Title Page

Title

BCR repertoire sequencing: different patterns of B cell activation after two Meningococcal vaccines

Running title

BCR repertoire analysis after different vaccines

Authors

Jacob D. Galson[1], Elizabeth A. Clutterbuck[1], Johannes Trück[1], Maheshi N. Ramasamy[1], Márton Münz[2], Anna Fowler[2], Vincenzo Cerundolo[3], Andrew J. Pollard[1], Gerton Lunter[2] & Dominic F. Kelly[1]

[1]Oxford Vaccine Group, Department of Paediatrics, University of Oxford and the NIHR Oxford Biomedical Research Center, Oxford, OX3 7LE, United Kingdom; [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom; and [3]Medical Research Council Human Immunology Unit, Weatherall Institute of Molecular Medicine, Oxford, OX3 9DS, United Kingdom

Corresponding author

Jacob D. Galson

Email: jacob.galson@paediatrics.ox.ac.uk

Telephone: +44 1865 857445 or +44 7805 243837

Abstract

Next-generation sequencing was used to investigate the B cell receptor heavy chain transcript repertoire of different B cell subsets (naïve, marginal zone, IgM memory and IgG memory) at baseline, and of plasma cells 7 days following administration of serogroup ACWY meningococcal polysaccharide and protein-polysaccharide conjugate vaccines. Baseline B cell subsets could be distinguished from each other using a small number of repertoire properties (clonality, mutation from germ-line and complementarity-determining region 3 length) that were conserved between individuals. However, analyzing the complementarity-determining region 3 amino acid sequence (which is particularly important for antigen binding) of the baseline subsets showed few sequences shared between individuals. In contrast, day 7 plasma cells demonstrated nearly tenfold greater sequence sharing between individuals than the baseline subsets, consistent with the plasma cells being induced by the vaccine antigen, and sharing specificity for a more limited range of epitopes. By annotating plasma cell sequences based on IgG subclass usage and mutation, and also comparing them to the sequences of the baseline cell subsets, we were able to identify different signatures after the polysaccharide and conjugate vaccines. Plasma cells produced after conjugate vaccination were predominantly IgG1, and most related to IgG memory cells. In contrast, after polysaccharide vaccination, the plasma cells were predominantly IgG2, less mutated, and were equally likely to be related to marginal zone, IgM memory or IgG memory cells. High-throughput B cell repertoire sequencing thus provides a unique insight into patterns of B cell activation not possible from more conventional measures of immunogenicity.

Introduction

For most vaccines, protection is achieved via activation of B cells with vaccine antigen-specific receptors, which subsequently differentiate into plasma cells (PCs) and produce functional antigen-specific antibody. Immunogenicity is conventionally assessed by measures of vaccine-specific antibody quantity and function, but this gives little insight into which B cell subsets were activated to generate the functional antibody response. Next-generation sequencing approaches to study B cell receptor (BCR) heavy chain repertoires can be used to measure the diversity of B cell populations, and allow resolution of vaccine response at the level of individual B cell clones.[1] Such methods have been used to demonstrate global changes in the BCR repertoire following vaccination with different antigens,[2–5] and that these changes are dependent on the type of vaccine given,[2,3] and the age of the individuals vaccinated.[2,4] One study investigated the repertoire after two successive annual influenza vaccinations, and showed that some identical sequences could be identified after both the first and second vaccines, indicating that BCR repertoire sequence data can be used to determine memory recall.[3] To date, BCR repertoire studies of vaccine response have focused on total B cells, or PCs. However, diverse B cell subsets may be involved in a response, including naïve, marginal zone (MZ) and memory (IgM and IgG) B cells, depending on the type of antigen, previous exposure, and route of immunization. Analysis of different B cell subsets revealed differences in their sequence, and VDJ gene segment composition,[6,7] and thus interrogation of the BCR repertoire on a subset-by-subset basis could potentially be used for fine delineation of which B cells are involved in vaccine responses.

Vaccines against polysaccharide-encapsulated pathogens (e.g. *Neisseria meningitidis* and *Streptococcus pneumoniae*) are of great importance in controlling these predominantly childhood diseases. Polysaccharides themselves are T-independent antigens, have poor immunogenicity in infants,[8] do not generate immunological memory,[9] and can lead to hyporesponsiveness with repeated exposure.[10] The conjugation of a carrier protein to a polysaccharide results in a T-dependent antigen,

which is immunogenic from early life, and primes for memory.[11] Licensed quadrivalent ACWY *N. meningitidis* vaccines contain either plain purified capsular polysaccharides (polysaccharide vaccine), or the same polysaccharides conjugated to a carrier protein (conjugate vaccine). The difference in immunogenicity, and the different B cell subsets involved in the response to these related vaccines are still being elucidated. The different B cell subsets activated by conjugate and polysaccharide vaccines, have previously been investigated during a comparative study of pneumococcal conjugate and polysaccharide vaccines.[12] The conjugate vaccine induced more circulating serotype-specific memory B cells than the polysaccharide vaccine.[12] However, despite previous suggestion that polysaccharide antigens stimulate MZ B cells.[13] there was no difference seen in the frequency of serotype-specific MZ B cells measured in peripheral blood after the two vaccines in this study, perhaps due to limitations in the sensitivity of the flow cytometry assay.[12]

We sought to determine the utility of BCR heavy chain repertoire sequencing as a tool for investigating vaccine responses, using meningococcal ACWY polysaccharide and conjugate vaccination as a model system. We immunized individuals with either a polysaccharide or conjugate vaccine, followed by a further immunization with a conjugate vaccine 4 weeks later (Fig. 1). We isolated naïve, MZ, IgM memory and IgG memory baseline B cell subsets, in addition to PCs 7 days after each vaccination (Fig. S1). This time point was chosen as previous work from our laboratory using the same vaccine regime has shown the presence of significant numbers of antigen-specific PCs 7 days after both vaccine doses[14], and induction of effective antibody responses in both groups.[15] Hence, sorting PCs at this time point would capture a population enriched for vaccine antigen-specificity. We further sorted day 7 PCs based on HLA-DR expression, to distinguish recently activated PCs (HLA-DR+) likely to be enriched for vaccine specificity, from long-lived PCs (HLA-DR-), which are considered to be displaced from the bone marrow by arrival of the newly generated PCs and likely have a broader range of specificities.[16] Characterizing mutation levels, and IgG

subclass usage of the PC sequences, as well as relating them to the sequences of the baseline B cell subsets, allowed us to determine the B cell subsets involved in the response to the two vaccines, and to distinguish T-independent from T-dependent responses.

Results

Cell sorting

Nine participants were enrolled into the study; four were assigned to the conjugate group, and five to the polysaccharide group (Fig. 1). Fluorescence-activated cell sorting was used to isolate baseline CD19+ B cell subsets bearing cell surface markers characteristic of naïve (CD27-, IgM+, IgD+), circulating MZ (CD27+, IgM+, IgD+)[17], IgM (only) memory (CD27+, IgM+, IgD-) and switched IgG memory (CD27+, IgM-, IgD-) for five participants at Visit 1 (Day 0). The sorting failed for the other four participants (2 from each vaccine group), so these subsets were sorted at Visit 3 (Day 28) instead, and treated as baseline. HLA-DR+ PCs (CD20-, CD27+, CD38+, HLA-DR+) and HLA-DR- PCs (CD20-, CD27+, CD38+, HLA-DR-) were successfully sorted at Visit 2 (Day 7), and HLA-DR+ PCs at Visit 4 (Day 35; day 7 after second vaccine) from all participants. The mean number of sorted cells per sample was 75 900 (12 000 – 140 000) for naïve cells, 6 753 (1 224 – 19 000) for MZ cells, 8 070 (1 888 – 19 000) for IgM memory cells, 27 145 (2 940 – 100 000) for IgG memory cells, 6 101 (110 – 50 000) for HLA-DR+ PCs and 4 311 (221 – 11 000) for HLA-DR- PCs (Table S1). For the Visit 2 (Day 7) HLA-DR+ and HLA-DR- PC populations, IgG and IgM transcripts were sequenced (except one IgG HLA-DR+ sample which failed). Only the IgG transcripts of the HLA-DR+ population were sequenced at Visit 4 (Day 35). For the baseline subsets, IgM transcripts were sequences from all naïve, MZ and IgM memory samples, and IgG transcripts sequences from all the IgG memory samples.

