9. Supplementary Material

9.1. Optional reformatting of read names after `flexbar`

```
zcat <sampleX.fasta.gz> |

  awk '{if(FNR%4==1) {
            if (NF==1) {
                print;
            } else {
                n=split($NF,v,"_");
                printf "%s_%s",$1,v[n];
                for (i=2;i<NF;i++) {
                    printf " %s",$i;
                }
                printf " ";
                for (i=1;i<(n-1);i++) {
                    printf "%s_",v[i];
                }
                print v[n-1];
            }
        } else {
            print
        }
  }' | gzip > <sampleX.fasta.gz>
```

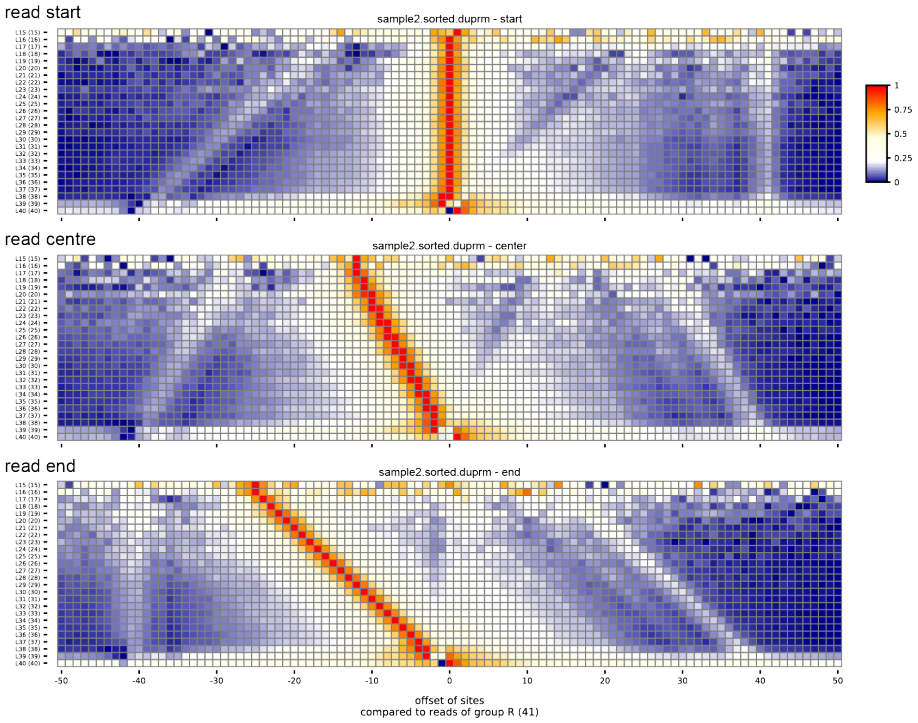
9.2. Supplementary Data Files

Supplementary Data 1: `Bash` code for all steps from raw reads until peak calling. This file provides the `bash` code for basic read processing, conversion into crosslink events and peak calling as described in Chapters 3-5.1. Details on the input files, preset variables and external tools required to run this code are listed in Chapters 3.1 and 2.1.

Supplementary Data 2: `R` code for postprocessing of `PureCLIP` output. This file provides the `R` code to postprocess the output of `PureCLIP` peak calling as described in Chapter 5.2. Running this code requires a bed file with 'crosslink sites' output by `PureCLIP` (`PureCLIP.crosslink_sites_short.bed`) and `bw` files with crosslink events (`sampleX.strand.bw`). The code to obtain these files is described in Supplementary Data 1 and Chapter 4.2.

Supplementary Data 3: `R` code for reproducibility and downstream analysis. This file provides the `R` code to reproducibility analyses and assignment of genes and transcript regions as described in Chapters 6.1 and 6.2. Running this code requires a `bed` file with binding sites (e.g. curated `PureCLIP` output, see Chapter 5), a `gtf` file with gene/transcript annotations, and `bw` files with crosslink events in the individual replicates (`sampleX.strand.bw`, see Chapter 4.2).

9.3. Supplementary Figure



Supplementary Figure 1: High-resolution read overlap heatmaps for sample 2 in our dataset, generated with `iCLIPro`. x-axis shows distance of start, centre and end positions of iCLIP reads that were trimmed to a given length (y-axis) relative to the start positions of reads that were not trimmed as reference positions (41-nt length, see Figure 5). In absence of read-end constraints, most reads starts align at the reference positions, irrespective of their length [9], whereas read centres and ends form a diagonal reflecting the read length after trimming. For more details on the underlying analysis steps and the visualisation, see [20] and `http://www.biolab.si/iCLIPro/doc/`.