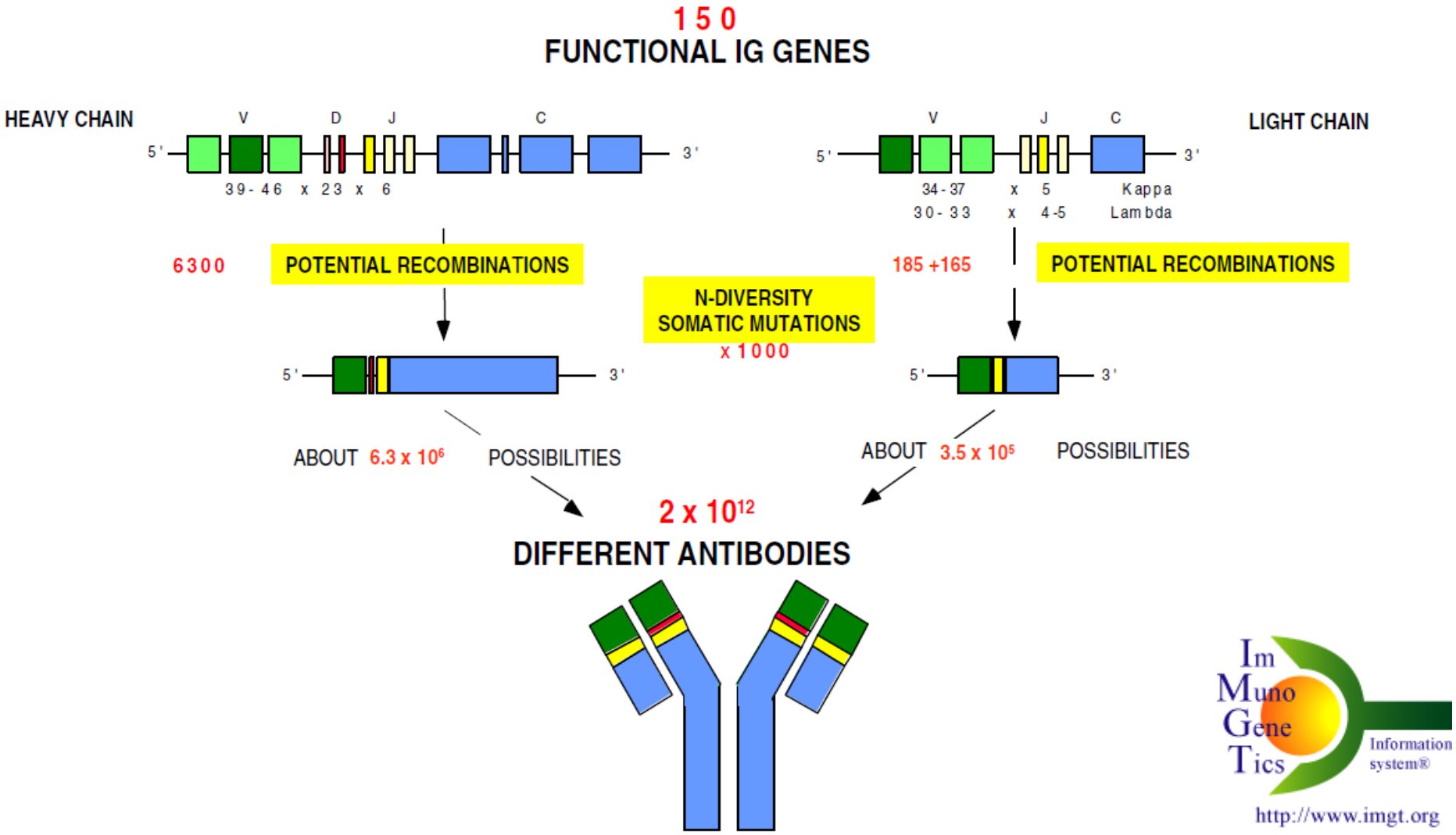




Flajnik Nat Rev Immunol 2002

Maizels Nat Rev Genet 2005

Antibody diversity



Im Muno Gene Tics Information system®

http://www.imgt.org

European Journal of Immunology

Wrestling with the repertoire: The promise and perils of next generation sequencing for antigen receptors

Paul D. Baum[1], Vanessa Venturi[2] and David A. Price[3]

Eur. J. Immunol. 2012. 42: 2834–2839

IMMUNOLOGY REVIEW ARTICLE

Rep-Seq: uncovering the immunological repertoire through next-generation sequencing

Jennifer Benichou,[1*] Rotem Ben-Hamo,[1*] Yoram Louzoun[2] and Sol Efroni[1]

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Systems analysis of adaptive immunity by utilization of high-throughput technologies

Sai T Reddy[1,2] and George Georgiou[1,2,3,4]

Current Opinion in Biotechnology 2011, 22:584–589

High-Throughput Sequencing of the Zebrafish Antibody Repertoire

Joshua A. Weinstein,[1*] Ning Jiang,[2*] Richard A. White III,[3]
Daniel S. Fisher,[1,4,5] Stephen R. Quake[1,2,3,4†]

SCIENCE VOL 324 8 MAY 2009

Scientific questions:

What fraction of the potential repertoire is expressed in an individual at any point in time?

How similar repertoires are among individuals who have lived in similar environments?

Zebrafish as a model organism to study antibody repertoires

- In evolutionary terms they have the earliest recognizable adaptive immune system whose features match the essential human elements

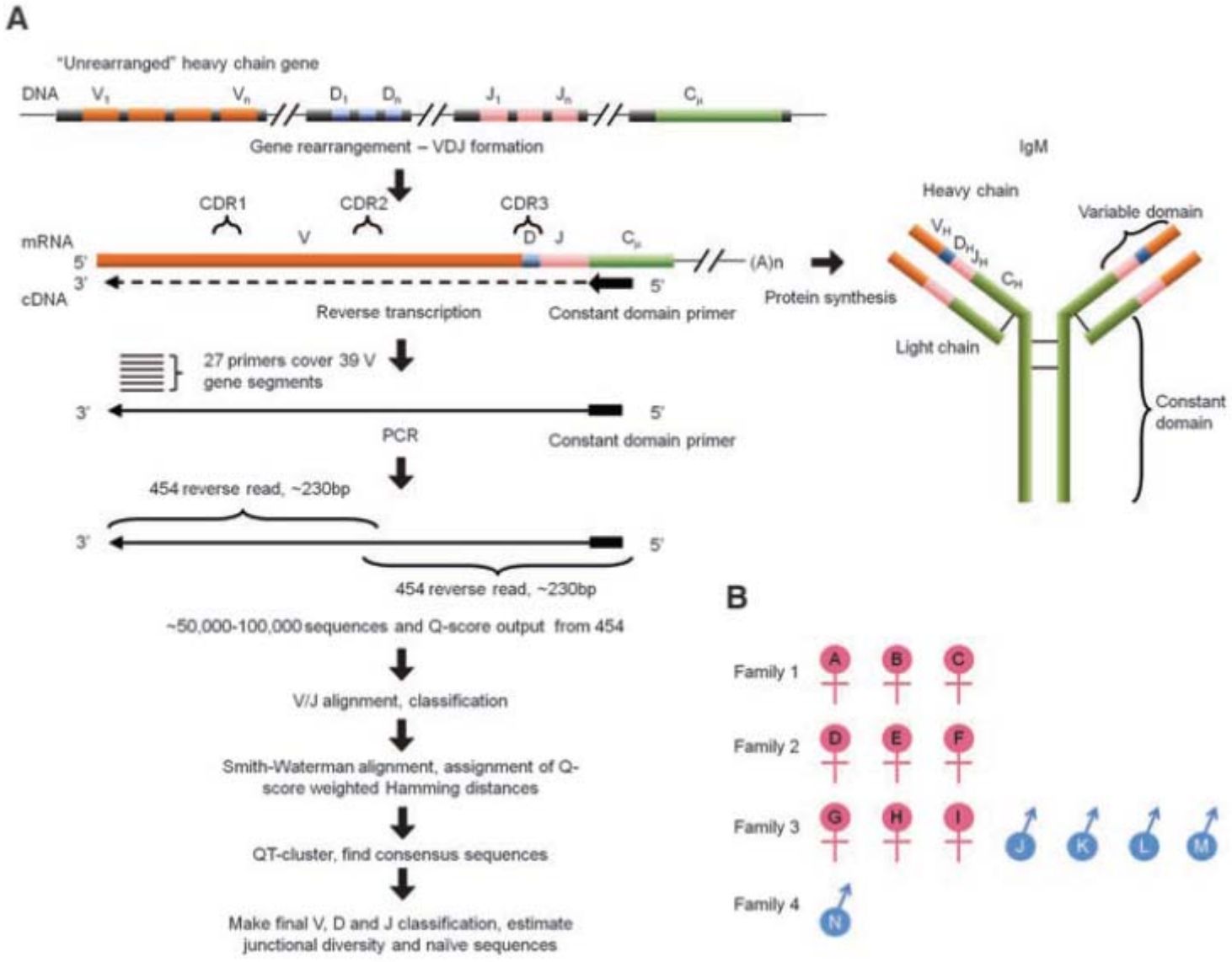
- They have a RAG and a combinatorial rearrangement of V, D, and J gene segments to create antibodies

- They have junctional diversity during recombination and somatic hypermutation of antibodies to improve specificity

- The organization of their Ig gene loci approximates that of human

- The zebrafish immune system has only ~300,000 antibody-producing B cells

Study overview



Main findings

- Zebrafish were found to use between 50 and 86% of all possible VDJ combinations and shared a similar frequency distribution, with some correlation of VDJ patterns between individuals

- Each individuals showed a few thousand unique heavy chains upon estimation

- Evidence of convergence, in which different individuals made the same antibody (CDR3 of IgM)

Paper	Model	Cell type	Receptor chain	Sequencing technology	PCR amplification	Main conclusions
Robins *et al.*[19]	Human	T cells, $CD4^+$/$CD8^+$, naive and antigen-experienced	β chain/CDR3	Illumina	Multiple primers	Nucleotide insertion bias towards G and C. High-frequency sequences have fewer nucleotide insertions and so are closer to germline sequences. Estimate of naive repertoire size is threefold to fourfold greater than previously estimated (10^6) and the antigen-experienced repertoire is 10-fold higher than $\sim 10^5$.
Weinstein *et al.*[33]	Zebrafish	B cells, IgM	Heavy chain, CDR3	454	Multiple primers	Analysed V–D–J usage. Zebrafish use 50–86% of possible VDJ combinations, individuals shared similar frequency distribution, evidences of convergence – same antibody in many fish. Repertoire size estimate – 5000–6000 unique antibodies (IgM and heavy chain only).
Freeman *et al.*[25]	Human	T cells, $CD4^+$ $CD8^+$	β chain, CDR3	Illumina	5′-RACE	An evidence of convergence – the same amino acid was translated from multiple nucleotide sequences. V gene usage ranged between 0·01 and 24·6% and J usage between 1·6 and 17·2%. CDR3 lengths ranged from 21 to 81 nucleotides.
Boyd *et al.*[21]	Human	B cells	Heavy chain	454	Multiple primers	Characterize IgH repertoire in healthy patients and in malignancies. V and J usage in healthy populations showed a diverse use, whereas samples that contained clonal IgH populations corresponding to lymphomas or chronic lymphocytic leukaemia specimens were readily identified. Showed that post-transplant lymphoproliferative disorder can develop multiple independent malignant clones.
Glanville *et al.*[17]	Human	B cells, IgM	Heavy/light chains, CDR1, CDR2, CDR3	454	Multiple primers	All germline gene families were represented within the sequences. In addition to CDR3, CDR1 and CDR3 also substantially contribute to the sequence variability caused by somatic hypermutations. They determined lower bounds of $2{\cdot}2 \times 10^5$ unique heavy chain sequences and $1{\cdot}6 \times 10^5$ unique light chains.
Robins *et al.*[18]	Human	T cells, $CD8^+$, naive and memory	β chain	Illumina	Multiple primers	CDR3 repertoire strongly biased towards specific V–J usage, most have few nucleotide insertions, only 0·1% of total estimated number of sequences. Overlap between two individuals is $\sim$ 7000 higher than expected.
Wang *et al.*[20]	Human	T cells, $CD4^+$ $CD8^+$, naive and memory	β/α chains, CDR3	454	Multiple primers	Observed 84% of possible Vα–Jα and Vβ–Jβ combinations in data. No significant difference was observed among T-cell subsets (V, J usage, CDR3 length, nucleotide additions, amino acid frequency). Estimated $0{\cdot}47 \times 10^6$ and $0{\cdot}35 \times 10^6$ unique TCR-α and TCR-β nucleotide sequences. Among the clonally expanded T cells, the majority of regulatory T cells and T helper type 1 and type 2 cells are related to memory T cells, whereas the majority of cytotoxic T cells are related to naive T cells.
Ben-Hamo and Efroni[41]	Zebrafish	B cells, IgM	Heavy chain	454	Multiple primers	Constructed sequences and mutations networks out of zebrafish IgM heavy chains. Fish population was divided into two groups based on V–J usage. The first group showed uni form usage, whereas the other revealed distinct subsets of sequences, suggesting that the latter had undergone a major antigen challenge.
Venturi *et al.*[10]	Human	T cells, naive and memory	β chain, CDR3	454	Multiple primers	Analysed TRBV12-4/TRBJ1-2 gene recombination only. High frequency clonotypes are shared among individuals. Convergent recombination shapes the TCR repertoire of the memory and naive T-cell pools, and also between individuals.
Warren *et al.*[14]	Human	T cells, naive and memory	β chain	Illumina	5′-RACE	Determined the sequencing error rate empirically and filtered the data. Multiple individuals share highly similar V and J gene usage frequencies, but only 1·1% of nucleotide sequences are shared. At the amino acid level, more sequences were shared. Shared amino acid sequences were encoded by a large diversity of nucleotide sequences.

Benichou et al Immunology 2011

nature biotechnology

LETTERS

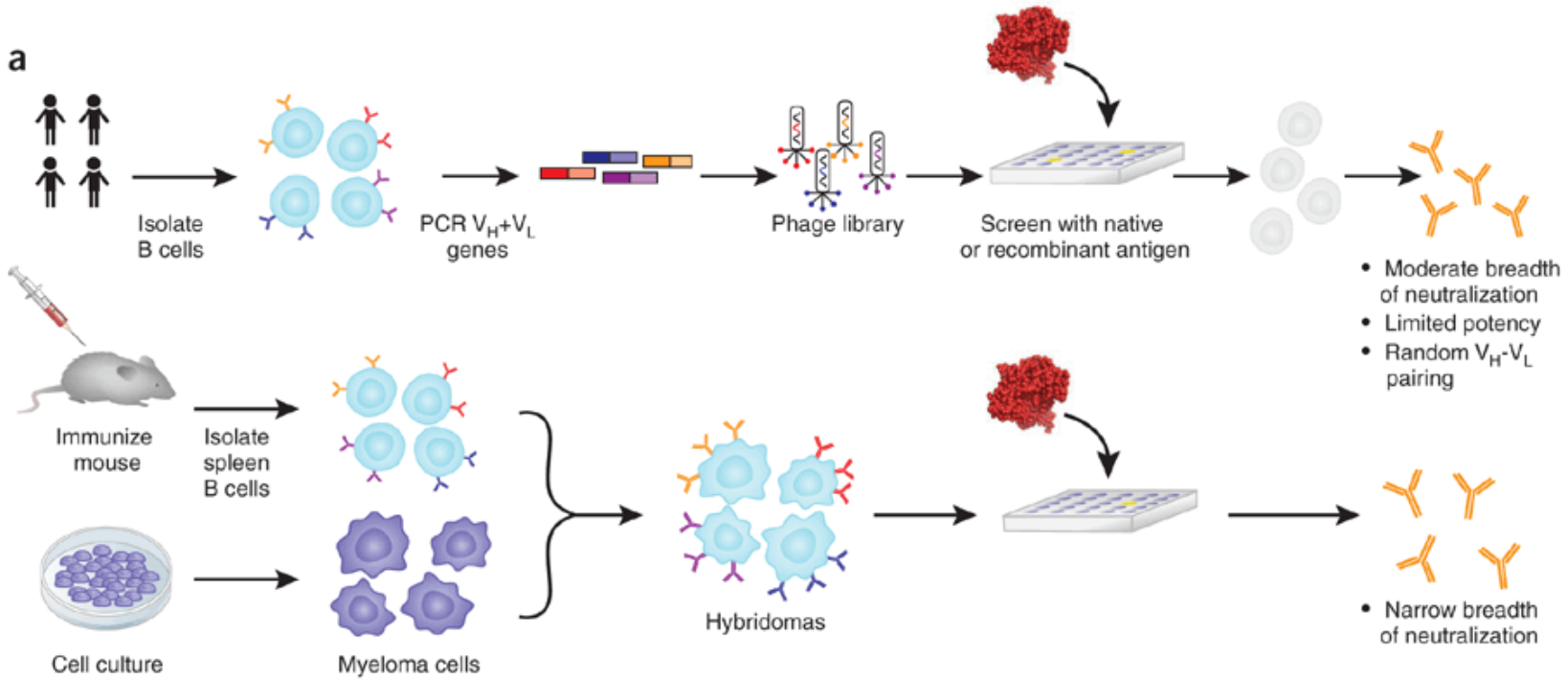
Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells

Sai T Reddy[1,2], Xin Ge[1], Aleksandr E Miklos[3,4], Randall A Hughes[3,4], Seung Hyun Kang[1], Kam Hon Hoi[2], Constantine Chrysostomou[1], Scott P Hunicke-Smith[3], Brent L Iverson[3,5], Philip W Tucker[3,6], Andrew D Ellington[3–5] & George Georgiou[1–3,6]

Scientific background

- Isolation of antigen-specific monoclonal antibodies (mAbs) and antibody fragments relies on screening of immortalized B cells, cloning of V genes by single-cell or approaches for *in vitro* discovery that involve the display and screening of recombinant antibody libraries

- Both *in vitro* and *in vivo* methods for antibody discovery are critically dependent on high-throughput screening to determine antigen specificity.

