The xGen™ Off-Target Quality Control (QC) Method increases accessible probe design space for hybridization capture when compared to the repeat annotation based QC method

Introduction

Targeted next generation sequencing (NGS) has been widely adopted for research purposes because it provides a more cost-effective method than whole genome sequencing to focus on regions of interest within a genome, such as the exome (whole exome sequencing, or WES), non-coding genes, or a subset of the exome [1]. Hybridization capture is an effective method to carry out targeted NGS [2], and IDT offers a complete stocked and custom panel portfolio to address hybridization capture needs. In a hybridization capture workflow, probes are designed to be complementary to the targeted regions so that they will hybridize to the targets of interest. Some targeted regions share high homology with DNA sequences outside the regions of interest. Non-specific capture of DNA sequences outside the targets of interest is known as off-target capture, and these off-target sequences contribute to the problem of inefficient use of sequencing data or even wasted sequencing reads. To reduce non-specific capture, one solution is to screen the designed probes with off-target quality control (QC) to mitigate the non-specific capture risk. However, computationally removing all probes with any non-specific capture risk from the final panel design may result in missed target regions of importance. The off-target QC performed by a software algorithm must balance two goals: minimize the number of off-target reads out of the total sequencing reads, and maximize the targeted regions covered (Figure 1). As will be described in this white paper, IDT uses a proprietary off-target QC method in our xGen Hyb Panel Design Tool.

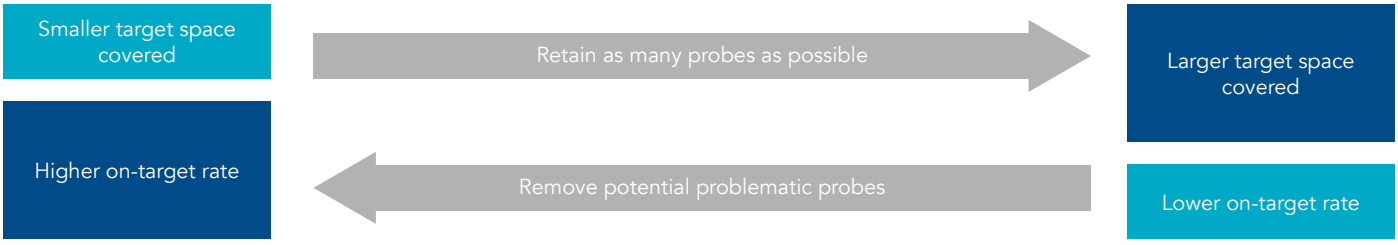


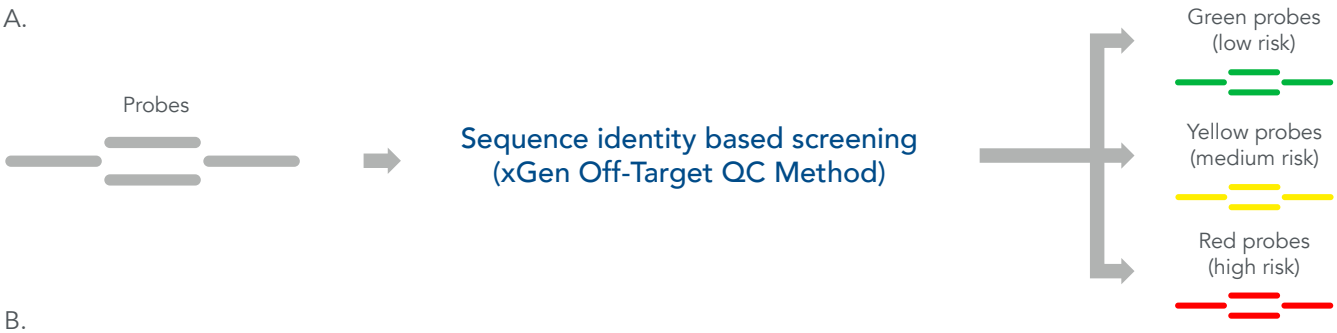
Figure 1. Goals of off-target QC. The ideal off-target QC method has two goals: (1) minimize the number of off-target reads out of the total sequencing reads (2) maximize the amount of target regions covered.

