Supporting Information

Vollmers et al. 10.1073/pnas.1312146110

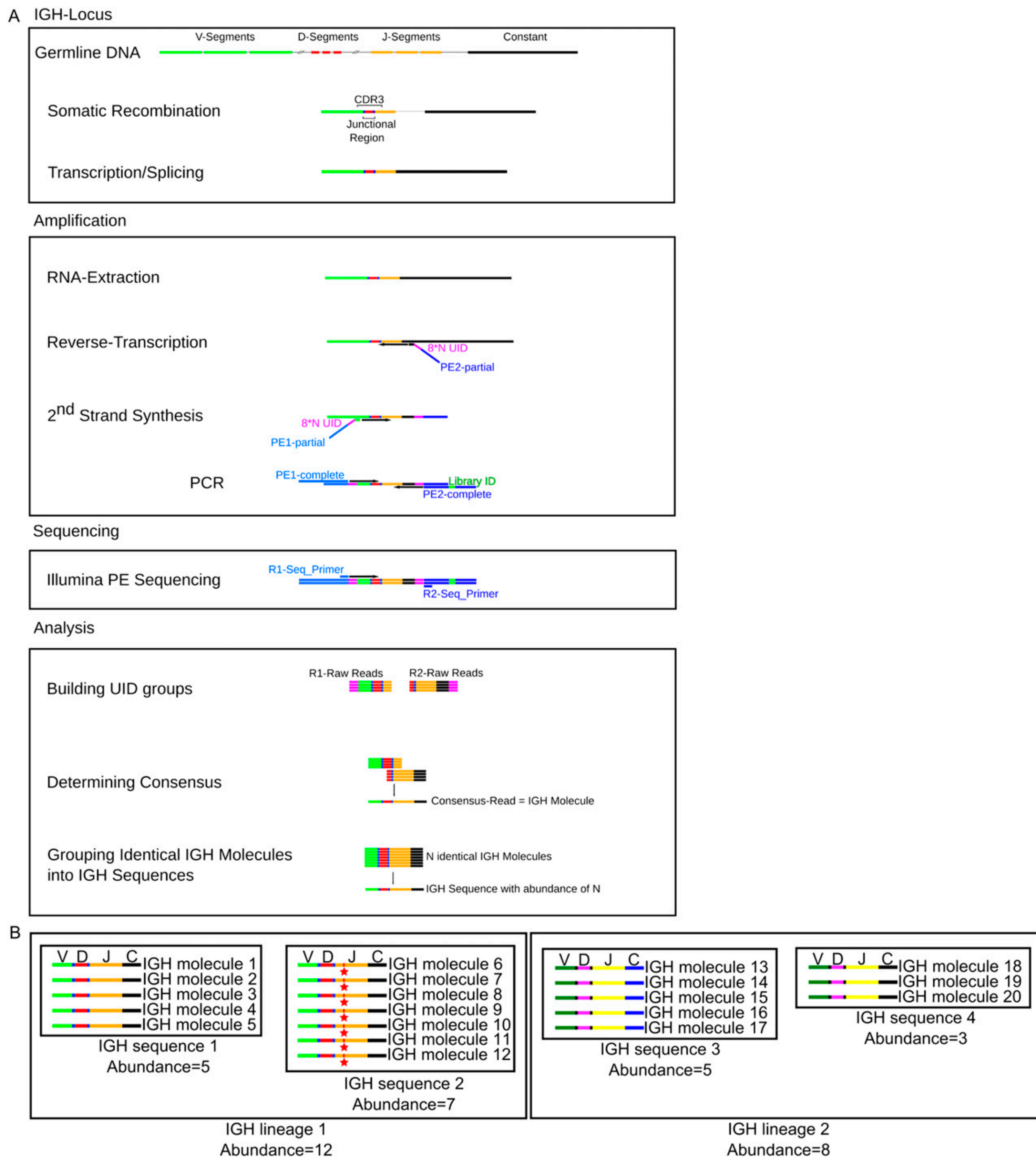

Fig. S1. Consensus read library generation and sequencing approach. (*A*) The human immunoglobulin heavy chain locus undergoes somatic recombination to form its variable region. After transcription, the isotype-determining constant region is spliced to the variable region to form IGH RNA molecules. IGH RNA molecules are reverse transcribed using primers specific for the 5′ end of the constant region using pool isotype-specific primers. Second-strand synthesis uses a primer pool specific to the framing region 3 (FR3) of IGH V segments. Both second-strand and reverse transcription primers contain 8 random nt and partial sequencing adapter sequences. A PCR is used to amplify the IGH sequences and complete the sequencing adapters. After sequencing, 16 random nt are used to build unique identifier (UID) groups of raw sequencing reads that all originated from the same original IGH RNA molecule. High-quality