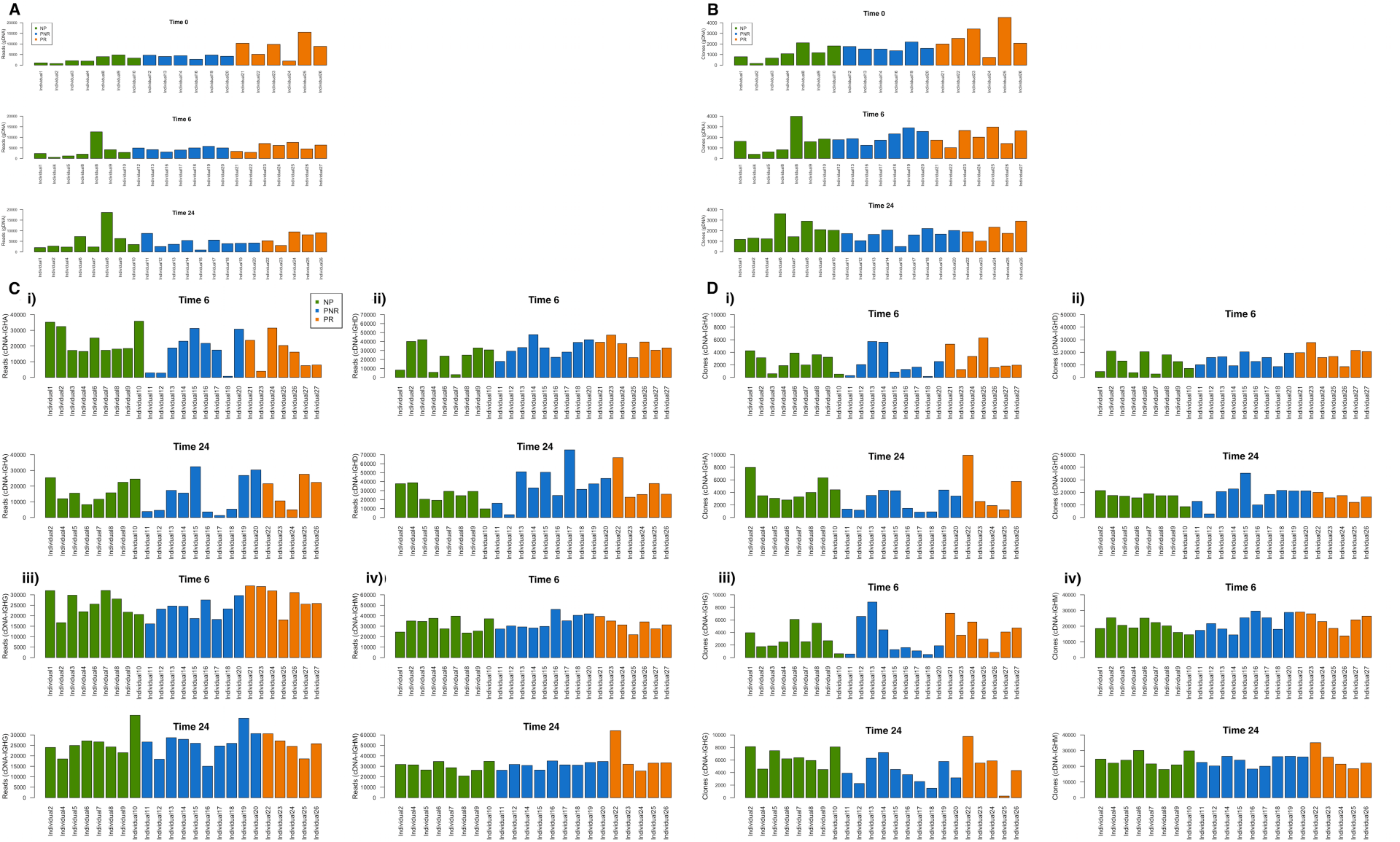
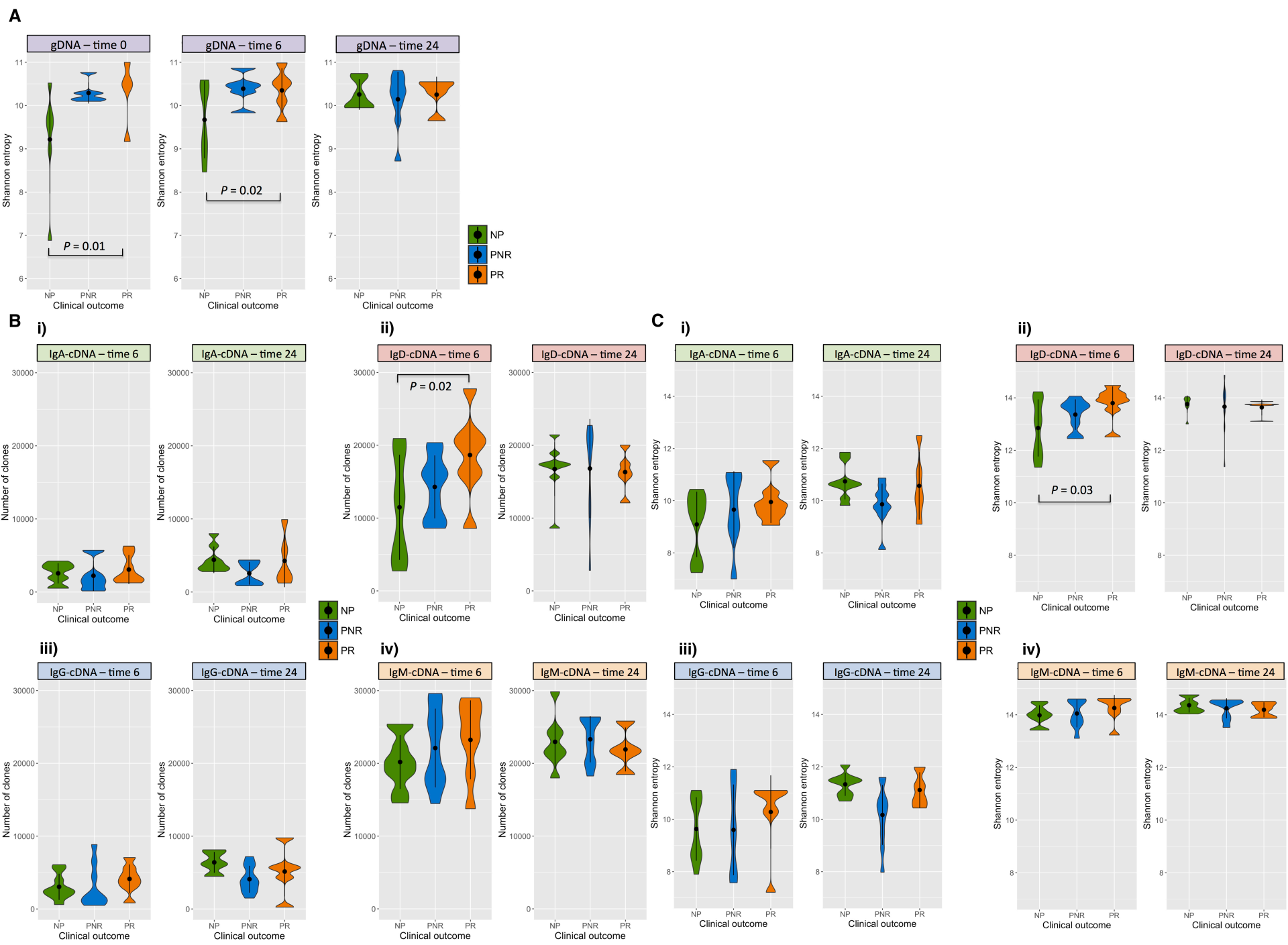
Supplementary Information

