Bionano EnFocus™ Fragile X Analysis JSON File Format Specification Sheet

Document Number: 30458

Document Revision: A

Table of Contents

Legal Notice

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

Revision History

Revision	Notes
A	Initial release of document.

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

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

Introduction

The Bionano EnFocus™ Fragile X 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 Fragile X analysis pipeline. For easy readability, JSON files can be opened in a text editor or specialized JSON viewers.

Format

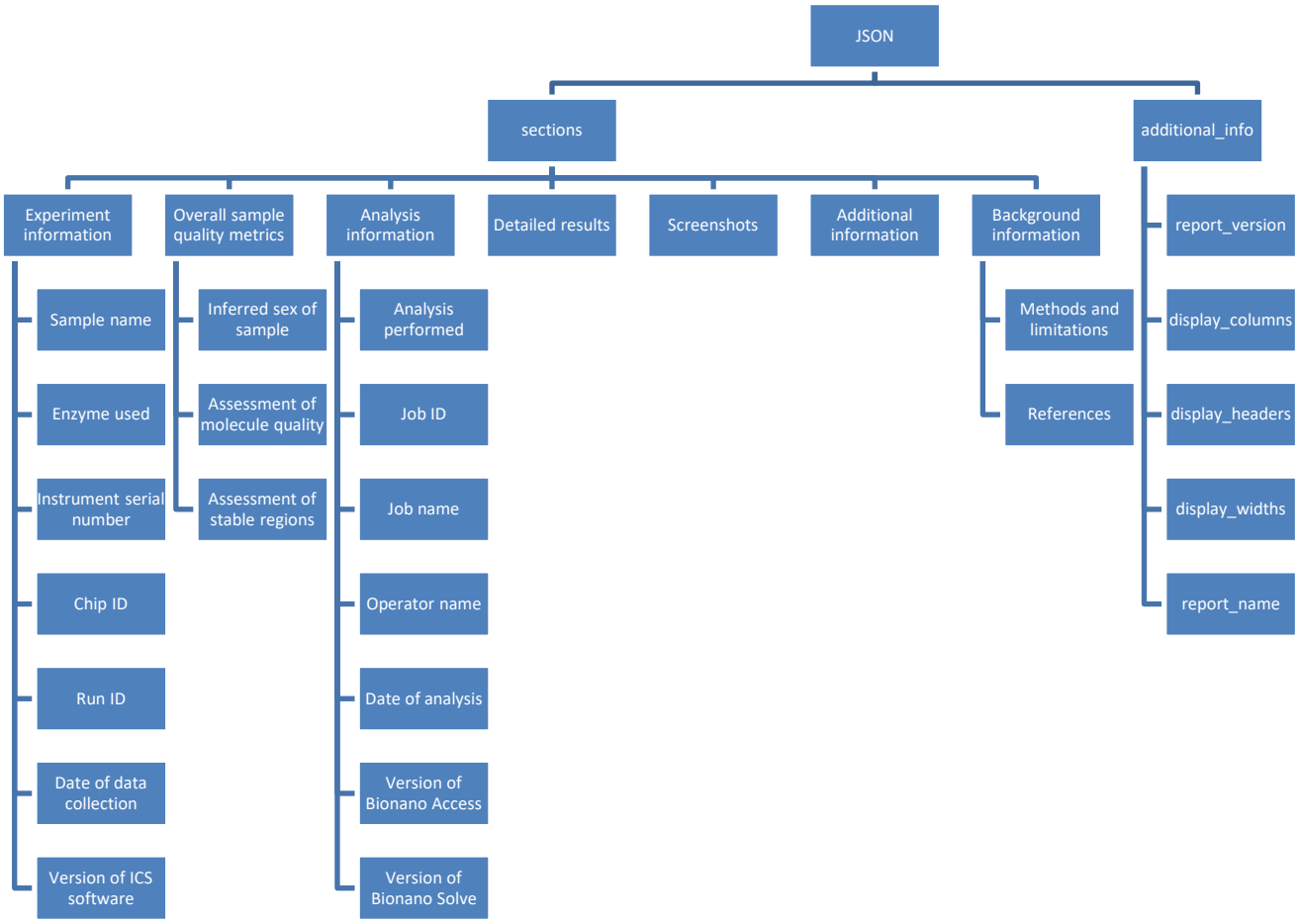
The data are organized in a hierarchy of key-value pairs. The top level has two main sections: “sections” and “additional_info”. The section “sections” contains data that Bionano Access uses for visualization and report generation. The section “additional_info” contains information that Bionano Access 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 Bionano Access
  - Version of Bionano Solve
- 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 Fragile X 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™ Fragile X Analysis (PN 30457).

