**3.4.2023**

**Parkour Meta Data – Nanopore Postprocessing**

The following data is recorded in Parkour regarding “Library Protocols”; “Library Types” and “Index Types”.

**Library Protocols** contain information on “Type” of nucleic acid (What do we receive, DNA or RNA?), Provider of the consumables, Catalogue number of the “Kit”, information on “Typical application” and a “Status” (obsolete or non-obsolete)



Schema 1: Example Library Protocols

Different protocols lead to different **library types**. For long read sequencing we have three different types of libraries that are ultimately being sequenced: Nanopore DNA library; Nanopore cDNA library and Nanopore RNA library. Library types are mapped to different protocols. One library type can be mapped to different protocols



Schema , Example Library type

If libraries contain indeces information on index sequence is contained in **index types**. To each index type respective i7 and i5 indeces are mapped. I7 and I5 are technical terms from “short read” sequencing. Long read sequencing libraries typically contain a single index, recorded under “I7”.



Schema Example index type

**Streamlining Nanopore Data Postprocessing and Data release**

The “fields” needed for automated data handling are existing and can be used by post processing scripts.

**Is DNA or RNA sequenced?**

Information is encoded in library type:

* Nanopore DNA and Nanopore cDNA library: “DNA” is sequenced.
* Nanopore RNA library “RNA” is sequenced.

**Which protocol was used?**

For nanopore data processing the correct name of the library preparation kit is needed. In Parkour a protocol can be called “PoreC v2” but for library preparation the library protocol “Nanopore DNAseq (SQK-LSK114)” was used. Explanation: a “protocol” consists of a sequence of different steps, one being the “library preparation” mainly used for adapter/index insertion.

Adjustments in Parkour (discussion points)

* Adjust in Library type the protocol name; use the exact “nanopore” name if possible. For improved overview we use as protocol name “Nanopore cDNA PCR”, correct would be “Nanopore PCR-cDNA Barcoding Kit”
* Catalogue or version in protocol name. Currently we use the catalogue number as part of the protocol name for instance: “Nanopore cDNA PCR (SQK-PCB111.24)”. This should help our users to follow protocol updating. Versioning is not very much in use but should be a handy tool for our custom workflows. For instance rename “Nanopore Custom ncRNA” in “Nanopore non-coding RNA v1”.
* Catalogue: Use the catalogue field with the exact catalogue number of the library preparation kit. For instance replace “Custom + SQK\_PCB109” by “SQK-PCB109”. Or the workflow “Nanopore Pore-C v2” contains simply the “SQK-LSK114” as entry in the catalogue fiked. Postprocessing scripts can query this filed.

**Which barcoding was used?**

This information can easily be retrieved from the field “Index Type”. Often library preparation protocol is also barcoding kit; but might not be true for all cases.

**Summary**

I believe that all information needed for nanopore postprocessing is in Parkour and we do not need new fields. We might need to rework our naming conventions including strict rules for usage of catalogue numbers and versioning. The field “catalogue” could play a more important role in future?