



UvA-DARE (Digital Academic Repository)

Chronological age prediction based on DNA methylation

Massive parallel sequencing and random forest regression

Naue, J.; Hoefsloot, H.C.J.; Mook, O.R.F.; Rijlaarsdam-Hoekstra, L.; van der Zwalm, M.C.H.; Henneman, P.; Kloosterman, A.D.; Verschure, P.J.

DOI

[10.1016/j.fsigen.2017.07.015](https://doi.org/10.1016/j.fsigen.2017.07.015)

Publication date

2017

Document Version

Other version

Published in

Forensic Science International. Genetics

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Naue, J., Hoefsloot, H. C. J., Mook, O. R. F., Rijlaarsdam-Hoekstra, L., van der Zwalm, M. C. H., Henneman, P., Kloosterman, A. D., & Verschure, P. J. (2017). Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. *Forensic Science International. Genetics*, 31, 19-28. <https://doi.org/10.1016/j.fsigen.2017.07.015>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Chronological Age Prediction based on DNA Methylation: Massive Parallel Sequencing and Random Forest Regression (Naue et al.)

Supplementary Tables S1, S3, S5 and Figures S1, S2, S3
(Tables S2 and S4 are provided as separate files)

Table S1 Selected public available 450K microarray datasets. Public available GSE files containing raw 450K microarray data were considered and divided into “whole blood”, and “buffy coat”. Not all samples of each study were used due to different reasons mentioned in the remarks column.

Characteristic	Reference	Samples in GSE	Whole blood	Buffy coat	Remarks
GSE42861	Liu et al. 2013 [1]	689	335		only healthy control individual used
GSE51032	<i>not published</i>	845		641	samples removed due to MethylAid QC fail
GSE55491	Prickett et al. 2015 [2]	24		6	only healthy control individual used
GSE61496	Tan et al. 2014 [3]	300		150	only one twin used
GSE66459	Fernando et al. 2015 [4]	22	11		cord blood, only healthy control individual used
GSE66552	Strong et al. 2015 [5]	45	15		age from author
GSE67444	<i>not published</i>	70	33		multiple measurements per individual and removal of samples with failed QC
GSE74548	Kok et al. 2015 [6]	174		55	samples removed due to MethylAid QC fail and multiple measurements
Total number of samples			394	852	
Min. age			0	9	
1st Qu. Age			37	47	
Median age			53	55	
Mean age			45.28	53.54	
3rd Qu. Age			60	60	
Max. age			70	75	
female			268	581	
male			126	271	

- [1] Y. Liu, M.J. Aryee, L. Padyukov, M.D. Fallin, E. Hesselberg, A. Runarsson, L. Reinius, N. Acevedo, M. Taub, M. Ronninger, K. Shchetynsky, A. Scheynius, J. Kere, L. Alfredsson, L. Klareskog, T.J. Ekström, A.P. Feinberg, Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis, *Nat. Biotechnol.* 31 (2013) 142–147. doi:10.1038/nbt.2487.
- [2] A.R. Prickett, M. Ishida, S. Böhm, J.M. Frost, W. Puszzyk, S. Abu-Amero, P. Stanier, R. Schulz, G.E. Moore, R.J. Oakey, Genome-wide methylation analysis in Silver–Russell syndrome patients, *Hum. Genet.* 134 (2015) 317–332. doi:10.1007/s00439-014-1526-1.
- [3] Q. Tan, M. Frost, B.T. Heijmans, J. von Bornemann Hjelmberg, E.W. Tobi, K. Christensen, L. Christiansen, Epigenetic signature of birth weight discordance in adult twins, *BMC Genomics.* 15 (2014) 1062. doi:10.1186/1471-2164-15-1062.
- [4] F. Fernando, R. Keijser, P. Henneman, A.-M.F. van der Kevie-Kersemaekers, M.M. Mannens, J.A. van der Post, G.B. Afink, C. Ris-Stalpers, The idiopathic preterm delivery methylation profile in umbilical cord blood DNA, *BMC Genomics.* 16 (2015) 736. doi:10.1186/s12864-015-1915-4.
- [5] E. Strong, D.T. Butcher, R. Singhanian, C.B. Mervis, C.A. Morris, D. De Carvalho, R. Weksberg, L.R. Osborne, Symmetrical dose-dependent DNA-methylation profiles in children with deletion or duplication of 7q11.23, *Am. J. Hum. Genet.* 97 (2015) 216–227. doi:10.1016/j.ajhg.2015.05.019.
- [6] D.E.G. Kok, R.A.M. Dhonukshe-Rutten, C. Lute, S.G. Heil, A.G. Uitterlinden, N. van der Velde, J.B.J. van Meurs, N.M. van Schoor, G.J.E.J. Hooiveld, L.C.P.G.M. de Groot, E. Kampman, W.T. Steegenga, The effects of long-term daily folic acid and vitamin B12 supplementation on genome-wide DNA methylation in elderly subjects, *Clin. Epigenetics.* 7 (2015) 121. doi:10.1186/s13148-015-0154-5.