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". Detailed information on the assessment of stable regions can be found in Bonano Solve Theory of Operation: Bionano EnFocus™ Fragile X Analysis (PN 30457)	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™ Fragile X 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 Fragile X analysis job when the analysis is run in Bionano Access. Defaulted to <object_name> if not provided.	string	Sample_1 DLE1 – FragileX Analysis_Solve3.7_09012021
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	2021-09-14 22:46
Version of Bionano Access	Version of Access	string	1.7
Version of Bionano Solve	Version of bioinformatics tools	string	Bionano Solve 3.7

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
Gene	Identifier of repeat gene	string	FMR1
Sample	Sample name	string	95552_6
Sex	Sex of sample	string	female
Chr	Chromosome where the repeat gene is located	string	X
Start_ref	Position of label in base pairs for the start of the interval of interest in the reference	int	147910189
End_ref	Position of label in base pairs for the end of the interval of interest in the reference	int	147918814
Interval_ref	Length of the interval of interest in the reference	int	8625
Count_repeat_ref	Repeat count in the reference	int	25
Repeat_unit_size	Repeat unit size	int	3
Irrelevant_ref	Length of the flanking irrelevant region in the reference	int	8550
MapID	Identifier of a particular map from the assembly	int	231
Start_repeat	Label ID for the repeat start	int	253
End_repeat	Label ID for the repeat end	int	254
I_2_start_ref	To correct for the irrelevant region, extra space needs to be added from the start of the observed interval (See FAQs in theory of operation for flanking region correction)	float	500.0
I_2_end_ref	To correct for the irrelevant region, extra space needs to be added from the end of the observed interval	float	500.0
Array_length	Length of the observed interval in kilo-base pairs	float	8.92
Unmatched_labels	Number of unmatched labels within the observed interval	int	1
Count_repeat_observed	Repeat count estimated by (observed interval – irrelevant_ref)/Repeat_unit_size	int	200
Count_repeat	Estimated repeat count with the maximum posterior probability	int	200
P >= expansion_cutoff	Probability that sample repeat number is greater than or equal to 200 units	string	99%

Expanded_repeat	Repeat cutoff for the full expansion	int	200
Realigned	If boundary labels are realigned	bool	False
CI_lower	Lower bound of repeat count for 99% credible interval	int	100
CI_upper	Upper bound of repeat count for 99% credible interval	int	300
Percentile	Percentage of negative control samples which have repeat number lower than the estimated repeat count	int	50
Repeat_spanning_coverage	Number of the molecules spanning the repeat region	int	30
Qry_contig_length	Total length of the consensus map	float	100000.0
ImageText	Text to be displayed in PDF report	string	Chromosome X, Map231 has a calculated repeat count of 548

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

The ***"Background information"*** section has two subsections: "Methods and limitations", which briefly describes the methods, and "References", which lists publications that introduce Fragile X and its analysis. The same text is shown in Bionano Access when a user sets up the Fragile X 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 Fragile X/JSON report	string	1.0.1
display_columns	Columns to be displayed in PDF report	list of string	[“Gene”,”Sample”, “Chr”, “Count_repeat”, “P >= expansion_cutoff”, “CI_lower”, “CI_upper”, “Repeat_spanning_coverage”]
display_headers	Column names to be used in PDF report	list of string	[“Gene”, “Sample”, “Chr”, “Calculated repeat count”, “Probability >= 200 repeat units”, “99% credible interval lower bound”, “99% credible interval upper bound”, “Repeat-spanning coverage (X)”]
display_width	Column widths to be used in PDF report	list of int	[40, 80, 25, 55, 55, 50, 50, 50]
report_name	Report name to be used in PDF report	string	Bionano EnFocus™ Fragile Analysis Report

