The rate of meiotic gene conversion varies by sex and age

Bjarni V Halldorsson\textsuperscript{1,2}, Marteinn T Hardarson\textsuperscript{1}, Birke Kehr\textsuperscript{1}, Unnur Styrkarsdottir\textsuperscript{1}, Arnaldur Gylfason\textsuperscript{1}, Gudmar Thorleifsson\textsuperscript{1}, Florian Zink\textsuperscript{1}, Adalbjorg Jonasdottir\textsuperscript{1}, Aslaug Jonasdottir\textsuperscript{1}, Patrick Sulem\textsuperscript{1}, Gisli Masson\textsuperscript{1}, Uunn Thorsteinsdottir\textsuperscript{1,3}, Agnar Helgason\textsuperscript{1,4}, Augustine Kong\textsuperscript{1}, Daniel F Gudbjartsson\textsuperscript{1,5} & Kari Stefansson\textsuperscript{1,3}

Meiotic recombination involves a combination of gene conversion and crossover events that, along with mutations, produce germline genetic diversity. Here we report the discovery of 3,176 SNP and 61 indel gene conversions. Our estimate of the non-crossover (NCO) gene conversion rate ($G$) is 7.0 for SNPs and 5.8 for indels per megabase per generation, and the GC bias is 67.6\%.

For indels, we demonstrate a 65.6\% preference for the shorter allele. NCO gene conversions from mothers are longer than those from fathers, and $G$ increases with the age of mothers, but not the age of fathers. A disproportionate number of NCO gene conversions in older mothers occur outside double-strand break (DSB) regions and in regions with relatively low GC content. This points to age-related changes in the mechanisms of meiotic gene conversion in oocytes.

In previous publications, we quantified the rate of chromosomal crossover in meiosis and its determinants and genomic distribution in humans\textsuperscript{17-19}. Here we report the discovery of allelic gene conversions in meiosis inferred from the germline genotypes of probands and their relatives. We adopted a study design similar to that of Williams et al.,\textsuperscript{20} based on probands in three-generation families, where genotypes are available for a proband, the proband’s spouse, both the proband’s parents, at least two of the proband’s siblings and at least one child. This family structure was chosen to limit the impact of genotyping errors that can mimic gene conversion, such that haplotypes carried by the parents could be independently verified in the siblings and gene conversion could be verified in the child. We searched for contiguous tracts of markers (no longer than 100 kb) consistent with non-reciprocal transfer of chromosome fragments in a transmission from parent to proband. This study design is adapted for the detection of NCO gene conversions, which are typically shorter than 100 kb and where the converted tract is flanked by sequence from the original recipient chromosome on both sides\textsuperscript{4,20}. Our study design does not allow for detection of the subset of CO gene conversions, where the converted tract is flanked by sequence from the original recipient chromosome on only one side\textsuperscript{4,8}.

However, we are able to detect crossover recombination events that are accompanied by one or more such gene conversion events—hereafter referred to as complex crossover (CCO) gene conversions\textsuperscript{20-22}.

Using this approach, we sought gene conversions in two overlapping data sets. The first consists of 7,219 proband–family sets genotyped on Illumina HumanHap and Omni BeadChip arrays (chip data set). The second consists of 101 whole-genome-sequenced (at $>30\times$ coverage) proband–family sets\textsuperscript{23} (sequencing data set), 91 of which are contained in the first data set (Fig. 1).

\textsuperscript{1}deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. \textsuperscript{2}School of Science and Engineering, Reykjavik University, Reykjavik, Iceland. \textsuperscript{3}Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. \textsuperscript{4}Department of Anthropology, University of Iceland, Reykjavik, Iceland. \textsuperscript{5}School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. Correspondence should be addressed to B.V.H. (bjarnih@decode.is) or K.S. (kstefans@decode.is).

Received 13 October 2015; accepted 16 August 2016; published online 19 September 2016; doi:10.1038/ng.3669
RESULTS
As gene conversions can only be detected at polymorphic markers, we restricted our analysis to high-quality SNPs and indels with a minor allele frequency over 0.5%. Indels were further restricted to be shorter than 10 bp, as longer indels are rare and yield less reliable genotypes. This restriction left 624,955 SNPs in the chip data set and 8,195,014 SNPs and 465,054 indels in the sequencing data set. For each proband–family set, we considered only variants where at least one parent was heterozygous, the haplotypes transmitted by both parents could be verified in a sibling and gene conversion could be verified in an offspring. Using this approach, we assessed 214,241,663 marker proband pairs (MPPs) in the chip data set, yielding 2,192 MPPs involved in a gene conversion (Tables 1 and 2). In the sequencing data set, 147,368,280 SNP and 8,715,873 indel MPPs were assessed, yielding 1,027 and 61 gene-converted MPPs, respectively (Supplementary Data).

