Bionano EnFocus™ FSHD Analysis JSON File Format Specification Sheet

Document Number: 30322

Document Revision: B

Table of Contents

Legal Notice

For Research Use Only. Not for use in diagnostic procedures.

Revision History

Revision	Notes
A	Initial release of document.
B	Added detail on additional_info fields

Bionano EnFocus™ FSHD Analysis JSON v1.0.1 File Format Specification Sheet

This file format specification sheet details the file format specifications for Bionano EnFocus™ FSHD Analysis JSON (*.json) file version 1.0.1.

Introduction

The Bionano EnFocus™ FSHD Analysis Pipeline generates a JSON file that includes information about the analysis and summarizes the results. JSON (JavaScript Object Notation) is a generic open-standard file format, which relates keys (or attributes) to values. Bionano Genomics has adapted this format to store summary information from the FSHD analysis pipeline. For easy readability, JSON files can be opened in a text editor or specialized JSON viewers.

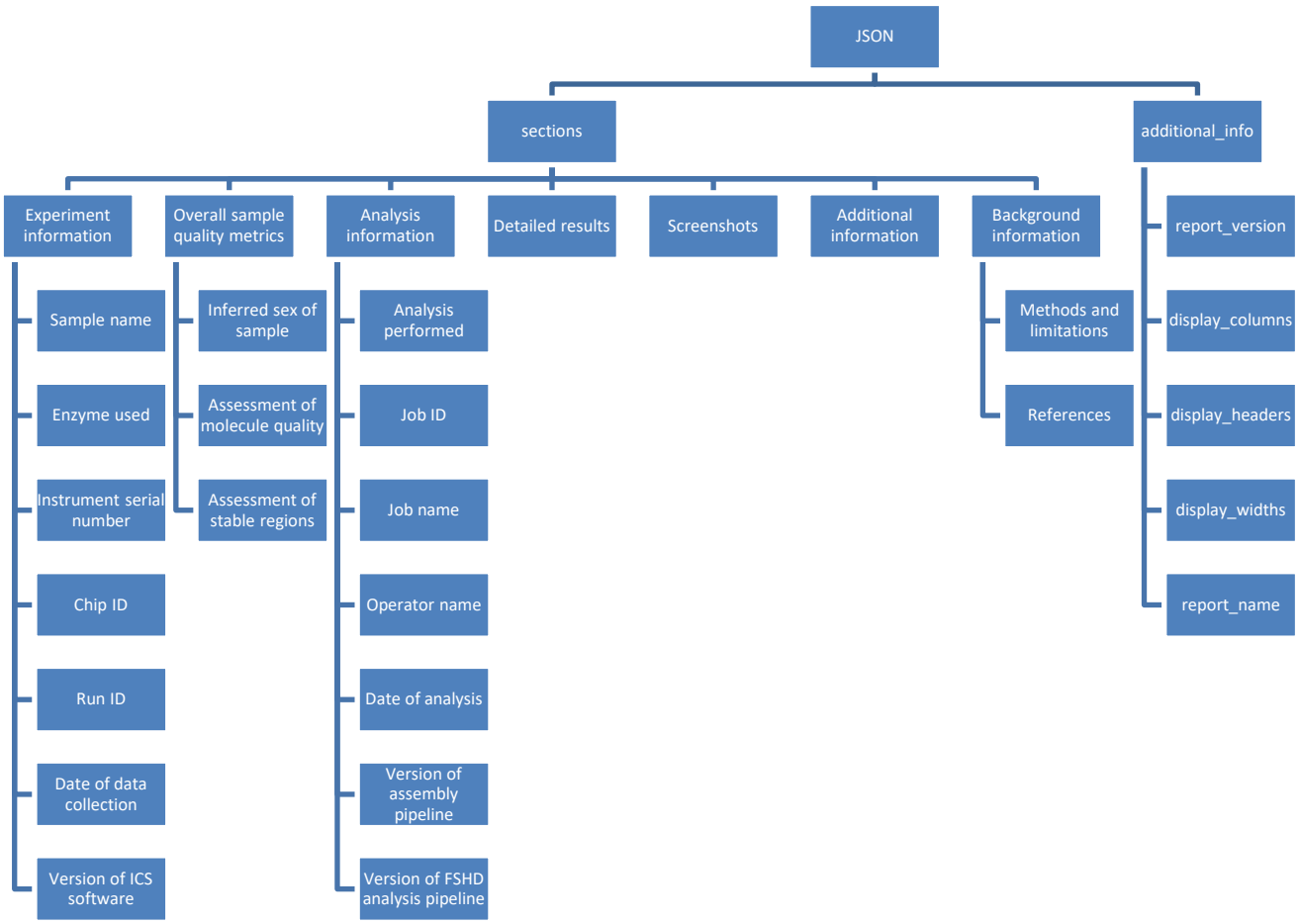
Format

The data are organized in a hierarchy of key-value pairs. The top level has two main sections: “sections” and “additional_info”. The section “sections” contains data that Bionano Access uses for visualization and report generation. The section “additional_info” contains data that Bionano Access uses to generate a PDF report. The report version (from the key report_version) is also contained in this section. The keys are numbered (0, 1, 2, and so forth; see example in “Example JSON Output” section) in order to define the order in which the sections should appear in the PDF report.

The JSON contains the following sections:

- sections
  - Experiment information
    - Sample name
    - Enzyme used
    - Instrument serial number
    - Chip ID
    - Run ID
    - Date of data collection
    - Version of ICS software
  - Overall sample quality metrics
    - Inferred sex of sample
    - Assessment of molecule quality

        - Assessment of stable regions
    - Analysis information
        - Analysis performed
        - Job ID
        - Job name
        - Operator name
        - Date of analysis
        - Version of assembly pipeline
        - Version of FSHD analysis pipeline
    - Detailed results
    - Screenshots
    - Additional information
    - Background information
        - Methods and limitations
        - References
- additional_info
    - report_version
    - display_columns
    - display_headers
    - display_widths
    - report_name



Specifications: "sections"

There are seven sub-sections under "sections": "Experiment information", "Overall sample quality metrics", "Analysis information", "Detailed results", "Screenshots", "Additional information", and "Background information".

