FISH Omics Format

4DN-DCIC

Jan 31, 2023

FOF-CT: CHROMATIN TRACING

CHAPTER

ONE

INTRODUCTION

A key output of the 4D Nucleome (4DN) project is the open publication of datasets related to the structure of the human cell nucleus and the genome, within. Recent years have seen a rapid expansion of FISH-omics methods, which quantify the spatial organization of DNA, RNA and protein in the cell and provide expanded understanding of how higher-order chromosome structure relates to transcriptional activity and cell development. Despite this progress, FISH-based image-data are not yet routinely made publicly available upon publication because of the lack of common specifications for data exchange. This challenge is experienced across the bioimaging community, as a result a solution built, tested and proven in 4DN can have a wide impact all over the world.

This document describes the **4DN FISH Omics Format - Chromatin Tracing (FOF-CT)**, a community data format designed for capturing and exchanging the results of chromosome imaging experiments produced within the context of the 4D Nucleome project. FOF-CT is directly compatible with several FISH omics techniques including, but not limited to, Optical Reconstruction of Chromatin Architecture (ORCA), Multiplexed Imaging of Nucleome Architectures (MINA), Hi-M, DNA Sequential Fluorescence In Situ Hybridization (seqFISH+), Oligonucleotide Fluorescent In Situ Sequencing (OligoFISSEQ), DNA Multiplexed error-robust fluorescence *in situ* hybridization (DNA-MERFISH), and *In-situ* Genomic Sequencing (IGS). In addition, the format is designed to be consistent with planned future extensions that will encompass single-molecule localization methods for volumetric imaging, such as OligoSTORM and OligoDNA-PAINT.

In chromatin tracing experiments, polymer tracing algorithms are used to string together the localization of individual DNA bright Spots to reconstruct the three-dimensional (3D) path of chromatin fibers. Thus, the format is organized around multiple tables. The core of the format consists of a Spot/Trace table that defines chromatin Traces as ensembles of individual DNA-FISH bright Spot localizations.

Additional tables support the integration of this core with additional properties such as quality metrics, physical coordinates placing the Spot/Trace in the context of cellular space, multiplexed RNA-FISH results and with additional data that is better captured at the global Trace (e.g., expression level of nascent RNA transcripts associated with a given Trace or overall localization of the Trace with respect to cellular or nuclear landmarks), Cell (e.g., boundaries and volume), sub-cellular Region of Interest (ROI; e.g., Nuclear feature or Nucleolus), or extracellular ROI (e.g., Tissue) level.

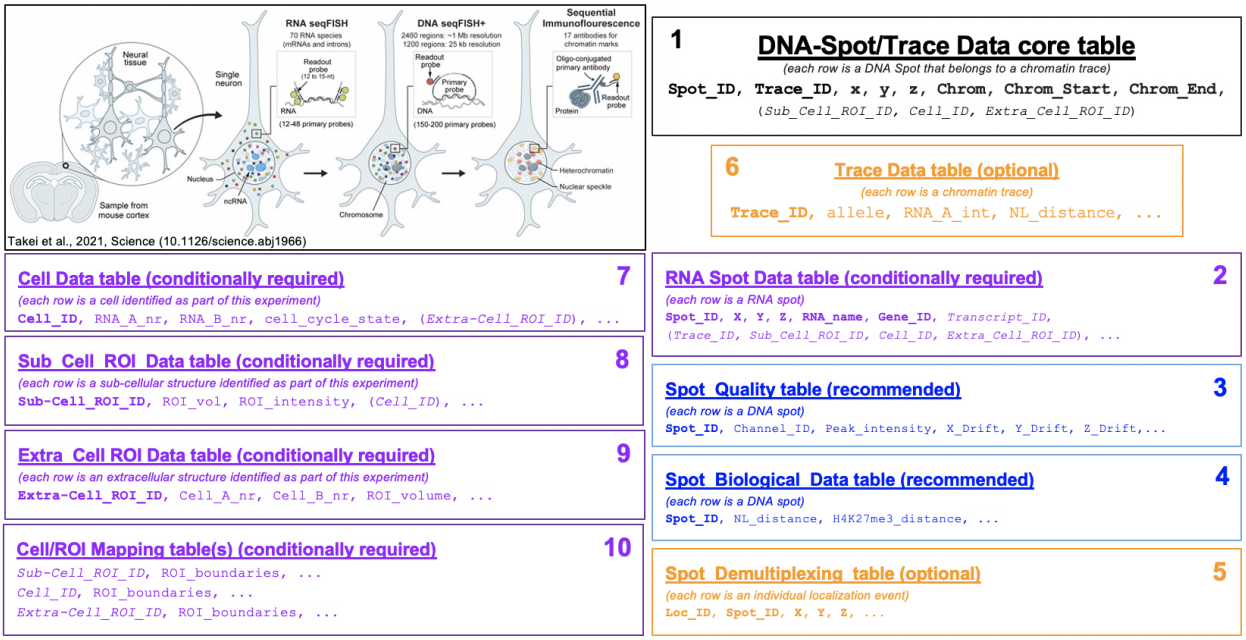


Fig. 1: Figure 1: Schematic representation of 10 tables composing the Fish Omics Format for Chromatin Tracing.

CHAPTER

TWO

TABLES

Number	Extended Name	Short Name	Namespace	Requirement Level
1	*DNA-Spot/Trace Data core table*	core	4dn_FOF-CT_core	required
2	*RNA-Spot Data table*	rna	4dn_FOF-CT_rna	conditionally required
3	*Spot Quality table*	quality	4dn_FOF-CT_quality	recommended
4	*Spot Biological Data table*	bio	4dn_FOF-CT_bio	recommended
5	*Spot Demultiplexing table*	demultiplexing	4dn_FOF-CT_demultiplexing	optional
6	*Trace Data table*	trace	4dn_FOF-CT_trace	optional
7	*Cell Data table*	cell	4dn_FOF-CT_cell	conditionally required
8	*Sub-Cell ROI Data table*	subcell	4dn_FOF-CT_subcell	conditionally required
9	*Extra-Cell ROI Data table*	extracell	4dn_FOF-CT_extracell	conditionally required
10	*Cell/ROI Mapping table*	mapping	4dn_FOF-CT_mapping	conditionally required

2.1 Introduction

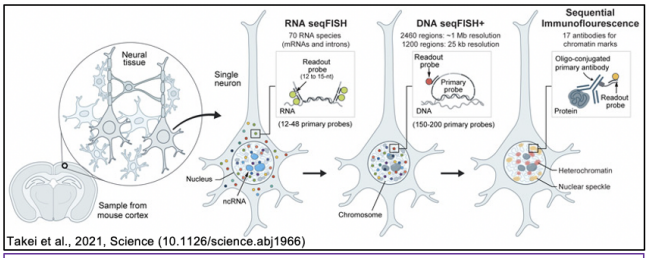
A key output of the 4D Nucleome (4DN) project is the open publication of datasets related to the structure of the human cell nucleus and the genome, within. Recent years have seen a rapid expansion of FISH-omics methods, which quantify the spatial organization of DNA, RNA and protein in the cell and provide expanded understanding of how higher-order chromosome structure relates to transcriptional activity and cell development. Despite this progress, FISH-based image-data are not yet routinely made publicly available upon publication because of the lack of common specifications for data exchange. This challenge is experienced across the bioimaging community, as a result a solution built, tested and proven in 4DN can have a wide impact all over the world.

This document describes the **4DN FISH Omics Format - Chromatin Tracing (FOF-CT)**, a community data format designed for capturing and exchanging the results of chromosome imaging experiments produced within the context of the 4D Nucleome project. FOF-CT is directly compatible with several FISH omics techniques including, but not limited to, Optical Reconstruction of Chromatin Architecture (ORCA), Multiplexed Imaging of Nucleome Architectures (MINA), Hi-M, DNA Sequential Fluorescence In Situ Hybridization (seqFISH+), Oligonucleotide Fluorescent In Situ Sequencing (OligoFISSEQ), DNA Multiplexed error-robust fluorescence *in situ* hybridization (DNA-MERFISH), and *In-situ* Genomic Sequencing (IGS). In addition, the format is designed to be consistent with planned future extensions that will encompass single-molecule localization methods for volumetric imaging, such as OligoSTORM and

OligoDNA-PAINT.

