



# Computational Model Reveals Limited Correlation between Germinal Center B-Cell Subclone Abundancy and Affinity: Implications for Repertoire Sequencing

## OPEN ACCESS

### Edited by:

Harry W. Schroeder,  
University of Alabama  
at Birmingham, USA

### Reviewed by:

Andrew M. Collins,  
University of New South Wales,  
Australia  
Alex Rosenberg,  
University of Rochester Medical  
Center, USA

### \*Correspondence:

Antoine H. C. van Kampen  
a.h.vankampen@amc.uva.nl

<sup>†</sup>Senior authors.

### \*Present address:

Paul-Peter Tak,  
Cambridge University, Cambridge,  
UK; Ghent University, Ghent,  
Belgium; GlaxoSmithKline,  
Stevenage, UK; Clinical  
Unit Cambridge, GlaxoSmithKline,  
Cambridge, UK

### Specialty section:

This article was submitted to  
B Cell Biology,  
a section of the journal  
Frontiers in Immunology

**Received:** 12 October 2016

**Accepted:** 16 February 2017

**Published:** 06 March 2017

### Citation:

Reshetova P, van Schaik BDC,  
Klarenbeek PL, Doorenspleet ME,  
Esveldt REE, Tak P-P, Guikema JEJ,  
de Vries N and van Kampen AHC  
(2017) Computational Model Reveals  
Limited Correlation between Germinal  
Center B-Cell Subclone Abundancy  
and Affinity: Implications for  
Repertoire Sequencing.  
*Front. Immunol.* 8:221.  
doi: 10.3389/fimmu.2017.00221

Polina Reshetova<sup>1,2</sup>, Barbera D. C. van Schaik<sup>2</sup>, Paul L. Klarenbeek<sup>3</sup>,  
Marieke E. Doorenspleet<sup>3</sup>, Rebecca E. E. Esveldt<sup>3</sup>, Paul-Peter Tak<sup>4†</sup>,  
Jeroen E. J. Guikema<sup>5</sup>, Niek de Vries<sup>3†</sup> and Antoine H. C. van Kampen<sup>1,2\*†</sup>

<sup>1</sup> Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands,

<sup>2</sup> Bioinformatics Laboratory, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>3</sup> Amsterdam  
Rheumatology and Immunology Center, Academic Medical Center, Amsterdam, Netherlands, <sup>4</sup> Department of Clinical  
Immunology and Rheumatology, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands,

<sup>5</sup> Department of Pathology, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands

Immunoglobulin repertoire sequencing has successfully been applied to identify expanded antigen-activated B-cell clones that play a role in the pathogenesis of immune disorders. One challenge is the selection of the Ag-specific B cells from the measured repertoire for downstream analyses. A general feature of an immune response is the expansion of specific clones resulting in a set of subclones with common ancestry varying in abundance and in the number of acquired somatic mutations. The expanded subclones are expected to have BCR affinities for the Ag higher than the affinities of the naive B cells in the background population. For these reasons, several groups successfully proceeded or suggested selecting highly abundant subclones from the repertoire to obtain the Ag-specific B cells. Given the nature of affinity maturation one would expect that abundant subclones are of high affinity but since repertoire sequencing only provides information about abundancies, this can only be verified with additional experiments, which are very labor intensive. Moreover, this would also require knowledge of the Ag, which is often not available for clinical samples. Consequently, in general we do not know if the selected highly abundant subclone(s) are also the high(est) affinity subclones. Such knowledge would likely improve the selection of relevant subclones for further characterization and Ag screening. Therefore, to gain insight in the relation between subclone abundancy and affinity, we developed a computational model that simulates affinity maturation in a single GC while tracking individual subclones in terms of abundancy and affinity. We show that the model correctly captures the overall GC dynamics, and that the amount of expansion is qualitatively comparable to expansion observed from B cells isolated from human lymph nodes. Analysis of the fraction of high- and low-affinity subclones among the unexpanded and expanded subclones reveals a limited correlation between abundancy and affinity and shows that the low abundant subclones are of highest affinity. Thus, our model suggests that selecting

highly abundant subclones from repertoire sequencing experiments would not always lead to the high(est) affinity B cells. Consequently, additional or alternative selection approaches need to be applied.

**Keywords:** computational modeling, repertoire sequencing, germinal center, clonal expansion, BCR affinity

## 1. INTRODUCTION

The adaptive immune system is a key component of our defense against pathogens and comprises highly specialized cells and processes. Its humoral component is responsible for memory B-cell formation and high-affinity antibody (Ab) production resulting from affinity maturation in germinal centers (GCs) (1, 2). During this process, GC B cells undergo multiple rounds of proliferation, somatic hypermutation (SHM), and selection to improve their affinity for the given antigen (Ag). This results in a dynamic ensemble of low- and high-affinity B-cell subclones comprising variants of a clone within a V(D)J family produced by SHM. Higher-affinity cells have increased chance to be positively selected for further rounds of proliferation and SHM, or for differentiation to memory cells and plasma cells.

Repertoire sequencing using high-throughput sequencing enables the determination of T- and B-cell repertoires in (clinical) samples by sequencing the expressed V, D, and J gene segments (3–6). Immune responses typically involve the initiation and coexistence of up to several hundreds of GCs, which emerge over an extended period of time (7–9). Consequently, B-cell repertoire sequencing of clinical samples typically identifies (sub)clones originating from a multitude of Ag-activated B cells and GCs or even from responses to multiple Ags. Despite this complexity, we and others successfully used repertoire sequencing for the identification of B cells involved in (auto) immune disorders or infection. One challenge is to select the Ag-specific B cells from the measured repertoire. A general feature of an immune response is the expansion of specific clones resulting in a set of subclones with common ancestry varying in abundance and the number of acquired somatic mutations. These expanded subclones will have BCR-binding affinities for the Ag that are expected to be higher than affinities of the naive B cells in the background population. This is a direct consequence of the higher initial affinities of the activated B cells for the Ag and the subsequent affinity maturation process. For this reason, several groups successfully proceeded or suggested selecting highly abundant subclones from the repertoire (10–15). Given the nature of affinity maturation, one would expect that abundant subclones are of high affinity, but since repertoire sequencing only provides information about abundancies, this can only be verified with additional experiments. Experimental determination of BCR affinities for a reasonable number of subclones is feasible as is, for example, demonstrated by vaccination studies but very labor intensive (5, 16). Since B cells are destructed in the sequencing experiment, affinity analysis requires either selective cloning of the individual B cells or expression of single BCRs in cloning systems. Currently, these requirements prohibit a routine analysis of affinity in BCR repertoire studies. Moreover, affinity measurement requires

knowledge of the Ag, which is often not available for clinical samples. Consequently, we do not know if highly abundant subclone(s) are generally also of the high(est) affinity. We developed a computational model of a single GC to gain insight in the putative affinity distribution among expanded and unexpanded subclones identified by B-cell repertoire sequencing. Inspired by existing models of affinity maturation [e.g., Ref. (17–20)], we implemented a mathematical model that comprises a large evolving set of ordinary differential equations (ODEs) providing information about the abundance and affinity of individual subclones emerging during the GCR. We did not use one of the published models, since existing ODE models do not track individual subclones, while agent-based models [e.g., Ref. (20)] are faced with the additional complexity of GC spatial dynamics, which we aimed to avoid. Moreover, most models are not available as a software implementation.

We show that our computational model is in agreement with typical GC dynamics. We also show that the amount of expansion of selected B-cell lineages from repertoire data acquired from a human lymph node is qualitatively comparable to the level of expansion observed in the simulated data. Given this support for our model, we subsequently inspected the affinities and abundancies of the individual subclones from the simulations and found that the expanded and unexpanded B-cell subclone compartments each comprise a mixture of high- and low-affinity cells, i.e., there is only partial correlation between affinity and abundance of subclones within a clonal family. Moreover, the low abundant subclones were of highest affinity. Although further work is required to experimentally validate these results, our simulations suggest that selection of highly abundant subclones from BCR repertoires will not necessarily lead to the highest affinity subclones. Therefore, additional or alternative selection strategies should be applied.

## 2. MATERIALS AND METHODS

### 2.1. Sample and Experimental Data

We selected a single sample for analysis and comparison to the simulation results. This sample represents leukocytes isolated from a lymph node from an otherwise healthy human individual, without ongoing infection (represented in biochemical parameters such as C-reactive protein). The sample was taken as described earlier (21). The needle biopsy was stored in liquid nitrogen until use. Total RNA was isolated using polytron tissue homogenizer (Kinematica AG, Littau-Lucerne, Switzerland) and AllPrep DNA/RNA mini kit (#80204, Qiagen, Venlo, The Netherlands) according to manufacturer's protocol and stored at  $-80^{\circ}\text{C}$  until further use. The BCR repertoire was analyzed using dedicated primers. This linear amplification protocol has been

extensively described earlier (3, 10, 22). Samples were prepared for sequencing according to the manual for amplicon sequencing (Roche FLX Titanium platform). Study candidates were informed about the background and purpose of the study and the biopsy procedure and possible complications (in particular hematoma). Written informed consent was obtained. Ethics approval was provided by the Ethics Committee Academic Medical Center/University of Amsterdam. Repertoire sequencing resulted in 7,771 reads (6,777 unique reads). Processing of the sequence data was performed as described in Ref. (3). In brief, reads from a multiplexed sequence run were separated by their multiplex identifiers (MID) and aligned against the IMGT database (23) with BLAT (24) to identify the corresponding V and J segments. Subsequently, each read was translated to a peptide sequence, and the CDR3 sequence was determined by identifying conserved motifs in the V and J segment that delineate the CDR3 (25). Consequently, only in-frame reads were used. Sequences with uncalled bases in their CDR3 region were excluded from analysis. This resulted in 4,454 unique subclones (clones within a VJ family defined as a peptide with a unique V and J assignment, and unique CDR3 sequence). This number of sequence reads is sufficient to represent most (expanded) subclones but may miss subclones occurring at very low frequencies. A full analysis and presentation of this and other lymph node samples will be part of future paper.

## 2.2. The Mathematical Model

We developed a mathematical model using ordinary differential equations (ODEs) to describe the dynamics of individual subclones during the GCR. This model is implemented in the R statistical environment version 3.2.2 (26) using R packages deSolve (version 1.12) (27), R6 (version 2.1.2), ggplot 2.0, and beeswarm 0.2.1. The software is available as open source (GPLv3) on request from the author.

