doi:10.1038/nature11396

Rate of *de novo* mutations and the importance of father's age to disease risk

Augustine Kong[1], Michael L. Frigge[1], Gisli Masson[1], Soren Besenbacher[1,2], Patrick Sulem[1], Gisli Magnusson[1], Sigurjon A. Gudjonsson[1], Asgeir Sigurdsson[1], Aslaug Jonasdottir[1], Adalbjorg Jonasdottir[1], Wendy S. W. Wong[3], Gunnar Sigurdsson[1], G. Bragi Walters[1], Stacy Steinberg[1], Hannes Helgason[1], Gudmar Thorleifsson[1], Daniel F. Gudbjartsson[1], Agnar Helgason[1,4], Olafur Th. Magnusson[1], Unnur Thorsteinsdottir[1,5] & Kari Stefansson[1,5]

Mutations generate sequence diversity and provide a substrate for selection. The rate of *de novo* mutations is therefore of major importance to evolution. Here we conduct a study of genome-wide mutation rates by sequencing the entire genomes of 78 Icelandic parent–offspring trios at high coverage. We show that in our samples, with an average father's age of 29.7, the average *de novo* mutation rate is 1.20×10^{-8} per nucleotide per generation. Most notably, the diversity in mutation rate of single nucleotide polymorphisms is dominated by the age of the father at conception of the child. The effect is an increase of about two mutations per year. An exponential model estimates paternal mutations doubling every 16.5 years. After accounting for random Poisson variation, father's age is estimated to explain nearly all of the remaining variation in the *de novo* mutation counts. These observations shed light on the importance of the father's age on the risk of diseases such as schizophrenia and autism.

The rate of *de novo* mutations and factors that influence it have always been a focus of genetics research[1]. However, investigations of *de novo* mutations through direct examinations of parent–offspring transmissions were previously mostly limited to studying specific genes[2,3] or regions[4–7]. Recent studies that used whole-genome sequencing[8,9] are important but too small to address the question of diversity in mutation rate adequately. To understand the nature of *de novo* mutations better we designed and conducted a study as follows.

Samples and mutation calls

As part of a large sequencing project in Iceland[10–12] (Methods), we sequenced 78 trios, a total of 219 distinct individuals, to more than 30× average coverage (Fig. 1). Forty-four of the probands (offspring) have autism spectrum disorder (ASD), and 21 are schizophrenic. The other 13 probands were included for various reasons, including the construction of multigeneration families. The probands include five cases in which at least one grandchild was also sequenced. In addition, 1,859 other Icelanders, treated as population samples, were also whole-genome sequenced (all at least 10×, 469 more than 30×). These were used as population samples to help to filter out artefacts. Sequence calling was performed for each individual using the Genome Analysis Toolkit (GATK) (Methods). The focus here is on single nucleotide polymorphism (SNP) mutations. The investigation was restricted to autosomal chromosomes.

Criteria for calling a *de novo* SNP mutation were as follows. (1) All variants that have likelihood ratio: lik(AR)/lik(RR) or lik(AA)/lik(RR) $> 10^4$, in which R denotes the reference allele and A the alternative allele, in any of the 1,859 population samples, were excluded. Some recurrent mutations could have been filtered out, but the number should be small. The *de novo* mutation calls further satisfy the conditions that (2) there are at least 16 quality reads for the proband at the mutated site; (3) the likelihood ratio lik(AR)/lik(RR) is above 10^{10}; and (4) for both parents, the ratio lik(RR)/lik(AR) is above 100. Applying criteria (1) to (4) gave 6,221 candidate mutations. Further examination led us to apply extra filtering (5) by including only SNPs in which the number of A allele calls is above 30% among the quality sequence reads of the proband. This was considered necessary because there was an abnormally high number of putative mutation calls in which, despite extremely high lik(AR)/lik(RR) ratios for the proband, the fraction of A calls was low (Supplementary Fig. 1). Applying (5) eliminated 1,285 candidate mutations (Supplementary Information). With high coverage, the false negatives resulting from (5) is estimated to be a modest 2% (Supplementary Information). After three more candidates were identified as false

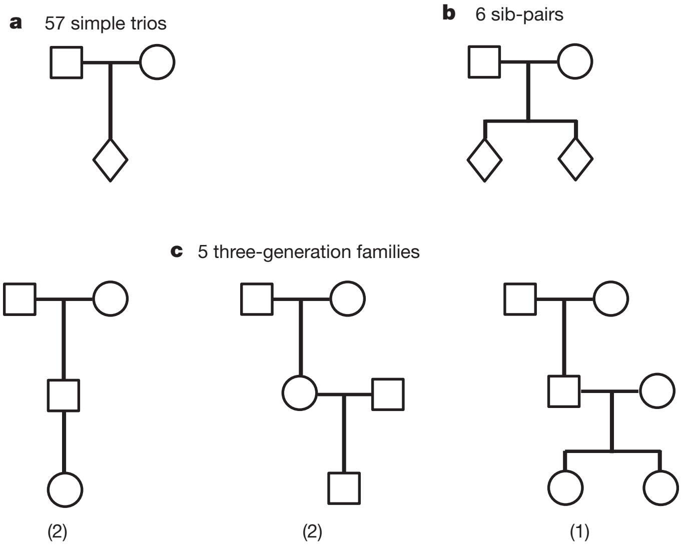


Figure 1 | A summary of the family types. a, Fifty-seven simple trios. **b**, Six sib-pairs accounting for 12 trios. **c**, Five three-generation families accounting for nine trios.

[1]deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland. [2]Bioinformatics Research Centre, Aarhus University, 8000 Aarhus, Denmark. [3]Illumina Cambridge Ltd, Chesterford Research Park, Little Chesterford, Essex CB10 1XL, UK. [4]University of Iceland, 101 Reykjavik, Iceland. [5]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

positives by Sanger sequencing (see section on validation), a total of 4,933 *de novo* mutations, or an average of 63.2 per trio, were called. (The *de novo* mutations are listed individually in Supplementary Table 1.)

Parent of origin and father's age

For the five trios in which a child of the proband was also sequenced, the parent of origin of each *de novo* mutation called was determined as follows. If the paternal haplotype of the proband was transmitted to his/her child, and the child also carries the mutation, then the mutation was considered to be paternal in origin. If the child carrying the paternal haplotype of the parent does not have the mutation, then it is inferred that the mutation is on the maternal chromosome of the proband. Similar logic was applied when the child inherited the maternal haplotype of the proband. In the five trios, the average number of paternal and maternal mutations is 55.4 and 14.2, respectively (Table 1). If mutations were purely random with no systematic difference between trios, their number should be Poisson distributed with the variance equal to the mean. The data, however, show overdispersion (Table 1). This is much more notable for the paternal mutations (variance = 428.8, $P = 1.2 \times 10^{-5}$) than the maternal mutations (variance = 48.7, $P = 0.016$). Moreover, the number of paternal mutations has a monotonic relationship with the father's age at conception of the child. Here, the mean number of paternal mutations is substantially higher than the mean number of maternal mutations (ratio = 3.9), but the difference is even greater for the variance (ratio = 8.8). Hence, variation of *de novo* mutation counts in these individuals is mostly driven by the paternal mutations.