Characterizing pre-transplant and post-transplant kidney rejection risk by B cell immune repertoire sequencing

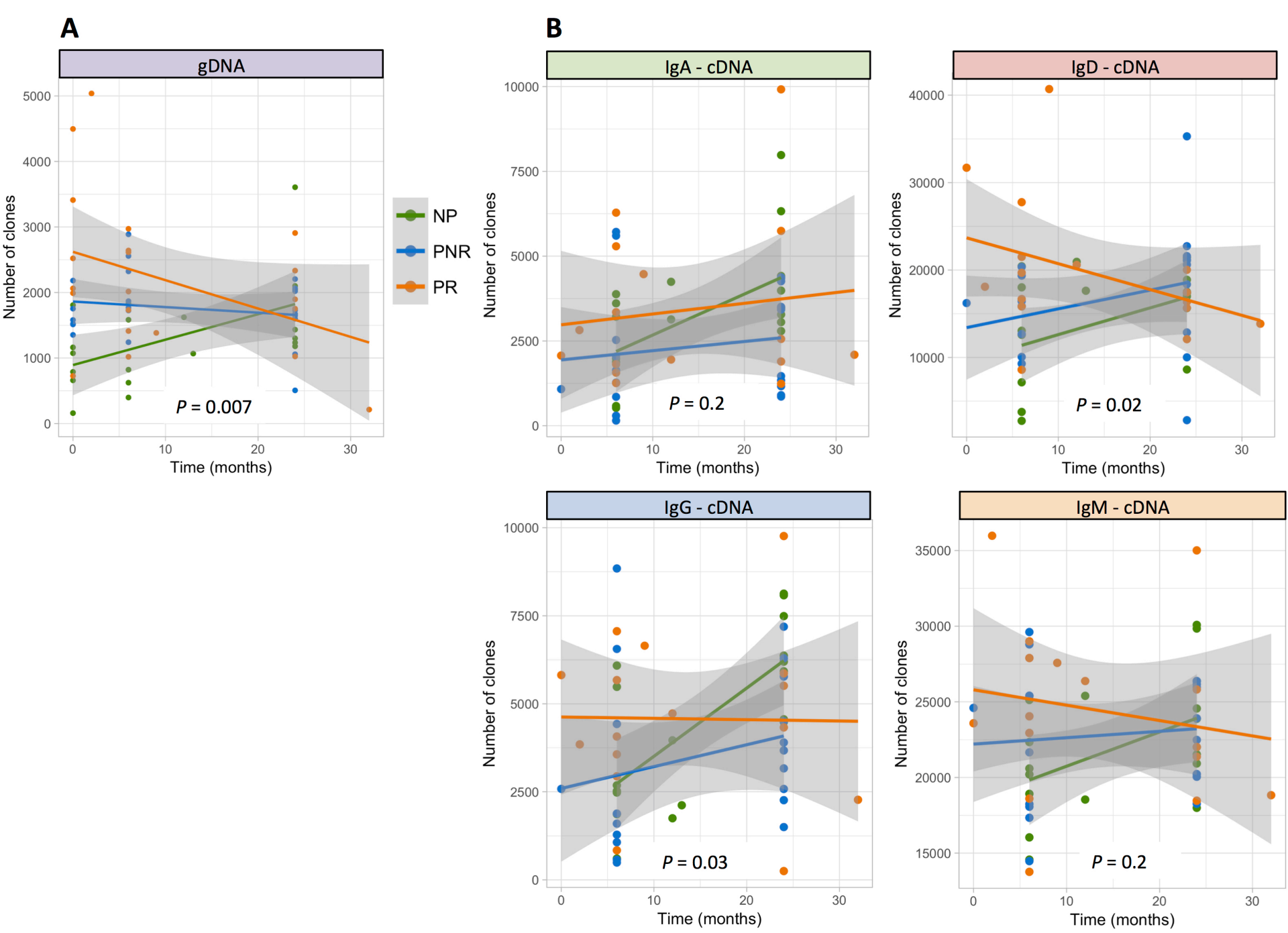
Pineda et al.