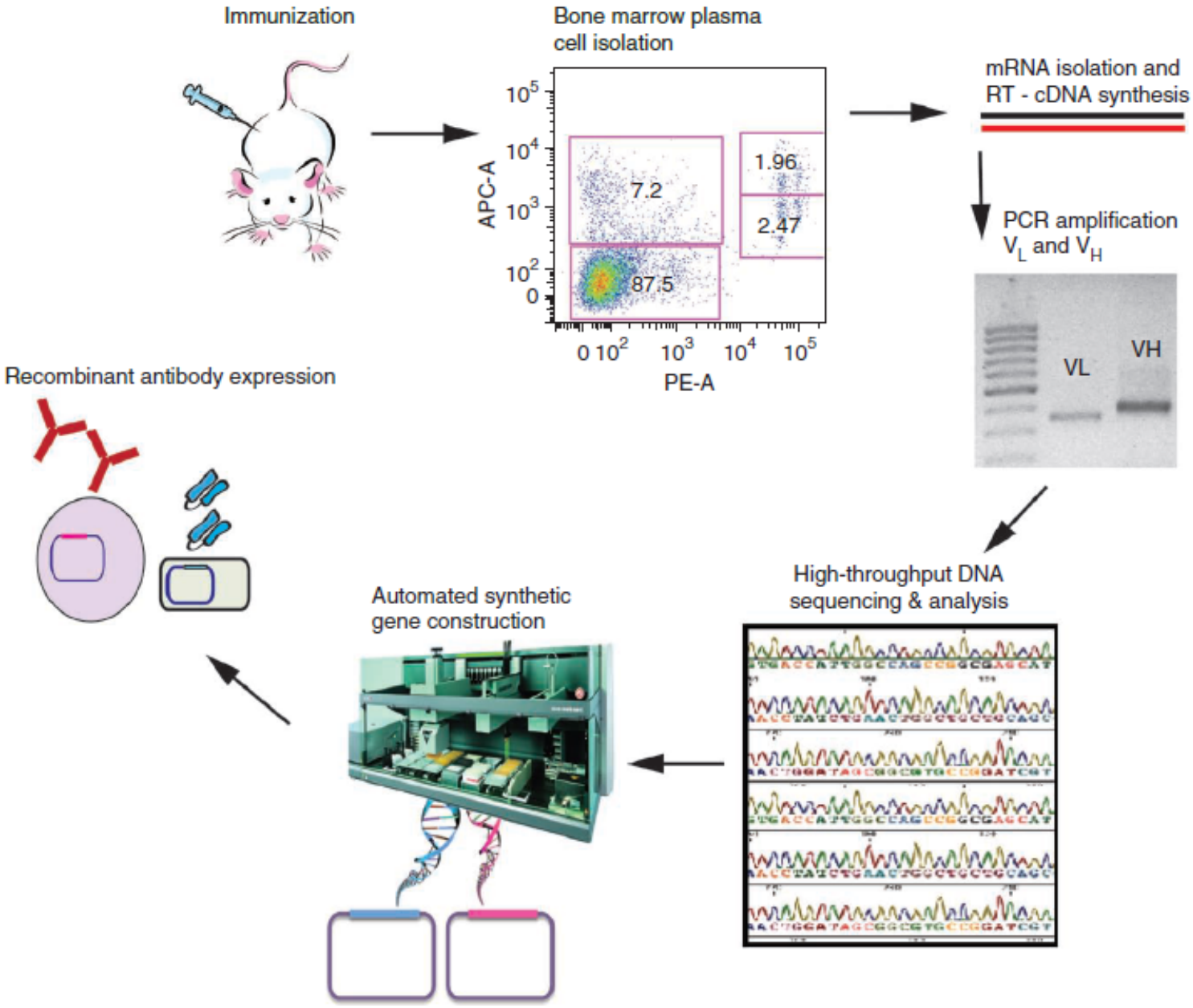


Joyce & ter Meulen Nat Biochen 2010

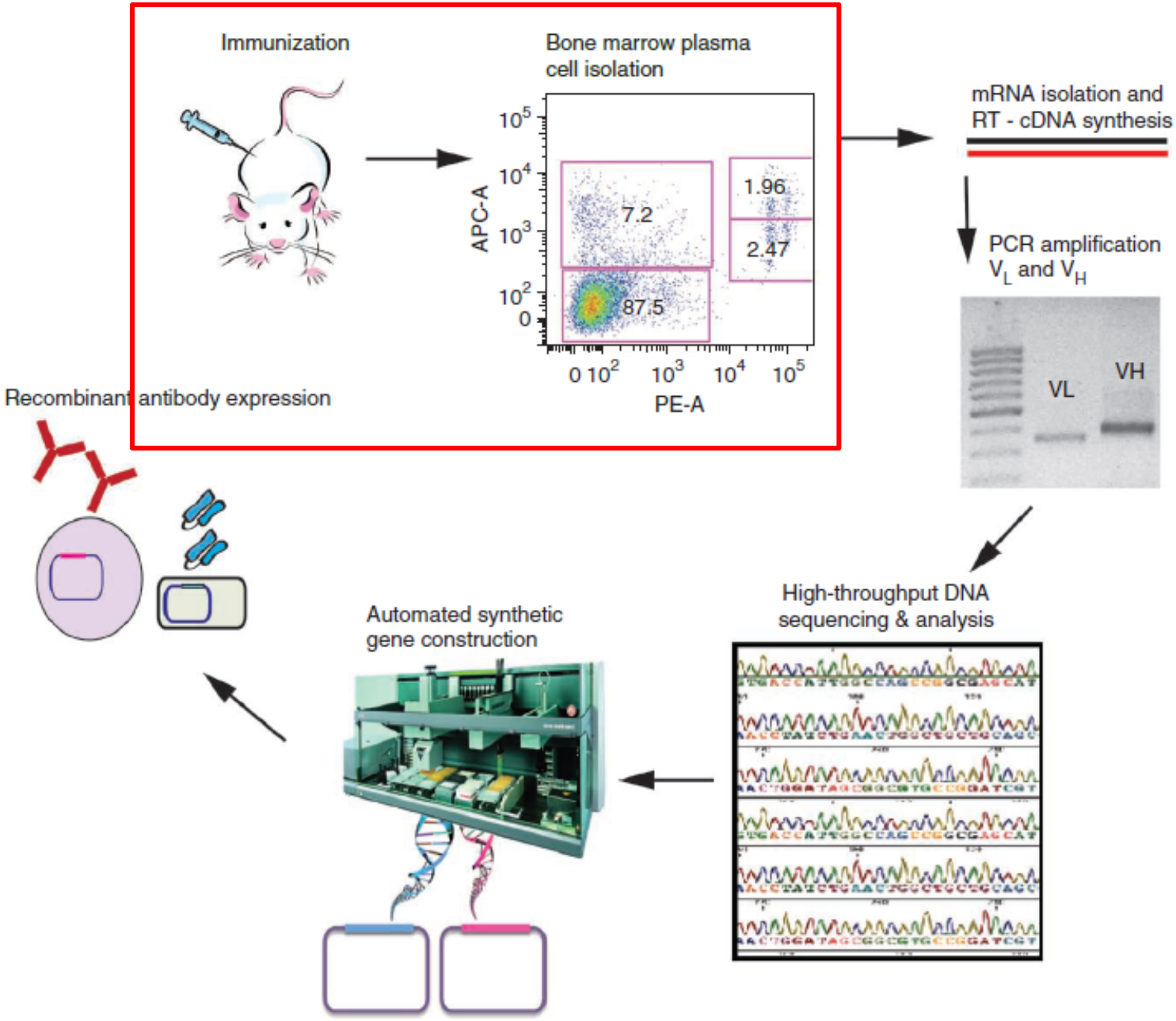
Aim of the study

To develop a simple and rapid method for Ab isolation without the need for screening, by exploiting high-throughput sequencing on antibody secreting cells from immunized mice

Study overview



Immunization and BMPCs isolation



Immunization and BMPCs isolation

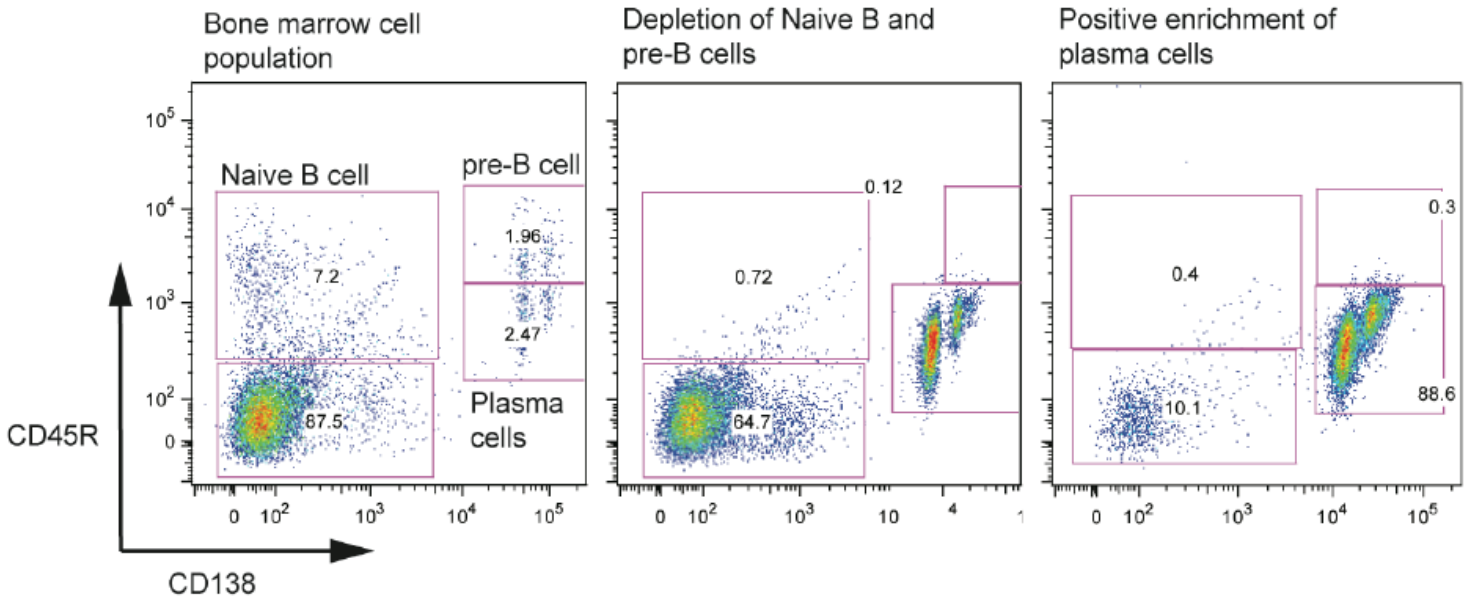
Pairs of mice were immunized with:

1) chicken egg ovalbumin + complete Freund's adjuvant (CFA)
2) human complement serine protease (C1s) + CFA
3) human B-cell regulator of IgH transcription (Bright) + CFA
4) CFA only.

(secondary booster immunization in incomplete Freund's adjuvant)

Mice were euthanized 6 d after secondary immunization

BMPCs were isolated by FACS



Why BMPCs?

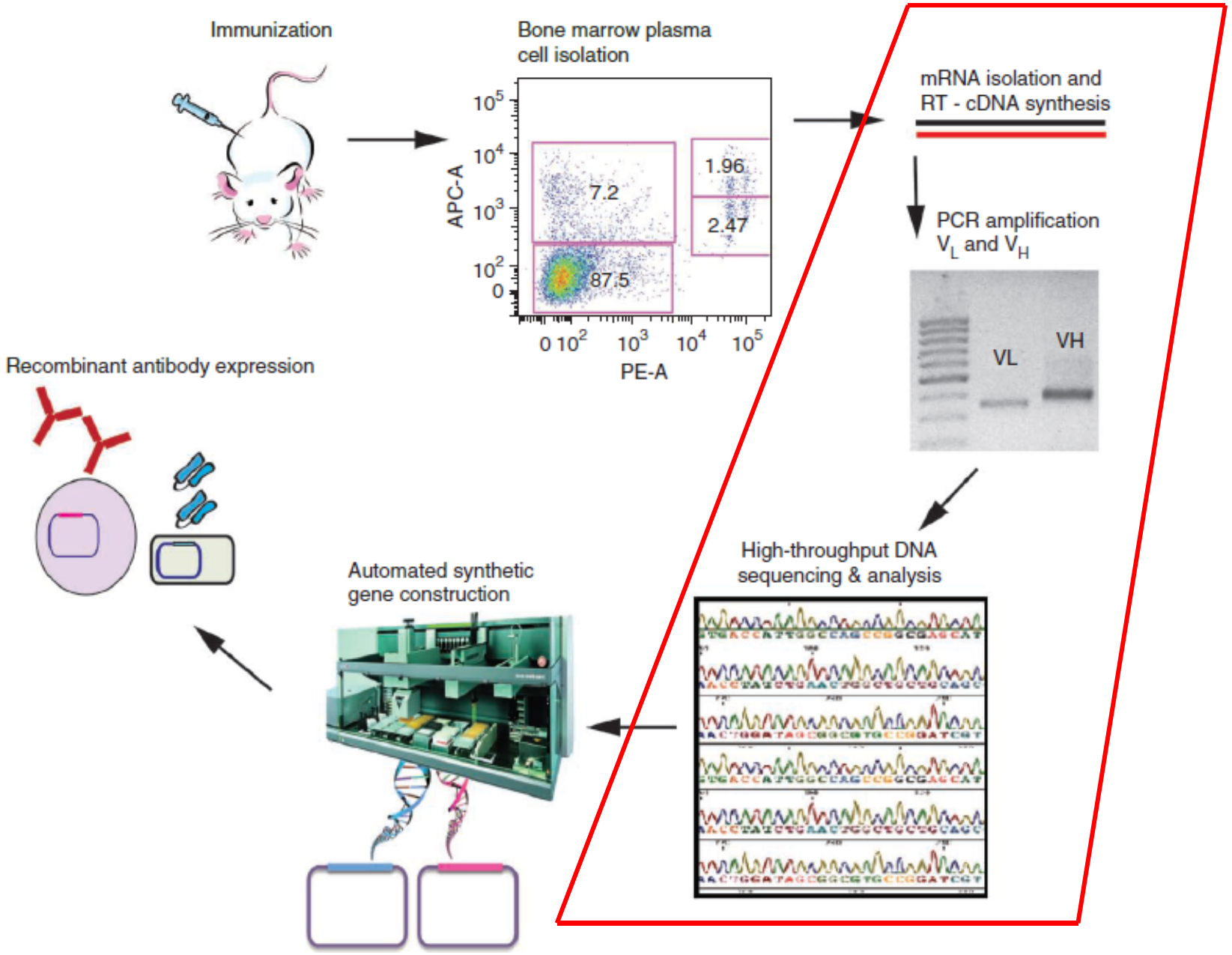
- The bone marrow constitutes the major compartment where plasma cells reside and produce antibodies for prolonged periods of time, whereas plasma cells present in secondary lymphoid organs are often short lived

- In mice, a stable and highly enriched antigen-specific BMPC population of ~10^5 cells (10–20% of all BMPCs) appears 6 d after secondary immunization and persists for prolonged periods

- In contrast, the increase in size of the splenic plasma cell population is highly transient, peaking at day 6 and rapidly declining to <10^4 cells by day 11

-BMPCs are responsible for making the stable circulating population of antibodies in serum, which in turn is likely to play a dominant role in pathogen neutralization and other protective humoral immune responses

Next-generation sequencing



Next-generation sequencing

- Total RNA was extracted and reverse transcribed for synthesis of first-strand cDNA

- Degenerate V-gene primer mixes were used for second-strand amplifications, resulting in VL and VH PCR products of high purity

- VL and VH amplicons were then submitted for high-throughput sequencing of long reads using the Roche 454-GS FLX technology

- No requirement for exhaustive coverage of the V gene repertoire: obtaining ~5,000 V-gene sequences per BMPC sample is sufficient to provide the information needed for antibody discovery

Frequency ranking of V genes

- For each mouse, frequency distributions of the CDR3s were calculated

- Sequencing of the same samples, from separate cDNA library preparations by different facilities, gave quantitatively similar rankings for the abundances of CDR3 sequences

- Frequency ranking is crucial for heavy/light chain pairing

Results: variability of the repertoire

- Of a total of 415,018 reads, 23.2% contained CDR3 of VH (CDRH3) and 26.6% contained CDR3 of VL (CDRL3) sequences, representing 6,681–16,743 and 7,112–21,241 CDRH3 and CDRL3 sequences reads per mouse, respectively

- In all immunized mice, including those receiving the same antigen, >92% of the CDRH3 sequences were unique to an individual mouse