In chromatin tracing experiments, polymer tracing algorithms are used to string together the localization of individual DNA bright Spots to reconstruct the three-dimensional (3D) path of chromatin fibers. Thus, the format is organized around multiple tables. The core of the format consists of a Spot/Trace table that defines chromatin Traces as ensembles of individual DNA-FISH bright Spot localizations.

Additional tables support the integration of this core with additional properties such as quality metrics, physical coordinates placing the Spot/Trace in the context of cellular space, multiplexed RNA-FISH results and with additional data that is better captured at the global Trace (e.g., expression level of nascent RNA transcripts associated with a given Trace or overall localization of the Trace with respect to cellular or nuclear landmarks), Cell (e.g., boundaries and volume), sub-cellular Region of Interest (ROI; e.g., Nuclear feature or Nucleolus), or extracellular ROI (e.g., Tissue) level.



1 DNA-Spot/Trace Data core table
(each row is a DNA Spot that belongs to a chromatin trace)
`Spot_ID, Trace_ID, x, y, z, Chrom, Chrom_Start, Chrom_End,`
`(Sub_Cell_ROI_ID, Cell_ID, Extra_Cell_ROI_ID)`

6 Trace Data table (optional)
(each row is a chromatin trace)
`Trace_ID, allele, RNA_A_int, NL_distance, ...`

Cell Data table (conditionally required) — 7
(each row is a cell identified as part of this experiment)
`Cell_ID, RNA_A_nr, RNA_B_nr, cell_cycle_state, (Extra-Cell_ROI_ID), ...`

RNA Spot Data table (conditionally required) — 2
(each row is a RNA spot)
`Spot_ID, X, Y, Z, RNA_name, Gene_ID, Transcript_ID,`
`(Trace_ID, Sub_Cell_ROI_ID, Cell_ID, Extra_Cell_ROI_ID), ...`

Sub_Cell_ROI_Data table (conditionally required) — 8
(each row is a sub-cellular structure identified as part of this experiment)
`Sub-Cell_ROI_ID, ROI_vol, ROI_intensity, (Cell_ID), ...`

Spot_Quality table (recommended) — 3
(each row is a DNA spot)
`Spot_ID, Channel_ID, Peak_intensity, X_Drift, Y_Drift, Z_Drift,...`

Extra_Cell ROI Data table (conditionally required) — 9
(each row is an extracellular structure identified as part of this experiment)
`Extra-Cell_ROI_ID, Cell_A_nr, Cell_B_nr, ROI_volume, ...`

Spot_Biological_Data table (recommended) — 4
(each row is a DNA spot)
`Spot_ID, NL_distance, H4K27me3_distance, ...`

Cell/ROI Mapping table(s) (conditionally required) — 10
`Sub-Cell_ROI_ID, ROI_boundaries, ...`
`Cell_ID, ROI_boundaries, ...`
`Extra-Cell_ROI_ID, ROI_boundaries, ...`

Spot_Demultiplexing_table (optional) — 5
(each row is an individual localization event)
`Loc_ID, Spot_ID, X, Y, Z, ...`

Fig. 1: Figure 1: Schematic representation of 10 tables composing the Fish Omics Format for Chromatin Tracing.

2.2 Tables

Number	Extended Name	Short Name	Namespace	Requirement Level
1	*DNA-Spot/Trace Data core table*	core	4dn_FOF-CT_core	required
2	*RNA-Spot Data table*	rna	4dn_FOF-CT_rna	conditionally required
3	*Spot Quality table*	quality	4dn_FOF-CT_quality	recommended
4	*Spot Biological Data table*	bio	4dn_FOF-CT_bio	recommended
5	*Spot Demultiplexing table*	demultiplexing	4dn_FOF-CT_demultiplexing	optional
6	*Trace Data table*	trace	4dn_FOF-CT_trace	optional
7	*Cell Data table*	cell	4dn_FOF-CT_cell	conditionally required
8	*Sub-Cell ROI Data table*	subcell	4dn_FOF-CT_subcell	conditionally required
9	*Extra-Cell ROI Data table*	extracell	4dn_FOF-CT_extracell	conditionally required
10	*Cell/ROI Mapping table*	mapping	4dn_FOF-CT_mapping	conditionally required

2.3 Format description: overview

Contents

2.3.1 General Info

- The format is organized in multiple individual tables.
- The only mandatory table is the *DNA-Spot/Trace Data core table*.
- All other tables are either recommended for all experiment types, or optional depending on the experiment design and type.
- Each file must contain a single table.
- Accepted file formats for storing Tables are txt, csv and tsv.

- An underscore must be used as a word separator in header field names and column headers to improve readability while not violating common name restrictions in coding environments (dash - may be mistaken as subtraction of variables).
- Each file has two parts: file header and data columns.

2.3.2 File Header

- In the file header, each line contains only one field.
- Header lines are denoted by #. In particular:
  - ## denotes machine readable header lines. These lines must follow the following format `##Key1=Value1` (e.g., `##FOF-CT_version=v0.1`).
  - # denotes human readable header lines. These lines should follow the following format, `#term: free text description` (e.g., `#lab_name: name of the lab where the experiment was performed`).
  - #^ denotes lines that define optional user specified columns. These lines provide the name of the column header and a description of the column content. Descriptions must be understandable and sufficient to ensure the interpretation and reproducibility of the results. These lines should follow the following format `#^term: free text description` (e.g., `#^optional_column_1: optional column 1 description`).
- Header names must use the underscore as a word separator (e.g., RNA_A_intensity).
- The file header contains **required**, *conditionally-required*, and optional fields.
- Conditionally-required fields are fields that are required when certains conditions are met (e.g., *##intensity_unit=* is required any time an intensity metric is reported).
- All tables have to contain a mandatory header section.

Mandatory header lines (all tables)

##FOF-CT_version= Data format version number. E.g. v0.2

##XYZ_unit= The unit used to represent the XYZ location of bright Spots in this table. Note: use micron (instead of µm) to avoid problems with special, Greek symbols. Other allowed values are: nm, mm etc.

#lab_name: name of the lab where the experiment was performed

#experimenter_name: name of the person performing the experiment

#experimenter_contact: email address of the person performing the experiment

#description: A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.

#additional_tables: AddTable1, AddTable2, AddTableN

##columns=(C1, C2, C3, Cn)

Additional mandatory header lines (DNA spot/trace core and RNA tables)

In addition to all of the above,

##genome_assembly= Genome build. Note that the 4DN data portal only accepts GRCh38 for human and GRCm38 for mouse.

#Software_Title: The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.

#Software_Type: The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Other

#Software_Authors: The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows: Doe, John; Smith, Jane; etc,.

#Software_Description: A free-text description of this Software. This description should provide a detailed understanding of the algorithm and of the analysis parameters that were used, in order to guarantee interpretation and reproducibility.

#Software_Repository: The URL of any repository or archive where the Software executable release can be obtained.

#Software_PreferredCitationID: The Unique Identifier for the preferred/primary publication describing this Software. Examples include Digital Object Identifier (DOI), PubMed Central Identifier (PMCID), ArXiv.org ID etc,.

2.3.3 Data Columns

- Tables contain **required**, *conditionally-required*, and optional columns.
- Conditionally-required columns are columns that are required when certain conditions are met (e.g., *Cell_ID* is required any time the experiment involves the identification of Cell boundaries).
- Column names should use the underscore as a word separator (e.g., Spot_ID).
- The first column is always either Spot_ID or another relevant ID (i.e., Trace_ID, Cell_ID, etc.). In the *DNA-Spot/Trace Data core table*, there are eight mandatory columns. All other columns are ordered at user's discretion.
- The order of the rows is at user's discretion.
- If an optional column does not contain any data (i.e., it is not used), it should be omitted.

2.4 DNA-Spot/Trace Data core table

Requirement level: required

2.4.1 Summary

This is the mandatory core table of the 4DN FISH-omics Format for Chromatin Tracing. This table is used to record and exchange the primary results of Chromatin Tracing experiments. The Table is organized around individual DNA bright Spots that are spatially linked together in a three-dimensional (3D) polymeric Trace using a 3D polymeric tracing algorithm. As a result, all Spots that share the same `Trace_ID`, by definition belong to the same Trace.