Sequence processing

Sequence data were obtained from 80 samples from 9 participants (Table S1). The mean number of raw reads obtained for each sample was 133 190, of which 72% remained on average after initial filtering steps, giving 95 610 reads per sample for analysis. As there was over-sequencing (more reads

than input cells) for nearly all samples, clustering was used to minimize the effect of PCR and sequencing error on the dataset. As the most important region in determining antigen-specificity of the sequence, we focused analysis on the complementarity-determining region 3 (CDR3).[18] Determining the distance of each CDR3 amino acid (AA) sequence to its nearest neighbor in each sample revealed a bimodal distribution (Fig. 2A), for all cell subsets except for the naïve cells, where there was the least over-sequencing. The first peak was at 0-2 AA, and the second at 3-10 AA. There was a strong positive correlation (r=0.72) between the proportion of reads in the repertoire in the first peak, and the degree of over-sequencing for each sample (Fig. 2B). Analysis of constant region sequences indicated a PCR and sequencing error rate of less than 0.5% per nucleotide. Given the median CDR3 length was 18 AA, it is likely the peak at 0-2 AA contained most of the variants due to sequencing and PCR error. Setting a clustering threshold of 1 AA mismatch per 12 AAs means that for 8% of reads, 1 AA mismatch was allowed during clustering, for 89% of reads 2 AA mismatches were allowed, and for 3% of reads 3 AA mismatches were allowed. Clusters should therefore encompass all sequences arising from the same cell that differ due to PCR and sequencing error, but will not contain reads more than 4 AA different which are more likely to have arisen from different cells. However, we can not rule out that we also cluster together reads from different, but clonally related cells that use the same V and J genes, and have a similar CDR3 sequence (Fig. 2A). After collapsing the clusters to give a single representative sequence (the cluster center) for each, duplicates were removed, and the number of sequences per sample was representative of the number of cells for that sample (Fig. 2C), and on average 6% of the original number of raw reads. For some samples, there were still more sequences than cells, but the discrepancy was small, and could be due to inaccuracies in the number outputted by the cell sorter, a small number of erroneous sequences not being incorporated into the same cluster, exogenous DNA contamination, or the formation of chimeric sequences during the PCR reaction (Table S1.). The mean number of sequences per sample

was 41 100 (19 240 – 52 440) for naïve cells, 7 873 (3 861 – 16 120) for MZ cells, 6 200 (3 096 – 9 673) for IgM memory cells, 3 652 (1 308 – 7 382) for IgG memory cells, 1 466 (436 – 2 786) for HLA-DR+ PCs, and 1 259 (601 – 2 277) for HLA-DR- PCs.

Post-vaccination plasma cell repertoires have distinct properties compared to the baseline cell subset repertoires

Baseline B cell subsets could be distinguished from each other by a small number of global repertoire properties: clonality, V gene mutation from germ-line and CDR3 length (Fig. 3A-C). Naïve B cells had the lowest clonality, few mutations and longest CDR3. Memory B cells had increased clonality, more mutations, and shorter CDR3 length compared to naïve cells, with IgM memory being less mutated and having longer CDR3 length than IgG memory cells. MZ cells had similar clonality to IgM memory cells, but had fewer mutations, and a shorter CDR3 length than IgM memory cells. PC subsets were the most clonal and most mutated populations. CDR3 length of PCs resembled that of memory cells with the same isotype. Within the PC populations, there were some differences between the IgG and IgM PCs (as with the memory cells, IgM were less mutated, and had shorter CDR3's than IgG), but no differences between the HLA-DR+ and HLA-DR- populations. Principal component analysis based on these variables showed distinct groupings of the different baseline subsets (Fig. 3D & Fig. S2). Strikingly, PCs had much greater inter-individual variation in these global repertoire properties than other subsets. Whilst the PCs could still be clearly distinguished from the naïve, marginal zone and IgM memory subsets, their global repertoire properties overlapped with the IgG memory cells. Within the PCs, populations could be distinguished based on isotype, but not HLA-DR expression.

There were differences in V, D, and J gene segment usage between all subsets analyzed, including numerous differences between PC and baseline subsets, and within the different baseline subsets (Fig. 3E & Fig. S3A). When comparing each PC population to other PC and baseline subsets,

the PCs were more similar to each other (whether IgG/IgM or HLA-DR+/-) than to any baseline subset. Differences in V, D and J gene usage between the HLA-DR+ and HLA-DR- PCs were further investigated by looking at the inter-individual variation in gene usage. A population with more restricted antigen-specificity is expected to have less variance in gene segment usage between individuals. For each gene segment, the variance in their usage between the participants was calculated for the two PC populations, and then compared between the two populations using an F-test for equality of variance (Fig. S3B). All J and D gene segments had similar variances in the two PC populations. For the IgG PCs, there were ten V genes that had greater inter-individual variation in the HLA-DR+ PCs compared to the HLA-DR- PCs, and only one V gene that had greater variation in the HLA-DR- PCs. For the IgM PCs, three V genes had greater variance in the HLA-DR+ PCs compared to two V genes that had greater variance in the HLA-DR- PCs. The HLA-DR+ PCs therefore appear to vary more between individuals in their V gene usage, while the HLA-DR- PCs have more similar V gene usage in different individuals.

Plasma cell repertoires show similarity between individuals at the sequence level

To determine if we could identify repertoire convergence between the different participants (the use of similar sequences to bind similar antigens following an immune stimulus) in response to vaccination, we searched for shared CDR3 AA sequences between samples (Fig. S4). The CDR3 region is the most variable of the three CDRs and is considered a key region for determining antigen binding.[18] Among the baseline subsets, on average 0.19% (range: 0.11-0.40%) of sequences were shared either between the vaccine groups, or by at least two participants within each vaccine group (Fig. 4). In contrast, PCs had greater sequence sharing, with on average 0.99% (range: 0.17-2.21%) of sequences being shared either within or between individuals in the two vaccine groups. The HLA-DR- PCs tended to have more sequence sharing than the HLA-DR+ PCs. Comparing the two vaccine groups the most striking difference was in the HLA-DR+ IgM PCs, where the percent of shared

sequences was more than two times greater within the polysaccharide group than the conjugate group. Other than this high level of sequence sharing in IgM HLA-DR+ PCs after polysaccharide vaccination, there was little difference in the number of shared sequences between IgM and IgG PCs. There were some differences in convergence at Visit 2 compared to Visit 4 for the HLA-DR+ IgG PCs; there was more convergence between the vaccine groups at Visit 4 (administered the same vaccine) compared to Visit 2 (administered different vaccines). However, it should be noted that it was a HLA-DR+ IgG Visit 2 sample from the conjugate group that failed sequencing, which would decrease the probability of sequence sharing within this population. Within the vaccine groups, the number of shared sequences increased approximately fourfold in the conjugate group, but decreased slightly in the polysaccharide group at Visit 4.

Analysis of plasma cell activation distinguishes polysaccharide from conjugate responses

We investigated whether there was evidence that the PCs isolated at day 7 following each vaccine, were related to the baseline subsets. Data from a previous study had demonstrated that the PC population on day 7 contained significant numbers of vaccine-specific PCs.[14] We therefore determined whether there were shared sequences (based on an exact CDR3 AA sequence match, and use of the same V and J gene segments) between the HLA-DR+ PC subsets, and the baseline subsets within the samples from a given individual. On average, 2.2% (30) of sequences of each PC subset were shared with one of the baseline subsets for each participant. Across all participants IgM PC sequences were predominantly shared with the MZ and IgM memory cells whereas the IgG PC sequences were predominantly shared with the IgG memory cells (Fig. 5A).