Chronological Age Prediction based on DNA Methylation: Massive Parallel Sequencing and Random Forest Regression (Naue et al.)

Table S3 Most informative 15 markers per dataset. The top 15 selected markers per dataset are shown. The top 6 markers are highlighted in yellow (only one marker per gene). The 15 final chosen markers are in bold. The dataset "stratified whole blood" and "whole blood without cord blood" represent subsets of the "whole blood dataset"

Whole blood		Buffy coat		Stratified whole blood		Whole blood without cord blood	
450K marker	Gene	450K marker	Gene	450K marker	Gene	450K marker	Gene
cg16867657	ELOVL2	cg16867657	ELOVL2	cg16867657	ELOVL2	cg16867657	ELOVL2
cg12934382	GRM2	cg08097417	KLF14	cg12934382	GRM2	cg12934382	GRM2
cg11807280	MEIS1_AS3	cg06784991	ZYG11A	cg13959344		cg06784991	ZYG11A
cg02872426	DDO	cg07553761	TRIM59	cg11807280	MEIS1_AS3	cg16054275	F5
cg06874016	NKIRAS2	cg16054275	F5	cg02872426	DDO	cg18473521	HOXC4
cg08097417	KLF14	cg16015712	ZYG11A	cg03224418	SAMD10	cg08262002	LDB2
cg24987259	CTSF	cg18473521	HOXC4	cg18473521	HOXC4	cg13959344	
cg05404236	IRS2	cg22156456	EIF1	cg26153045	ELN	cg17372101	CNTNAP2
cg26153045	ELN	cg21776419		cg07553761	TRIM59	cg01528542	
cg19722847	IPO8	cg01100784	MAPK8IP3	cg25410668	RPA2	cg10835286	
cg18473521	HOXC4	cg16690859	PLXNA4	cg16054275	F5	cg03224418	SAMD10
cg01528542		cg05093315	SAAL1	cg15893346	GUSB	cg25410668	RPA2
cg25410668	RPA2	cg02228185	ASPA	cg01528542		cg09243824	EHD2
cg04080625	KIAA1026	cg24847230	UBE2Z	cg05404236	IRS2	cg05156137	RCAN1
cg07553761	TRIM59	cg05156137	RCAN1	cg22156456	EIF1	cg12580096	C19orf57
cg22160883	MATR3	cg03293770	SDHAF2	cg20988565	ZFPM2	cg05404236	IRS2

Chronological Age Prediction based on DNA Methylation: Massive Parallel Sequencing and Random Forest Regression (Naue et al.)

Table S5 SNP detection for the 13 age-dependent markers. Observed SNP sites and number of affected samples. * SNP not present in dbSNP150. Exact location within the amplicon can be found within Figure S2.

Marker/ Gene	Samples with SNP / total samples (training set, test set)	dbSNP 150	Direct effect on CpG sites
F5	64/208, 32/104	rs2269648	none
GRM2	17/208, 6/104	rs149387441	none
KLF14	1/104	rs542947804	loss of another CpG site, new CpG site ahead
NKIRAS	3/208, 1/104	rs34326614	additional CpG site
SAMD10	7/104, 11/208	rs75097987	none
ZYG11A	15/104, 36/208	rs533935	additional CpG site
	87/104, 177/208	rs534070	none
	0/208, 1/104	rs1012331325	none
	0/208, 1/104	53308826 G>A*	none

Chronological Age Prediction based on DNA Methylation: Massive Parallel Sequencing and Random Forest Regression (Naue et al.)