Each row reports the `X`, `Y`, `Z` localization, and the Trace assignment (i.e., `Trace_ID`) of a FISH-omics bright Spot and corresponds to a specific genomic DNA target sequence identified by chromosome ID (`Chrom`), and by start (`Chrom_Start`) and end (`Chrom_End`) chromosome coordinates. In this table the reported `X`, `Y`, `Z` coordinates are assumed to result from post-processing and quality control procedures and therefore correspond to the final localization of the DNA target under study.

At a minimum the Table has to have 8 columns in the following order: **Spot_ID**, **Trace_ID**, **X**, **Y**, **Z**, **Chrom**, **Chrom_Start**, **Chrom_End**. These are required. Additionally in case sub-cellular structures, cells or extra cellular structures (e.g., Tissue) are identified as part of this experiment, this table has to mandatorily include the ID of the Sub_Cellular, Cell or Extra Cellular Structure Region of Interest (ROI) each Spot/Trace is associated with.

All other spot properties must be kept in the two additional tables *Spot Quality table* and *Spot Biological Data table*, indexed by Spot_ID and as described in the instructions for those tables. Additionally, in the case in which the final localization of DNA target results from combining multiple detection events (e.g., by combining localization events from different focal planes or times), the underlying raw data can be recorded in the corresponding *Spot Demultiplexing table* table as described in the instructions of that table.

Finally, `Spot_ID` identifiers are unique across the entire dataset, thus allowing to identify unambiguously a Spot in the *Spot Quality table*, *Spot Biological Data table* and *Spot Demultiplexing table*.

NOTE: Also RNA Spots have a `Spot_ID` (in the *RNA-Spot Data table*). Thus, when assigning an identifier to each Spot, make sure that this is unique not only within the *DNA-Spot/Trace Data core table*, but also in the *RNA-Spot Data table* if present.

2.4.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_core
##genome_assembly=GRCh38
##XYZ_unit=micron
#Software_Title: ChrTracer3
#Software_Type: SpotLoc+Tracing
#Software_Authors: Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN
#Software_Description: ChrTracer3 software was developed for analysis of raw DNA labeled␣
↪images. As an input, it takes an.xlsx table containing information and folder names of␣
↪the DNA experiment. As an output, it returns tab delimited.txt les with drift-
↪corrected x, y, z positions for all labeled barcodes. These can be used directly to␣
↪calculate the nm scale distances between all pairs of labeled loci. The current␣
↪version of the software as of this writing is ChrTracer3.
#Software_Repository: https://github.com/BoettigerLab/ORCA-public
#Software_PreferredCitationID: https://doi.org/10.1038/s41596-020-00478-x
#lab_name: Nobel
#experimenter_name: John Doe
#experimenter_contact: john.doe@email.com
#additional_tables: 4dn_FOF-CT_quality, 4dn_FOF-CT_rna, 4dn_FOF-CT_trace, 4dn_FOF-CT_cell
##columns=(Spot_ID, Trace_ID, X, Y, Z, Chrom, Chrom_Start, Chrom_End, Cell_ID)
1, 1, 14.43, 41.43, 1.23, chr1, 0001, 1000, 1
2, 1, 14.83, 41.83, 1.83, chr1, 1001, 2000, 1
3, 1, 15.83, 42.83, 1.33, chr1, 2001, 3000, 1
4, 2, 20.43, 50.43, 1.23, chr1, 0002, 2000, 1
5, 2, 21.83, 60.83, 1.83, chr1, 1002, 3000, 1
```

2.4.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_core"

The header MUST contain a mandatory set of fields that describe the algorithm(s) that were used to identify and localize bright Spots and to connect them to form Traces. In case more than one algorithm were used, please use the same set of fields for each of the algorithm used.

The columns for this table are mandatory and do not need to be described in the header.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version:	Version of the FOF format used in this release.	v0.1	
##Table_namespace:	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_core	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	
#Software_Author(s):	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understanding of the algortithm and of the analysis parameters that were used, in order to guarantee interpretation and reproducibility.	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled barcodes. These can be used directly to calculate the nm scale distances between all pairs of labeled loci. The current version of the software as of this writing is ChrTracer3.	
#Software_Repository:	The URL of any repository or repositories where the Software executable release can be obtained.	https://github.com/BoettigerLab/ORCA-public	
#Software_	The Unique Identifier for the	https://doi.org/10.1038/s41596-020-00478-x	

2.4.4 Data Columns

As with all other Spot Data tables in this format, each row corresponds to data associated with an individual Spot.

The first columns are always: **Spot_ID**, **Trace_ID**, **X**, **Y**, **Z**, **Chrom**, **Chrom_Start**, **Chrom_End**. Additionally in case sub-cellular structures, cells or extra cellular structures are identified as part of this experiment, the subsequent columns must mandatorily be *Sub_Cell_ROI_ID*, *Cell_ID* or *Extra_Cell_ROI_ID*, respectively.

The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Spot_ID	A unique identifier for this bright Spot.		
Trace_ID	In case multiple DNA Spots are connected to form 3D polymer traces of chromatin fibers (such as in ORCA; https://doi.org/10.1038/s41596-020-00478-x), this fields reports a unique identifier for the DNA trace the Spot belongs to. Note: this is used to connect Spots that are part of the same polymeric Trace. It is also used to connect data in this table with any Trace specific measurements such as nascent RNA expression, recorded in the corresponding Trace Data table.	1	
X	The sub-pixel X coordinate of this bright Spot. NOTE: the reported X position is understood to be the one resulting from any performed post-processing correction procedures (i.e. drift correction, chromatic correction etc).		
Y	The sub-pixel Y coordinate of this bright Spot. NOTE: the reported Y position is understood to be the one resulting from any performed post-processing correction procedures (i.e. drift correction, chromatic correction etc).		
Z	The sub-pixel Z coordinate of this bright Spot. NOTE: the reported Z position is understood to be the one resulting from any performed post-processing correction procedures (i.e. drift correction, chromatic correction etc).		
Chrom	Chromosome name. Because BED (Browser Extensible Data) is the de facto exchange bioinformatics format for genomic data, the BED terminology was used here.	chr3, chrY, chr2_random	
Chrom_Start	Start coordinate on the Chromosome for the sequence associated with this bright Spot (the first base on the chromosome is numbered 0). Because BED (Browser Extensible Data) is the de facto exchange bioinformatics format for genomic data, the BED terminology was used here.	0	
Chrom_End	End coordinate on the Chromosome for the sequence associated with this bright Spot. This position is non-inclusive, unlike Chrom_Start. Because BED (Browser Extensible Data) is the de facto exchange bioinformatics format for genomic data, the BED terminology was used here.	1000	
Sub_Cellular_ROI_ID	If known, this field reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a sub-cellular structure a given Spot/Trace is associated with. Note: this is used to connect individual Spot/Traces that are part of the same ROI. It is also used to connect data in this table with any ROI specific measurements such as boundaries, intensities or volume, recorded in the corresponding Sub-Cell ROI Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a Sub_Cell_ROI identified as part of this experiment.
Cell_ID	If known, this field reports the unique identifier for the Cell a given Spot/Trace is associated with. Note: this is used to connect individual Spot/Traces that are part of the same Cell. It is also used to connect data in this table with any Cell specific measurements such as boundaries, intensities and volume, recorded in the corresponding Cell Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a Cell identified as part of this experiment.
Extra_Cellular_ROI_ID	If known, this field reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a extracellular structure (e.g., Tissue) a given Spot/Trace is associated with. Note: this is used to connect individual Spot/Traces that are part of the same ROI. It is also used to connect data in this table with any ROI specific measurements such as boundaries, intensities and volume, recorded in the corresponding Extra-Cell ROI Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a extracellular structure ROI (e.g., Tissue) identified as part of this experiment.

2.5 RNA-Spot Data table

Requirement level: conditionally required

2.5.1 Summary

This table is used to store and share the results of RNA FISH-omics experiments and it is **conditionally required** in the case RNA data was collected as part of this experiment. Each row represents a detected RNA bright Spot and corresponds to the location of a specific RNA transcript.

At a minimum, one needs to know the **Spot_ID**, the **X**, **Y**, **Z** coordinates of each spot, the **Gene_ID** and an additional ID used to link this data with other tables in this format (i.e., *Trace_ID*, *Sub_Cell_ROI_ID*, *Cell_ID* and/or *Extra_Cell_ROI_ID*). In addition, in case multiple transcripts are associated with the same Gene_ID and the FISH probes are capable of distinguishing them, *Transcript_ID* MUST also be reported. Thus, at a minimum there needs to be 6 (or 7) data columns. These are required. All other data columns are optional.

In this table the reported X, Y and Z coordinates are assumed to result from post-processing and quality control procedures performed on primary localization events and therefore correspond to what is considered the best-bet location of the RNA molecule under study.

In the case of multiplexed FISH experiments (i.e., MERFISH) in which the final location of RNA molecule results from combining multiple detection events (e.g., by combining individual Localization events detected in separate planes or images), the underlying raw data can be recorded in the corresponding *Spot Demultiplexing table* as described in the instructions of that table.

`Spot_ID` identifiers are unique across the entire dataset, thus allowing to identify unambiguously a Spot in the *Spot Quality table*, *Spot Biological Data table* and *Spot Demultiplexing table*.

NOTE: Also DNA Spots have a `` `Spot_ID `` (in the *DNA-Spot/Trace Data core table*). Thus, when assigning an identifier to each Spot, make sure that this is unique not only within the *RNA-Spot Data table*, but also in the *DNA-Spot/Trace Data core table*.

2.5.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_rna
##genome_assembly=GRCh38
##XYZ_unit=micron
##Gene_ID_type=Ensemble_V38
#Software_Title: Xyz
#Software_Type: SpotLoc
#Software_Authors: Janet Doette
#Software_Description: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas␣
↪sagittis est mollis, pulvinar tortor mattis, dignissim nisi. Nunc tincidunt volutpat␣
↪lacus vitae bibendum.
#Software_Repository: https://xyz.com
#Software_PreferredCitationID: https://doi.org/xyz
#lab_name: Nobel
#experimenter_name: John Doe
#experimenter_contact: john.doe@email.com
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_quality, 4dn_FOF-CT_cell
##columns=(Spot_ID, X, Y, Z, RNA_name, Gene_ID, Transcript_ID, Cell_ID)
1, 14.43, 41.43, 1.23, ACTB, ENSG00000075624, ENST00000646664.1, 1
```

