

ThruPLEX Tag-seq: how deep should I sequence?

Confident minor allele frequency detection

When using [ThruPLEX Tag-seq](#) in your research, you may have questions about how deep you need to sequence to detect the Minor Allele Frequency (MAF) of interest. In order to determine the detectable MAF, there are several factors to consider, including input amount, depth of sequencing, and capture panel size. ThruPLEX Tag-seq provides confident MAF detection by including 16 million unique molecular tags (UMTs) to label each DNA molecule. Using the UMTs, bioinformatics software groups the duplicates into amplification families and constructs a consensus sequence, thus reducing false positives.

Input amount

Input amount plays a critical role in MAF. An appropriate input amount should be selected to ensure that an adequate number of copies of the variant in question is present for detection. Table I indicates the input amount, total haploid genome copies, and total variant copies available for library preparation at various allele frequencies. Note that the number of copies available for detection will be lower than the number shown, as there is loss during the library preparation and enrichment process.

Table 1. Estimated genome copies available for library preparation

Input amount	Total haploid genome copies*	Total variant copies at the indicated allele frequency		
		5%	1%	0.5%
50 ng	16,666	833	166	83
30 ng	10,000	500	100	50
10 ng	3,333	166	33	16
5 ng	1,666	83	16	8
1 ng	333	16	3	1

*Calculated using 3 pg as the mass of a haploid genome. The genomic complexity of plasma samples is highly variable. All numbers are rounded down to the nearest whole number.

Sequencing depth

Another factor that affects detection sensitivity is sequencing depth. Generally, to detect lower MAFs, a greater amount of sequencing is required. ThruPLEX Tag-seq uses UMTs to bioinformatically group duplicates into amplification families. An amplification family size of 8–10 reads is recommended for maximum specificity, but can be changed based on experimental needs (Kennedy et al. 2014).

In order to estimate the amount of sequencing needed to detect the MAF desired, the number of unique molecules required to make a variant call must be determined. For example, for an allele frequency of 1%, if three unique molecules are required to make a variant call and each amplification family has approximately 10 reads, then you would need to sequence to roughly 3,000X coverage. Table II and equation below can be used as a reference.

Sequencing depth = (number of unique variants required to make a variant call ÷ allele frequency) × (approximate number of reads in each amplification family)

For example: $(3 \div 0.01) \times 10 = 3,000\text{X}$ coverage required

Table 2. Estimated mean raw sequencing depth required*

Minimum number of unique molecules to make a variant call	Allele frequency		
	5%	1%	0.5%
3	600X	3,000X	6,000X
5	1,000X	5,000X	10,000X
10	2,000X	10,000X	20,000X

*Raw sequencing depth includes all reads prior to removal of duplicates. This is calculated using a target peak amplification family size of 10 reads per unique molecule.

Target enrichment

One way to decrease the amount of sequencing needed is to perform target enrichment using hybrid capture. Targeted panels enrich the genes of interest, which decreases the total amount of bases needing coverage. Additional major considerations of hybrid capture, then, are the desired coverage (AKA sequencing depth), the size of the target panel, the sequencing read length, and the fraction of on-target reads (defined as reads mapping to your target of interest). The following equation provides an estimate for the number of reads required for each sample:

$$\text{Millions of reads required} = (\text{coverage} \times \text{target panel size [in mb]}) \div (\text{read length} \times \text{on-target fraction})$$

Example using 150-bp paired-end reads: $(600 \times 5) \div (300 \times 0.50) = 20$ million reads required

Important factors to consider

Input amount, depth of sequencing, and capture panel size are all factors to consider when determining MAF, and here we have provided several guidelines for designing your Tag-seq experiments. Additionally, it is important to note that other factors such as sample quality and your choice of data processing algorithms may play a role.

References

Kennedy, S. et al. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nature Protocols* **9**, 2586–2606 (2014).

Takara Bio USA, Inc.

United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2023 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com.