To investigate which baseline subsets were related to the vaccine-stimulated PCs, we calculated the probability of the shared PC sequences mapping to each baseline subset (Fig. 5B-C). There was a low probability of IgG or IgM PC sequences being shared with the naïve baseline subset

at either visit, in any vaccine group (Fig. 5B-C). In the conjugate group, at Visit 2, the HLA-DR+ IgG PC sequences were more likely to be related to the IgG memory sequences than naïve, MZ or IgM memory sequences (Fig. 5B). In contrast, in the polysaccharide group, HLA-DR+ IgG PC sequences were equally likely to be related to MZ, IgM memory and IgG memory sequences. Therefore for the HLA-DR+ IgG PCs, the baseline subset that they were most likely to be shared with was related to the vaccine that was given (Fig. 5B-C); we investigated the potential effect of having baseline subsets from Day 0 vs Day 28 on this difference between the vaccine groups, but it was not significant (Fig. S5). We also saw no difference between the vaccine groups for the HLA-DR- IgG PCs, which should not be specific for the vaccine (Fig. S6). By comparison, for the HLA-DR+ IgM PC sequences, there was a similar probability of being similar to the MZ and IgM memory cells for both vaccine groups. Interestingly after the second dose of vaccine (conjugate for both groups), the probability of HLA-DR+ PCs being shared with each baseline subset was similar for both groups (Fig. 5C); the profile resembled that of the conjugate group after the first vaccination.

Vaccine response can further be distinguished based on IgG subclass usage, and mutation

We then further characterized differences in the response between the vaccine groups by comparing V gene mutation from germline, and IgG subclass usage between the vaccine groups for the HLA-DR+ PC sequences generated after vaccination. IgG sequences extended on average 126 nucleotides into the constant region. This was sufficient to allow mapping of the IgG subclass for all sequences in our dataset. In the IgG memory cells, IgG subclasses were present in uneven frequencies, with IgG1 (58%) and IgG2 (34%) being the most common, and IgG3 (7%) and IgG4 (1%) being the least common (Fig. S7A). By comparison, In the PCs produced after the first vaccination, in the conjugate group, IgG1 remained the most common subclass (58%), followed by IgG2 (33%), IgG4 (7%) and IgG3 (2%) (Fig. S7A). In the polysaccharide group on the other hand, the subclass distribution was

different; IgG2 was the dominant subclass (58%), followed by IgG1 (37%), IgG3 (5%) and IgG4 (<1%). After the second vaccine, when both groups were given conjugate vaccine, the subclass usage of the PCs resembled that of the conjugate group in both vaccine groups.

Considering V gene mutation from germ-line for total IgG compared to IgM, IgG PCs were more mutated than IgM PCs ($p = 0.0024$). After the first vaccination, there was no significant difference in V gene mutation between the vaccine groups for either the total IgG or the IgM PCs (Fig. 6B). After the second vaccination (when both groups were given the conjugate vaccine), V gene mutation in the IgG PCs was greater in the group primed with the polysaccharide rather than conjugate vaccine. There was a significant increase in IgG PC V gene mutation in the polysaccharide group after the second vaccination (conjugate) compared to the first vaccine (polysaccharide). In the conjugate group, there was no difference in IgG PC V gene mutation after the second dose compared to the first dose of the two conjugate vaccines.

Calculating average V gene mutation separately for each IgG subclass revealed further differences between the vaccine groups. In the IgG memory cells, there are slight differences in V gene mutation in the different subclasses going from IgG3 (7.04%) to IgG1 (7.19%), IgG2 (7.64%) and IgG4 (7.87%) (Fig. S7B). Comparing the vaccine groups showed the only significant difference to be in IgG1 PC sequences. IgG1 sequences were significantly more mutated in the conjugate group after the first vaccine (7.3% in the conjugate group vs. 5.9% in the polysaccharide group), but significantly more mutated in the polysaccharide group after the second vaccine (7.2% in the conjugate group vs. 8.7% in the polysaccharide group) (Fig. 6C). Mutation for all four subclasses increased between the first vaccination and the second vaccination in the polysaccharide group. In the conjugate group, there was only a significant increase in mutation in the IgG4 sequences after the second vaccination.

Discussion

The critical events in the generation of an immune response occur largely in the lymphoid tissue, which is challenging to sample directly in humans. After vaccination with meningococcal conjugate or polysaccharide vaccines, antigen-specific PCs are detected in peripheral blood at high levels 7 days post-vaccination.[19] Using BCR repertoire analysis of the response to two different meningococcal serogroup ACWY vaccines, this study suggest that insights into B cell activation can be obtained by relating BCR sequence data from post-vaccination PCs to a variety of baseline B cell subsets. The study has shown that a limited number of global BCR repertoire properties (mean CDR3 length, mutation and clonality) can be used to define different baseline subsets, and that these properties are well conserved between individuals. In contrast, antigen stimulated PCs display more inter-individual variation in these global repertoire properties. Despite this diversity in global repertoire properties, the PCs show a higher degree of sharing between individuals at the level of the CDR3 AA sequence (the most critical for antigen binding), when compared to the baseline subsets. Although we recognize that such sharing can come from cross-sample PCR contamination, laboratory work was conducted under stringent conditions with this in mind, and we find no contaminant sequences present in all samples, indicating that this would not have had a large effect on our results. The finding of PCs expressing the greatest difference between individuals in their global repertoire properties is consistent with them potentially being activated from a variety of different baseline B cell subsets, which will differ for each individual. The greater proportion of shared sequences in the PCs suggests that there is then some degree of convergence at the sequence level to the vaccine antigens stimulating their production. Furthermore our data suggest that immunisation with protein-polysaccharide conjugate vaccines stimulates the development of PCs from IgG memory cells, but plain polysaccharide vaccination results in significant numbers of B cells being derived from IgM memory

and MZ B cells as well. Further dissection of the PC sequences gives insight into differential IgG subclass usage, and V gene mutation induced by the two vaccines.

Differences have previously been observed in the BCR repertoire between naïve, IgM memory and IgG memory cell subsets[20,21] that are conserved across multiple individuals. This study extends these observations by including, MZ, and a variety of PC subsets in the same analysis. By measuring clonality, CDR3 length, and V gene mutation from germline (Fig. 3), we show that B cell subsets appear to have a characteristic sequential pattern of global repertoire properties in relation to their maturation state (naïve → (MZ) → IgM Memory → IgG Memory → PCs), with the more mature subsets tending towards increased clonality, shorter CDR3's, and more V gene mutation. The presence of the circulating MZ cells being a distinct grouping from memory cells, yet located between the naïve and IgM cells is interesting in light of recent work detailing the developmental pathway of MZ cells.[22] It may be that MZ cells are a distinct cell subset arising from germinal center independent pathways, or alternatively that they are just in a transitional state between naïve and IgM memory cells.[23]

Increased clonality and mutation are expected during B cell maturation, due to the processes of antigen-driven selection and proliferation of naïve cells to the more mature states. The observation of mature B cells having shorter CDR3's than naïve cells has previously been reported,[6,21] and as some autoreactive antibodies have long CDR3's, this may be a side effect of a mechanism to reduce autoreactivity.[24] We,[25] and others,[26,27] have previously shown that after exposure to a common antigen, there is a degree of CDR3 sequence convergence between different individuals, with convergent sequences likely to be antigen-specific; the demonstration here that the greatest degree of CDR3 AA sequence sharing was between the PC subsets further supports this hypothesis. The dichotomy between PCs being the most similar population between individuals at the sequence level, but most divergent in terms of their general repertoire properties is consistent with the hypothesis that

a variety of precursor cells belonging to different baseline B cell subsets are activated to form the PCs of different individuals, but the process of affinity maturation then leads them to converge at the sequence level.

