Recombination rate and reproductive success in humans

Augustine Kong[1], John Barnard[2], Daniel F Gudbjartsson[1], Gudmar Thorleifsson[1], Gudrun Jonsdottir[1], Sigrun Sigurdardottir[1], Bjorgvin Richardsson[1], Jonina Jonsdottir[1], Thorgeir Thorgeirsson[1], Michael L Frigge[1], Neil E Lamb[3], Stephanie Sherman[3], Jeffrey R Gulcher[1] & Kari Stefansson[1]

Intergenerational mixing of DNA through meiotic recombinations of homologous chromosomes during gametogenesis is a major event that generates diversity in the eukaryotic genome. We examined genome-wide microsatellite data for 23,066 individuals, providing information on recombination events of 14,140 maternal and paternal meioses each, and found a positive correlation between maternal recombination counts of an offspring and maternal age. We postulated that the recombination rate of eggs does not increase with maternal age, but that the apparent increase is the consequence of selection. Specifically, a high recombination count increased the chance of a gamete becoming a live birth, and this effect became more pronounced with advancing maternal age. Further support for this hypothesis came from our observation that mothers with high oocyte recombination rate tend to have more children. Hence, not only do recombinations have a role in evolution by yielding diverse combinations of gene variants for natural selection, but they are also under selection themselves.

We recently constructed a high-resolution genetic map that highlights the variation in recombination rates between the sexes and across the genome[1]. We confirmed a previous observation[2] that recombination rates among mothers can differ substantially and observed that even gametes of one mother have different recombination rates; a gamete with a high recombination count in one chromosome tends also to have high recombination counts in other chromosomes[1]. Here, we focused on determining whether recombination rate is related to the age of the mother. The chiasma frequency of mouse oocytes is reported to decrease as the mouse ages[3]. It was suggested that a reduction in crossing over leading to formation of univalents might explain age-dependent nondisjunction. As chiasma formation occurs prenatally, however, the 'production line' hypothesis was proposed. This hypothesis states that there is a gradient in the fetal ovary, so that the first-formed oocytes have a higher chiasma frequency than those formed later, and that oocytes are ovulated in the same order that they enter meiosis. Attempts to validate this model have been equivocal[4,5];

however, studies in the mouse suggest that the last-formed oocytes are also the last to be ovulated[6].

In humans, a number of studies have been done to estimate recombination counts using genetic data from families (that is, parent-offspring transmissions), but none has provided convincing evidence that the recombination count in an ooctye is correlated with maternal age. A reported decrease in recombination with increasing maternal age using the Venezuelan Reference Pedigree[7] could not be replicated by further analysis using the same data source[8]. Most earlier studies were based on small sample sizes and were not genome-wide investigations[9–11]. Two genome-wide studies[1,2] did not detect a statistically significant age effect. Suspecting that the failures of previous studies to detect an effect were due to the lack of power, we carried out a large study using two primary resources: a genetic database with genotypic data on ~1,000 microsatellite markers typed in 70,000 individuals and a genealogy database covering the entire Icelandic nation. We used these to construct a data set consisting of 5,463 families, with 23,066 individuals genotyped (average yield >800 genotypes per person) and providing information on 14,140 maternal and paternal meioses each. These are nuclear families with two or more siblings and at least one parent genotyped (**Table 1**). Our genealogy database provides the birth years of the individuals, rounded up to the nearest five years to protect privacy[12]. We calculated the approximate age of the mother at the birth of every child and determined the total number of children a mother had, regardless of genotype status. We chose families in which the mothers were

Table 1 Count of families according to the number of genotyped children and the status of parental genotyping

Parent genotyped	Number of genotyped children				
	2	3	4	5 or more	Total
Father only	342	178	58	31	609
Mother only	885	345	102	59	1,391
Father and mother	2,059	956	307	141	3,463
Total	3,286	1,479	467	231	5,463

Table 2 Distribution of mother's age at birth for all offspring

Mother's age at birth	Count	Percent
15	317	2.2
20	3,340	23.6
25	4,662	33.0
30	3,386	23.9
35	1,770	12.5
40	595	4.2
45–50	70	0.5
Total	14,140	100.0

born between 1925 and 1955. **Table 2** gives the distribution of maternal age at birth.

Genotype data do not provide complete information on recombination counts, which complicates the analysis. To handle this missing-data problem[13], we applied two different statistical methods. The first method, called 'mean imputation', imputes the recombination counts using the best guesses. The way we implemented this method makes it robust, meaning that the calculated P values are insensitive to model mis-specifications or potential artifacts in the data. There is some loss of efficiency, however, and effects tend to be underestimated. The second method is likelihood-based, is fully efficient but computationally intensive, and can be sensitive to model mis-specification.

