



Guide

Advyser Sequence coverage calculator

TABLE OF CONTENTS

TABLE OF CONTENTS	2
1. INTRODUCTION	3
2. USING THE CALCULATOR	3
2.1 Select sequencing system and kit	3
2.2 Select Devyser kit	3
2.3 Circle diagram	5
2.4 Report	6
3. COMPUTE SAMPLE COVERAGE	6
4. REFERENCES	7
5. REVISION HISTORY	7

1. INTRODUCTION

The identification of variants from sequencing data, known as variant calling, requires that a certain number of sequencing reads has been generated that contain the variant to be called. The number of sequencing reads covering a certain nucleotide is referred to as the depth of coverage, sequencing depth, or read depth, or simply coverage. The higher the number of high-quality sequence reads, i.e. the higher the coverage, the more confident is the base calling. The coverage will primarily depend on the sequencing chemistry and reagents used, i.e. the flow cell and the number of samples that are run together in one sequencing run. Multiple other factors will however also influence the coverage over a nucleotide, such as sample quality, DNA concentration, nature of the target region, GC-content, library preparation method, sequencing cluster density, sequencing quality etc.

The Devyser sequence coverage calculator, hereafter referred to as “Calculator”, has been designed as an easy to use web application that facilitates the efficient and cost-effective planning of each Devyser sequencing run. The Calculator can be used to determine:

- The number of samples that can be pooled in one flow cell in order to achieve the required coverage.
- How multiple library pools should be combined in order to achieve required coverage of each sample when sequenced together.

The recommendations given by the Calculator are based on theoretical calculations described in the literature and confirmed in real sample preparation and sequencing runs. However, since the coverage can be affected by several parameters, we recommend each laboratory to use the numbers generated by the Calculator as a starting point in their internal validation process and adjust the number of samples per sequencing run to ensure enough coverage as well as optimal usage of the sequencing flow cell.

NOTE

The performance claims for the CE-IVD version of the respective Devyser kits are based on data generated by strictly following the protocol and where the minimum required coverage of each amplicon has been reached.

2. USING THE CALCULATOR

2.1 Select sequencing system and kit

In this first section of the Calculator, use the dropdown menus to first select a **Sequencing system** and then a **Kit/Flow cell**. Consult the Instruction for use (IFU)/Handbook for the Devyser NGS kit to be sequenced to ensure that it is compatible with the selected sequencing system.

It is possible to change kit/flow cell in the Calculator at any time if a different size or type is preferred.

2.2 Select Devyser kit

In the second section of the Calculator, select the Devyser kit(s) used to prepare the library pool(s) by clicking the kit-specific button(s). Choose the appropriate coverage calculation mode to detect germline, or somatic variants, or to select copy number variant (CNV) analysis.

It is also possible to add a library pool generated using a kit from a different vendor by clicking the **Other** button. In this case, it is assumed that the library concentration is 2 nM and the user needs to manually enter the total number of reads required for this pool in the **Number of reads/read pairs** column of the library table. Consult the respective vendor for instructions on how to calculate the required number of reads.

A library pool may be deleted by first ticking the check box to the left of the library table before clicking the **Remove** button.

NOTE

Contact Devyser technical support at techsupport@devyser.com prior to pooling Devyser libraries with libraries from a different vendor to verify if the different libraries can be sequenced together in the same run.

The library table consist of six columns described below:

Devyser kit / variant

The name of the kit and coverage calculation mode selected by the user.

Amplicon pool name

A custom name for each added library pool can be entered or left as default.

Samples

The number of samples included in each library pool can be adjusted by entering a number or using the up/down arrows (arrows are not visible in all internet browsers).

Minimal coverage per allele

The minimal number of reads per allele can be adjusted by entering a number or using the scrolling bar. This number is fundamental in the computation of the required number of reads per sample.

For reliable calling of single nucleotide variants (SNVs) and insertion/deletion variants (indels), NGS guidelines recommend a minimal coverage between 10-125 reads per allele, corresponding to 20-250 read coverage per amplicon^{1,2,3}. The default value in the Calculator is set to 50 reads per allele (100 per amplicon).

For calling CNVs, a higher coverage is favorable. The default values in the Calculator is set to 100 reads per allele (200 per amplicon).

The default values in the Calculator correspond to Devyser's minimum coverage requirements.

NOTE

It is not recommended to decrease the default values as it increases the risk of not reaching the required minimal coverage needed for to confidently call a variant.

Variant allele frequency (VAF)

Germline variants occur at an allele frequency of 50% (heterozygous), or 100% (homozygous). The VAF value for germline calculation modes in the Calculator is fixed to 50%.

Somatic variants can occur at VAF that differ from 50% or 100% since the percentage of tumor cells in a tumor sample is often lower than 100%, and since tumor heterogeneity may occur. If a somatic coverage calculation mode is selected, specify the desired VAF.