The acquisition of sequence data from both baseline B cell subsets, as well as vaccine-activated PCs, allowed us to determine whether baseline B cell subsets were related to the PCs, by measuring the probability of sequence sharing between them. Such sequence sharing can be used to indicate which baseline B cell subsets are stimulated to form PCs by the vaccine. Whilst the current study does not confirm the antigenic specificity of these PCs, a previous study using an identical vaccine has demonstrated that vaccine-specific PCs are detectable at high levels 7 days following either the first dose or second dose of the ACWY vaccine. Therefore, the PC populations we sorted are likely to be enriched for vaccine-specific PCs.[14] In addition, we focused our analysis on HLA-DR+ PCs, which are considered to be a recently activated population, and therefore likely to be further enriched for vaccine-specificity.[16] Pre-existing immunity to *N. meningitidis* is common in adults, so the vaccinated individuals in this study are unlikely to mount naïve responses.[29] The observation that the lowest probability of sequence sharing was between the naïve cells and the PCs in this study therefore supports our approach for determining B cell activation by showing that such sequence sharing is unlikely to occur by chance alone. IgM PCs appeared to be mainly derived from the MZ, or IgM memory cells, whereas the IgG PCs appeared to be mainly derived from the IgG memory cells, suggesting that PCs are mainly activated from memory cells as part of a recall response. Comparing the vaccine groups, the IgG PCs after polysaccharide vaccine are more likely to be related to MZ and IgM memory cells, and less likely to be related to IgG memory cells than after conjugate vaccine. This supports the hypothesis that the polysaccharide vaccine stimulates a more T-independent like response, while the conjugate vaccine stimulates a more T-dependent like response. However, there was still a high probability of IgG PCs being related to IgG memory cells even after

polysaccharide vaccination, indicating that categorizing responses in this way is not clear-cut, and an antigen may activate both T-dependent and T-independent pathways. While B cell memory recall has previously been assessed by testing antibody affinity,[30] this method does not indicate which B cell subsets are activated in the recall response. The technique presented here for monitoring memory recall at the sequence level has clear application in vaccine development, where long-term protection requires such memory formation.

Knowing the exact sequences of the recalled cells not only allows relationships to precursor memory cell subsets to be explored, but also allows identification of IgG subclass and mutation levels, giving further insight into the vaccine response. Determination of IgG subclass is only possible when B cell heavy chain mRNA transcripts are sequenced rather than genomic DNA, as there is a large intron between the variable region gene segments, and the constant region gene segments.[31] Since the four IgG subclasses differ in their ability to activate complement, and bactericidal activity in different antigenic contexts,[32] monitoring IgG subclass usage is useful. Normally, IgG2 is produced in response to polysaccharide antigens, IgG1 and IgG3 produced in response to protein antigens,[33] while IgG4 serves a more regulatory function.[34] The data here show that after polysaccharide vaccination, IgG2 was the predominant subclass produced, but after conjugate vaccination, IgG1 was the predominant subclass produced (even in individuals primed with the polysaccharide vaccine). The ability to detect these differences in isotype usage after the different vaccines suggests that we have significant enrichment of vaccine-specific PCs in the HLA-DR+ PC population, which are able to drive these differences. As mutational levels are correlated with cellular proliferation in the germinal center,[35] measuring changes in BCR sequence mutation may be a proxy for the magnitude of a high affinity PC response following vaccination. PC levels of IgG1 mutation were greater in the conjugate compared to the polysaccharide group after the first vaccine, which is consistent with increased germinal center activity and switching to IgG1 in the response to conjugate vaccination. The increase

in frequency of mutations in all IgG subclasses between the first (polysaccharide) and second (conjugate) vaccines in the polysaccharide group further suggests that conjugate vaccination may initiate a more marked germinal center response than polysaccharide vaccination. In the conjugate group, only IgG4 sequences increased in mutation after the second vaccine, indicating that after repeat vaccination with such a highly stimulating T-dependent antigen, IgG4 plays an active role in the response. Considering IgG4 have unique structural and functional properties compared to the other IgG subclasses,[34] the precise role of IgG4 in terms of regulating the response following antigen administration would be interesting to explore further. It has previously been suggested that B cells have a predictable sequential pattern of IgG class-switching while in the germinal center, going from IgG3 to IgG1 to IgG2 to IgG4.[36] Our data support this hypothesis in that we see increasing levels of mutation in the IgG subclasses in the IgG memory cells going from IgG3 to IgG1 to IgG2 to IgG4 (Fig. S7), indicative of proliferation followed by subclass switching occurring in the same order.[37] This would suggest that although there appears to be less PC proliferation following polysaccharide vaccine, there is a greater rate of subclass switching than following conjugate vaccine. However, it is interesting to note that when administered the conjugate vaccine following the polysaccharide vaccine, there is a decrease in IgG2 and an increase in IgG1 PCs after the conjugate vaccine compared to the polysaccharide vaccine, which is hard to reconcile with the sequential switching hypothesis.

The sequence analysis presented here requires sequencing an enriched population of vaccine-specific PCs, and in this light we focused our analysis on HLA-DR+ PCs. Previous work has demonstrated that PCs show a gradient of HLA-DR expression from high levels in lymphoid tissue (tonsils) through to lower levels in blood and lowest expression in bone-marrow.[38] This is thought to correlate with maturation towards more long-lived PCs in the bone marrow. Studies of tetanus-specific PCs 6 days following tetanus immunization suggested that the vaccine antigen-specific population were HLA-DR+ and constituted a significant proportion of total HLA-DR+ PCs at that

time point.[16] HLA-DR- PCs at day 6 were not vaccine antigen-specific and were thought to have been displaced from their residence in the bone marrow by the vaccine generated PCs. In our study, the acquisition of data from both HLA-DR+ and HLA-DR- PC populations gave us an opportunity to compare the properties of the BCR repertoire in these two populations thought to be from divergent origins. We saw no differences in clonality, V gene mutation or CDR3 length between HLA-DR+ and HLA-DR- PCs, which might be expected considering that high clonality, increased mutation and short CDR3s appear to be features of all mature B cells regardless of antigen-specificity.[25] In addition, there were limited differences observed in V, D and J gene usage between the two PC populations that were conserved between individuals. Recently activated PCs have a limited stimulus (vaccine antigens), whereas previously generated PCs could potentially have been activated by a lifetime of different antigenic stimuli, so one may expect the HLA-DR+ PCs to show the greatest similarity between individuals. Intriguingly, we saw more similarity between individuals both at the level of CDR3 AA sequence, and V gene usage, for the HLA-DR- than the HLA-DR+ population. It therefore appears that either there is a degree of selection for incorporation into the HLA-DR- compartment, or that there are certain common antigens that all individuals are exposed to during the course of their lifetime and sequences against these common antigens dominate the HLA-DR- compartment. Another possibility is that in addition to vaccine-specific PCs in the HLA-DR+ population, additional inter-individual arises from the presence of PCs generated in incidental responses against non-vaccine antigens or due to 'bystander stimulation'.[39] Perhaps the most compelling evidence we have of the HLA-DR+ and HLA-DR- PCs comprising distinct populations is that when analyzing which baseline B cell subsets are activated to produce the IgG PCs, we see differences between the vaccine groups for the HLA-DR+, but not the HLA-DR- PCs. This indicates that the HLA-DR+ population is enriched for vaccine-specificity, allowing us to detect these differences, but the HLA-DR- population is not. Further work would be beneficial to elucidate the exact antigenic-specificity of the HLA-DR+

and HLA-DR- sequences by generating monoclonal antibodies for characterization. As we only sequenced the BCR heavy chain, this is not currently possible, but advances in heavy-light chain pairing technology mean that this may be possible in the future.[40]

Cell sorting is a technically challenging process, and failures meant that for some participants, baseline subsets had to be sorted 28 days post-vaccination instead of pre-vaccination. There were no differences in the baseline repertoire properties depending on which timepoints they were taken (Fig. S2 & Fig. S5), and data from our laboratory suggests that by 28 days post-vaccination, the repertoire returns to a pre-vaccination state (Galson *et al.* unpublished), so these cell-sorting failures are unlikely to have affected our results. Nevertheless, developing a process for in-silico cell sorting of bulk sequence data from peripheral blood would improve the accessibility and ease of our methods for determining subset-specific memory recall, and characterizing vaccine responses at the sequence level. The differences we found in the sequence composition of the different cell subsets should aid the development of such methods. Furthermore, although the vaccine-specific patterns of IgG isotype usage seen suggest we already have significant enrichment of vaccine-specific sequences in the HLA-DR+ PC population, further enrichment for sequences with known vaccine-specificity either by antigen-specific cell sorting, or in-silico cell sorting would increase the validity of our results. We were limited to just 9 participants in our study, but were nevertheless able to find a number of significant differences between the two vaccine groups. Using in-silico cell sorting would also allow such work to be carried out more easily in larger cohorts, and likely shed further light on the different effects of conjugate and polysaccharide vaccines on the immune response. Since conducting this study, improved methods to account for PCR error in B cell repertoire data have been developed.[3] While the clustering method used can account for PCR and sequencing error to an extent, it may also erroneously cluster together sequences that arise from distinct, but clonally related cells. Use of these new methods would improve the specificity of measuring memory recall.

