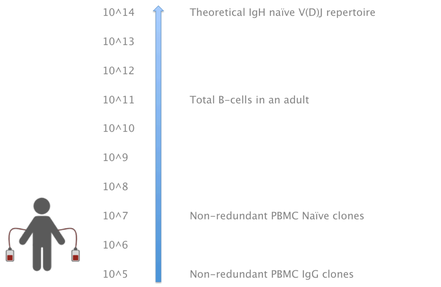
Call of the wild: a new generation of antibody discovery from human populations

Jacob Glanville, David Louis, David Maurer, Usha Mulvaney, Casey Keyes, Gayathri Shankar, Amy Geurts, Maureen Kagimbi, Distributed Bio, jake@distributedbio.com

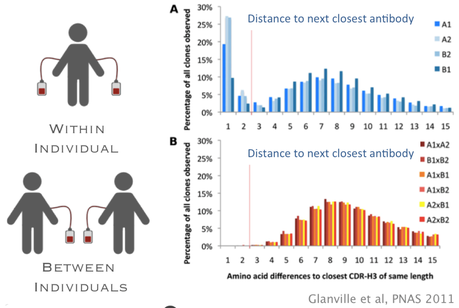
How diverse is the human repertoire?

High throughput sequencing has shown the human repertoire, as sampled by peripheral blood draw, to be biased and limited by clonal expansion and homeostatic proliferation. It also has nearly no overlap in clones between different individuals. This argues for natural libraries to be constructed from sorted cells, with diversity for each CDR sourced from naive for CDR-H3, but memory for all others.

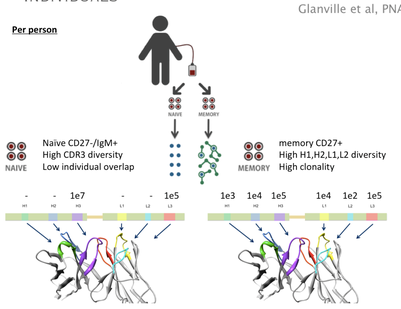
Individual adults only sample a small part of the antibody repertoire



Humans produce dispersed receptor repertoires

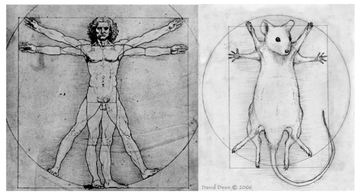


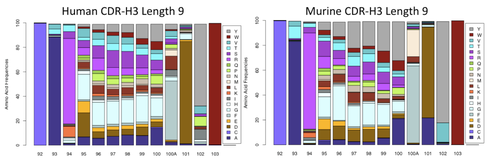
Per person



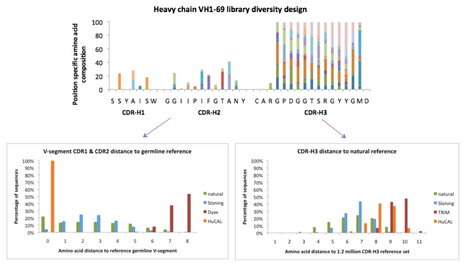
What's so great about natural repertoires?

Natural repertoires have been sculpted by evolutionary forces to be populated by diversity that can fold up and recognize antigen.



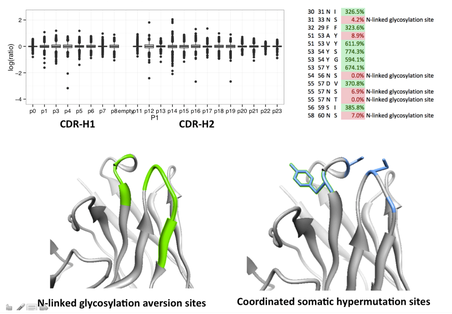


Our previous research has shown that mimicking nature's fitness landscapes can dramatically improve synthetic libraries



Zhoi & Glanville et al, "Synthetic antibodies designed on natural sequence landscapes", Journal of Molecular Biology, 2011

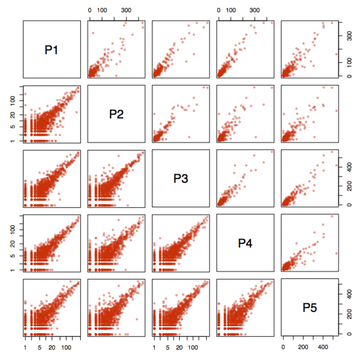
New analysis reveals complex rules of fitness that are encoded in the Natural repertoire but not available elsewhere.