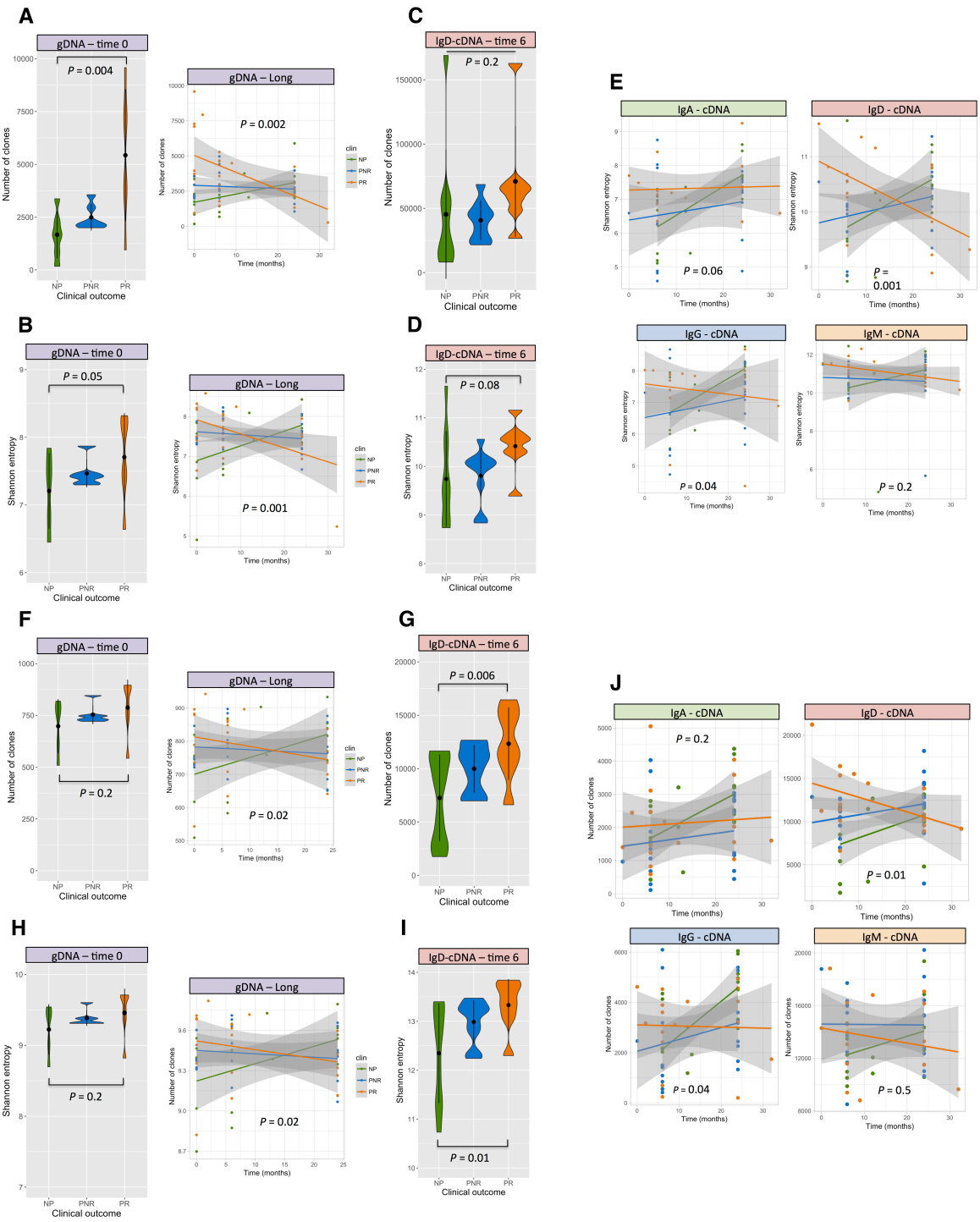
Supplementary Figure 1. Barplots showing the number of reads (A) and number of clones (B) from the gDNA B cell sequencing output and number of reads (C) and number of clones (D) from the cDNA B cell sequencing output. Each bar shows the number of reads or clones per individual after QC at time 0, 6 and 24 for gDNA and time 6 and 24 for cDNA (No samples had available RNA at time 0). The cDNA output is represented by isotypes: IgA (i), IgG (ii) IgD (iii) and IgM (iv). All samples show here passed the 100 clones minimum threshold.



Supplementary Figure 2. Violin plots showing the diversity measured by Shannon Entropy for gDNA (A) at time 0, 6 and 24 months, by number of clones for cDNA (B) and by Shannon Entropy for cDNA (C) at time 6 and 24 months defining the repertoire across the three clinical outcomes (NP, PNR and PR). The cDNA output is represented by isotypes: IgA (i), IgD (ii) IgG (iii) and IgM (iv). The p-values are obtained from the adjustment of a linear model considering the number of clones as a dependent variable and clinical outcome as an independent factor variable. Only p-values<0.05 are shown.