Our study shows that B cell repertoire sequencing opens a novel window onto vaccine responses, not possible through conventional immunological measures. When combined with samples taken at different time points and from different B cell subsets, individual B cell clones can be tracked to determine activation and memory recall. Accumulating such data in vaccine development programs can be used to show which B cell subsets (or even specific B cell sequences known to produce protective antibodies) are activated by the vaccine, so that the vaccine can then be altered to drive generation of optimal responses.

Methods

Participants and vaccinations

This study was approved by the Oxfordshire Research Ethics Committee (OXREC reference number: 12/SC/0275) with informed consent obtained from participants before enrolment in accordance with the Declaration of Helsinki. Nine healthy volunteers (aged 30-70 years), with no meningococcal vaccine history, were recruited to receive two doses of meningococcal serotypes A, C, W and Y vaccine. Following a first dose of either ACWY polysaccharide vaccine (MenACWY-PS, ACWYVax®, GSK) given to 5 individuals, or ACWY conjugate vaccine (MenACWY-CRM_{197}, Menveo®, Novartis) given to 4 individuals (Fig. 1), all individuals received a dose of MenACWY-CRM_{197} at day 28. All vaccines were administered as intramuscular injections. A sample of 50 ml peripheral blood was taken from participants on the day of the first vaccine (Visit 1), 7 days after the first vaccine (Visit 2), 28 days after the first vaccine (Visit 3) and 7 days after the second vaccine (Visit 4), and transferred to a heparinized tube for processing within 4 hours of collection.

Cell sorting

Peripheral blood mononuclear cells were isolated by density gradient centrifugation over lymphoprep (Axis-Shield Diagnostics). CD19+ B cells were then enriched using CD19 microbeads (Miltenyi Biotec), and the AutoMACS magnetic cell separator. On Visit 1, and Visit 3, cells were stained for isolation of naïve (viable, CD19+, CD27-, IgM+, IgD+), MZ (viable, CD19+, CD27+, IgM+, IgD+), IgM memory (viable, CD19+, CD27+, IgM+, IgD-) and IgG memory (viable, CD19+, CD27+, IgM-, IgD-) B cell subsets (Fig. S1). On Visit 2 and Visit 4, cells were stained for isolation of recently activated PCs (viable, CD19+, CD20-, CD27+, CD38+, HLA-DR+), and bystander PCs (viable, CD19+, CD20-, CD27+, CD38+, HLA-DR-) (Fig. S1). Cell sorting was performed on a MoFlo

(Beckman Coulter) cell sorter; sorted cells were re-suspended in RLT buffer (Qiagen), and frozen at -80 ^{o}C.

Repertoire sequencing

Sorted cells were defrosted, and total RNA extracted using the RNeasy Mini Kit (Qiagen). cDNA synthesis was performed using the SuperScript III reverse transcriptase system (Invitrogen), with random hexamer primers (Applied Biosciences). Reverse transcription was performed at 42°C for 60 min and 95°C for 10 min. VH genes were amplified from cDNA using Multiplex PCR MasterMix (Qiagen) containing 200 nM each VH-family specific forward primers, and either IgM or IgG-specific reverse primers depending on the cell population.[21] PCR conditions were 94 ^{o}C for 15 min, 33 cycles of 94 ^{o}C for 30 s, 58 ^{o}C for 90 s and 72 ^{o}C for 30 s, and 72 ^{o}C for 10 minutes. The PCR product was gel-extracted, purified, and quantified using a Qubit fluorometer (Invitrogen, Grand Island, NY, USA) prior to library preparation. The 80 samples obtained were end-repaired, A-tailed and adaptor ligated prior to size-selection and amplification. All samples were multiplexed, and sequenced using one 2x250 bp MiSeq run (Illumina, San Diego, CA).

Sequence processing

Raw reads were de-multiplexed, and paired-ends joined using fastq-join (ea-utils) based on a minimum overlap of 6 bp, and 8% maximum mismatches, to give one continuous read that spanned from framework region 1 to 53 bp or 126 bp inside the constant region for IgM and IgG sequences respectively. Joined reads were filtered for a minimum Phred quality of 30 over 75% of bases and then submitted to IMGT/HighV-Quest for mapping to germline VDJ genes, and annotation.[41] Only reads defined as ‘productive’ by IMGT were considered for further analysis. To obtain isotype subclass information, constant region sequences (as defined by IMGT) were mapped using Stampy[42] to a reference table of constant region sequences obtained from the VEGA database[43]. PCR and

sequencing error rates were estimated from the number of nucleotide mismatches in the constant region, as these regions are not subject to somatic hypermutation.[44] Percent V gene mutation was calculated based on the number of V gene nucleotide mismatches determined by IMGT, and the length of the V gene that was sequenced. For each read, CDR3 AA distance to its nearest neighbor was calculated by comparing it to all other reads from the same sample with the same CDR3 length. The read with the fewest AA mismatches in the CDR3 region is termed the nearest neighbor, and the distance is the number of AA differences in the CDR3.

Clustering

Related reads from each sample were clustered using parameters that best clustered together reads arising from the same cell. To be considered part of the same cluster, reads were required to have the same V and J gene annotation, the same length CDR3, and a similar CDR3 AA sequence, with one AA mismatch per 12 AAs allowed (length < 12 AA = 1 mismatch, 12 AA <= length < 24 AA = 2 mismatches, etc). The same D gene annotation was not required for inclusion in the same cluster, as somatic hypermutation in this region makes accurate assignment of a gene segment difficult. We focused on the CDR3 region of the sequence, as it is highly variable, and has the dominant role in determining antigenic specificity of the sequence.[18] Clusters were iteratively defined using an approach to identify cluster centers that gave the largest possible clusters. Briefly, clustering started with a set of unique sequences of the same length U(L) and an empty set of clusters C(L). The first cluster center is defined as the sequence $x \in U(L)$ that had the most neighbor's $|N(x)|$; the set N(x) is then added to C(L), with cluster center x. N(x) is then removed from U(L), and the process repeated until U(L) is empty and C(L) is full. Clusters were then collapsed into single reads, with each cluster represented by its cluster center sequence, V D and J gene usage, isotype subclass, and average V gene mutation. Where a single cluster contained multiple reads with different isotype subclass or D

gene usage, the most frequent usage within the cluster was used for cluster annotation. These collapsed clusters were used in all subsequent analysis, and defined as "sequences".

Statistical analysis and graphing

Statistical analysis was performed using R.[45] The ggplot2[46] package for R was used for constructing graphs. Circos plots were created using circos software (http://www.cirocs.ca). To calculate the clonality of each sample, similar clones were defined as having identical CDR3 AA sequences, and clonality was calculated using a modified version of the Simpson index (λ), where clonality equals 1- λ. Principal component analysis was performed using the prcomp function in the stats R package based on clonality, average CDR3 length, and mutation for each sample.

For analysis of shared sequences, we used two different definitions for sequence sharing: exact CDR3 AA sequence matches, or exact CDR3 AA sequence matches plus exact V and J gene usage. The reason for doing this is to distinguish sequences that are shared due to a common antigenic-specificity (shared CDR3 only), and sequences that are shared due to common ancestry (shared CDR3 and VJ). Requiring identical V and J gene segment usage for the identification of shared sequences will also ensure that chimeric sequences formed during PCR (where the CDR3 is the same, but either the V or J gene annotation may differ), do not artificially increase the number of shared sequences we find. The definition used for specific figures is noted in the legend. We searched for shared sequences between all samples individually, and also between samples grouped by vaccine group.

The probability of PC sequences being related to sequences from each baseline subsets was calculated separately for each participant and each PC subset. PC sequences that were shared with any baseline subset were found, and the proportion of these sequences that mapped to each baseline subset was then determined (proportion shared PC shared with baseline subset = P_{SHARED}). To account for the different representation of baseline subsets, the proportion of each baseline subset of the total

baseline sequence pool was calculated (proportion baseline subset of total baseline = P_{BL}). The probability of PC's being related to a specific baseline subset was then determined as P_{SHARED} / P_{BL}. The resulting value was normalized so that the sum of the probabilities of being related to each baseline subset was equal to 1.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The NIHR Oxford Biomedical Research Centre provided study funding. The Nuffield Department of Medicine and Wellcome Trust Centre for Human Genetics provided funding for sequencing through their Audacious Experiment fund. Sequence data was generated by the High Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, which is subsidized by Wellcome trust grant reference 090532/Z/09/Z. A.J.P. is a Jenner Investigator and James Martin Senior Fellow. J.D.G. receives funding from the BBSRC and GlaxoSmithKline in the form of an iCASE studentship. V.C. is funded by the Medical Research Council. D.F.K. receives salary support from the NIHR Oxford Biomedical Research Centre.