How to effectively cover light chain diversity

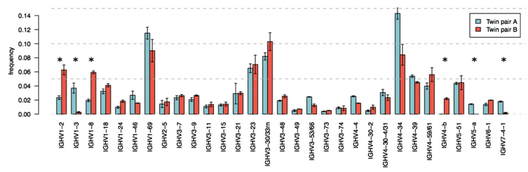
In humans, the light chain really isn't that diverse. In an example set of five unrelated individuals P1-P5, we see high correlations in usage of the same set of roughly 5000 common light chains, and a total repertoire of less than 500k SHM variants.

In "How to barcode an army of cancer survivors," this feature is exploited in calculating H/L coverage of immune response events.

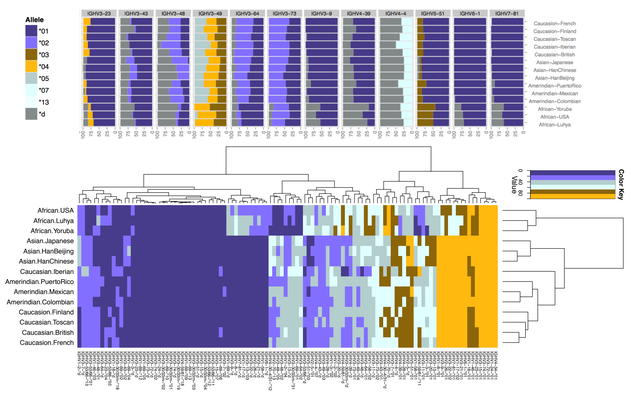


The population genetics of good scaffold selection

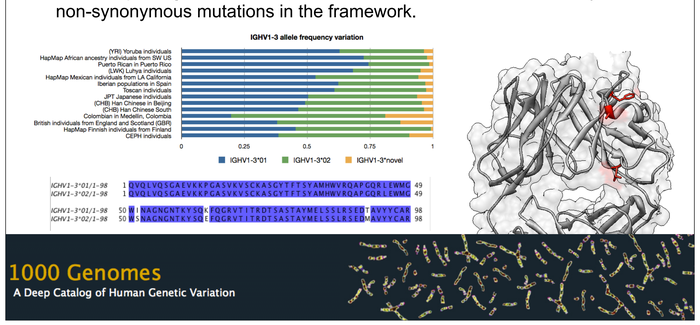
Some V-genes are only found in a subset of individuals. These will look Non-human in individuals without them.



Most V-genes have some allele variants, and these variants often have different frequencies in different racial groups. To avoid racist medicine, it is better to select V-gene scaffolds with a single dominant allele.



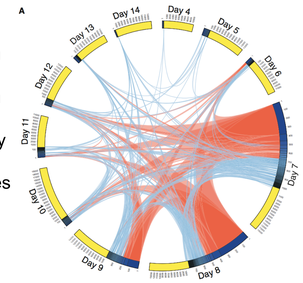
An example V-gene IGHV1-3, with two dominant alleles that differ by three non-synonymous mutations in the framework.

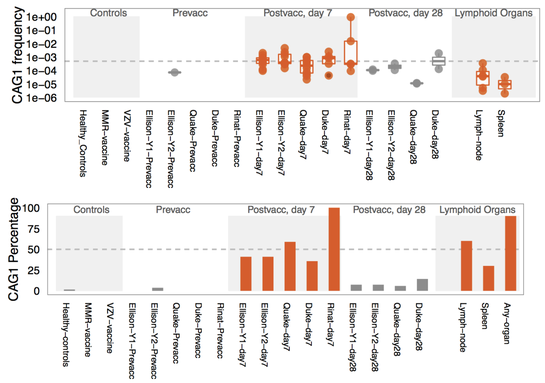


1000 Genomes
A Deep Catalog of Human Genetic Variation

How to catch the repertoire at the right time

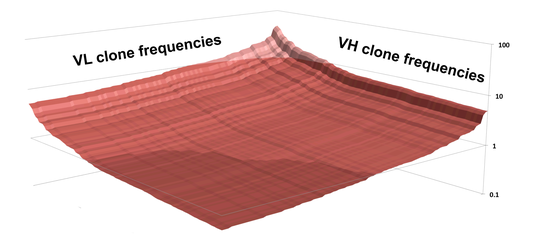
Relevant immune memory is only transiently available in the periphery. A) vaccine at day 0 results in an expansion at day 7 that peaks and becomes undetectable by day 14. B & C) Expansion of vaccine-specific shared clones are found in secondary lymphoid organs outside of the acute activation period.