Relationships between parents' age and the number of mutations (paternal and maternal combined, as they could not be reliably separated without data from a grandchild) were examined using all 78 trios (Fig. 2). The number of mutations increases with father's age ($P = 3.6 \times 10^{-19}$) with an estimated effect of 2.01 mutations per year (standard error = 0.17). Mother's age is substantially correlated with father's age ($r = 0.83$) and, not surprisingly, is also associated with the number of *de novo* mutations ($P = 1.9 \times 10^{-12}$). However, when father's age and mother's age were entered jointly in a multiple regression, father's age remained highly significant ($P = 3.3 \times 10^{-8}$), whereas mother's age did not ($P = 0.49$). On the basis of existing knowledge about the mutational mechanisms in sperm and eggs[2], the results support the notion that the increase in mutations with parental age manifests itself mostly, maybe entirely, on the paternally inherited chromosome.

Given a particular mutation rate, due to random variation, the number of actual mutations is expected to have a Poisson distribution. After taking Poisson variation into account, with a linear fit (effect = 2.01 mutations per year), father's age explains 94.0% (90% confidence interval: 80.1%, 100%) (Supplementary Information) of the remaining variation in the observed mutation counts. When an exponential model is fitted (red curve in Fig. 2), the number of paternal and maternal mutations combined is estimated to increase by 3.23% per year. This model explains 96.6% (90% confidence interval: 83.2%, 100%) of the remaining variation. A third model fitted (blue curve in Fig. 2) assumes that the maternal mutation rate is constant at 14.2 and paternal mutations increase exponentially. This explains 97.1% (90% confidence interval: 84.3%, 100%) of the remaining variation and the rate of paternal mutations is estimated to increase by 4.28% per year, which corresponds to doubling every 16.5 years and increasing by 8-fold in 50 years. Seventy-six of the 78 trios have father's ages between 18 and 40.5, a range in which the differences between the three models are modest. Hence, although it seems that the number of paternal *de novo* mutations increases at a rate that accelerates with father's age, more data at the upper age range are needed to evaluate the nature of the acceleration better.

Table 1 | *De novo* mutations observed with parental origin assigned

			Number of *de novo* mutations in proband		
	Father's age (yr)	Mother's age (yr)	Paternal chromosome	Maternal chromosome	Combined
Trio 1	21.8	19.3	39	9	48
Trio 2	22.7	19.8	43	10	53
Trio 3	25.0	22.1	51	11	62
Trio 4	36.2	32.2	53	26	79
Trio 5	40.0	39.1	91	15	106
Mean	29.1	26.5	55.4	14.2	69.6
s.d.	8.4	8.8	20.7	7.0	23.5
Variance	70.2	77.0	428.8	48.7	555.3

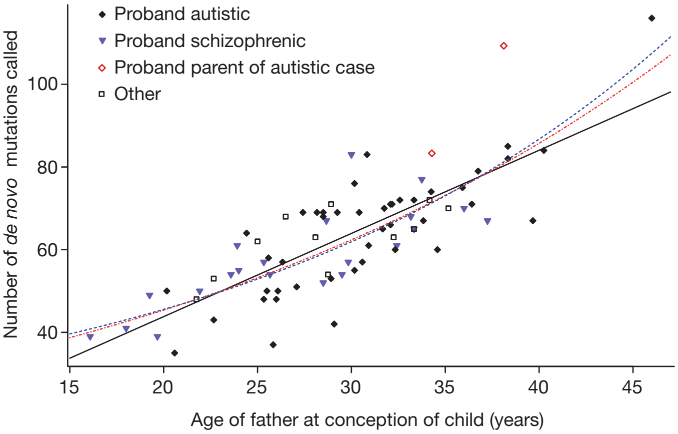


Figure 2 | Father's age and number of *de novo* mutations. The number of *de novo* mutations called is plotted against father's age at conception of child for the 78 trios. The solid black line denotes the linear fit. The dashed red curve is based on an exponential model fitted to the combined mutation counts. The dashed blue curve corresponds to a model in which maternal mutations are assumed to have a constant rate of 14.2 and paternal mutations are assumed to increase exponentially with father's age.

Validation and the nature of errors

Among the *de novo* mutations originally called, two were observed twice, both in siblings, one on chromosome 6 and one on chromosome 10. These cases were examined by Sanger sequencing. The mutation on chromosome 6 is not actually *de novo* as it was seen in the mother also. The one on chromosome 10 was confirmed, that is, it was observed in both siblings, who share the paternal haplotype in this region, but not the parents. This supports the theory that *de novo* mutations in different sperms of a man are not entirely independent[2]. Our trios include seven sib-pairs with 921 *de novo* mutations called. A false *de novo* mutation call for one sib resulting from a missed call in the parent would also show up in the other sib about 50% of the time. Only one such false positive was detected, indicating that this type of error accounts for a small percentage (2/(920/2) = 0.43%) of the called mutations. To evaluate the overall number of false positives, 111 called *de novo* mutations were randomly selected for Sanger sequencing. Eleven failed primer design. Six did not produce results of good quality in at least one member of the corresponding trio (Supplementary Information). For the remaining 94 cases, 93 were confirmed as *de novo* mutations—that is, the mutated allele was observed in the proband but not in the parents. One false positive, in which the putative mutation was not observed in the proband, was identified. The 17 cases that could not be verified are more likely to be located in genomic regions that are more difficult to analyse and hence probably have higher false-positive rates than average. Even so, the overall false-positive rate for the *de novo* mutation calls cannot be high.

The variance of the number of false positives is as important as the mean. False positives that are Poisson distributed, although adding noise, would not create bias for the effect estimates in either the linear or the exponential models for father's age, nor would they bias the estimate of the fraction of variance explained after accounting for Poisson variation. In general, they do not create substantial bias for analyses of differences and ratios. However, if the variance of the false positives is higher than the mean, resulting from systematic effects that affect trios differently, such as DNA quality and library construction, it would increase the unexplained variance and reduce the fraction of variance explained by father's age. The candidates filtered out by criterion (5), if kept, would have introduced false positives of this kind (Supplementary Information). Because father's age explains such a high fraction of the systematic variance of the currently called *de novo* mutations, false positives with this property cannot be common. A similar discussion about false negatives[13] is in Supplementary Information.

Father's age and diseases

Consistent with other epidemiological studies[14,15], in Iceland, the risk of schizophrenia increases significantly with father's age at conception ($n = 569$, $P = 2 \times 10^{-5}$). Father's age is also associated with the risk of ASD. The observed effect is limited to non-familial cases ($n = 631$, $P = 5.4 \times 10^{-4}$), defined as those in which the closest ASD relative is farther than cousins. The epidemiological results, the effect of father's age on *de novo* mutation rate shown here, together with other studies that have linked *de novo* mutations to autism and schizophrenia, including three recent studies of autism through exome sequencing[4–6], all point to the possibility that, as a man ages, the number of *de novo* mutations in his sperm increases, and the chance that a child would carry a deleterious mutation (not necessarily limited to SNP mutations) that could lead to autism or schizophrenia increases proportionally. However, this model does not indicate that the relationship observed here between mutation rate and father's age would have been much different if the probands studied were chosen to be all non-ASD/schizophrenic cases instead. For example, assume that autism/schizophrenia is in each case caused by only one *de novo* mutation. Then autism/schizophrenia cases would on average have more *de novo* mutations than population samples. The magnitude could be substantial if the distribution of father's age has a large spread in the population, but then most of the difference would be caused by the cases having older fathers. If we control for the age of the father at the conception of the individual, then this difference in the average number of *de novo* mutations between control individuals and those with autism/schizophrenia would be reduced to approximately one (Supplementary Information).