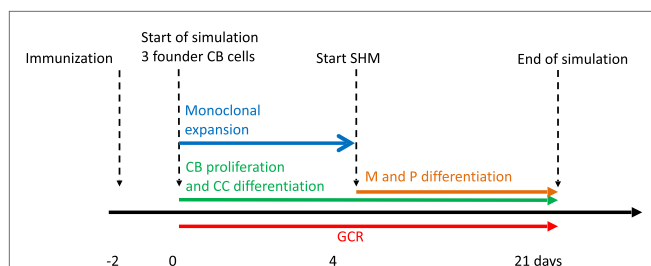
### 2.2.1. Overall Simulation Setup

Our simulation framework represents a simplified but adequate model of the GCR (1, 2) (Figure 1). Briefly, prior to the GCR, B cells and T cells are activated by recognition of their cognate antigen in the primary follicle and T-cell zone, respectively (day -2 in Figure 1). Activated B cells and T cells migrate to the interfollicular region and interact resulting in the full activation

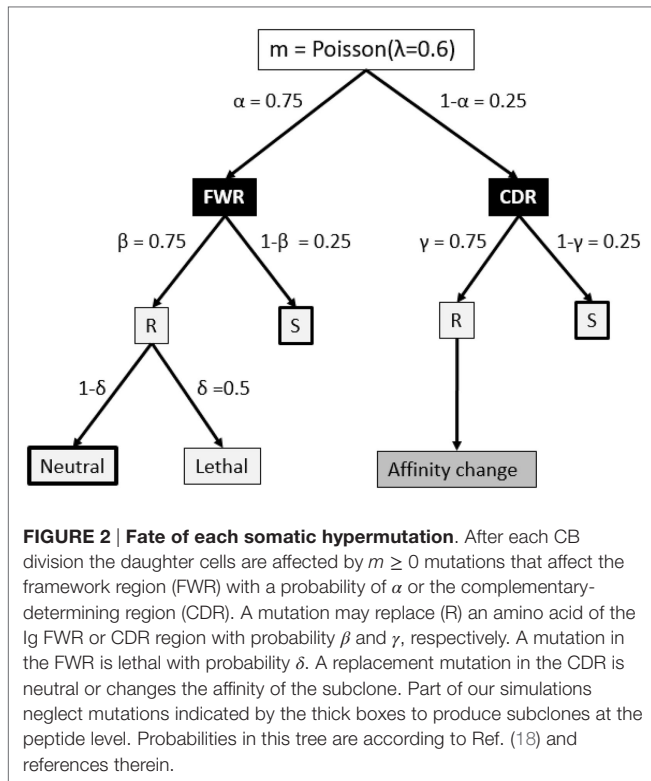
of B cells, while the T cells differentiate to T follicular helper cells (Tfh). Two days after immunization, the GCR is initiated (day 0 in our simulation) with the Tfh cells and activated B cells migrating into the follicle, which is characterized by a network of follicular dendritic cells (FDCs). Here, the B cells engage in a rapid monoclonal expansion to over 10,000 cells at day 4 forming the GC. During this expansion, a dark zone comprising centroblasts (CBs) and a light zone comprising centrocytes (CC), FDCs, and Tfh cells are established. The dark zone is the site of B-cell clonal expansion and BCR diversification through SHM, producing novel subclones. The GC light zone is the site of positive B-cell selection through Ag and Tfh binding and signaling. Together, these processes are responsible for B-cell affinity maturation. SHM has been reported to start at day 7 post-immunization in mice (28). Oprea and Perelson (17) assumed that the GC is initiated 3 days after immunization and, correspondingly, start SHM at day 4 of the GCR in their model. Others reported that SHM starts 2 days post-immunization (29) or even prior to GC formation (30). Following Oprea and Perelson, we also start SHM at day 4 in our simulations. Following monoclonal expansion, memory cells and plasma cells are being produced (day 4 in our simulation). Although the precise mechanisms and timing of the output cells are not well understood (31), it has been proposed that initially memory B cells are produced while at later stages the GCR is dedicated toward (higher-affinity) long-lived plasma cells (32, 33). In our model the production of memory and plasma cells starts at the same moment (day 4), but we made the rate of plasma cell differentiation dependent on the absolute affinity of the CCs resulting in a low plasma cell output during early stages of the GCR. Since we were not interested in the production of output cells, these are not further discussed in this paper. Our simulation starts at day 0 with three founder B cells (CBs) with different affinities and terminates after 21 days, the life span of an average GC. Consequently, we do not model GC shutdown since its mechanisms remain to be established. Our model does not explicitly include the dark/light zones, Ags, FDCs, or Tfh cells since we are neither interested in the spatial dynamics nor in the precise selection mechanisms but rather in modeling subclonal diversity, expansion, and affinity. Therefore, to avoid an overly complex model, we represent the Ag and Tfh survival signals with sigmoidal functions as explained below.

### 2.2.2. Somatic Hypermutation, Subclones, and Affinity

The V, D, and J segments that make up the BCR cover four framework regions (FWRs) providing the Ab structural framework and three Ag-binding complementary determining regions (CDRs) (34, 35). Our model considers the FWR and CDR regions without an explicit nucleotide representation of the BCR, but instead, using a decision tree that decides on the fate of each individual SHM (18, 36, 37) (Figure 2). This tree involves probabilities for silent (synonymous mutations), lethal FWR, and affinity-changing CDR mutations. The probabilities for replacement and silent mutations were determined from many mice germline sequences. The probability of the lethal mutations was based on studies that analyzed mutation patterns in real sequences. To determine the



**FIGURE 1 | Simulation time line of the germinal center reaction (GCR).** The GCR starts with 3 founder B cells (affinities 0.1, 0.2 and 0.3  $\mu\text{M}$ ) 2 days after immunization and continues for 21 days.



number of mutations during each CB cell division, we defined the BCR to have a length of 600 nucleotides (i.e., one light and heavy chain). Given that the SHM rate is  $10^{-3}$  per bp per division this results in 0.6 mutations per division. We model this as a Poisson distribution  $m = \text{Poisson}(\lambda = 0.6)$ , and consequently, each cell acquires 0, 1, or more mutations after each cell division. The mutation decision tree distinguishes the CDRs and their surrounding structural FWRs but does not differentiate between CDR1, CDR2, and CDR3 (34, 35).

In repertoire sequencing one is usually interested determining the population of (sub)clones in an immune response. Each of these subclones has its own binding affinity for the Ag. Since the CDR3 region is the main determinant in Ag-binding, one generally defines and discriminates these subclones on the basis of their unique CDR3 peptide sequence within a VJ family. Alternatively, we can also define a subclone as having a unique BCR nucleotide sequences (i.e., V-CDR3-J). In the first situation, only non-synonymous SHMs in the CDR3 region produce new subclones, while in the second situation each non-lethal SHM results in a new subclone. The mutation decision tree (Figure 2) is defined at the level of the nucleotide sequence, and consequently, in our simulation we implicitly define and track subclones at the nucleotide level throughout the GCR. Consequently, each SHM generates a new subclone that is initially represented as a single CB that subsequently proliferates and differentiates to coexist as CB, CC, memory cell, and plasma cell at succeeding time points. Alternatively, we may consider only CDR replacement mutations to define and track subclones at the peptide level. In this situation, only non-lethal replacement mutations in the CDR generate

new subclones. Since the tree does not specifically distinguish CDR3 from CDR1 and CDR2, our simulations at the peptide level effectively includes all three CDRs, which may give an overestimation of the number of unique clones compared to only considering the CDR3 as is done in repertoire sequencing experiments. However, since all three CDR regions are involved in Ag binding, the simulation might be more realistic. Subclones with CB cell counts less than one (a result from using continuous differential equations; see below) are kept in our simulation but are not further be affected by SHM to avoid the generation of new subclones from these cells.

Each subclone in our model has a unique BCR with an absolute affinity  $\sigma$  that specifies the interaction strength with the Ag. The affinities of the three single cell founder CBs are set to arbitrary but different low-affinity values (0.1, 0.3, and 0.5  $\mu\text{M}$ ). Three different values were chosen to establish an initial level competition between the founder cells. The magnitude of the initial affinities does not affect the dynamics of our model since this depends on relative affinities (see below). Only plasma cell output depends on absolute affinities. For each affinity changing mutation (Figure 2) the affinity of the affected subclone is updated according to  $\sigma_{\text{new subclone}} = \sigma_{\text{parent}} + \Delta\sigma$  where  $\Delta\sigma$  is drawn from a distribution  $f(\sigma)$  with probability density function:

$$f(\sigma) = g(s, r) - \mu - (\sigma_{\text{parent}} * 0.1), \quad (1)$$

where  $g(s, r)$  is the inverse gamma distribution with  $s = 3$  and  $r = 0.3$  representing the shape and rate parameter, respectively.  $\mu$  is the expected value of  $g(s, r)$  and subtracted from  $g(s, r)$  to center the distribution  $g$  around zero resulting in about equal chances for decreasing and increasing the affinity of mutated subclones. We used the gamma distribution because it is right skewed and, therefore, allows for a small chance for making larger affinity improvements representing key mutations (36, 38). We do not distinguish between one or multiple affinity changing mutations. To account for the fact that mutations in higher-affinity subclones have less chance to further improve affinity we shift distribution  $f$  to the left as a function of the parent cell affinity (Figure S2 in Supplementary Material). The distribution shape and rate parameters (3 and 0.3) and the affinity shift (0.1) were chosen by trial and error such as to obtain the dynamics of a typical GC.

### 2.2.3. Positive and Negative Selection of Subclones

Following cell division and SHM, the CBs differentiate to CCs that are programmed to undergo apoptosis (negative selection) unless they receive survival signals (positive selection) through interactions with the Ag (presented by FDCs) and Tfh cells (1). These selection mechanisms impose competition between the B-cell subclones, which are assumed to be based on their relative BCR affinities  $\sigma_{\text{rel}}$  (1). CCs bind Ag to acquire their first survival signal. Subsequently, the Ag is internalized and presented to Tfh cells. Higher-affinity B cells present more Ag and, therefore, compete favorably for the limited number of Tfh cells to acquire a second survival signal. Positively selected CCs recycle to the dark zone for further rounds of division and SHM, or they differentiate into memory cells or plasma B cells.