- The CDRL3 repertoires were less diverse, and in some instances, BMPCs from mice immunized with different antigens expressed high levels of the same CDRL3

- ~10–20% of the total repertoire of all immunized mice were on average composed of only four CDRH3 sequences

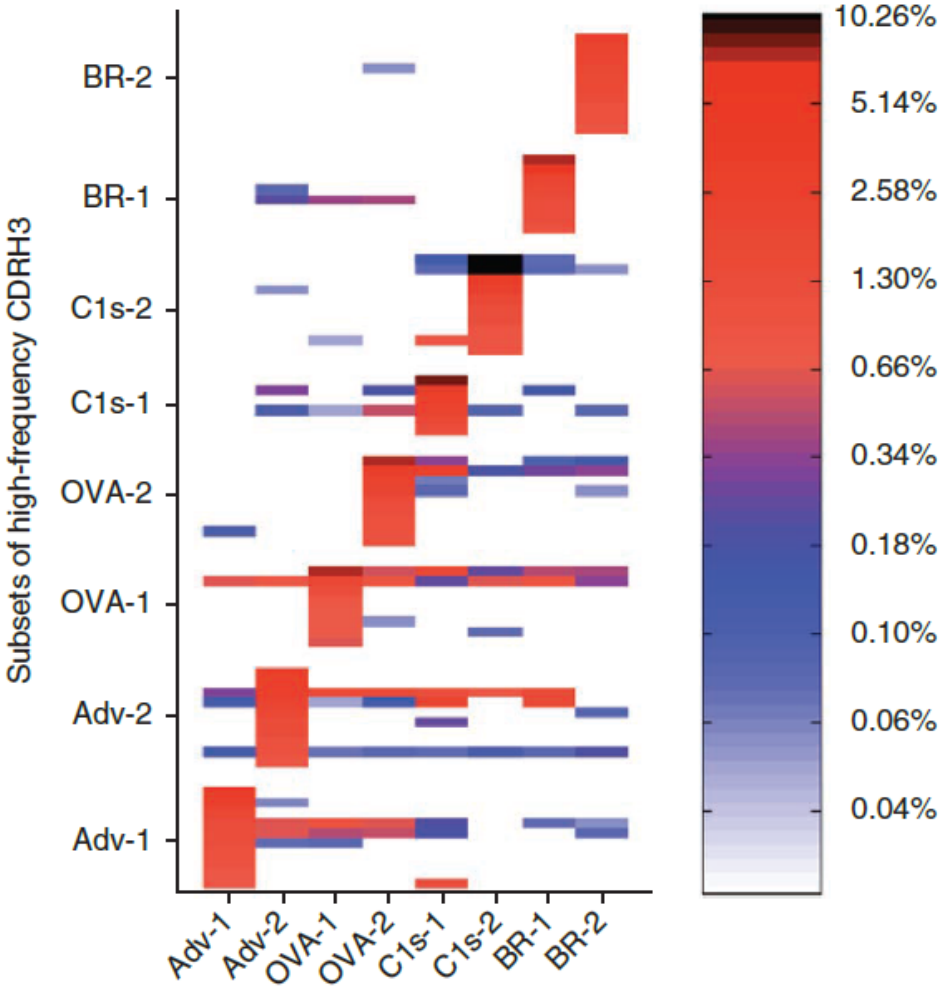
- As expected for early responses, the most highly abundant CDR3s were assembled from a diverse array of germline V-gene segments, with an average somatic mutation rate of only two and five amino acid substitutions for VL and VH, respectively

- Certain germline V-gene families were represented preferentially in mice responding to particular antigens

Results: variability of the repertoire

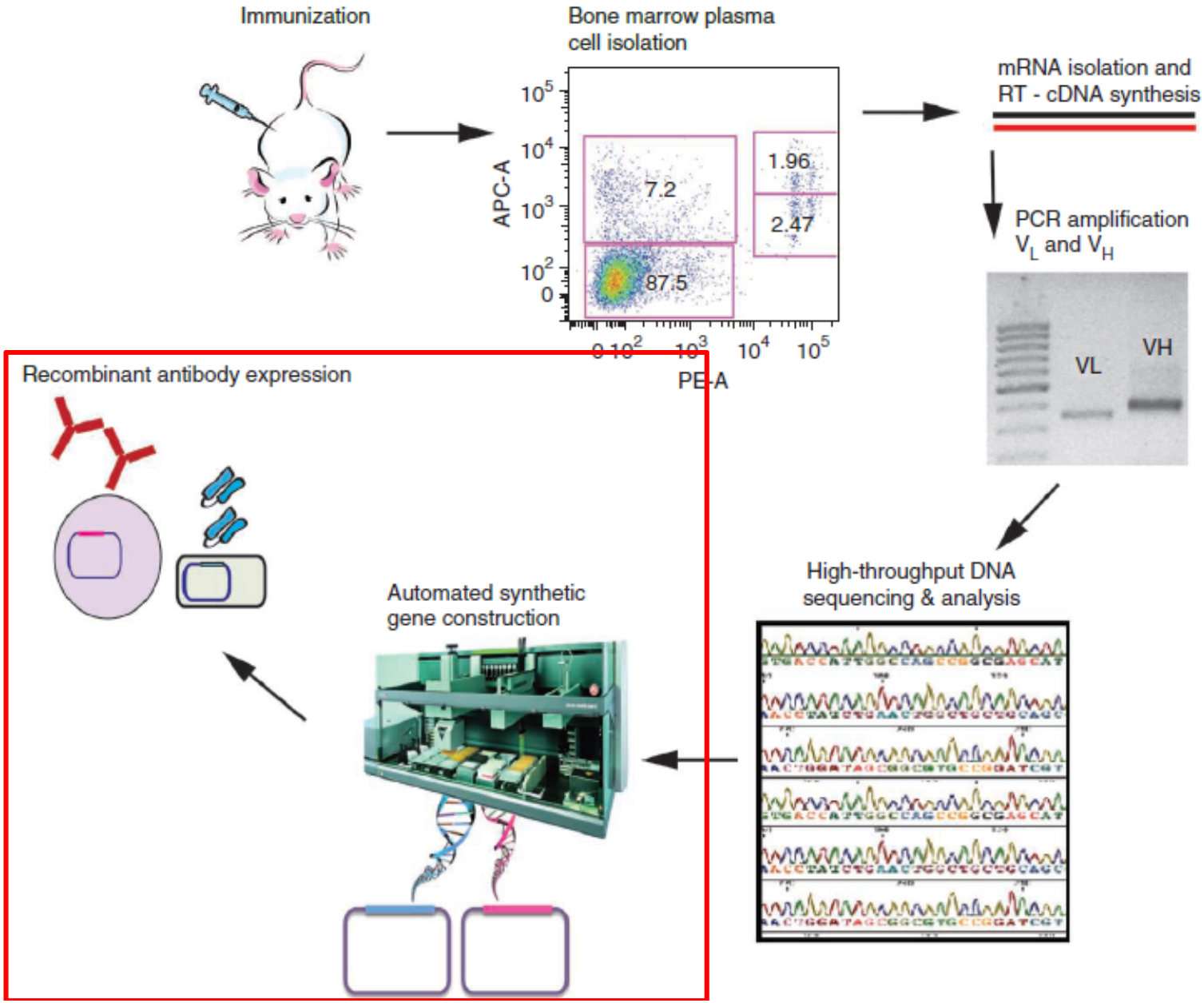
- The VH repertoires were quite distinct even among genetically identical littermates immunized with the same antigen on the same day

- This suggests that each mouse generates its own unique and highly expressed VH-gene repertoire, which may allow for the discovery of a panel of diverse antibodies



The y axis represents the ten highest frequency CDRH3 sequences identified in each mouse
The x axis compares the frequency of these prevalent CDRH3 sequences across all other mice

Synthetic gene construction and Ab expression



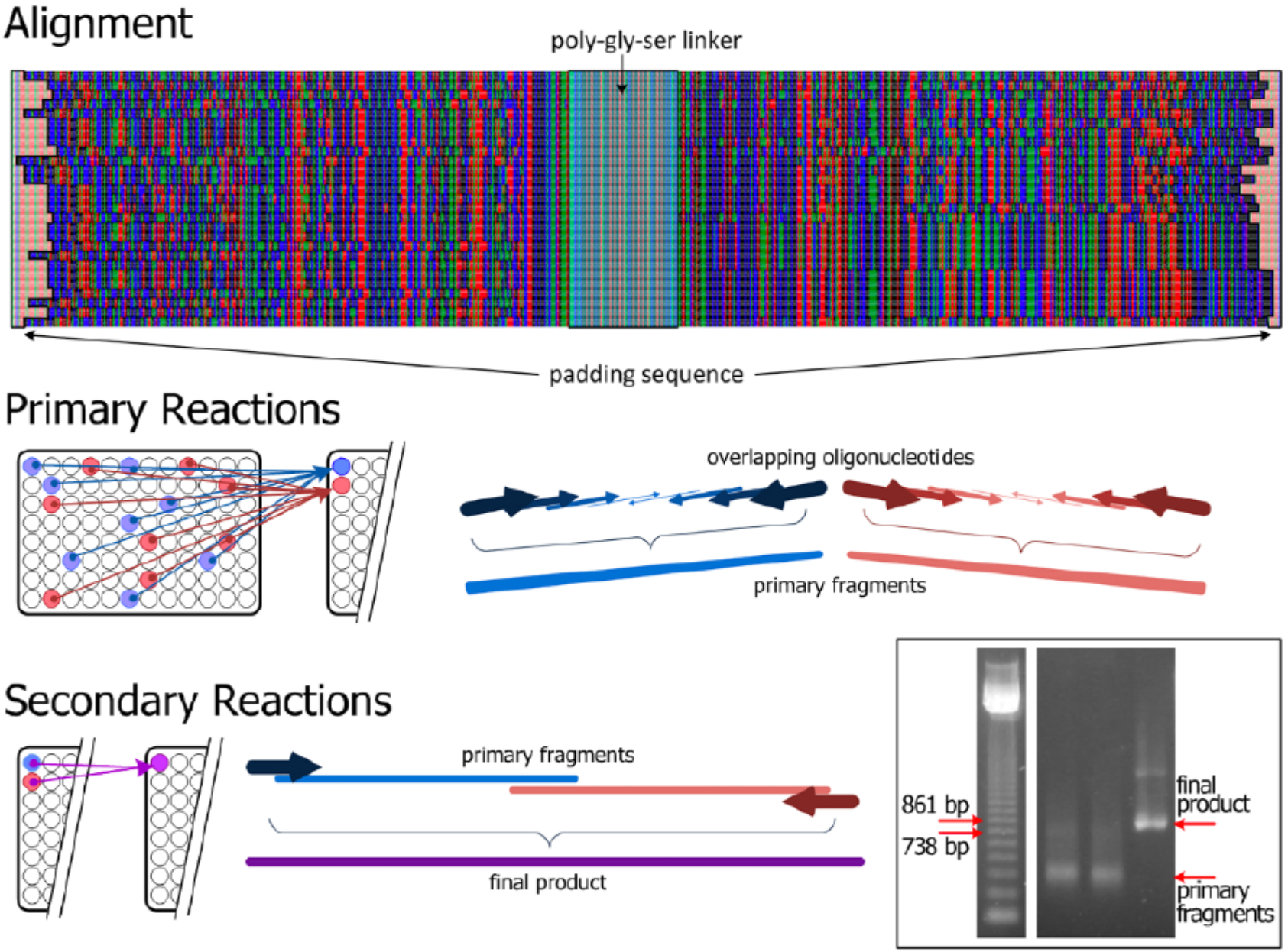
Pairing VH and VL genes and gene synthesis

- Hypothesis: VL and VH genes represented at approximately the same frequency likely arise from the same plasma cell and, hence, are naturally paired

- To test this hypothesis, four or five of the most abundant full-length VL and VH genes from each mouse (excluding VH sequences that were cross-represented in adjuvant-only mice) were synthesized and tested for antigen binding

- In cases where two VL or VH genes were found at very similar frequencies, multiple VL-VH combinations where constructed

Robotically assisted, high-throughput DNA synthesis



Results: antigen-binding of predicted VH-VL pairs (as scFv)

- Predicted VH-VL pairs, expressed as scFv, were antigen specific in most of the cases : we obtained 21/27 (~78%) antigen-specific antibodies from six mice immunized with three different protein antigens

Combinatorial library of scFvs comprising the four most abundant VL and VH genes from each of the two mice immunized with C1s

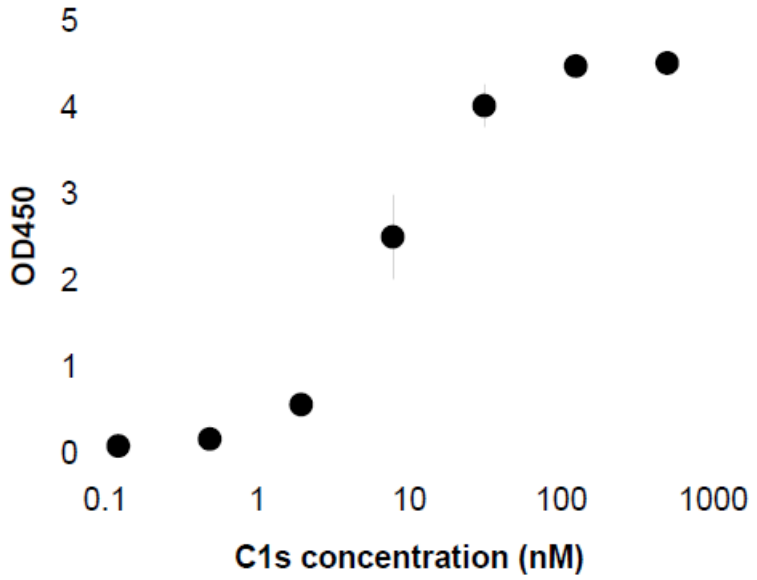
C1s-1	
V_L-V_H Pairing	ELISA Signal[a]
1L-1H	4.36
2L-1H	19.92
3L-2H	4.36
4L-3H	6.8
4L-4H	3.7
C1s-2	
1L-1HB	63.3
3L-1HB	4.65
2L-2H	3.1
3L-2H	4.22
3L-3H	8.8

Table 1 Antigen binding of antibody single-chain variable fragments (scFvs) from high frequency V_L and V_H genes

V_L-V_H pair	% V_L	CDRL3	% V_H	CDRH3	scFv binding
α–OVA					
1.1L-1.1H	11.70	WQGTHFPLT	7.11	GSSYYAMDY	+
1.2L-1.2H	4.40	QQYNSYPLT	1.10	LLWLYAMDY	+
1.3L-1.3H	3.38	QQSNSWYT	0.57	DVYDGYAMDY	+
1.4L-1.4H	2.20	QHHYGTPPWT	0.54	NPYAMDY	–
2.1L-2.1H	5.32	WQGTHFPLT	7.61	RTTVSRDWYFDV	+
2.2L-2.2H	4.05	QQYNSYPLT	3.23	YYYGSSAMDY	+
2.3L-2.3H	3.46	QQYSSYPLT	2.22	DGWYYFDY	+
2.4L-2.4H	2.01	QQHYSTPWT	2.10	EDDYDLFAY	+
α–C1s					
1.1L-1.1H	12.95	WQGTHFPQT	7.93	GNYYYAMDY	+
1.2L-1.1H	6.94	QQWSSYPQLT	7.93	GNYYYAMDY	+
1.3L-1.2H	3.81	QNDHSYPLT	2.64	DMISYWYFDV	+
1.4L-1.3H	3.16	QQGQSYPFT	1.67	EDYGNYWYFDV	+
1.4L-1.4H	3.16	QQGQSYPFT	1.67	EGYYYGSSYFDY	–
2.1L-2.1HA	17.10	FQGSHVPLT	10.99	SDRYDGYFDY	+
2.1L-2.1HB	17.10	FQGSHVPLT	9.93	SDRFDGYFDY	+
2.2L-2.2H	2.62	QQSNEDPWT	3.30	WLLLAY	+
2.3L-2.2H	2.20	WQGTHFPH	3.30	WLLLAY	+
2.3L-2.3H	2.20	WQGTHFPH	1.65	SDGYYYFDY	+
2.4L-2.4H	1.64	QQHYSTPFT	1.15	YYDYDKAYYFDY	–
α–Br					
1.1L-1.1H	6.64	LQYASSPFT	7.20	HDYGNYVDY	+
1.2L-1.2H	4.73	WQGTHFPRT	5.62	DGNYQEDYFDY	–
1.3L-1.3H	4.51	QQNNEDPRT	1.91	EGYAYDVDY	+
1.4L-1.4H	3.59	QQRSSYPLT	1.20	YDYGKDFDY	+
2.1L-2.1H	7.24	WQGTHFPQT	2.57	RGDGNYFFDY	+
2.2L-2.2H	4.50	QQGQSYPWT	2.27	GDEAWFAY	–
2.3L-2.3H	3.12	LQYASSPYT	2.03	EGDFDY	–
2.4L-2.4H	2.58	FQGSHVPWT	1.63	GGNYDYAMDY	+

Results: antigen-binding of predicted VH-VL pairs (as IgG)