Mutations by type and by chromosome

Examination of the 4,933 *de novo* mutations showed that 73 are exonic, including two stop-gain SNPs and 60 non-synonymous SNPs (Supplementary Table 2). One non-familial schizophrenic proband carries a *de novo* stop-gain mutation (p.Arg113X) in the neurexin 1 (*NRXN1*) gene, previously associated with schizophrenia[16–20]. One non-familial autistic proband has a stop-gain *de novo* mutation (p.R546X) in the cullin 3 (*CUL3*) gene. *De novo* loss of function mutations in *CUL3* have been reported to cause hypertension and electrolyte abnormalities[21]. Recently, a separate stop-gain *de novo* mutation (p.E246X) in *CUL3* was reported in an autistic case[5]. Another one of our mutations is a non-synonymous variant (p.G900S) two bases from a splice site in the EPH receptor B2 (*EPHB2*), a gene implicated in the development of the nervous system. A *de novo* stop-gain mutation (p.Q858X) in this gene has recently been described in another autistic case[6]. Given the small number of loss of function *de novo* mutations we and others have reported (approximately 70 genes in the three autism exome scans[4–6]), the overlap is unlikely to be a coincidence. Hence, *CUL3* and *EPHB2* can be added to the list of genes that are relevant for ASD. Effective genome coverage, computed by discounting regions that have either very low (less than half genome average) or very high (more than three times genome average) local coverage, the latter often a symptom of misaligning reads, was estimated to be 2.63 billion base pairs (Supplementary Information). From that, 4,933 mutations correspond to a germline mutation rate of 1.20×10^{-8} per nucleotide per generation, falling within the range between 1.1×10^{-8} and 3.8×10^{-8} previously reported[3,7,8,22,23]. Tables 2 and 3 summarize the nature of the *de novo* mutations with respect to sequence context. Approximately two-thirds ($3{,}344/4{,}933 = 67.8\%$) are transitions. Moreover, there is a clear difference between mutation rates at CpG and non-CpG sites. CpG dinucleotides are known to be mutational hotspots in mammals, ostensibly because spontaneous oxidative deamination of methylated cytosines leads to an increase in transition mutations[24]. The observed rate of transitions here is 18.2 times that at non-CpG sites, higher than but not inconsistent with previous estimates of 13.3 (ref. 23) and 15.4 (ref. 3). The transversion rate is also higher at CpG sites, 2.55-fold that at non-CpG sites. Most of this increased transversion rate at CpG sites is presumably due to general mutation bias favouring mutations that decrease G+C content. The rate of mutations that change a strong (G:C) base pair to a weak (A:T) one is 2.15-times higher than mutations in the opposite direction. This mutational pressure in the direction of A+T is observed for both transitions (ratio = 2.24) and transversions (ratio = 1.82), and cannot be solely explained by CpG mutations. The father's age does not seem to affect the ratios between the rates of these different classes of mutations, that is, as a man ages rates of all mutation types increase by a similar factor.

Table 2 | Germline mutation rates at CpG and non-CpG sites

Type of mutation	*n*	Rate per base per generation
Transition at non-CpG	2,489	6.18×10^{-9}
Transition at CpG	855	1.12×10^{-7}
Transversion at non-CpG	1,516	3.76×10^{-9}
Transversion at CpG	73	9.59×10^{-9}
All	4,933	1.20×10^{-8}

Mutation rates are per generation per base. For non-CpG sites, the effective number of bases examined is taken as 2.583 billion, whereas for CpG sites the number is 48.8 million. These numbers take into account the variation of local coverage in sequencing (Supplementary Information).

The average number of mutations for each chromosome separately and the effect of father's age are displayed in Fig. 3. The effect of father's age is significant ($P < 0.05$) for 14 of the 22 chromosomes when evaluated individually. The solid line in the figure corresponds to a model in which the linear effect of father's age is proportional to the mean number of mutations on the chromosome, or that father's age has a uniform multiplicative effect across the chromosomes. All 22 95% confidence intervals overlap the line, indicating that the results are consistent with the model.

Discussion

The recombination rate is higher for women than men, and children of older mothers have more maternal recombinations that those of young mothers[25]. However, men transmit a much higher number of mutations to their children than women. Furthermore, even though our data also show some overdispersion in the number of maternal *de novo* mutations, it is the age of the father that is the dominant factor in determining the number of *de novo* mutations in the child. Seeing an association between father's age and mutation rate is not surprising[2], but the large linear effect of more than two extra mutations per year, or the estimated exponential effect of paternal mutations doubling every 16.5 years, is striking. Even more so is the fraction of the

Table 3 | Strong-to-weak and weak-to-strong mutation rates

Mutation type	S→W (*n*) rate	W→S (*n*) rate	S→W rate/W→S rate
Transition	(2,025) 1.21×10^{-8}	(1,319) 5.42×10^{-9}	2.24
Transversion	(446) 2.67×10^{-9}	(358) 1.47×10^{-9}	1.82
All	(2,471) 1.48×10^{-8}	(1,677) 6.89×10^{-9}	2.15

n denotes observed mutation counts, and mutation rates are calculated per generation per base. For strong (S; G:C) to weak (W; A:T), the effective number of sites examined is taken as 1.071 billion, and for weak to strong the number is 1.56 billion.

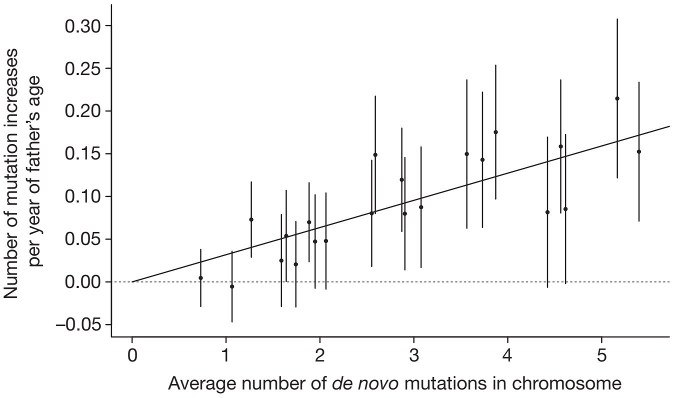


Figure 3 | Effect of father's age by chromosome. By chromosome, the estimated increase in the number of *de novo* mutations per year of father's age is plotted against the average number of mutations observed. The 95% confidence intervals are given. The solid straight line corresponds to the model in which the additive effect of father's age on the number of *de novo* mutations is assumed to be proportional to the mean number of mutations on the chromosome. From left to right, the points correspond to chromosome 21, 22, 19, 20, 15, 17, 18, 14, 16, 13, 12, 9, 10, 11, 8, 7, 6, 3, 5, 4, 2 and 1.

variation it explains, which limits the possible contribution by other factors, such as the environment and the genetic and non-genetic differences between individuals, to mutation rate on a population level. Given the results, it may no longer be meaningful to discuss the average mutation rate in a population without consideration of father's age. Also, even though factors other than father's age do not seem to contribute substantially to the mutation rate diversity in our data, it does not mean that hazardous environmental conditions could not cause a meaningful increase in mutation rate. Rather, the results indicate that, to estimate such an effect for a specific incident, it is crucial to take the father's age into account.

