Supplementary Tables

Cell type	Cell line	Replicate	Linear regression		M-W younger vs older fathers		M-W younger fathers vs simulations		M-W older fathers vs simulations	
			p	β	p	β	p	β	p	β
Lymphoblastoid cells	6 lines	-	**0,0022**	0,158	**1,3E-04**	-0,070	**4,9E-04**	-0,053	0,6839	0,004
	C0202	1	**0,0305**	0,110	**1,7E-03**	-0,042	**2,7E-03**	-0,033	0,83	0,002
	C0202	2	**0,0285**	0,111	**9,9E-04**	-0,050	**2,8E-04**	-0,046	0,70	-0,004
Neural precursor cell	BG01	1	**0,0265**	0,108	**1,5E-03**	-0,057	**9,5E-04**	-0,049	0,84	-0,002
	BG01	2	**0,0257**	0,111	**7,5E-04**	-0,051	**7,6E-04**	-0,042	0,93	-0,001
Embryonic stem cell	BG01	-	**7,7E-04**	0,171	**3,6E-04**	-0,057	**5,0E-05**	-0,055	0,37	-0,009
	BG02	1	**0,0036**	0,148	**5,2E-04**	-0,048	**4,4E-05**	-0,047	0,27	-0,009
	BG02	2	**0,0047**	0,146	**4,7E-03**	-0,044	**1,1E-03**	-0,042	0,35	-0,009
	H7	-	**0,0036**	0,146	**3,2E-03**	-0,051	**1,5E-04**	-0,055	0,16	-0,015
	H9	-	**0,0021**	0,155	**7,1E-04**	-0,050	**1,3E-05**	-0,054	0,15	-0,013
Induced pluripotent stem cell	iPS4	1	**6,1E-04**	0,172	**1,1E-03**	-0,054	**5,6E-05**	-0,056	0,31	-0,011
	iPS4	2	**0,0046**	0,143	**3,0E-04**	-0,065	**8,4E-05**	-0,059	0,68	-0,004
	iPS5	1	**8,9E-04**	0,167	**3,6E-04**	-0,060	**3,2E-04**	-0,051	0,93	-0,001
	iPS5	2	**8,7E-04**	0,171	**5,5E-03**	-0,045	**2,7E-03**	-0,040	0,65	-0,004

Supplementary Table 1. Paternal age effect on *de novo* mutation replication timing measured in 6 cell types

The source of the replication timing data was Koren *et* al. [1] for the 6 lymphoblastoid cell lines (1st row) and Ryba *et al*. [2] for all other cell lines. The linear regression column contains p-values and estimates (β) for the parsimonious model described in Methods. The M-W test columns contain the p-values and estimated difference using a Mann-Whitney test between the distribution of mutation replication timing values of: offspring of younger (<28 years old) vs older (≥28 years old); offspring of younger fathers vs simulations; offspring of older fathers vs simulations. Significant p-values are highlighted in bold font

	$P(\beta_{r,t} \neq 0)$
T→G	0.54
C→T	**0.00696**
G→C	0.0502
T→C	0.221
T→A	**0.025**
C→A	**6.06×10^{-8}**
CpG→TpG	**1.98×10^{-8}**

Supplementary Table 2. Predictive power of local substitution rates.

Predictive power of primate substitution rates for local *de novo* mutation rates using the Poisson regression model described in the Supplementary Note. Only S→W and W→W substitutions have significant predictive power for local *de novo* mutation rates.

	$\widehat{\beta_{\rho,t}}$	$P(\widehat{\beta_{\rho,t}} \neq 0)$	$\hat{f}_t$
G→A	-2.485×10^{-5}	0.23	1.034
T→A	2.597×10^{-6}	0.649	0.891
C→A	-9.857×10^{-5}	$< 2 \times 10^{-16}$	1.014
CpG→CpA/TpG	-0.00449	$< 2 \times 10^{-16}$	0.988

Supplementary Table 3. Dependency of local mutation rates on recombination rates

Summarizes the estimated dependency of local mutation rates on recombination rates. Only C→A and CpG→TpG exhibit a significant dependency.

