Supplement Table of Contents

Supplemental Tables

Supplemental Figures

Supplemental Notes

Supplement

Supplementary Tables

Supplementary Table 1: Further description of cohorts

	Cohort #1	***Cohort #2***	***Cohort #3***	***Cohort #4***
No. offspring	816	1291	3180	420
Mean no. offspring per family	1.05	1.04	2	9.3
No. multi-offspring families	36	43	1590	45
Thereof no. of dizygotic twin families	35	28	48	0
Study	Goldmann, 2016	Goldmann, 2018	Wilfert, 2020	Sasani, 2019
Sequencing technology	Complete Genomics, 60x	Illumina HiSeq, 40x	Illumina X Ten, 30x	Illumina HiSeqX, 30x
Mapping	Complete Genomics, GRCh37/hg19 genome	Illumina Whole Genome Sequencing Service Informatics Pipeline, GRCh37/hg19 genome	BWA, GRCh38/hg38	BWA-MEM, GRCh37/hg19
Mutation calling	Complete Genomics	GATK	GATK	GATK
DNM detection	custom pipeline, Gilissen 2012	custom pipeline	custom pipeline, Wilfert 2020	cyvcf
No. DNMs per individual (mean)	43.9	57.1	58.1	68.7
R^2 of age model	0.39	0.37	0.42	0.51
Intercept	4.2	8.7	10.8	16.3
Paternal age	Mean: 33.66	Mean: 33.7	Mean: 32.91	Mean: 31.08

	Median: 33.31	Median: 33.51	Median: 39.62	Median: 30.4
	Range: 17.22 – 63.22	Range: 18.74 59.00	Range: 16.10 – 57.62	Range: 18.4 – 49.4
Maternal age	Mean: 31.51	Mean: 31.61	Mean: 30.89	Mean: 28.29
	Median: 31.72	Median: 31.72	Median: 31.02	Median: 27.4
	Range: 17.32 – 43.48	Range: 18.06 – 47.60	Range: 17.10 – 47.78	Range: 16.4 – 45.
Parental age correlation	0.72	0.66	0.71	0.90
Paternal age coefficient	0.93	1.1	1.1	1.4
Maternal age coefficient	0.3	0.4	0.4	0.4

Supplementary Table 2: DNM counts in twins and PAMUCs by cohort.

	Dizygotic twins	***PAMUCs***
Mean DNMs	Cohort #1: 46.49	Cohort #1: 40.91
	Cohort #2: 59.04	Cohort #2: 56.44
	Cohort #3: 62.95	Cohort #3: 57.35
Median DNMs	Cohort #1: 45.5	Cohort #1: 41.00
	Cohort #2: 58.5	Cohort #2: 55.00
	Cohort #3: 63.0	Cohort #3: 57.00
Range DNMs	Cohort #1: 24 – 71	Cohort #1:27 – 65
	Cohort #2: 28 – 94	Cohort #2: 28 – 84
	Cohort #3: 35 – 92	Cohort #3: 30 – 90

Supplementary Table 3: Variance component estimations of all cohorts. Estimations and 95% confidence intervals for influences on variance. “Fixed” denotes the combined effect of both paternal age and maternal age. These values are visualized in **Supplementary Figure 4**.

Cohort	***Factor***	***Variance component estimate***	***Lower c.i. border***	***Upper c.i. border***
Cohort #1	Family	0.029	0.00	0.246
	Batch	0.0047	0.00	0.143
	Residual	0.545	0.320	0.610
	Fixed	0.379	0.327	0.435
Cohort #2	Family	0.158	0.00	0.326
	Residual	0.472	0.269	0.632
	Fixed	0.370	0.226	0.406
Cohort #3	Family	0.0538	0.0024	0.084
	Batch	0.001	0.00	0.006
	Residual	0.521	0.484	0.555
	Fixed	0.424	0.400	0.445
Cohort #4	Family	0.0381	0.0006	0.079
	Residual	0.430	0.37	0.49
	Fixed	0.532	0.477	0.590

Supplementary Table 4: P-values for significant differences between observed and simulated mutation count distributions (not corrected for multiple testing).