It is well known that demographic characteristics shape the evolution of the gene pool through the forces of genetic drift, gene flow and natural selection. With the results here, it is now clear that demographic transitions that affect the age at which males reproduce can also have a considerable effect on the rate of genomic change through mutation. There has been a recent transition of Icelanders from a rural agricultural to an urban industrial way of life, which engendered a rapid and sequential drop in the average age of fathers at conception from 34.9 years in 1900 to 27.9 years in 1980, followed by an equally swift climb back to 33.0 years in 2011, primarily owing to the effect of higher education and the increased use of contraception (Fig. 4). On the basis of the fitted linear model, whereas individuals born in 1900 carried on average 73.7 *de novo* mutations, those born in 1980 carried on average only 59.7 such mutations (a decrease of 19.1%), and the mutational load of individuals born in 2011 has increased by 17.2% to 69.9. Demographic change of this kind and magnitude is not unique to Iceland, and it raises the question of whether the reported increase in ASD diagnosis lately is at least partially due to an increase in the average age of fathers at conception. Also, the observations here are likely to have important implications for the use of genetic variation to estimate divergence times between species or populations, because the mutation rate cannot be treated as a constant scaling factor, but rather must be considered along with the paternal generation interval as a time-dependent variable.

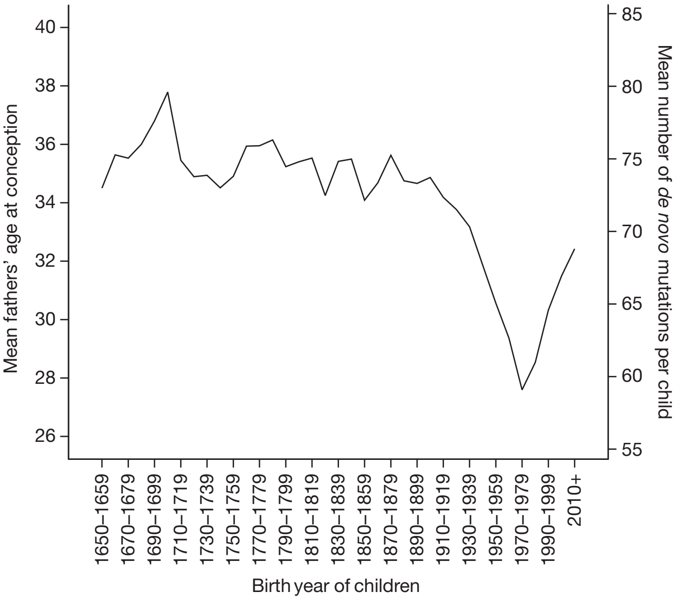


Figure 4 | Demographics of Iceland and *de novo* mutations. The deCODE Genetics genealogy database was used to assess fathers' age at conception for all available 752,343 father–child pairs, in which the child's birth year was ≥1650. The mean age of fathers at conception (left vertical axis) is plotted by birth year of child, grouped into ten-year intervals. On the basis of the linear model fitted for the relationship between father's age and the number of *de novo* mutations, the same plot, using the right vertical axis, shows the mean number of expected mutations for each ten-year interval.

METHODS SUMMARY

Whole-genome sequence data for this study were generated using the Illumina GAII$_x$ and HiSeq2000 instruments. The sequencing reads were aligned to the hg18 reference genome with Burrows–Wheeler aligner (BWA)[26] and duplicates were marked with Picard (http://picard.sourceforge.net/). Quality score recalibration, indel realignment and SNP/indel discovery were then performed on each sample separately, using GATK 1.2 (ref. 27). Likelihoods presented are based on the normalized Phred-scaled likelihoods that are calculated by the GATK variant calling. Statistical analysis was performed in part using the R statistical package. Estimates and confidence intervals for the fraction of variance explained after accounting for Poisson variation were calculated using Monte Carlo simulations (Supplementary Information). Variants were annotated using SNP effect predictor (snpEff2.0.5, database hg36.5) and GATK 1.4-9-g1f1233b with only the highest-impact effect (P. Cingolani, 'snpEff:Variant effect prediction', http://snpeff.sourceforge.net, 2012). More details are in Supplementary Information.

Received 28 February; accepted 4 July 2012.

1. Keightley, P. D. Rates and fitness consequences of new mutations in humans. *Genetics* **190,** 295–304 (2012).
2. Crow, J. F. The origins, patterns and implications of human spontaneous mutation. *Nature Rev. Genet.* **1,** 40–47 (2000).
3. Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* **21,** 12–27 (2003).
4. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485,** 242–245 (2012).
5. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485,** 246–250 (2012).
6. Sanders, S. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485,** 237–241 (2012).
7. Xue, Y. *et al.* Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* **19,** 1453–1457 (2009).
8. Conrad, D. F. *et al.* Variation in genome-wide mutation rates within and between human families. *Nature Genet.* **43,** 712–714 (2011).
9. Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328,** 636–639 (2010).
10. Holm, H. *et al.* A rare variant in *MYH6* is associated with high risk of sick sinus syndrome. *Nature Genet.* **43,** 316–320 (2011).
11. Rafnar, T. *et al.* Mutations in *BRIP1* confer high risk of ovarian cancer. *Nature Genet.* **43,** 1104–1107 (2011).
12. Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nature Genet.* **43,** 1127–1130 (2011).
13. Keightley, P. D. *et al.* Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* **19,** 1195–1201 (2009).
14. Malaspina, D. Paternal factors and schizophrenia risk: de novo mutations and imprinting. *Schizophr. Bull.* **27,** 379–393 (2001).
15. Croen, L. A., Najjar, D. V., Fireman, B. & Grether, J. K. Maternal and paternal age and risk of autism spectrum disorders. *Arch. Pediatr. Adolesc. Med.* **161,** 334–340 (2007).

16. Duong, L. *et al.* Mutations in *NRXN1* in a family multiply affected with brain disorders: *NRXN1* mutations and brain disorders. *Am. J. Med. Genet.* **159B,** 354–358 (2012).
17. Gauthier, J. *et al.* Truncating mutations in *NRXN2* and *NRXN1* in autism spectrum disorders and schizophrenia. *Hum. Genet.* **130,** 563–573 (2011).
18. Kirov, G. *et al.* Comparative genome hybridization suggests a role for *NRXN1* and *APBA2* in schizophrenia. *Hum. Mol. Genet.* **17,** 458–465 (2008).
19. Levinson, D. F. *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and *VIPR2* duplications. *Am. J. Psychiatry* **168,** 302–316 (2011).
20. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18,** 988–996 (2009).
21. Boyden, L. M. *et al.* Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482,** 98–102 (2012).
22. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl Acad. Sci. USA* **107,** 961–968 (2010).
23. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156,** 297–304 (2000).
24. Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* **274,** 775–780 (1978).
25. Kong, A. *et al.* Recombination rate and reproductive success in humans. *Nature Genet.* **36,** 1203–1206 (2004).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
27. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was partly funded by The National Institutes of Health grant MH071425 (K.S.); the European Community's Seventh Framework Programme, PsychCNVs project, grant agreement HEALTH-F2-2009-223423, and NextGene project, grant agreement IAPP-MC-251592; The European Community IMI grant EU-AIMS, grant agreement 115300.