Log Likelihood	**GoNL uniform**	**Primate $r_{t,i}$**	**Corrected $\mu_{t,i}$(%/%)***		
Recomb. rates	No	No	Male	Sex-Averaged	Female
log $L_{T>G}$	-1,591.0	-1,605.3	-1,584.6 (0.4% / 1.3%)	-1,584.6 (0.4% / 1.3%)	-1,584.6 (0.4% / 1.3%)
log $L_{G>A}$	-2,974.5	-2,961.7	-2,959.2 (0.5% / 0.1%)	-2,959.2 (0.5% / 0.1%)	-2,957.7 (0.6% / 0.1%)
log $L_{G>C}$	-1,735.6	-1,736.5	-1,728.3 (0.4% / 0.5%)	-1,728.4 (0.4% / 0.5%)	-1,728.4 (0.4% / 0.5%)
log $L_{T>C}$	-3,176.2	-3,183.5	-3,152.0 (0.8% / 1.0%)	-3,152.0 (0.8% / 1.0%)	-3,152.0 (0.8% / 1.0%)
log $L_{T>A}$	-1,439.8	-1,437.7	-1,434.5 (0.4% / 0.2%)	-1,434.5 (0.4% / 0.2%)	-1,434.5 (0.4% / 0.2%)
log $L_{C>A}$	-1,852.2	-1,836.7	-1,832.5 (1.1% / 0.2%)	-1,831.9 (1.1% / 0.3%)	-1,831.7 (1.1% / 0.3%)
log $L_{CpG>CpA/TpG}$	-2,696.2	-2,439.6	-2,435.8 (9.7% / 0.2%)	-2,435.0 (9.7% / 0.2%)	-2,434.9 (9.7% / 0.2%)
Total	**-15,465.6**	**-15,201.0**	**-15,126.9 (2.2% / 0.5%)**	**-15,125.6 (2.2% / 0.5%)**	**-15,123.8 (2.2% / 0.5%)**

* In parenthesis is the percent change of log likelihood of $\mu_{t,i}$ compared to GoNL uniform model and uncorrected primate rate model $r_{t,i}$.

Supplementary Table 4. Likelihood of the observed data under different mutation rate models

Likelihood of the observed *de* novo mutation data by substitution type based on (a) a uniform mutation rate model derived from the observed mutations, (b) the uncorrected primate rate matrix $\boldsymbol{r_{t,i}}$ and (c) the computed mutation rate matrix $\boldsymbol{\mu_{t,i}}$.

Supplementary Note

Mutation rate map

For each 1Mb window *i*, a substitution rate matrix was inferred using the context-dependent primate substitution model described in Duret *et al*. [3]for seven types of substitutions, parameterized by $r_{t,i}$ with *t* in {T→G, G→C, T→C, T→A, C→A, C→T, CpG→TpG (to account for hyper-mutability of CpG sites)}.

First, we tested if the observed *de novo* mutation rates co-vary with primate substitution rates across the genome using the following Poisson regression model with log link function:

$$\log(n_{t,i}) = \beta_{r,t} r_{t,i} + \beta_{r,t,0} + \log(N_{t,i}) \quad \text{for } t \neq \text{CpG→TpG}$$

$$= \beta_{r,t}(r_{t,i} + r_{C\to T,i}) + \beta_{r,t,0} + \log(N_{t,i}) \quad \text{for } t = \text{CpG→TpG}$$

where $n_{t,i}$ is the observed count of *de novo* mutations of type *t* in window *i*, $r_{t,i}$ is the substitution rate of type *t* in window *i,* and $N_{t,i}$ is the number of sites at which *de novo* mutations of type *t* can be detected with high confidence in window *i*. The offset term $\log(N_{t,i})$ was added since the number of called *de novo* mutations is dependent on detection power. The C→T mutation of CpG sites requires special treatment since it can be attributed to context-independent C→T substitution as well as hyper-mutability of CpG.

