**Quality control survey for GEUVADIS, WP2**

i) Partner ID (Short name and number)

ii) Name(s):

iii) Sequencing technology/system? How many machines of each system do you have?

(e.g. one HiSeq2000, two SOLiD4, three 454)

iv) Capacity? Approximately how many samples per week do you run? (or number of runs?)

**RNA sequencing**

**Sample prep**

1. Which RNA extraction method do you use or recommend
   1. For small RNA?

* 1. For mRNA?

1. Do you use DNA/phiX control and/or spike in?

1. What kind of QC do you perform on the RNA material? (RIN, concentration)

1. How do you quantify the input RNA?

1. What measures do you take to minimise laboratory errors?

(e.g. use of automatisation robots, LIMS, GLP compliance)

**Library prep, RNA**

1. Which Library preparation protocol …
   1. … for small RNA?

* 1. … for mRNA?

1. What fragmentation method(s) do you use?

1. How do you check your library QC? (concentration, gel, test runs?)

1. How do you quantify the DNA-library and ensure equimolar pooling when multiplexing ?

**Data quality, RNA… “How do you assure good alignment? ” etc..**

1. If defined, what is your tolerated sequencing error rate? (derived from errors in control DNA/phiX sequence)

1. Describe your mostly used RNA-Seq data analysis pipeline(s). Mention what they are, when you use it (eg. ‘for differential expression analysis’, ‘allele specific expression’, ‘splice variant/fusion transcripts detection in cancer’ etc.)  
   Please try to include information on ..

Read length and what coverage and number of mappable reads you aim for during sequencing?

Softwares and parameters used, e.g. for base calling, alignment, expression levels etc.?

To what reference is the reads aligned (masked in any way, pseudogenes, repeats)?

Filtering of reads? (based on what? Quality scores, alignments, read duplicates?)

Normalization method used?

Method for estimate/determine average sequence coverage?

1. How is the confidence score cut-off determined for differential expression levels?

(e.g. such that the top 30% remains)?

1. Reproducibility?

Do you e.g. use replicates for comparison (biological or technological replicates)?

**Exome sequencing**

**Library prep, Exome-sequencing**

1. Which Exome-enrichment protocol(s) do you use / recommend?

1. Which Library preparation protocol(s) do you use / recommend?

1. Do you use DNA/phiX control (or equivalent)?

1. How much DNA do you need for QA purposes? What kind of QC do you do on the input DNA?

1. What fragmentation method(s) do you use)?

1. Library QC (concentration, gel, test runs, staining in flowcell)?

1. How do you quantify the DNA-library and ensure equimolar pooling when multiplexing ?

1. What measures do you take to minimise laboratory errors?

(e.g. use of automatisation robots, LIMS, GLP compliance)

**Data quality, exome Sequencing.. “How do you assure good alignment?”**

1. Please describe briefly your mostly used Exome-Seq data analysis pipeline.   
   Try to include information on ..

What coverage and number of reads you aim for during sequencing?

Softwares and parameters used for alignment, variant calling etc.?

Reference used for alignment? Normalisation methods?

Filtering of reads? (based on what? Read duplicates, pairwise alignment, quality scores?)

Method for estimate/determine average sequence coverage?

1. Do you compare and consider differences in read depths between case and control samples?

1. Exome enrichment: Do you assess the …
   1. … enrichment rate of exomes across the genome?

* 1. … reproducibility of the enrichment rates at a given genomic region across samples?

1. Additional comments?