The authors are grateful to all the study participants, and to Dr. Alastair MacGregor for assisting with sample collection at the Oxford Vaccine Group. The authors also thank Craig Waugh (Weatherall Institute for Molecular Medicine, Oxford) for help with cell sorting. The authors acknowledge the support of the National Institute for Health Research Clinical Research Network.

References

1 Galson JD, Pollard AJ, Trück J, Kelly DF. Studying the antibody repertoire after vaccination: practical applications. *Trends Immunol* 2014; **35**: 319–331.

2 Jiang N, He J, Weinstein JA, Penland L, Sasaki S, He X-S *et al.* Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci Transl Med* 2013; **5**: 171ra19.

3 Vollmers C, Sit R, Weinstein JA, Dekker CL, Quake SR. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc Natl Acad Sci USA* 2013; **110**: 13463–8.

4 Ademokun A, Wu Y-C, Martin V, Mitra R, Sack U, Baxendale H *et al.* Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging Cell* 2011; **10**: 922–30.

5 Laserson U, Vigneault F, Gadala-Maria D, Yaari G, Uduman M, Vander Heiden JA *et al.* High-resolution antibody dynamics of vaccine-induced immune responses. *Proc Natl Acad Sci U S A* 2014; **111**: 4928–33.

6 Mroczek ES, Ippolito GC, Rogosch T, Hoi KH, Hwangpo TA, Brand MG *et al.* Differences in the Composition of the Human Antibody Repertoire by B Cell Subsets in the Blood. *Front Immunol* 2014; **5**: 1–14.

7 Kaplinsky J, Li A, Sun A, Coffre M, Koralov SB, Arnaout R. Antibody repertoire deep sequencing reveals antigen-independent selection in maturing B cells. *Proc Natl Acad Sci* 2014; **111**: 2622–9.

8 Gold R, Lepow ML, Goldschneider I, Draper TL, Gotschlich EC. Clinical evaluation of group A and group C meningococcal polysaccharide vaccines in infants. *J Clin Invest* 1975; **56**: 1536–47.

9 De Wals P, De Serres G, Niyonsenga T. Effectiveness of a mass immunization campaign against serogroup C meningococcal disease in Quebec. *JAMA* 2001; **285**: 177–81.

10 Jokhdar H, Borrow R, Sultan A, Adi M, Riley C, Fuller E *et al.* Immunologic hyporesponsiveness to serogroup C but not serogroup A following repeated meningococcal A/C polysaccharide vaccination in Saudi Arabia. *Clin Diagn Lab Immunol* 2004; **11**: 83–8.

11 Kelly DF, Snape MD, Cutterbuck EA, Green S, Snowden C, Diggle L *et al.* CRM197-conjugated serogroup C meningococcal capsular polysaccharide, but not the native polysaccharide, induces persistent antigen-specific memory B cells. *Blood* 2006; **108**: 2642–7.

12 Clutterbuck EA, Lazarus R, Yu L-M, Bowman J, Bateman EAL, Diggle L *et al.* Pneumococcal conjugate and plain polysaccharide vaccines have divergent effects on antigen-specific B cells. *J Infect Dis* 2012; **205**: 1408–16.

13 Weill J-C, Weller S, Reynaud C-A. Human marginal zone B cells. *Annu Rev Immunol* 2009; **27**: 267–85.

14 Ramasamy M. B cell responses to conjugate and polysaccharide meningococcal vaccines. *PhD Thesis Univ Oxford, UK* 2012.

15 Ramasamy MN, Clutterbuck EA, Bowman J, Omar O, Amber J, Blanchard-rohner G *et al.* United Kingdom Randomized Clinical Trial To Evaluate the Immunogenicity of Quadrivalent Meningococcal Conjugate and Polysaccharide Vaccines. *Clin Vaccine Immunol* 2014; **21**: 1164–8.

16 Odendahl M, Mei H, Hoyer BF, Jacobi AM, Hansen A, Muehlinghaus G *et al.* Generation of migratory antigen-specific plasma blasts and mobilization of resident plasma cells in a secondary immune response. *Blood* 2005; **105**: 1614–21.

17 Weller S, Braun MC, Tan BK, Rosenwald A, Cordier C, Conley ME *et al.* Human blood IgM 'memory' B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 2004; **104**: 3647–54.

18 Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 2000; **13**: 37–45.

19 Mitchell R, Kelly DF, Pollard AJ, Trück J. Polysaccharide-specific B cell responses to vaccination in humans. *Hum Vaccin Immunother* 2014; **10**: 1661–1668.

20 Briney BS, Willis JR, McKinney BA, Crowe JE. High-throughput antibody sequencing reveals genetic evidence of global regulation of the naïve and memory repertoires that extends across individuals. *Genes Immun* 2012; **13**: 469–73.

21 Wu Y-C, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 2010; **116**: 1070–8.

22 Cerutti A, Cols M, Puga I. Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat Rev Immunol* 2013; **13**: 118–32.

23 Berkowska M, Driessen GJ, Bikos V, Grosserichter-wagener C, Cerutti A, He B *et al.* Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways. *Blood* 2011; **118**: 2150–2159.

24 Aguilera I, Melero J, Nuñez-Roldan A, Sanchez B. Molecular structure of eight human autoreactive monoclonal antibodies. *Immunology* 2001; **102**: 273–80.

25 Trück J, Ramasamy MN, Galson JD, Rance R, Parkhill J, Lunter G *et al.* Identification of Antigen-Specific B Cell Receptor Sequences Using Public Repertoire Analysis. *J Immunol* 2014; **194**: 252–261.

26 Parameswaran P, Liu Y, Roskin KM, Jackson KKL, Dixit VP, Lee J-Y *et al.* Convergent Antibody Signatures in Human Dengue. *Cell Host Microbe* 2013; **13**: 691–700.

27 Jackson KJL, Liu Y, Roskin KM, Glanville J, Hoh RA, Seo K *et al.* Human Responses to Influenza Vaccination Show Seroconversion Signatures and Convergent Antibody Rearrangements. *Cell Host Microbe* 2014; **16**: 1–10.

28 Saada R, Weinberger M, Shahaf G, Mehr R. Models for antigen receptor gene rearrangement: CDR3 length. *Immunol Cell Biol* 2007; **85**: 323–332.

29 Christensen H, May M, Bowen L, Hickman M, Trotter CL. Meningococcal carriage by age: a systematic review and meta-analysis. *Lancet Infect Dis* 2010; **10**: 853–61.

30 Li G-M, Chiu C, Wrammert J, McCausland M, Andrews SF, Zheng N-Y *et al.* Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. *Proc Natl Acad Sci U S A* 2012; **109**: 9047–52.

31 Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T *et al.* The Complete Nucleotide Sequence of the Human Immunoglobulin Heavy Chain Variable Region Locus. *J Exp Med* 1998; **188**: 2151–2162.

32 Michaelsen TE, Garred P, Aase A. Human IgG subclass pattern of inducing complement-mediated cytolysis depends on antigen concentration and to a lesser extent on epitope patchiness, antibody affinity and complement concentration. *Eur J Immunol* 1991; **21**: 11–6.

33 Hjelholt A, Christiansen G, Sørensen US, Birkelund S. IgG subclass profiles in normal human sera of antibodies specific to five kinds of microbial antigens. *Pathog Dis* 2013; **67**: 206–13.

34 Aalberse RC, Stapel SO, Schuurman J, Rispens T. Immunoglobulin G4: an odd antibody. *Clin Exp Allergy* 2009; **39**: 469–77.

35 Gitlin AD, Shulman Z, Nussenzweig MC. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature* 2014; **509**: 637–40.