consensus reads are generated for each IGH molecule, and identical IGH molecules are grouped into IGH sequences. (*B*) Schematic of how IGH molecules were grouped into IGH sequences, which were then clustered into IGH lineages. Identical IGH molecules are grouped into IGH sequences. Clustering of IGH sequences requires 90% similarity in the junctional region but allows for differences in the other parts of the IGH sequences, like mutations (red stars) or different isotypes.

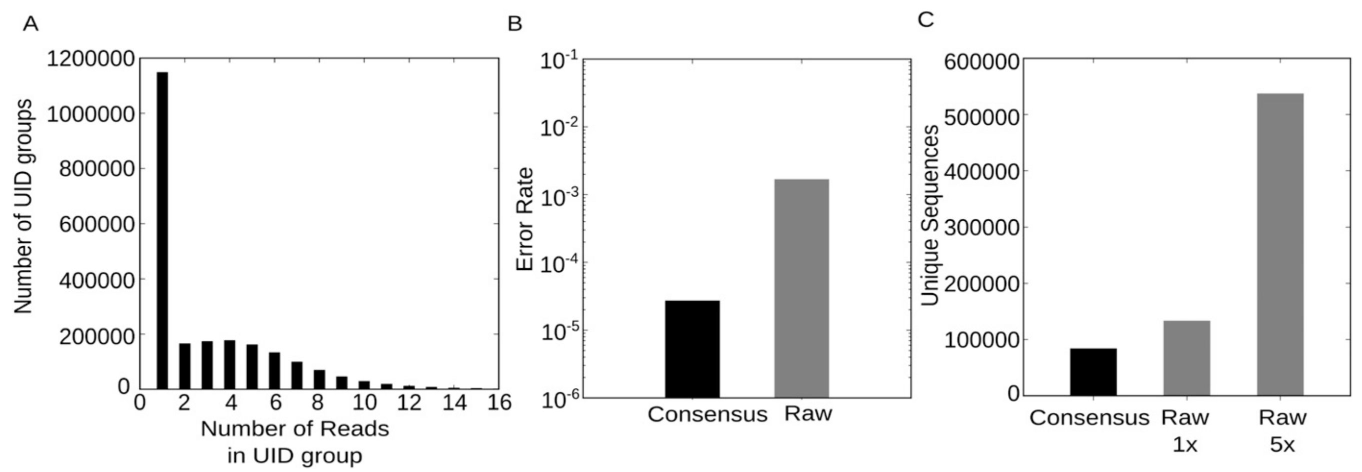

Fig. S2. Consensus read validation. (*A*) Histogram showing the number of consensus groups containing the indicated number of reads. (*B*) Estimated error rates for consensus reads and raw Illumina reads are shown. (*C*) The number of unique IGH sequences among ~500,000 consensus reads, ~500,000 raw reads, or ~2,500,000 raw reads is shown. Average mutation rate and abundance of sequences separated by isotype. Average abundance of mutated or naive IgM, IgA, or IgM sequences.