Using the robust method, we estimated the effect of maternal age on recombination rate to be 0.043 recombinations per year (s.e. = 0.011; $P = 0.00016$). Because we used family-adjusted recombination counts and the ages of mothers at birth, the age trend that we detected existed 'within family' (*i.e.*, a child born to a mother later in life tends to have more maternal recombinations than a child born earlier in her life) and was not simply a consequence of the possibility that some mothers tend to have children later in life and also happen to have higher recombination rates. The likelihood-based method gives an estimate of 0.082 recombinations per year (s.e. = 0.012; $P < 1 \times 10^{-8}$).
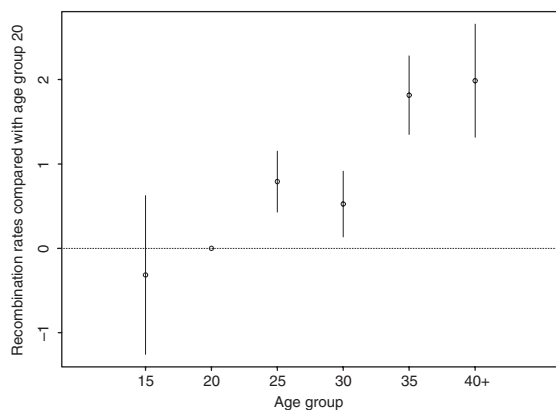

Figure 1 Recombination rate and maternal age. Using the age group 20 as the reference, the estimates and 95% confidence intervals for the differences in recombination rates between the other age groups and age group 20 are shown. Maternal ages of 40 and more were grouped into a single bin. There is a trend towards increase, but the data deviate from linearity with a slight drop in the estimate from age 25 to 30 followed by a big jump from age 30 to 35. Although the drop from age 25 to 30 is not statistically significant and probably not real, the data do support a relatively big incremental increase from age 30 to 35. An exact linear relationship between recombination rate and maternal age is not consistent with the data and is rejected by a goodness of fit test ($P < 0.005$).

Although the effect is significant even with the conservative method, the higher estimate based on the likelihood method is probably better. To determine whether the age effect is well fitted by a linear relationship, we fitted a model treating maternal age as a categorical variable using the likelihood method (**Fig. 1**; the distributions of maternal recombination counts of individual offspring are shown in **Supplementary Fig. 1** online). The age effect is already apparent for relatively young women, and there is a marked increase in recombination rate from age 30 to 35. Notably, the rate of increase of maternal nondisjunction accelerates during this time frame.

The maternal age effect translates into only an additional two recombinations, or ∼4% of the average maternal recombination rate, over a period of 25 years. But the relevance and importance of the observed effect depends on the underlying causes. There are at least two possible explanations for the results: first, recombination rate among the eggs of a woman increases as she ages; and second, the recombination rate of eggs does not increase, but there is a selection effect that increases the chance of an egg with more recombinations to produce a successful live birth. This selection effect probably exists even early on, but becomes stronger as the woman ages.

The first explanation is unlikely to be true, because recombinations take place prenatally and a 'production line' hypothesis would have to be invoked as outlined above. Moreover, this increase contradicts the observed decrease in chiasma frequency reported for mice. The second explanation, related to selection forces, is more plausible. A higher number of recombinations along a chromosome might reduce the chance of maternal age–related nondisjunction, the leading cause of pregnancy loss due to aneuploidy in the fetus. Maternal nondisjunction is associated with maternal age and reduced levels or altered placement of recombination[14]. Altered recombination has been identified for all examined cases of trisomy arising at the first stage of maternal meiosis. Consequently, increasing amounts of meiotic recombination may be protective for certain forms of nondisjunction, depending on the location of the additional exchanges. There is evidence, at least for chromosome 15, that multiple recombinants may be more resistant to nondisjunction because of increased stability of the bivalent over time[15]. Age-related abnormalities in spindle morphology and chromosome alignment at the meiotic plate have been reported[16], suggesting that some components of the meiotic apparatus are susceptible to the effects of aging. It has also been proposed that the sister chromatid cohesion complex may suffer an age-related breakdown[17]. If this is true, then meiotic tetrads from older oocytes may retain their integrity on the basis of their chaismata alone. Greater numbers of recombination would then provide additional protection from age-related meiotic breakdown.

