SOPHiA™

SOPHiAGENETICS.COM

Genomic Integrity Report

Run name: 21752-000072

Genomic Integrity Report created: May 5, 2023

Number of samples tested : 4

Configuration : HRD_v1_spikeinWGS_v2, A2

Contents

Genomic integrity detection by NGS

1.1 Background

Homologous Recombination Repair (HRR) Deficiency (HRD) is a phenotype of cancer cells, particularly in Ovarian cancers. One method for determining the presence of HRD is to sequence known HRR genes and look for deleterious, or pathogenic, mutations. Disruptions of genes involved in the HRR pathway can cause HRD. Another method for determining the presence of HRD is to look for the effect (or phenotype) of the loss of HRR function by quantifying the degree of genomic aberrations or scars in the genome. This is also called "genomic integrity". HRR is necessary for the error-free repair of double strand breaks and for maintaining genome integrity. Double strand breaks are a particularly genotoxic form of DNA damage. A consequence of HRD is the accumulation of genomic aberrations due to the cell's inability to repair double strand breaks. One method for quantifying these genomic aberrations is through the use of Whole Genome Sequencing (WGS). More specifically, our method uses a deep learning algorithm that has been specifically trained to recognize patterns of Homologous Repair Deficiency in the genome of Ovarian cancer samples. The algorithm analyzes low-pass WGS (lpWGS) data and produces a "Genomic Integrity" (GI) index that is a quantitative measurement of the level of genomic aberrations present in the sample being tested. A sample with a GI index above a specific threshold indicates that the sample is GI positive and has low genomic integrity. A sample with a GI index below the threshold indicates that the sample is GI negative and has a high genomic integrity.

1.2 Method

1.2.1 Overview

The GI indices presented in this report are computed using a proprietary deep learning approach that takes as input the lpWGS coverage profile measured in each sample of interest. To compute the final GI status, the following algorithmic steps are performed sequentially:

- Processing of lpWGS data. WGS paired-end reads are mapped to the human reference genome and processed to trim adaptors and low-quality base calls. The WGS coverage profile is computed and normalized.
- Quality assessment (QA). The normalized WGS coverage profile undergoes QA, in which the metrics reported in Table 1.1 are computed and one of the following QA statuses is assigned to the sample:
  - High quality: the quality of the data is sufficient to confidently compute a GI index and a GI status.
  - Medium quality: the quality of the data is lower (compared to high quality) and, as a consequence, the deep learning algorithm may not succeed in computing a GI index.
  - Low quality: the quality of the data does not meet the criteria required to compute a GI index.
- GI index calculation. The GI index is obtained by processing the normalized WGS coverage profile with a deep learning algorithm. The GI index quantifies the level of genomic integrity.
- GI status determination. A GI status that applies to Ovarian Cancer samples is determined by combining the sample QA status and the GI index. Five outcomes are possible:
  - GI positive: samples with a GI index larger or equal than 0.
  - GI negative: samples with a GI index smaller than 0.

- GI negative*: samples with a GI index smaller than 0 but featuring an increased risk of false negative calls due to low signal to noise ratio.
- GI inconclusive: medium quality samples for which the deep learning algorithm did not succeed in computing a GI index due to insufficient signal to noise ratio. A common cause that leads to a GI inconclusive status is insufficient tumor content.
- GI rejected: low quality samples discarded from GI analysis. Samples are rejected if the number of DNA fragments available for WGS coverage profile calculation is insufficient, if the noise of the WGS coverage profile is excessive, or if the proportion of coverage outliers is excessive.

1.2.2 Definition of QA and GI metrics

Definition of metrics reported in Table 1.1:

- **Total nb. of fragments**: the total number of DNA fragments (paired-end reads) that are properly mapped.
- **Nb. WGS fragments**: the total number of DNA fragments available for the raw coverage WGS profile calculation. DNA fragments mapping to the genomic regions enriched for variant calling are excluded. If the number of WGS fragments is smaller than 4 millions, the QA status is deemed low.
- **Percentage WGS fragments**: fraction of the total number of WGS fragments over the total number of fragments.
- **Proportion of Coverage Outliers**: percentage of WGS regions considered for GI analysis which feature an artefactual and excessive localised coverage which is compensated by the coverage normalisation algorithm. If the proportion of coverage outliers is larger than 20%, the QA status is deemed low.
- **Purity/ploidy ratio**: the ratio between sample tumor content and sample ploidy, estimated by measuring the strength of the signal induced in the normalized WGS coverage profile by a copy number change. If the purity/ploidy ratio estimation fails or if the estimated value is lower than 0.1 (suggesting insufficient tumor content), the QA status is deemed medium.
- **Residual noise**: residual noise is computed by measuring the standard deviation of the normalized WGS coverage profile (black dots in Fig. 1.1) with respect to the smoothed WGS coverage profile (orange line in Fig. 1.1). If residual noise is larger than 0.17, the QA status is deemed low.
- **SNR**: strength of the signal induced in the normalized WGS coverage profile by all copy number aberrations present in the sample divided by the residual noise. Samples with SNR smaller than 0.55 will be classified as medium quality GI inconclusive. SNR is also considered by the algorithm to assign the GI Negative* status.