Author Contributions A.K. and K.S. planned and directed the research. A.K. wrote the first draft and together with K.S., S.B., P.S., A.H. and U.T. wrote the final version. O.T.M. and U.T. oversaw the sequencing and laboratory work. G. Masson, G. Magnusson and G.S. processed the raw sequencing data. A.K. and M.L.F. analysed the data, with W.S.W.W., H.H., G.B.W., S.S., G.T. and D.F.G. providing assistance. P.S. and S.A.G. performed functional annotations. S.B. analysed the mutations with respect to sequence content. A.S., Aslaug J. and Adalbjorg J. did the Sanger sequencing. A.H. investigated the contribution of demographics.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.K. (kong@decode.is) or K.S. (kari.stefansson@decode.is).

doi:10.1038/nature11396

1. Study samples.

A total of 2078 samples from a large sequencing project at deCODE were used in this study, 219 samples from 78 trios with two grandchildren who were not also members of other trios, along with 1859 population samples. For the offspring members of each trio, 44 were classified with Autism Spectrum Disorder (ASD) according to ICD-10 criteria using the Autism Diagnostic Interview-Revised (Lord, C., Rutter, M. & Le Couteur, A. 1994), and 21 were classified as having schizophrenia as diagnosed according to Research Diagnostic Criteria (Spitzer, R.L., Endicott, J. & Robins, E. 1978) using the Schedule for Affective Disorders and Schizophrenia Lifetime Version (Spitzer, R.L. & Endicott, J. 1977). The probands from the remaining 13 trios have neither diagnosis.

All biological samples used in this study were obtained according to protocols approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Informed consent was obtained from all participants and all personal identifiers were encrypted with a code that is held by the Data Protection Commission of Iceland.

2. Preparation of samples for whole genome sequencing.

The TruSeq™ sample preparation kit (Illumina) was employed for the preparation of libraries for whole genome sequencing (WGS). In short, approximately 1 µg of genomic DNA, isolated from frozen blood samples, was fragmented to a mean target size of approximately 300-400 bp using a Covaris E210 instrument. The resulting fragmented DNA was end repaired using T4 and Klenow polymerases and T4 polynucleotide kinase with 10 mM dNTP followed by addition of an 'A' base at the ends using Klenow exo fragment (3′ to 5′-exo minus) and dATP (1 mM). Sequencing adaptors containing 'T' overhangs were ligated to the DNA products followed by agarose (2%) gel electrophoresis. Fragments of about 450-500 bp were isolated from the gels (QIAGEN Gel Extraction Kit), and the adaptor-modified DNA fragments were PCR enriched for ten cycles using Phusion DNA polymerase (Finnzymes Oy) and PCR primers PE 1.0 and PE 2.0 needed for paired-end sequencing. Enriched libraries were purified using AMPure XP beads. The quality and concentration of the libraries were assessed with the Agilent 2100 Bioanalyzer using the DNA 1000 LabChip. Libraries were stored at −20 °C. All steps in the workflow were monitored using an in-house laboratory information management system (LIMS) with barcode tracking of all samples and reagents.

3. DNA whole genome sequencing.

Template DNA fragments were hybridized to the surface of paired-end (PE) flowcells (either for $\mathrm{GAII_x}$ or HiSeq 2000 sequencing instruments) and amplified to form clusters using the Illumina cBot™. In brief, DNA (3–12 pM) was denatured, followed by hybridization to grafted adaptors on the flowcell. Isothermal bridge amplification using Phusion polymerase was then followed by linearization of the bridged DNA, denaturation, blocking of 3′ ends and hybridization of the sequencing primer.

Sequencing-by-synthesis (SBS) was performed on either Illumina $\mathrm{GAII_x}$ or HiSeq 2000 instruments, respectively. Paired-end libraries were sequenced using 2x120 cycles of incorporation and imaging with Illumina SBS kits, TruSeq™ v5 for the GAIIx. For the HiSeq 2000, 2x101 cycles with SBS kits v2.5 or v3 were employed. Each library was initially run on a single lane on a $\mathrm{GAII_x}$ for validation, assessing optimal cluster densities, insert size, duplication rates and comparison to chip genotyping data. Following validation, the desired sequencing depth (either 10X or 30X) was then obtained using either sequencing platform. Targeted raw cluster densities ranged from 500–800 K/mm^2, depending on the version of both the sequencing chemistry and the data imaging/analysis software packages (SCS.2.8/RTA1.8 or SCS2.9/RTA1.9 for the $\mathrm{GAII_x}$ and HCS1.3.8. or HCS1.4.8 for HiSeq 2000). Real-time analysis involved conversion of image data to base-calling in real-time.

4. Sequence alignments and variants calling.

For each lane in the DNA sequencing output, the resulting qseq files were converted into fastq files using an in-house script. All output from sequencing was converted, and the Illumina quality filtering flag was retained in the output. The fastq files were then aligned against Build 36 of the human reference sequence using bwa version 0.5.9 (Li, H. & Durbin, R. 2009).

SAM file output from the alignment was converted into BAM format using samtools version 1.1.18 (Li, H. *et al* 2009), and an in-house script was used to carry the Illumina quality filter flag over to the BAM file. The BAM files for each sample were then merged into a single BAM file using samtools. Finally, Picard (versions from 1.17 to 1.55) (http://picard.sourceforge.net) was used to mark duplicates in the resulting BAM files.

GATK 1.2 (McKenna, A. *et al* 2010) was used for quality score recalibration and indel realignment. SNP/Indel discovery was then performed by GATK 1.2 on each sample separately using standard filtering parameters as recommended (http://www.broadinstitute.org/gsa/wiki/index.php/Best_Practice_Variant_Detection_with_the_GATK_v3). For variant discovery, a confidence level threshold of 50.0 was used, which was slightly higher than was recommended for DEEP (>10X) coverage. The discovery set of SNPs and indels for the individuals, restricted to variants with lik(RR)/lik(RA) > 10^4 , max(lik(RA)/lik(AA), lik(AA)/lik(RA)) > 10^3 and local coverage less than three times the sample's average coverage, were merged using a combination of in-house scripts (similar to the CombineVariants tool in GATK) and the individuals were then recalled for the merged variant set. Variant sites were investigated as potential *de novo* mutations for each trio if none among the other sequenced individuals (excluding first degree relatives of the trio proband) had a lik(RR)/lik(RA) ratio greater than 10^4. All likelihoods evaluated here are based on the normalized Phred-scaled likelihoods calculated by the GATK variant caller (UnifiedGenotyper).