DIFFERENT OFF-TARGET QC METHODS HAVE DIFFERENT QC OUTCOMES

In order for bioinformatics algorithms in research to perform off-target QC on designed probes, the off-target capture risk needs to be evaluated. The repeat annotation (RA) based off-target QC method checks the targeted regions against a pre-computed repeat annotation to avoid designing probes in the regions masked by repeats. This method depends on the existence and quality of the pre-computed repeat annotation, which is not always available for the organism being studied. Additionally, the annotated repeats can be short or have low copy number, in which case there may not be a significant off-target capture risk. It is also possible that there are regions sharing significant homology outside of the annotated repeat regions, which may lead to significant off-target capture.

The bioinformatics scientists at IDT have developed a proprietary off-target QC method, the xGen Off-Target QC Method, which utilizes a sequence identity based approach and thus does not have the same limitations as the RA-based off-target QC method (Figure 2A). Our off-target QC software method conducts sequence identity based screening on designed probes and categorizes their off-target capture risk (Figure 2A). We categorize probes into three categories—low risk (green), medium risk (yellow), and high risk (red). If there are targets that are of high interest but fall into the medium risk (yellow) category, it may be beneficial to add these in addition to low risk (green) probes. Overall, the xGen off-target QC method provides greater flexibility to the user to expand design space while minimizing off-target risk. When the xGen Off-Target QC Method is compared with the RA-based off-target QC method, the off-target QC outcomes from the two methods can differ (Figure 2B). On the one hand, probes failing the RA-based off-target QC method may be categorized as medium or low risk in the xGen Off-Target QC Method due to partial or minimal overlap with the annotated repeat (Figure 2B). On the other hand, probes passing the RA-based off-target QC method could be categorized as high risk in the xGen Off-Target QC Method due to homology outside of the annotated repeats (Figure 2B).

A.



B.