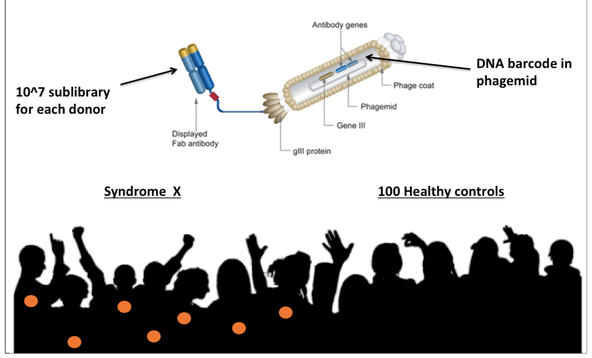


How to barcode an army of cancer survivors

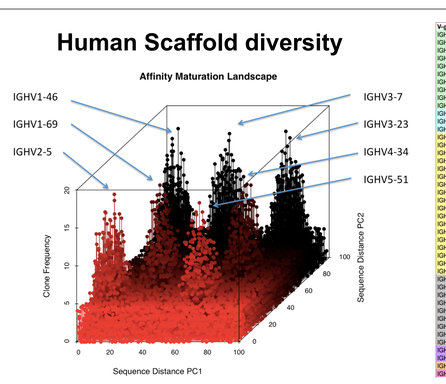
A peripheral sample of human blood will only recover a few hundred thousand memory clones and plasmablasts, with a set of less than 10,000 that dominate the sample. The top 100k expanded clones can be captured in a 10^7 sub-library with excellent chain pairing.