5. The reason for and the effect of applying filter (v).

As noted in the main text, after applying criteria (i) to (iv), there were 6,221 candidate *de novo* mutations remained. Two of these, on chromosome 6, were identical and seen in two siblings. Validation by Sanger sequencing revealed that the variant is actually also carried by the mother, and hence not really *de novo*. Removing these left us with 6,219 candidate mutations. (One of the 6,219, as noted in the main text, was revealed at a much later stage as a false positive by Sanger sequencing. This case is included in the analysis described here because we think this better reflects what led us to apply filter (v) in the first place. But, of course, removing it from this analysis would make very little difference.) For each of these called variants, among the quality reads, the fraction of A (alternative allele) calls was calculated. **Supplementary Fig. 1** is a histogram of the 6219 fractions. The histogram has two modes, one at 50%, and one at 20% to 25%. This suggests a mixture of two distributions, one representing true heterozygotes with a mode at 50%, and one representing erroneous calls with a mode at a much lower percentage. Many of the cases contributing to the latter probably resulted from having reads from two or more different, but highly similar, regions mixed up together. For example, if two sites are mixed together, one is heterozygous and the other homozygous reference (RR), the fraction of A reads would be 25% in expectation. It could sometimes be one in 6, or 16.7%, if reads from three different sites were misaligned to one location. Out the 6219 candidates, fraction of A calls are at or below 30% for 1285 of them. Filtering these out from the set of 6219 gave a set 4,934. It is interesting to note that, if we did not eliminate the 1285 cases and performed the analysis with 6219 *de novo* mutations called, the estimated effect of father's age would be very similar, actually a little higher (2.30 mutations per year as opposed to 2.01), but the significance and fraction of variance explained after accounting for Poisson variation would be substantially reduced ($P = 7.6\times10^{-14}$ and variance explained = 67.7%, as opposed to $P = 3.9\times10^{-19}$ and variance explained = 93.9% when using the 4934 called mutations). That the estimate is higher is possibly because there are some true positives in the 1285 cases. The *P* value is less significant and variance explained is substantially lower because the 1285 cases are introducing a lot of noise. In particular, the 1285 cases exhibit substantial over dispersion, variance/mean = 3.9, and only a small fraction of that could be accounted for by father's age. We can get a rough estimate of how many true positives are in the 1285 cases in two ways. Firstly, with 30 reads, with a true heterozygote, the probability of having 9 or less A reads is 2.1%. Given that there are about 5000 *de novo* mutations in our trios, that corresponds to about 105 true positives filtered away, or 105 false negatives introduced. Secondly, from the histogram (**Supplementary Fig. 1**), there are 71 mutation calls with the fraction of A reads greater than or equal to 70%. Assuming symmetry, it would imply that about 71 true positives were filtered away by (v), and corresponds to a false negative rate of about 1.4%. Taking these two estimates into account, we believe that the filter (v) is likely to be responsible for about 2% of false negatives.

6. Models fitted, estimating fraction of variance explained and confidence intervals.

As noted in the main text, we fitted 3 models to evaluate the relationship between father's age and number of *de novo* mutations. Let Y denote the number of *de novo* mutations, and let X be the age of the father at conception of the child. The linear model was fitted by performing a simple regression of Y on X. The first exponential model fitted was done by regressing log(Y) on X. The second exponential model fitted was performed by regressing log(Y – 14.2) on X, noting that 14.2 was chosen because that is the mean number of maternal *de novo* mutations observed in the 5 trios for which parent of origin of the mutations could be determined. For the exponential fits, residual sum of squares and variance explained were calculated by converting the fitted values back to the original scale. Note that the same number of parameters, an intercept and a slope, were fitted in all 3 cases. The difference is just the scale under which the regression was performed. Because the 3 models are not nested, one cannot test one against another in a standard frequentist manner and compute *P*-values. But we note the following. If we add a quadratic term of X to the linear fit, the quadratic term is marginally significant with $P = 0.07$. But even with the quadratic term added, R^2, the fraction of variance explained, is still slightly lower than those resulting from fitting the two exponential models. Hence it is reasonable to say the exponential models fit the data better than the linear model.

For a Poisson distribution the variance is equal to the mean. Hence, using the data, a simple estimate of the fraction of variance explained after accounting for Poisson variation is

$$R^2/[1 - \text{mean}(Y)/\text{var}(Y)] \quad (*)$$

where R^2 is the fraction of total variance explained by father's age obtained from the model fit. For the linear fit, to slightly improve this estimate and to construct confidence intervals, we performed Monte Carlo simulations based on the following model:

$$Y \sim \text{Poisson}(A + B{*}\text{age} + \text{Normal}(0,\text{SIGMA})).$$

We set A and B to the fitted values. By varying SIGMA, we could set the theoretical value of the fraction of systematic variance explained by father‘s age to any value we like. From the simulations, we found that the simple estimate (*) is slightly biased, in the sense that its sampling distribution has a mean/median that is a little higher (about 0.5%) than the actual value used to do the simulation. So we centered the estimate by choosing the value so that when it is used to perform the simulations, the median of the simulated values of (*) will correspond to the observed value calculated from the real data. Similarly, the lower bound of the 90% CI is the value so that, when it is used to perform the simulations, the 95th percentile of the simulated values of (*) will correspond to the observed value.

A similar method is used to obtain estimates and confidence intervals for the exponential fits.

7. Some details on the Sanger sequencing results.

One hundred and eleven of the *de novo* mutations called were randomly selected for validation using Sanger sequencing. Eleven failed primer design. For the 100 cases where we obtained primers, the first run generated reliable results for 86 of them, and all confirmed as *de novo* mutations. The 14 cases that failed to generate reliable results were rerun. Results were obtained for 8 cases, with 7 confirmations and one false positive identified where the putative variant was not observed in the proband. Hence, overall, we have 93 = 86+7 confirmations and one identified false positive. Among the other six cases, two of them had problems with the PCR and did not generate any useful results at all. For the other four cases, the mutation was seen in proband and not in the mother, but reliable results could not be obtained for the father due apparently to problems with low quality DNA. Hence, the data for these four cases, while not conclusive, are consistent with true *de novo* mutations.

8. The impact of false negatives on various analyses.

In the main text, we discussed how false positives of various types could impact the analyses. Here is a similar discussion on false negatives. Because of the limitations of current sequencing technology and that the methods used to call the variants are still far from perfect, we had to apply filters to limit the false positives. As a result, false negatives are unavoidable. While not attempting to give a precise estimate of its overall magnitude, we note that the overall mutation rate observed here is not inconsistent with other estimates reported for trio data. For the analyses of father's age, false negatives that are Poisson in nature will bias the effect estimate of the linear model downwards, implying that the actual effect genomewide is very likely to be above the current estimate of 2.01 per year. However, the effect estimate for the exponential model, and, in general estimates of ratios, should not be substantially affected. Non-Poisson false negatives would add to the unexplained variance, and, following the same argument applied to false positives, their magnitude is likely to be modest.

9. Average number of *de novo* mutations in cases and in controls.

Suppose non-familial ASD/SZ cases are in each case caused by one (and only one) *de novo* mutation. Suppose, while father's age has an effect on the number of *de novo* mutations, its effect is in a multiplicative sense uniform over the genome. And suppose there is no other systematic factors influencing the number of mutations (or that their contributions are very small on a population level) other than Poisson variation. Then the chance of an individual being a case is essentially proportional to the number of *de novo* mutations they carry, e.g. a person carrying 120 *de novo* mutations will have 3 times the chance of being a case than those that carry 40. Of interest here is the reverse question --- what is the average number of *de novo* mutations in cases, and more specifically, on average how many more *de novo* mutations do cases have compared to population controls. The answer depends on the spread of the population distribution of the number of *de novo* mutations, the greater the spread the greater the difference. Mathematically, if X is a random variable having the mutation count distribution, then the difference is

$$[\text{mean}(X^2)/\text{mean}(X)] \quad - \text{mean}(X)$$

But the spread of *de novo* mutation count distribution is driven, apart from Poisson variation, by father's age. Using the father's age distribution for 97,095 births in Iceland in the last century (mean = 31.7, SD = 6.31), and assuming the exponential model that was fitted for paternal mutations assuming that maternal mutation rate is fixed at 14.2, the above difference is estimated to be 4.70. However, a large fraction of the effect is the consequence of the cases on average having older fathers than the controls. If we condition/adjust for father's age, that essentially means we are comparing cases and controls at a fixed age. In that case all variation comes from Poisson variation. So the difference above can be calculated by assuming that X has a Poisson distribution with some fixed mean. It happens that, regardless of the mean of the Poisson distribution, the above difference is 1.

10. Classifying *de novo* mutations by function and with respect to genes.

See **Supplementary Tables 1** and **2**.