To avoid an overly complex model, Ag and Tfh survival signals are modeled with a sigmoidal function:

$$S(\sigma_{rel,i}) = \frac{\sigma_{rel,i}^n}{k^n + \sigma_{rel,i}^n}, \quad (2)$$

where  $i$  denotes a subclone. This function converts relative affinities  $\sigma_{rel,i}$  to a signal strength between 0 and 1. Relative affinities are obtained by scaling absolute affinities  $\sigma$  to values between 0 and 1. Signal  $S$  affects the CB to CC differentiation rate ( $\eta_{CB \rightarrow CC}$ ) and the CC apoptosis rate ( $\mu_{CC}$ ) [equations (3a) and (3b)]. Recently, it was shown that higher-affinity cells stay longer in the dark zone further facilitating their expansion and diversification resulting in less apoptosis (39). This is accommodated in our model by multiplying the CB to CC differentiation rate with  $(1 - S)$  resulting in a rate between 0 and its maximum value  $\eta_{CB \rightarrow CC}$  (Table 1). Similarly, a higher signal reduces the apoptosis rate. We assume that  $S$  does not affect these rates to the same extent and, therefore, we parameterized  $S$  separately for differentiation and apoptosis. We set  $k = 0.06$  and  $n = 1$  for differentiation ( $S_d$ ), and  $k = 0.1$  and  $n = 4$  for apoptosis ( $S_a$ ; Figure 3). The parameters  $k$  and  $n$  were chosen to obtain a typical GC response that attains a maximum number of cells during the first phase of the GCR. During our simulation the emergence of new subclones with higher absolute affinity will “push” existing subclones with lower affinities to lower relative affinities as result of the scaling and, hence, to smaller survival signals resulting in the vanishing of these subclones.

#### 2.2.4. Ordinary Differential Equations

Each subclone  $i$  assumes 4 phenotypes: centrocytes ( $CC_i$ ), centroblasts ( $CB_i$ ), memory cells ( $M_i$ ), and plasma cells ( $P_i$ ) (Figure 4). The temporal dynamics of each individual subclone is described by a set of ordinary differential equations (ODEs) representing these four phenotypes [equations (3a)–(3d)].

$$\frac{dCB_i}{dt} = \rho_{CB} \cdot \left( \frac{A^h}{CB_{total}^h + A^h} \right) \cdot CB_i + \eta_{CC \rightarrow CB} \cdot CC_i - (1 - S_d(\sigma_{rel,i})) \cdot \eta_{CB \rightarrow CC} \cdot CB_i \quad (3a)$$

$$\frac{dCC_i}{dt} = (1 - S_d(\sigma_{rel,i})) \cdot \eta_{CB \rightarrow CC} \cdot CB_i - \eta_{CC \rightarrow CB} \cdot CC_i - (1 - S_a(\sigma_{rel,i})) \cdot \mu_{CC} \cdot CC_i - \eta_{CC \rightarrow M} \cdot CC_i - \eta_{CC \rightarrow P} \cdot \sigma_i \cdot CC_i \quad (3b)$$

$$\frac{dM_i}{dt} = \eta_{CC \rightarrow M} \cdot CC_i - \mu_M \cdot M_i \quad (3c)$$

$$\frac{dP_i}{dt} = \eta_{CC \rightarrow P} \cdot \sigma_i \cdot CC_i - \mu_P \cdot P_i \quad (3d)$$

To allow the GC to grow to a sufficient number of cells during monoclonal expansion, the signal  $S_{(d,a)}$  is set to 0.9 for the first 4 days of the simulation to reduce differentiation of CBs to CCs and apoptosis of these initial CCs. The CB equation includes a density-dependent expansion term defining non-specific resource competition between the B cells, reducing their

proliferation rate if the number of cells approaches  $A$ . The CC apoptosis rate and the CB to CC differentiation rate are multiplied by  $(1 - S_{(d,a)}(\sigma_{rel,i}))$  for reasons explained above. Plasma cell differentiation depends on the absolute affinity  $\sigma_i$  to reduce their production at earlier stages of the GCR. During the simulation we calculate the differential equations for periods of 6 h (the duration of one CB division). After each division we impose SHM and update the population of subclones as described above. For each non-lethal SHM, a new subclone and an additional set of four ODEs are created. The CB cell count for new subclones is set to one, while the corresponding cell counts for the CCs, memory cells, and plasma cells are set to zero. The CB cell count of the parent subclone is reduced by one. If the sum of CC and CB counts for subclone  $i$  is less than 0.1 cell, we remove the subclone and corresponding equations from the system. Since SHM is a stochastic process that affects the subclone population and their (relative) affinities, we repeated the simulations 15 times with the same initial conditions (three founder B cells with initial affinities 0.1, 0.3, and 0.5).

TABLE 1 | Model parameters.

B-cell type	Proliferation rate (day <sup>-1</sup> )	Differentiation rate (day <sup>-1</sup> )	Apoptosis rate (day <sup>-1</sup> )
Centroblast (CB)	$\rho_{CB} = 4$ (40–42)	$\eta_{CB \rightarrow CC} = 6$ (1)	
Centrocyte (CC)		$\eta_{CC \rightarrow M} = 1$ (43) $\eta_{CC \rightarrow P} = 0.1$ (43) $\eta_{CC \rightarrow CB} = 1$ (1)	$\mu_{CC} = 4$ (17)
Plasma cell (P)			$\mu_P = 0.25$ (43)
Memory cell (M)			$\mu_M = 0.01$ (43)
<b>Other parameters</b>			
Capacity $A = 8,000$	$k = 0.06, n = 1$ ( $S_d$ )	$s = 3.0$	
Number of founder cells: 3	$k = 0.1, n = 4$ ( $S_a$ )	$r = 0.3$	
Initial affinities: 0.1, 0.3, 0.5 $\mu M$	$h = 20$	Affinity shift = 0.1 $\mu M$	

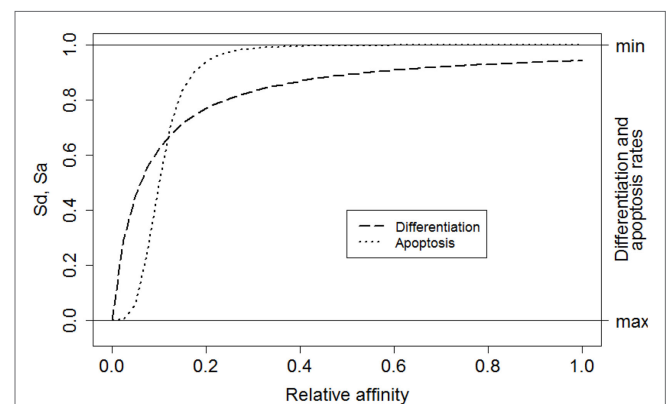
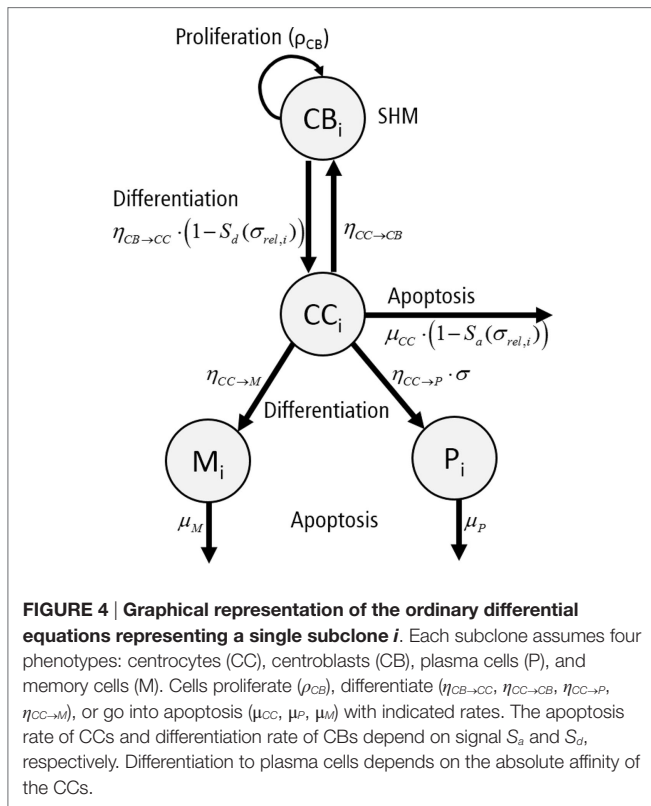


FIGURE 3 | The relative affinity of each subclone determines the magnitude of the overall survival signal. Left axis:  $S_d$  corresponds to signal affecting the CB to CC differentiation rate (dashed line).  $S_a$  corresponds to the signal affecting CC apoptosis (dotted line). Right axis: effect of signal  $S$  on the differentiation and apoptosis rates. A high signal results in low differentiation and apoptosis rates.



### 2.2.5. Model Parameters

Parameter values for proliferation, differentiation, and apoptosis were obtained from literature (Table 1). Parameters  $A$  and  $h$  were chosen to limit the maximum size of the GC. Values for parameters  $k$ ,  $n$ ,  $s$ ,  $r$ , and affinity shift were acquired by trial and error aiming to produce a typical GC response with a peak of at least 10,000 cells during the first phase of the GCR. There are very limited (quantitative) data describing the GC response. We are not aware of any data obtained from human samples describing the dynamics of GC volume (number of cells) during the GCR. Consequently, the precise timing and magnitude of the maximum GC response, its decay, the biological variation of this response across samples and organisms, and the factors affecting this response remain to be established. The canonical GC response has, for example, been observed by tracking follicle center volume as fraction of total splenic volume in mice (41) or as fraction of the total volume of the GC in rat (44), which may be used as GC cell count substitutes. These volumes showed a peak during the first phase of the GC. Such measurements have been used previously to validate a GC model (20). However, other studies showed that there might not exist a typical GC in terms of size (45), and that GCs in a single immune response might not be synchronized (8). The lack of precise quantitative data, current uncertainties in GC dynamics, and our decision not the model GC termination limits the possibilities and value of a compute-intensive parameter inference strategy to obtain values for the aforementioned parameters. However, instead of our trial and error approach, Approximate Bayesian

Computation algorithms (46), MEANS (47), or other methods may be used to fit parameters on complex stochastic models such as ours.