V_L-V_H pair	2.1L-2.1HB (scFv)	2.1L-2.1HB (IgG)	2.3L-2.2H (IgG)
% V_L	17.10	17.10	2.20
% V_H	9.93	9.93	3.30
CDRL3	FQGSHVPLT	FQGSHVPLT	WQGTHFPH
CDRH3	SDRFDGYFDY	SDRFDGYFDY	WLLLAY
k_{on} (M^{-1} sec^{-1})	2.3×10^4	2.4×10^4	4.5×10^5
k_{off} (sec^{-1})	5.0×10^{-4}	1.2×10^{-3}	1.9×10^{-4}
K_D (nM)	20	50	0.43



Summary

- Successful generation of scFV or moAb specific for the desired antigen

- Procedure favorably compares in terms of costs and duration with other methods for moAb generation (hybridomas, B-cell immortalization, and B-cell screening and/or single-cell cloning)

- This technology could be extended to primates, included humans

- Codon optimization for scFv or moAB production could be easily implemented

- The high efficiency of the approach might depend on the use of BMPCs, which could allow to focus on antibodies selected by the immune system for their potent pathogen neutralization

- The applicability of this technology to conditions characterized by more complex antigens (e.g. bacteria, viruses) and multiple highly immunogenic epitopes will have to be explored

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]

Scientific background

- Each B-cell receptor consists of a pair of heavy and light chains

- High-throughput sequencing can identify large numbers of VH and VL sequences in a given B-cell repertoire, but information about endogenous pairing of heavy and light chains is lost after bulk lysis of B-cell populations

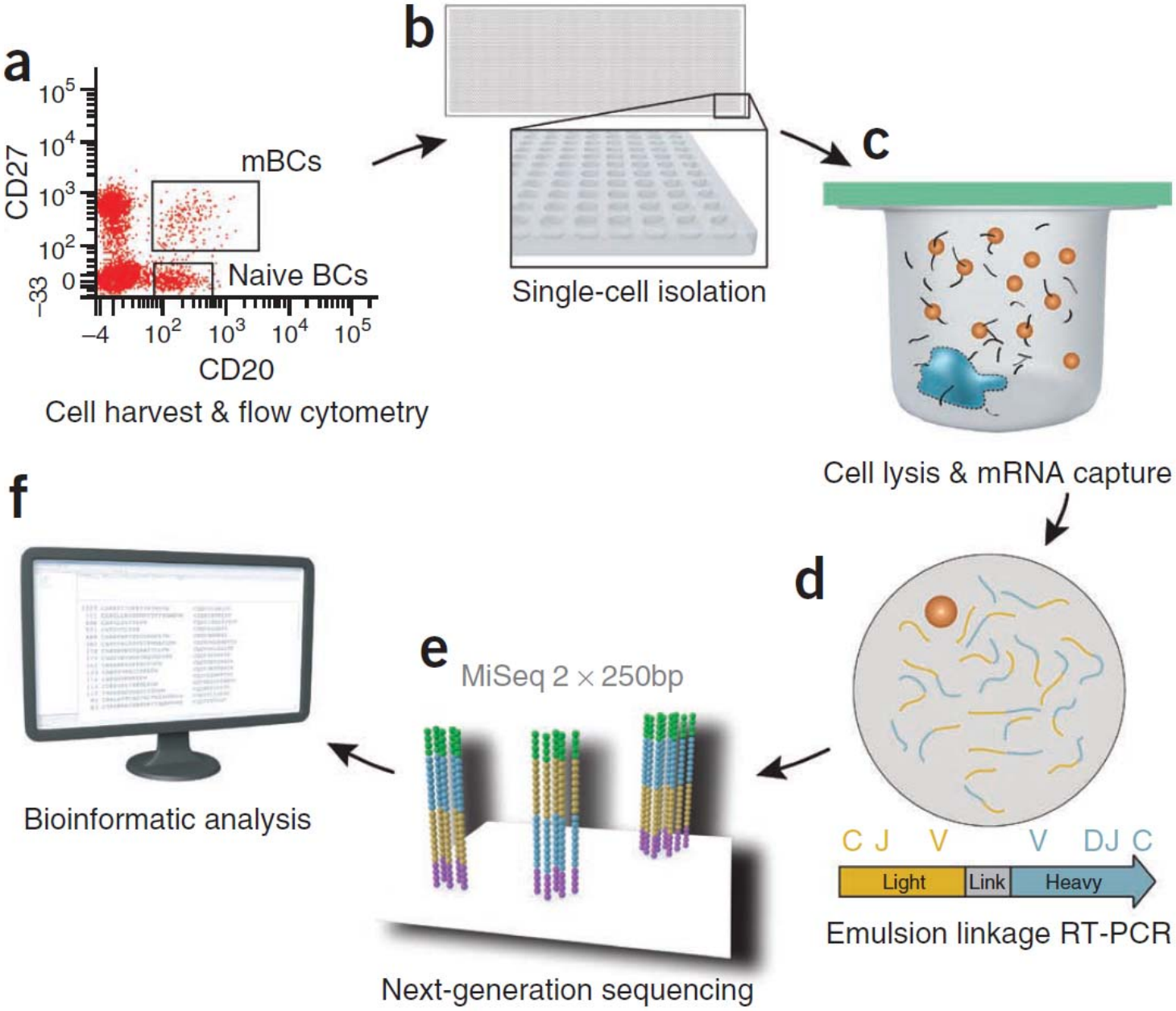
- Sequence analysis of VH:VL pairs is currently done by microtiter-well sorting of individual B cells followed by single-cell RT-PCR (scRT-PCR) and Sanger sequencing

- However, at most a few hundred VH:VL pairs are identified by means of scRT-PCR

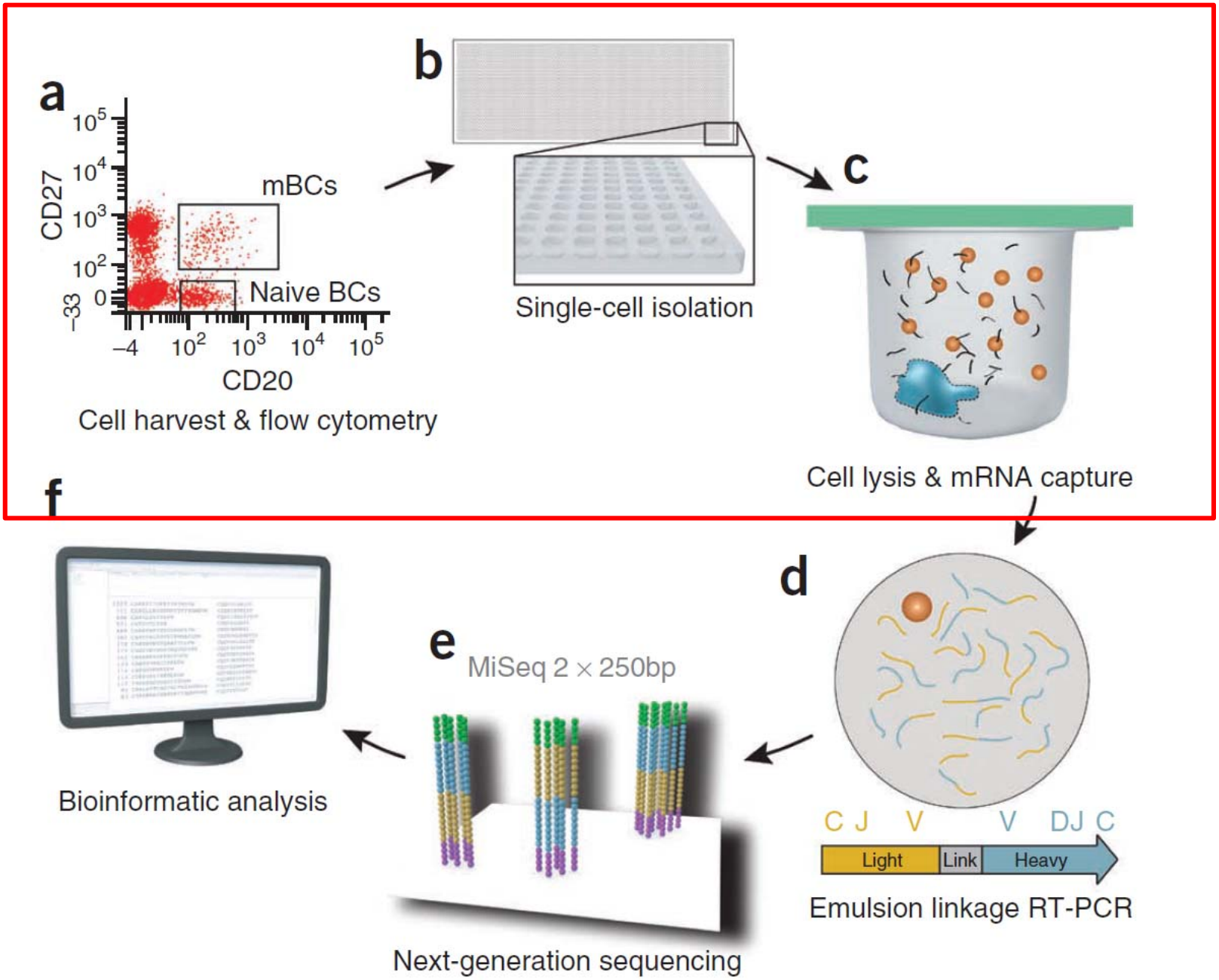
Aim of the study

To develop an accessible and scalable technology for the high-throughput sequencing of VH:VL pairs from individual human B lymphocytes

Study overview



Cell-sorting, plating and lysis



Cell-sorting, plating and lysis

-The desired population of sorted B cells was deposited by gravity into 125-pl wells (56-µm diameter) molded in polydimethylsiloxane (PDMS) slides

- Each slide contained 1.7×10^5 wells; four slides processed concurrently accommodated 68,000 lymphocytes at a ≥1:10 cell/well occupancy, which gave at least a 95% probability of there being only one cell per well

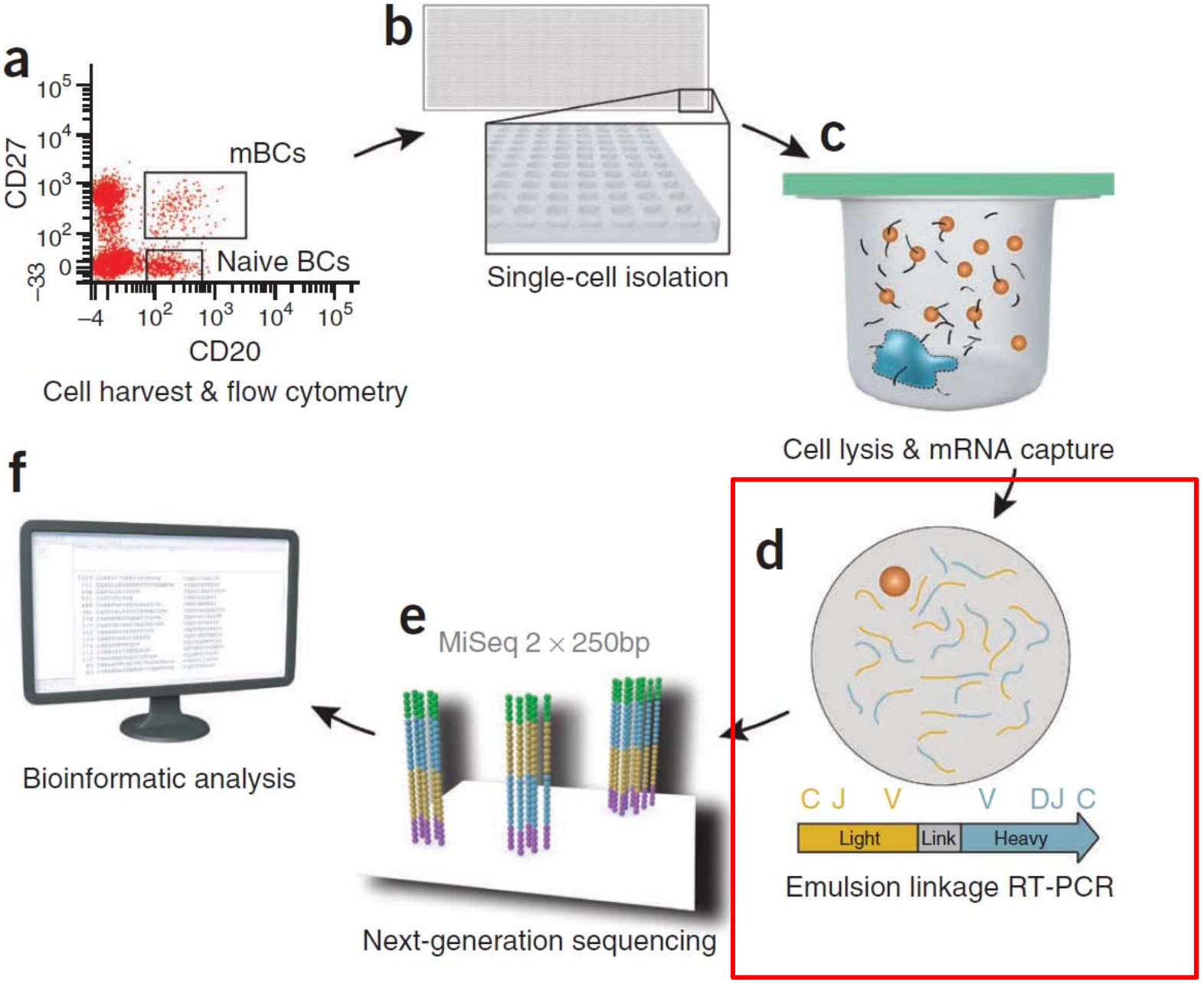
- Poly(dT) magnetic beads with a diameter of 2.8 µm were deposited into the microwells (55 beads/well)

- The slides were covered with a dialysis membrane

- The membrane-covered slides were incubated with an optimized cell lysis solution containing 1% lithium dodecyl sulfate that resulted in complete cell lysis within <1 min

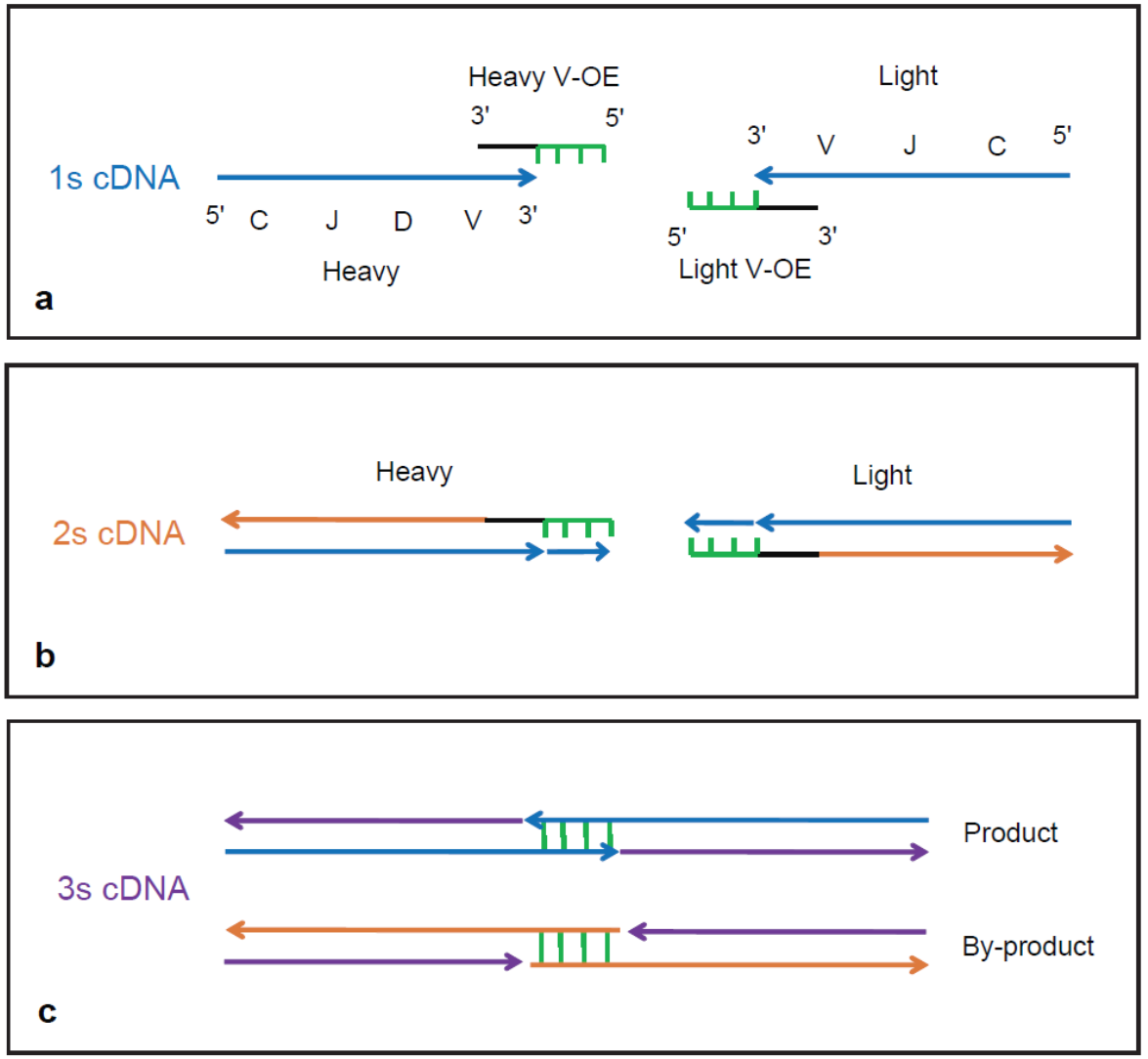
-- The mRNA annealed to the poly(dT) magnetic beads, which were then collected, washed and emulsified with primers, reverse transcriptase and thermostable DNA polymerase to carry out reverse transcription followed by linkage PCR