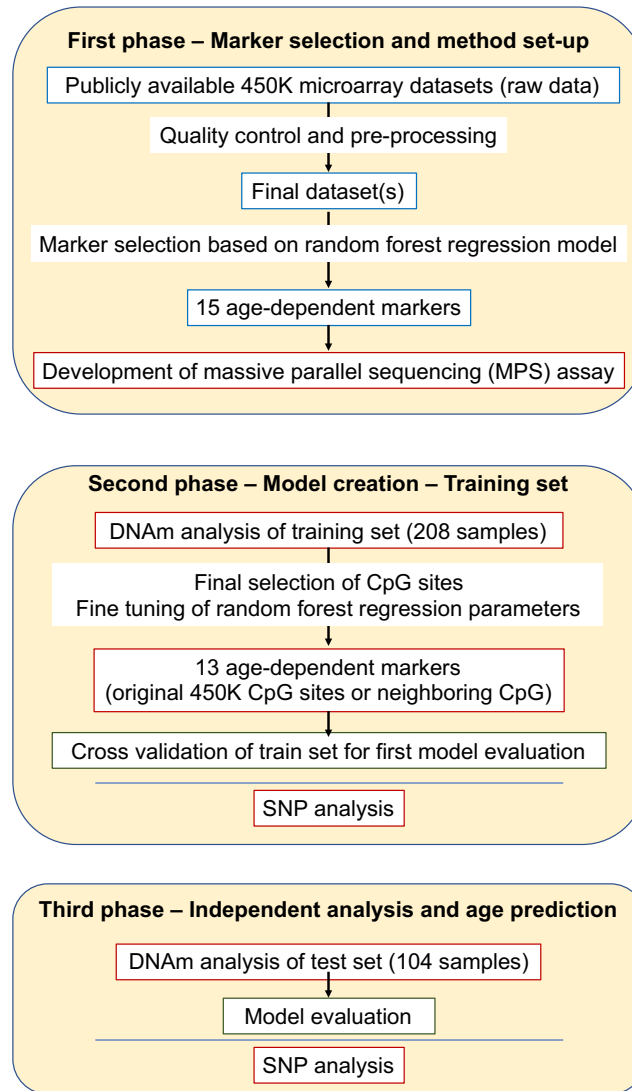


Figure S1. Project workflow. The project can be divided into marker selection, Marker testing and model creation as well as independent model evaluation

Chronological Age Prediction based on DNA Methylation: Massive Parallel Sequencing and Random Forest Regression (Naue et al.)

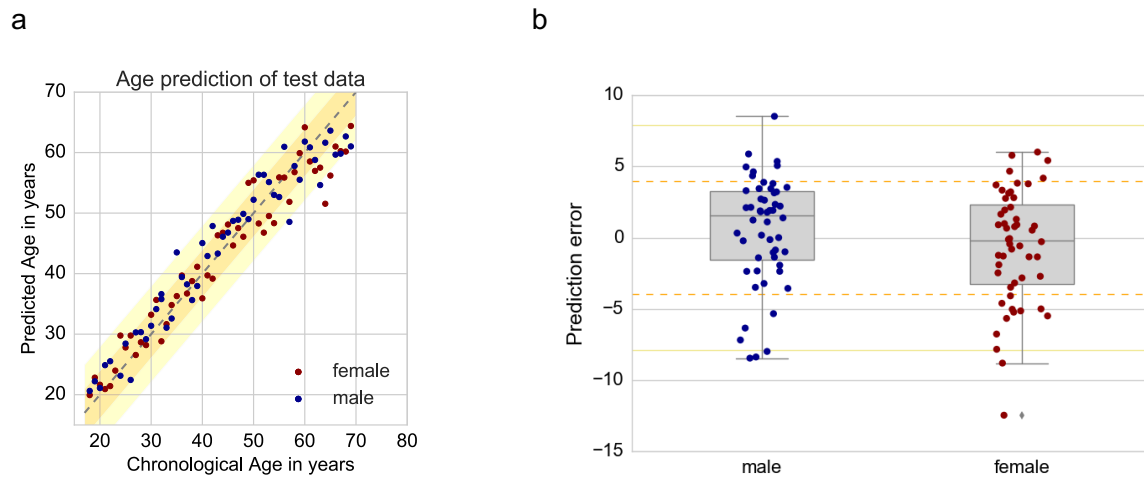


Figure S3. Age prediction for female and male individuals. a: Grouping of the samples according to their gender. b: Deviation from the chronological age grouped by gender. No statistically significance in the prediction error between female and male individuals was found. Nevertheless, a tendency for rather an overestimation of age in males can be seen. Vertical lines: RMSE 3.93 years (orange), 2xRMSE (yellow)