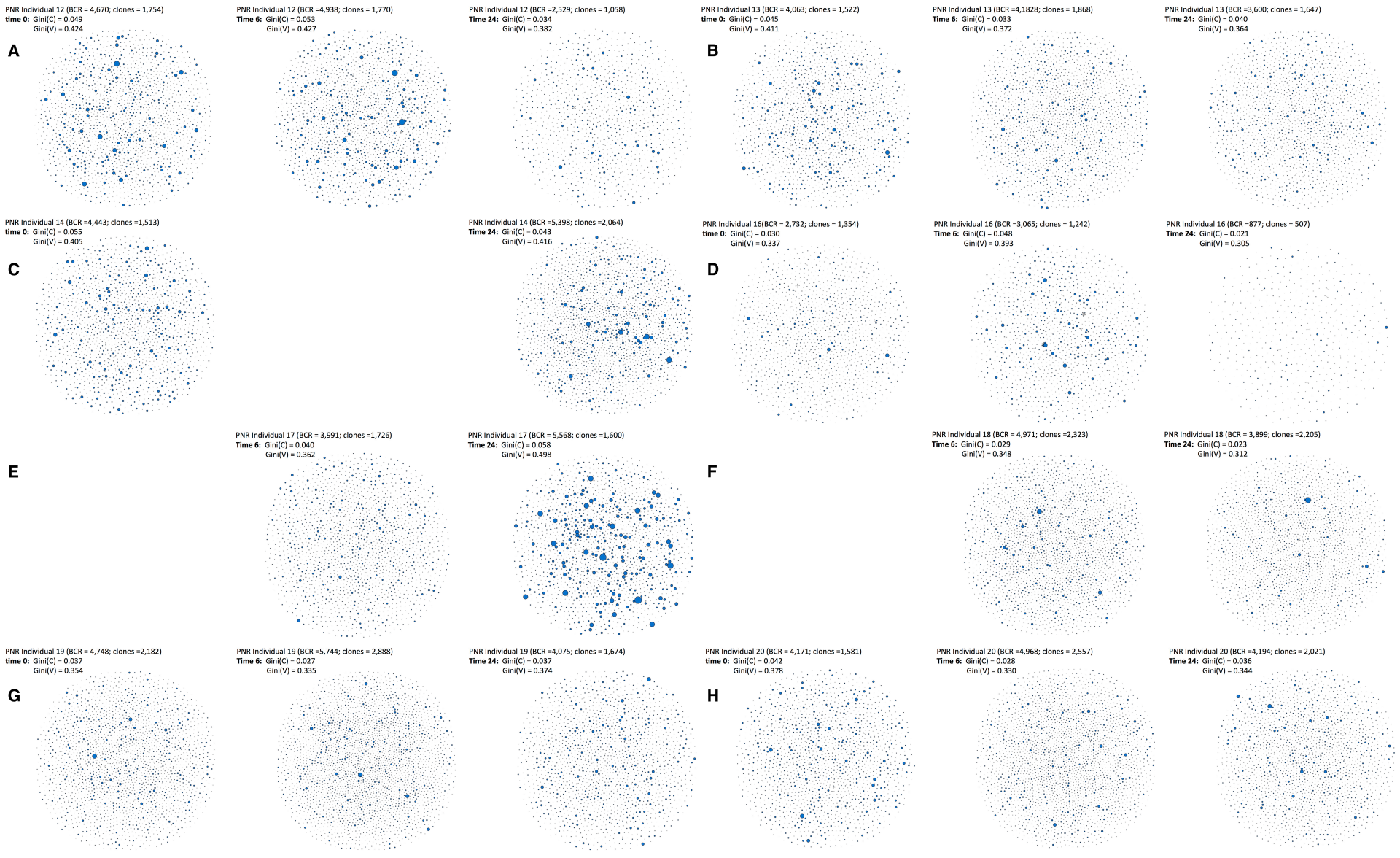


Supplementary Figure 3. Longitudinal data plotted with the fitted line for each clinical outcome. (A) Richness measured by number of clones represented across time points by the three clinical outcomes (NP, PNR, PR) in gDNA. (B) Richness measured by number of clones represented across time points by the three clinical outcomes (NP, PNR, PR) by isotypes in cDNA. The p-values correspond to the interaction term defined by time*clinical outcome adjusting a linear mixed effect model.



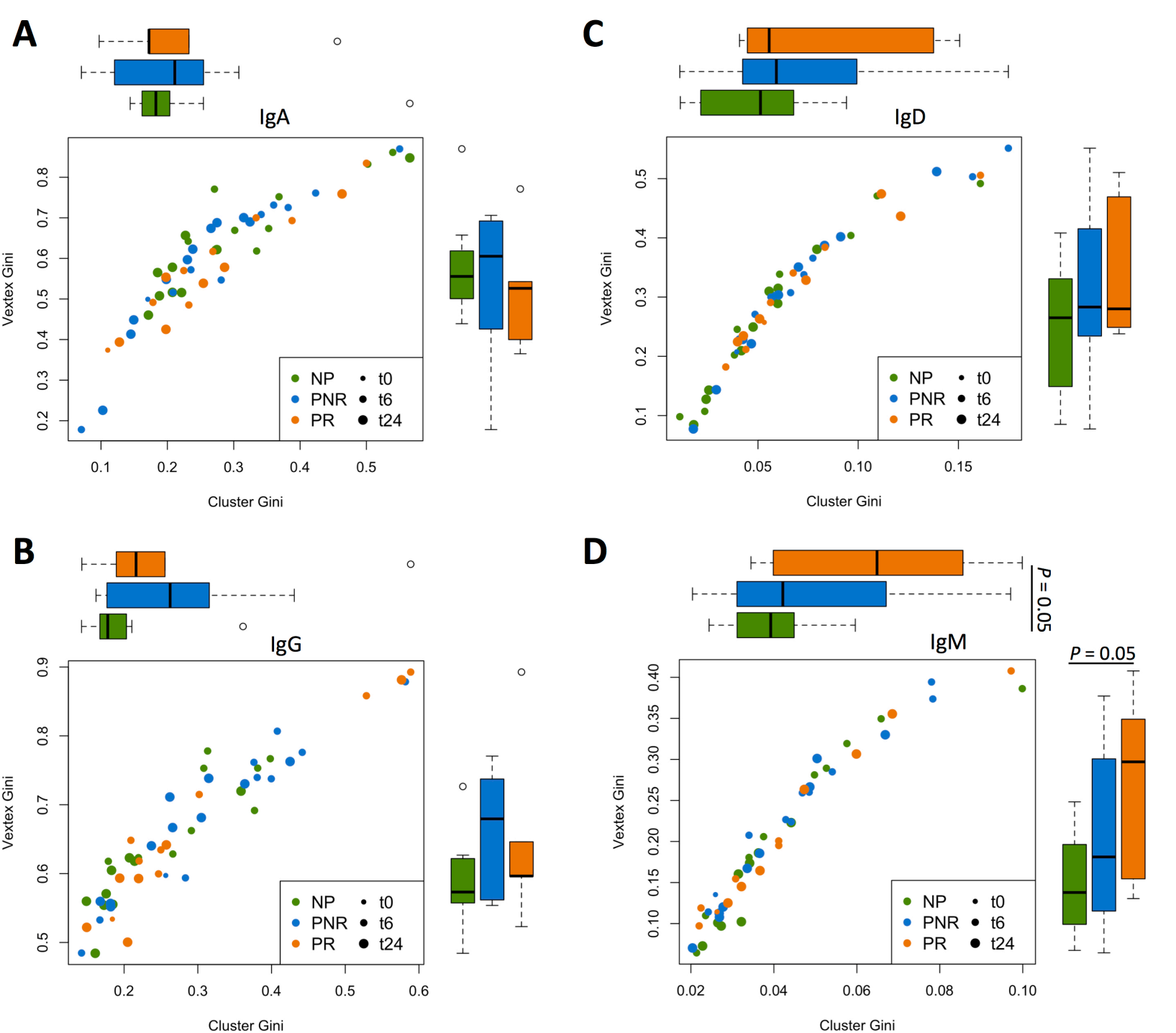
Supplementary Figure 4. Diversity analysis results using Recon (A-E) and using down-sampling (F-J). Violin plots showing the number of clones (richnes) at time 0 and the longitudinal data plotted with the fitted line for each clinical outcome measured by richness (number of clones) for gDNA using recon (A) and using down-sampling (F). Violin plots showing diversity measured by Shannon Entropy at time 0 for gDNA using recon (B) and using down-sampling (H). Violin plots showing the number of clones (richnes) at time 6 for IgD isotype for cDNA using recon (C) and using down-sampling (G). Violin plots showing diversity measured by Shannon Entropy at time 6 for IgD isotype for cDNA using recon (D) and using down-sampling (I). Longitudinal data plotted with the fitted line for each clinical outcome measured by Shannon entropy (diversity) for all isotypes in cDNA using recon (E) and using down-sampling (J). To perform the down-sampling strategy we had to filter out 9 samples with < 1000 clones for gDNA, therefore the sample size for the single time points analysis is reduced losing statistical power.

Supplementary Figure 5. Network representation of B cell repertoires from all individuals representing the NP group across time points. Each vertex represents a unique BCR being the vertex size defined by the number of identical BCRs considering the nucleotide sequences. An edge exists between vertices when they belong to the same clone as defined before, so clusters are groups of interconnected vertices forming a clone. Each sample shows the gini index obtained for the vertex size (*Gini(V)*) and cluster size (*Gini(C)*). BCR reflects the total B cell receptors for that specific sample and clones reflect the total number of unique clones. The data that is not represented are those that did not passed the quality control threshold (clones >100). Individual 8 (I) is the patient who developed EBV+ post-transplant lymphoproliferative disease (PTLD) at 2.2 yrs post Ktx, characterized by proliferation of Epstein-Barr virus (EBV)-infected B cells.