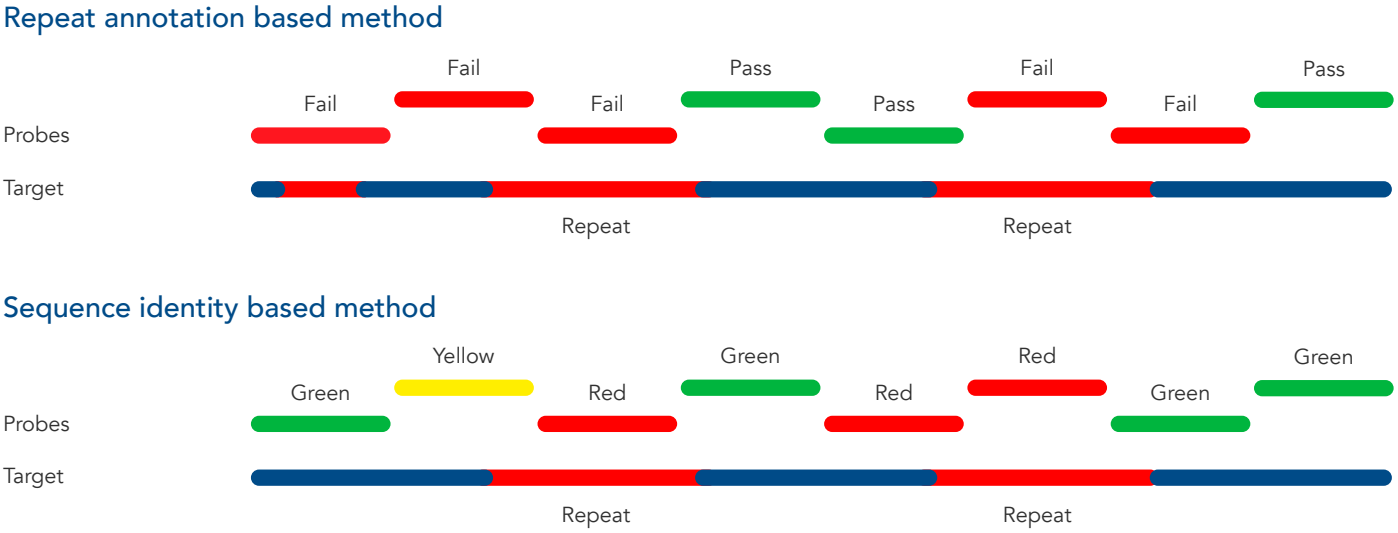


Figure 2. Different off-target QC methods. (**A**) The xGen Off-Target QC Method utilizes a sequence identity based approach to screen designed probes for off-target capture risk. (**B**) Example diagram showing potentially different outcomes from the two off-target QC methods. Probes failing the RA-based off-target QC method may be categorized as mid or low risk in the xGen Off-Target QC Method due to partial or minimal overlap with the annotated repeat, and probes passing the RA-based off-target QC method could be categorized as high risk in the xGen Off-Target QC Method.

THE xGen OFF-TARGET QC METHOD IMPROVES DESIGN SPACE ACCESSIBILITY

To understand what the difference in off-target QC outcomes means in terms of design space accessibility, approximately 30 Mb of target space from human chromosome 21 was examined as an example. The two off-target QC methods were used to screen probes in the example target space, and the off-target QC outcomes from the two different methods were summarized in a Venn diagram (Figure 3). We found that 63.9% of the example target space passed (as green) both the xGen Off-Target QC Method and the RA-based off-target QC method (Figure 3). However, the xGen Off-Target QC Method allowed 35.5% more out of the total example target space examined, compared to the RA-based off-target QC method, while covering almost all (99.5%) of the target space passing the RA-based off-target QC method (Figure 3).

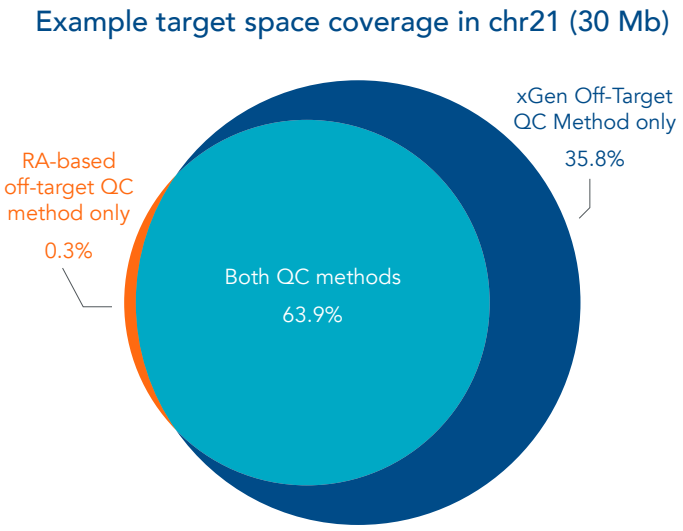


Figure 3. The xGen Off-Target QC Method improves design space accessibility. Probes in the example target space from human chromosome 21 were screened using the two off-target QC methods. As shown here, 63.9% of the example target space passed both off-target QC methods. Notably, 35.8% of the example target space passed only the xGen Off-Target QC Method, whereas 0.3% of the example target space passed only the RA-based off-target QC method. Therefore, the xGen Off-Target QC Method expands the accessible target space by 35.8% but loses only 0.3% compared to the RA-based off-target QC method.