The immune memory of 10,000 people can be loaded into a barcoded library



Human Scaffold diversity



V-gene	FW1	CDR-H1	FW2	CDR-H2	FW3		H1 length	H2 length	H1 canon	H2 canonical	migA	migG	migM	migM
IGHV1-69*01	QVQLVQSGAEVKKPGSSVKVSCKASGGTFS	STAIS	WVRQAPGQGLEWMG	GIIPIFGTANYAQKFQG	AYIIIADESTDTAYMELSSLRSEDTAVYYC	AR	5	17	1	2	4.0%	8.0%	2.4%	11.6%
IGHV1-18*01	QVQLVQSGAEVKKPGASVKVSCKASGYTFT	SYGIS	WVRQAPGQGLEWMG	WISAYNGNTNYAQKLQG	RVTMTTDTSTSTAYMELRSLRSDDTAVYYC	AR	5	17	1	1	5.3%	4.9%	3.5%	5.3%
IGHV1-2*01	QVQLVQSGAEVKKPGASVKVSCKASGYTFT	GTYMH	WVRQAPGQGLEWMG	RINPNSGGTNYAQKFQG	RVTSTRDTSISTAYMELSRLRSDDTVVYYC	AR	5	17	1	3	3.6%	3.3%	3.2%	3.2%
IGHV1-2*01	QVQLVQSGAEVKKPGASVKVSCKASGYTFT	ETAMH	WVRQAPGQGLEWMG	WINAGNGNTKYSQKFQG	RVTITRDTSASTAYMELSSLRSEDTAVYYC	AR	5	17	1	3	2.4%	2.5%	2.4%	2.7%
IGHV1-46*01	QVQLVQSGAEVKKPGASVKVSCKASGYTFT	STYMH	WVRQAPGQGLEWMG	IINPSGGSTSYAQKFQG	RVTMTRDTSTSTVYMELSSLRSEDTAVYYC	AR	5	17	1	3	2.4%	2.2%	1.6%	1.9%
IGHV1-8*01	QVQLVQSGAEVKKPGASVKVSCKASGYTFT	STGIN	WVRQATGQGLEWMG	WMNPNSGNTGYAQKFQG	RVTMTRNTSISTAYMELSSLRSEDTAVYYC	AR	5	17	1	3	1.5%	1.9%	1.6%	3.2%
IGHV1-24*01	QVQLVQSGAEVKKPGASVKVSCKVSGYTLT	ELSMH	WVRQAPGKGLEWMG	GFDPEDGETIYAQKFQG	RVTMTEDTSTDTAYMELSSLRSEDTAVYYC	AT	5	17	1	U	0.9%	1.7%	0.4%	1.6%
IGHV1-f*01	EVQLVQSGAEVKKPGATVKISCKVSGYTFT	DYTMH	WVQQAPGKGLEWMG	LVDPEDGETIYAEKFQG	RVTITADTSTDTAYMELSSLRSEDTAVYYC	AT	5	17	1	2	0.2%	0.3%	0.0%	0.2%
IGHV1-58*01	QMQLVQSGPEVKKPGTSVKVSCKASGFVFT	SSAVQ	WVRQARGQRLEWIG	WIVVGSGNTNYAQKFQE	RVTITRDMSTSTAYMELSSLRSEDTAVYYC	AA	5	17	1	3	0.2%	0.2%	0.2%	0.3%
IGHV1-46*01	QMQLVQSGAEVKKPGASVKVSCKASGYTFT	YRTLH	WVRQAPGQALEWMG	WITPFNGNTNYAQKFQD	RVTITRDRSMSTAYMELSSLRSEDTAVYYC	AR	5	17	1	3	0.1%	0.0%	0.0%	0.0%
IGHV2-5*01	QVTLKESGPALVKPTQTLTLTCTFSGFSLS	TSGVGVG	WIRQPPGKALEWLA	LIYWDGDKRYSPSLKS	RLTITKDTSKNQVVLTMTNMDPVDTATYYC	AH	7	16	3	1	1.9%	2.3%	1.9%	1.2%
IGHV2-70*01	QVTLRESGPALVKPTQTLTLTCTFSGFSLA	TSGMCVS	WIRQPPGKALEWLA	LIDWDDDKYYSTSLKT	RLTISKDTSKNQVVLTMTNMDPVDTATYC	AR	7	16	3	1	0.3%	0.3%	0.3%	0.2%
IGHV2-26*01	QVTLKESGPVLVKPTETLTLTCTVSGFSLS	NARMGVS	WIRQPPGKALEWLA	HIFSNDEKSYSTSLKS	RLTISKDTSKSQVVLTMTNMDPVDTATYC	AR	7	16	3	1	0.2%	0.2%	0.1%	0.3%
IGHV3-30-3*01	QVQLVESGGGVVQPGRSLRLSCAASGFTFS	SYGMH	VIWYDGSNKYYADSVKG	RFTISRDNSKNTLYLQMNSLRAEDTAVYYC	AR	5	17	1	1	4	10.3%	10.8%	9.5%	9.6%