### 2.3. Identification of Expanded Subclones

To determine a threshold that identifies expanded subclones we follow an approach that is similar to the method we applied in our previous repertoire sequencing studies, e.g., Ref. (3, 10). First, a histogram of counts  $c$  (cell counts for simulated data and read counts for experimental data) for all (un)expanded subclones is constructed to reflect their cell/read count frequencies  $F(c)$  (Figure S3 in Supplementary Material). In general, subclones with low counts (e.g.,  $c = 1$ ) occur much more frequently (high  $F$ ) than subclones with high count (e.g.,  $c = 100$ ). Next, we define  $T$  as lowest count  $c$  for which  $F(c) = 0$ . That is, no subclones with  $c$  cells/reads are observed. We assume that  $F(c \geq T) = 0$  for the underlying but unknown null distribution of unexpanded subclones. We define subclones with  $c > T$  ( $F(c) \geq 1$ ) to be expanded. That is, subclones observed with cell/read counts  $c > T$  are larger than expected based on the distribution of unexpanded subclones. The threshold  $T$  is stringent but could be relaxed by defining the threshold  $T$  as the lowest count  $c$  for which  $F(T) < p$ , with  $p \geq 1$ .

The expansion threshold  $T$  was estimated for each individual simulation. We assumed that repertoire sequencing experiments measure mainly CCs since CBs do not, or at very low levels, express BCRs. Consequently, for the simulated data we determine threshold  $T$  from CC cell counts only. CC cell counts were taken from the last time point of the simulation.

### 2.4. Comparison of Simulated and Experimental Data

We qualitatively compare subclone cell counts from our simulations to read counts from a single sample repertoire sequencing experiment. Since our computational model does not explicitly represent the BCR as a nucleotide (or protein) sequence we do not consider multiple (back) mutations occurring at previously mutated positions. Consequently, the number of different mutations and, hence, subclones in our simulation is slightly overestimated.

Each unique nucleotide read obtained from repertoire sequencing (RNAseq) can be considered as a unique subclone representing a set of mutations acquired during affinity maturation. Statistics calculated for these subclones can be compared to statistics calculated for the nucleotide-level subclones generated in our simulations. Alternatively, we can define subclones measured in the sample at the peptide level as having unique combination of V and J segments (determined by alignment) together with a unique CDR3. The peptide level definition allows to compare the statistics from the experimental data to the peptide-level simulations. In contrast to subclones analyzed at the nucleotide-level, this definition considers any mutations in the CDR3 for the experimental data and any affinity changing mutation in CDR1,2,3 for the simulated data.

Repertoire sequencing experiments performed on tissue (e.g., lymph node) generally results in a set of subclones from multiple

GCs and, most likely, from different Ag responses, while in our simulation we generate subclones from a single GCR initiated by three founder clones. We account for this by selecting subclones corresponding to three lineages from sample LN25. We first map all reads against reference sequences extracted from the IMGT database to determine their V and J segments. Subsequently, combinations of V and J are counted, and reads corresponding to the three most abundant V–J combinations (V3.7–J4, V3.74–J4, and V3.23–J4) are selected. The resulting three groups of reads still comprise subclones from multiple lineages. Therefore, we subsequently aligned all pairs of reads within each V–J group to determine the number of nucleotide differences (mutations) between them. Each pair of reads with two or fewer differences is connected to form clusters of subclones that are assumed to belong to the same lineage. Finally, the largest cluster (lineage) for each V–J combination was selected. The same procedure was followed at the peptide level. Since these three clusters did not include the most abundant subclone, we also selected the second largest cluster from the V3.7–J4 subclones. The results of this procedure are shown in **Table 2**. Note that the number of differences between pairs of not connected reads within a cluster (lineage) may be larger than 2. These clusters of reads could in principle be subjected to further phylogenetic analysis to determine a lineage tree establishing their relationships (48).

### 3. RESULTS

First, we confirm that the computational model produces the dynamics of a typical GC response that currently is not very well defined as discussed in the method section. We performed 15 repeated simulations with subclones defined at the nucleotide level. In agreement with previous work, the GC response peaks

around day 8 (**Figure 5A**) (41, 44, 45). The size of the GC reaches approximately 14,000 cells, which is in agreement with estimations from histological sections of two GCs (49). The CB to CC ratio (not shown) after day 8 remains between 1.4 and 2.0 and is in agreement with data obtained from intravital microscopy (50). The maximum number of SHMs in subclones emerging from our simulation ranges from 4 (day 10) to 11 (day 21) and is in good agreement with the 9 somatic mutations found in a single Ab after affinity maturation (51); with the 8 to 18 mutations found in an analysis of BCR sequences obtained from cells from GC sections derived from human lymph nodes (49); and with the 4 to 9 mutations observed in B cells from single GCs obtained from mice lymph nodes (52). Monoclonal expansion of the 3 founder cells results in many low-affinity subclones at the initial GC stage, but gradually higher-affinity clones start to appear and outcompete lower affinity subclones. As expected from affinity maturation, and in agreement with other computational models [e.g., Ref. (17, 20)], the subclone population evolves to higher affinities (**Figure 5B**). The drop in the CB cell count after day 4 is caused by the initiation of SHM and the subsequent differentiation to CCs that may go into apoptosis. Since we do not model GC shutdown, the cell counts remain relatively stable after 14 days. These results show that our computational model adequately captures the dynamics of a typical GCR.

#### 3.1. Subclonal Diversity

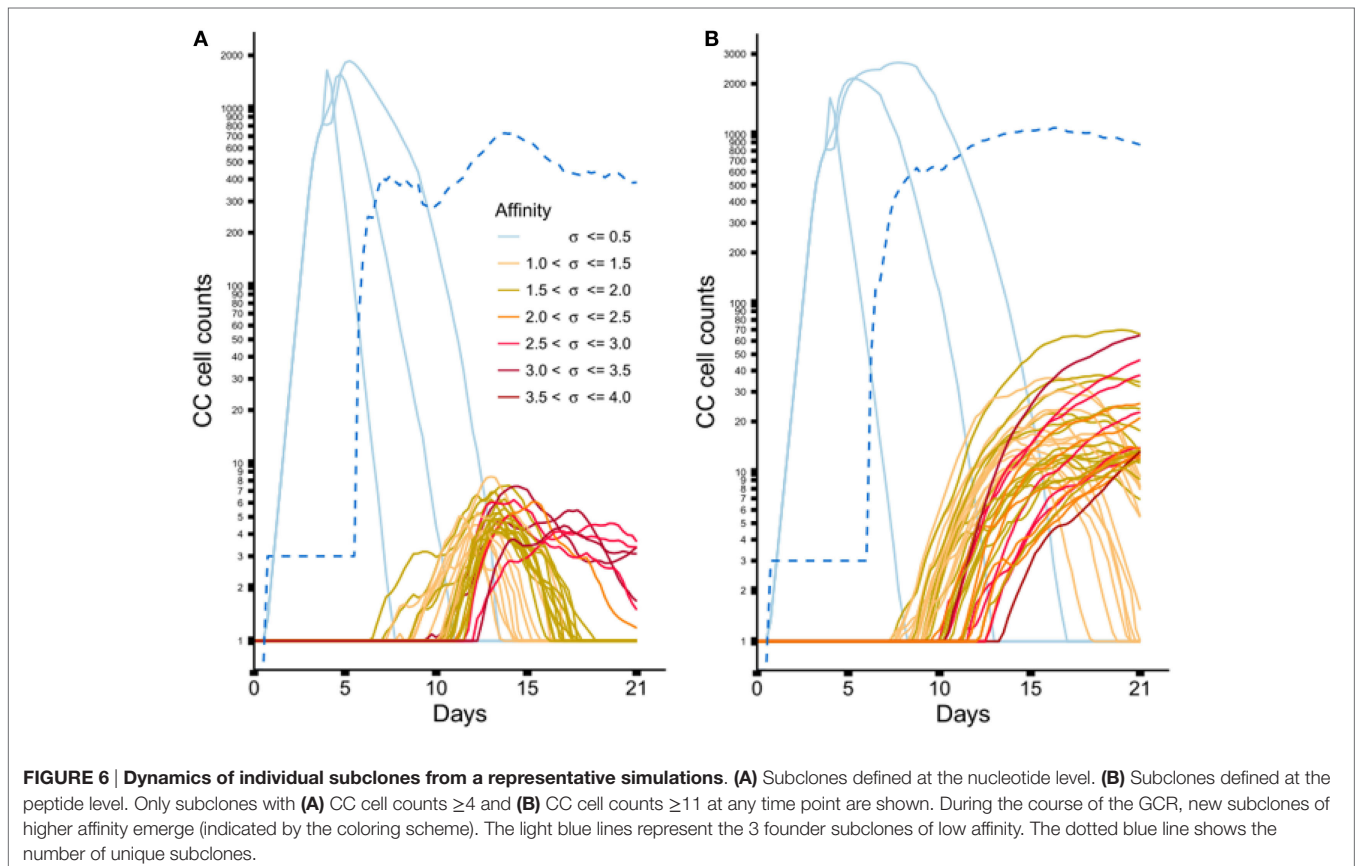
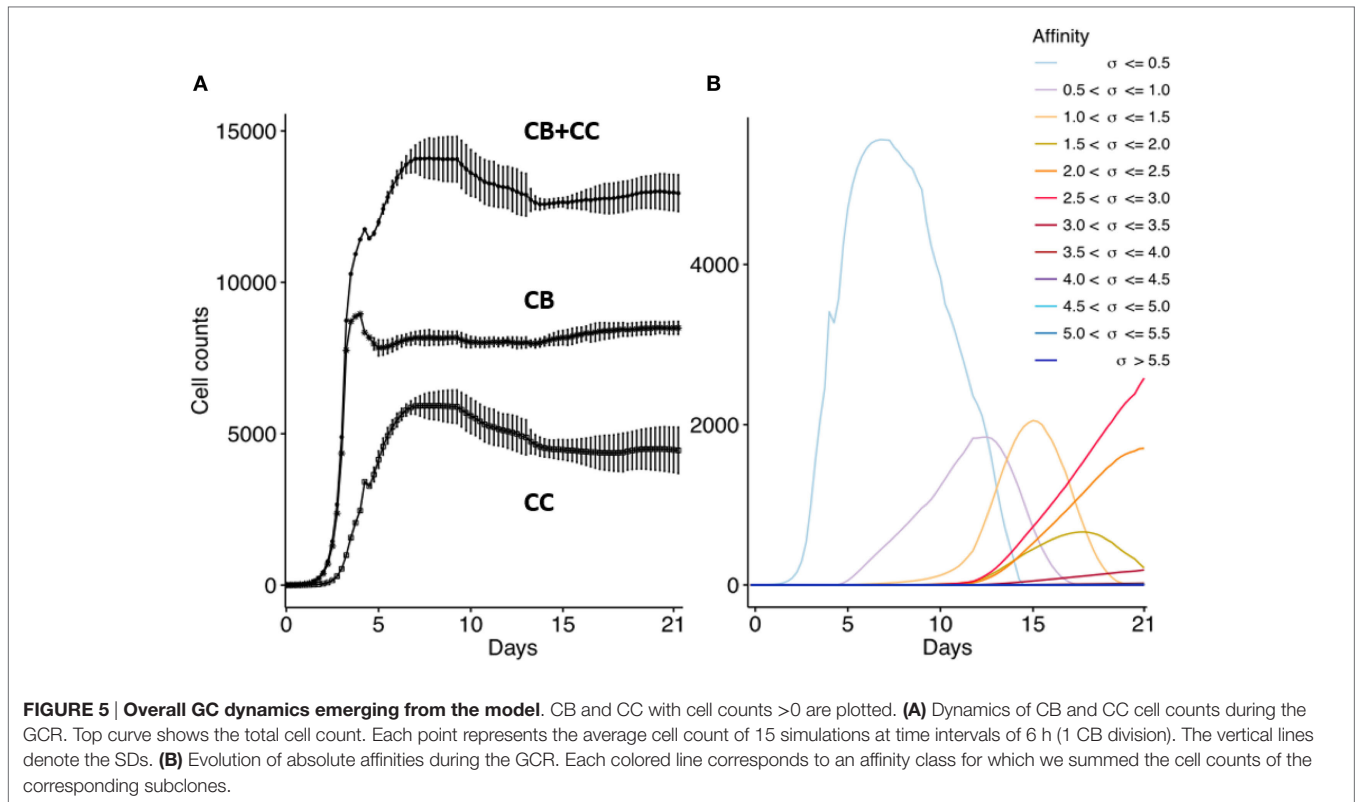
**Figure 6** shows the dynamics of individual subclones during the GCR at the nucleotide and peptide level. Initially, 3 founder clones expand monoclonally until day 4 after which SHM is initiated and new subclones with higher affinities start to be produced. The three low-affinity founder subclones reach high cell counts since, during monoclonal expansion, no lethal SHM occurs and  $S_{(d,a)}$  assumes a large value (0.9) resulting in a very low rate of CB differentiation and CC apoptosis. New (higher-affinity) subclones realize much lower cell counts because they start as single proliferating cells but are also reduced in count due to new mutation events and apoptosis as a result of competition with higher-affinity subclones. Interestingly, although the population of subclones evolves to higher affinities (**Figure 5B**) there is neither a single nor a small set of subclones that dominates this population during the later stages of the GCR. In fact, the number of unique subclones (**Figure 6A**) remains around 550 during the second half of the GCR.