THE xGen OFF-TARGET QC METHOD ADDS ADDITIONAL DESIGN SPACE WITHOUT SACRIFICING PANEL PERFORMANCE

The design space accessibility comparison in the previous section shows that the xGen Off-Target QC Method improves design space accessibility. An experiment was designed to assess the impact on panel functionality when additional target regions from the two off-target QC methods were included (Figure 4). The design space that was made accessible exclusively by either of the two off-target QC methods was tested with a backbone panel previously shown to pass both off-target QC methods (Figure 4A). The backbone panel was used as a control, and the exclusively allowed design space from each of the two off-target QC methods was spiked into the backbone panel (Figure 4B).

A.

B.

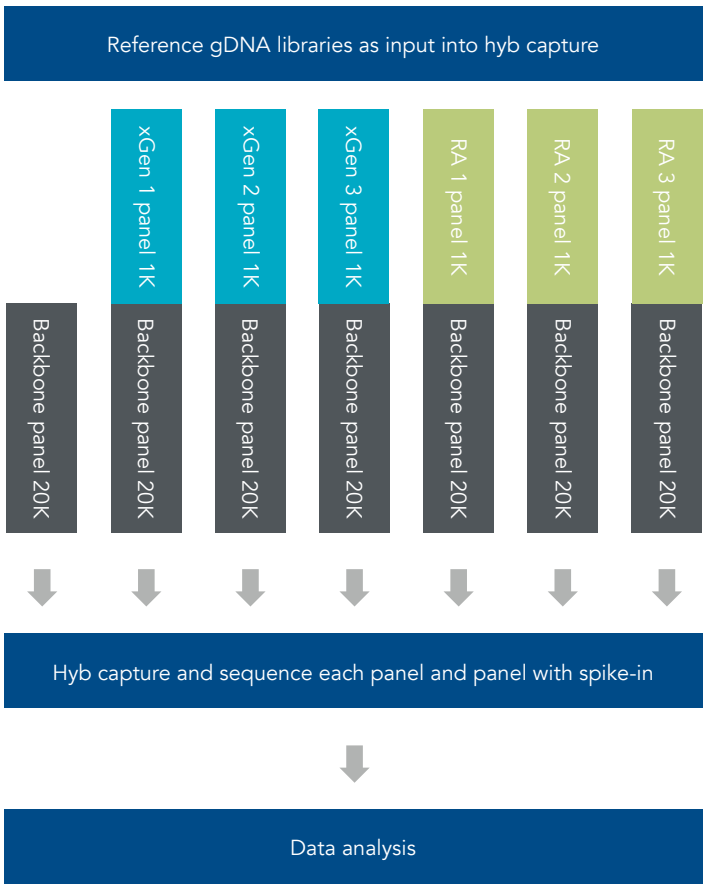


Figure 4. Comparison of off-target QC methods. (**A**) Design space definition of the panels. The backbone panel passes both off-target QC methods. The "xGen" panel design space passes the xGen Off-Target QC Method but fails the repeat annotation (RA) based off-target QC method, whereas the "RA" panel design space passes the repeat annotation based off-target QC method but fails the xGen Off-Target QC Method. (**B**) Experimental setup of the off-target QC method comparison. Reference gDNA samples (100 ng) prepared with the xGen™ cfDNA & FFPE DNA Library Prep Kit were taken through hybridization capture. The backbone panel with approximately 20,000 probes was first used as control itself, and then the six spike-in panels (three "xGen" panels and three "RA" panels) with 1,000 probes each were individually captured together with the backbone panel. The samples were then sequenced and analyzed.