VH and VL amplification

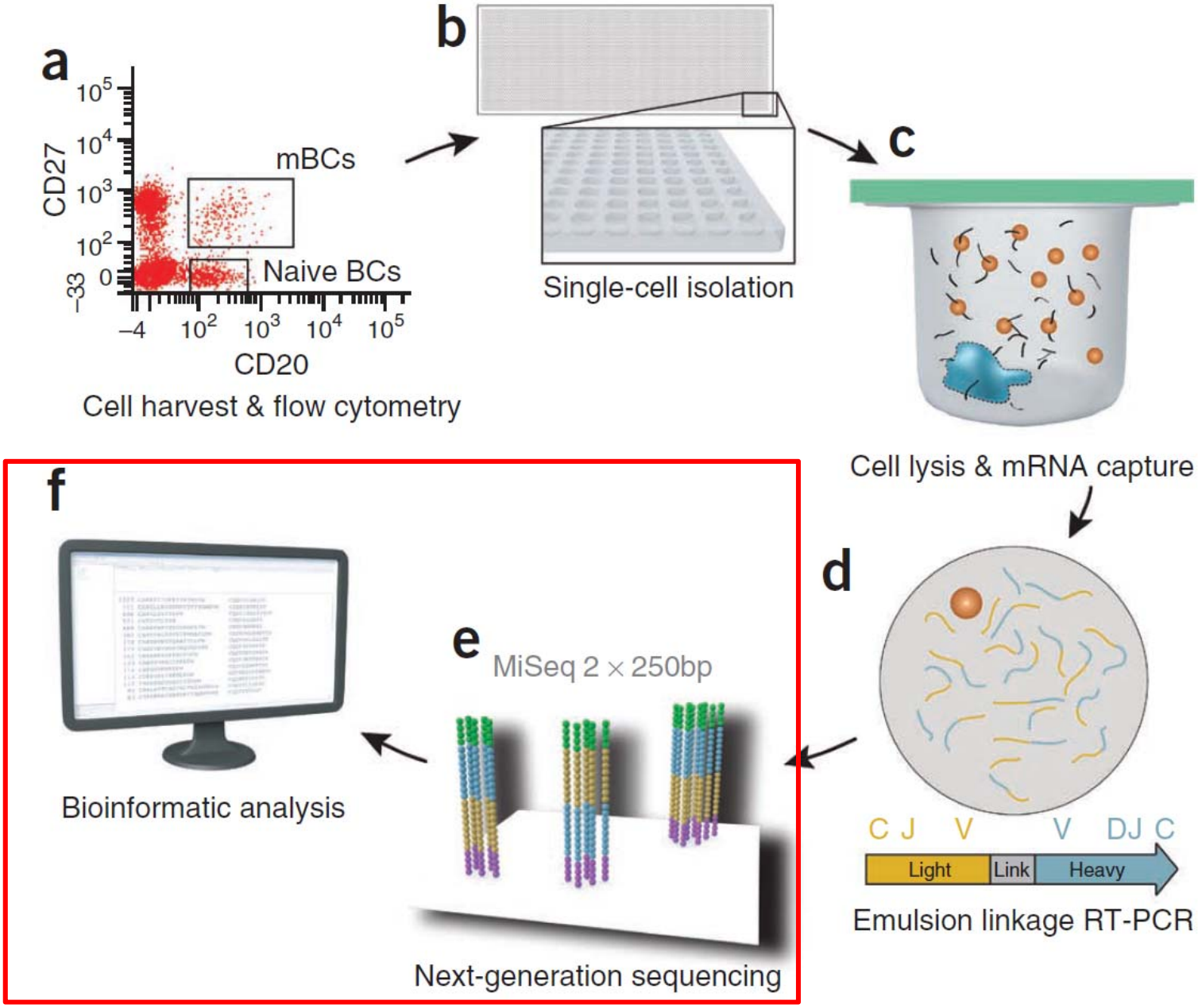


VH and VL amplification

a) V-region primers (black) with a 5’ complementary heavy/light overlap region (green) anneal to first strand cDNA
b) Second strand cDNA is formed by 5’ to 3’ extension; the overlap region is incorporated into all cDNA
c) After denaturation, heavy and light chains with first strand sense anneal to generate a complete 850 bp product through 5’ to 3’ extension. The CDR-H3 and CDR-L3 are located near the outside of the final linked construct to allow CDR3 analysis by 2x250 paired-end Illumina sequencing.



Next-generation sequencing



Next-generation sequencing

-The most informative 500 bp of this fragment, which encompassed the CDR-H3 and CDR-L3, was then sequenced on a long-read next-generation sequencing platform such as the 2 × 250 Illumina MiSeq

-This approach also provided the framework region FR3 and FR4 sequences and constant region N termini amino acid sequences that can be used for isotype assignment

- If FR1 to CDR2 region sequences were also desired, the VH and VL gene repertoires were analyzed by separate 2 × 250-bp sequencing runs.

- This latter step was required because of read-length limitations with existing technology; whereas single-molecule sequencing techniques allow for longer reads, the error rate is too high to enable robust classification of VH:VL sequences

Experimental samples and controls for VH:VL pairing

Experimental sample	Spiked control cells	Number of PDMS slides (total wells)
IgG^+ B cells from fresh blood donated by healthy individuals (61000 cells)	immortalized IM-9 lymphoblast cells	4 slides (6.8×10^5 total wells)
human plasmablasts ($CD19^+CD3^-CD14^-CD38^{++}CD27^{++}CD20^-$) from a healthy volunteer were collected 7 d after TT immunization (400 cells)	immortalized ARH-77 cell line as an internal control	1 slide (1.7×10^5 total wells)
Peripheral $CD19^+CD3^-CD27^+CD38^{int}$ memory B cells were isolated from a healthy volunteer 14 d after vaccination with the 2010-2011 trivalent FluVirin influenza vaccine (8000 cells)*	immortalized IM-9 lymphoblast cells	2 slides (3.4×10^5 total wells)

* An identical aliquot freshly processed for microtiter-well sorting of individual B cells followed by single-cell RT-PCR (scRT-PCR) and Sanger sequencing

Results: VH:VL gene family usage in different B cell populations

IgG+ B cells

Plasmablasts

Ten of the identified VH:VL pairs were expressed as IgG proteins in HEK293K cells

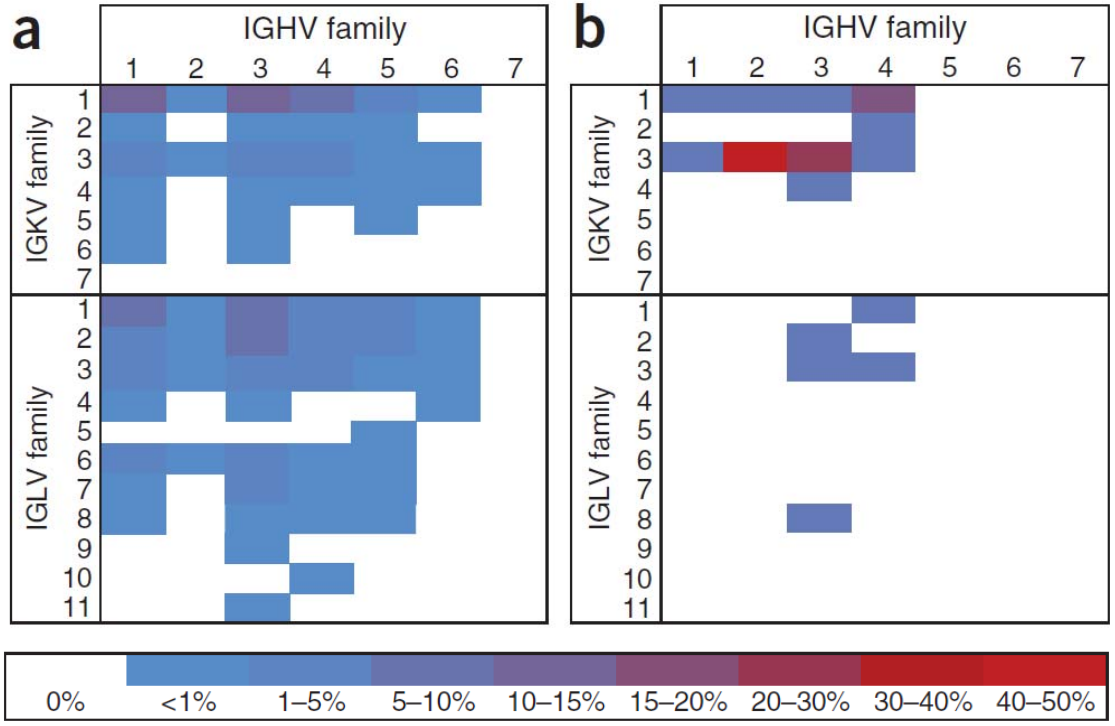


Table 1 TT-binding affinities of IgG antibodies sequenced from TT⁺ peripheral plasmablasts

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

Color indicates percentage of unique pairs from the sample

- Spiking controls demonstrated accurate VH:VL pairing

- VH:VL pairs containing rare families (e.g., IGHV7; IGKV5, 6, and 7; IGLV4, 10, and 11), indicating that this technique can identify rare B-cell clones present at physiological levels together with much more abundant clones

Results: comparison with scRT-PCR approach

scRT-PCR:

168 single B cells were sorted into four 96-well plates

168 RT and 504 nested PCR reactions were carried out individually to separately amplify the VH and VL (κ and λ) genes

DNA products were resolved by gel electrophoresis and sequenced to yield a total of 51 VH:VL pairs, of which 50 were unique.

High-throughput sequencing of the paired repetoire:

A total of 240 unique CDR-H3:CDR-L3 pairs were recovered

Four CDR-H3 sequences detected in the high-throughput pairing set were also observed in the single-cell RT-PCR analysis

VH:VL pairs identified with one, but not with the other technique presumably represent unique or very low abundance B-cell clones

Summary

- Successful analysis of the paired heavy and light chain repertoire and generation of moAb specific for the desired antigen

- Procedure favorably compares in terms of costs and duration with other methods for moAb generation (hybridomas, B-cell immortalization, and B-cell screening and/or single-cell cloning)

- This method cannot distinguish somatic variants originating from clonally related B cells that contain upstream mutations between FR1 and CDR2 regions

- The use of primers against the FR1 region can represent a bias in cases in which antibodies contain somatic mutations in all regions, including FR1 (e.g. chronic infections)

nature biotechnology

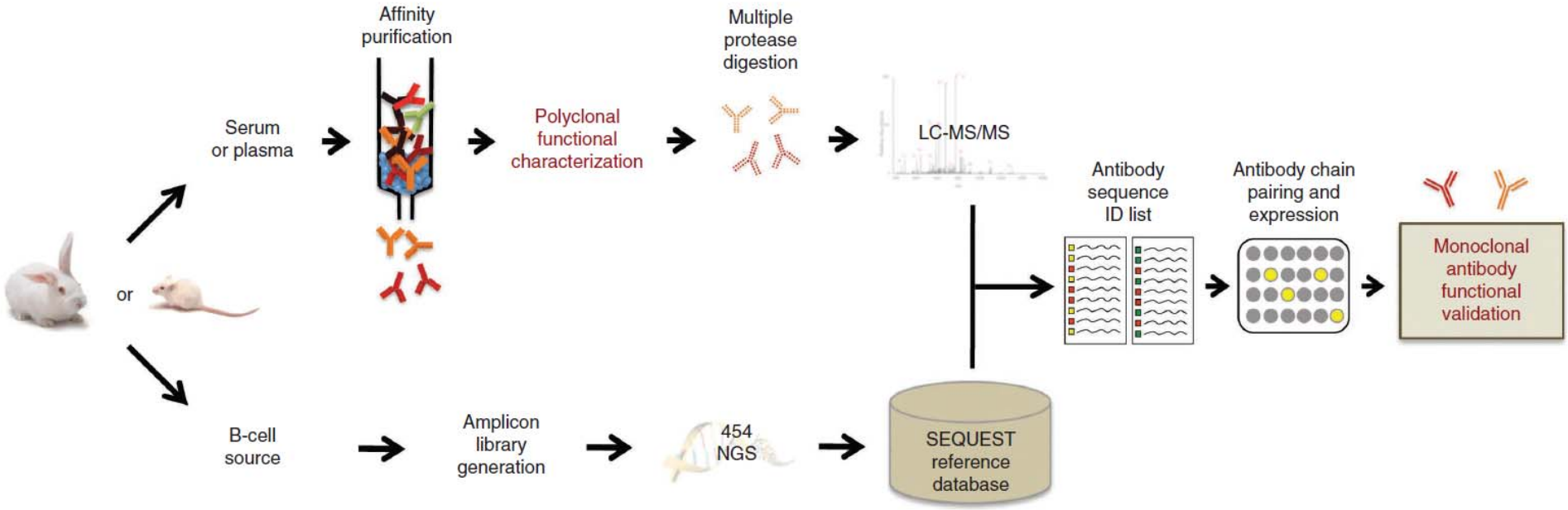
A proteomics approach for the identification and cloning of monoclonal antibodies from serum

Wan Cheung Cheung[1,2], Sean A Beausoleil[1,2], Xiaowu Zhang[1], Shuji Sato[1], Sandra M Schieferl[1], James S Wieler[1], Jason G Beaudet[1], Ravi K Ramenani[1], Lana Popova[1], Michael J Comb[1], John Rush[1] & Roberto D Polakiewicz[1]

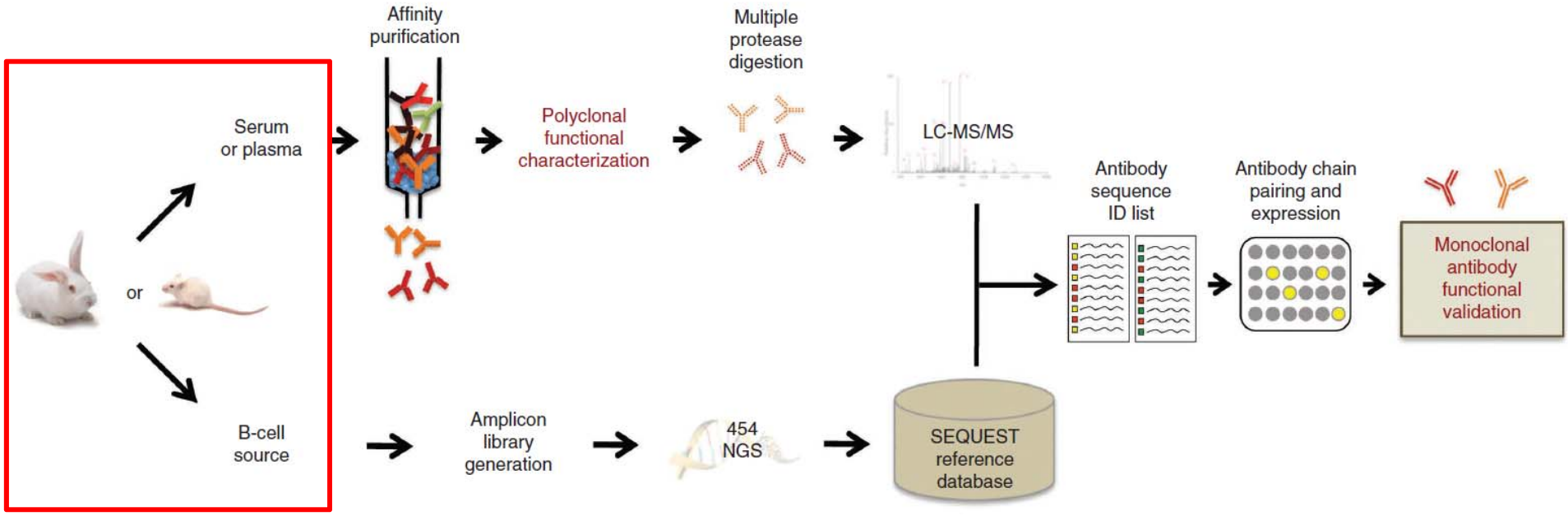
Aim of the study

To develop a proteomics approach that identifies antigen-specific antibody sequences directly from circulating polyclonal antibodies in the serum of an immunized animal

Study overview



Immunization



Immunization

- New Zealand white rabbits with human progesterone receptor A/B (PR A/B) peptides conjugated to keyhole limpet hemocyanin