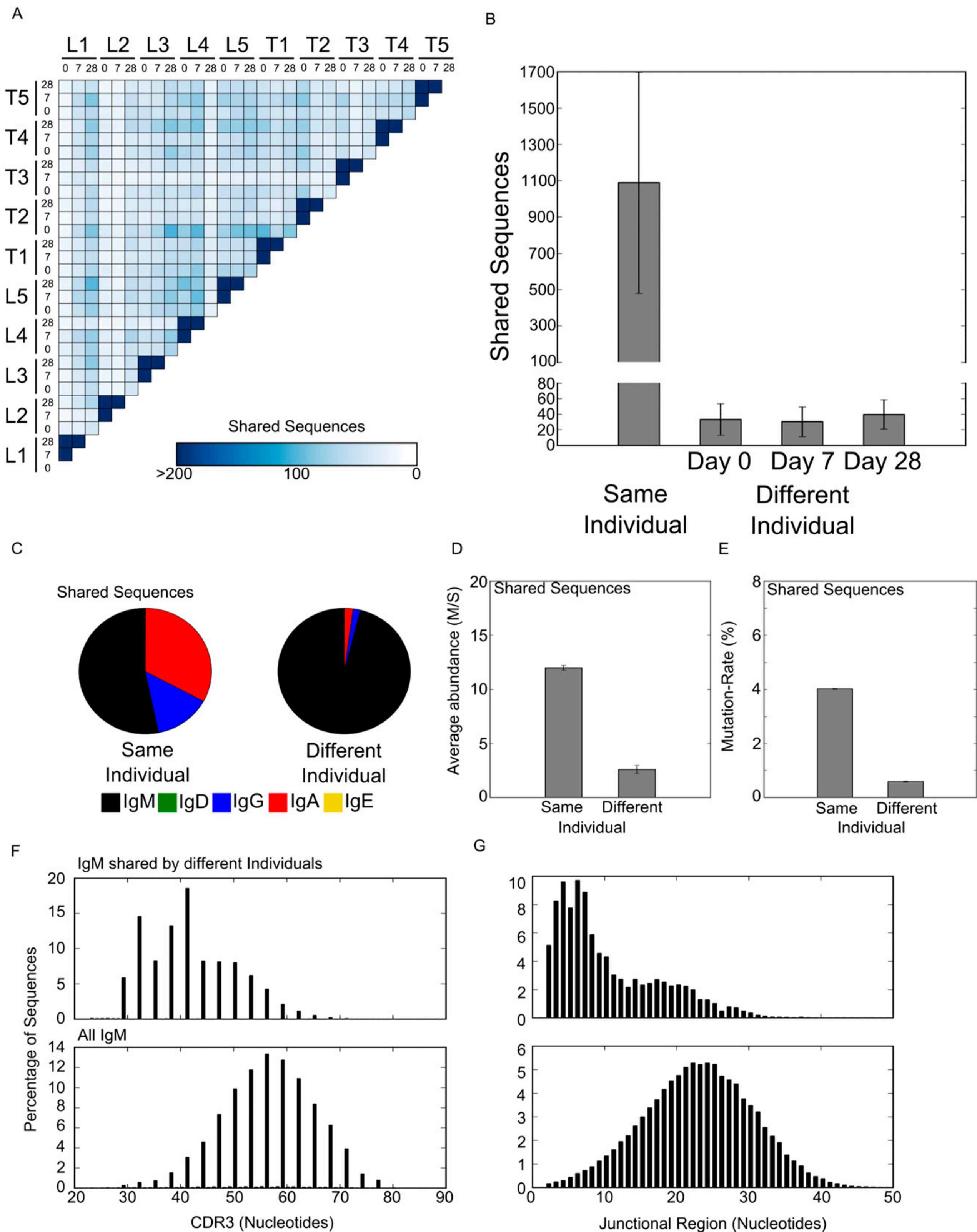

Fig. S3. IGH sequences shared between samples and individuals. (*A*) Annotated heatmaps indicating the rate of amino acid sequence overlap between two samples. White indicates low overlap, and blue indicated high overlap. The number of shared sequences is indicated for each sample pair. (*B*) Average number of amino acid sequences between different samples of the same individual or same days of different individuals. (*C*) Isotype distribution, (*D*) average abundance, and (*E*) mutation rate of sequences shared between different samples of the same individual or different individuals. (*F* and *G*) Histograms of (*F*) CDR3 length and (*G*) junctional region length distribution of all IgM sequences and the IgM sequences shared between different individuals.