Since this experiment was designed to compare two off-target QC methods, we focused on evaluating on-target rates. The Picard HsMetrics algorithm offers a standard on-target rate metric (PCT_SELECTED_BASES) that can be conveniently referenced by users to get a standard estimate of the on-target rate [3]. The xGen Off-Target QC Method adds design space with minimal impact to the on-target rate compared to the repeat annotation based off-target QC method (Figure 5A). The number of off-target reads contributed by probes can also be used to measure the off-target capture risk of the designed probes. The result shows that the xGen Off-Target QC Method minimizes the number of off-target reads contributed by probes while adding design space compared to the RA-based off-target QC method (Figure 5B). Additionally, the result demonstrates that the xGen Off-Target QC Method adds design space without increasing percentage of ambiguity and zero coverage targets (Figures 5C–D). Overall, for this experiment, the xGen Off-Target QC Method achieved comparable panel performance as the control backbone panel and improved the accuracy of off-target QC when compared to the repeat annotation based off-target QC method (Figure 5).

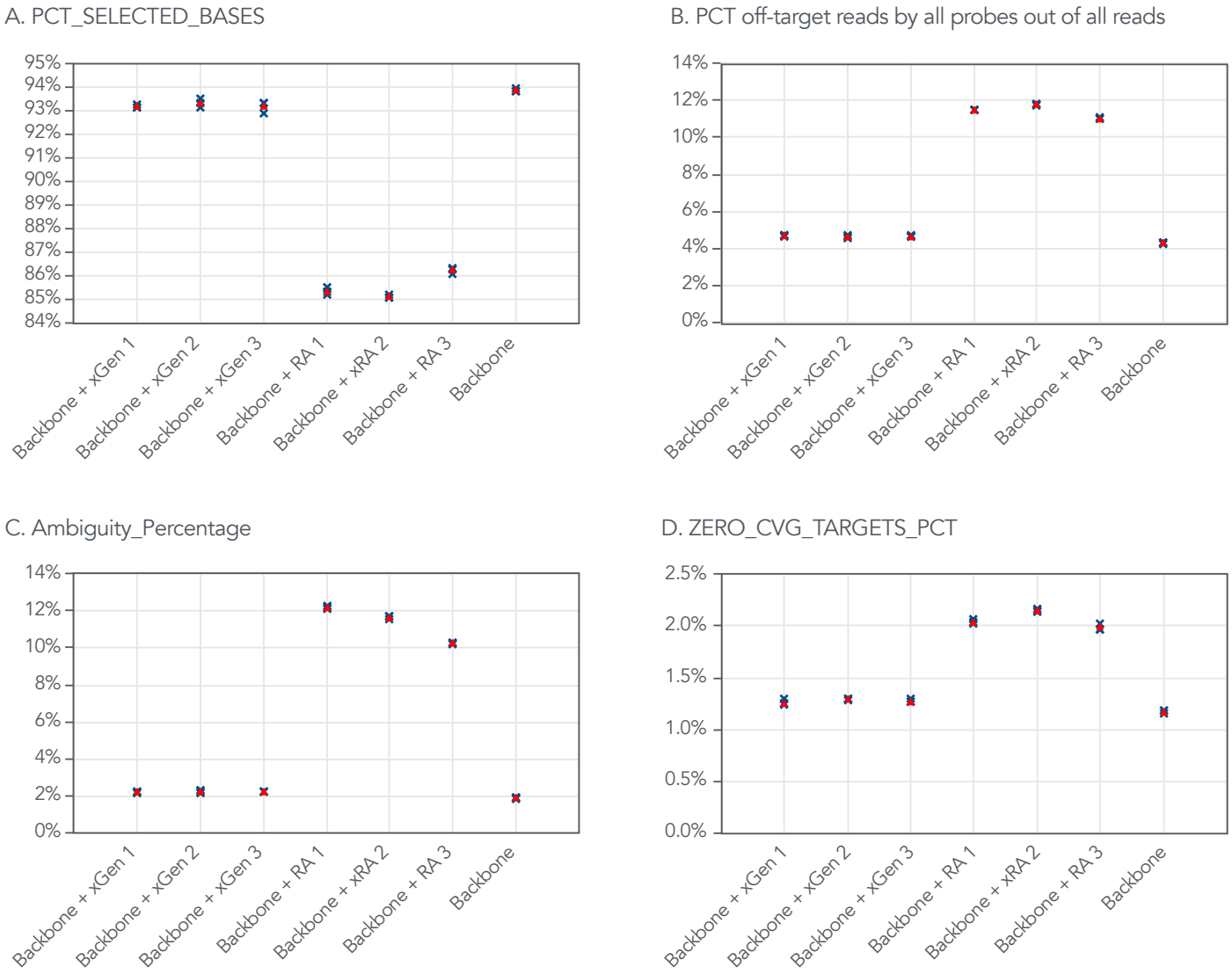


Figure 5. The xGen Off-Target QC Method adds design space without sacrificing panel performance. (**A**) The xGen Off-Target QC Method adds design space with minimal impact on the on-target rate compared to the repeat annotation based off-target QC method. (**B**) The xGen Off-Target QC Method minimizes the number of off-target reads contributed by probes compared to the repeat annotation based off-target QC method. (**C-D**) The xGen Off-Target QC Method adds additional design space without increasing the percentage of ambiguity and zero coverage targets. The blue X's represent experimental replicates, and the red triangles represent the average of three replicates.

CONCLUSION

Off-target QC is a critical step to mitigate the off-target capture risk of designed probes. However, removing designed probes increases the amount of target regions of interest potentially not covered by probes. The xGen Off-Target QC Method allows probe design against an increased number of target regions without sacrificing panel results compared to the RA-based off-target QC method, for the regions and probes tested. This improvement over the RA-based off-target method allows users to cover more target regions of interest without worrying about the off-target capture risk, which ultimately helps users obtain an efficient targeted sequencing assay. Therefore, the xGen Off-Target QC Method is used in our xGen Hyb Panel Design Tool.