(continues on next page)

(continued from previous page)

```
2, 14.83, 41.83, 1.83, GAPDH, ENSG00000111640, ENST00000229239.10, 1
3, 15.83, 42.83, 1.33, MB, ENSG00000198125, ENST00000397326.7, 1
```

2.5.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_rna"

The header MUST contain a mandatory set of fields that describe the algorithm(s) that were used to identify and localize bright Spots. In case more than one algorithm were used, please use the same set of fields for each of them.

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format version in this case.	v0.1	
##Table_namespace=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_rna	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	
#Software_Authors:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understanding of the algorithm and of the analysis parameters that were used, in order to guarantee interpretation and reproducibility.	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled barcodes. These can be used directly to calculate the nm scale distances between all pairs of labeled loci. The current version of the software as of this writing is ChrTracer3.	
#Software_Repository:	The URL of any repository where the Software executable release can be obtained.	https://github.com/BoettigerLab/ORCA-public	
#Software_	The Unique Identifier for the	https://doi.org/10.1038/s41596-020-00478-x	

2.5.4 Data Columns

As with all other Spot Data tables in this format, each row corresponds to data associated with an individual Spot.

The first columns are always: **Spot_ID**, **X**, **Y**, **Z**, **RNA_name**, **Gene_ID**, followed by *Transcript_ID* if applicable, and by **one or more** of the following *Trace_ID*, *Sub-Cell_ROI_ID*, *Cell_ID* and/or *Extra_Cell_ROI_ID*. The order of the other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Spot_ID	A unique identifier for this bright Spot.	1	
X	The sub-pixel X coordinate of this bright Spot. NOTE: the reported X position is understood to be the one resulting from any performed post-processing procedures (i.e. drift correction, chromatic correction etc).	14.43	
Y	The sub-pixel Y coordinate of this bright Spot. NOTE: the reported Y position is understood to be the one resulting from any performed post-processing procedures (i.e. drift correction, chromatic correction etc).	14.43	
Z	The sub-pixel Z coordinate of this bright Spot. NOTE: the reported Z position is understood to be the one resulting from any performed post-processing procedures (i.e. drift correction, chromatic correction etc).	1.23	
RNA_name	This is the official name of the Gene the targeted RNA is transcribed from.	ACTB	
Gene_ID	This is the official ID for the Gene encoding for the targeted RNA transcript.	ENSG00000075624	
Transcript_ID	This is the official ID for the targeted RNA transcript. This field is required in case the same Gene has multiple different Transcripts and the FISH probe used in this case is capable of distinguishing between them.	ENST00000466644.1	Conditional requirement: this MUST be reported if multiple transcripts are associated with the same Gene_ID and the FISH probes are capable of distinguishing them.
Trace_ID	This fields reports the unique identifier for a DNA Trace identified as part of this experiment. Note: this is used to connect data in this table with a given Trace and with Trace specific measurements as recorded in the corresponding Trace Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a Trace identified as part of this experiment.
Sub_Cell_ROI_ID	If known, this fields reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a sub-cellular structure a given Spot is associated with. Note: this is used to connect individual Spots that are part of the same ROI. It is also used to connect data in this table with any ROI specific measurements such as boundaries, intensities or volume, recorded in the corresponding Sub-Cell ROI Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a Sub_Cell_ROI identified as part of this experiment.
Cell_ID	If known, this fields reports the unique identifier for the Cell a given Spot is associated with. Note: this is used to connect individual Spots that are part of the same Cell. It is also used to connect data in this table with any Cell specific measurements such as boundaries, intensities and volume, recorded in the corresponding Cell Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a Cell identified as part of this experiment.
Extra_Cellular_ROI_ID	If known, this fields reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a extracellular structure (e.g., Tissue) a given Spot is associated with. Note: this is used to connect individual Spots that are part of the same ROI. It is also used to connect data in this table with any ROI specific measurements such as boundaries, intensities and volume, recorded in the corresponding Extra-Cell ROI Data table.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a extracellular structure ROI (e.g., Tissue) identified as part of this experiment.

2.6 Spot Quality table

Requirement level: recommended

2.6.1 Summary

This table is highly recommended and it is designed to provide quality metrics for the Spot localization, information about the optical Channel that was used to image the Spot, and various aberration corrections that have been applied prior to localization (e.g., drift correction, chromatic correction, etc.).

Because the metrics used to quantify Spot detection accuracy and precision are not trivial and lacking a widely shared consensus, the specific columns in this table remain largely at the user's discretion and should be described with sufficient details to ensure interpretation and reproducibility.

However, in order to align with existing 4DN-BINA-OME Microscopy Metadata specifications, the use of specific column names and descriptions is **conditionally required** in case the described metric is reported. As an example, the column name **X_Drift** is conditionally required in case the user intends to report a comparison between the Observed vs. Expected (i.e., based on a fiducial reference) positions of a detected Spot.

The table is indexed by Spot_ID and each row corresponds to a DNA or RNA bright Spot. The order of all other columns (including those conditionally required) and of the rows are at the user's discretion.

2.6.2 Example

Spot fit quality

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_quality
##XYZ_unit=micron
##intensity_unit=photonCount
#Software_Title: SpotQualityCheck
#Software_Type: QualityControl
#Software_Authors: John Doe et al.
#Software_Description: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do␣
↪eiusmod tempor incididunt ut labore et dolore magna aliqua.
#Software_Repository: https://eample.com
#Software_PreferredCitationID: https://doi.org/00000
#lab_name: Nobel
#experimenter_name: John Doe
#experimenter_contact: john.doe@email.com
#^Channel_ID: A unique identifier that refers to the Channel that was used to image this␣
↪Spot.
#^Peak_Intensity: The signal intensity of the brightest pixel within a bright Spot (i.e.␣
↪local maximum).
#^Raw_X: the original fit x-position relative to the camera and objective, (prior to␣
↪drift correction, chromatic correction, or conversion to stage coordinates). This is␣
↪the appropriate coordinate system for correcting optical aberrations.
#^Raw_Y: the original fit y-position relative to the camera and objective, (prior to␣
↪drift correction, chromatic correction, or conversion to stage coordinates). This is␣
↪the appropriate coordinate system for correcting optical aberrations.
#^Raw_Z: the original fit z-position relative to the camera and objective, (prior to␣
↪drift correction, chromatic correction, or conversion to stage coordinates). This is␣
```

(continues on next page)

(continued from previous page)

```
↪the appropriate coordinate system for correcting optical aberrations.
#^X_Drift: the distance in nm the spot was moved in x based on fiducial tracking
#^Y_Drift: the distance in nm the spot was moved in y based on fiducial tracking
#^Z_Drift: the distance in nm the spot was moved in z based on fiducial tracking
#^X_Chromatic_Shift: the distance in nm the spot was moved in x based on chromatic␣
↪correction map
#^Y_Chromatic_Shift: the distance in nm the spot was moved in y based on chromatic␣
↪correction map
#^Z_Chromatic_Shift: the distance in nm the spot was moved in z based on chromatic␣
↪correction map
#^X_Loc_Precision: lower and upper bound of 95% confidence interval on X-position after␣
↪fit
#^Y_Loc_Precision: lower and upper bound of 95% confidence interval on Y-position after␣
↪fit
#^Z_Loc_Precision: lower and upper bound of 95% confidence interval on Z-position after␣
↪fit
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_rna, 4dn_FOF-CT_trace, 4dn_FOF-CT_cell
##columns=(Spot_ID, Channel_ID, Peak_Intensity, Raw_X, Raw_Y, Raw_Z, X_Drift, Y_Drift, Z_
↪Drift, X_Chromatic_Shift, Y_Chromatic_Shift, Z_Chromatic_Shift, X_Loc_Precision, Y_Loc_
↪Precision, Z_Loc_Precision)
1, 1, 100, 1.1, 1.05, 1.2, 0.1, 0.05, 0.2, 0.2, 0.2, 0.2, 0.01, 0.01, 0.01
2, 1, 200, 1.11, 1.055, 1.22, 0.11, 0.055, 0.22, 0.22, 0.22, 0.22, 0.012, 0.012, 0.012
3, 2, 500, 1.12, 1.054, 1.21, 0.12, 0.054, 0.21, 0.22, 0.22, 0.22, 0.012, 0.012, 0.012
4, 3, 333, 1.13, 1.15, 1.202, 0.13, 0.15, 0.202, 0.23, 0.23, 0.23, 0.013, 0.013, 0.013
```

2.6.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_quality"

The header MUST contain a mandatory set of fields that describe any algorithm that was used to produce/process data in this table. In case more than one algorithm were used, please use the same set of fields for each of them.

The header MUST include a detailed description of each optional columns used.

Name	Description
##FOF-CT_version=	Version of the FOF format used in this case.
##Table_namespace=	Identifier for this type of table. Value must be as in the example.
#lab_name:	name of the lab where the experiment was performed.
#experimenter_name:	name of the person performing the experiment.
#experimenter_contact:	email address of the person performing the experiment.
#description:	A free-text, description of the experiment and of the data recorded in this table. This description
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bri
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation,
#Software_Authors:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Autho
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understandir
#Software_Repository:	The URL of any repository or archive where the Software executable release can be obtained.
#Software_PreferredCitationID:	The Unique Identifier for the preferred/primary publication describing this Software. Examples i
#additional_tables:	list of the additional tables being submitted. Note: use a comma to separate each table name fron
#Intensity_Measurement_Method:	If relevant, the method that was used to performed intensity measurements. In particular, sufficie

Name	Description
#^Channel_ID:	A unique identifier that refers to the Channel that was used to image this Spot.
#^Fluorophore_ID:	A unique identifier that refers to the Fluorophore whose Emission is utilized to detect this Spot.
#^Centroid_Intensity:	The signal intensity of the pixel occupying the center-of-mass within a bright Spot (i.e. centroid)
#^Peak_Intensity:	The signal intensity of the brightest pixel within a bright Spot (i.e. local maximum).
#^Raw_X:	The Raw sub-pixel X coordinate of this bright Spot relative to the optical system (i.e., Objective
#^Raw_Y:	The Raw sub-pixel Y coordinate of this bright Spot relative to the optical system (i.e., Objective
#^Raw_Z:	The Raw sub-pixel Z coordinate of this bright Spot relative to the optical system (i.e., Objective
#^X_Drift:	This field captures the offset in the observed X-coordinate of the Intensity maxima or the Intensi
#^Y_Drift:	This field captures the offset in the observed Y-coordinate of the Intensity maxima or the Intensi
#^Z_Drift:	This field captures the offset in the observed Z-coordinate of the Intensity maxima or the Intensi
#^X_Chromatic_Shift:	This field captures the offset in the observed X-coordinate of the Intensity maxima or the Intensi
#^Y_Chromatic_Shift:	This field captures the offset in the observed Y-coordinate of the Intensity maxima or the Intensi
#^Z_Chromatic_Shift:	This field captures the offset in the observed Z-coordinate of the Intensity maxima or the Intensi
#^X_Loc_Error:	Metric used to quantify the Error associated with the estimation of the X-axis localization of this
#^Y_Loc_Error:	Metric used to quantify the Error associated with the estimation of the Y-axis localization of this
#^Z_Loc_Error:	Metric used to quantify the Error associated with the estimation of the Z-axis localization of this
#^X_Loc_Precision	Metric used to quantify the Precision associated with the estimation of the X-axis localization of
#^Y_Loc_Precision	Metric used to quantify the Precision associated with the estimation of the Y-axis localization of
#^Z_Loc_Precision	Metric used to quantify the Precision associated with the estimation of the Z-axis localization of
#^optional_column_1:	
#^optional_column_2:	
#^optional_column_3:	
##XYZ_unit=	The unit used to represent XYZ locations or distances in this table. Note: use micron (instead of
##time_unit=	If relevant, the unit used to represent a time interval. Note: use 'sec' for seconds, 'msec' for mill
##intensity_unit=	If relevant, the unit used to represent intensity measurements.
##columns=	list of the data column headers used in the table. Note: enclose the column headers and use a co

2.6.4 Data Columns

As with all other Spot Data tables in this format, each row corresponds to data associated with an individual Spot.

The first column of this table is always Spot_ID. The content and order of all other columns is largely at user's discretion. However, in order to align with existing Microscopy Metadata specifications, the use of specific column names and descriptions is **conditionally required** as indicated below. The order of all other columns (including those conditionally requried) and of the rows are at the user's discretion.

Name	Description	Example	Conditional requirement conditions
Spot_ID	A unique identifier for this bright Spot.	1	
conditionally_required_column_1:			one of the conditionally required columns desribed in the header
conditionally_required_column_2:			one of the conditionally required columns desribed in the header
conditionally_required_column_3:			one of the conditionally required columns desribed in the header
optional_column_1:			
optional_column_2:			
optional_column_3:			

2.7 Spot Biological Data table

Requirement level: recommended

2.7.1 Summary

This table is highly recommended and it is designed to store and share biological properties associated with individual Spots (e.g., distance from the nuclear lamina (NL) or the nuclear pore complex (NPC), etc.; Su et al 2020 Cell and Takei et al 2021 Nature) identified as part of this experiment. In the absence of a consensus regarding biological properties to be recorded in association with individual bright Spots, the specific columns in this table remain at the user's discretion and should be described with sufficient details to ensure interpretation and reproducibility.

This table is mandatorily indexed by Spot_ID.

2.7.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_bio
##XYZ_unit=micron
#^NL_distance:
#^H4K27me3_distance:
#additional_tables: 4dn_FOF-CT_rna, 4dn_FOF-CT_cell
##columns=(Spot_ID, NL_distance, H4K27me3_distance)
1, 1.345, 0.445
2, 1.245, 0.005
3, 1.005, 0.150
```

2.7.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_bio"

This Table can be indexed mandatorily by Spot_ID.

The header MUST contain a mandatory set of fields that describe any algorithm that was used to produce/process data in this table. In case more than one algorithm were used, please use the same set of fields for each of them.

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format used in this case.	v0.1	
##Table_name=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_bio	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Author:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understanding of the algortithm and of the analysis parameters	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled	Conditional requirement: this MUST be reported any time a soft-

2.7.4 Data Columns

Each row corresponds to data associated with an individual Spot. The first column is always **Spot_ID**. The order of the other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Spot_ID	A unique identifier for this bright Spot.	1	
optional_column_1:			
optional_column_2:			
optional_column_3:			

2.8 Spot Demultiplexing table

Requirement level: optional

2.8.1 Summary

This table is optional and is designed to be used in the case of multiplexed FISH experiments (i.e., MERFISH) in which the final localization of a bright DNA or RNA Spot results from the combination of multiple individual localization events (e.g., by combining particles detected and localized in separate images). In such a case the final Spot localization data is recorded in the *DNA-Spot/Trace Data core table*, while the underlying primary localization data can be recorded by using this table, as shown for DNA Spots in the example below.

This table is indexed by **Loc_ID**, mandatorily reports the **X**, **Y**, **Z** coordinates of the Localization event, and it has a mandatory **Spot_ID** column that is used to link individual localization events to the resulting Spot.

Other columns are at user's discretion.

2.8.2 Example

DNA spots detected with multiplexed barcodes

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_demultiplexing
##XYZ_unit=micron
#Software_Title: ExampleLocalizationSoftware
#Software_Type: SpotLoc
#Software_Authors: Doe, J.
#Software_Description: A pretty clear description
#Software_Repository: https://github.com/repo_name_goes_here
#Software_PreferredCitationID: https://doi.org/doi_goes_here
#lab_name: Nobel
#experimenter_name: John Doe
#experimenter_contact: john.doe@email.com
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_quality
#^Hyb: the labeling round in which this localization occurred
```

(continues on next page)

(continued from previous page)

```
#^Fluor: the fluorescent channel in which this localization was detected
#^Brightness: the photon count for this localization event
#^Fit_Quality: the quality of fit for this localization, on a relative scale of 0-1
##columns=(Loc_ID, Spot_ID, X, Y, Z, Hyb, Fluor, Brightness, Fit_Quality)
1, 1, 2342, 2354, 545, 2, cy3, 1003, 0.83
2, 1, 2342, 2354, 545, 2, cy5, 2000, 0.93
3, 1, 2342, 2354, 545, 3, cy5, 1233, 0.85
4, 2, 3345, 5432, 654, 3, cy3, 2324, 0.95
5, 2, 3345, 5432, 654, 3, cy5, 2324, 0.95
6, NA, 4345, 432, 100, 4, cy3, 2324, 0.95
7, 2, 3345, 5432, 654, 4, cy3, 2324, 0.95
```

2.8.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_demultiplexing"

The header MUST contain a mandatory set of fields that describe any algorithm that was used to produce/process data in this table. In case more than one algorithm were used, please use the same set of fields for each of them.

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format used in this case.	v0.1	
##Table_name=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_demultiplexing	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Author:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Description:	A free-text, description of this software. This description should provide a detailed understanding of the algortithm and of the analysis parameters	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled	Conditional requirement: this MUST be reported any time a soft-

2.8.4 Data Columns

This table is indexed by Loc_ID and therefore each row corresponds to data associated with an individual Localization event.

The first columns are always: **Loc_ID**, **Spot_ID**, **X**, **Y**, **Z**. The content and order of all other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Loc_ID	A unique identifier for this individual Localization event.	1	
Spot_ID	A unique identifier for the bright DNA or RNA Spot with which this individual localization event is associated.	1	
X	The sub-pixel X coordinate of this Localization event.		
Y	The sub-pixel Y coordinate of this Localization event.		
Z	The sub-pixel Z coordinate of this Localization event.		
#^optional_column_1:			
#^optional_column_2:			
#^optional_column_3:			

2.9 Trace Data table

Requirement level: optional

2.9.1 Summary

This table is used to document properties that are globally associated with individual Traces rather than individual bright Spots (e.g., Physical coordinates, RNA transcription, or Allele). These are properties that are shared by all bright Spots that constitute a Trace.

Each row in the table corresponds to an individual Trace and is indexed by a unique **Trace_ID** that links the data reported in this table with data stored in one of the other tables (e.g., *DNA-Spot/Trace Data core table*, *RNA-Spot Data table*, etc.).

2.9.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_trace
##XYZ_unit=micron
##intensity_unit=a.u.
#^allele: This field records the Allele to which this Trace was mapped. This can be one␣
↪of the following values: BL6, CAST.
#^RNA_A_intensity: This records the intensity of the nascent RNA A expression signal␣
↪associated with this Trace.
#^NL_distance: This field records the distance of this Trace to the Nuclear Lamina.
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_cell
```

(continues on next page)

(continued from previous page)

```
##columns=(Trace_ID, allele, RNA_A_int, NL_distance)
1, BL6, 43253, 0.235
2, CAST, 40001, 0.563
3, BL6, 1000, 0.135
4, CAST, 1500, 0.633
```

2.9.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_trace"

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version:	Version of the FOF format used in this case.	v0.1	
##Table_name:	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_trace	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISHomics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Author:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understanding of the algortithm and of the analysis parameters	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled	Conditional requirement: this MUST be reported any time a soft-

2.9.4 Data Columns

Each row corresponds to data associated with an individual Trace.

The first column of this table is always **Trace_ID**. The content and order of all other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Trace_ID	This field reports the unique identifier for a DNA Trace identified as part of this experiment. Note: this is used to connect data in this table with a given Trace as recorded in the corresponding DNA-Spot/Trace Data core table.	1	
optional_column_1			
optional_column_2			
optional_column_3			

2.10 Cell Data table

Requirement level: optional

2.10.1 Summary

This table is used to document properties that are globally associated with individual Cells (e.g., cell size, cell volume, cell type) and it is required in the case Cell segmentation data was collected as part of this experiment. These are properties that are shared by all bright Spots and Traces that belong to an individual Cell. Each row in the table corresponds to a different Cell studied in the experiment and is identified by a unique **Cell_ID** that links the data reported in this table with data stored in one of the other tables (e.g., *DNA-Spot/Trace Data core table*, *Sub-Cell ROI Data table*, *Cell/ROI Mapping table*, etc.).

2.10.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_cell
##XYZ_unit=micron
##Extra_Cell_ROI_type=Organoid
#^RNA_A_nr: the number of detected bright Spots corresponding to RNA transcript A␣
↪detected in this Cell, see also RNA Spot Data table
#^RNA_B_nr: the number of detected bright Spots corresponding to RNA transcript B␣
↪detected in this Cell, see also RNA Spot Data table
#^cell_cycle_state: the Cell Cycle state in which this Cell is found as measured with␣
↪the Fucci system. This column can contain one of the following values: G1, S, G2 or M.
#^cell_volume: the volume of this Cell expressed in micron^3.
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_rna, 4dn_FOF-CT_trace
##columns=(Cell_ID, Extra_Cell_ROI_ID, RNA_A_nr, RNA_B_nr, cell_cycle_state, cell_volume)
1, 1, 10, 22, 1041.5, 12354.24, G1, 13453
```

(continues on next page)

(continued from previous page)

```
2, 1, 0, 11, 2041.3, 32234.24, G2, 35545
3, 2, 10, 33, 101.5, 12354.24, S, 10010
4, 3, 0, 44, 201.1, 32234.24, M, 25340
```

2.10.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_cell"

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format used in this case.	v0.1	
##Table_name=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_cell	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Author:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Description:	A free-text, description of this software. This description should provide a detailed understanding of the algortithm and of the analysis parameters	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled	Conditional requirement: this MUST be reported any time a soft-

2.10.4 Data Columns

Each row corresponds to data associated with an individual Cell.

The first column of this table is always **Cell_ID**. The content and order of all other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Cell_ID	This fields reports the unique identifier for Region of Interest (ROI) that represent the boundaries of a Cell identified as part of this experiment. Note: this is used to connect individual Spots or Traces that are part of the same Cell.	1	
Extra_Cellular_ROI_ID	In case multiple Cells are localized within a given extracellular (e.g., Tissue) Region of Interest (ROI), this fields reports the unique identifier that allows to identify such as ROI. Note: this is used to connect individual Cells that are part of the same extracellular ROI.	1	Conditional requirement: this column is mandatory if data in this table can be associated with an extracellular ROI identified as part of this experiment.
optional_column_1			
optional_column_2			
optional_column_3			

2.11 Sub-Cell ROI Data table

Requirement level: conditionally required

2.11.1 Summary

This table is used to document properties that are globally associated with individual sub-cellular ROIs that typically correspond to sub-nuclear features (e.g., Nucleoli, Nuclear Lamina, Chromosome Domains, PML bodies, etc.) and it is required in the case sub-cellular ROI segmentation data was collected as part of this experiment. These are properties that are shared by all bright Spots and Traces that are associated with individual ROIs. Each row in the table corresponds to a different Subcell ROI studied in the experiment and is identified by a unique **Sub_Cell_ROI_ID** that links the data reported in this table with data stored in one of the other tables (e.g., *DNA-Spot/Trace Data core table*, *Cell Data table*, etc.).

2.11.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_subcell
##XYZ_unit=micron
##intensity_unit=a.u.
##Sub_Cell_ROI_type=Nucleolus
#^ROI_volume: the volume of this ROI expressed in micron^3.
#^ROI_intensity: the integrated average signal intensity of the marker of interest as␣
```