11. Effective coverage of whole genome sequencing.

The effective genome coverage is based on the sum of the read depth over all 2,078 sequenced individuals. The initial coverage includes 2.628 billion non-CpG bases and 53.40 million CpG bases, a total of 2.681 billion. To calculate effective coverage, we applied one lower bound: (i) local coverage has to be above 50% of average genome coverage, and one upper bound: (ii) local coverage is no more than 3 times average genome coverage. After filtering using (i) and (ii), 2.583 billion non-CpG bases remained and 48.80 million CpG bases remained. Note that while less than 2% of non-CpG bases were filtered out, over 8% of CpG bases were filtered out. This is in part due to the fact that CpG bases are in regions that are GC rich, locations where current sequencing technology tends to have lower coverage (Wang et al. 2011). Note that according to (Keightley et al. 2009), mutation rate estimates stabilize at sites with a read depth above 4. Therefore, since our average sequencing depth is high, we expect only a small fraction of the genome needed to be removed when considering the coverage.

Criteria (i) and (ii) are chosen do deal with problematic regions for the sequencing technology we are employing. The boundaries were chosen after looking at the quality of SNP calls at different read depths. Variant calls in genomic regions with low coverage can both inflate the false positive rate (Keightley et al. 2009) and overlook mutations, which increases the false negative rate. Excessive coverage can correspond to regions with low complexity (e.g., in the vicinity of the centromeres and telomeres) and sequence repeats, imposing challenges for read alignment and increasing the chance of calling false mutations.

12. The list of 4,933 *de novo* mutations

The attached excel file **Supplementary Table 1** contains information for each of the 4,933 *de novo* mutations individually. They correspond to the summary in **Supplementary Table 2**. The positions are based on Human Assembly Build 36.

References

Spitzer, R.L., Endicott, J. & Robins, E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* **35**, 773-782 (1978).

Spitzer, R.L. & Endicott J. *Schedule for Affective Disorders and Schizophrenia-Lifetime version (SADS-L)* 3rd edition, New York. New York State Psychiatric Institute (1977).

Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**, 659-695 (1994).

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

McKenna, A. *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**,1297-1303 (2010).

Wang, W., Wei, Z., Lam T.-W., and Wange, J. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Sci Rep.* 1:55 (2011).

Keightley, P.D., Trivedi, U., Thomson, M., Oliver, F., Kumar S., and Blaxter, M.L. Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Res.* 19: 1195-1201 (2009)

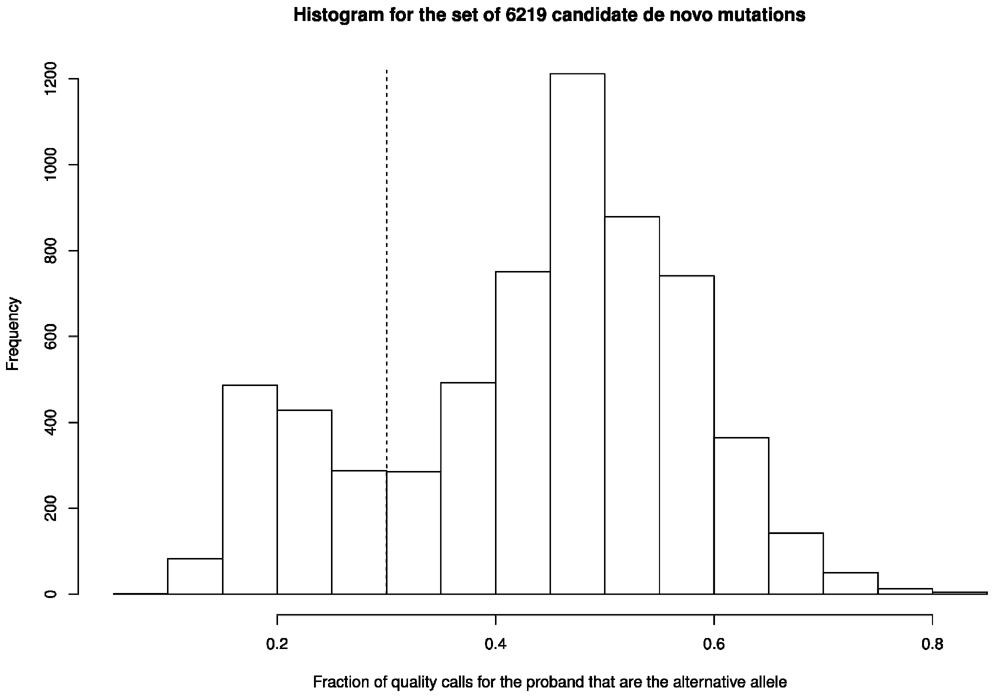
Supplementary Table 2. Breakdown by gene context

Gene Content	Count of Mutations
Non_synonymous coding	60
Stop_gained	2
Synonymous coding	11
UTR_3_prime	16
Upstream	175
Downstream	267
Intergenic	2589
Intron	1808
Transcript*	5

*It includes 4 pseudogenes and 1 Immunoglobulin gene.

Variants were annotated using SNP effect predictor (snpEff2.0.5, database hg36.5) and Genome Analysis Toolkit 1.4-9-g1f1233b with only the highest-impact effect (Cingolani, P. "snpEff:Variant effect prediction", http://snpeff.sourceforge.net, 2012).

Supplementary Figure 1.



BBC NEWS
HEALTH

22 August 2012 Last updated at 13:11 ET

Older dads linked to rise in mental illness

By Pallab Ghosh
Science correspondent, BBC News

A genetic study has added to evidence that the increase in some mental disorders may be due to men having children later in life.

An Icelandic company found the number of genetic mutations in children was directly related to the age of their father when they were conceived.

One prominent researcher suggested young men should consider freezing their sperm if they wanted to have a family in later life.

The research is published in Nature.

According to Dr Kari Stefansson, of Decode Genetics, who led the research, the results show it is the age of men, rather than women, that is likely to have an effect on the health of the child.

"Society has been very focussed on the age of the mother. But apart from [Down's Syndrome] it seems that disorders such as schizophrenia and autism are influenced by the age of the father and not the mother".

Male driven

Dr Stefansson's team sequenced the DNA of 78 parents and their children.

This revealed a direct correlation between the number of mutations or slight alterations to the DNA, of the child and the age of their father.

The results indicate that a father aged 20 passes, on average, approximately 25 mutations, while a 40-year-old father passes on about 65. The study suggests that for every year a man delays fatherhood, they risk passing two more mutations on to their child.

What this means in terms of the impact on the health of the child is unclear. But it does back studies that also show fathers are responsible for mutations and that these mutations increase with age.

And, for the first time, these results have been quantified and they show that 97% of all mutations passed on to children are from older fathers.

"No other factor is involved which for those of us working in the field is very surprising," said Dr Stefansson.

He added that the work backed other studies that have found links between older fathers and some mental disorders.

"The average age of fathers has been steeply rising [in industrialised countries] since 1970. Over the same period there has been an increase in autism and it is very likely that part of that rise is accounted for by the increasing age of the father," he said.

The findings should not alarm older fathers. The occurrence of many of these disorders in the population is very low and so the possible doubling in risk by having a child later in life will still be a very low risk.

Nearly all children born to older fathers will be healthy. But across the population the number of children born with disorders is likely to increase if this theory holds true.

Older fathers and therefore genetic mutations have been linked with neurological conditions because the brain depends on more genes for its development and regulation.