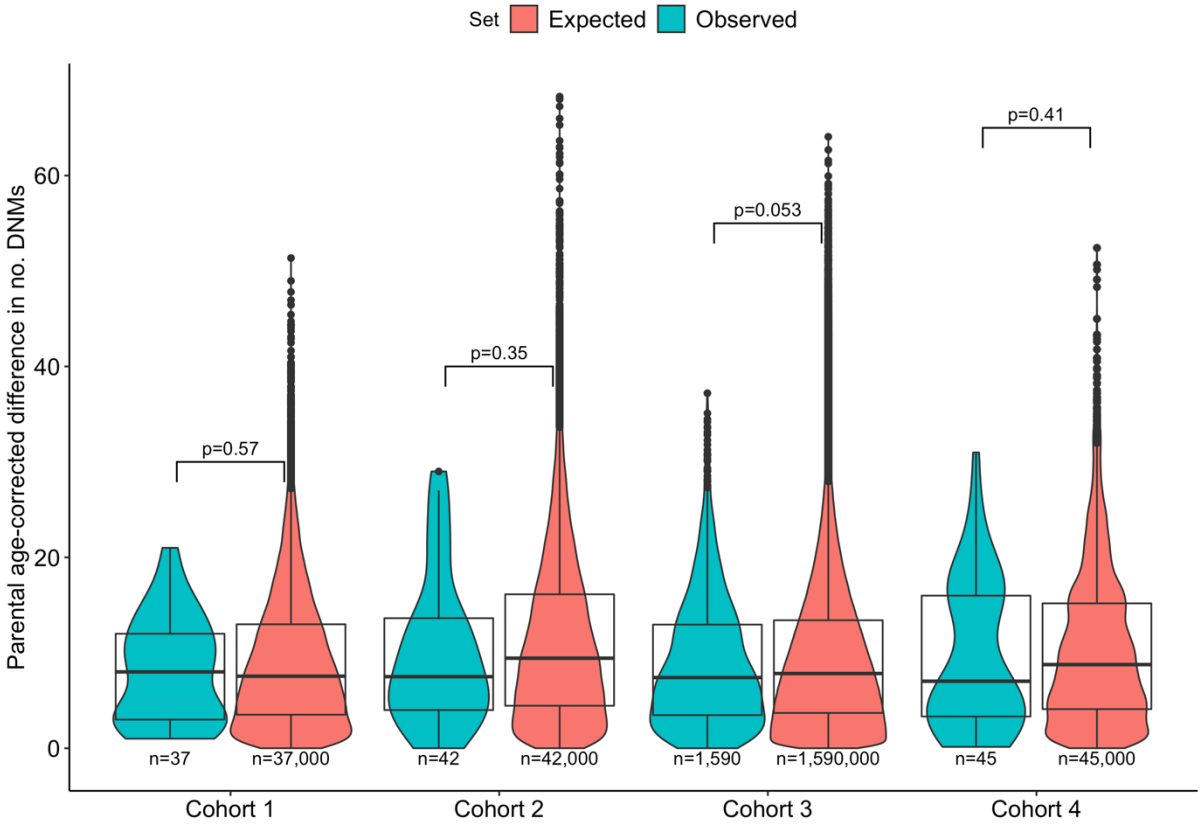
Cohort	*Wilcoxon Rank Sum p*	*Levene's p*	*Fligner-Killeen p*	*Ansari-Bradley p*	*Mood's p*	*F-test p*
Cohort #1	0.722	0.986	0.954	0.968	0.785	0.502
Cohort #2	0.320	0.993	0.008	0.336	0.264	0.343
Cohort #3	0.736	0.999	0.626	0.465	0.667	0.596
Cohort #4	0.939	0.978	0.851	0.951	0.889	0.991

Supplementary Table 5: P-values for significant differences between observed and simulated mutation count distributions by nucleotide substitution (not corrected for multiple testing).

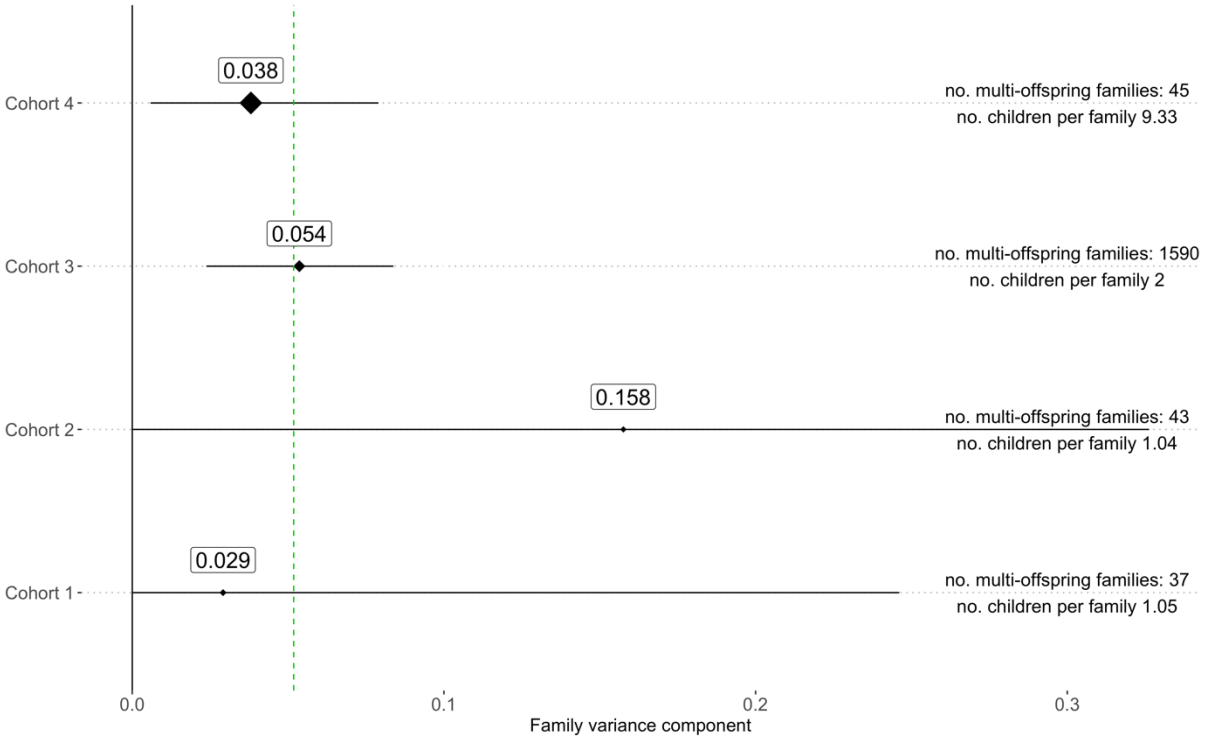
Cohort	*Substitution*	*Wilcoxon Rank Sum p*	*Levene's p*	*Fligner-Killeen p*	*Ansari-Bradley p*	*Mood's p*	*F-test p*
Cohort #1	C - A	0.85	1.00	0.91	0.85	0.88	0.83
	C - G	0.85	1.00	0.92	0.86	0.89	0.80
	C - T	0.98	1.00	0.99	0.98	0.95	0.83
	C - T CpG	0.87	1.00	0.89	0.88	0.90	0.91
	T - A	0.83	1.00	0.86	0.82	0.84	0.98
	T - C	0.89	1.00	0.87	0.90	0.94	0.86
	T - G	0.84	1.00	0.87	0.84	0.85	0.96
Cohort #2	C - A	0.87	1.00	0.89	0.87	0.90	0.89
	C - G	0.86	1.00	0.93	0.86	0.88	0.80
	C - T	0.51	1.00	0.67	0.54	0.69	0.62
	C - T CpG	0.99	1.00	0.97	0.98	0.96	0.76
	T - A	0.99	1.00	0.96	0.99	0.97	0.88
	T - C	0.71	1.00	0.83	0.74	0.85	0.77
	T - G	1.0	1.00	0.97	1.00	1.00	0.92
Cohort #3	C>A	0.99	1.00	0.99	0.99	0.97	0.85
	C>G	0.99	1.00	0.97	0.99	0.98	0.79
	C>T	0.90	1.00	0.97	0.91	0.94	0.80
	C>T CpG	0.89	1.00	0.93	0.90	0.94	0.91
	T>A	0.85	1.00	0.86	0.85	0.87	0.99
	T>C	0.82	1.00	0.90	0.86	0.96	0.89
	T>G	0.86	1.00	0.86	0.86	0.88	0.98
Cohort #4	C>A	0.99	1.00	0.97	0.99	0.98	0.80
	C>G	0.99	1.00	0.97	0.99	0.98	0.82
	C>T	1.0	0.99	0.95	1.00	0.99	0.95
	CpG>TpG	0.77	0.99	0.78	0.78	0.82	0.69
	T>A	1.0	1.0	0.98	1.00	1.00	0.88
	T>C	0.90	0.99	0.97	0.90	0.94	0.89
	T>G	0.86	1.00	0.92	0.86	0.89	0.66

