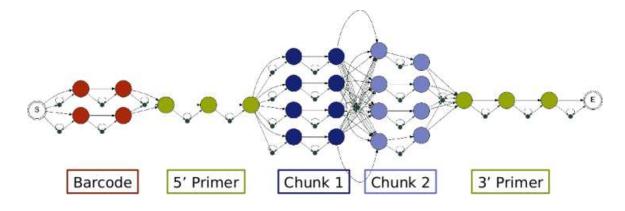
**Table S1.** X-aptamer library construction<sup>a</sup> input file (.csv format) data illustrating the available choices at ten random regions during a bead-based X-aptamer synthesis procedure, and the subsequent bar codes, later appended on the 5'-side, for labeling during next-generation sequencing.

Name	Bar	Random	Alternate	Alter-	Alter-	Alter-	Alter-	Alter-	Alter-	Alter-	Alter-	Alter-	Alter-	Alter-
	Code	Region		nate	nate	nate	nate	nate	nate	nate	nate	nate	nate	nate
Bar Code	1	0	GCTA	TAGC	GTCA	AGTC	CGTA	TGCA	ACTG	GACT	TCGA	CTGA	ATGC	CAGT
5'-primer	0	0	GGGTGAACTGACTCCAGTTGACTGG											
Random Region	0	1	GC	TC	TG	XT								
Random Region	0	1	GTG	GCG	TGG	TXG								
Random Region	0	1	TG	GT	$\mathbf{XC}$	TA								
Random Region	0	1	GG	GX	GG	AC								
Random Region	0	1	AT	GT	CC	CX								
Random Region	0	1	TGC	TTC	XAC	<mark>GCC</mark>								
Random Region	0	1	GTT	GTG	GXA	GAG								
Random Region	0	1	GTG	CCG	GAG	GXC								
Random Region	0	1	GT	TX	GC	TG								
Random Region	0	1	GCG	GCC	GGC	GXG								
3'-primer	0	0	ATGCGAACTGGTGTCACAGCTTA											
<sup>a</sup> Example result	= 5'- <mark>T</mark>	<mark>'AGC</mark> -GG	GTGAACTGACTCCAGTTGACTGG - TG	- <mark>TXG</mark> -X	C- <mark>AC-</mark> G	T- <mark>GCC</mark> -C	GXA- <mark>CC</mark>	G-TX-G	GC-ATC	GCGAAC	TGGTG	TCACA	GCTTA	-3'

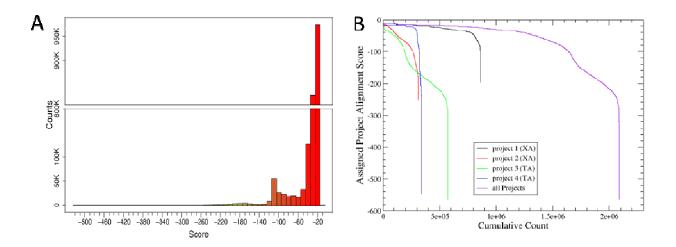
**Table S2.** Example format of closely related sequences within the top 5000 compared to the top 50. In the mutant format in the bottom half of the table deletions (all relative to Seq\_1) are denoted by an asterisk (\*), mutations are noted by the letter of the mismatched base, and insertions are noted in the second column in parentheses.

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>						
>>Close Sequences			EditDist	Occurr	Seq_id	
"GCGTGGTGXTTCGTG	TXGCC"		0	2200	Seq_1	
"GCGTGGTGXTTCGTG	TXTCC"		1	23	Seq_3177	
"TCGTGGTGXTTCGTG	TXGCC"		1	21	Seq_3392	
"GCGTGTTGXTTCGTG	TXGCC"		1	20	Seq_3509	
"GCGTGGTGXTTCTTG	TXGCC"	1	15	Seq_4253		
"GCGTGGTGXTTCGTT	TXGCC"	1	13	Seq_4648		
"GCGTGGTGXTTCGTG	TGGGC"		2	25	Seq_2965	
"GCGTGGTGXTCGTT	TXGCGC"		2	19	Seq_2543	
>>Mutation_format (* = c	lel)					
		-insert	EditDist	t -Occurr	-Seq_id	
"GCGTGGTGXTTCGTG	TXGCC"	"	" 0	2200	Seq_1	
TT	т "	"	" 1	23	Seq_3177	
"т	"	"	" 1	21	Seq_3392	
"Т	"	"	" 1	20	Seq_3509	
" Т	"	"	" 1	15	Seq_4253	
" Т	"	"	" 1	13	Seq_4648	
11	G G "	"	" 2	25	Seq_2965	
" *	"	"(G25)	" 2	19	Seq_2543	

