

## ER-flow Application Description Template

<p><b>Application Name:</b> SNP annotation with Annovar</p>									
<p><b>Application domain:</b> bioinformatics, life sciences, dna sequencing</p> <p><b>Brief description of application</b> Adds annotation to variant lists, e.g. gene name, prediction of damage, conservation, known variants</p> <p>data:</p> <p style="padding-left: 20px;">input data format: annotation database, variants (VCF format)</p> <p style="padding-left: 20px;">output data format: text, csv files</p> <p>sample data: tbd application (link) <a href="http://www.openbioinformatics.org/annovar/">http://www.openbioinformatics.org/annovar/</a> documentation (link) see above publication (link) Wang K, Li M, Hakonarson H. <a href="#">ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data</a> <i>Nucleic Acids Research</i>, 38:e164, 2010</p>									
<p><b>Execution environment</b> DCI: (computing, data, VO, etc) grid, cluster middleware: glite, pbs workflow system: moteur, ws-pgrade/guse</p>									
<p><b>Execution characteristics</b></p> <p>data size (per unit, typical number of units):</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 33%;">input</td> <td style="width: 33%; text-align: center;">temporary</td> <td style="width: 33%; text-align: right;">output</td> </tr> <tr> <td colspan="3">input: ~10MB (SNP file), 1.4GB (database)</td> </tr> <tr> <td colspan="3">output: ~50MB (annotated variants)</td> </tr> </table> <p>processing time (per unit): tbd memory usage: tbd disk usage:</p>	input	temporary	output	input: ~10MB (SNP file), 1.4GB (database)			output: ~50MB (annotated variants)		
input	temporary	output							
input: ~10MB (SNP file), 1.4GB (database)									
output: ~50MB (annotated variants)									
<p><b>Target users</b> Community, projects: (link) number of users: community, related projects: bioinformatics, dna sequencing number: of general interest to researchers working with genome re-sequencing data profile (programmers, end-users): developers and end-users</p> <p>user type: developer and end-user</p>									

**Usage scenario for workflow in the ER-FLOW** (how workflow will be reused, metaworkflow, how expected to contribute to project indicators, etc.).

Especially useful for annotating many datasets, like in population studies

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.

We do not foresee end-users running workflows from the SHIWA platform

**Contact information** (author)

name:

e-mail

Barbera van Schaik, Bioinformatics Laboratory, AMC, Amsterdam, the Netherlands

b.d.vanschaik@amc.uva.nl