#295

openPrimeR, a Primer Design Tool for Amplifying B Cell cDNA Encoding Human Antibodies

Matthias Döring[1]*, Christoph Kreer[2,3]*, Nathalie Lehnen[2,3,4], Philipp Schommers[2,5], Meryem Seda Ercanoglu[2,3], Nico Pfeifer[1,6,7]* and Florian Klein[2,3,4]*

* These authors have made equal contributions to the work

[1]Max Planck Institute for Informatics, Saarland Informatics Campus; [2]Institute of Virology, University Hospital Cologne; [3]Center for Molecular Medicine, University Hospital of Cologne; [4]German Center for Infection Research, Cologne; [5]Department I of Internal Medicine, University Hospital Cologne; [6]Methods in Medical Informatics, University of Tübingen; [7]Medical Faculty, University of Tübingen

Introduction

A small fraction of HIV-1 infected individuals (~1%) develops potent broadly neutralizing antibodies (bNAbs) against HIV-1 [1, 2, 3, 4]. These antibodies have been isolated by PCR-based single B-cell cloning methods and have been demonstrated to be promising candidates for HIV-1 prevention and therapy [5, 6].

However, HIV-1 directed bNAbs exhibit exceptionally high levels of somatic mutations. Therefore, designing single primer sets to amplify gene segments encoding bNAbs is particularly challenging. Since leader regions of immunoglobulin variable gene segments (IGHV, IGKV, IGLV) are less mutated, we set out to develop a computational approach to find the minimal primer set of primers that binds to all leaders of functional variable gene segments.

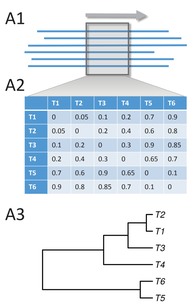
Here, we present openPrimeR, a tool for designing and evaluating multiplex PCR primers. Using openPrimeR, we designed a novel IGHV primer set that exhibits superior performance for amplifying highly mutated variable gene segments from B cells when compared to primer sets binding within the V region.

Data used for primer design

Primers for amplifying IGHVs (*openPrimeR-IGHV*) were designed based on an IMGT data set of 53 functional genes comprising 152 germline alleles [7]. We chose the first 32 nucleotides of the leader region as templates for primer design. Partial leader sequences were completed by our own next-generation sequencing data. To reduce the total number of primers, we allowed binding with 1 mismatch as long as no stop codons were introduced into the amplicons.

Three-phase primer design algorithm

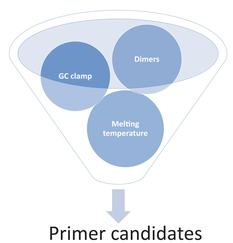
1.



	T1	T2	T3	T4	T5	T6
T1	0	0.05	0.1	0.2	0.7	0.9
T2	0.05	0	0.2	0.4	0.6	0.8
T3	0.1	0.2	0	0.3	0.9	0.85
T4	0.2	0.4	0.3	0	0.65	0.7
T5	0.7	0.6	0.9	0.65	0	0.1
T6	0.9	0.8	0.85	0.7	0.1	0

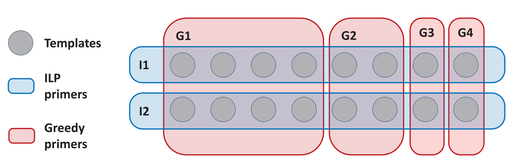
In the initialization phase, a set of potential primer candidates is generated. After the template sequences have been aligned, a phylogenetic tree (A3) is determined for each possible subalignment via hierarchical clustering of the respective dissimilarity matrix (A2). Per traversal of the tree's internal nodes from leaves to root, it is possible to construct degenerate primers by forming the consensus sequence of all descendants of the currently selected node.

2.



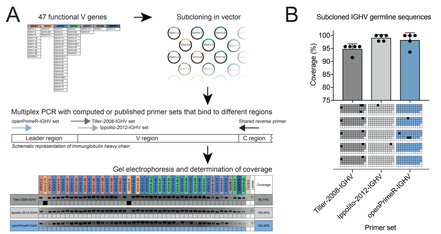
In the filtering phase, primers with unfavorable properties are removed. Up to 12 physicochemical properties can be considered and only those primers fulfilling all active constraints are retained. The user-defined target constraints are automatically adjusted to ensure that the set of remaining primers can attain a high coverage of template sequences.

3.