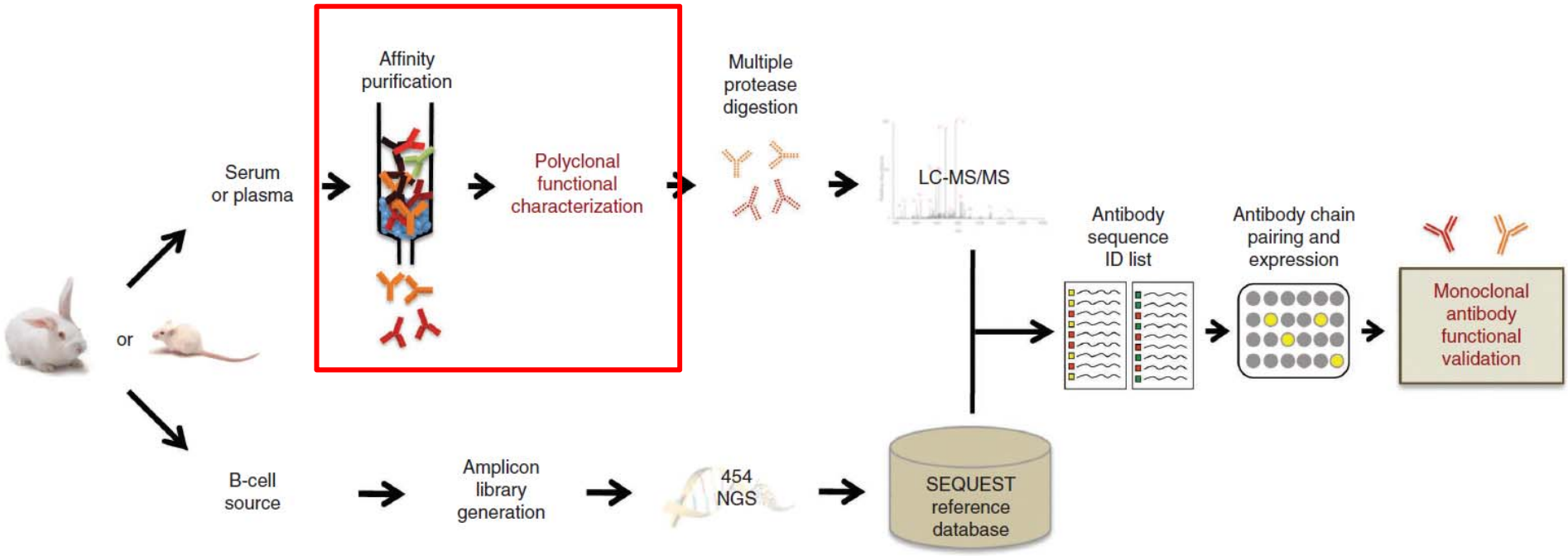
- Antigen-specific antibody activity in the crude serum of each animal was screened

- The rabbit with the strongest ELISA and WB signals to PR A/B (data not shown)

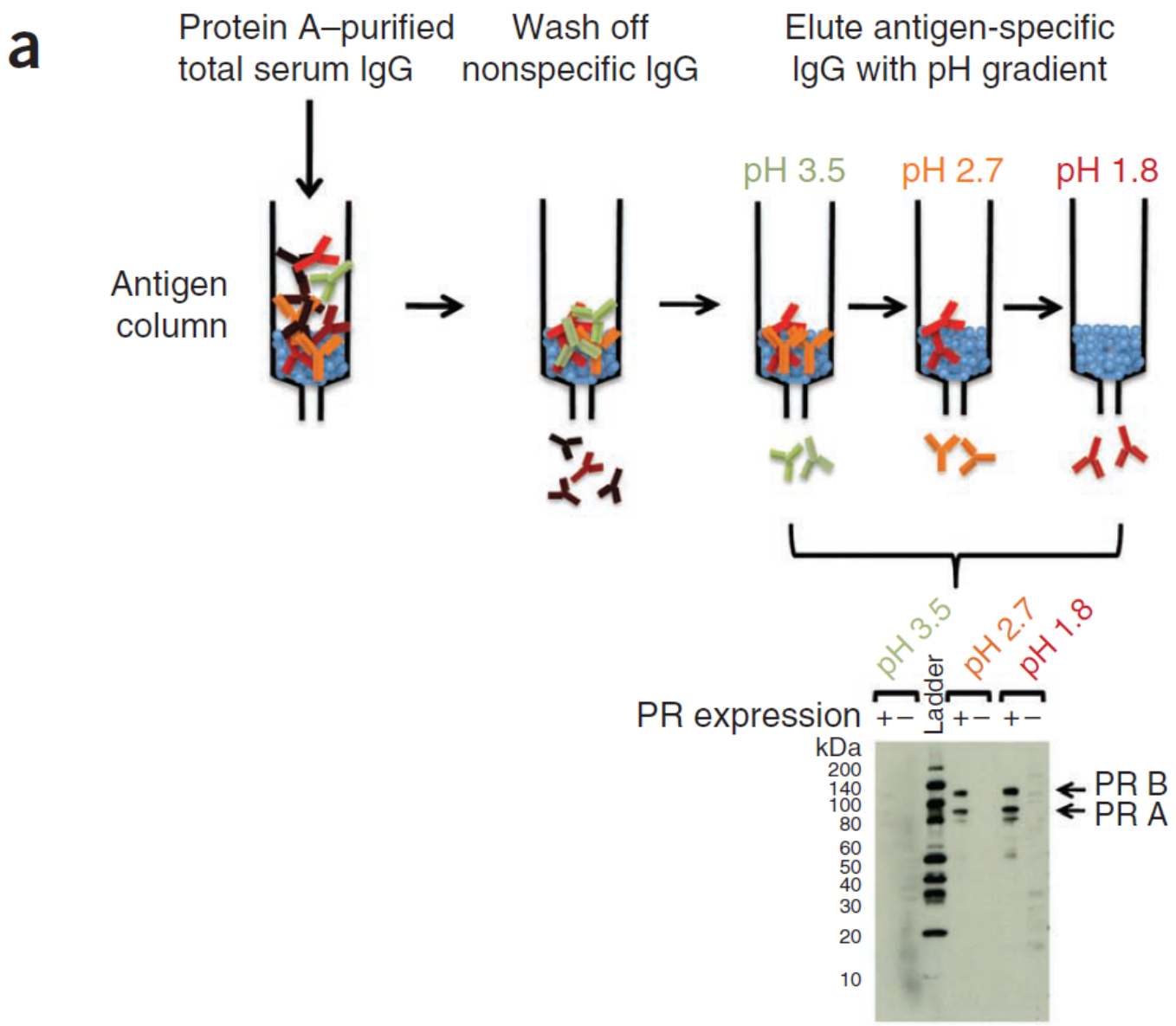
- Serum from this animal was collected from 20 ml of blood

- RNA was obtained from splenic B cells

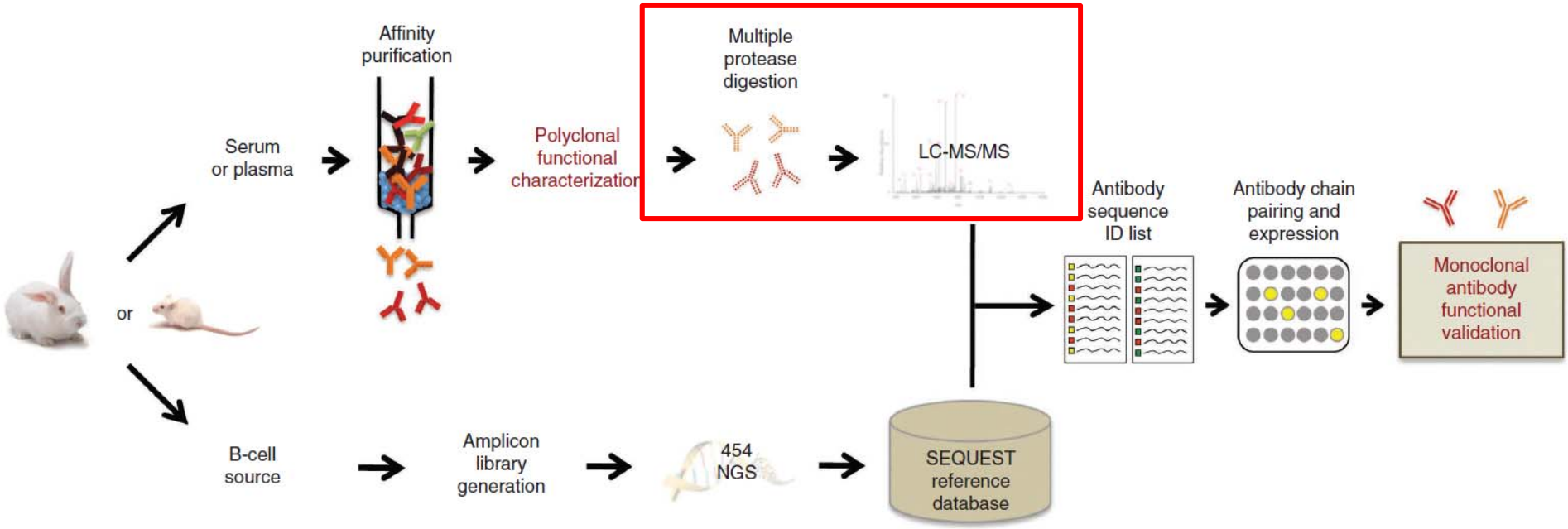
Polyclonal antibodies isolation and characterization



Polyclonal antibodies isolation and characterization

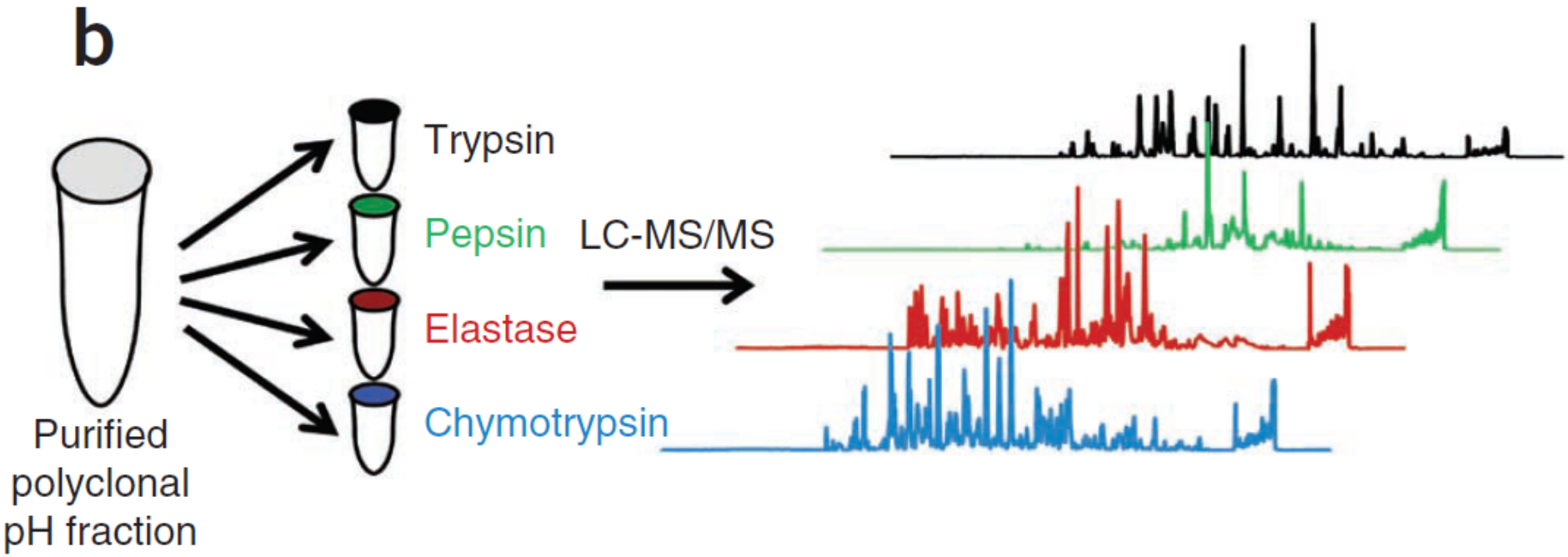


LC-MS/MS analysis



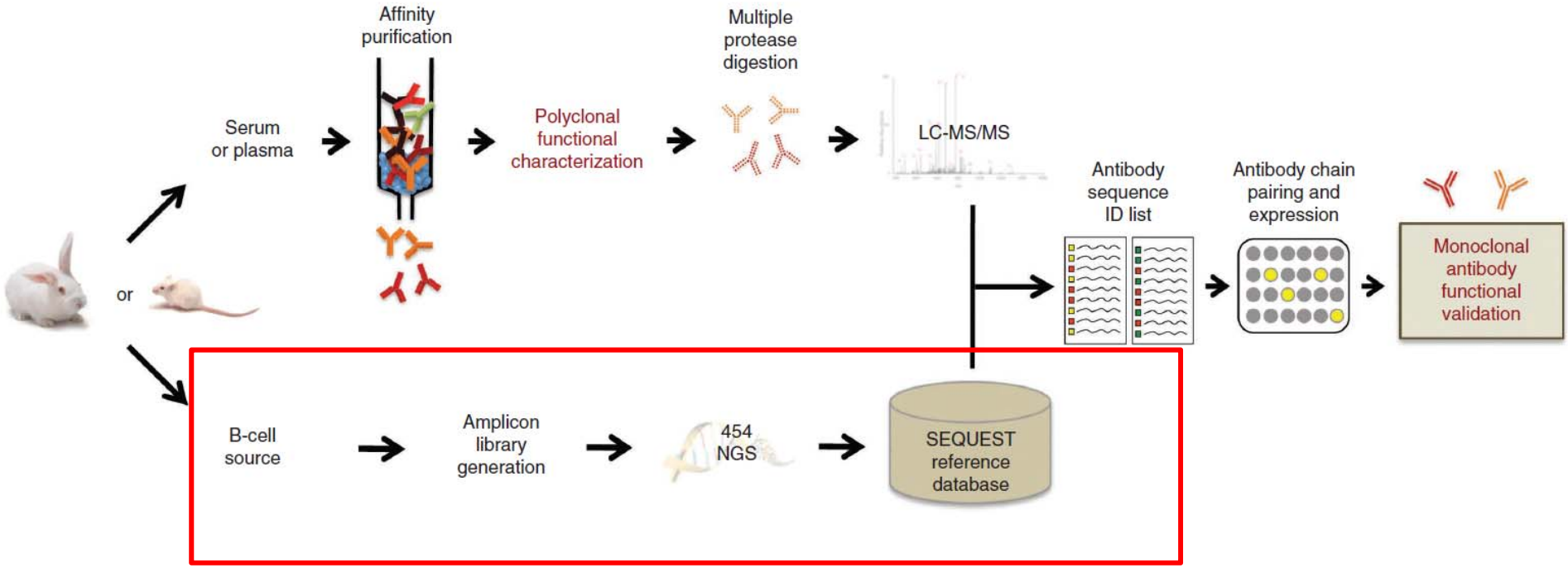
LC-MS/MS analysis

On pH 1.8 fraction



Four runs with 45-min gradient, producing ~10000 spectra per run

Generation of a custom database for rearranged Ig genes



Generation of a custom database for rearranged Ig genes

- RNA isolated from total splenocytes collected from the same animal that showed strong specific activity to PR A/B

-VH and VL region amplicons generated using barcoded primers specific to rabbit Ig γ- and κ-chains

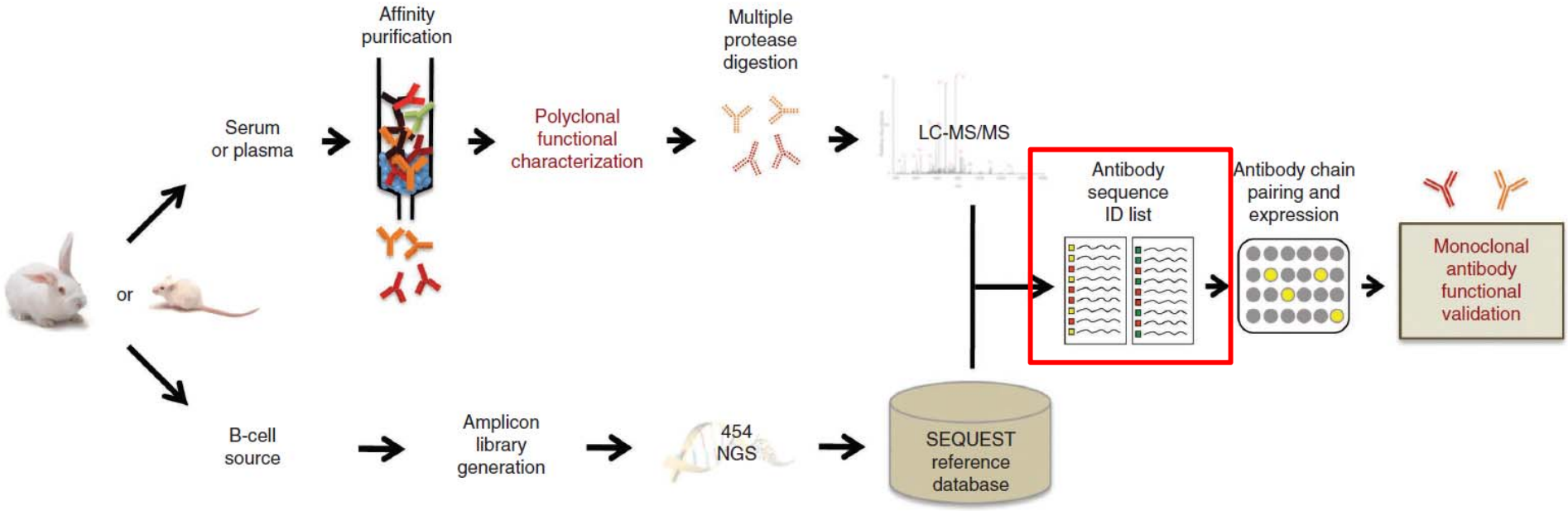
- Sequencing performed on Roche 454 next-generation sequencing platform

- 80,000 high-confidence reads obtained, 44,363 of which contained the entire V-region:

