# Data analysis procedure for SMURF-seq reads

## Contents

# 1 Introduction

SMURF-seq is a protocol to efficiently sequence short DNA molecules on a long-read sequencer by randomly ligating them to form long molecules. The SMURF-seq protocol involves cleaving the genomic DNA into short fragments. These fragmented molecules are then randomly ligated back together to form artificial, long DNA molecules. The long re-ligated molecules are sequenced following the standard MinION library preparation protocol. After (or possibly concurrent with) sequencing, the SMURF-seq reads are mapped to the reference genome in a way that simultaneously splits them into their constituent fragments, each aligning to a distinct location in the genome (for most fragments).

This manual explains how to map SMURF-seq reads, generate copy-number profiles from the mapped fragments, and perform additional optional analysis of the sequenced read or the mapped fragments.

# 2 Mapping SMURF-seq reads

At this point, we assume that the reads generated from a SMURF-seq experiment are base-called and are in a fastq or fasta file.

The reads sequenced using SMURF-seq protocol needs to be mapped to the reference genome to identify the fragment locations. The reads can be aligned leveraging long-read mapping tools that are designed for split-read alignment.

We present the option of aligning SMURF-seq reads using BWA-MEM, Minimap2, and LAST, and we present several parameter recommendations for each tool. A user has an option of using either of these tools following the procedure in section aaa, bbb, or ccc. We recommend using BWA-MEM as this produced higher fragment counts.

## 2.1 Prerequisites

**Software required:**

**Environment variables:**

## 2.2 Mapping SMURF-seq reads to the reference genome

### 2.2.1 Mapping SMURF-seq reads with BWA-MEM

**Reference genome index creation**

**Aligning SMURF-seq reads**

**Parameter recommendations**

### 2.2.2 Mapping SMURF-seq reads with Minimap2

**Reference genome index creation**

**Aligning SMURF-seq reads**