Supplementary Figures

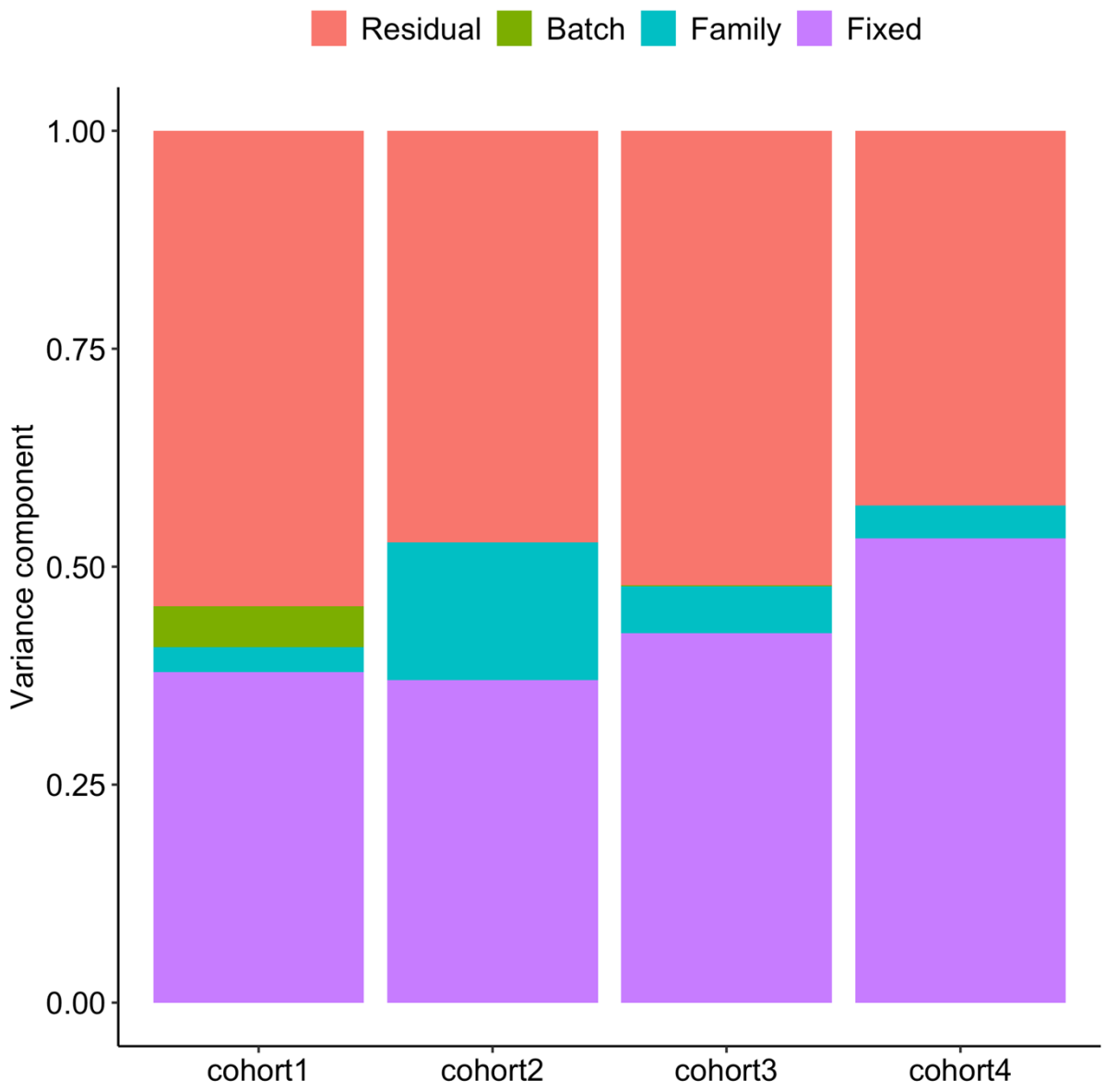
Supplementary Figure 1: *Absolute differences in DNM counts between siblings, corrected for parental age.* Absolute differences of siblings were compared to differences of random combination of children. Numbers indicate sizes of sets, boxes indicate interquartile range and bold line indicates median. Cohort #4 was included by randomly sampling two siblings per family to create 45 sibling pairs because cohort #4 contained no families with exactly 2 offspring. To ensure robustness of this method, we also sampled sibling pairs from cohort4 using the oldest and youngest sibling per family (45 pairs, p = 0.68, not shown), and by randomly sampling families into all possible pairs of two offspring (169 pairs, p = 0.52, not shown).