The ***"Experiment information"*** section includes information about the extracted and labeled DNA sample ("Sample name" and "Enzyme used"), the map data collection process ("Instrument serial number", "Chip ID", "Run ID", and "Date of data collection"), and the version of the imaging analysis software used to convert the image data into molecule data ("Version of ICS software"). Some of the information is passed into the pipeline by Bionano Access, so they may be absent if the pipeline is run on the command line.

Key	Description	Format	Example
Sample name	Name of the sample; corresponds to "Name" in Bionano Access. Defaulted to <sample_name> if not provided.	string	Sample_1
Enzyme used	Enzyme used to label the DNA; only DLE-1 is supported in Bionano Access	string	DLE-1
Instrument serial number	Serial number of the Bionano Saphyr instrument	string	SAPHYR_A1

Chip ID	Serial number of the chip followed by the flowcell number in parentheses	string	3RSBCYWNPMKXRNWU (Flowcell 2)
Run ID	Unique identifier for a chip run	string	4ba6a250-c593-41fe-b8bf-fd56ecee9e33
Date of data collection	Date and time when the data from the first scan is generated	datetime	2019-07-29 10:20:39 AM
Version of ICS software	Version of the ICS software used for analyzing the image data. Defaulted to "unknown" if unable to get information from input bnx file.	string	ICS 4.8.19085.2

The FSHD analysis pipeline assesses sample quality metrics in order to provide users information about the data quality; the data is summarized in ***"Overall sample quality metrics"***. The metrics and the results are divided into three subsections: "Inferred sex of the sample", "Assessment of molecule quality", and "Assessment of stable regions". For more information, see Bionano Solve Theory of Operation Bionano EnFocus™ FSHD Analysis (PN 30321).

Key	Description	Format	Example
Inferred sex of sample	Sex of the sample as inferred from the copy number analysis pipeline based on the molecule alignment ("coverage") data. "NULL" if data is not available; otherwise, "male" or "female".	string	female
Assessment of molecule quality	Quality of the molecules based on three criteria: molecule N50 (> 150 kbp) has to be at least 200 kbp, effective coverage has to be at least 87.5X, and map rate has to be at least 70%. "NULL" if data is not available; otherwise, "PASS" or "FAIL".	string	PASS
Assessment of stable regions	Quality of the consensus based on evaluation of regions considered stable. "NULL" if data is not available; otherwise, "PASS" or "FAIL".	string	PASS

The ***"Analysis information"*** section includes information about the analysis being performed. Some of the information is passed into the pipeline by Bionano Access, so they may be absent if the pipeline is run on the command line.