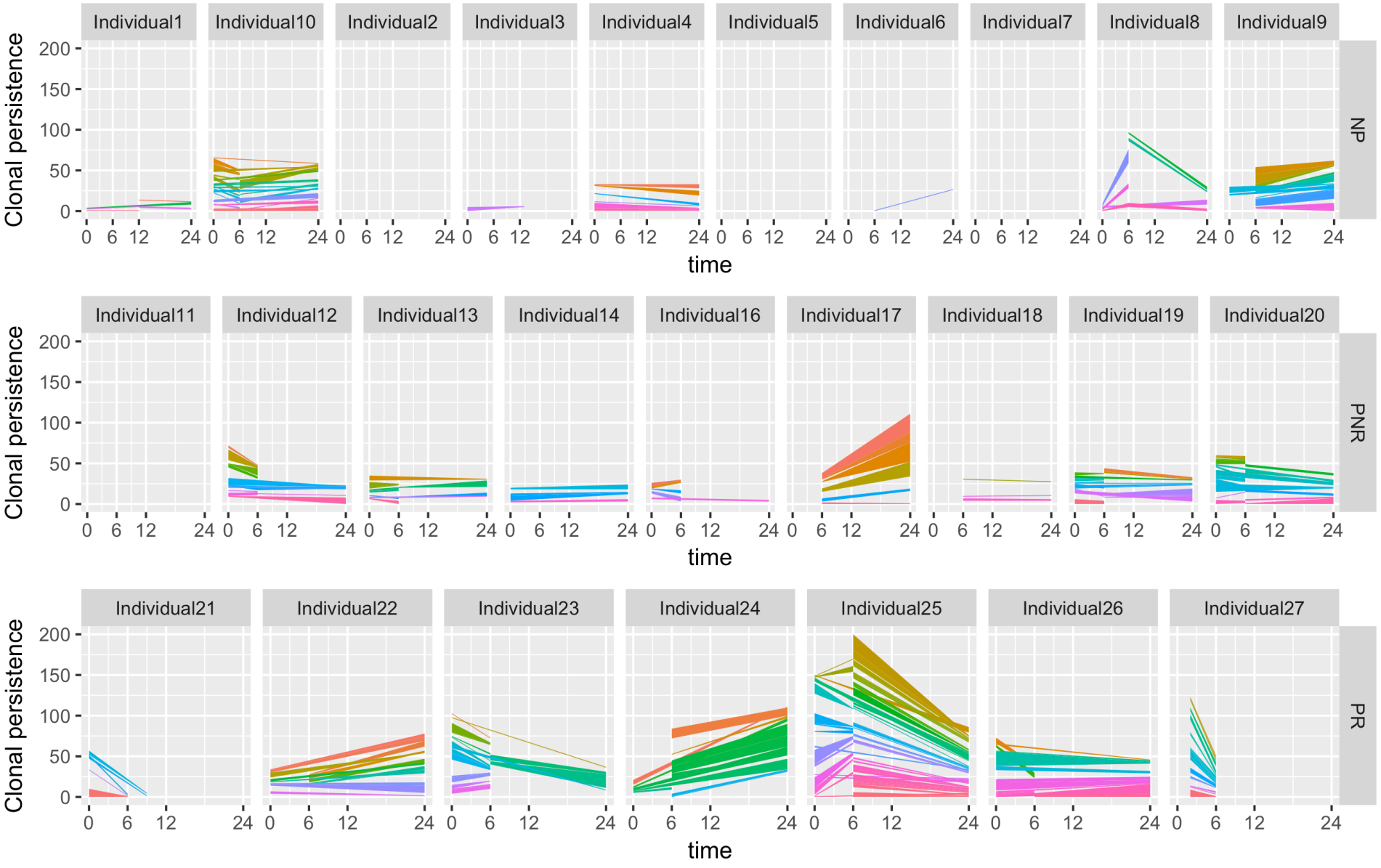


Supplementary Figure 6. Network representation of B cell repertoires from all individuals representing the PNR group across time points. Each vertex represents a unique BCR being the vertex size defined by the number of identical BCRs considering the nucleotide sequences. An edge exists between vertices when they belong to the same clone as defined before, so clusters are groups of interconnected vertices forming a clone. Each sample shows the gini index obtained for the vertex size (*Gini(V)*) and cluster size (*Gini(C)*). BCR reflects the total B cell receptors for that specific sample and clones reflect the total number of unique clones. The data that is not represented are those that did not passed the quality control threshold (clones >100).