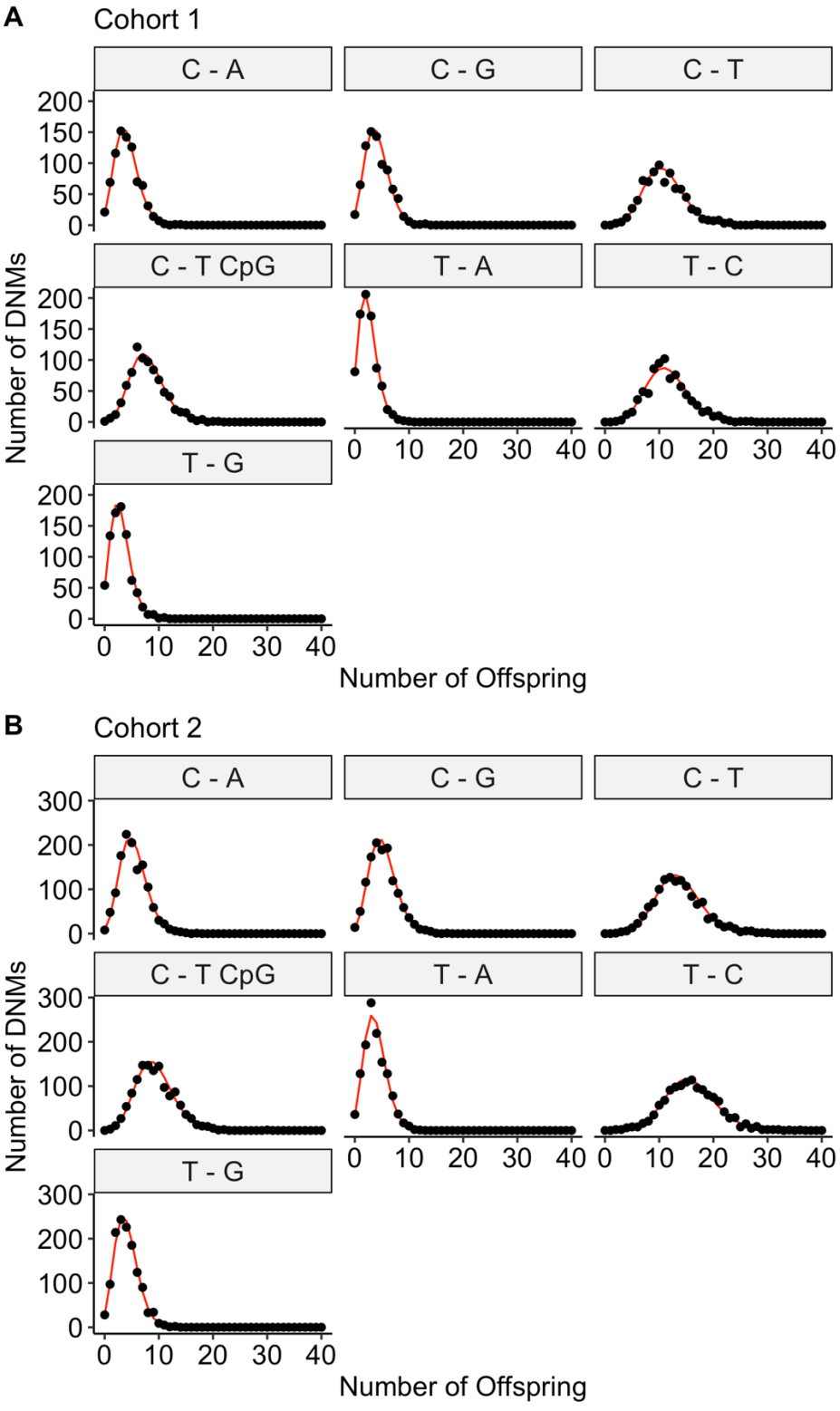
Supplementary Figure 2: *Family variance component estimates across all cohorts.* The error bars denote the 95% confidence intervals. The diamonds indicating the estimates are scaled according to the mean number of children per family. The vertical green line indicates the weighted mean between the point estimates of the four cohorts.