36 Collins AM, Jackson KJL. A Temporal Model of Human IgE and IgG Antibody Function. *Front Immunol* 2013; **4**: 235.

37 Jackson KJL, Wang Y, Collins AM. Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels. *Immunol Cell Biol* 2014; : 1–5.

38 Medina F, Segundo C, Jiménez-Gómez G, González-García I, Campos-Caro A, Brieva J a. Higher maturity and connective tissue association distinguish resident from recently generated human tonsil plasma cells. *J Leukoc Biol* 2007; **82**: 1430–1436.

39 Oliveira TG, Milani SR, Travassos LR. Polyclonal B-cell activation by Neisseria meningitidis capsular polysaccharides elicit antibodies protective against Trypanosoma cruzi infection in vitro. *J Clin Lab Anal* 1996; **10**: 220–8.

40 DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD *et al.* In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* 2015; **21**: 86–91.

41 Brochet X, Lefranc M-P, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* 2008; **36**: W503–8.

42 Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 2011; **21**: 936–9.

43 Wilming LG, Gilbert JGR, Howe K, Trevanion S, Hubbard T, Harrow JL. The vertebrate genome annotation (Vega) database. *Nucleic Acids Res* 2008; **36**: D753–60.

44 Longerich S, Tanaka A, Bozek G, Nicolae D, Storb U. The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. *J Exp Med* 2005; **202**: 1443–1454.

45 Team RDC. R: A language and environment for statistical computing. *R Found Stat Comput Vienna, Austria* 2008.

46 Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer; 1st ed. 2009. Corr. 3rd printing 2010 edition, 2009.

Figure Legends

Fig. 1. Study plan and sampling protocol. Four participants were given a conjugate, and five participants given a polysaccharide MenACWY vaccine at day 0. All participants were given the conjugate vaccine at day 28. All participants had blood taken 7 days after each vaccine for PC repertoire analysis. Each participant also had blood taken on either the day of the first vaccine or the day of the second vaccine (but not both) for repertoire analysis of baseline B cell subsets. Fluorescence-activated cell sorting was used to isolated the different B cell subsets at each visit.

Fig. 2. Validation of clustering parameters. A) Distribution of reads with different CDR3 AA distances to their closest neighbor. Calculated separately for each participant, and mean values plotted for each cell subset. B) Correlation between degree of over-sequencing (number of sequences / number of cells) and the proportion of reads with a distance of 2 or fewer AA differences to their closest neighbor, calculated for each cell subset. Error bars indicate ± SEM. Regression line, and correlation coefficient with corresponding p value are shown. C) Correlation between cell number for a particular sample, and the number of sequences (after clustering) for that sample.

Fig. 3. Repertoire differences in the different cell subsets. Differences in A) clonality, B) V gene mutation and C) CDR3 AA length between the different cell subsets. For A,B and C, boxes show locations of 25, 50 and 75th percentiles. Whiskers show data within 1.5x the interquartile range. Samples from all participants and timepoints were included together in the analysis. D) Principal component analysis including just the baseline cell subsets, just the plasma cell subsets, or all cell subsets together, utilizing the clonality, V gene mutation, and CDR3 length

variables. E) Proportion of sequences utilizing different V gene families in the different cell subsets. Error bars indicate ± SEM.

Fig. 4. Percent of CDR3 AA sequences shared either within or between the two vaccine groups in the different subsets on the different days. Within vaccine group shared sequences are shared between at least two participants. Percent shared between A and B = (A∩B/sum(A,B)*100).

Fig. 5. Relationships between baseline subsets and PCs produced after vaccination. A) Circos diagrams from one representative participant in the conjugate group, and one representative participant in the polysaccharide group showing the relationship between baseline subsets, and HLA-DR+ IgM and IgG PCs after the first vaccination. Colored sections represent baseline subsets, and black section represents PCs after vaccination. The length of each section represents the number of different sequences comprising that subset. Sequences are ordered clockwise by abundance, which is represented by the grey histogram. Colored lines join sequences that are present in the PCs, and any one of the baseline subsets. Sequences shared within the different baseline subsets are not shown. B) Probability of shared PC sequences at visit 2 being shared with each baseline cell subset for HLA-DR+ IgG and IgM PCs after the first vaccine (either conjugate or polysaccharide depending on vaccine group). C) Same as B, but with HLA-DR+ PCs at visit 4 after the second vaccine (conjugate for both vaccine groups). Error bars indicate ± SEM. For sequences to be defined as shared they had to have an exact CDR3 AA sequence match, and use the same V and J genes.

Fig. 6. IgG usage and mutation in HLA-DR+ PC sequences after vaccination in the two study groups. A) Proportion of sequences of each IgG subclass after the first and second vaccines. B)

Percent of mutated nucleotides in the V gene after the first and second vaccines for IgM and IgG PCs. C) Same as B, but split by IgG subclass. Error bars indicate ± SEM. Comparisons performed using the t test. *p<0.05, **p<0.005.

Figure 1

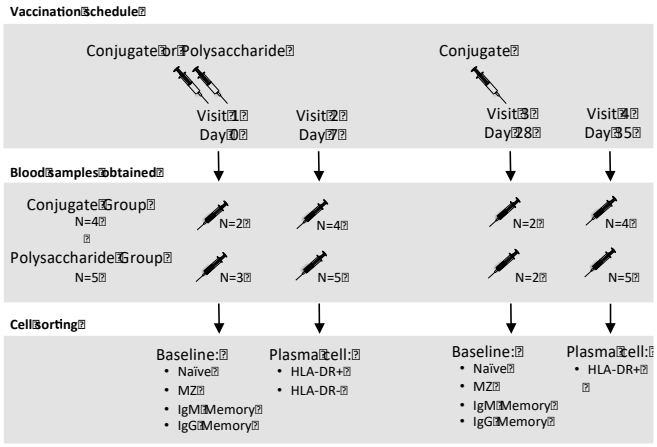

Vaccination schedule
Conjugate or Polysaccharide
Conjugate
Visit 1
Day 0
Visit 2
Day 7
Visit 3
Day 28
Visit 4
Day 35
Blood samples obtained
Conjugate Group
N=4
N=2
N=4
N=2
N=4
Polysaccharide Group
N=5
N=3
N=5
N=2
N=5
Cell Sorting
Baseline:
Naïve
MZ
IgM Memory
IgG Memory
Plasma cell:
HLA-DR+
HLA-DR-
Baseline:
Naïve
MZ
IgM Memory
IgG Memory
Plasma cell:
HLA-DR+

Figure 2

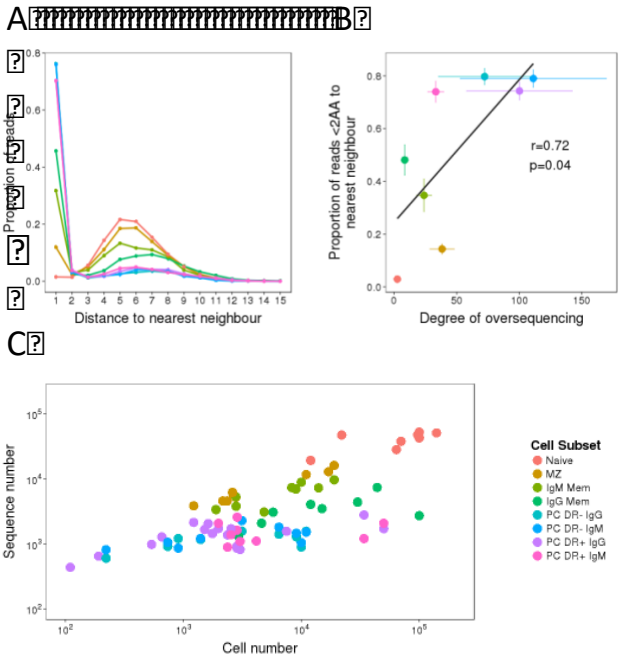

A
B
C
Distance to nearest neighbour
Degree of oversequencing
r=0.72
p=0.04
Sequence number
Cell number
Cell Subset
Naive
MZ
IgM Mem
IgG Mem
PC DR- IgG
PC DR- IgM
PC DR+ IgG
PC DR+ IgM