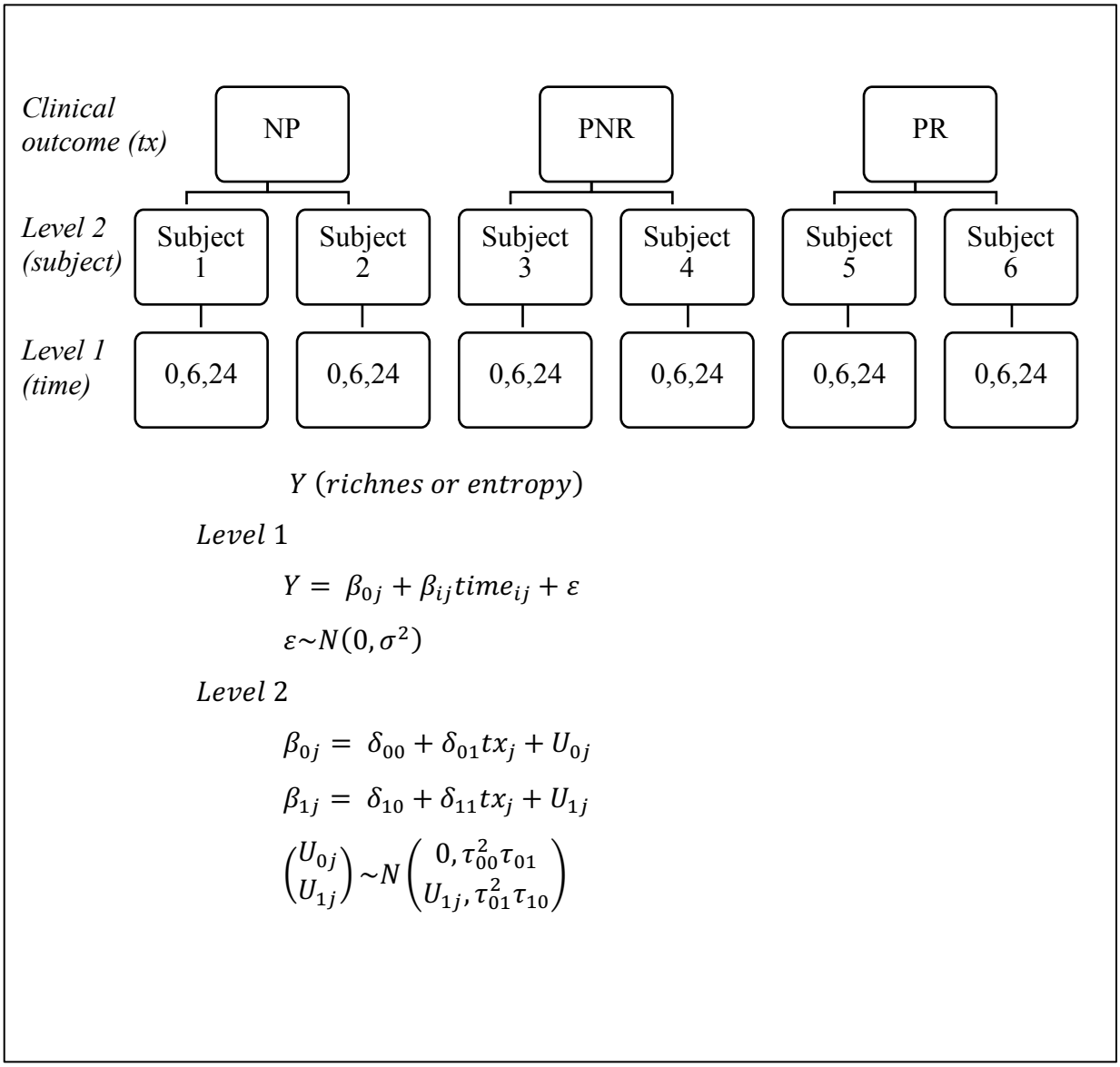
Supplementary Figure 7. Network representation of B cell repertoires from all individuals representing the PR group across time points. Each vertex represents a unique BCR being the vertex size defined by the number of identical BCRs considering the nucleotide sequences. An edge exists between vertices when they belong to the same clone as defined before, so clusters are groups of interconnected vertices forming a clone. Each sample shows the gini index obtained for the vertex size (*Gini(V)*) and cluster size (*Gini(C)*). BCR reflects the total B cell receptors for that specific sample and clones reflect the total number of unique clones. The data that is not represented are those that did not passed the quality control threshold (clones >100).



Supplementary Figure 8. Vertex Gini Index plotted against Cluster Gini Index for all the individuals in NP, PNR and PR differentiated by time points at the four isotypes from cDNA (A:IgA, B: IgD, C: IgD and D:IgM). Boxplots shows the *Gini(V)* and *Gini(C)* differences at time 24. The p-values are obtained from the adjustment of a linear model considering the *Gini(V)* and *Gini(C)* as a dependent variable and clinical outcome as an independent factor variable for each time point. Only p-values<0.05 are shown.



Supplementary Figure 9. Stacked area charts showing the clonal persistence for each individual and clinical outcome. Colors correspond to different clones and the size on the y-axis reflects the counts of each clone. More persistent clones in PR than in NP is observed (p-value = 0.01 calculated using a linear model with the counts of each clone as dependent variable and clinical outcome as the predictor variable). Higher number of persistent clones in PR compared with NP is observed (p-value = 0.09 calculated using linear model with the number of persistent clones as dependent variable and clinical outcome as the predictor variable).



$$Y\ (richnes\ or\ entropy)$$

$$Level\ 1$$

$$Y = \beta_{0j} + \beta_{ij} time_{ij} + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$Level\ 2$$

$$\beta_{0j} = \delta_{00} + \delta_{01} tx_j + U_{0j}$$

$$\beta_{1j} = \delta_{10} + \delta_{11} tx_j + U_{1j}$$

$$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N \begin{pmatrix} 0, \tau_{00}^2 \tau_{01} \\ U_{1j}, \tau_{01}^2 \tau_{10} \end{pmatrix}$$

Supplementary Figure 10. Conditional growth model for linear-mixed effect model

Supplementary Table 1. Clones associated with clinical outcome at each time point. The second column shows the number of individuals that had the clone (present) in each clinical outcome. The p-values are obtained from Fisher's exact test.

CDR3 amino acid sequence	Present / Absent			p-value	time	IGHV gene	IGHJ gene
	NP	PNR	PR				
ARDRWFDAFDI ARDRGGDAFDI ARDLGFDAFDI ARADTDDAFDI ARVGGWEAFDI	0	0	3				
ARQLAGYAFDI	9	10	3	0.009	0	*IGHV1-18*	*IGHJ3*
ARTSGWYFDY AREWDDYFDY ARERGYYFDY AREIGYYFDY AREDGDYFDY AREVFNYFDY ARDRGGYFDY ARDEGYYFDY ARDWRYYFDY ARVYGDYEGY ARVRGSYYDY	0	0	3				
ARYSGSYFDY	9	10	3	0.009	0	*IGHV3-21*	*IGHJ4*
ANSYGDNYYYGMDV AREGIGNYYYGMDV ARESLHYYYYGMDV ARDMGDYYYYGMDV ARDRLYYYYYGMDV ARADDFYYYYGMDV ARGDDYYYYYGMDV ARGGRLGYYYDMDV ARVGNYYYYYGMDV ARVEFYYYYYGMDV ARVGNYYYYCGMDV ASVGNYYYYYGMDV ARKGSYYYYYGMDV ARRGIGYYYYGMDV	0	1	3				
AGGDDYYYYYGMDV	9	9	3	0.04	0	*IGHV3-48*	*IGHJ6*
AREQYYYYGMDV ARDRDYYYGMDV ARDDLYYYGMDI ARDDLYYYGMDV ARDGSYYYGMDV ARAPGYYYGMDV ARGLDYYYGMDV	0	1	3				
ARVDYYYYGMDV	9	9	3	0.04	0	*IGHV3-53*	*IGHJ6*
TRDAY AREYY	0	0	3				
ARDDY	9	10	3	0.009	0	*IGHV3-7*	*IGHJ4*