PNAS PNAS PNAS PNAS

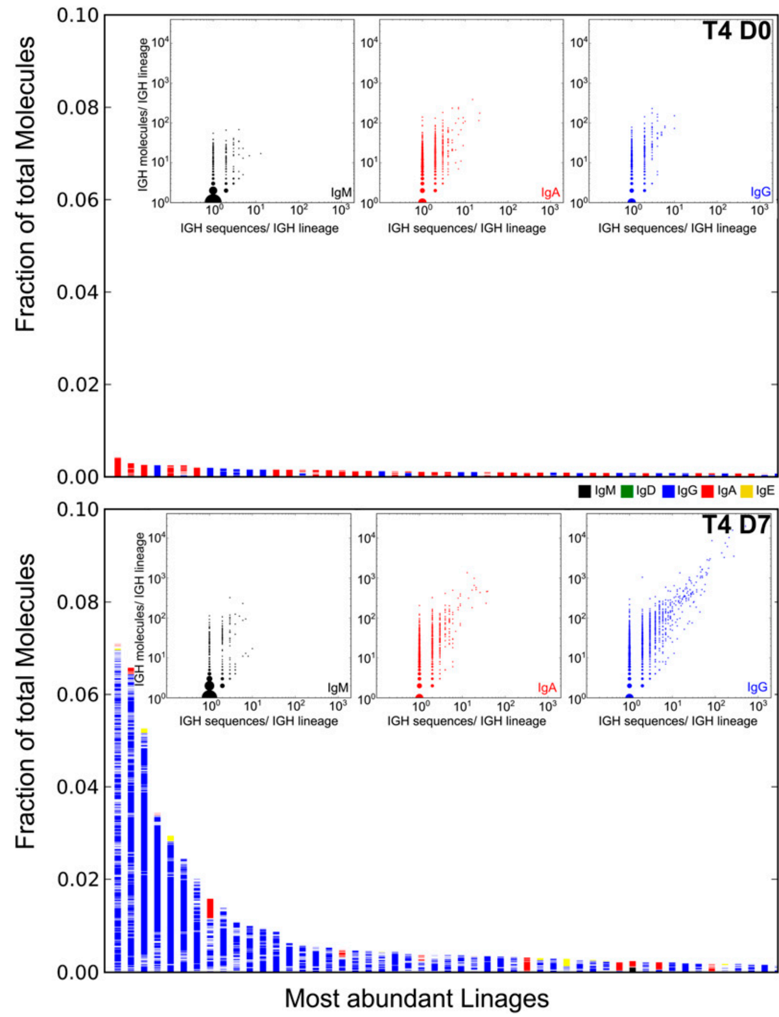

Fig. S4. Lineage structure. The most abundant lineages are shown for T4 D0 and T4 D7. Each horizontal bar represents an IGH lineage, which is separated into its IGH sequences by white vertical lines. The color of the bar segments indicates the isotype, whereas their length indicates the abundances of IGH sequences. The inlets for T4 D0 and T4 D7 show scatterplots of IGH molecules/IGH lineage and IGH sequences/IGH lineage for each isotype separately, with the area of each dot proportional to the square root of its abundance.