    - 5,279 unique γ-chain CDR3 sequences

    - 11,681 unique κ-chain CDR3 sequences

Antibody identification



Antibody identification

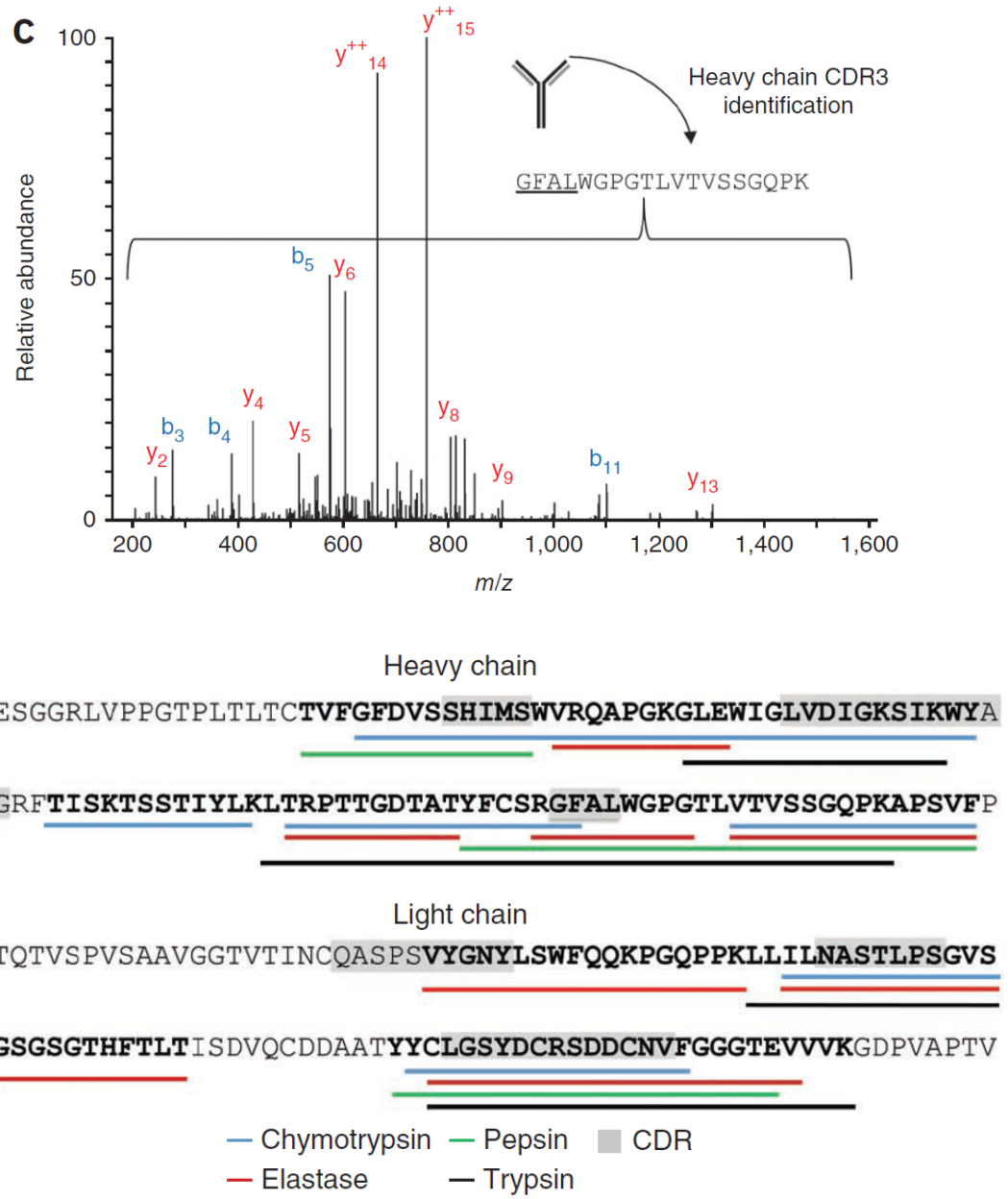
High-confidence V-region sequence identification based on:

(i) overall high coverage (≥65%)

(ii) at least 12 unique peptides due to high degree of homology of V-region sequences

(iii)high hypervariable region coverage, specifically, ≥95% coverage of CDR3

Antibody identification



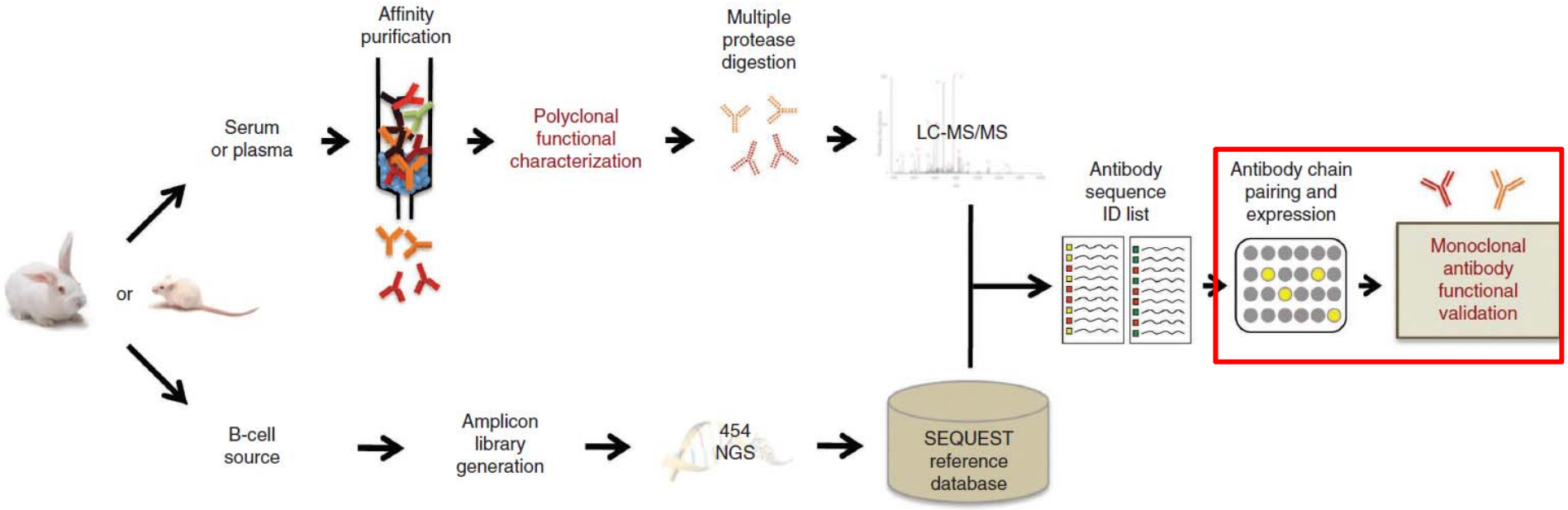
Results: Antibody identification

Table 1 Identification of high-confidence heavy and light chains

NGS ref. no.	Total peptide count	Percent variable region coverage	CDR3 sequence	NGS rank by CDR3 frequency	Germline V(D)J
G2JXQJ001A2Q81	101	95.69	KLGL	212	IGHV1S45, D4-2, J4
G2JXQJ001AGJSJ	91	92.04	GFSL	76	IGHV1S69, *, J4
G2JXQJ001BJE8R	78	98.26	DLGDL	423	IGHV1S45, D3-1, J4
G2JXQJ001BT2NA	70	86.21	DLGNL	461	IGHV1S45, D4-1, J4
G2JXQJ001AFBNC	61	87.27	GNL	58	IGHV1S44, D4-1, J4
G2JXQJ001AL49Y	59	87.72	DFHL	237	IGHV1S45, *, J4
G2JXQJ001BWR23	56	89.17	GSLGTLPL	103	IGHV1S45, D8-1, J2
G2JXQJ001BN8MH	50	82.14	GFAL	109	IGHV1S69, *, J4
G2JXQJ001BPNUG	48	81.51	GHDDGYNYVYKL	123	IGHV1S69, D6-1, J4
G2JXQJ001BZA42	35	95.54	GFTL	1,417	IGHV1S69, *, J4
G2JXQJ001BJ8KJ	93	87.27	LAGYDCTTGDCFA	2,769	IGKV1S15, J1-2
G2JXQJ001BQM6D	47	95.5	LGGYDCDNGDCFT	85	IGKV1S15, J1-2
G2JXQJ001A9VP3	33	92.79	LGTYDCRRADCNT	5,654	IGKV1S19, J1-2
G2JXQJ001BQJFD	28	98.15	QSTLYSSTDEIV	86	IGKV1S10, J1-2
G2JXQJ001BJCLS	28	96.23	QCSYVNSNT	4,518	IGKV1S44, J1-2
G2JXQJ001AG4TB	24	65.45	LGSYDCRSDDCNV	179	IGKV1S2, J1-2
G2JXQJ001AIZ32	17	86.11	LGAYDDAADNS	252	IGKV1S19, J1-2
G2JXQJ001BJYR5	15	72.07	LGTYDCNSADCNV	1,128	IGKV1S15, J1-2

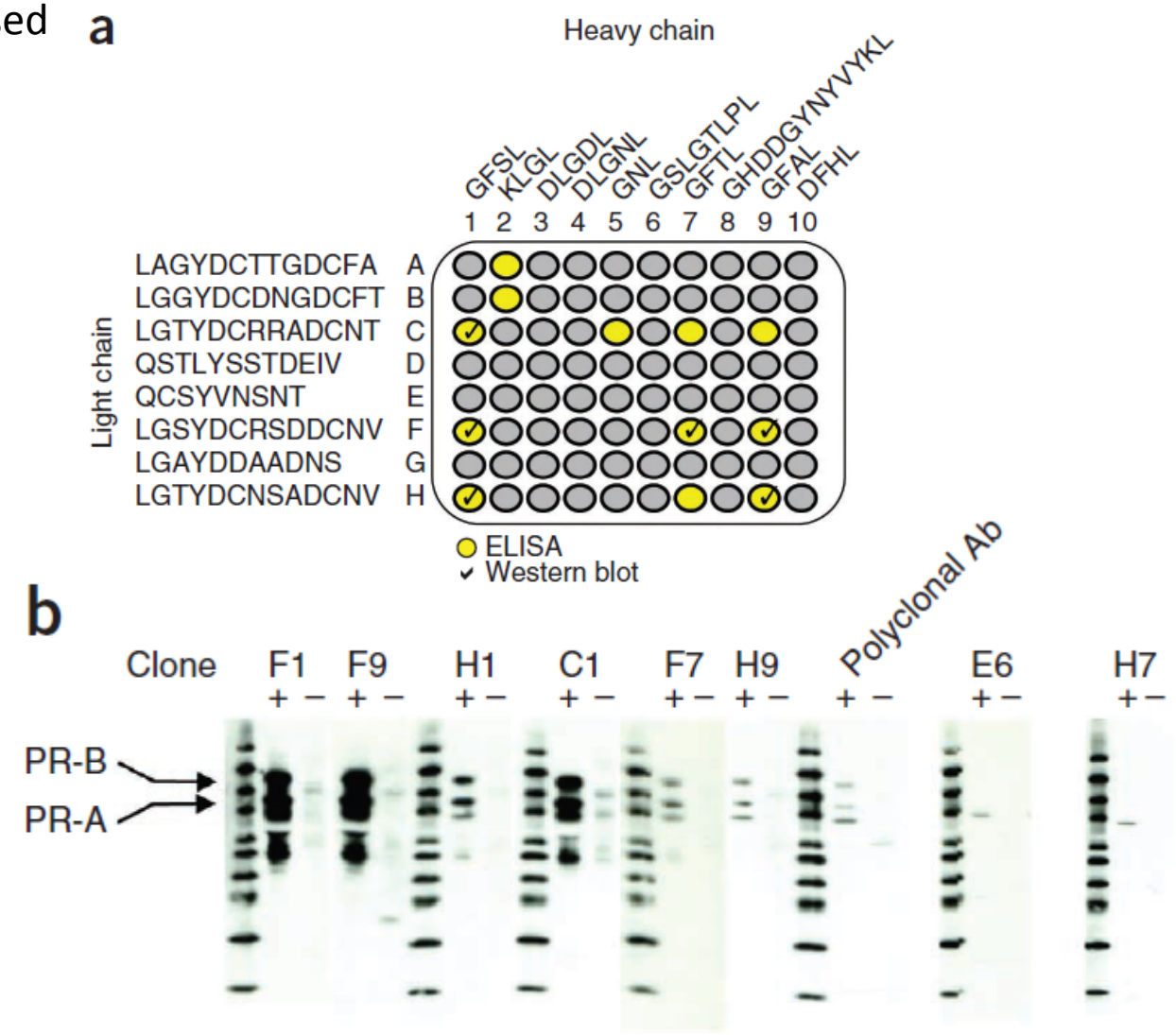
No information on natural pairing of the identified VH and VL genes.

Antibody chain pairing, expression and characterization

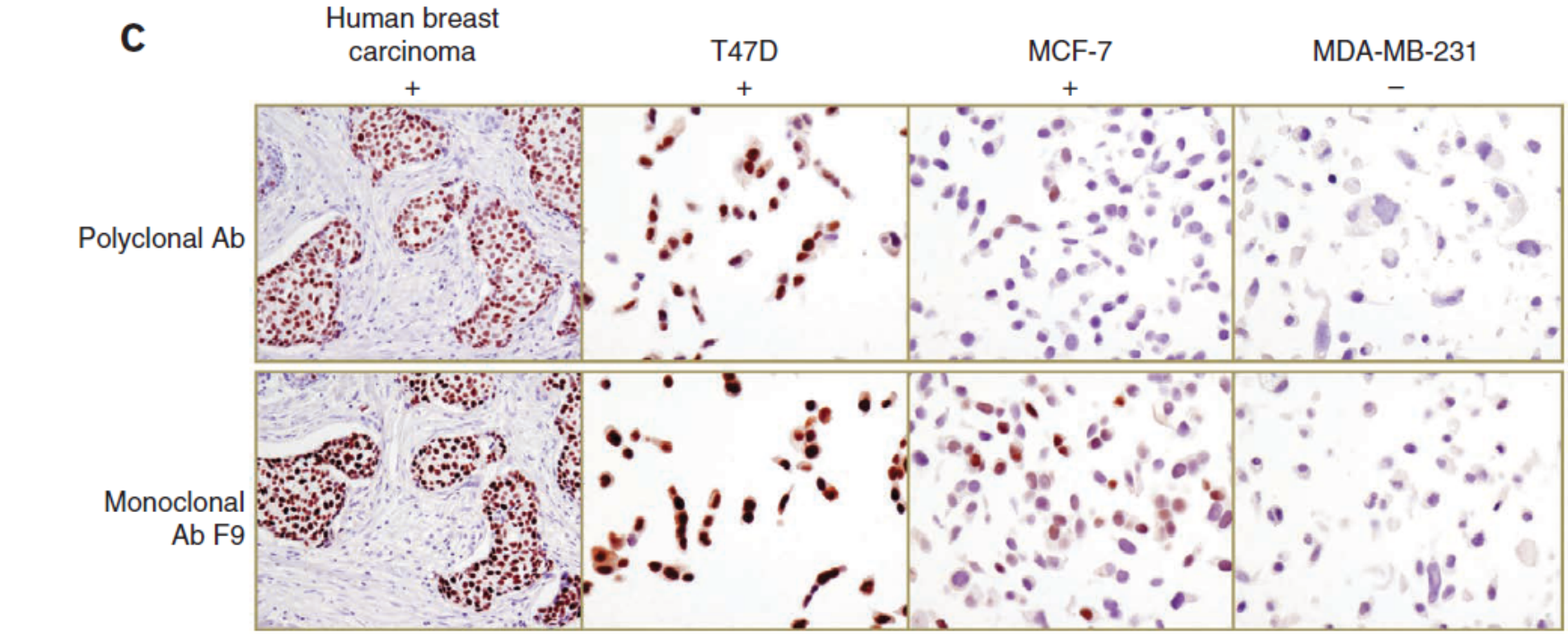


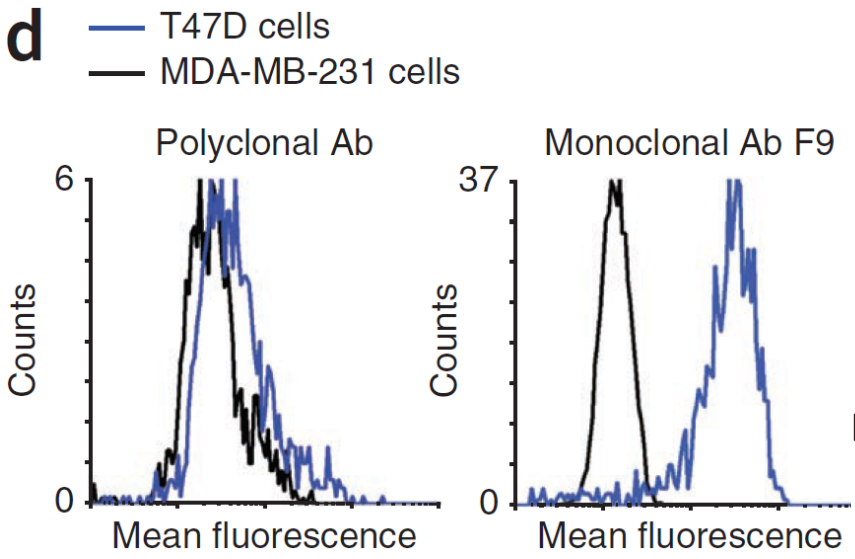
Antibody chain pairing, expression and characterization