As most sequenced individuals also had chip data, it follows that a subset of 5,233,293 chip data MPPs can be informative about the sensitivity and specificity of gene conversion calling in the two sets. From this subset, 1,060,127 were omitted from the sequencing data set because of quality control thresholds, leaving 4,173,166 overlapping MPPs that were assessed in both data sets. All yielded concordant results, including an overlap of 43 gene conversions (Supplementary Note). We further confirmed phased haplotypes using read pairs (Supplementary Table 1 and Supplementary Note) and genotypes using Sanger sequencing and whole-genome sequencing (Supplementary Table 2 and Supplementary Note), which showed error rates between 0.0 and 1.1%.

NCO gene conversion in DSB regions, recombination hotspots
We first examined the distribution of NCO gene conversions in the genome and their rate over recombination hotspots. As most such events are thought to be due to programmed DSBs, we compared our predicted NCO gene conversions to a map of DSB regions generated from the spermatocyte samples of five human males24, which have a mean size of 1,464 bp, s.d. = 586 bp. Table 1 shows that NCO gene conversions are highly over-represented in spermatocyte DSB regions (odds ratio (OR) > 10).

In paternal transmissions, the over-representation was 42.3-fold and 45.7-fold in the chip and sequencing data sets, respectively. In maternal transmissions, the over-representation was only 5.4-fold and 7.1-fold in the chip and sequencing data sets, respectively, albeit highly significant (P < 0.001, for both data sets). As the locations of crossover recombination hotspots are known to differ between male and female meioses18 and hotspots are largely determined by DSB regions24, it follows that a stronger elevation of maternal G would be expected against a map of DSB regions from oocytes (such a map is currently not available). A previous study20 also showed an over-representation of NCO gene conversions in male DSB regions and indications of a sex difference in their localization. Recent research has shown that PRDM9 alleles carried by the parent strongly influence the locations of DSBs24 and, consequently, the locations of crossovers and NCOs in chromosomes transmitted to the offspring. Our results confirm these findings. The PRDM9 allele of the proband strongly affected the distribution of NCO gene conversions, but not their over-representation in DSB regions (Supplementary Table 3 and Supplementary Note).

We next compared NCO gene conversions to a sex-specific map of crossover recombination hotspots25. Crossover recombination, when estimated in the sexes separately, were enriched 38.0- and 27.6-fold in male and female crossover recombination hotspots, respectively17. Table 1 shows that, for males, the enrichment of G in male crossover recombination hotspots was 8.2- and 12.7-fold in the chip and sequencing data sets, respectively. For females, the enrichment of G was 4.6- and 8.0-fold in female crossover recombination hotspots.

Figure 1 The study design and method used for detecting gene conversions. (a) Family structure. (b) The two sources of genotypes. (c) The genotypes of the siblings are necessary to verify the haplotypes of both parents. We require at least one child of the proband to verify gene conversion. The method is shown for NCO gene conversions.
Table 1 Estimates of NCO gene conversion of SNPs