Under the selection hypothesis, women with higher recombination rates would have more children. To examine this possibility, we regressed the total number of children of the mother on (i) the (estimated) recombination rate of the mother; (ii) the number of genotyped children of the mother; (iii) the mother's mean age at the

Table 3 Estimated effects for four predictors of family size

Predictor	Estimated effect	Standard error (s.e.)	P value
Recombination rate of mother	0.0109	0.0041	0.0076
Number of genotyped children of the mother	0.6815	0.0212	0.0000
Mother's average age at birth of her children	0.0002	0.0045	0.9637
Birth date of mother	−0.0469	0.0022	0.0000

Table 4 Estimated effect of mother's age on recombination rate for different parts of the genome

	Centromeric half	Telomere minus 6 cM	Telomeric 6 cM
Female genetic length (cM)	23.904	20.483	1.152
Recombination increase per year (s.e.)	0.038 (0.0072)	0.041 (0.0065)	0.009 (0.0014)
Percent increase per year (s.e.)	0.158 (0.0299)	0.201 (0.0316)	0.761 (0.1231)

The centromeric half consists of the part of each chromosome arm that is next to the centromere and accounts for roughly one-half of that chromosome arm in female genetic distance. The telomeric 6-cM region is defined as in **Supplementary Table 1** online.

times of birth of the genotyped children; and (iv) the mother's birth date (**Table 3**). As expected, the number of genotyped children of the mother is correlated with the total number of children of a mother, but the correlation is not perfect ($R^2 = 0.19$). Its inclusion in the regression ensures that any correlation observed between family size and recombination rate of a mother is not spurious: a higher number of recombinations is not estimated or detected simply because more children are genotyped. After accounting for the generational trend, recombination rate has a positive and statistically significant effect ($P = 0.0076$) on family size. With the mother's mean age at the times of birth of the genotyped children, which happens to be non-significant, also included in the regression, mothers who have a larger number of children have a higher recombination rate not simply because they have more children at a later age. Although it is significant, the effect of recombination rate on family size is modest. This is not surprising, as many factors affect family size.

We investigated whether the maternal-age effect is specific to certain genomic regions, as data from nondisjoined chromosomes indicate that there is selection against specific chiasmatic configurations[18]. The age effect is roughly the same for long and short chromosomes. Dividing each chromosome arm into two roughly equal parts on the basis of female genetic distance, the telomeric halves have a slightly higher percentage increase per year than the centromeric halves, but the difference is not significant. Focusing on marker intervals within 6 cM of our most telomeric marker (**Supplementary Table 1** online), we determined that the percentage increase per year for these telomeric regions is roughly four times higher than that of the rest of the genome ($P < 0.0001$; **Table 4**). Because these regions account for only ~2.5% of the genome in genetic length, however, ~90% of the yearly increase of recombinations observed is accounted for by the other parts of the genome.

We observed no association of recombination rates with paternal age (**Supplementary Table 2** and **Supplementary Fig. 2** online), nor did we identify a systematic difference in recombination rates between or within men. A previous report using an immunofluorescence method to examine exchanges in human spermatocytes described significant variation in recombination rates within and among men, but no age effect[19]. The observed variation identified among spermatocytes, but not live births, suggests that selection occurs at the level of spermatocytes. Presumably, the checkpoints for such meiotic disruptions are more stringent in spermatocytes than in oocytes[17].

The proposed selection hypothesis explains the maternal-age effect and the correlation of maternal recombination rate with family size. But there could be alternative explanations. A recent paper[20] challenged the doctrine that all the oocytes of a woman are produced when the woman is still at her fetal stage and suggested that follicular renewal may occur in the postnatal mammalian ovary. If true, this would provide a natural time ordering of the oocytes that corresponds to the dates of birth of the children and an alternative explanation for the age effect we observed. But this theory does not explain why mothers who have higher recombination rates have more children.

Our observations and hypotheses do not contradict this new theory; there could be both follicle renewal and selection associated with recombination counts.

Among the 5,463 families studied, 1,090 mothers make up 545 independent sister pairs. Based on the correlation of estimated recombination rates of these sisters, the heritability[21] of recombination rate is estimated to be 30.4% (s.e. = 8.5%; $P = 0.0004$), which supports the idea that there is a large genetic component to recombination rate. Together with the hypothesis that reproductive success of eggs is dependent on the number of recombinations they have across the genome, these data imply that not only do recombinations have a role in evolution by yielding diversity of combinations of gene variants for natural selection, but they are also under selective forces acting at the level of chromosome segregation and reduced survival of mis-segregated oocytes.

METHODS

Data collection and genotyping. We obtained all the biological samples used in this study according to protocols approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. We obtained informed consent from all participants. All personal identifiers were encrypted using a code that is held by the Data Protection Commission of Iceland[12]. Details concerning genotyping, allele-calling and genotype quality control can be found in the supplemental material of our previous study[1].