Key	Description	Format	Example
Analysis performed	Name of the analysis	string	Bionano EnFocus™ FSHD Analysis
Job ID	Unique Job ID assigned by Bionano Access when the analysis is run. Defaulted to <job_id> if not provided.	string	123456
Job name	Name of the FSHD analysis job when the analysis is run in Bionano Access. Defaults to <object_name> if not provided.	string	Sample_1 DLE1 - FSHD Analysis_Solve3.5_11212019"
Operator name	Name of the user when the analysis is run in Bionano Access. Defaulted to <operator_name> if not provided.	string	John Doe
Date of analysis	The date and time when the FSHD analysis is run	datetime	2019-11-21 15:11
Version of assembly pipeline	Version of the assembly pipeline used for targeted assembly of the regions of interested	string	Bionano Solve 3.5
Version of FSHD analysis pipeline	Version of the Bionano EnFocus™ FSHD Analysis pipeline	string	Bionano EnFocus™ FSHD Analysis 1.0

The ***"Detailed results"*** section contains the necessary data for generating the results table in the PDF output report. The dataframe/table-like data is represented in a list of key-value pairs format. The keys correspond to column names in the table; the values correspond to the cell entries in the table. Each row contains data for a particular map that represents an allele.

The columns of the data are subject to change; the specific columns that are used in report generation are defined in the “additional_info” section as documented below. Selected columns are described below.

Key	Description	Format	Example
MapID	Identifier of a particular map from the assembly	int	22
Chr	Chromosome which the map (referenced in MapID) is from; either 4 or 10	int	4
Haplotype	Haplotype of the allele; 4qA or 4qB if the map is from chr4, and 10qA or 10qB if the map is from chr10. “unknown” if undermined	string	4qA
Count_repeat	Repeat count. If repeat is fully spanned, the pipeline would output an integer value. If not, the pipeline would output a lower bound value (for example >= 20).	int or string	5
Repeat_spanning_coverage	Number of molecules spanning the repeat region	int	30
Start_repeat	Label ID for the repeat start	int	56
End_repeat	Label ID for the repeat end	int	57
Start_haplotype	Position of label in basepairs for the haplotype start	float	600000.0
End_haplotype	Position of label in basepairs for the haplotype end	float	690000.0
Confidence	Deprecated	NA	NA
Map_alignment_confidence	Statistical confidence of alignment: Negative Log10 of p-value of alignment which is the same as the confidence value in XMAP	float	100.0
Anchor_to_mapend_map	Distance from the anchor label to the end of map. The anchor label refers to the label previous to repeat start.	float	10000.0
Array_length	Length of the repeat array interval	float	16.11
Count_length_consistency	Ratio of repeat count between before and after counting shift (See theory of operation for repeat shift)	float	1.01
Contains_SV	Indicates whether the map contains SVs proximal to the D4Z4 region; true/false	bool	true
ImageText	Text to be displayed in PDF report	string	Chromosome 4, Map 22 whose haplotype is 4qA has a calculated repeat count of 5
Count_repeat_mol	Deprecated	NA	NA
Merged	If other redundant maps have the same repeat number	bool	false
truncated_bool	If the map is truncated or not	bool	false
parsed_repeat_counts	Only applicable for the truncated map. Convert string “>= repeat number” to a	int	10

	numeric value (see theory of operation for ">=" sign in the truncated map)		

The ***"Screenshots"*** section indicates where the screenshots (shown in PDF report) should be inserted. It does not contain data.

The ***"Additional information"*** section includes statements indicating whether there may be additional SVs and/or CNVs of interest. The text can vary depending on whether there is presence or absence of such SVs and/or CNVs. The first statement is related to the presence or absence of SVs and/or CNVs proximal to the chr4 D4Z4 region; the second statement is related to the presence or absence of CNVs proximal to the SMCHD1 gene.

The ***"Background information"*** section has two subsections: "Methods and limitations", which briefly describes the methods, and "References", which lists publications that introduce FSHD and its analysis. The same text is shown in Bionano Access when a user sets up the FSHD analysis.

Specifications: "additional_info"

There are five key-value pairs under "additional_info": "report_version", "display_columns", "display_headers", "display_widths", and "report_name". These are used by Bionano Access, and they impact the PDF report generation.