IGHV3-23*01	EVQLLESGGGLVQPGGSLRLSCAASGFTFS	STAMS	WVRQAPGKGLEWVS	AISGSGGSTYYADSVKG	RFTISRDNSKNTLYLQMNSLRAEDTAVYYC	AK	5	17	1	3	9.4%	8.1%	11.1%	6.6%
IGHV3-7*01	QVQLVESGGGLVQPGGSLRLSCAASGFTFS	SYWMS	WVRQAPGKGLEWVA	NIKQDGSEKYYVDSVKG	RFTISRDNAKNSLYLQMNSLRAEDTAVYYC	AR	5	17	1	3	4.8%	4.8%	6.2%	2.6%
IGHV3-53/66	QVQLVESGGGLIQPGGSLRLSCAASGFTVS	SNYMS	VRQAPGKGLEWVS	VIYSGGSTYYADSVKG	RFTISRDNSKNTLYLQMNSLRAEDTAVYYC	AR	5	16	1	1	3.0%	2.9%	3.5%	2.3%
IGHV3-48*01	QVQLVESGGGLVQPGGSLRLSCAASGFTFS	SYEMN	WVRQAPGKGLEWVS	YISSSSSTIYYADSVKG	RFTISRDNAKNSLYLQMNSLRAEDTAVYYC	AR	5	17	1	3	3.3%	2.9%	4.2%	2.8%
IGHV3-74*01	QVQLVESGGGLVQPGGSLRLSCAASGFTFS	SYMSH	WVRQAPGKGLVWVS	RINSDGSSTSYADSVKG	RFTISRDNAKNTLYLQMNSLRAEDTAVYYC	AR	5	17	1	3	3.9%	2.4%	4.0%	1.2%
IGHV3-21*01	QVQLVESGGGLVKPGGSLRLSCAASGFTFS	SYSMN	VRQAPGKGLEWVS	SISSSSSYIYYADSVKG	RFTISRDNAKNSLYLQMNSLRAEDTAVYYC	AR	5	17	1	3	2.5%	2.2%	2.7%	3.3%
IGHV3-9*01	QVQLVESGGGLVQPGRSLRLSCAASGFTFD	DYAMS	WVRQAPGKGLEWVS	GISWNSGSIGYADSVKG	RFTISRDNAKNSLYLQMNSLRTEDTALYYC	AK	5	17	1	3	2.7%	2.0%	2.0%	1.9%
IGHV3-15*01	QVQLVESGGGLVKPGGSLRLSCAASGFTFS	NAWMS	WVRQAPGKGLEWVG	RIKSKTDGGTTDYAAPVKG	RFTISRDGSKNTLYLQMNSLKTEDTAVYYC	TT	5	19	1	U	2.4%	1.9%	2.1%	1.4%
IGHV3-11*01	QVQLVESGGGLVKPGGSLRLSCAASGFTFS	SYMSW	VRQAPGKGLEWVS	YISSSGSTIYYADSVKG	RFTISRDNAKNSLYLQMNSLRAEDTAVYYC	AR	5	17	1	3	1.4%	1.3%	1.4%	1.2%
IGHV3-49*01	QVQLVESGGGLVQPGRSLRLSCTASGFTFG	DYAMS	WFRQAPGKGLEWVG	FIRSKAYGGTTEYTASVKG	RFTISRDGSKSIAYLQMNSLKTEDTAVYYC	TR	5	19	1	U	1.0%	0.8%	0.9%	0.8%
IGHV3-73*01	QVQLVESGGGLVQPGGSLKLSCAASGFTFR	SSAMH	RIRSKANSYATAAAYAVSVKG	RFTISRDDSKNTAYLQMNSLKTEDTAVYYC	TR	5	19	1	4	0.6%	0.7%	0.9%	0.5%
IGHV3-64*01	QVQLVESGGGLVQPGGSLRLSCAASGFTFS	SYAMH	WVRQAPGKGLEYVS	AISSNGGSTYYANSVKG	RFTISRDNSKNTLYLQMGSLRAEDMAVYYC	AR	5	17	1	3	0.6%	0.6%	0.6%	0.3%
IGHV3-72*01	QVQLVESGGGLVQPGGSLRLSCAASGFTFS	DYHMD	WVRQAPGKGLEWVG	RTRNKANSYTTEYAASVKG	RFTISRDDSKNSLYLQMNSLKTEDTAVYYC	AR	5	19	1	4	0.7%	0.5%	0.8%	0.1%
IGHV3-13*01	QVQLVESGGGLVQPGGSLRLSCAASGFTFS	STMIH	WVRQATGKGLEWVS	AIGTAGDTYYPGSVKG	RFTISRENAKNSLYLQMNSLRAGDTAVYYC	AR	5	16	1	1	0.4%	0.5%	0.5%	0.3%
IGHV3-43*01	QVQLVESGGVVQPGGSLRLSCAASGFTFD	DYIMH	WVRQAPGKGLEWVS	LISWDGGSTYYADSVKG	RFTISRDNSKNSLYLQMNSLRTEDTALYYC	AK	5	17	1	3	0.5%	0.4%	0.2%	0.3%