To confidently call somatic variants that have a low VAF, a higher coverage per amplicon is needed to achieve the required minimal coverage of 50 reads per variant allele. Therefore, the specified VAF greatly affects the minimal coverage for each amplicon.

As an example, in order to reliably identify a variant at VAF 10%, at least 50 reads carrying the variant are required, which corresponds to at least 500 reads covering this position.

$$\frac{COV_{min}}{VAF} = \frac{50}{0.1} = 500 \text{ reads}$$

NOTE VAF below 5% has not been tested with the Devyser BRCA kit.

Number of reads/read pairs

This field shows the number of sequencing reads (when sequencing in Single Read mode), or read pairs (when sequencing in Paired End mode) required to achieve the minimal coverage for each amplicon in each sample in the pool. This number is computed automatically for Devyser libraries, using the entered values. Refer to the section **3. Compute sample coverage** for details on how this number is computed.

When **Other** is selected as library pool, the default starting value is 0, and the user must enter a value or use the up/down arrows (arrows are not visible in all browsers).

2.3 Circle diagram

After addition of library pools, the estimated flow cell capacity for the individual library pool as well as the total used flow cell capacity is indicated in the circle diagram.

Hover the mouse over the circle diagram to see the respective percentages from each library pool. The color(s) in the circle diagram correspond to the colored box(es) of each library pool displayed in the **Amplicon pool name** column of the library table.

If the total used capacity exceeds the capacity of the selected flow cell, the user can modify the coverage parameters for one or more library pools, reduce the number of samples, remove entirely a library pool, and/or select a larger flow cell.

NOTE The capacity of the kit/flow cell is set to 90% of the capacity specified by Illumina.

NOTE 1% of the reads are automatically allocated to the PhiX control to obtain sufficient nucleotide diversity. The PhiX portion is represented in the circle diagram.

2.4 Report

Clicking the **Generate report** button on the lower right corner creates a report that summarizes the parameters and values previously selected and described above.

When there is remaining capacity on the sequencing kit/flow cell, each library pool will generate proportionally more reads in the sequencing run, which is outlined in the column “% of run” in the report table.

The last column of the report table includes volume information to correctly combine different library pools. Refer to the note written below the report table for instructions on how to combine different libraries pools together. If only one library pool was entered in the Calculator, “N/A” will be displayed. In this case, proceed directly to the sequencing step as described in the Devyser kit IFU/Handbook.

3. COMPUTE SAMPLE COVERAGE

When selecting a Devyser kit, the field **Number of reads/read pairs** is automatically computed using the following variables:

- The minimal number of reads per allele needed to ensure correct variant calling

minimal coverage per allele = COV_{min}

- The desired variant limit of detection (LOD)

variant allele frequency = VAF

- The uniformity of the library preparation assay

uniformity factor = U

The uniformity factor (U) is among others dependent on sample type. DNA extracted from blood is for example expected to have a higher uniformity factor compared to DNA extracted from FFPE samples.

- The size of the target region, i.e. number of amplicons in the assay

number of amplicons = AMP

The required number of reads per sample is calculated by taking the listed variables into account and by using the formula below. Variables U and AMP are specific to each Devyser assay and are fixed in the Calculator.

$$COV_{SAMPLE} = \frac{COV_{min}}{VAF \times U} \times AMP$$

The mean coverage per amplicon in the assay might occasionally be a more informative parameter. It is calculated by taking the minimal coverage per allele and variant allele frequency

into account but also includes the uniformity factor. The mean coverage value per amplicon is listed in the report.

$$COV_{MEAN} = \frac{COV_{min}}{VAF \times U}$$

4. REFERENCES

1. Weiss MM, Van der Zwaag B, Jongbloed JD, Vogel MJ, Brüggewirth HT, Lekanne Deprez RH, Mook O, Ruivenkamp CA, van Slegtenhorst MA, van den Wijngaard A, Waisfisz Q, Nelen MR, van der Stoep N. **Best practice guidelines for the use of next-generation sequencing applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories.** Hum Mutat 2013; 34:1313-21.
2. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E. **Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee: ACMG clinical laboratory standards for next-generation sequencing.** Genet Med 2013; 15:733-47.
3. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. **Guidelines for Validation of Next-Generation Sequencing Based Oncology Panels - A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists.** J Mol Diagn 2017; 19(3):341-365.

5. REVISION HISTORY

Version 2021-06-30

- Editorial changes

Version 2020-01-20

- Editorial changes

Version 2018-10-18

- Modified to reflect graphical changes in the Calculator
- Editorial changes
- Calculations based on 90% of flow cell capacity

Version 2017-11-14

- New (v1.0)