Definition of metrics reported in Table 1.2:

- **QA Status**: the quality assessment status of a sample (see section 1.2.1).
- **GI Index**: the GI index is a scalar value, ranging between -20 and 20, that quantifies the level of genomic integrity. High GI indices reflect low levels of genomic integrity. Low GI indices reflect high levels of genomic integrity.
- **GI Status**: the genomic integrity status of a sample (see section 1.2.1).

1.3 Recommendations and Limitations

For interpreting results, please refer to the Instructions For Use (IFUs), and in particular the limitations listed in the relevant section.

1.4 Results

1.4.1 Sample QA

Sample	Total nb. fragments [M]	Nb. WGS fragments [M]	Percentage WGS fragments	Proportion coverage outliers	Purity ploidy ratio	Residual noise	SNR	QA status
200080981-345-S14	28.5	20.2	71.0%	0.0%	0.23	0.13	1.88	**High**
200080982-347-S15	28.7	20.3	70.8%	0.0%	0.25	0.08	3.01	**High**
200080983-348-S16	25.9	18.8	72.8%	0.0%	0.12	0.08	1.71	**High**
200080984-346-S17	24.6	21.7	87.9%	0.0%	0.3	0.12	1.83	**High**

Table 1.1: Sample QA (metrics defined in Sec: 1.2). Metrics leading to a low quality QA status are highlighted in red whereas metrics leading to a medium quality QA status are highlighted in orange. Dashes '-' denote values that the QA algorithm could not compute.

1.4.2 Genomic Integrity

Sample	GI index	QA status	GI status
200080981-345-S14	7.7	**High**	**Positive**
200080982-347-S15	-4.1	**High**	**Negative**
200080983-348-S16	2.8	**High**	**Positive**
200080984-346-S17	-2.1	**High**	**Negative**

Table 1.2: Genomic integrity results (metrics defined in Sec: 1.2)

1.5 Results per sample

1.5.1 Results for sample 200080981-345-S14

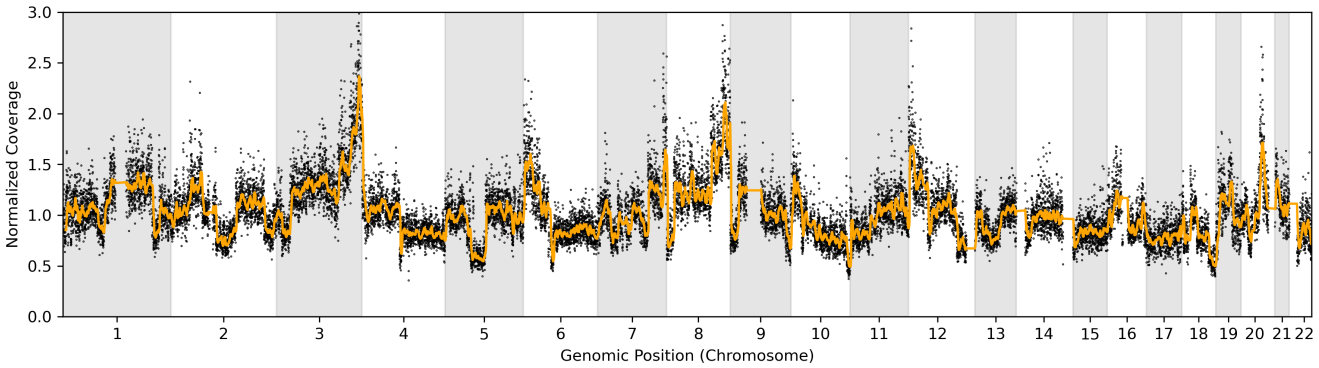

Figure 1.1: Normalized WGS coverage profile (black) and smoothed normalized WGS coverage profile (orange). White and gray shading denotes chromosome boundaries. QA status: High , GI index: 7.7 , GI status: Positive