METHOD

The design space accessibility comparison was performed by examining ~30 Mb of target space from human genome hg38 chromosome 21. Probes in the example target space were screened using the two off-target QC methods and categorized based on their passing and failing status from the two off-target QC methods. The total number of bases covered by the probes was counted to calculate the percentages shown in the Venn diagram.

A backbone panel with 20,038 probes was generated that passed both off-target QC methods. Then, three xGen Hyb Panels each containing 1,000 probes passing only the xGen Off-Target QC Method and three RA panels each containing 1,000 probes passing only the repeat annotation based off-target QC method were generated with no overlap to the backbone panel. Panels were ordered as xGen custom hybridization panels (Accel option) and mixed together during hybridization following the xGen hyb and wash protocol. Coriell genomic DNA was used to generate libraries using the xGen cfDNA & FFPE DNA Library Prep Kit, and the xGen hybridization and wash protocol was carried out using a 4-hour hybridization. Post-capture libraries were pooled and sequenced on an Illumina NextSeq™ instrument using a High-Output flowcell.

The samples with the backbone panel itself and the samples with spike-in panels were subsampled to different total numbers of reads to account for their different numbers of probes, so that they could theoretically achieve similar mean target coverages. PCT_SELECTED_BASES and ZERO_CVG_TARGETS_PCT were calculated using Picard HsMetrics [3]. Ambiguity_Percentage was calculated with our in-house script using a mapping quality of 20 as the cutoff after the duplicate reads were excluded. The percent of off-target reads contributed by all probes out of total reads was calculated using our proprietary in-house analysis tool.

REFERENCES

1. Yohe S and Thyagarajan B (2017). *Arch Pathol Lab Med* 141(11): 1544-1557.
2. Samorodnitsky E, Datta J, et al. (2015). *J Mol Diagn* 17(1): 64-75.
3. Institute B (2019) Picard Toolkit. Broad Institute, GitHub repository http://broadinstitute.github.io/picard/.

The xGen™ Off-Target Quality Control (QC) Method increases accessible probe design space when compared to the repeat annotation based QC method

Technical support: applicationsupport@idtdna.com

MAPPS Doc ID# RUO21-0257_001 09/21