Transcriptomic Age Prediction - Tutorial

GTTACACATGGACTGGAGATCACACOMPLEX GENETICSGTGTACGAATTTTAAGGGGATAGT
GTCAGCTTAAGENETIC LABORATORYATGGTCACCATGATCGATGGTTACGAACTTAAGAAA
TACCGAAATTCGGATGINTERNAL MEDICINEGTACTTGGCCCATGAGCTGAGTTCCCAAACGG
CCGATTGCACCATAATCGCGACCAATATGWASGCATAATAGCCGGGSNPATTGCTGAGATAC
GATAACCTTTAAACACACAEXOME SEQUENCINGGATCCGTAGGGACCGCGAATCCGTTTTTT

TUTORIAL for Transcriptomic Age Prediction (TRAP)

Marjolein J. Peters - m.peters@erasmusmc.nl

This tutorial illustrates how to calculate Transcriptomic Age using the online calculator.

Mandatory input: A gene expression file, e.g. measured on the Illumina HumanHT-12 (v3 or v4) Expression BeadChip or the Affymetrix Human Exon 1.0 ST GeneChips. The file should be tab separated (.txt), and the gene expression levels should be normalized and standardized: we prefer quantile-normalization to the median distribution, log2 transformation, probe centering, and sample Z transformation. Additionally, your gene expression data should be adjusted for a number of covariates (if available):

- Sex (0=male, 1=female)
- Fasting (0=fasting, 1 = non-fasting)
- Smoking (0= non-smoking, 1 = smoking)
- RNA quality score (RQS or RIN)
- Batch effects (*e.g.,* plate ID)
- Cell counts (#granulocytes, #lymphocytes, #monocytes, #erythrocytes, #platelets)

The easiest way to do so is to calculate the residuals:

residuals[,i] <- lm(expression[,i] ~ as.factor(sex]) + as.factor(fasting) + as.factor(smoking) + RQS + as.factor(batch) + nrgranulocytes + nrlymphocytes + nrmonocytes + nrerythrocytes + nrplatelets, na.action=na.omit)$residuals

The residuals can be used as gene expression input files for the Transcriptomic Age Prediction.

The gene expression file needs to contain either Illumina IDs (ILMNID.txt) or gene symbols (GENEID.txt). The header should contain "ILMNID" or "GENEID" (both with capitals) and non-numeric

sample IDs (i.e. "Sample1", "Sample2", *etc.*) for which the gene expression levels were measured. Genes that were not measured in your samples should be included as zeros.

Example gene expression file using Illumina IDs:

ILMNID	Sample1	Sample2	Sample3	Sample4	*etc.*
ILMN_1651209	0.299273	0.303709	-0.16879	0.292448	
ILMN_1651229	0.565635	1.459124	-0.95011	-0.0631	
ILMN_1651254	2.330158	0.080845	0.373774	-0.73325	
etc.					

Example gene expression file with Gene IDs:

ILMNID	Sample1	Sample2	Sample3	Sample4	*etc.*
A1BG	-0.63126	0.9331	0.525583	-0.65303	
A1CF	0.088495	0.004431	0.358768	0.842198	
AAAS	1.617508	-0.43129	0.562675	1.374531	
etc.					

The age file needs to contain the chronological age for each sample. The header should contain "SAMPLEID" and "AGE" (both with capitals), and the sample IDs should be identical to the ones used in the gene expression file. The age file should have samples as rows, and age as column.

Example Chronological Age file:

SAMPLEID	AGE
Sample1	50.6
Sample2	57.3
Sample3	85.8
Sample4	63.5
etc.	

Make sure that the sample IDs in the chronological age file are identical to the sample IDs used in the gene expression file.

Start your job:

Go to the tab "Define new job", and complete the online form:

- **Name** = give a name to your analysis
- **Expression type** = select the type of identifier used: *Gene Symbols* or *Illumina IDs*?
- **Expression file** = choose your *tab-delimited* gene expression input file
- **Predictor type** = select the type of predictor: *GENERAL* or *SCALED*?
- **Age file** = if you selected the *GENERAL* predictor, you need to choose your chronological age input file. If you selected the *SCALED* predictor, you can leave this file location empty.

Results:

After submitting your files, these will be uploaded to our data analysis server.

Under the tab "Job Administration", the status of your analysis is shown. If your analysis is finished, the status will be "finished", and you can download the "results" file.

If you ran the GENERAL predictor (for which you uploaded chronological age), you will obtain the following variables:

- **CA** = Chronological Age
- **TA** = Transcriptomic Age
- **DA** = Delta Age (the difference between the predicted TA and CA)

If you ran the SCALED predictor, you will only obtain Transcriptomic Age (**TA**).