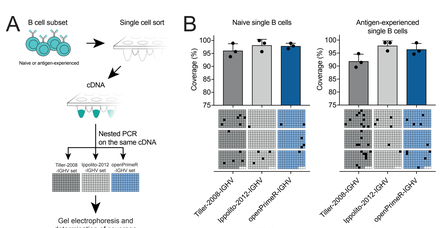
In the last phase, an optimization algorithm (either a greedy algorithm or an integer linear program, ILP) finds a minimal set of primers maximizing the coverage. The figure is based on an example provided by Pearson et al. [7], which illustrates the impact of the optimization algorithm on the size of designed primer sets. Templates are shown as circles. Rectangles represent sets of templates that are covered by an individual primer. Primers selected by ILP or a greedy algorithm are shown in blue and red, respectively. Here, the ILP found the minimal set cover of size 2, while the greedy algorithm found an approximate solution, a cover of size 4 .

Validation on an IGHV library



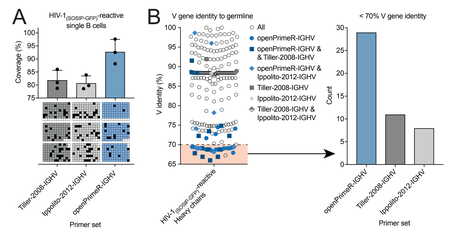
(A) Generation of an IGHV template library by cloning one representative allele of each V gene into a vector backbone. This library was used for comparing and validating three different primer sets: The novel *openPrimeR-IGHV* primer set and two published primer sets, *Tiller-2008-IGHV* [9] and *Ippolito-2012-IGHV* [10]. **(B)** *openPrimeR-IGHV* attains a coverage that is comparable to the previously described sets and successfully amplified all functional IGHV gene segments, as predicted by openPrimeR.

Amplifying single B-cell IGHVs



(A) Naive and antigen-experienced B cells from healthy individuals were FACS-sorted into 96 well plates and cDNA was generated as previously described [9]. Efficacy of nested PCR amplification with *openPrimer-IGHV*, *Tiller-2008-IGHV*, and *Ippolito-2012-IGHV* was compared using the same cDNA templates. **(B)** The *openPrimer-IGHV* set exhibits a comparable performance to the other primer sets, thereby demonstrating the effective amplification of both germline and low-mutated cDNA templates.

HIV-1-specific B cells



Single HIV-1-reactive B cells were isolated from an HIV-1 infected individual with a broad HIV-1 neutralizing serum activity. **(A)** cDNA was generated and the primer set performance was compared on the same cDNAs in three individual experiments. **(B)** V gene germline identity was determined on 211 heavy chain amplicons using IgBLAST [11]. The chance of amplifying a higher mutated sequence (i.e. less identity) increases using *openPrimer-IGHV*. The results show that *openPrimer-IGHV* is capable of amplifying a greater number of highly-mutated B cell receptor sequences than primer sets used for comparison.

Conclusions

- openPrimeR is a versatile tool for designing and evaluating multiplex PCR primer sets
- The *openPrimer-IGHV* primers perform equally well as established primer sets when amplifying single cells
- The *openPrimer-IGHV* primer set is superior to established primer sets when amplifying highly-mutated sequences

References

1. Scheid et al. Science. 2011 Sep 16;333(6049):1633-7. doi: 10.1126/science.1207227.
2. Wu et al. Science. 2010 Aug 13;329(5993):856-61. doi: 10.1126/science.1187659.
3. West et al. Cell. 2014 Feb 13;156(4):633-48. doi: 10.1016/j.cell.2014.01.052.
4. Simek et al. J Virol. 2009 Jul;83(14):7337-48. doi: 10.1128/JVI.00110-09.
5. Moldt et al. Proc Natl Acad Sci U S A. 2012 Nov 13;109(46):18921-5. doi: 10.1073/pnas.1214785109.
6. Klein et al. Science. 2013 Sep 13; 341(6151):1199-204. doi: 10.1126/science.1241144.
7. Lefranc M.P. Cold Spring Harb Protoc. 2011 Jun 1;2011(6):595-603. doi: 10.1101/pdb.top115.
8. Pearson, William R., et al. Discrete Applied Mathematics 71.1-3 (1996): 231-246.
9. Tiller et al. J Immunol Methods. 2008 Jan 1;329(1-2):112-24. Epub 2007 Oct 31.
10. Ippolito et al. PLoS One. 2012;7(4):e35497. doi: 10.1371/journal.pone.0035497.
11. Ye et al. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W34-40. doi: 10.1093/nar/gkt382.

max planck institut informatik
Contact: mdoering@mpi-inf.mpg.de

UNIKLINIK KÖLN

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

openPrimeR

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS