

File Format: CLF (Affymetrix Power Tools v1.6.1)

The CLF (cel layout file) maps probe IDs to a particular location in the CEL file. The CLF file is based on version 2 of the [TSV file format](#).

The CLF file is based on version 1 of the [TSV file format](#) .

Specs & Features

- Required headers
 - chip_type: indicates the chip type (as stored in the CEL file) which the clf file describes. Multiple chip_type headers may be present indicating the clf file can be used with any one of the listed chip types.
 - lib_set_name: indicates the name of a collection of related library files for a given chip. For example, PGF and CLF files intended to be used together should have the same lib_set_name. Only a single lib_set_name is allowed.
 - lib_set_version: indicates the version of a collection of related library files for a given chip. For example, PGF and CLF files intended to be used together should have the same lib_set_version. Only a single lib_set_version is allowed.
 - clf_format_version: currently the only documented and supported version is 1.0.
 - rows: the number of rows in the CEL file (1 base)
 - cols: the number of cols in the CEL file (1 base)
 - header0: indicates the values in the CLF file. A CLF file must contain probe_id, x, and y columns. The probe_id column must be unique. Other columns may be present.
- Optional headers
 - sequential: When present, this header indicates that the relationship between probe IDs and x/y location is deterministic. The value of this key is the starting ID. See note below about Deterministic IDs. If sequential is specified, then the order header must also be supplied. The value must be an integer ≥ 0 .
 - order: Indicates whether the IDs are enumerated in a column or row major order. Valid values are "col_major" and "row_major". If order is specified, then sequential must also be specified.
 - create_date: timestamp for when the file was created
 - guid: a unique identifier for the fileA
 - other headers may be present
- Required columns
 - per TSV format, order of columns is not guaranteed
 - per TSV format, additional columns may be present
 - probe_id: A positive integer ID for the probe. Must be unique within the CLF file.
 - x: the x position of the probe in the CEL file. (0 base)
 - y: the y position of the probe in the CEL file. (0 base)

Deterministic IDs

In many cases, probe IDs are simply enumerated starting with a fixed number by walking through the CEL file in a column or row major order. When this is the case, hints may be provided in the headers which allow the location to be computed from the probe_id. This saves the work of reading the file. When sequential is specified along with col_major order, the the following determines probe IDs:

```
probe_id = y*cols + x + sequential
```

And for row_major:

```
probe_id = x*rows + y + sequential
```

Parsing and Writing

The official C++ parser used by affy can be found in APT under sdk/file/TsvFile/ClfFile.h. When possible, parsing and writing of CLF files should be done using this code.

Notes

ChipLayout used by APT (in sdk/chipstream) will currently only accept CLF files for which:

- sequential=1
- order=col_major

Example 1

```
##chip_type=HuEx-1_0-st-v2
##chip_type=HuEx-1_0-st-v1
##chip_type=HuEx-1_0-st-ta1
##lib_set_name=HuEx-1_0-st
##lib_set_version=r2
##create_date=Tue Sep 19 15:18:05 PDT 2006
##guid=0000008635-1158704285-0732263232-1857033251-0689718480
##clf_format_version=1.0
##rows=2560
##cols=2560
##sequential=1
##order=col_major
##header0=probe_id      x      y
1      0      0
2      1      0
3      2      0
```

4	3	0
5	4	0
6	5	0
7	6	0
...		

Related Pages

- [TsvFile Design Notes](#)
- [TSV File Format](#)