From sample LN25 we identified 112 nucleotide-level defined subclones (i.e., unique sequence reads) corresponding to 300 reads in the three largest lineages (**Table 2**). Since multiple sequence reads may originate from a single B-cell it is not possible to scale these numbers to 14,000 GC cells but obviously 300 reads do not represent this many GC cells. Therefore, these 112 subclones underestimate the true number of subclones in a single GC. Although this number does not provide a validation for the 550 subclones observed in our simulations, it does show that the diversity of subclones in the experiment and the simulations is high. Using multiphoton microscopy and sequencing, it was recently shown that efficient affinity maturation can occur without homogenizing selection, and that loss of clonal diversity during the GCR varies widely from one GC to the other (52). Note

**TABLE 2 | Selected B-cell subclones from sample LN25.**

Subclone	Total		Largest cluster (lineage)	
	Subclones	Reads	Subclones	Reads
V3.7–J4 (nt)	171	334	84	232
V3.74–J4 (nt)	125	249	23	56
V3.23–J4 (nt)	37	60	5	12
Total (nt)	333	643	112	300
V3.7–J4 (pep)	89	606	19	36
V3.74–J4 (pep)	97	519	9	14
V3.23–J4 (pep)	76	193	7	12
Total (pep)	262	1,318	35	62
			Second largest cluster (lineage)	
V3.7–J4 (pep)	89	606	9	417

*V and J nomenclature follow IMGT (23, 34). Subclones are defined as unique nucleotide sequences (nt) or as peptides (pep) with unique V and J assignment and a unique CDR3 sequence. For each V–J family, the number of subclones and corresponding number of sequence reads are shown. The selected clusters for the given V–J segments correspond to the largest cluster of subclones having  $\leq 2$  differences at nucleotide or peptide level. For V3.7–J4, the second largest cluster, which contains the most abundant subclone, is also included.*

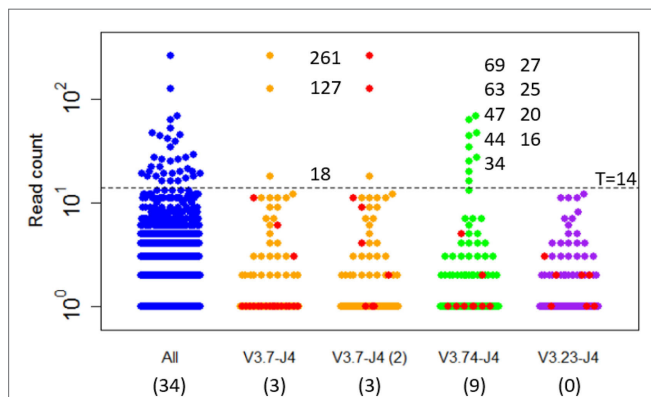




that when comparing **Figure 6A** (nucleotide level) to **Figure 6B** (peptide level), the overall dynamic behavior is similar but the cell counts of higher-affinity peptide-level subclones are about five times larger. An increase in cell count is expected since, in this scenario, neutral and synonymous somatic mutations do not result in new subclones and, hence, no reduction of cell counts occurs. The number of unique subclones is still in the same order of magnitude as the previous simulation but counterintuitively increased compared to previous situation since a decrease is expected due to the fewer mutations imposed on these subclones. The observed increase is, however, a result of plotting and summing only the subclones with CC cell counts  $\geq 1$ . Including cell counts  $< 1$  shows that the number of subclones indeed decreases (data not shown).

### 3.2. Subclonal Expansion

Expanded subclones are derived from experimental data on the basis of their peptide-level definition and relative abundance. Basically, this definition neglects any mutation in the V and J region and any synonymous mutations in the CDR3. We identified expanded subclones from the experimental data (**Figure 7**). First, the expansion threshold was determined using all subclones from the LN25 sample resulting in 34 expanded subclones. Using this threshold ( $T = 14$ ), a total of 3 and 9 subclones from the V3.7-J4 and V3.23-J4 subclones, respectively, are expanded. For each V-J family, **Figure 7** also shows the subclones corresponding to the largest cluster (read counts ranging from 1 to 11), and for V3.7-J4, the subclones corresponding to the second largest cluster (read counts ranging from 1 to 261). This shows that subclones within a B-cell lineage may exhibit a wide range of read counts, which is in agreement with our simulated data. It also shows that the most abundant subclones do not necessarily belong to the largest cluster within a V-J family.



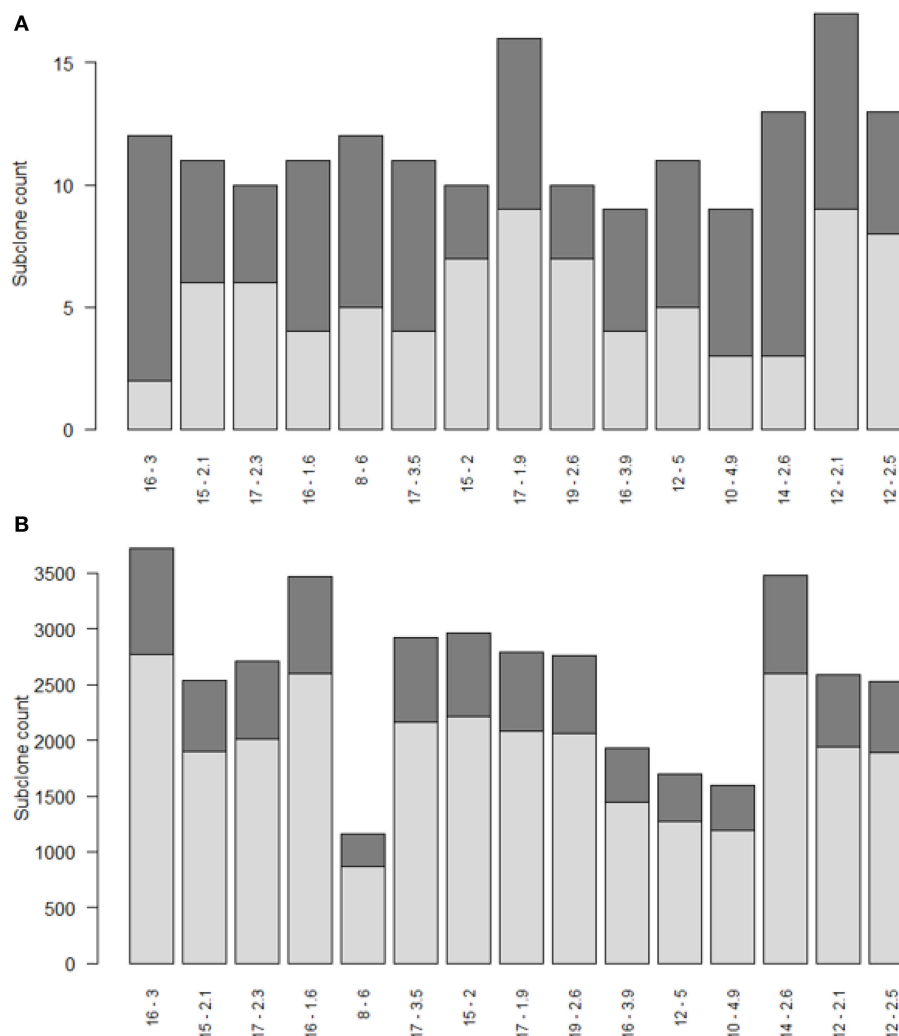
**FIGURE 7 | Subclones measured in a lymph node sample (LN25) from a healthy individual.** The blue points show the read counts for all 4,454 subclones measured in this sample (34 expanded subclones). The expansion threshold ( $T = 14$ ) is determined from the all LN25 subclones and indicated by the dashed line. Subclones of the three most abundant V-J combinations are shown in orange, green, and purple. The red dots indicate the subclones of the largest clusters and, for V3.7-J4, also the second largest cluster. Read counts of the expanded subclones are shown. The numbers in the parenthesis show the number of expanded subclones in the selected V-J subsets.