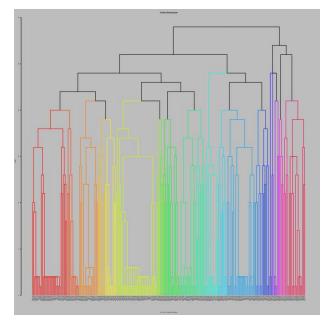
**Figure S1.** Schematic representation of the Markov model topology for a library with only two bar codes, a 3-base 5'-primer, two random fragments with four choices each, and a 3-base 3'-primer. The bar code may be on either end of the 5'- or 3'-primers.



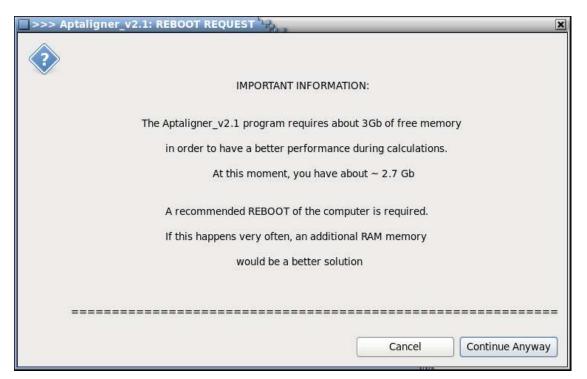
**Figure S2.** Sequence counts by score. A) Histogram showing scores for a single project when noise and cut-off filters were not used. Even when not using length and noise cut-off filters, 75% or more of all sequences in a project have good scores. B) In total, about 1.5 million good sequences (scores of about - 60 or higher) were obtained from a data chip containing NGS sequences from four different projects. Projects 1 through 4, respectively, had good sequence scores for approximately 800k, 200k, 200k, and 300k sequences. The sharp drops at the end of each curve represent a very small number of really poor sequences (noise), which are forced into a project when noise and length cut-off filters are not used (older Aptaligner v1.0). XA and TA refer to X-Aptamers and normal aptamers, respectively.



**Figure S3**. Example clustering diagram produced by Aptaligner for the top 250 sequences. Clustering information is also reported in text files for each project.



**Figure S4**. Example warning when RAM *may* be an issue with running Aptaligner successfully. The actual amount of RAM needed will depend on the complexity of the library being analyzed. This warning can sometimes be ignored, but will often lead to failure. When analyzing 5 million sequences, 3+ GB of RAM is recommended.



**Figure S5**. An example GUI comparing needed disk space (up to 12 GB) versus current free disk space. Disk space requirements depend on the number of sequences analyzed and the number of library design files against which they are compared. If the number of free GB are not shown in this box ("you have about ~ GB), and the user continues anyway, the program is highly likely to crash in within a few minutes with a resulting error declaring that the project 1 sequence file "HAS NOT THE EXPECTED FORMAT" – because it is empty.

>>> Aptaligner_v2.1: HARD DRIVE FREE SPACE REQUIREMENT
IMPORTANT INFORMATION:
The Aptaligner_v2.1 program might requires about 12 Gb of Hard Drive free space, per CHIP in order to save all its data.
At this moment, you have about ~ 73 Gb
a) backup and delete some files or
b) move your JOB directory to another Hard Drive and re-run the program or
c) continue and fix it later
Cancel Continue Anyway

**Figure S6**. Error resulting from an incorrect Python build. Re-installing python has been shown to fix this error, which appears to result from the C code not compiling correctly during Aptaligner set up. Reducing the RAM to 0.25 GB and CPU to 1 could not reproduce this error when analyzing the "medium test file provided at .