

## ER-flow Application Description Template

|  |
|--|
| <b>Application Name:</b><br>Call insertions and deletions from sequence alignments (INDEL calling)   |
| <b>Application domain:</b><br>bioinformatics, life-sciences, DNA sequencing  |
| <b>Brief description of application</b><br>Determine insertions, deletions, inversions, and other larger variants in a DNA sequencing experiment.<br><br>data:<br>input data format: Raw sequence fragments (fastq format), reference database (fasta format), parameters<br>output data format: list of variants<br>sample data: (link) tbd<br>application (link) Dindel and Pindel<br>documentation (link) <a href="http://www.sanger.ac.uk/resources/software/dindel/">http://www.sanger.ac.uk/resources/software/dindel/</a> ,<br><a href="https://trac.nbic.nl/pindel/">https://trac.nbic.nl/pindel/</a><br>publication (link) <a href="http://www.ncbi.nlm.nih.gov/pubmed/20980555">http://www.ncbi.nlm.nih.gov/pubmed/20980555</a> ,<br><a href="http://www.ncbi.nlm.nih.gov/pubmed/19561018">http://www.ncbi.nlm.nih.gov/pubmed/19561018</a> , |
| <b>Execution environment</b><br>DCI: (computing, data, VO, etc)<br>middleware: pbs, glite (vlemed) workflow system: guse   |
| <b>Execution characteristics</b><br>data size (per unit, typical number of units):<br>input 1-100GB temporary tbd output tbd<br>processing time (per unit): tbd<br>memory usage: tbd disk usage: tbd   |
| <b>Target users</b><br>Community, projects: (link) number of users:<br>Biomedical and bioinformatics users<br>Of general interest to researchers that re-sequence human or other genomes<br>user type: developer and end-user  |
| <b>Usage scenario for workflow in the ER-FLOW</b> (how workflow will be reused, metaworkflow, how expected to contribute to project indicators, etc.).<br><br>Dindel application will be used for determining structural variants in single-end sequence runs, and Pindel will be used for paired-end sequence runs.<br><br>First the workflows will be developed and disseminated among the advanced users, then these will become available for end-users at the AMC ebioinfra gateway.<br>We do not foresee end-users running workflows from the SHIWA platform   |
| <b>Contact information</b> (author)<br>name: Barbera van Schaik<br>e-mail <a href="mailto:b.d.vanschaik@amc.uva.nl">b.d.vanschaik@amc.uva.nl</a> , <a href="mailto:support@ebioscience.amc.nl">support@ebioscience.amc.nl</a><br>Bioinformatics Laboratory, KEBB, Academic Medical Center, Amsterdam, NL   |