Statistical methods. The first method we applied to study the age effect is called mean imputations[13] and is similar to the method we used previously[1]. With all the family data, we first fitted a male and a female genetic map using maximum likelihood and the EM algorithm[22]. We then calculated the expected paternal and maternal recombination counts of each child conditional on the observed genotype data and the fitted maps. We calculated the conditional expected recombination counts using the simulation option of our linkage program Allegro[23]; we carried out 100 simulations and computed the averages. We then treated these estimated recombination counts as though they were the actual recombination counts in the subsequent analysis. To study the age effect, we regressed the family-adjusted recombination counts on the family-adjusted age of mother at birth of the child (the family-adjusted value is the difference between the value of a child and the value averaging over all the children in the same family). Using the family-adjusted values not only ensures that any potential artifacts are eliminated, because any bias would have the same effect on all children in the family, but it also implies that any age effect detected will not be confounded by the differences between mothers. We obtained the P values through a randomization test. The children in a family were permuted and the analysis repeated. We did this 25,000 times, and the two-sided P value reported is twice as large as the fraction of times that the permutations produced an estimated effect bigger than or equal to the observed effect of 0.043. This method is robust and completely insensitive to model mis-specifications. There is, however, loss of efficiency and the effect tends to be underestimated. This is because the mean imputations are done under the model of no age effect; hence, the estimated effect has a tendency to shrink towards the null hypothesis. The amount of shrinkage is proportional to the amount of missing information, which is expected to be quite small for the data set in our previous study[1] as over 5,000 markers were used there; but, as only ~1,000 markers were used here, the shrinkage is expected to be greater. Also,

when there are only two siblings genotyped and the genotypes of grandparents are not available, the data can provide a good estimate of the total maternal and paternal recombination counts in the family, but the data are completely uninformative regarding whether a crossover occurs with one child or the other, and, as a consequence, the mean imputations would be the same for both siblings. Hence, when using mean imputations in conjunction with a family-adjusted analysis, families with two siblings genotyped are completely uninformative. As including these families would not add information but would further shrink the estimated effect, we used only the 2,177 families with three or more siblings genotyped when applying this method. But we used all 5,463 families for the likelihood approach described below, and also when the mean imputations were used to study the relationship between family size and recombination rate of a mother.

The second method we used was a full-likelihood approach, which is maximally efficient but may not be as robust as our first method. It is based on a model that assumes a multiplicative maternal effect which is constant across the genome, that is, any effect on recombination rate is assumed to affect all genomic regions equally, unless otherwise specified. The multiplicative effect is modeled as a function of the mother, the gamete and the age of the mother at birth. The mother and gamete effects are modeled as random effects and the age effect is modeled as a fixed effect. We employed the following model for the mean number of maternal crossovers per gamete per chromosome: $\mu_{mgc} = \exp(\alpha_c + \beta \times age_{mg}) \times \exp(U_m + U_{mg})$, where m indexes mother, g indexes gamete, c indexes chromosome and U_m and U_{mg} are assumed to be independent and normally distributed. The full model can be viewed as a generalized linear mixed model[24] with a Poisson random conditional component, log link and normally distributed random effects. To simplify the analysis, we assumed the absence of any crossover interference throughout the model. This can create some biases as interference exists[25], but the effect is likely to be modest for the parameters of interest in this study. To transform from the multiplicative scale to the additive scale, we took $(\exp(\beta) - 1)$ times the total estimated length of the maternal genome as the additive effect of maternal age on recombination rate. Because the recombination counts are not directly observed, even with the assumption of no interference, maximizing the likelihood under both the null hypothesis (no age effect) and the alternative hypothesis is challenging, going beyond the difficulties found in standard generalized linear mixed models. To meet these computational challenges we applied various computational techniques that include the Monte Carlo Newton-Raphson algorithm, Monte Carlo EM-algorithm and importance-reweighting of the samples simulated from Allegro under the null hypothesis[24,26,27]. Standard errors for the model parameters are determined on the basis of the observed Fisher information, and P values are obtained on the basis of likelihood ratio tests. When studying potential differences between genomic regions, a separate α_c and β are assigned to each genomic group.