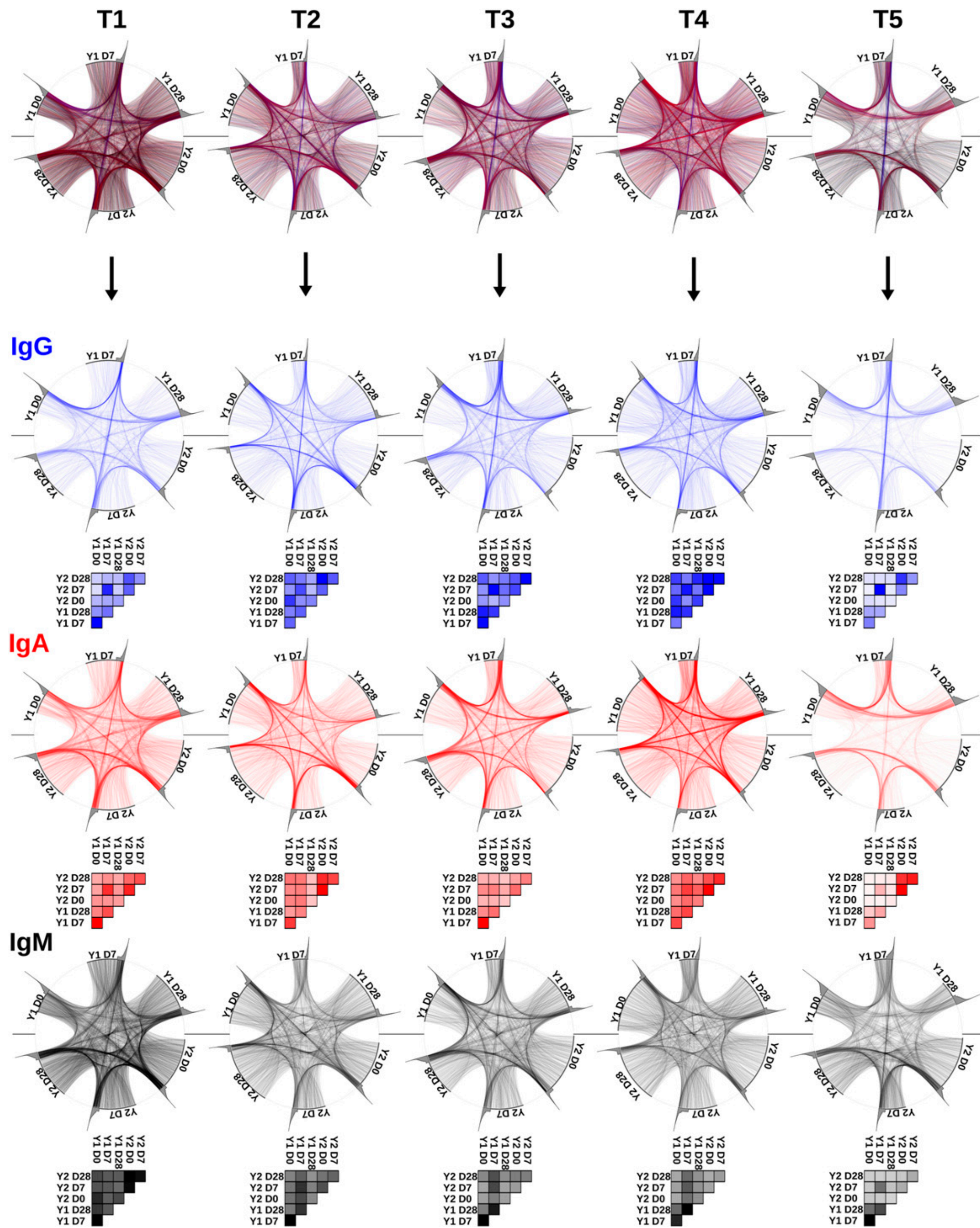

Fig. S5. Lineage recall. Visualization of lineages shared between time points in five individuals. Data were subsampled to 50,000 IGH molecules for each time point to normalize for variability in sampling depth. Lineages of each time point are plotted on circumference of the circle, with the gray area representing abundance of the respective lineages. Lineages present in two time points are connected with lines colored according to their isotype. Shared lineages of all or only IgG, IgA, and IgM are shown. Numbers of lineages shared between indicated time points normalized to the overall amount of lineages are shown as heatmaps for all isotypes and five individuals.

Table S1. Primer sequences used in this study

Step/relative amount in pool	Name	Sequence (5′–3′)
Reverse transcription		
1	G	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNAAGACCGATGGGCCCTTG
1	A	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAAGACCTTGGGGCTGGT
1	M	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGGGAATTCTCACAGGAGACG
1	D	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGGGTGTCTGCACCCTGATA
1	E_1	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAAGACGGATGGGCTCTGT
1	E_2	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNTTGCAGCAGCGGGTCAAGGG
Second-strand synthesis		
1	V1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNAGCCTACATGGAGCTGAGC
1	V2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNAGGTGGTCCTTACAATGACCAAC
1	V3_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTCTGCAAATGAACAGCCTGA
0.2	V3_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTGTTCAAATGAGCAGTCTGAGAG
0.2	V3_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTCTGCAAATGGGCAGCCTGA
1	V4/6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTTCTCCCTGAAGCTGAACTCTG
1	V5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGCCTACCTGCAGTGGAGCAG
1	V6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTTCTCCCTGCAGCTGAACTCTG
1	V7_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGCATATCTGCAGATCAGCAGC
1	V7_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNCAGATCAGCAGCCTAAAGGC
PCR		
1	PE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
1	PE2	CAAGCAGAAGACGGCATACGAGATAACGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATT

Partial paired end (PE) 1 adapters are in red, and partial PE2 adapters are in blue. Random nucleotides are in pink, and Illumina barcodes are in turquoise. Gene-specific nucleotides are in black, and complete PE1 and PE2 are in green.