Key	Description	Format	Example
report_version	Version of the FSHD/JSON report	string	1.0.1
display_columns	Columns to be displayed in PDF report	list of string	["Chr", "MapID", "Count_repeat", "Haplotype", "Repeat_spanning_coverage"]
display_headers	Column names to be used in PDF report	list of string	["Chr", "Map ID", "Calculated repeat count (units)", "Haplotype", "Repeat-spanning coverage (X)"]
display_width	Column widths to be used in PDF report	list of int	[35, 40, 80, 60, 80]
report_name	Report name to be used in PDF report	string	Bionano EnFocus™ FSHD Analysis Report

Example JSON output


```
{
  "sections": {
    "0": {
      "Experiment information": {
        "0": {
          "Sample name": "Sample_1"
        },
```

```
      "1": {
        "Enzyme used": "DLE-1"
      },
      "2": {
        "Instrument serial number": "SAPHYR_A1"
      },
      "3": {
        "Chip ID": "3RSBCYWNPMKXRNWU (Flowcell 2)"
      },
      "4": {
        "Run ID": "4ba6a250-c593-41fe-b8bf-fd56ecee9e33"
      },
      "5": {
        "Date of data collection": "2019-07-29 10:20:39 AM"
      },
      "6": {
        "Version of ICS software": "ICS 4.8.19085.2"
      }
    }
  },
  "1": {
    "Overall sample quality metrics": {
      "0": {
        "Inferred sex of sample": "male"
      },
      "1": {
        "Assessment of molecule quality": "PASS"
      },
      "2": {
        "Assessment of stable regions": "PASS"
      }
    }
  },
  "2": {
    "Analysis information": {
      "0": {
        "Analysis performed": "Bionano EnFocus™ FSHD Analysis"
      },
      "1": {
        "Job ID": 123456
      },
      "2": {
        "Job name": "Sample_1 DLE1 - FSHD Analysis_Solve3.5_11212019"
      },
      "3": {
```

```
      "Operator name": "John Doe"
    },
    "4": {
      "Date of analysis": "2019-11-21 15:11"
    },
    "5": {
      "Version of assembly pipeline": "Bionano Solve 3.5"
    },
    "6": {
      "Version of FSHD analysis pipeline": "Bionano EnFocus™ FSHD Analysis 1.0"
    }
  }
},
"3": {
  "Detailed results": [
    {
      "MapID": 22,
      "Chr": 4,
      "Haplotype": "4qA",
      "Count_repeat": 5,
      "Repeat_spanning_coverage": 27,
      "Start_repeat": 110,
      "End_repeat": 111,
      "Start_haplotype": 769862.1,
      "End_haplotype": 787920.9,
      "Confidence": -1,
      "Map_alignment_confidence": 124.07,
      "Anchor_to_mapend_map": 32504,
      "Array_length": 16.11,
      "Count_length_consistency": 0.98,
      "Contains_SV": true,
      "ImageText": "Chromosome 4, Map 22 whose haplotype is 4qA has a calculated repeat count of 5",
      "Count_repeat_mol": -1,
      "Merged": false,
      "truncated_bool": false,
      "parsed_repeat_counts": -1
    },
    {
      "MapID": 290,
      "Chr": 4,
      "Haplotype": "4qB",
      "Count_repeat": 17,
      "Repeat_spanning_coverage": 23,
      "Start_repeat": 51,
      "End_repeat": 52,
```

```
      "Start_haplotype": 361995.5,
      "End_haplotype": 388109.3,
      "Confidence": -1,
      "Map_alignment_confidence": 54.83,
      "Anchor_to_mapend_map": 65341.09999999998,
      "Array_length": 56.85,
      "Count_length_consistency": 1.01,
      "Contains_SV": true,
      "ImageText": "Chromosome 4, Map 290 whose haplotype is 4qB has a calculated repeat count of
17",
      "Count_repeat_mol": -1,
      "Merged": false,
      "truncated_bool": false,
      "parsed_repeat_counts": -1
    },
    {
      "MapID": 11,
      "Chr": 10,
      "Haplotype": "10qA",
      "Count_repeat": 6,
      "Repeat_spanning_coverage": 43,
      "Start_repeat": 1017,
      "End_repeat": 1018,
      "Start_haplotype": 7713036.3,
      "End_haplotype": 7731665.1,
      "Confidence": -1,
      "Map_alignment_confidence": 1208.81,
      "Anchor_to_mapend_map": 35881.89999999944,
      "Array_length": 20.04,
      "Count_length_consistency": 1.01,
      "Contains_SV": true,
      "ImageText": "Chromosome 10, Map 11 whose haplotype is 10qA has a calculated repeat count of