Figure 3

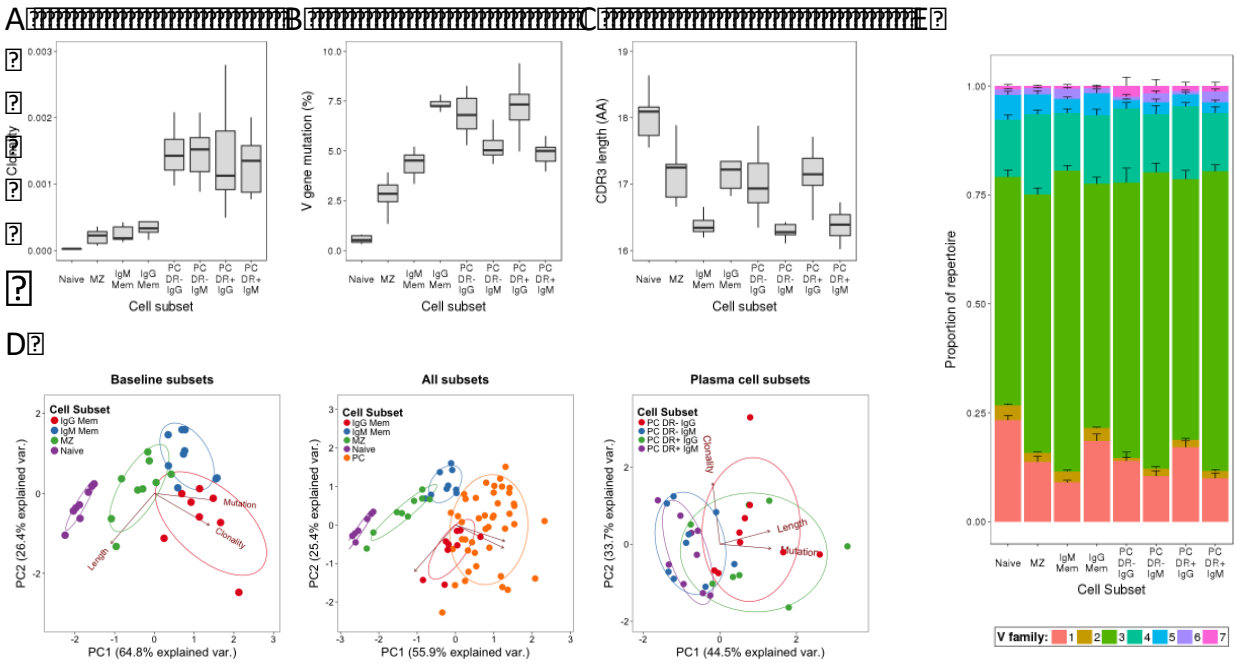

A
B
C
E
D
Cell subset
V gene mutation (%)
CDR3 length (AA)
Naive
MZ
IgM Mem
IgG Mem
PC DR- IgG
PC DR- IgM
PC DR+ IgG
PC DR+ IgM
Baseline subsets
All subsets
Plasma cell subsets
Cell Subset
IgG Mem
IgM Mem
MZ
Naive
PC
PC DR- IgG
PC DR- IgM
PC DR+ IgG
PC DR+ IgM
Mutation
Clonality
Length
PC1 (64.8% explained var.)
PC2 (26.4% explained var.)
PC1 (55.9% explained var.)
PC2 (25.4% explained var.)
PC1 (44.5% explained var.)
PC2 (33.7% explained var.)
Proportion of repertoire
V family: 1 2 3 4 5 6 7

Figure 4

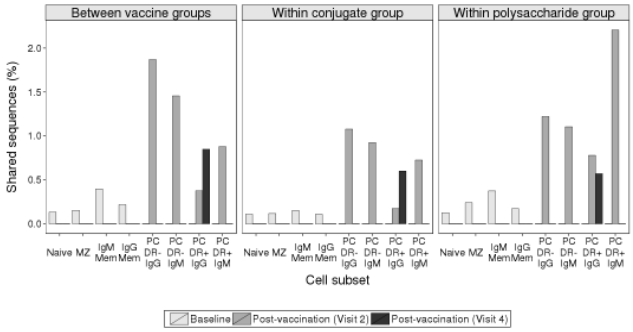

Between vaccine groups
Within conjugate group
Within polysaccharide group
Shared sequences (%)
2.0
1.5
1.0
0.5
0.0
Naive
MZ
IgM Mem
IgG Mem
PC DR- IgG
PC DR- IgM
PC DR+ IgG
PC DR+ IgM
Cell subset
Baseline
Post-vaccination (Visit 2)
Post-vaccination (Visit 4)

Figure 5

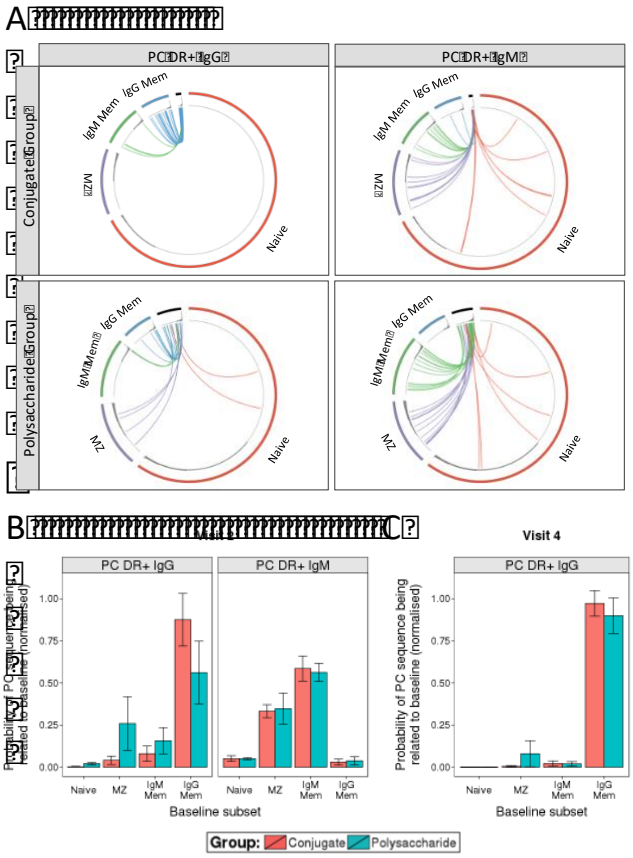
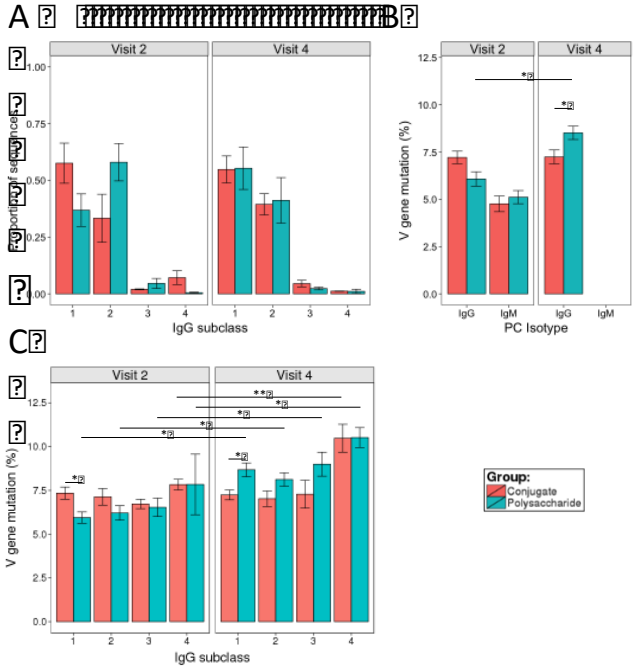

A
PC DR+ IgG
PC DR+ IgM
Conjugate Group
Polysaccharide Group
IgM Mem
IgG Mem
MZ
Naive
B
C
Visit 4
PC DR+ IgG
PC DR+ IgM
Probability of PC sequence being related to baseline (normalised)
Naive
MZ
IgM Mem
IgG Mem
Baseline subset
Group: Conjugate Polysaccharide

Figure 6


A
B
Visit 2
Visit 4
IgG subclass
V gene mutation (%)
PC Isotype
IgG
IgM
C
Group:
Conjugate
Polysaccharide