ARDVH ARGAY ARGGY ARGVY ARLDY						
AKDRYYDSSGYLDY AKARYYDSSGYLDY AKGNYYDSSGYYDY AKGRDYDSSGYFDY AKGYYYDSSGYYDY	1 8	0 10	3 3	0.02	0	*IGHV3-9*	*IGHJ4*
AKDRYYYYYYGMDV AKDSGSSYYYGMDV AKDIAGHYYYGMDV AKDMKVVYYYGMDV AKDIEGYYYYGMDV AKDIGSYYGNGMDV AKDPDSNYYYGMDV AKDAMAYYYYGMDV AKDFWAYYYYGMDV AKVGAAYYYYGMDV	0 9	1 9	3 3	0.04	0	*IGHV3-9*	*IGHJ6*
ATGTDYYYYYGMDV ARSAASYYYYYMDV ARQPSYYYYYGMDV ARPANYYYYYYMDV ARLANYYYYYYMDV ARDLPNYYYYGMDV ARADYDYYYYGMDV ARGQGYYYYYGMDV ARGQDLYYYYGMDV ARGRGVHYYYGMDV ARGLPNYYYYYMDV ARGGHYYYYYGMDV ARGTGDYYYYGMDV ARVRIHYYYYGMDV ARASRRYYYYGMDV	0 9	1 10	3 3	0.04	0	*IGHV4-34*	*IGHJ6*
AREEYYDSSGYSDY ARDNYYDSSGPNDY ARDYYYDSSGSFDY ARAYYYDSSGYYVY ARGDYYDSSGSFDY	1 8	0 10	3 4	0.04	6	*IGHV1-2*	*IGHJ-4*
TTDQYYYDSSGYYRRDY TTDPYYYDSSGYYFFDY TTDAYYYDSSGYFYFDY TTDPYYYDSSGYSILDY TTDVHYYDSSGYKYFDY TTDLDYYDSSGYYPFDY TTVAYYYDSSGYYYGHY	0 9	0 10	3 4	0.01	6	*IGHV3-15*	*IGHJ4*
ARTSGWYFDY AREWDDYFDY ARERGYYFDY AREIGYYFDY AREDGDYFDY AREVFNYFDY	1 8	0 10	3 4	0.04	6	*IGHV3-21*	*IGHJ4*

ARDRGGYFDY ARDEGYYFDY ARDWRYYFDY ARVYGDYEGY ARVRGSYYDY ARYSGSYFDY							
AKDADYYDSSGYFDY AKDAYYYDSSGYLDY AKAPGYYDSSGYFDY AKAPSYYDSSGYVDY ARERNYYDSSGYYDY AREEYYYDSSGYYDY ARDTYYYDSSGAFDY ARDQYYYDSSGYFDY ARDPFYYDSSGYVAY ARDRHYYDNTGYVDY ARDLGYYDSSGSLDY ARDLGYYDSSGYLAY ARDGGYYDSSGHLDY ARDSLYYDSSGYPDY ARDLRYYDSSGYLDY ARAVDYYDSSGYLDY ARGLDYYDSSGTLDY ARVTGYYDSSGYFDY	1 8	0 10	3 4	0.04	6	*IGHV3-30*	*IGHJ4_45*
ARDSYYDSSGYYPHYYGMDV AREYYYDSSGYYTYYYGMDV	0 8	0 10	2 3	0.04	24	*IGHV1-46*	*IGHJ6*
ASESYSSSFGY	0 8	0 10	2 3	0.04	24	*IGHV1-8*	*IGHJ4*
ARGWYCSSTSCDYDY	0 8	0 10	2 3	0.04	24	*IGHV1-8*	*IGHJ4*
ARGLSSGWYDPYPYYYYYYGMDV	0 8	0 10	2 3	0.04	24	*IGHV1-8*	*IGHJ6*
ARTNLGYCSGGSCYYWFDP	0 8	0 10	2 3	0.04	24	*IGHV2-26*	*IGHJ5*
ARDASLKWGGMD	0 8	0 10	2 3	0.04	24	*IGHV3-11*	*IGHJ6*
TTEIFDY TTEIFGY TTEWLDY TTDRTDY TTDAPDY TTAYFDY TTGPLDY	0 8	0 10	2 3	0.04	24	*IGHV3-15*	*IGHJ4*
AKERYYYDSSGYYGY AKEADYYDSSGLFDY AKDPGYYDSSGYWDY	0 8	1 9	3 2	0.03	24	*IGHV3-23*	*IGHJ4*

AKDPGYYDSSGYYHY AKDPGYYDSSGYCHY AKDPFYYDSSGYWDY AKDRQYYDSSGYFDY AKDLYYYDSSGYYSY AKGQDYYDSSGYYDY AKLTYYYDSSGYSDY							
AKEYYDSSGYYYPFDY	0	0	2				
AKDYYDSSGYYSLFDY ATYYYDSSGYYYYFDY	8	10	3	0.04	24	*IGHV3-23*	*IGHJ4*
AKEGIYGDYNYYYGMDV	0	0	2				
AKGRVYGDYVYYYGMDV ARASDYGDYYYYYGMDV	8	10	3	0.04	24	*IGHV3-23*	*IGHJ6*
	0	0	2				
AKDTAHHGSGSSWMAPIDY	8	10	3	0.04	24	*IGHV3-30*	*IGHJ6*
TKDSSSSNYYFGLDV	0	0	2				
AKDRGSSFYYYGMDV AKYSSSSNYYYGMDV	8	10	3	0.04	24	*IGHV3-30*	*IGHJ6*
ARTGIDY ARDSFDY ARDGHDY	0	0	2				
ARDGGDY ASDYFDY	8	10	3	0.04	24	*IGHV3-33*	*IGHJ4*
TRDRRGGLDY ARTNGSYYDY ARERVGAIDY AREGEGYFDY AREGSGLFDY AREGGSYFDY AREVGGSIDY ARDRLDYFDY ARDRGGYFDY ARDSGGGFDY ARDIMGYFDY ARDHSGSFDY ARDRGIHFDY ARDRGDSFDY ARDRNGEFDY ARDRGYYFDY ARDREGYFDY ARDVWYYFDY ARDYRYYFDY ARDWAYYFDY ARGSGSYLDY							
ASDYGVNFDY ASYYGVNFDY	8	8	2	0.04	24	*IGHV3-33*	*IGHJ4*
	0	0	2				
TRVSGPYGSGSQDY	8	10	3	0.04	24	*IGHV3-49*	*IGHJ4*
ARDRYCSGGSCSHFDY	0	0	2				
ARVGYCSGGSCNDLDY				0.04	24	*IGHV3-7*	*IGHJ4*