The primate substitution rates in the above Poisson regression model only had significant predictive power for local *de novo* mutation rates for S→W and W→W substitutions (Supplementary Table 2). For this reason, we only estimated local mutation rates based on the primate substitution rate for substitutions in t_{SW} = {T→A, C→A, C→T, CpG→TpG}. For the rest of substitutions (*t* in {T→G, T→C, G→C}), we used the genome-wide averaged mutation rates r'_t estimated from our observed mutations:

$$r'_t = \frac{\sum_i n_{t,i}}{\sum_i N_{t,i}} \cdot \frac{1}{c}$$

$$c = \frac{\sum_i \sum_{t \epsilon t_{SW}} n_{t,i}}{\sum_i \sum_{t \epsilon t_{SW}} r_{t,i} N_{t,i} + \sum_i r_{C\to T,i} N_{CpG\to TpG,i}}$$

where *c* is a scaling factor to convert between *de novo* mutation rates and instantaneous substitution rates.

Second, we corrected for the biases due to local recombination rates. The observed local *de novo* mutation rates were not significantly correlated with recombination rates when considering each type of substitutions separately (Bonferroni-corrected p-value > 0.05). However, local substitution rates $r_{t,i}$ depend significantly on local sex-averaged recombination rates ρ_i for *t* = C→A and CpG→TpG (Supplementary Table 3). To eliminate the dependency on recombination rate, we fit the following linear regression model:

$$r_{t,i} = \beta_{\rho,t}\, \rho_i + \beta_{0,t}$$

and residualized $r_{t,i}$ by subtracting the ρ_i-dependent term.

The final formula we used to compute the mutation rates for each 1Mb window *i* is then:

$$\mu_{t,i} = (r_{t,i} - \beta_{\rho,t}\, \rho_i) \cdot f_t \quad \text{for } t \text{ in } \{\text{C→A, CpG→TpG}\}$$

$\mu_{t,i} = r_{t,i} \cdot f_t$ for *t* in {T→A, C→T}

$\mu_{t,i} = r'_t$ for *t* in {T→G, T→C, G→C}

where f_t is a global scaling factor for substitution of type *t* to match the observed frequencies of different types of *de novo* mutations (Supplementary Table 4). In particular, A→T mutation is over-represented in primate substitutions by 12% compared to our *de novo* data. For each *t* in $t_{SW,}$ f_t is defined to satisfy the following conditions:

$\sum_i \mu_{t,i} N_{t,i} = \frac{1}{c} \sum_i n_{t,i}$ for *t* in {T→A, C→T, C→A}

$\sum_i (\mu_{t,i} + \mu_{C \to T,i}) N_{t,i} = \frac{1}{c} \sum_i n_{t,i}$ for *t* = CpG→TpG

Finally, the mutation rate *μ* was scaled so that the overall mutation rate across the autosome is 1.2 x 10^{-8} per nucleotide per generation.

To evaluate the fit of the estimated mutation rates to observed *de novo* mutations, we examined the likelihood L_t of the observed data given mutation rates, assuming homogenous Poisson process for each type of mutation *t* within each window *i*:

$$L_t(data|\mu_{t,i})$$

$$= \prod_i Poisson\left(n_{t,i} \middle| \lambda = \frac{1}{c'} \mu_{t,i} N_{t,i}\right) \quad \text{for } t \neq \text{CpG} \to \text{TpG}$$

$$= \prod_i Poisson\left(n_{t,i} \middle| \lambda = \frac{1}{c'} (\mu_{t,i} + \mu_{C \to T,i}) N_{t,i}\right) \quad \text{for } t = \text{CpG} \to \text{TpG}$$

$$c' = \frac{\sum_i \sum_t n_{t,i}}{\sum_i \sum_t \mu_{t,i} N_{t,i} + \sum_i \mu_{C \to T,i} N_{CpG \to TpG,i}}$$

The likelihood of the observed data under different models is summarized in Supplementary Table 4.

We estimated functional mutation rates in protein-coding region for autosomal protein-coding transcripts (downloaded from Ensembl [4] v74). Excluding 24,508 transcripts (3,808 genes) outside our analysis windows for bias correction, we computed bias-corrected mutations rates for a total of 54,310 transcripts (15,462 genes). For maximum coverage of genes, however, we provide two additional functional mutation rates based on uncorrected local primate substitution rates $r_{t,i}$ and the uniform genome-wide averaged mutation rates r'_t derived from our observed data.