1.5.2 Results for sample 200080982-347-S15

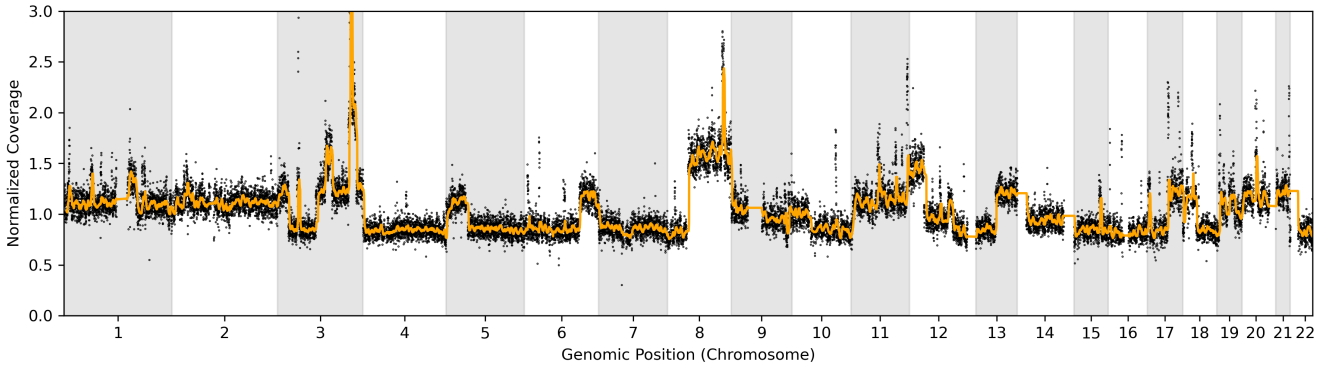

Figure 1.2: Normalized WGS coverage profile (black) and smoothed normalized WGS coverage profile (orange). White and gray shading denotes chromosome boundaries. QA status: High , GI index: -4.1 , GI status: Negative

1.5.3 Results for sample 200080983-348-S16

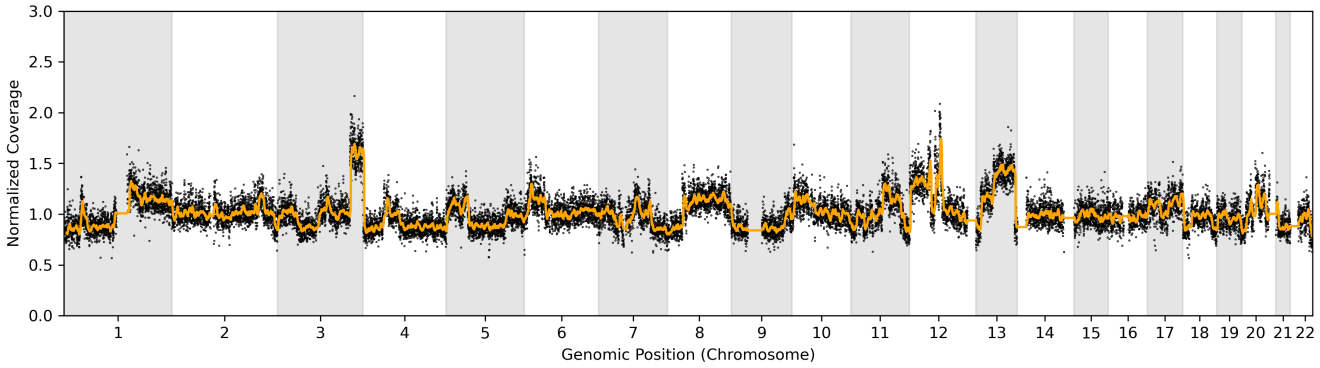

Figure 1.3: Normalized WGS coverage profile (black) and smoothed normalized WGS coverage profile (orange). White and gray shading denotes chromosome boundaries. QA status: High , GI index: 2.8 , GI status: Positive

1.5.4 Results for sample 200080984-346-S17

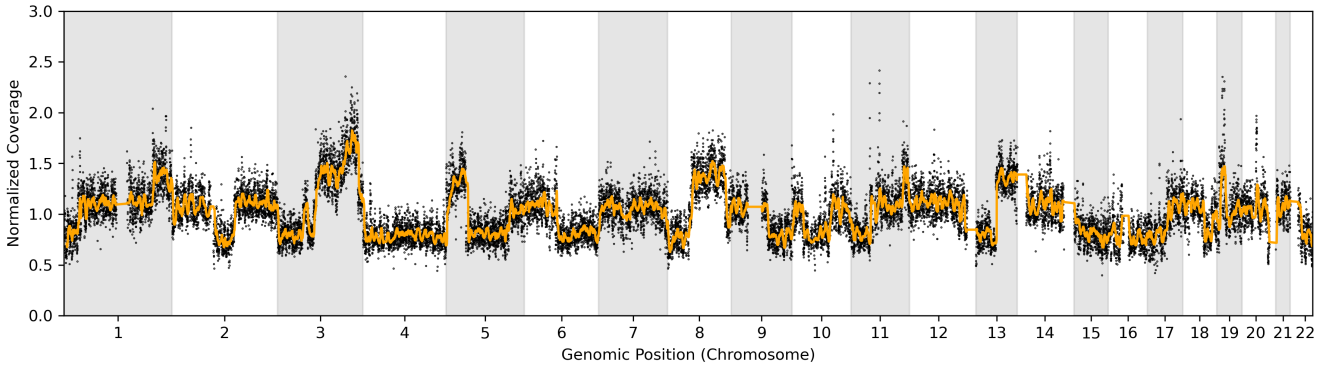

Figure 1.4: Normalized WGS coverage profile (black) and smoothed normalized WGS coverage profile (orange). White and gray shading denotes chromosome boundaries. QA status: High , GI index: -2.1 , GI status: Negative