The clonal size (number of reads of a subclone divided by total number of reads) of the expanded LN25 subclones varies from 0.2 to 3.4%. Together, these represent 0.8% (34 out of 4,454) of all subclones. This is similar to the amount of expansion found in one of our previous studies where clonal sizes  $\geq 0.5\%$  were found to represent expanded subclones representing 0.3 and 1.9% of the subclones in peripheral blood and synovial tissue of RA patients, respectively (10). Since our computational model does not explicitly consider V and J segments, and because we cannot distinguish CDR3 from CDR1 and CDR2 mutations, we cannot group subclones resulting from our simulation in a way similar to the experimental data. However, by neglecting neutral and silent FWR/CDR mutations we can simulate subclones at the peptide level. The resulting subclones differ only in their CDR regions. The expanded peptide-level subclones in our simulation represent clonal sizes ranging from 0.3 to 8.7% representing 0.3 to 1.0% of the subclones. This degree of expansion is in the same order of magnitude as expansion observed in our experimental data.

### 3.3. BCR Affinity of (Un)Expanded Subclones

Repertoire sequencing provides only information about the relative abundance of B-cell subclones in a sample. In contrast, our computational model provides also information about the (relative) affinity of each subclone, which we use to gain insight in the affinity distributions among expanded and unexpanded subclones. High absolute affinity was defined by setting a threshold at the 75th percentile of absolute affinities of all subclones produced during the course of the GCR (range 1.53–10.6; 75th percentile is 3.00). **Figure 8** shows the number of high- and low-affinity subclones among (un) expanded subclones for each of the 15 simulations. In these simulations we defined the subclones at the peptide level. In these 15 simulations, the percentage of low-affinity subclones among high-abundant subclones varies from 17 to 70%. The percentage of high-affinity subclones among the low abundant subclones is about 25% in each of the simulations. In 14 out of 15 simulations, the affinity of most abundant subclones belongs to the highest 25% of affinities (**Figure 9A**), but the most abundant subclones never correspond to the highest affinity subclone (**Figure 9B**). **Figure 9B** shows that the affinity tends to increase with subclone abundance (spearman rank correlation is 0.6) but that the largest affinities correspond to the low abundant subclones. Increasing the affinity threshold to 95% results in more low-affinity subclones among the expanded subclones (data not shown).

Although the affinity distribution depends on the expansion and affinity thresholds, the results demonstrate that lower affinity cells will be among the expanded subclones and *vice versa*. However, in a repertoire sequencing experiment one might not detect the very low abundant (high-affinity) subclones. The high-affinity cells in the unexpanded fraction are either new subclones that have undergone significant affinity improvement but did not yet have sufficient time to proliferate or are high-affinity subclones previously expanded but now being outcompeted by new subclones.



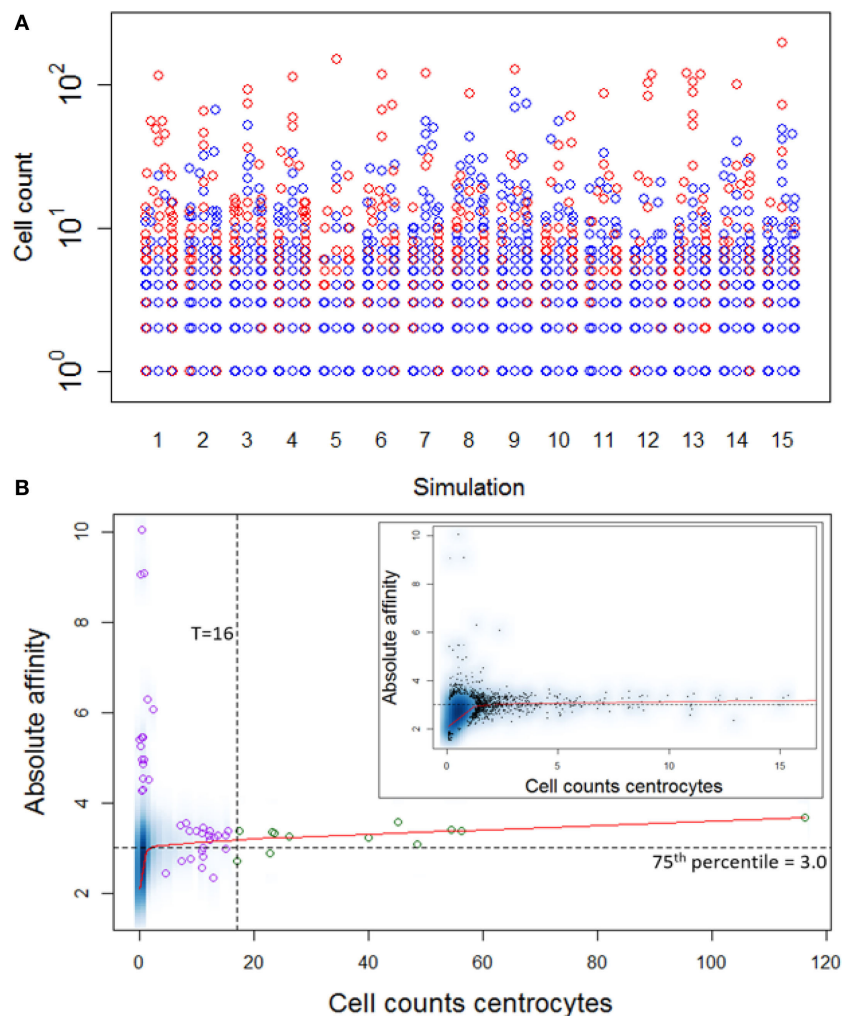
**FIGURE 8 | Numbers of high (dark gray) and low (light gray) affinity subclones among expanded (A) and unexpanded (B) subclones in 15 simulations (x-axis).** Subclones were defined at the peptide level. There are many more unexpanded subclones compared to expanded subclones. Subclones with CC cell counts >0 were included. The numbers at the x-axis denote the thresholds for expansion ( $T$ ) and absolute affinity (75th percentile).

## 4. DISCUSSION

The identification of autoreactive B cells is important for understanding the pathogenesis of autoimmune diseases and developing therapies that target specific B cells to improve clinical outcome. However, for many autoimmune disorders the Ags are unknown, which makes screening approaches challenging. Repertoire sequencing strategies have been developed as an alternative Ag-agnostic approach to identify autoreactive B cells while relying on the assumption that expanded B cells measured in blood or tissue are involved in the pathogenesis of the disease. B-cell subclones identified by sequencing can be cloned and functionally characterized and used to identify the autoantigen. In previous work we demonstrated that expanded clones identified by repertoire sequencing of synovium samples from RA patients point to putative autoreactive B cells (10). This

potentially provides the opportunity to develop novel therapeutic approaches targeting these cells.

(Deep) Repertoire sequencing is successfully used for the identification of antigen-specific B cells involved in immune disorders or infection. Since clonal expansion is a general feature of an immune response, the selection of highly abundant subclones to identify the Ag-specific B cells has been suggested and used by several groups (10, 12–15). Selected subclones can subsequently be characterized or can facilitate the identification of the Ag (11, 53). However, repertoire sequencing itself provides no information about affinity. Therefore, we developed a computational model to investigate the relation between subclone abundance and affinity. Although our computational model was not expected to provide precise quantitative results, we showed that the fraction of low-affinity cells among expanded subclones and the fraction of high-affinity subclones among unexpanded B



**FIGURE 9 | (A)** Distribution of high-affinity subclones (red) among all subclones for 15 simulations. **(B)** Density plot of CC cell counts and absolute affinity for simulation 1. Inset shows only the low abundant subclones. Data points show a selection of subclones imposed on the density plot. Green points denote the expanded subclones. Purple points indicate a selection of low abundant subclones. The red line shows a lowest regression to indicate the overall relation between abundance and affinity.

cells are substantial (Figure 8). There exists a moderate positive correlation between subclone abundance and affinity, but we also showed that the highest affinity subclones are of very low abundance (Figure 9). We performed a sensitivity analysis to assess the robustness of the model for changes in model parameters (see Supplementary Material). This analysis showed that changes in parameter values have mostly small to moderate effects on the output of the model but do not change the main results of the simulations. Therefore, based on our simulations, we conclude that selection of highly abundant subclones may not necessarily lead to the high(est) affinity B cells. We realize that this conclusion will need further experimental validation by simultaneous measurement of abundance and affinity although this will remain difficult for clinical samples. Evolving experimental technologies and approaches may make this less labor intensive in the future. Using a tractable immunization mouse model and a well-defined

Ag might be a first step toward validation. In this case, a single cell strategy is required to sequence both the heavy and light Ig chains. Subsequently, the Igs must be cloned and expressed followed by measuring antibody–antigen binding kinetics using surface plasmon resonance (54).

However, to support the results from the simulations we analyzed a lymph node sample (LN25) to determine (i) the variability of subclone frequencies within a lineage, (ii) the number of expanded subclones, and (iii) the subclonal diversity. Although each of these analyses supported the simulation, the LN25 sample could have contained plasma cells (PCs) that express as much more RNA than GC B cells (55) thereby distorting the relationship between sequence read counts and cellular abundances. We could not verify this experimentally since no sample was left available. In addition, we did not account for possible sequencing and PCR errors using molecular barcodes or

other methods (56–60). Such errors may result in (low abundant) artificial subclones that potentially affect the interpretation of the data analysis. PCs and sequencing/PCR errors can affect the variation subclone frequencies that we observe in the data, but it is unlikely that all variability is explained by these artifacts as it is an intrinsic characteristic of affinity maturation. The number of expanded subclones can be inflated by PCs, but it is unlikely that all subclones above our threshold are PCs and, hence, the percentages of expanded clones determined from the data and simulations would probably stay comparable if PCs would be removed. Sequencing/PCR errors artificially increase subclonal diversity but naively extrapolating the number of subclones observed in the selected LN25 VJ families to 14,000 GC cells in the simulation demonstrates that removing these errors would in fact bring the observed variability closer to the simulation results. Moreover, it is unlikely that the creation of these artificial low abundant subclones largely effect the frequencies of the expanded subclones from which they could have originated. Thus, we expect that sample LN25 after accounting for PCs and/or sequencing/PCR errors, would still support the simulation results. Moreover, note that we do not use the sample to directly verify the relationship between cell count and affinity, which is solely derived from the simulations. However, to further validate our simulations we selected and analyzed a public BCR repertoire data set obtained from a cervical lymph node from a chronic multiple sclerosis patient (61). In this sample no CD188+ PCs were observed. Results from this analysis are similar as those obtained from the LN25 sample and also support our simulation results (see Supplementary Material).