For each transcript, the local mutation rate was determined by the 1Mb genomic window that overlapped the coordinate of midpoint between transcription start and end sites. Based on this rate, all possible nonsense, missense, synonymous and 4-fold degenerate synonymous mutations were examined with respect to the reference genome, and their mutation rates were aggregated over the entire transcript.

While we assumed the equal rate of $\mu_{A \to G,i}$ and complementary $\mu_{T \to C,i}$ in non-coding region, we adjusted for their strand bias in protein-coding region as follows:

$$\mu^{tx}_{A \to G,i} = \frac{N^{tx}_A + N^{tx}_T}{N^{tx}_A} \frac{\gamma_{sb}}{1 + \gamma_{sb}} \mu^{nc}_{T \to C,i}$$

$$\mu^{tx}_{T \to C,i} = \frac{N^{tx}_A + N^{tx}_T}{N^{tx}_A} \frac{1}{1 + \gamma_{sb}} \mu^{nc}_{T \to C,i}$$

$$\gamma_{sb} = \frac{\mathrm{n}^{tx}_{A\to G}}{\mathrm{n}^{tx}_{T\to C}}$$

where $\mu^{nc}_{T\to C,i}$ (= $\mu^{nc}_{A\to G,i}$) is the local mutation rate of A:T→G:C in non-coding region, $\mu^{tx}_{T\to C,i}$ and $\mu^{tx}_{A\to G,i}$ are the local mutation rates of T→C and A→G in protein-coding with respect to the transcribed strand, N^{tx}_A and N^{tx}_T are the total numbers of protein-coding A and T bases in transcribed strand across the autosomes, and $\mathrm{n}^{tx}_{T\to C}$ and $\mathrm{n}^{tx}_{A\to G}$ are the genome-wide counts of observed T→C and A→G *de novo* mutations with respect to the transcribed strand in our dataset. γ_{sb} was estimated to be 1.389 and $N^{tx}_A/(N^{tx}_A + N^{tx}_T)$ to be 0.543 in our data.

The Genome of the Netherlands consortium

Steering committee

Cisca Wijmenga[1,2] (Principal Investigator), Morris A. Swertz[1,2], Cornelia M. van Duijn[3], Dorret I. Boomsma[4], P. Eline Slagboom[5], Gertjan B. van Ommen[6], Paul I.W. de Bakker[7,8-10]

Analysis group

Morris A. Swertz[1,2] (Co-Chair), Laurent C. Francioli[7], Freerk van Dijk[1,2], Androniki Menelaou[7], Pieter B.T. Neerincx[1,2], Sara L. Pulit[7], Patrick Deelen[1,2], Clara C. Elbers[7], Pier Francesco Palamara[11], Itsik Pe'er[11,12], Abdel Abdellaoui[4], Wigard P. Kloosterman[7], Mannis van Oven[13], Martijn Vermaat[14], Mingkun Li[15], Jeroen F.J. Laros[14], Mark Stoneking[15], Peter de Knijff[16], Manfred Kayser[14], Jan H. Veldink[17], Leonard H. van den Berg[17], Heorhiy Byelas[1,2], Johan T. den Dunnen[14], Martijn Dijkstra[1,2], Najaf Amin[3], K. Joeri van der Velde[1,2], Jouke Jan Hottenga[4], Jessica van Setten[7], Elisabeth M. van Leeuwen[3], Alexandros Kanterakis[1,2], Mathijs Kattenberg[4], Lennart C. Karssen[3], Barbera D.C. van Schaik[18], Jan Bot[19], Isaäc J. Nijman[7], Ivo Renkens[7], David van Enckevort[20], Hailiang Mei[20], Vyacheslav Koval[21], Karol Estrada[21], Carolina Medina-Gomez[21], Kai Ye[22,5], Eric-Wubbo Lameijer[5], Matthijs H. Moed[5], Jayne Y. Hehir-Kwa[23], Robert E. Handsaker[10,24], Steven A. McCarroll[10,24], Shamil R. Sunyaev[9,10], Paz Polak[9], Dana Vuzman[9], Mashaal Sohail[9], Fereydoun Hormozdiari[25], Tobias Marschall[26], Alexander Schönhuth[26], Victor Guryev[27], Paul I.W. de Bakker[7,8-10] (Co-Chair);