(continues on next page)

(continued from previous page)

```
↪measured within the boundaries of this ROI.
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_rna, 4dn_FOF-CT_trace
##columns=(Sub_Cell_ROI_ID, Cell_ID, ROI_volume, ROI_intensity)
1, 1, 1345, 3500
2, 1, 3554, 1500
3, 2, 1001, 2500
4, 3, 2534, 3498
```

2.11.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_subcell"

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format used in this case.	v0.1	
##Table_name=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_subcell	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Author:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understanding of the algortithm and of the analysis parameters	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled	Conditional requirement: this MUST be reported any time a soft-

2.11.4 Data Columns

Each row corresponds to data associated with an individual subcellular ROI.

The first column of this table is always **Sub_Cell_ROI_ID**. The content and order of all other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Sub_Cell_ROI_ID	This fields reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a sub-cellular structure identified as part of this experiment. Note: this is used to connect all Spots, and Traces that belong to the same ROI.	1	
Cell_ID	This fields reports the unique identifier for Region of Interest (ROI) that represent the boundaries of a Cell identified as part of this experiment. Note: this is used to connect individual Spots or Traces that are part of the same Cell.	1	Conditional requirement: this column is mandatory if data in this table can be associated with a Cell identified as part of this experiment.
optional_column_1			
optional_column_2			
optional_column_3			

2.12 Extra-Cell ROI Data table

Requirement level: conditionally required

2.12.1 Summary

This table is used to document properties (i.e., volume, mean fluorescence intensity) that are globally associated with individual extracellular structures (e.g., Tissue, Organoid, etc.) Regions of Interest (ROI), and it is required in the case extracellular ROI segmentation data was collected as part of this experiment. These are properties that are shared by all bright Spots, Traces and Cells that belong to an individual extracellular structure identified as part of this study. Each row in the table corresponds to a different extracellular structure studied in the experiment and is identified by a unique **Extra_Cell_ROI_ID** that links the data reported in this table with data stored in one of the other tables (e.g., *DNA-Spot/Trace Data core table*, *RNA-Spot Data table*, etc.).

2.12.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_extracell
##XYZ_unit=micron
##Extra_Cell_ROI_type=Organoid
#^Cell_A_nr: the number of identified Cells of type A identified to belong to this␣
↪extracellular ROI.
#^Cell_B_nr: the number of identified Cells of type B identified to belong to this␣
```

(continues on next page)

(continued from previous page)

```
↪extracellular ROI.
#^ROI_volume: the volume of this extracellular ROI expressed in micron^3.
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_rna, 4dn_FOF-CT_trace
##columns=(Extra_Cell_ROI, Cell_A_nr, Cell_B_nr, ROI_volume)
1, 10, 22, 13453
2, 0, 11, 35545
3, 10, 33, 10010
4, 44, 0, 25340
```

2.12.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_extracell"

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format used in this case.	v0.1	
##Table_name=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_extracell	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) that were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Author:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	Conditional requirement: this MUST be reported any time a software is used to produce data associated with this table.
#Software_Description:	A free-text, description of this software. This description should provide a detailed understanding of the algortithm and of the analysis parameters	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled	Conditional requirement: this MUST be reported any time a soft-

2.12.4 Data Columns

Each row corresponds to data associated with an individual extracellular ROI.

The first column of this table is always **Extra_Cell_ROI_ID**. The content and order of all other columns is at user's discretion. The order of the rows is at user's discretion.

Name	Description	Example	Conditional requirement conditions
Extra_Cell_ROI_ID	This fields reports the unique identifier for an extracellular structure (e.g., Tissue, Organoid) Region of Interest (ROI) identified as part of this experiment. Note: this is used to connect individual Cells that are part of the same extracellular ROI.	1	
optional_column_1			
optional_column_2			
optional_column_3			

2.13 Cell/ROI Mapping table

Requirement level: conditionally required

2.13.1 Summary

This table is used to provide the boundaries of Cells and other ROIs identified as part of this experiment, and it is required in case Cell and other ROI segmentation data were collected as part of this experiment.

This table is mandatory in case a *Sub-Cell ROI Data table*, *Cell Data table*, and/or *Extra-Cell ROI Data table* tables are deposited with this submission.

The table is organized on a Cell or ROI basis via a Cell or ROI ID and provides the Cell or ROI boundaries in global coordinates as specified by the OME ROI data model.

This table might be organized in one of the following manner:

- **Cell_ID** → Cell boundaries in global coordinates (following the OME Data Model for Polygon - ROI, the Cell boundaries are defined as a list of comma separated x,y coordinates separated by spaces like "x1,y1 x2,y2 x3,y3" e.g. "0,0 1,2 3,5").
- **Sub_Cell_ROI_ID** → Sub-cellular ROI (e.g., Nuclear feature, Nucleolus, etc.) boundaries x/y/z in global coordinates (following the OME Data Model for Polygon - Sub_Cell ROI, boundaries are defined as a list of comma separated x,y coordinates separated by spaces like "x1,y1 x2,y2 x3,y3" e.g. "0,0 1,2 3,5"). This table might also report the feature brightness.
- **Extra_Cell_ROI_ID** → Extracellular ROI boundaries (e.g., Tissue) in global coordinates (following the OME Data Model for Polygon - ROI, Super-Cell ROI boundaries are defined as a list of comma separated x,y coordinates separated by spaces like "x1,y1 x2,y2 x3,y3" e.g. "0,0 1,2 3,5").

In addition, this table might be used to report additional vectorial properties such as:

- Lists of RNA Spot x/y/z in global coordinates
- Lists of barcode sequence ID

- Lists of channels

2.13.2 Example