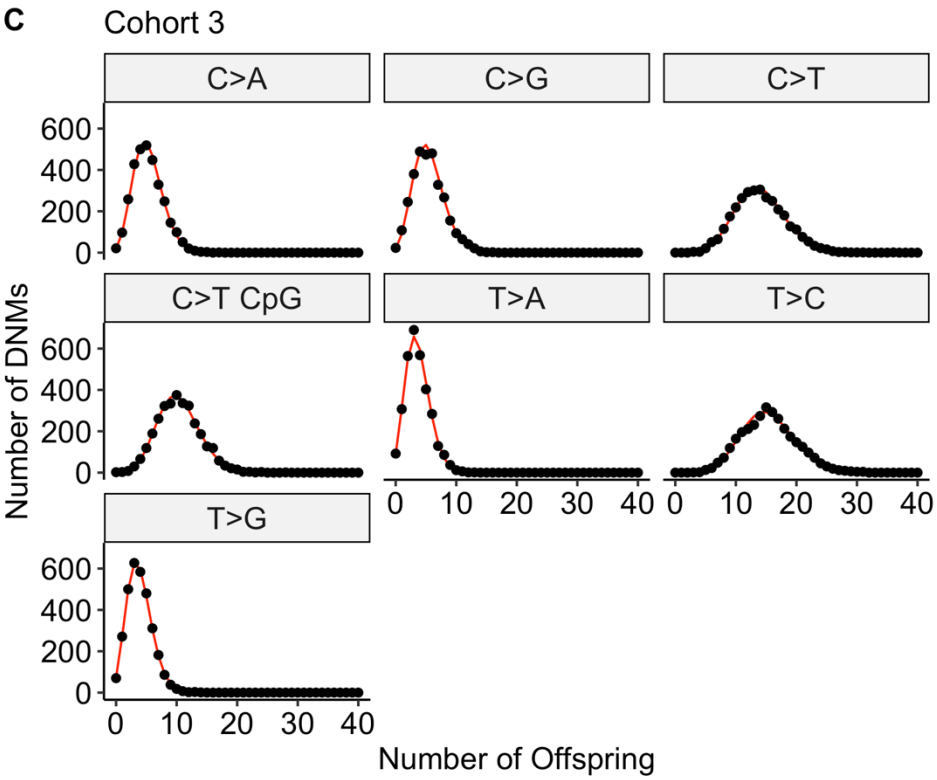


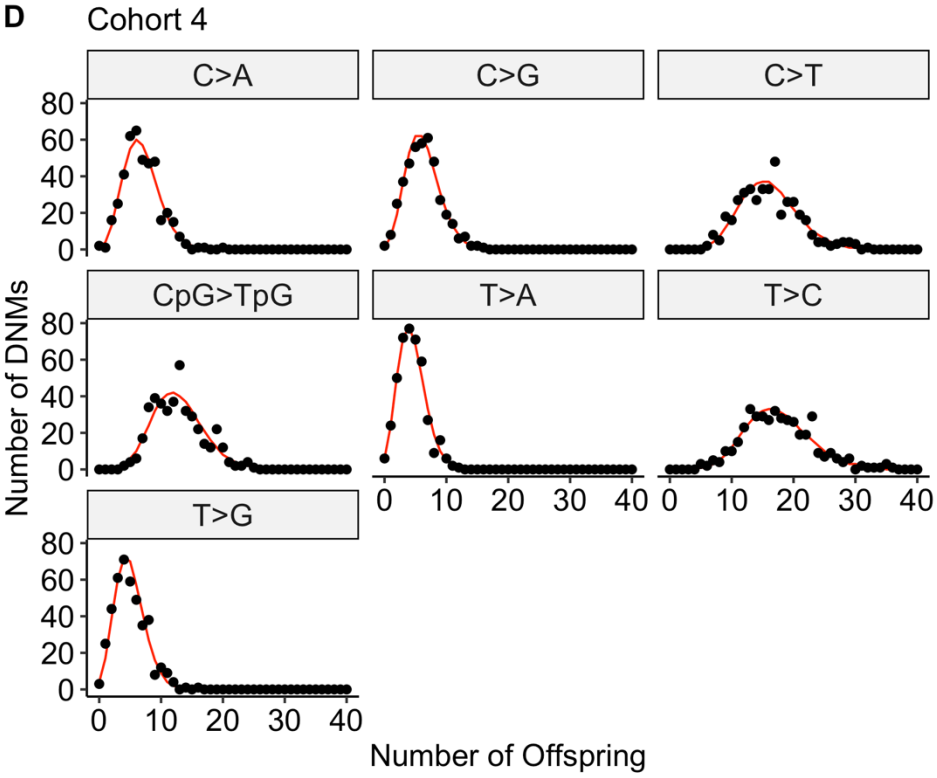
Supplementary Figure 3: *All variance component estimates.* “Fixed” denotes the combined effect of both paternal age and maternal age. “Residual” denotes variance that cannot be attributed to any of the other factors.



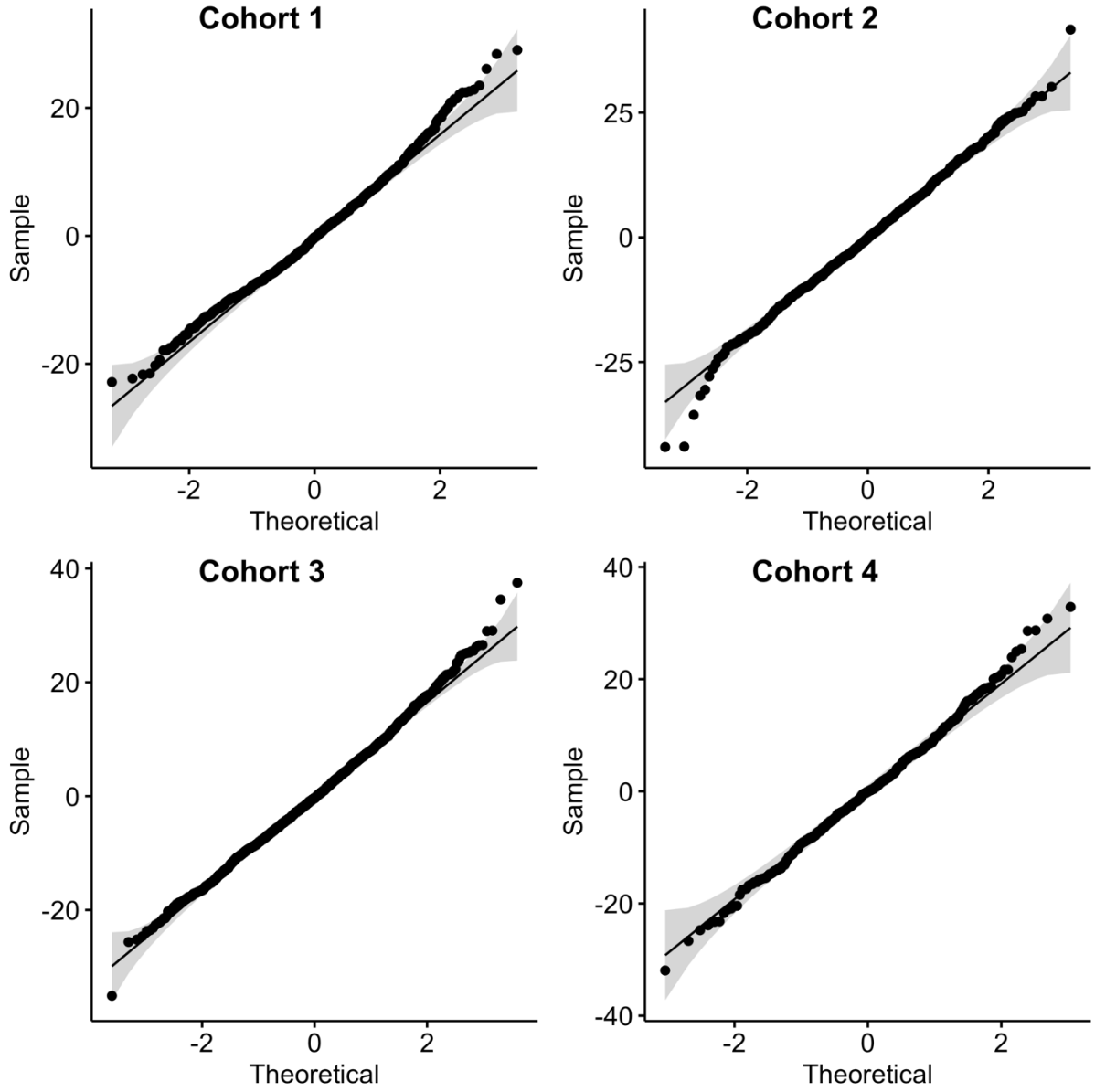
Supplementary Figure 4: *Modelling DNM as family-independent Poisson process for each nucleotide substitution independently*. A,B,C,D: Simulations from cohorts #1, #2, #3 and -#4, respectively. Red lines depict Poisson-based predictions, black dots denote observations. **Supplementary Table 4** lists p-values for various test comparing predictions to observations.