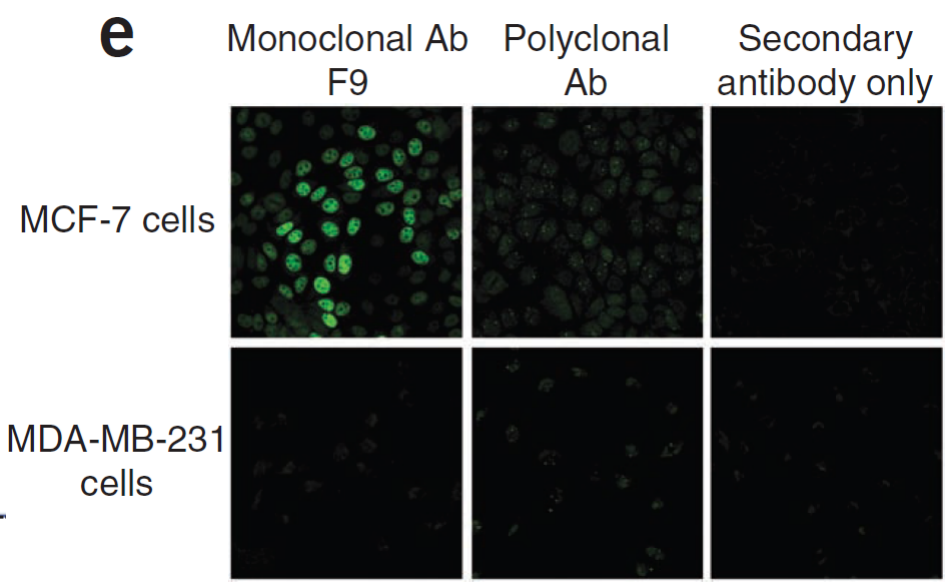
- All possible combinations of heavy and light chain pairings (80 antibodies) in addition to the heavy and light chain sequences most frequently observed in the B-cell sequencing data were cloned and expressed



Antibody chain pairing, expression and characterization







Further validation

Table 2 Functionally relevant monoclonal antibodies against multiple targets identified by the LC-MS/MS platform as tested by ELISA and western blot analysis

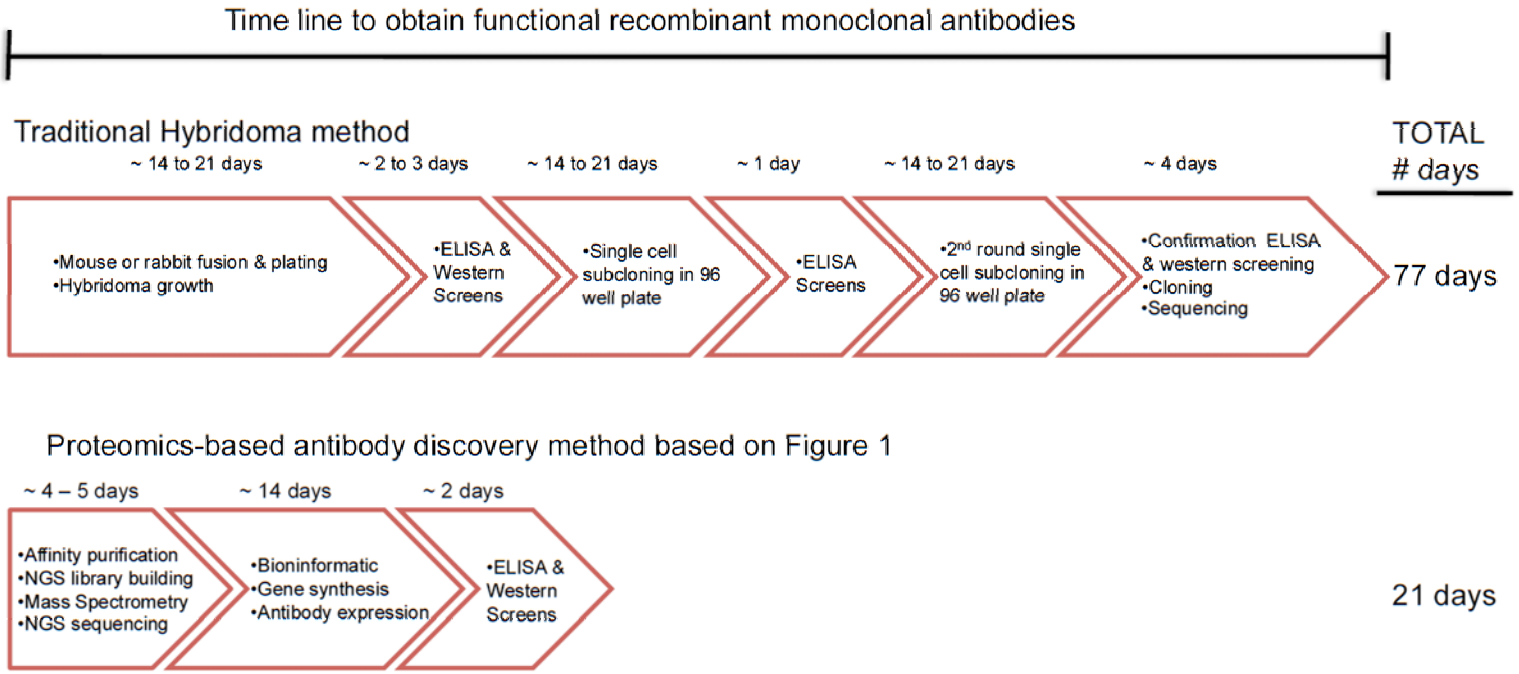
Antigen	Immunized species	High-confidence heavy + light chains	Unique ELISA^{+} clones	Unique WB^{+} clones
PR A/B	Rabbit	8 + 10	12	6
pMET	Rabbit	11 + 10	6	4
Lin28A	Rabbit	7 + 4	5	5
Sox1	Rabbit	9 + 5	12	1
p-p44/42	Mouse	12 + 13	15	3

NGS, next-generation sequencing; WB, western blot analysis.

Summary

- Successful application of this proteomics and next-generation sequencing approach for the identification of antigen-specific antibodies when using different antigens in both rabbits and mice

- Procedure favorably compares in terms of costs and duration with other methods for moAb generation



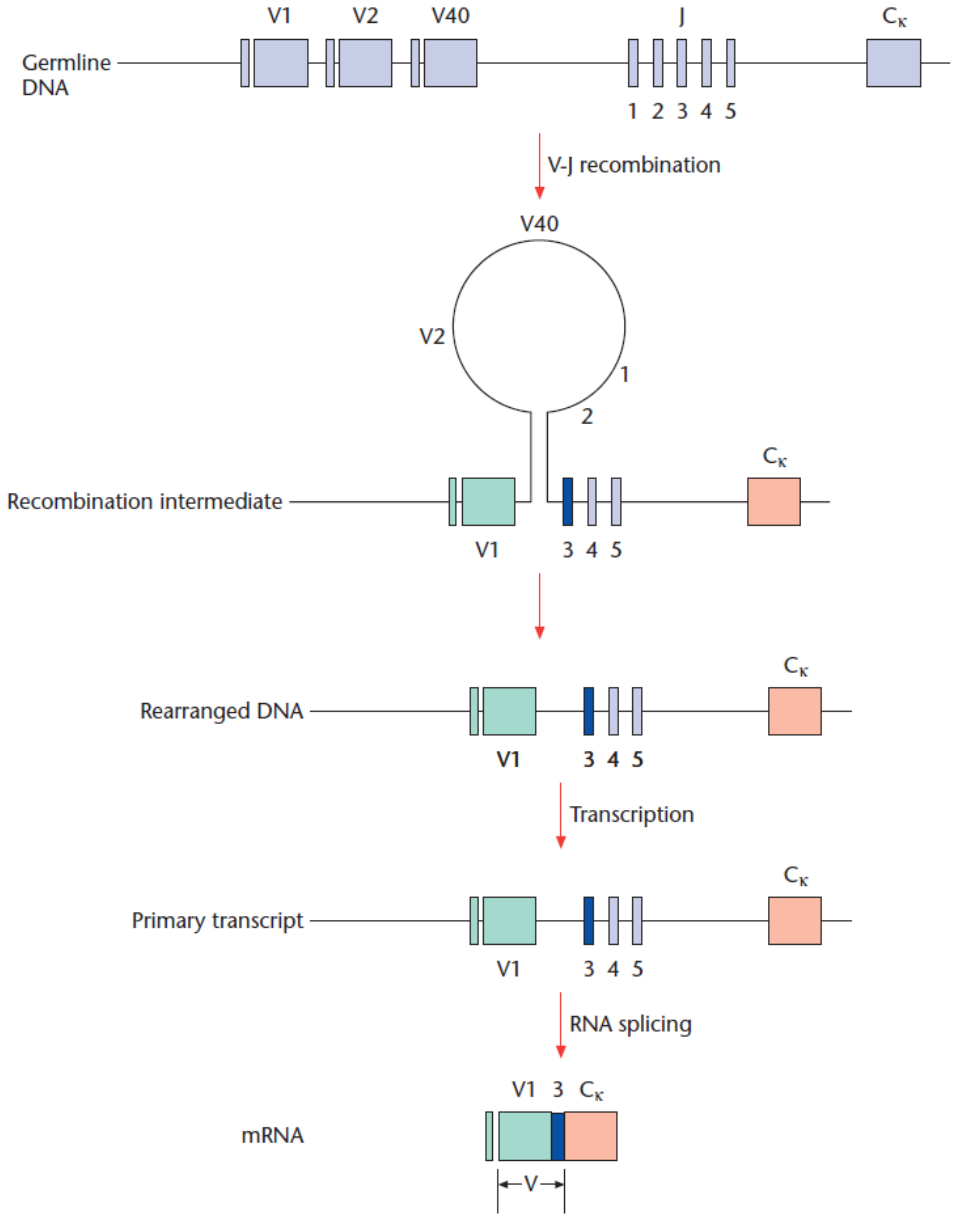
Summary

- VH:VL pairing based on high frequency in next-generation sequencing experiments did not result in antigen-specific antibodies

-The approach disrupts natural VH:VL pairing, however this did not hamper the ability to identify antibody pairs with functional properties equal or superior to the polyclonal antibody from which they were derived

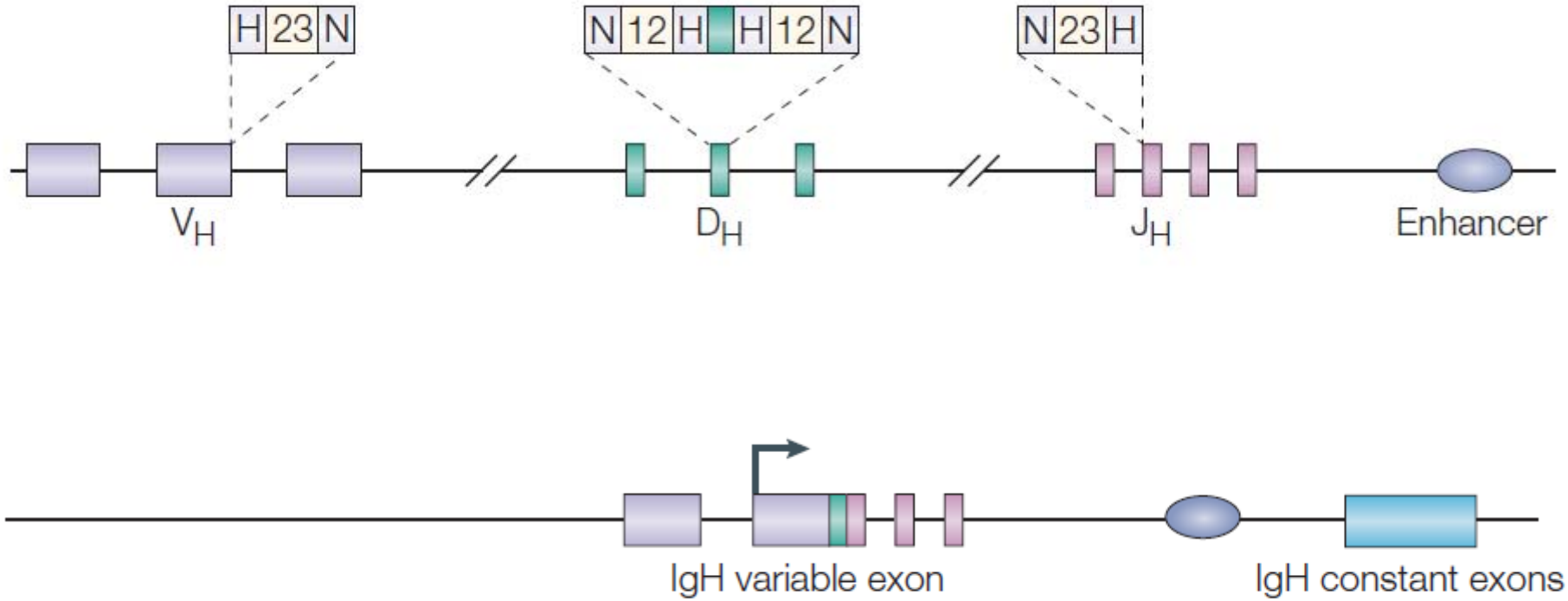
- The approach is focus on IgGk antibodies, however it could be adapted to include also different isotypes

L chain production

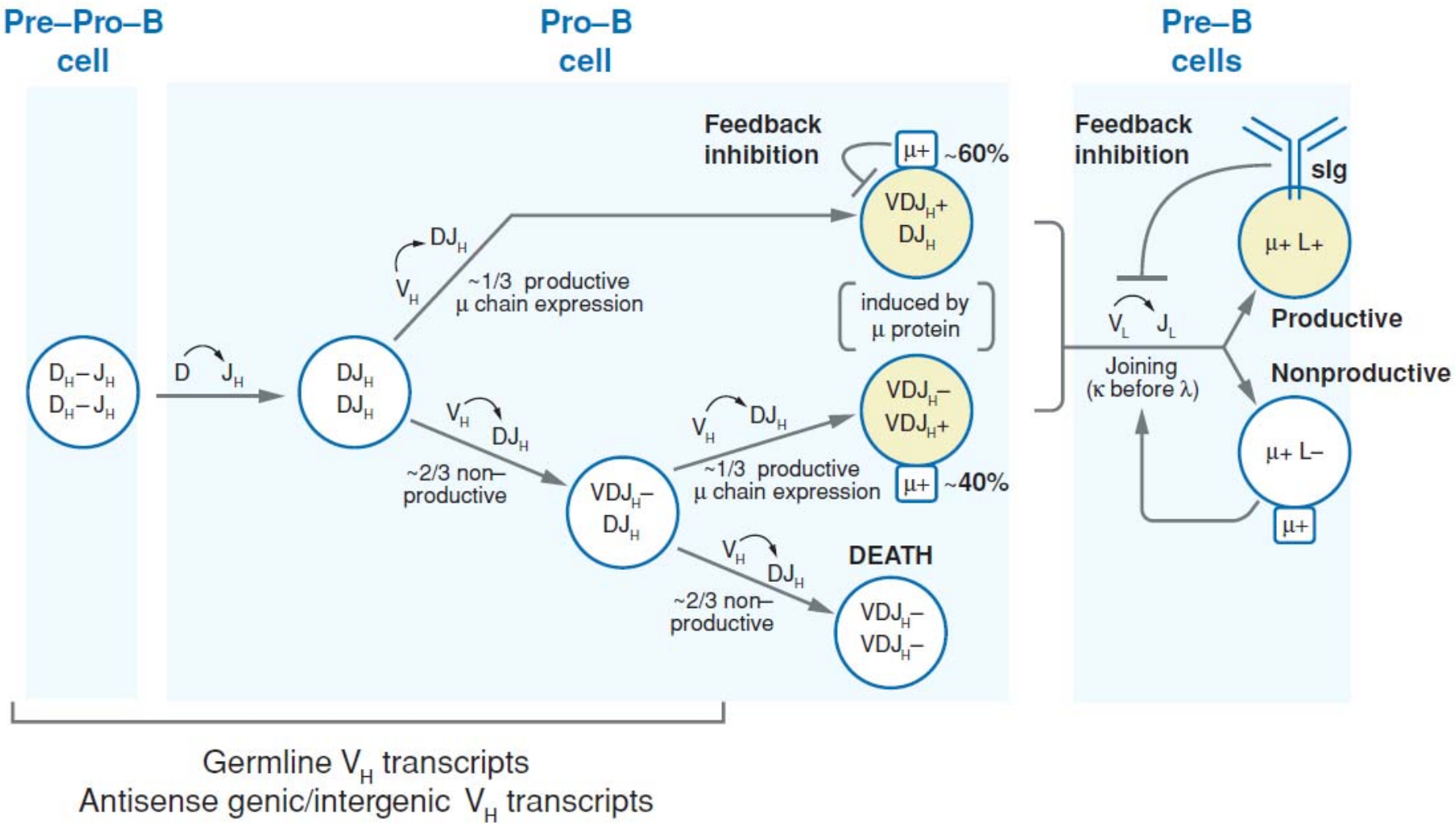


Lucas Encyclopedia of life sciences 2003

Recombination signal sequences



Schlissel Nat Rev Immunol 2003



Jung et al Ann Rev Immunol 2006

Bioinformatics pipeline for V gene analysis