development of broadly neutralizing antibodies (bNAbs) in order to aid the development of preventative HIV-1 vaccines [13]. Interestingly, bNAb precursors may be preferred in vaccine design over observed bNAbs because they may have a wider binding profile, and the sequences of such precursors can be inferred using methods of phylogenetic ancestral sequence reconstruction [14]. This requires an understanding of the diversity and dynamics of B-cell clones during infection, and significant strides have been made, both in tracking the coevolution of B-cell and viral lineages, [15–17] and in finding potential bNAb precursors using phylogenetic methods [14].

Despite these advances, more general trends in B-cell diversity and clonal dynamics during HIV-1 infection are still poorly understood, particularly during early infection. While some studies have contrasted the plasmablast, naive and memory B-cell content between early and chronic infections [18], the diversity of patient 'antibodyomes' has not yet been characterized, either in comparison to healthy controls or through time in individuals that have recently seroconverted. Although early anti-retroviral therapy (ART) has been shown to have a significant effect on viral divergence (e.g. [19]), the corresponding effect on B-cell clonal diversity under ART is unknown. While it may be expected that B-cell populations in HIV+ individuals are more clonal than in uninfected individuals, and that the degree of B-cell clonal relatedness is related to the size of the concurrent viral population, these associations have not been explicitly tested.

Most previous studies that investigated the repertoire of B-cell receptor (BCR) diversity in peripheral blood (as opposed to those that focused on specific clonal lineages, e.g. [15]) have typically used non-phylogenetic approaches to summarize B-cell diversity. One successful approach has been to use single-linkage clustering to group sequences into clusters (or clones), and then to infer clonal expansions by measuring properties of the size distribution of clones using entropy scores, such as the Gini index [8]. Others studies have used alternative statistics, including mean clone size, the number of unique IgHV-D-J allele combinations, and genetic distances between IgV segment sequences and their respective germline homologues (e.g. [7]). Both sets of approaches have shown promise in studying B-cell cancers, which result from significant clonal expansions of usually one BCR lineage [20], but it is unknown if a similar approach will be informative when studying immune responses to HIV infection.

To better understand B-cell repertoire dynamics during early HIV infection and the degree to which it is modulated by ART, we used deep-sequencing to capture the diversity of Ig heavy-chain sequences from eight HIV patients enrolled in the short pulse anti-retroviral therapy at seroconversion (SPARTAC) trial [21]. This is, to our knowledge, the first time high-throughput BCR sequencing has been applied to studying general B-cell clonal diversity during ART and early HIV infection. Patients were enrolled an estimated 12–95 days after seroconversion, and were sampled at up to eight time points over approximately 2 years. At each time point, B-cell repertoire sequencing was performed and both viral load and $CD4^+$ T-cell counts were measured. Three patients were untreated, while five received ART for the first 48 weeks of the study only. This study design not only allows us to track individual B-cell clones during early infection, but also to test for associations between the dynamics of B-cell sequence diversity and clinical variables, including treatment status.

For each patient and each time point, we used a high-throughput Illumina MiSeq platform to obtain paired-end reads from Ig heavy-chain sequences that represent the mixture of antibody classes in peripheral blood. Within each patient, we extended a previous single-linkage clustering approach [8] to classify sequences into clones and track their relative frequencies through time. We also explored a number of statistics in order to quantify BCR sequence diversity from HIV-1 infected patients, and to compare this diversity to that observed in a cohort of HIV-negative controls. While some general patterns were observed, overall we found a high degree of heterogeneity in B-cell clonal dynamics both among patients and through time.

2. Material and methods

(a) HIV patients

Peripheral blood mononuclear cells (PBMCs) were isolated from eight patients with primary HIV-1 infection recruited from the SPARTAC study [21]. All had recently seroconverted before enrolment in the trial (an estimated 12–93 days before enrolment; median = 56 days). Patients 1–3 were untreated during the study period. Patients 4–8 received an ART regimen from week 0 to week 48, after which treatment was suspended. Patients were sampled between six and eight times over 108 weeks of the study (all time points are defined as weeks after start of the study, defined as week 0). All patients were sampled at weeks 4, 16, 24, 52, 60 and 108, whereas four patients were also sampled at week 0, and six at week 12. To ensure consistency among patients, and to ensure an equal number of time points during and after ART, only the former time points were used in analysis.

(b) RT-PCR

RT-PCR reagents were purchased from Invitrogen and primers (supplied by Sigma Aldrich) are described by Van Dongen *et al.* [22] and in electronic supplementary material, table S1. Reverse transcription (RT) was performed using 500 ng of total PBMC RNA mixed with 1 µl JH reverse primer (10 µM), 1 µl dNTPs (0.25 mM) and RNase-free water added to make a total volume of 11 µl. This was incubated for 5 min at 65°C, and 4 µl First strand buffer, 1 µl DTT (0.1 M), 1 µl RNaseOUT™ Recombinant Ribonuclease Inhibitor and 1 µl SuperScript™ III reverse transcriptase (200 units μl^{-1}) was added. RT was performed at 50°C for 60 min before heat-inactivation at 70°C for 15 min. PCR amplification of cDNA (5 µl of the RT product) was performed with the JH reverse primer and the FR1 forward primer set pool (0.25 µM each), using 0.5 µl Phusion® High-Fidelity DNA Polymerase (Finnzymes), 1 µl dNTPs (0.25 mM), 1 µl DTT (0.25 mM), per 50 µl reaction. The following PCR programme was used: 3 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C, with a final extension cycle of 7 min at 72°C on an MJ Thermocycler.

(c) Sequencing and reference-based V-D-J assignment

MiSeq libraries were prepared using Illumina protocols and sequenced by 150 bp paired-ended MiSeq (Illumina). MiSeq reads were filtered for base quality (median more than 32) using QUASR (http://sourceforge.net/projects/quasr) [23]. Sequences were concatenated and a gap inserted between the forward and reverse reads (average gap length approx. 35 nucleotides; electronic supplementary material, figure S2). Non-Ig sequences were removed; only those reads with significant similarity to reference IgHV and IgHJ genes from the ImMunoGeneTics (IMGT) database [24] were retained, as determined using BLAST [25] with *E*-value thresholds of 1×10^{-10} and 1×10^{-3}