Example JSON output


```
{
  "sections": {
    "0": {
      "Experiment information": {
        "0": {
          "Sample name": "95552_6"
        },
        "1": {
          "Enzyme used": "DLE1"
        },
        "2": {
          "Instrument serial number": "SAPHYR_D08"
        },
        "3": {
          "Chip ID": "C7B2OJGNPPRSJNWU (Flowcell 3)"
        },
        "4": {
          "Run ID": "be595a62-82d3-4f8a-a95b-f097f720ba7b"
        },
        "5": {
          "Date of data collection": "2021-04-14 11:16:50 PM"
        },
        "6": {
          "Version of ICS software": "ICS 5.1.21018.2"
        }
      }
    },
    "1": {
      "Overall sample quality metrics": {
        "0": {
          "Inferred sex of sample": "female"
        },
        "1": {
          "Assessment of molecule quality": "PASS"
        },
        "2": {
          "Assessment of stable regions": "PASS"
        }
      }
    },
    "2": {
      "Analysis information": {
        "0": {
          "Analysis performed": "Bionano EnFocusTM Fragile X Analysis"
        },
        "1": {
          "Job ID": "123456"
        },
        "2": {
          "Job name": "Sample_1 DLE1 – FragileX Analysis_Solve3.7_09012021"
        },
        "3": {
          "Operator name": "John Doe"
        },
```

```
    "4": {
      "Date of analysis": "2021-09-14 22:46"
    },
    "5": {
      "Version of Bionano Access": "1.7"
    },
    "6": {
      "Version of Bionano Solve": "Bionano Solve 3.7"
    }
   }
  },
  "3": {
   "Detailed results": [
    {
     "Gene": "FMR1",
     "Sample": "95552_6",
     "Sex": "female",
     "Chr": "X",
     "Start_ref": 147910189,
     "End_ref": 147918814,
     "Interval_ref": 8625,
     "Count_repeat_ref": 25,
     "Repeat_unit_size": 3,
     "Irrelevant_ref": 8550,
     "MapID": 231,
     "Start_repeat": 213,
     "End_repeat": 214,
     "I_2_start_ref": -1,
     "I_2_end_ref": -1,
     "Array_length": 8.92,
     "Unmatched_labels": 0,
     "Count_repeat_observed": 122,
     "Count_repeat": 122,
     "P >= expansion_cutoff": "0.08%",
     "Expanded_repeat": 200,
     "Realigned": false,
     "CI_lower": 63,
     "CI_upper": 184,
     "Percentile": 99,
     "Repeat_spanning_coverage": 50,
     "Qry_contig_length": 2025503.5,
     "ImageText": "Chromosome X, Map231 has a calculated repeat count of 122"
    },
    {
     "Gene": "FMR1",
     "Sample": "95552_6",
     "Sex": "female",
     "Chr": "X",
     "Start_ref": 147910189,
     "End_ref": 147918814,
     "Interval_ref": 8625,
     "Count_repeat_ref": 25,
     "Repeat_unit_size": 3,
     "Irrelevant_ref": 8550,
     "MapID": 232,
     "Start_repeat": 214,
     "End_repeat": 215,
```

```
        "I_2_start_ref": -1,
        "I_2_end_ref": -1,
        "Array_length": 8.92,
        "Unmatched_labels": 0,
        "Count_repeat_observed": 122,
        "Count_repeat": 122,
        "P >= expansion_cutoff": "0.08%",
        "Expanded_repeat": 200,
        "Realigned": false,
        "CI_lower": 63,
        "CI_upper": 184,
        "Percentile": 99,
        "Repeat_spanning_coverage": 49,
        "Qry_contig_length": 2024381,
        "ImageText": "Chromosome X, Map232 has a calculated repeat count of 122"
      }
    ]
  },
  "4": {
    "Screenshots": "Screenshots to be inserted here"
  },
  "5": {
    "Background information": {
      "0": {
        "Methods and limitations": "The Bionano EnFocusTM Fragile X Syndrome analysis is performed based on optical genome mapping (OGM) data collected on the Bionano Genomics Saphyr genome imaging instrument. Based on specific labeling, alignment, and assembly of ultra-long DNA molecules in nanochannel arrays, OGM enables for assessment of CGG expansions in the FMR1 gene locus.\n\nMolecules aligning to the region of interest on chrX are extracted and assembled. The resulting consensus maps are used for the Bionano EnFocusTM Fragile X syndrome analysis. The size of the CGG repeat array in the FMR1 gene is inferred based on the measured distance between two neighboring labels on the assembled map(s) that contain the FMR1 gene. By incorporating a known set of control map measurements, the pipeline estimates the most likely CGG repeat count as well as the credible intervals for the uncertainty of the repeat count. We also compute the probability that the repeat count exceeds the pathogenic threshold of >200 repeat unit. The expanded CGG segment silences the FMR1 gene expression, which in turn disrupts nervous system functions leading to learning and cognitive impairment seen in Fragile X syndrome.\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. Because it is impossible to exclude other sequence insertions within the expansion interval or other repeat expansion (i.e. containing AGG interruptions), repeat expansion is inferred and any increase in length is assumed to be CGG expansion. \n\nOptical Genome mapping is intended for research use only; it is not a diagnostic test.  "
      },
      "1": {
        "References": "Hunter JE, Berry-Kravis E, Hipp H, Todd PK. FMR1 Disorders. 1998 Jun 16 [updated 2019 Nov 21]. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2021. Available from http://www.ncbi.nlm.nih.gov/books/NBK1384/\nSahajpal, N. et. al. Optical Genome Mapping as a Next-Generation Cytogenomic Tool for Detection of Structural and Copy Number Variations for Prenatal Genomic Analyses. Genes (Basel). 2021 Mar; 12(3): 398."
      }
    }
  }
 },
 "additional_info": {
  "0": {
    "report_version": "1.0.1"
  },
```

```
    "1": {
      "display_columns": [
        "Gene",
        "Sample",
        "Chr",
        "Count_repeat",
        "P >= expansion_cutoff",
        "CI_lower",
        "CI_upper",
        "Repeat_spanning_coverage"
      ]
    },
    "2": {
      "display_headers": [
        "Gene",
        "Sample",
        "Chr",
        "Calculated repeat count",
        "Probability >= 200 repeat units",
        "99% credible interval lower bound",
        "99% credible interval upper bound",
        "Repeat-spanning coverage (X)"
      ]
    },
    "3": {
      "display_widths": [
        40,
        80,
        25,
        55,
        55,
        50,
        50,
        50
      ]
    },
    "4": {
      "report_name": "Bionano EnFocusTM Fragile X 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