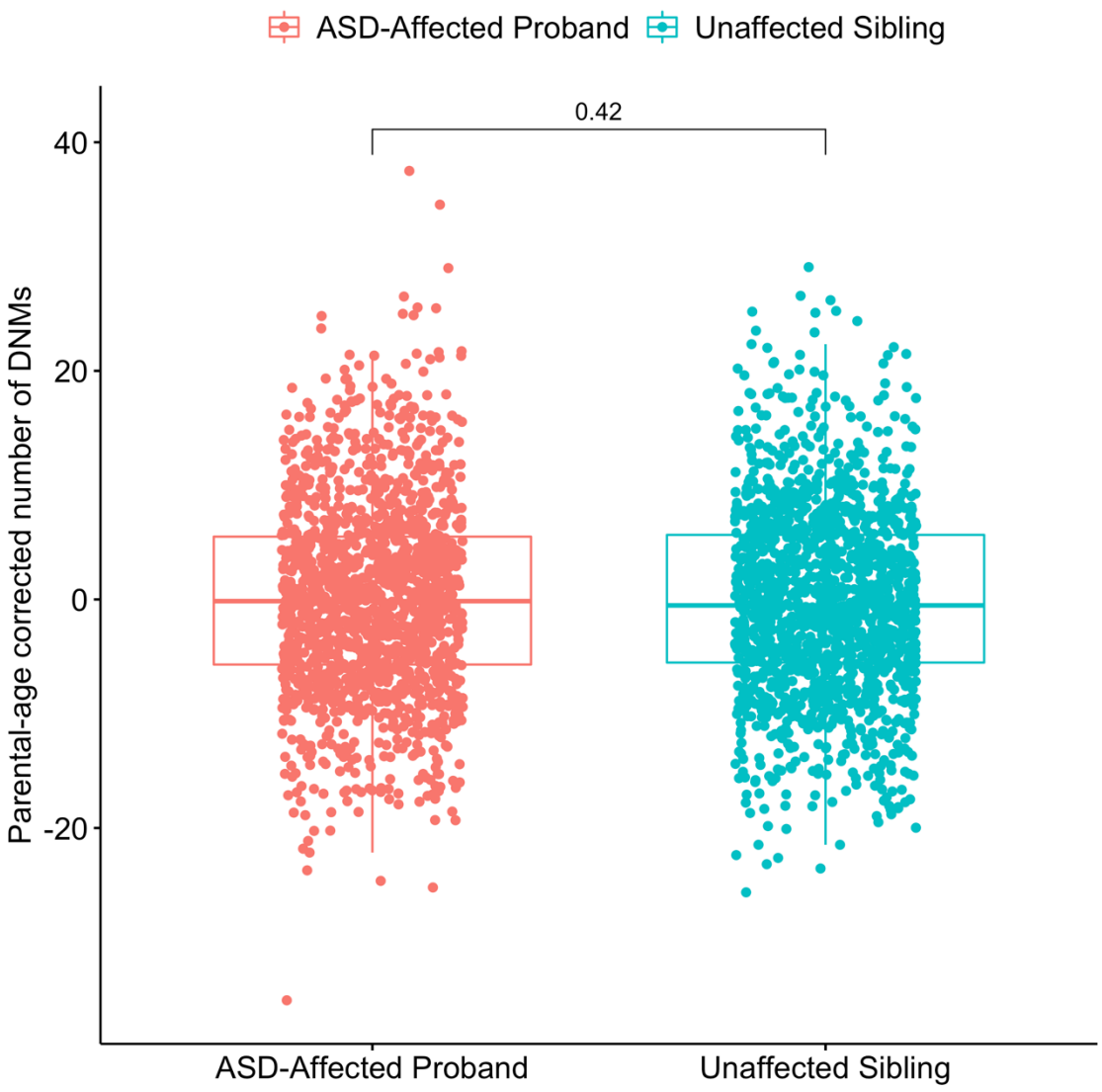

C
Cohort 3
C>A
C>G
C>T
C>T CpG
T>A
T>C
T>G
Number of DNMs
Number of Offspring


D
Cohort 4
C>A
C>G
C>T
CpG>TpG
T>A
T>C
T>G
Number of DNMs
Number of Offspring

Supplementary Figure 5: *Quantile-quantile plots for residuals of the linear models of the number of DNMs against parental age for all cohorts.*



Supplementary Figure 6: *Residual number of DNMs after parental age correction of autism-spectrum disorder ASD patients versus their unaffected siblings of cohort 3.* The number indicates Wilcoxon Rank Sum test p-value.