IGHV3-20*01	QVQLVESGGGVVRPGGSLRLSCAASGFTVS	DYDMS	WVRQAPGKGLEWVS	GINWNGGSTGYADSVKG	RFTISRDNAKNSLYLQMNSLRAEDTALHYC	AR	5	17	1	3	0.3%	0.2%	0.2%	0.2%
IGHV3-d*01	QVQLVESGGVLVQPGRSLRLSCTASGFTFS	SNEMS	WVRQAPGKGLEWVS	SI-SGGSTYYADSVKG	RFTISRDNSKNTLHLQMNSLRAEDTAVYYC	KK	5	16	1	6	0.0%	0.0%	0.0%	0.0%
IGHV4-61*01	QVQLQESGPGLVKPSETLSLTCAVSGGSVS	SGSYYWS	WIRQPPGKGLEWIG	YIYYSGSTNYNPSLKS	RVTISVDTSKNQFSLKLSSVTAADTAVYYC	AR	5	16	3	1	7.1%	7.1%	7.1%	6.8%
IGHV4-59*01	QVQLQESGPGLVKPSETLSLTCTVSGGSIS	SYTWS	WIRQPPGKGLEWIG	YIYYSGSTNYNPSLKS	RVTISVDTSKNQFSLKLSSVTAADTAVYYC	AR	5	16	1	1	6.7%	6.9%	7.1%	6.8%
IGHV4-39*01	QVQLQESGPGLVKPSETLSLTCTVSGGSIS	SSSYYWG	WIRQPPGKGLEWIG	SIYYSGSTYYNPSLKS	RVTISVDTSKNQFSLKLSSVTAADTAVYYC	AR	5	16	3	1	5.8%	4.4%	5.6%	5.2%
IGHV4-34*01	QVQLQQWGAGLLKPSETLSLTCAVYGGSFS	GTYWS	WIRQPPGKGLEWIG	EINHSGSTNYNPSLKS	RVTISVDTSKNQFSLKLSSVTAADTAVYYC	AR	5	16	1	1	2.3%	3.9%	4.3%	7.1%
IGHV4-30-4*01	QVQLQESGPGLVKPSQTLSLTCTVSGGSIS	SGSYYWS	WIRQPPGKGLEWIG	YIYYSGSTYYNPSLKS	RVTISVDTSKNQFSLKLSSVTAADTAVYYC	AR	5	16	3	1	3.5%	3.0%	2.3%	3.5%
IGHV4-4*01	QVQLQESGPGLVKPSQTLSLTCAVSGGSIS	SSNWS	WIRQPPGKGLEWIG	EIYHSGSTNYNPSLKS	RVTISVDKSKNQFSLKLSSVTAADTAVYYC	AR	5	16	2	1	1.6%	1.7%	1.5%	1.3%
IGHV4-30-2*01	QVQLQESGSGLVKPSQTLSLTCAVSGGSIS	SGGYSWS	WIRQPPGKGLEWIG	YIYHSGSTYYNPSLKS	RVTISVDRSKNQFSLKLSSVTAADTAVYYC	AR	7	16	3	1	0.6%	0.8%	0.6%	0.9%
IGHV4-b*01	QVQLQESGPGLVKPSETLSLTCAVSGYSIS	SGYYWG	WIRQPPGKGLEWIG	SIYHSGSTYYNPSLKS	RVTISVDTSKNQFSLKLSSVTAADTAVYYC	AR	6	16	2	1	0.5%	0.3%	0.4%	0.4%
IGHV4-28*01	QVQLQESGPGLVKPSQTLSLTCAVSGYSIS	SSNWG	WIRQPPGKGLEWIG	YIYYSGSTYYNPSLKS	RVTMSVDTSKNQFSLKLSSVTAVDTAVYYC	AR	6	16	2	1	0.0%	0.0%	0.1%	0.0%
IGHV4-31*01	QVQLQESGPGLVKPSQTLSLTCTVSGGSIS	SYWIG	IIYPGDSDTRYSPSFQG	QVTISADKSISTAYLQWSSLKASDTAMYYC	AR	5	17	1	3	3.9%	4.2%	4.1%	3.5%
IGHV5-A*01	EVQLVQSGAEVKKPGESLKISCKGSGYSFT	SYWIG	WVRQMPGKGLEWMG	RIDPSDSYTNYSPSFQG	HVTISADKSISTAYLQWSSLKASDTAMYYC	AR	5	17	1	2	0.6%	0.5%	0.5%	0.6%
IGHV6-1*01	QVQLQQSGPGLVKPSQTLSLTCAISGDSVS	SNSAAWN	WIRQSPSRGLEWLG	RTYYRSKWYNDYAVSVKS	RITINPDTSKNQFSLQLNSVTPEDTAVYYC	AR	5	17	3	1	1.5%	1.8%	3.1%	1.7%
IGHV7-4-1*01	QVQLVQSGSELKKPGASVKVSCKASGYTFT	STAMN	WVRQAPGQGLEWMG	WINTNTGNPTTAQQFTG	RFVFSLDTSVETAYLQICSLKAEDTAVYYC	AR	5	17	1	2	1.7%	1.3%	1.2%	1.3%