In **Figure 1** and **Supplementary Figure 2** online, the confidence intervals have incorporated the uncertainties of both the recombination rate of the particular age bin and the recombination rate of age bin 20. In **Supplementary Figure 1** online, the box plots are constructed using a central box, indicating the range of the middle 50% of recombination (from the first to the third quartiles) with the median value indicated by a horizontal bar within the box, and whiskers (the dashed vertical lines and the hinged horizontal lines at their edge) that extend to the furthest data point that is not more than one-and-a-half times the width of the interquartile range beyond the central box. All other recombination values further away are indicated individually by horizontal lines.

Note: Supplementary information is available on the Nature Genetics website.

1. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
2. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. & Weber, J.L. Comprehensive human genetic map: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
3. Henderson, S.A. & Edwards, R.G. Chiasma frequency and maternal age in mammals. *Nature* **218**, 22–28 (1968).
4. Speed, R.M. & Chandley, A.C. Meiosis in the foetal mouse ovary. II. Oocyte development and age-related aneuploidy. Does a production line exist? *Chromosoma* **88**, 184–189 (1983).
5. Polani, P.E. & Jagiello, G.M. Chiasmata, meiotic univalents and age in relationship to aneuploid imbalance in mice. *Cytogenet. Cell. Genet.* **16**, 505–529 (1976).
6. Polani, P.E. & Crolla, J.A. A test of the production line hypothesis of mammalian oogenesis. *Hum. Genet.* **88**, 64–70 (1991).
7. Tanzi, R.E. *et al.* A genetic linkage map of human chromosome 21: analysis of recombination as a function of sex and age. *Am. J. Hum. Genet.* **50**, 551–558 (1992).
8. Haines, J.L. *et al.* A genetic linkage map of chromosome 21: a look at meiotic phenomena. *Prog. Clin. Biol. Res.* **384**, 51–61 (1993).
9. Elston, R.C., Lange, K. & Namboodiri, K.K. Age trends in human chiasma frequencies and recombination fractions. II method for analyzing recombination fractions and applications to the ABO:nail-patella Linkage. *Am. J. Hum. Genet.* **28**, 69–76 (1976).
10. Renwick, J.H. & Schulze, J. Male and female recombination fractions, for the nailpatella:ABO linkage in man. *Ann. Hum. Genet.* **28**, 379–392 (1965).
11. Weitkamp, L.R. The relation of parental sex and age to recombination in the HLA system. *Hum. Hered.* **23**, 197–205 (1973).
12. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
13. Little, R.J.A. & Rubin, D.B. *Statistical Analysis with Missing Data* (John Wiley & Sons, New York, 1987).
14. Hassold, T.J., Sherman, S. & Hunt, P. Counting cross-overs: characterizing meiotic recombination in mammals. *Hum. Mol. Genet.* **9**, 2409–2419 (2000).
15. Robinson, W. *et al.* Maternal meiosis I nondisjunction of chromosome 15: dependence of the maternal age effect on the level of recombination. *Hum. Mol. Genet.* **7**, 1011–1109 (1998).
16. Hodges, C.A. *et al.* Experimental evidence that changes in oocyte growth influence meiotic chromosome segregation. *Hum. Reproduction.* **17**, 1171–1180 (2002).
17. Hunt, P.A. & Hassold, T.J. Sex matters in meiosis. *Science* **296**, 2181–2183 (2002).
18. Lamb, N.E. *et al.* Characterization of susceptible chiasma configurations that increase the risk for maternal nondisjuction of chromosome 21. *Hum. Mol. Genet.* **6**, 1391–1399 (1997).
19. Lynn, A. *et al.* Covariation of synaptonemal complex length and mammalian meiotic exchange rates. *Science* **296**, 2222–2225 (2002).
20. Johnson, J., Canning, J., Kaneko, T., Pru, J.K. & Tilly, T.L. Germline stem cells and follicular renewal in the postnatal mammalian ovary. *Nature* **428**, 145–150 (2004).
21. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Massachusetts, 1998).
22. Dempster, A.P., Laird, N.M. & Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodol.* **39**, 1–38 (1977).
23. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* **25**, 12–14 (2000).
24. McCulloch, C.E. Maximum likelihood algorithms for generalized linear mixed models. *J. Am. Stat. Assoc.* **92**, 162–170 (1997).
25. Broman, K.W. & Weber, J.L. Characterization of human crossover interference. *Am. J. Hum. Genet.* **66**, 1911–1926 (2000).
26. Irwin, M., Cox, N.J. & Kong, A. Sequential imputation for multilocus linkage analysis. *Proc. Natl. Acad. Sci. USA* **91**, 11684–11688 (1994).
27. Quintana, F.A., Liu, J.S. & del Pino, G.E. Monte carlo EM with importance reweighting and its applications in random effects model. *Comput. Stat. Data Anal.* **29**, 429–444 (1999).