In addition to merely selecting highly abundant subclones from B-cell repertoires, there exist alternative selection strategies that can be used to identify the Ag-specific B cells (6). For example, it has been shown that representative Abs selected from clonal families, reconstructed by phylogenetic analysis, neutralize influenza more effectively than “singleton” Abs that use heavy-chain V(D)J and/or light-chain VJ gene segments that are not used in any other Ab in the repertoire (5). This study showed that Abs from clonal families have significantly higher affinities than did singleton antibodies. Such strategy could be combined with subclone abundance. In previous work we have shown that the identification of pathogenic subclones in RA benefits from the selection of high-abundant subclones that are present in multiple joints within a patient (10, 22). It would be interesting to determine the affinity of these overlapping subclones in comparison to high-abundant non-overlapping clones. Marcatili and coworkers (62) used BCR repertoires from a large number of CLL patients to cluster the receptors into groups with similar sequence properties that potentially can be used for prognostics. Alternatively, one can compare repertoires across patients to identify consistent Ab sequence features (63). One might be tempted to identify high-affinity clones by selecting the clone with the highest number of somatic mutations since multiple rounds of proliferation, SHM, and selection increases the overall affinity of the GC B-cell population, but such correlation between affinity and number of mutations is not observed in our simulation results (data not shown). In addition, mutation and affinity

measurements the study of Tan and coworkers did also not reveal such correlation (5).

Surprisingly, our model shows that the number of unique subclones in a single GC remains remarkably constant throughout the GCR and does not evolve to a single or few high-affinity dominating subclones although the affinity of the population as a whole increases as has been shown in previous studies (17, 20). Moreover, the cell counts of individual subclones remain very low. Adding additional mechanistic detail (e.g., GC shutdown) is unlikely to change this observation. Moreover, this observation is in agreement with repertoire sequencing data and also seems in agreement with a recent study that showed that many clones may mature in parallel, and sporadic clonal bursts generates many SHM variants of a clone (52).

The parameters used in the model (Table 1) and mutation probabilities in the decision tree (Figure 2) are not all based on recent experiments and data. Moreover, we have taken parameters and other information (e.g., the typical GC response) mostly from mouse studies since human data are often not available. It might therefore be necessary to use more up-to-date experimental approaches and/or (public) data to revisit these parameters and probabilities in order to establish them with more precision, under a variety of different conditions, and for different organisms. For example, the immunogenetics community made large progress in establishing comprehensive immunoglobulin V-gene databases. Recent work based on these data showed that there exist large differences between different mouse strains, and that mouse repertoire is more germline-focused than the human repertoire suggesting that affinity maturation is less important for mouse than it is for human (64, 65). Therefore, a typical GC response for human might be different from mouse. It also suggests that our mutation probabilities that were based on limited mouse data generated prior to 1998 (18) need to be updated as part of future work.

Our model can be improved in several ways. Given the current results it would be interesting to investigate if our results hold with more detailed GC models since with the current model it is very difficult to control the amount of expansion by changing the sigmoid functions without distorting the overall GC dynamics (although this might happen also *in vivo*). It would be interesting to investigate what exactly controls selection pressures and how this affects subclonal expansion and the BCR affinity distribution. Nevertheless, as we have shown, the current magnitude of expansion observed from the model is in the same order of magnitude as observed in experimental data from LN25. However, the amount of expansion shown in sample M5 (Supplementary Material) can currently not be reproduced with the model. To allow a better comparison to the experimental data we plan to include an explicit representation of the BCR as a nucleotide sequence in our future model. This would allow to distinguish between the different CDR regions, to account for multiple (back) mutations at identical positions, and to more precisely specify subclones at both the nucleotide and protein level. In analogy to Ref. (7, 66), this would allow to explore the clonal composition and subclonal dynamics in a system where the BCR sequence with the highest Ag affinity is known and can be reached in few (key) mutations such as in the response against (4-hydroxy-3-nitrophenyl)acetyl.

However, in general, the incorporation of realistic affinities in GC models will remain a challenge until these really can be measured on a large scale. Another interesting extension would include the egress of B cells to investigate the (sub)clonal composition in blood and to compare this to repertoire sequencing data obtained from blood samples. Finally, we did not model GC shutdown since its mechanism is unknown, although mechanisms have been hypothesized such as Ab feedback (67). It is difficult to predict how GC shutdown would affect the results presented in this paper since this would depend on the timing that such mechanism would affect the different cell types and how it would differentiate between subclones with different abundance and affinity.

## AUTHOR CONTRIBUTIONS

NV and AK designed the study. PR, AK, JG, and PK defined the model. PR performed the simulations. PR and AK analyzed the results from the simulations. P-PT developed the protocol for acquiring LN samples and provided the sample used in this study. RE conducted the repertoire sequencing experiment. PK, MD, BS, AK, and NV analyzed the experimental data. PR and AK wrote the manuscript. All authors critically read, contributed, and approved the manuscript.

## REFERENCES

1. Victora GD, Nussenzweig MC. Germinal centers. *Annu Rev Immunol* (2012) 30(1):429–57. doi:10.1146/annurev-immunol-020711-075032
2. De Silva NS, Klein U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* (2015) 15(3):137–48. doi:10.1038/nri3804
3. Klarenbeek PL, Tak PP, van Schaik BDC, Zwinderman AH, Jakobs ME, Zhang Z, et al. Human T cell memory consists mainly of unexpanded clones. *Immunol Lett* (2010) 133(1):42–8. doi:10.1016/j.imlet.2010.06.011
4. Calis JJA, Rosenberg BR. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends Immunol* (2014) 35(12):581–90. doi:10.1016/j.it.2014.09.004
5. Tan YC, Blum LK, Kongpachith S, Ju CH, Cai X, Lindstrom TM, et al. High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clin Immunol* (2014) 151(1):55–65. doi:10.1016/j.clim.2013.12.008
6. Yaari G, Benichou JIC, Vander Heiden JA, Kleinstein SH, Louzoun Y. The mutation patterns in B-cell immunoglobulin receptors reflect the influence of selection acting at multiple time-scales. *Philos Trans R Soc Lond B Biol Sci* (2015) 370(1676):20140242. doi:10.1098/rstb.2014.0242
7. Kleinstein SH, Singh JP. Toward quantitative simulation of germinal center dynamics: biological and modeling insights from experimental validation. *J Theor Biol* (2001) 211(3):253–75. doi:10.1006/jtbi.2001.2344
8. Wittenbrink N, Weber TS, Klein A, Weiser AA, Zuschratter W, Sibila M, et al. Broad volume distributions indicate nonsynchronized growth and suggest sudden collapses of germinal center B cell populations. *J Immunol* (2010) 184(3):1339–47. doi:10.4049/jimmunol.0901040
9. Or-Guil M, Wittenbrink N, Weiser AA, Schuchhardt J. Recirculation of germinal center B cells: a multilevel selection strategy for antibody maturation. *Immunol Rev* (2007) 216(1):130–41. doi:10.1111/j.1600-065X.2007.00507.x
10. Doorenspleet ME, Klarenbeek PL, de Hair MJH, van Schaik BDC, Esveldt REE, van Kampen AHC, et al. Rheumatoid arthritis synovial tissue harbours dominant B cell and plasma cell clones associated with autoreactivity. *Ann Rheum Dis* (2014) 73(4):756–62. doi:10.1136/annrheumdis-2012-202861

## ACKNOWLEDGMENTS

Prof. Dr. Age Smilde (Biosystems Data Analysis Group) is acknowledged for critically reading and improving the manuscript. Dr. Huub Hoefsloot, Dr. Johan Westerhuis (Biosystems Data Analysis Group) and Prof. Rogier Sanders (Laboratory for Experimental Virology) are acknowledged for their suggestions during this research. The work will be included in the Ph.D. thesis (Integration of prior knowledge into biological systems modeling) of PR (68).

## FUNDING

This work was carried out on the Dutch national e-infrastructure of SURFsara with the support of SURF Foundation. Research was supported by the Netherlands Bioinformatics Center (NBIC).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fimmu.2017.00221/full#supplementary-material>.

11. Robinson WH. Sequencing the functional antibody repertoire – diagnostic and therapeutic discovery. *Nat Rev Rheumatol* (2015) 11(3):171–82. doi:10.1038/nrrheum.2014.220
12. Hershberg U, Meng W, Zhang B, Haff N, St Clair EW, Cohen PL, et al. Persistence and selection of an expanded B-cell clone in the setting of rituximab therapy for Sjögren's syndrome. *Arthritis Res Ther* (2014) 16(1):R51. doi:10.1186/ar4481
13. Galson JD, Kelly DF, Truck J. Identification of antigen-specific B-cell receptor sequences from the total B-cell repertoire. *Crit Rev Immunol* (2015) 35(6):463–78. doi:10.1615/CritRevImmunol.2016016462
14. Hershberg U, Luning Prak ET. The analysis of clonal expansions in normal and autoimmune B cell repertoires. *Phil Trans R Soc B* (2015) 370:20140239. doi:10.1098/rstb.2014.0239
15. Maillette de Buy Wenniger LJ, Doorenspleet ME, Klarenbeek PL, Verheij J, Baas F, Elferink RP, et al. Immunoglobulin G4+ clones identified by next-generation sequencing dominate the B cell receptor repertoire in immunoglobulin G4 associated cholangitis. *Hepatology* (2013) 57(6):2390–8. doi:10.1002/hep.26232
16. Briney B, Sok D, Jardine JG, Kulp DW, Skog P, Menis S, et al. Tailored immunogens direct affinity maturation toward HIV neutralizing antibodies. *Cell* (2016) 166(6):1459–70. doi:10.1016/j.cell.2016.08.005
17. Oprea M, Perelson AS. Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts. *J Immunol* (1997) 158(11):5155–62.
18. Shlomchik MJ, Watts P, Weigert MG, Litwin S. Clone: a Monte-Carlo computer simulation of B cell clonal expansion, somatic mutation, and antigen-driven selection. *Curr Top Microbiol Immunol* (1998) 229:173–97.
19. Shahaf G, Barak M, Zuckerman NS, Swerdlin N, Gorfine M, Mehr R. Antigen-driven selection in germinal centers as reflected by the shape characteristics of immunoglobulin gene lineage trees: a large-scale simulation study. *J Theor Biol* (2008) 255(2):210–22. doi:10.1016/j.jtbi.2008.08.005
20. Meyer-Hermann M. A mathematical model for the germinal center morphology and affinity maturation. *J Theor Biol* (2002) 216(3):273–300. doi:10.1006/jtbi.2002.2550
21. de Hair M, Zijlstra I, Boumans M, van de Sande M, Maas M, Gerlag D, et al. Hunting for the pathogenesis of rheumatoid arthritis: core-needle biopsy