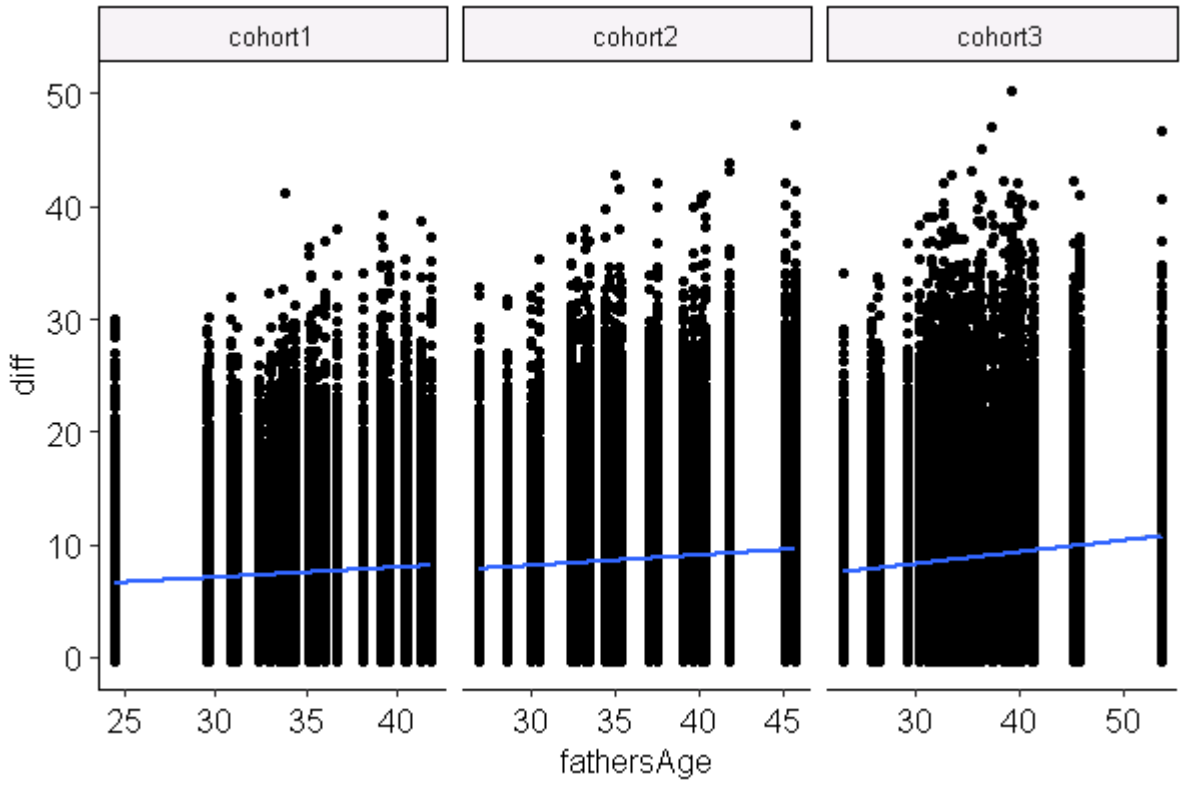


Supplementary Notes

Supplementary Note 1: Poisson-distributed DNM counts and trends in sibling differences with paternal age

We argue in this paper that the number of DNMs of an individual can be modelled as combination of parental age effects and by Poisson-distributed stochasticity. Such a model would imply that the DNM count differences of dizygotic twins would be larger for twin pairs from older parents that for twin pairs from younger parents, as the variance of the Poisson distribution is larger if the expectation value is larger. Yet, in contrast to this intuition, we did not observe a significant trend with paternal age. To understand why such a trend could not be observed we implemented a simulation.

We simulated Poisson-distributed DNM counts based on parental age for cohorts 1-3. Each cohort was simulated 1000 times. In the 1000-fold simulations, there is a moderate but significant increase of the twin differences with paternal age, as suggested by the above intuition (p for linear slope being different from zero: p=0.005, p=0.006, p=0.004 for simulated cohorts 1-3).