6",
      "Count_repeat_mol": -1,
      "Merged": true,
      "truncated_bool": false,
      "parsed_repeat_counts": -1
    },
    {
      "MapID": 260,
      "Chr": 10,
      "Haplotype": "10qA",
      "Count_repeat": 15,
      "Repeat_spanning_coverage": 25,
      "Start_repeat": 40,
```

```
        "End_repeat": 41,
        "Start_haplotype": 417067.2,
        "End_haplotype": 435718.9,
        "Confidence": -1,
        "Map_alignment_confidence": 51.28,
        "Anchor_to_mapend_map": 66380.79999999999,
        "Array_length": 50.54,
        "Count_length_consistency": 1.02,
        "Contains_SV": true,
        "ImageText": "Chromosome 10, Map 260 whose haplotype is 10qA has a calculated repeat count of 15",
        "Count_repeat_mol": -1,
        "Merged": false,
        "truncated_bool": false,
        "parsed_repeat_counts": -1
      }
    ]
  },
  "4": {
    "Screenshots": "Screenshots to be inserted here"
  },
  "5": {
    "Additional information": "Structural variants and other copy number variants were detected in the proximal chr4 region. No copy number variants were detected proximal to SMCHD1."
  },
  "6": {
    "Background information": {
      "0": {
        "Methods and limitations": "The Bionano EnFocus™ FSHD Analysis is performed based on whole-genome optical mapping data collected on the Bionano Genomics Saphyr Genome Imaging Instrument. Based on specific labeling and mapping of ultra-high molecular weight DNA in nanochannel arrays, optical mapping enables high-resolution analysis of the D4Z4 repeat array.\n\nMolecules aligning to regions of interest in chr4 and chr10 are extracted and assembled. The resulting consensus maps are used for the Bionano EnFocus™ FSHD Analysis. The repeat arrays are sized, and the permissive and non-permissive alleles (4qA and 4qB) assigned. Additional structural variants and copy number gains and losses are noted in the proximity of the D4Z4 repeat array on chr4. Copy number gains and losses in the proximity of the SMCHD1 gene on chr18 are also noted.\n\nThe analysis data can be imported into Bionano Access, a graphical user interface tool for visualization and curation. This method cannot detect single-nucleotide variants that do not impact sequence motif sites and may miss small variants with potential functional impacts."
      },
      "1": {
        "References": "Wijmenga et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature Genetics (1992).\nDeidda et al. Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). J Med Genetics (1996).\nZhang et al. Clinical application of single-molecule optical mapping to a multigeneration FSHD1 pedigree. Molecular Genetics and Genomic Medicine (2019)."
      }
```

```
      }
     }
    },
    "additional_info": {
     "0": {
      "report_version": "1.0.1"
     },
     "1": {
      "display_columns": [
       "Chr",
       "MapID",
       "Count_repeat",
       "Haplotype",
       "Repeat_spanning_coverage"
      ]
     },
     "2": {
      "display_headers": [
       "Chr",
       "Map ID",
       "Calculated repeat count (units)",
       "Haplotype",
       "Repeat-spanning coverage (X)"
      ]
     },
     "3": {
      "display_widths": [
       35,
       40,
       80,
       60,
       80
      ]
     },
     "4": {
      "report_name": "Bionano EnFocus™ FSHD Analysis Report"
     }
    }

}
```

Technical Assistance

For technical assistance, contact Bionano Genomics Technical Support.

You can retrieve documentation on Bionano products, SDS's, certificates of analysis, frequently asked questions, and other related documents from the Support website or by request through e-mail and telephone.

Type	Contact
Email	**support@bionanogenomics.com**
Phone	**Hours of Operation:** **Monday through Friday, 9:00 a.m. to 5:00 p.m., PST** **US: +1 (858) 888-7663**
Website	**www.bionanogenomics.com/support**

Bionano Genomics, Inc.
9540 Towne Centre Drive, Suite 100
San Diego, CA 92121