- of inguinal lymph nodes as a new research tool. *Ann Rheum Dis* (2012) 71(11):1911–2. doi:10.1136/annrheumdis-2012-201540
22. Klarenbeek PL, de Hair MJH, Doorenspleet ME, van Schaik BDC, Esveldt REE, van de Sande MGH, et al. Inflamed target tissue provides a specific niche for highly expanded T cell clones in early human autoimmune disease. *Ann Rheum Dis* (2012) 71(6):1088–93. doi:10.1136/annrheumdis-2011-200612
  23. Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res* (2005) 33(Database issue):D256–61. doi:10.1093/nar/gki010
  24. Kent WJ. BLAT – the BLAST-like alignment tool. *Genome Res* (2002) 12:656–64. doi:10.1101/gr.229202
  25. Lefranc MP. Immunoglobulin and T cell receptor genes: IMGT and the birth and rise of immunoinformatics. *Front Immunol* (2014) 5:22. doi:10.3389/fimmu.2014.00022
  26. Core Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing (2016).
  27. Soetaert K, Petzoldt T, Setzer RW. Solving differential equations in R: package deSolve. *J Stat Softw* (2010) 33(9):1–25. doi:10.18637/jss.v033.i09
  28. Jacob J, Kassir R, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl) acetyl. I. The architecture and dynamics of responding cell populations. *J Exp Med* (1991) 173(5):1165–75. doi:10.1084/jem.173.5.1165
  29. Levy NS, Malipiero UV, Lebecque SG, Gearhart PJ. Early onset of somatic mutation in immunoglobulin VH genes during the primary immune response. *J Exp Med* (1989) 169:2007–19. doi:10.1084/jem.169.6.2007
  30. Berek C, Berger A, Apel M. Maturation of the immune response in germinal centers. *Cell* (1991) 67(6):1121–9. doi:10.1016/0092-8674(91)90289-B
  31. Zotos D, Tarlinton DM. Determining germinal centre B cell fate. *Trends Immunol* (2012) 33(6):281–8. doi:10.1016/j.it.2012.04.003
  32. Shlomchik MJ, Weisel F. Germinal center selection and the development of memory B and plasma cells. *Immunol Rev* (2012) 247(1):52–63. doi:10.1111/j.1600-065X.2012.01124.x
  33. Weisel FJ, Zuccarino-Catania GV, Chikina M, Shlomchik MJ. A temporal switch in the germinal center determines differential output of memory B and plasma cells. *Immunity* (2016) 44(1):116–30. doi:10.1016/j.immuni.2015.12.004
  34. Lefranc M, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* (2003) 27(1):55–77. doi:10.1016/S0145-305X(02)00039-3
  35. Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, et al. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* (2005) 29(3):185–203. doi:10.1016/j.dci.2004.07.003
  36. Kleinstein SH, Singh JP. Why are there so few key mutant clones? The influence of stochastic selection and blocking on affinity maturation in the germinal center. *Int Immunol* (2003) 15(7):871–84. doi:10.1093/intimm/dxg085.sgm
  37. Hershberg U, Uduman M, Shlomchik MJ, Kleinstein SH. Improved methods for detecting selection by mutation analysis of Ig V region sequences. *Int Immunol* (2008) 20(5):683–94. doi:10.1093/intimm/dxn026
  38. Radmacher MD, Kelsoe G, Kepler TB. Predicted and inferred waiting times for key mutations in the germinal centre reaction: evidence for stochasticity in selection. *Immunol Cell Biol* (1998) 76(4):373–81. doi:10.1046/j.1440-1711.1998.00753.x
  39. Gitlin AD, Shulman Z, Nussenzweig MC. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature* (2014) 509(7502):637–40. doi:10.1038/nature13300
  40. Beyer T, Meyer-Hermann M, Soff G. A possible role of chemotaxis in germinal center formation. *Int Immunol* (2002) 14(12):1369–81. doi:10.1093/intimm/dfx104
  41. Liu YJ, Zhang J, Lane PJ, Chan EY, MacLennan IC. Sites of specific B cell activation in primary and secondary responses to T cell-dependent and T cell-independent antigens. *Eur J Immunol* (1991) 21(12):2951–62. doi:10.1002/eji.1830211209
  42. MacLennan ICM. Germinal centers. *Annu Rev Immunol* (1994) 12:117–39. doi:10.1146/annurev.iy.12.040194.001001
  43. Shannon M, Mehr R. Reconciling repertoire shift with affinity maturation: the role of deleterious mutations. *J Immunol* (1999) 162(7):3950–6.
  44. Hollowood K, Macartney J. Cell kinetics of the germinal center reaction – a stathmokinetic study. *Eur J Immunol* (1992) 22:261–6. doi:10.1002/eji.1830220138
  45. Wittenbrink N, Klein A, Weiser AA, Schuchhardt J, Or-Guil M. Is there a typical germinal center? A large-scale immunohistological study on the cellular composition of germinal centers during the hapten-carrier-driven primary immune response in mice. *J Immunol* (2011) 187(12):6185–96. doi:10.4049/jimmunol.1101440
  46. Luciani F, Sanders MT, Oveissi S, Pang KC, Chen W. Increasing viral dose causes a reversal in CD8+ T cell immunodominance during primary influenza infection due to differences in antigen presentation, T cell avidity, and precursor numbers. *J Immunol* (2013) 190(1):36–47. doi:10.4049/jimmunol.1200089
  47. Fan S, Geissmann Q, Lakatos E, Lukauskas S, Ale A, Babbie AC, et al. MEANS: python package for Moment Expansion Approximation, iNference and Simulation. *Bioinformatics* (2016) 32(18):2863–5. doi:10.1093/bioinformatics/btw229
  48. Barak M, Zuckerman NS, Edelman H, Unger R, Mehr R. IgTree: creating immunoglobulin variable region gene lineage trees. *J Immunol Methods* (2008) 338(1–2):67–74. doi:10.1016/j.jim.2008.06.006
  49. Küppers R, Zhao M, Hansmann ML, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *EMBO J* (1993) 12(13):4955–67.
  50. Victora GD, Schwickert TA, Fooksman DR, Kamphorst AO, Meyer-Hermann M, Dustin ML, et al. Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* (2010) 143(4):592–605. doi:10.1016/j.cell.2010.10.032
  51. Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC. Structural insights into the evolution of an antibody combining site. *Science* (1997) 276(5319):1665–9. doi:10.1126/science.276.5319.1665
  52. Tas JMJ, Mesin L, Pasqual G, Targ S, Jacobsen JT, Mano YM, et al. Visualizing antibody affinity maturation in germinal centers. *Science* (2016) 351(6277):1048–54. doi:10.1126/science.aad3439
  53. Dornmair K, Meinel E, Hohlfeld R. Novel approaches for identifying target antigens of autoreactive human B and T cells. *Semin Immunopathol* (2009) 31(4):467–77. doi:10.1007/s00281-009-0179-y
  54. Hearty S, Leonard P, O’Kennedy R. Measuring antibody-antigen binding kinetics using surface plasmon resonance. *Methods Mol Biol* (2012) 907:411–42. doi:10.1007/978-1-61779-974-7\_24
  55. Kelley DE, Perry RP. Transcriptional and posttranscriptional control of immunoglobulin mRNA production during B lymphocyte development. *Nucleic Acids Res* (1986) 14(13):5431–47. doi:10.1093/nar/14.13.5431
  56. Yaari G, Kleinstein SH. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med* (2015) 7(1):121. doi:10.1186/s13073-015-0243-2
  57. Zhang W, Du Y, Su Z, Wang C, Zeng X, Zhang R, et al. IMonitor: a robust pipeline for TCR and BCR repertoire analysis. *Genetics* (2015) 201(2):459–72. doi:10.1534/genetics.115.176735
  58. Safonova Y, Bonissone S, Kurpilyansky E, Starostina E, Lapidus A, Stinson J, et al. IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis. *Bioinformatics* (2015) 31(12):53–61. doi:10.1093/bioinformatics/btv238
  59. Kuchenbecker L, Nienen M, Hecht J, Neumann AU, Babel N, Reinert K, et al. IMSEQ – a fast and error aware approach to immunogenetic sequence analysis. *Bioinformatics* (2015) 31(18):2963–71. doi:10.1093/bioinformatics/btv309
  60. Turchaninova MA, Davydov A, Britanova OV, Shugay M, Bikos V, Egorov ES, et al. High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nat Protoc* (2016) 11(9):1599–616. doi:10.1038/nprot.2016.093
  61. Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med* (2014) 6(248):248ra107. doi:10.1126/scitranslmed.3008879
  62. Marcatili P, Ghiotto F, Tenca C, Chailyan A, Mazzarello AN, Yan XJ, et al. Igs expressed by chronic lymphocytic leukemia B cells show limited binding-site

- structure variability. *J Immunol* (2013) 190(11):5771–8. doi:10.4049/jimmunol.1300321
63. Parameswaran P, Liu Y, Roskin KM, Jackson KK, Dixit VP, Lee JY, et al. Convergent antibody signatures in human dengue. *Cell Host Microbe* (2013) 13(6):691–700. doi:10.1016/j.chom.2013.05.008
64. Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJ. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Phil Trans R Soc B* (2015) 370:20140236. doi:10.1098/rstb.2014.0236
65. Corcoran MM, Phad GE, Bernat NV, Stahl-Hennig C, Sumida N, Persson MA, et al. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nat Commun* (2016) 7:13642. doi:10.1038/ncomms13642
66. Jacob J, Przylepa J, Miller C, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *J Exp Med* (1993) 178(4):1293–307. doi:10.1084/jem.178.4.1293
67. Zhang Y, Meyer-Hermann M, George LA, Figge MT, Khan M, Goodall M, et al. Germinal center B cells govern their own fate via antibody feedback. *J Exp Med* (2013) 210(3):457–64. doi:10.1084/jem.20120150
68. Reshetova P. *The Title of the Work*. Ph.D. thesis, University of Amsterdam, Amsterdam (2016). Thesis in preparation.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Reshetova, van Schaik, Klarenbeek, Doorenspleet, Esveldt, Tak, Guikema, de Vries and van Kampen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.