So mutations in genes are more likely to show up as problems in the brain than in any other organ. But it is unclear whether the age of fathers has an effect on any other organ or system. The research has not yet been done.

The reason that men rather than women drive the mutation rate is that women are born with all their eggs whereas men produce new sperm throughout their adult life. It is during sperm production that genetic errors creep in, especially as men get older.

Writing a commentary in the Journal Nature, Prof Alexey Kondrashov, of Michigan University, said young men might wish to consider freezing their sperm if future studies showed there were other negative effects on a child's health.

"Collecting the sperm of young adult men and cold storing it for later use could be a wise individual decision. It might also be a valuable for public health, as such action could reduce the deterioration of the gene pool of human populations," he said.

Dr Stefansson, however, told BBC News that from a long-term perspective the decision by some men to have children later in life might well be speeding up the evolution of our species.

"The high rate of mutations is dangerous for the next generation but is generating diversity from which nature can select and further refine this product we call man," he said.

"So what is bad for the next generation may be good for our species in general."

Follow Pallab **on Twitter**

More Health stories

UK trials heart failure impant op [/news/health-19344001]

A pioneering operation in the UK to fit a nerve-stimulating implant in a patient with heart failure is due to finish later.

DRC Ebola outbreak 'kills ten' [/news/world-africa-19346753]

Antibiotics link to child weight [/news/health-19341639]

NATURE | NEWS

Fathers bequeath more mutations as they age

Genome study may explain links between paternal age and conditions such as autism.

Ewen Callaway

22 August 2012

In the 1930s, the pioneering geneticist J. B. S. Haldane noticed a peculiar inheritance pattern in families with long histories of haemophilia. The faulty mutation responsible for the blood-clotting disorder tended to arise on the X chromosomes that fathers passed to their daughters, rather than on those that mothers passed down. Haldane subsequently proposed[1] that children inherit more mutations from their fathers than their mothers, although he acknowledged that "it is difficult to see how this could be proved or disproved for many years to come".

That year has finally arrived: whole-genome sequencing of dozens of Icelandic families has at last provided the evidence that eluded Haldane. Moreover, a study published in *Nature* findsthat the age at which a father sires children determines how many mutations those offspring inherit[2]. By starting families in their thirties, forties and beyond, men could be increasing the chances that their children will develop autism, schizophrenia and other diseases often linked to new mutations. "The older we are as fathers, the more likely we will pass on our mutations," says lead author Kári Stefánsson, chief executive of deCODE Genetics in Reykjavik. "The more mutations we pass on, the more likely that one of them is going to be deleterious."

Older fathers' sperm have more mutations — as do their children.

V. PEÑAFIEL/FLICKR/GETTY

Haldane, working years before the structure of DNA was determined, was also correct about why fathers pass on more mutations. Sperm is continually being generated by dividing precursor cells, which acquire new mutations with each division. By contrast, women are born with their lifelong complement of egg cells.

Stefánsson, whose company maintains genetic information on most Icelanders, compared the whole-genome sequences of 78 trios of a mother, father and child. The team searched for mutations in the child that were not present in either parent and that must therefore have arisen spontaneously in the egg, sperm or embryo. The paper reports the largest such study of nuclear families so far.

Nature Podcast

deCODE's Kári Stefánsson explains to Ewen Callaway how a father's age might affect a baby's risk of disease.

00:00

Go to full podcast

Fathers passed on nearly four times as many new mutations as mothers: on average, 55 versus 14. The father's age also accounted for nearly all of the variation in the number of new mutations in a child's genome, with the number of new mutations being passed on rising exponentially with paternal age. A 36-year-old will pass on twice as many mutations to his child as a man of 20, and a 70-year-old eight times as many, Stefánsson's team estimates.

The researchers estimate that an Icelandic child born in 2011 will harbour 70 new mutations, compared with 60 for a child born in 1980; the average age of fatherhood rose from 28 to 33 over that time.

Most such mutations are harmless, but Stefánsson's team identified some that studies have linked to conditions such as autism and schizophrenia. The study does not prove that older fathers are more likely than younger ones to pass on disease-associated or other deleterious genes, but that is the strong implication, Stefánsson and other geneticists say.

Previous studies have shown that a child's risk of being diagnosed with autism increases with the father's age. And a trio of papers[3–5] published this year identified dozens of new mutations implicated in autism and found that the mutations were four times more likely to originate on the father's side than the mother's.

Related stories

- Autism linked to hundreds of spontaneous genetic mutations
- Human-chimp interbreeding challenged
- Human mutation rate revealed

More related stories

The results might help to explain the apparent rise in autism spectrum disorder: this year, the US Centers for Disease Control and Prevention in Atlanta, Georgia, reported that one in every 88 American children has now been diagnosed with autism spectrum disorder, a 78% increase since 2007. Better and more inclusive autism diagnoses explain some of this increase, but new mutations are probably also a factor, says Daniel Geschwind, a neurobiologist at the University of California, Los Angeles. "I think we will find, in places where there are really old dads, higher prevalence of autism."

However, Mark Daly, a geneticist at Massachusetts General Hospital in Boston who studies autism, says that increasing paternal age is unlikely to account for all of the rise in autism prevalence. He notes that autism is highly heritable, but that most cases are not caused by a single new mutation — so there must be predisposing factors that are inherited from parents but are distinct from the new mutations occurring in sperm.

Historical evidence suggests that older fathers are unlikely to augur a genetic meltdown. Throughout the seventeenth and eighteenth centuries, Icelandic men fathered children at much higher ages than they do today, averaging between 34 and 38. Moreover, genetic mutations are the basis for natural selection, Stefánsson points out. “You could argue what is bad for the next generation is good for the future of our species,” he says.

Nature **488,** 439 (23 August 2012) doi:10.1038/488439a

See News & Views, page 467 and Article, page 471

References

1. Haldane, J. B. S. *Ann. Eugen.* **13**, 262–271 (1947).
   Show context Article PubMed ChemPort
2. Kong, A. *et al*. *Nature* **488**, 471–475 (2012).
   Show context Article
3. Sanders, S. J. *et al*. *Nature* **485**, 237–241 (2012).
   Show context Article PubMed ISI ChemPort
4. O’Roak, B. J. *et al*. *Nature* **485**, 246–250 (2012).
   Show context Article PubMed ISI ChemPort
5. Neale, B. M. *et al*. *Nature* **485**, 242–245 (2012).
   Show context Article PubMed ISI ChemPort

Related stories and links

From nature.com

- **Autism linked to hundreds of spontaneous genetic mutations**
  09 June 2011
- **Human-chimp interbreeding challenged**
  28 August 2009
- **Human mutation rate revealed**
  27 August 2009
- **Genetic hotspot for autism found**

07 January 2008

- **Nature News special: The autism enigma**

From elsewhere

- **deCODE Genetics**
- **Mark Daly**
- **Daniel Geschwind**

Comments

2012-08-22 06:33 AM

Hamid Sadeghi said: If a disease can be linked to gene mutation in sperm it needs to take place exactly at the same spot where gene defect for that disease is located. Even if a point mutation is responsible for a disease to emerge so the chance for that is 1 in 3 billion and increasing the likelihood 8 folds it will be 1 in 375 million, not a big difference, so for autism most likely it is a combination of hereditary and environmental factors and both parents seems to contribute equally.

You need to be registered with *Nature* and agree to our Community Guidelines to leave a comment. Please log in or register as a new user. You will be re-directed back to this page.

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687