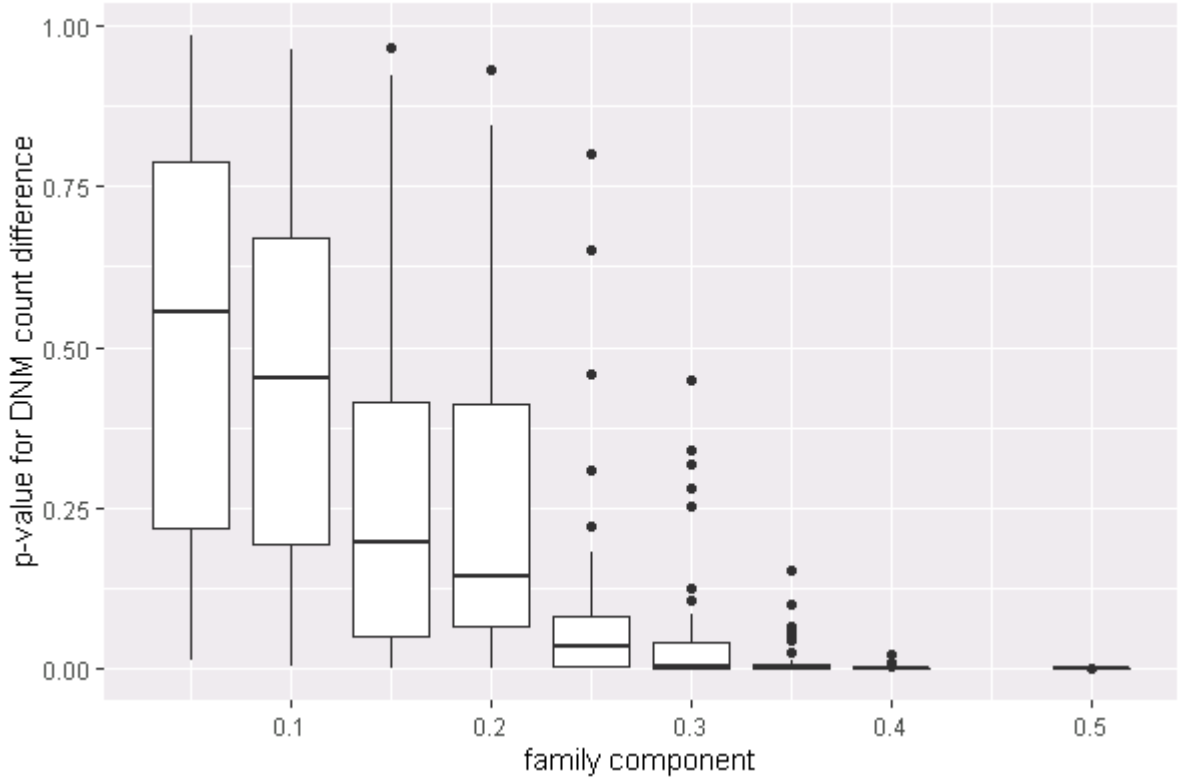


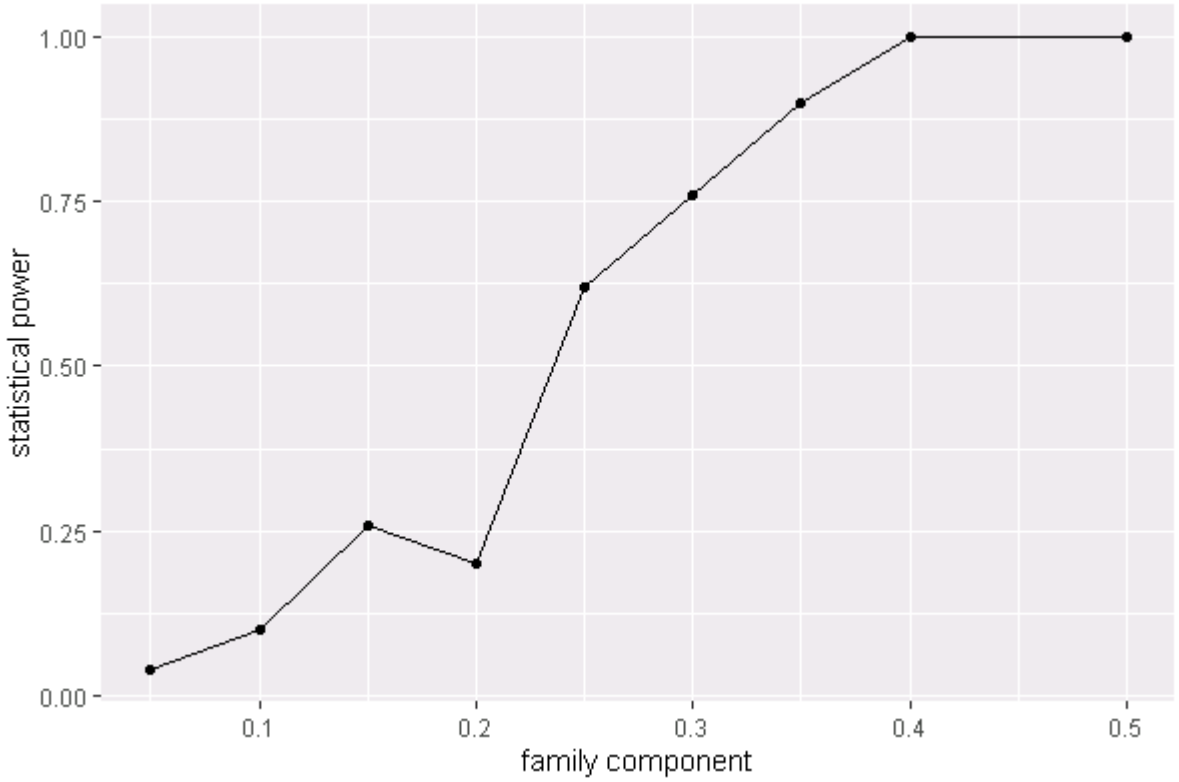
Nevertheless, the cohorts in this study might be too small to significantly detect this slope. In order to test this hypothesis, we pooled the samples of cohorts 1-3 and again simulated 1000 DNM counts for every parent-offspring trio. During each simulation, we calculated the p-value for the slope being different from zero. We obtained values smaller than or equal to 0.05 only in less than 14% of the simulations. This shows that we lack the statistical power to detect an effect of the expected size.

Supplementary Note 2: Power simulations of twins versus PAMUC analysis

In order to estimate our statistical power for detecting a family effect we performed a simulation analysis.

We generated datasets of the same size as cohort 3 by predicting the number of DNMs from a Poisson model of parental ages. To these predicted DNM numbers, we added two terms as normally-distributed stochastic noise. These two factors are simulations of family effects (being the same for family members) and residual variance (independently for every individual). The variances of these simulated terms were set such that the total variance is of the same magnitude as in cohort 3 and the ratio of residual and familial variance can be set to fixed values. We ran 50 of these simulations for with family effects set to 0.05, 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5. With each simulated dataset, we again compared the absolute DNM number differences of twin an PAMUC pairs and assessed potential differences by running Wilcoxon Rank Sum tests.





With a family component of 0.05 (as estimated in this study), our statistical power for detecting this effect in the large cohort 3 is only 0.04. However, with a statistical power of 80% we can exclude a family effect of >= 0.35.