Cohort collection and sample management group

P. Eline Slagboom[5], Marian B. Beekman[5], Anton J.M. de Craen[5], H. Eka D. Suchiman[5], Albert Hofman[3], Cornelia M. van Duijn[3], Ben Oostra[28], Aaron Isaacs[3], Najaf Amin[3], Fernando Rivadeneira[21], André G. Uitterlinden[21], Dorret I. Boomsma[9], Gonneke Willemsen[9], LifeLines Cohort Study[29], Mathieu Platteel[1], Steven J. Pitts[30], Shobha Potluri[30], Purnima Sundar[30], David R. Cox[30,36];

Whole-genome sequencing

Qibin Li[31], Yingrui Li[31], Yuanping Du[31], Ruoyan Chen[31], Hongzhi Cao[31], Ning Li[32], Sujie Cao[32], Jun Wang[31,33,34];

Ethical, Legal, and Social Issues

Jasper A. Bovenberg[35], Margreet Brandsma[6];

Affiliations

[1] Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[2] Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[3] Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

[4] Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands

[5] Section of Molecular Epidemiology, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

[6] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

[7] Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

[8] Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands

[9] Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[10] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[11] Department of Computer Science, Columbia University, New York, NY, USA

[12] Department of Systems Biology, Columbia University, New York, NY, USA

[13] Department of Forensic Molecular Biology, Erasmus Medical Center, Rotterdam, The Netherlands

[14] Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

[15] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[16] Forensic Laboratory for DNA Research, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

[17] Department of Neurology, University Medical Center Utrecht, Utrecht, The Netherlands

[18] Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Medical Center, Amsterdam, The Netherlands

[19] SURFsara, Science Park, Amsterdam, The Netherlands

[20] Netherlands Bioinformatics Centre, Nijmegen, The Netherlands

[21] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

[22] The Genome Institute, Washington University, St. Louis, MO, USA

[23] Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

[24] Department of Genetics, Harvard Medical School, Boston, MA, USA

[25] Department of Genome Sciences, University of Washington, Seattle, WA, USA

[26] Centrum voor Wiskunde en Informatica, Life Sciences Group, Amsterdam, The Netherlands

[27] European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[28] Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands

[29] A full list of the LifeLines Cohort Study members can be found in the Supplemental Note

[30] Rinat-Pfizer Inc, South San Francisco, CA, USA

[31] BGI-Shenzhen, Shenzhen, China

[32] BGI-Europe, Copenhagen, Denmark

[33] Department of Biology, University of Copenhagen, Copenhagen, Denmark

[34] The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark

[35] Legal Pathways Institute for Health and Bio Law, Aerdenhout, The Netherlands

[36] Deceased

The LifeLines Cohort Study

Behrooz Z Alizadeh[1], Paul de Bakker[2], H Marike Boezen[1], Pim van der Harst[2], Gerjan Navis[3], Marianne Rots[4], Salome Scholtens[5], Harold Snieder[1], Ronald P Stolk[1,5], Morris Swertz[6], Bruce HR Wolffenbuttel[7], Cisca Wijmenga[6]

Affiliations

[1] Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands

[2] Department of Cardiology, University Medical Center Groningen, University of Groningen, The Netherlands

[3] Department of Nephrology, University Medical Center Groningen, University of Groningen, The Netherlands

[4] Department of Medical Biology, University Medical Center Groningen, University of Groningen, The Netherlands

[5] LifeLines Cohort Study, University Medical Center Groningen, University of Groningen, The Netherlands

[6] Department of Genetics, University Medical Center Groningen, University of Groningen, The Netherlands

[7] Department of Endocrinology, University Medical Center Groningen, University of Groningen, The Netherlands

Supplementary References

1. Koren, A. *et al.* Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91,** 1033–40 (2012).
2. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20,** 761–70 (2010).
3. Duret, L. & Arndt, P. F. The impact of recombination on nucleotide substitutions in the human genome. *PLoS genetics* **4,** e1000071 (2008).
4. Flicek, P. *et al.* Ensembl 2013. *Nucleic acids research* **41,** D48–55 (2013).