```
##FOF-CT_version=v0.1
##Table_namespace=4dn_FOF-CT_mapping
##XYZ_unit=micron
##intensity_unit=a.u.
##Sub_Cell_ROI_type=PML_body
##ROI_boundaries_format=(X1,Y1 X2,Y2 Xn,Yn)
#^ROI_volume: the volume of this ROI expressed in micron^3.
#^ROI_intensity: the integrated average signal intensity measured within the boundaries␣
↪of this ROI, of the marker used to identify this nuclear feature.
#additional_tables: 4dn_FOF-CT_core, 4dn_FOF-CT_rna, 4dn_FOF-CT_trace
##columns=(Sub_Cell_ROI_ID, ROI_boundaries, ROI_volume, ROI_intensity)
1, (0,0 1,2 3,5), 100, 1.00
2, (0,0 2,3 4,6), 48, 0.90
3, (0,0 3,2 7,5), 63, 0.67
4, (0,0 9,2 9,5), 88, 0.10
```

2.13.3 File Header

- The first line in the header is always "##FOF-CT_version=vX.X"
- The second line in the header is always "##Table_namespace=4dn_FOF-CT_mapping"

The header MUST include a detailed description of each optional columns used.

Name	Description	Example	Conditional requirement conditions
##FOF-CT_version=	Version of the FOF format used in this case.	v0.1	
##Table_name=	Identifier for this type of table. Value must be as in the example.	4dn_FOF-CT_mapping	
#lab_name:	name of the lab where the experiment was performed.	Nobel	
#experimenter_name:	name of the person performing the experiment.	John Doe	
#experimenter_contact:	email address of the person performing the experiment.	john.doe@email.com	
#description:	A free-text, description of the experiment and of the data recorded in this table. This description should provide a clear understanding of the process utilized to produce the data and contain sufficient details to ensure interpretation and reproducibility.		
#Software_Title:	The name of the Software(s) were used in this case for localizing individual FISH-omics bright Spots and/or to produce three-dimensional (3D) polymeric chromatin Traces.	ChrTracer3	
#Software_Type:	The type of this Software. Allowed values: SpotLoc, Tracing, SpotLoc+Tracing, Segmentation, QC, Other	SpotLoc+Tracing	
#Software_Authors:	The Name(s) of the individual Author(s) of this Software. In case there are more than one Authors, individual names should be listed as follows, Doe, John; Smith, Jane; etc,.	Mateo, LJ; Sinnott-Armstrong, N; Boettiger, AN	
#Software_Description:	A free-text, description of this Software. This description should provide a detailed understanding of the algorithm and of the analysis parameters that were used, in order to guarantee interpretation and reproducibility.	ChrTracer3 software was developed for analysis of raw DNA labeled images. As an input, it takes an.xlsx table containing information and folder names of the DNA experiment. As an output, it returns tab delimited.txt les with drift-corrected x, y, z positions for all labeled barcodes. These can be used directly to calculate the nm scale distances between all pairs of labeled loci. The current version of the software as of this writing is ChrTracer3.	
#Software_Repository:	The URL of any repository or Repository where the Software executable release can be obtained.	https://github.com/BoettigerLab/ORCA-public	
#Software_PreferredCitationID:	The Unique Identifier for the preferred Citation publication describing this Software. Examples include, Digital Object Identifier (DOI), PubMed Central Identifier (PMCID), ArXiv.org ID etc,.	https://doi.org/10.1038/s41596-020-00478-x	

2.13.4 Data Columns

Each row corresponds to data associated with an individual **Cell_ID**, **Sub_Cell_ROI_ID**, or **Extra_Cell_ROI_ID**.

The first column of this table is always the relevant ID. The content and order of all other columns is at user's discretion. The order of the rows is at user's discretion.

It is mandatory to choose one of the three types of ID.

Name	Description	Example	Conditional requirement conditions
Sub_Cell_ROI_ID	This fields reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a sub-cellular structure identified as part of this experiment. Note: this is used to connect all Spots, and Traces that belong to the same ROI.	1	Conditional requirement: This table must have at least one of the ID columns. Sub_Cell_ROI_ID MUST be reported if this table contains sub-cellular ROI data
Cell_ID	This fields reports the unique identifier for Region of Interest (ROI) that represent the boundaries of a Cell identified as part of this experiment. Note: this is used to connect individual Spots or Traces that are part of the same Cell.	1	Conditional requirement: This table must have at least one of the ID columns. Cell_ID MUST be reported if this table contains Cell data
Extra_Cell_ROI_ID	This fields reports the unique identifier for a Region of Interest (ROI) that represents the boundaries of a extracellular structure (e.g., Tissue) identified as part of this experiment. Note: this is used to connect all Spots, and Traces that belong to the same ROI.	1	Conditional requirement: This table must have at least one of the ID columns. Extra_Cell_ROI_ID MUST be reported if this table contains extra-cellular ROI data.
optional_column_1			
optional_column_2			
optional_column_3			

2.14 Miscellaneous

Contents

* *Example Tables*

2.14.1 Contributors

Contributors, listed alphabetically: Sarah Aufmkolk, Bogdan Bintu, Alistair Boettiger, Andrea Cosolo, Adam Jussila, Caterina Strambio De Castillia, Steven Wang.

2.14.2 Older revision history

Note: Older versions of this document are available in the following Google Doc: https://docs.google.com/document/d/1z7rIYsQnbeS7y_SMuwoa8qsWKBD_BpV88vR79WiH_XI/edit?usp=sharing and Google Sheet: https://docs.google.com/spreadsheets/d/1GvqokS5w8Yw2tAngsqDC8YcLdRha5cGr/edit?usp=sharing&ouid=103316056144222958298&rtpof=true&sd=true

Feb 1, 2021 Alistair Boettiger

Feb 2, 2021 Bogdan Bintu, Steven Wang, Alistair Boettiger

Feb 8, 2021 Bogdan Bintu, Steven Wang, Alistair Boettiger

Feb 9, 2021 Steven Wang, Andrea Cosolo, Andrew Schroeder, Alistair Boettiger

Feb 12, 2021 Alistair

Feb 26, 2021 Caterina Strambio De Castillia

July 6, 2021 Alistair, Andrea

Aug, 2021, Sarah + Alistair

Sept 10, 2021 Alistair

Sept 16, 2021 Sarah (addition of SMLM data example #3 and #4)

October 18-29, 2021 Caterina (various comments and changes)

October 25, 2021 Discussion between Alistair and Caterina to address several comments/issues. The main clarification point was that this format is used specifically to define Chromatin Tracing results. This is a subtype of a more generic FISH Omics Format. Other subtypes will be defined ASAP.

November, 2021 Caterina (various comments and changes)

February 9, 2022 Caterina and Andrea: Change name and description for tables #4 and #5 and add Table# to table header.

2.14.3 4DN Experimental and Microscopy Metadata

- Project =
- Center =
- Lab =
- Experiment protocol description =
- Date collected =
- Date submitted =

- Experiment Type = FISH Omics - Chromatin Tracing
- Experiment Set Type = Replicate
- Organism = D. melanogaster
- Biosource Type = tissue culture cell line
- Biosource = IMR90
- Modification Type = none
- Treatment Type = none
- Microscopy Metadata (including Provenance and Quality Control) conforming to 4DN-BINA-OME data model
- Browsable probe map, (bed file, see example)
- Probe sequences, (fasta file, see example)

2.14.4 Useful information

OME-NGFF and OME-Zarr

- https://www.biorxiv.org/content/10.1101/2021.03.31.437929v4
- https://zarr.readthedocs.io/en/stable/

Browsable probe map, example bed file

```
track name="AllRegions" description="mm10 AllRegions" visibility=1 itemRgb="On"
chr12 113100000 113130000 IgH_001 1 + 113100000 113130000 255,0,0
chr12 113130001 113160001 IgH_002 1 + 113130001 113160001 255,14,0
chr12 113160002 113190002 IgH_003 1 + 113160002 113190002 255,28,0
chr12 113190003 113220003 IgH_004 1 + 113190003 113220003 255,42,0
...
```

Probe sequence, example fasta file

```
>FwdPrimer01__BarcodeName__SecondBarcodeName__probeTargetName_p001__RevPrimer01
GCGGGACGTAAGGGCAACCGcatcaacgccacgatcagctGCTATCGTTCGTTCGAGGCCaggcaattcgagtggcgccctcgaagacgtctcgcaccttCCG
>FwdPrimer01__BarcodeName__SecondBarcodeName__probeTargetName_p002__RevPrimer01
GCGGGACGTAAGGGCAACCGcatcaacgccacgatcagctGCTATCGTTCGTTCGAGGCCagactttggaagccaccctcattgattgctcgtgctccatCCG
...
```

Example published / available data sets

- Wang…Zhuang 2016, *Science* (IMR90)
- Bintu,Mateo…Boettiger,Zhuang, 2018, *Science* (IMR90, K562, A549, HCT116)
- Mateo…Boettiger 2019, *Nature* (mESC + D. mel)
- Liu…Wang 2020, *Nat. Com.* (mouse liver)
- Saw…Wang,Mango 2020, *Mol Cell* (C. elegans)
- Su…Bintu,Zhuang 2020 *Cell* (IMR90)
- Takei…Cai 2021 *Nature* (mESC)
- Takei…Cai 2021 *bioRxiv* (mouse brain)
- Wiggins…Boettiger,Crabtree. 2021 *NSMB*, (mESC)

Example Tables

[Other publications with potentially accessible and similar data to aggregate]

- Bintu and Ren Sox2 paper
- Nir…Wu 2018, (localization data is published: https://data.4dnucleome.org/experiment-set-replicates/4DNESQN4JCAS/ - but data format discussion ongoing)
- Wu lab FISSEQ Nat. Methods chr tracing paper,
- Joyce lab (mostly STORM so far?)
- Nollman lab data