ARVGYCSGGSCNDFDY ASEGYCSGGSCYWFDY	8	10	3				
ARVDDSSGYYLYYFDY ARVDDSSGYYLCYFDY ARVDDSSGYYQYYFDY	0 8	0 10	2 3	0.04	24	*IGHV3-7*	*IGHJ4*
ARDRSYGMDV ARDLSYGMHV ARGNYYGMDV ARGLAYGMDV ARGIYYAMDV ARGHYYGMDV ARVGYYGMDV	0 8	0 10	2 3	0.04	24	*IGHV3-74*	*IGHJ6*
ARERTVTTPAYYYGTDV ARERTVTTPAYYYGMDV ARDSTVTTGYYYYGMDV	0 8	0 10	2 3	0.04	24	*IGHV4-31*	*IGHJ6*
ARVNVVVPSGSYSG	0 8	0 10	2 3	0.04	24	*IGHV4-39*	*IGHJ5*
ARDRRDYGDYKYYFDY	0 8	0 10	2 3	0.04	24	*IGHV4-4*	*IGHJ4*

Supplementary Table 2. List of persistence clones shared across individuals

CDR3 amino acid sequence	Clinical Outcome (shared individuals)		
	NP	PNR	PR
ARDRRWSFDY ARDRRWPFDY ARDLSWAFDY VRDYLWGFDY VRDYYYRFDY ARDSVYAFDY VRDRDWGFDF ARDLLGAFDY ARDHNWAFDY ARDVNWAFDY ARAYGGNYDY ARDWNWAFDY VRDYYYRFDY ARDHDWAFDC VRDFDWNFDY ARDRGHYFDY	Individual 9	Individual 12, 13, 14,16, 19	
ARVQGHYYYYGMDV ARTLNAYYYYGMDV ARDRRDYYYYGMDV AKEIRGYYYYGMDV AGNYGDYDYYGMDV ARDLSYYYYYGMDV AKDPAWDYYYGMDV AKDIGVYYYYGMDV AKAMADYYYYGMDV AKDKGSYYYYGMDV ARETGDYYYYGMDV AKDLTDYYYYGMDV ARARGDYYYYGMDV ARDSHYEYYYGMDV AKVRATLYYYGMDV	Individual 10	Individual 12, 14, 19, 20	Individual 27
ARDYYDYGMDV ARDSGDYGMDV AREDGEYGMDV ARVSGSYGMDV	Individual 9	Individual 13	
ARGSRKYYYYYGMDV ARGRLRVYYYYGMDV ARGEGDYNYYYGMDV ARGVILYYYYYGMDV ARGNGRYYYYYGMDV	Individual 9	Individual 18	
ARDVWYYFDY ARERVGAIDY ARTNGSYYDY		Individual 12, 19	Individual 24

ARDSGGGFDY ARDRLDYFDY ARDRGGYFDY ARDRGDSFDY			
ARAINYYYGMDV ARDYYYYYGMDV ASNYYYYYGMDV AREEGYYYGMDV ARDQDYYYGMDV ARDRSYYYGMDV		Individual 19, 20	Individual 26
AKAPSYYDSSGYVDY ARGLDYYDSSGTLDY ARDSLYYDSSGYPDY ARDLRYYDSSGYLDY ARDLGYYDSSGSLDY AREEYYYDSSGYYDY AKDAYYYDSSGYLDY ARERNYYDSSGYYDY ARAVDYYDSSGYLDY ARDLGYYDSSGYLAY	Individual 9	Individual 20	Individual 25
ARDQRFYFDY AKDSGSYYDY AKDNGDYFDY AKDRDTCFDY AKDYGFYFDY AKDRGFGFDY AKGYSGSLDY AKDRVHYFDY AKDRGPYFDY AKYRGPYFDY		Individual 20	Individual 22, 26
ARGRRDYYYGMDV ARGYYYYYYGMDV ARGRYYYYYGMDV ARGYHYYYYGMDV ARVYYYYYYGMDV ARRSDYYYYGMDV ARGAFYYYYGMDV		Individual 14	Individual 23, 25
AKLSGYYYYYGMDV AKASGYYYYYGMDV AKLLDYYYYYGMDV AKVRGPYYYYGMDV ARDPEGYYYYGMDV	Individual 10		Individual 23
ARLYDYGDYYYYYGMDV ARDRSGYEYYYYYGMDV ARASDYGDYYYYYGMDV ARDRRDGDYYYYYGMDV AREGGGYDFYYYYGMDV ARDRGYDGYYYYYGMDV ARYGYGDYGYYYYGMDV ARDRGDYVLGYYYGMDV	Individual 9		Individual 23, 26, 27

AKDIAYYFDY AKSRDYYFDY AKDEIYYFDY AKDLYGYFDY AKDRWNYVDY			Individual 22, 24
ARYNYDSSGYCDY AREDYDSSGYTDY AKNYYDSSGYLRY ARDYYDSSGYFDY	Individual 10		Individual 25
ARDTPNYYYYGMDV ARDRPRYYYYGMDV ARAAPYYYYYGMDV AREIHYYYYYGMDV		Individual 14	Individual 25
ARVFEYYFDY ARGSGWLVDY AGGSGWIFTN ARGFGDYFDY AAGSGWLIDN ARDSSGSSDY			Individual 22, 25
ARHRQYYYYYGMDV ARRLGGYYYYGMDV ARGDRYYYYYGMDV ARQPSYYYYYGMDV		Individual 12	Individual 25
AKEADYYDSSGLFDY AKDRQYYDSSGYFDY AKGQDYYDSSGYYDY AKDPGYYDSSGYYHY AKDPGYYDSSGYWDY AKDPGYYDSSGYCHY		Individual 23, 25	
ARQAGSSFDY ARRVRGNFDY ARGGTYYFDY ARDWGGHFDY		Individual 20	Individual 26
ARQPSYYYYYGMDV ARGQGYYYYYGMDV ARGGHYYYYYGMDV ARSAASYYYYYMDV ARGRGVHYYYGMDV ARPANYYYYYYMDV ARLANYYYYYYMDV ARGTGDYYYYGMDV			Individual 25, 26
ARFYYYYYGMDV ARISTYYYGMDV ASGRNYYYGMDV ARWSYYYYYMDV		Individual 20	Individual 26
AREARLYYYYGMDV ARDLDYYYYYGMDV ARSAGHYYYYGMDV ARVLRYYYYYGMDV	Individual 10		Individual 27

ARDHYYDSSGYLDY ARCYYYDSSGPIDY AKDSYYDSSGPFDY AKAYYYDSSGYFNY ASSDYYDSSGYLDY AKDYYYDSSGFLGY ARDLYYDSSGYFDY		Individual 20	Individual 27
ARDGYKTYYYYGMDV ARDVGAAYYYYGMDV ARAHGDLYYYYGMDV ARDAGSYYYYYGMDV ARAGGLYYYYYGMDV ARSGGPYYYYYGMDV AKDRELDYYYYGMDV		Individual 20	Individual 27