|                | Paternal  | Maternal | Total | Paternal  | Maternal | Total |
|----------------|-----------|----------|-------|-----------|----------|-------|
| Converted MPPs | 641       | 1,393    | 2,034 | 300       | 565      | 865   |
| n MPPs considered | 106,740.626 | 106,761.366 | 213,501.992 | 73,919.056 | 73,016.289 | 146,935.345 |
| Observed gene conversion rate per Mb (95% CI) | 6.0 (5.5–6.5) | 13.0 (12.0–14.1) | 9.5 (9.0–10.1) | 4.1 (3.4–4.8) | 7.7 (5.7–9.9) | 5.9 (4.9–7.0) |
| Corrected gene conversion rate per Mb (95% CI) | 3.9 (3.5–4.4) | 10.0 (8.5–11.6) | 7.0 (6.0–8.0) | – | – | – |
| n events | 618       | 1,018    | 1,636 | 244       | 241      | 485   |
| Short | 606       | 849      | 1,455 | 234       | 196      | 430   |
| Long | 12        | 169      | 181   | 10        | 45       | 55    |
| Gene-converted MPPs/event (SE) | 1.04 (0.01) | 1.37 (0.03) | 1.24 (0.02) | 1.23 (0.06) | 2.34 (0.28) | 1.78 (0.15) |
| Short | 1.01 (0.00) | 1.04 (0.01) | 1.03 (0.01) | 1.09 (0.02) | 1.17 (0.04) | 1.13 (0.02) |
| Long | –         | 3.01 (0.13) | 2.96 (0.12) | –         | 7.44 (1.15) | 6.93 (0.99) |
| GC bias (%) (95% CI) | 64.4 (60.7–68.3) | 69.1 (66.6–71.6) | 67.6 (65.7–69.8) | 63.1 (57.1–69.1) | 72.5 (69.1–76.1) | 69.3 (65.8–72.3) |
| Short | 64.7 (61.1–68.5) | 69.5 (66.4–72.3) | 67.5 (65.3–69.7) | 65.3 (58.3–72.0) | 73.0 (66.5–79.7) | 68.9 (64.1–74.1) |
| Long | –         | 68.6 (64.5–72.3) | 67.9 (64.0–71.7) | –         | 72.2 (67.2–77.4) | 69.8 (64.9–74.5) |
| OR DSB (95% CI) | 42.3 (35.2–50.9) | 5.4 (4.7–6.2) | 10.7 (9.6–11.9) | 45.7 (32.8–67.1) | 7.1 (5.2–9.7) | 14.9 (11.3–19.8) |
| Short | 49.7 (41.2–61.1) | 9.1 (7.9–10.5) | 17.5 (15.9–19.4) | 66.9 (49.4–92.1) | 20.3 (15.7–25.8) | 37.0 (30.2–45.7) |
| Long | –         | 1.1 (0.7–1.5) | 1.1 (0.7–1.5) | –         | 1.6 (0.9–2.3) | 2.0 (1.2–3.0) |
| OR crossover recombination hotspots (95% CI) | 8.2 (6.6–10.1) | 4.6 (3.8–5.4) | –         | 12.7 (8.6–18.3) | 8.0 (5.7–11.1) | – |
| Short | 8.8 (7.1–10.6) | 6.5 (5.4–7.7) | –         | 14.8 (10.7–20.5) | 14.1 (9.8–19.5) | – |
| Long | –         | 2.0 (1.1–3.2) | –         | –         | 4.0 (1.5–6.9) | – |

The corrected gene conversion rate is corrected for the crossover recombination rate. This correction and the odds ratio for hotspots are computed using sex-specific crossover recombination maps. Short events, <1,000 bp; long events, ≥1,000 bp. Results for long NCO events are not presented in paternal transmissions because of small sample size. As crossover maps are sex-specific, no joint analysis is provided. NCO, non-crossover; OR, odds ratio; DSB, double-strand break; MPP, marker proband pair; CI, confidence interval; SE, standard error.

NCO gene conversion rate

Williams et al. estimated G as 5.9 conversions/Mb per generation, using a genome-wide approach similar to the one presented here (but restricting to events shorter than 5 kb). Comparable estimates have been obtained based on sperm genotyping6,7 (not across the genome) and coalescent inference14,15,26. Our genome-wide estimate of G is 9.5 conversions/Mb per generation in the chip data set (unadjusted for SNP ascertainment) and 5.9 conversions/Mb per generation in the sequencing data set (Table 1). While the sequencing data set is minimally affected by ascertainment bias, the markers on Illumina BeadChip arrays are preferentially selected sequencing data set is minimally affected by ascertainment bias, per generation in the sequencing data set (Table 1).

Genomic distribution and sex differences

An assessment of the genomic distribution of G in the two sexes (Supplementary Note) showed that G was elevated near telomeres and therefore decreased with chromosome length (Fig. 2b). This pattern is analogous to that observed for crossover recombinations14,28 and is consistent with higher stationary GC content near telomeres29. Moreover, as in the case of crossover recombinations, the proportion of events near telomeres was greater in fathers than in mothers. Overall, G was 2.17 (95% CI = 1.94–2.45) and 1.91 (95% CI = 1.34–2.58) times higher in maternal transmissions than in paternal transmissions when assessed in the chip and sequencing data sets, respectively. This sex difference is very similar to that observed for crossover recombination rate17, which is 2.03, although the difference in G can largely be attributed to longer events in maternal transmissions.

G increases with maternal age

As previous studies have reported an age-related increase in the crossover recombination rate in females that is not seen in males19,30, we examined the impact of age on G. In the chip data set, there was a

Table 2 Estimates of CCO gene conversions of SNPs

|                | Chip data set | Sequencing data set |
|----------------|--------------|---------------------|
| Converted MPPs | 8            | 5                   |
| n MPPs considered | 289,364     | 170,612             |
| Observed gene conversion rate per Mb (95% CI) | 28 (10–52) | 29 (6–58) | 598 (289-1,011) | 374 (182–624) |
| n events | 7           | 5                   | 34               |
| Gene-converted MPPs/event (SE) | 1.14       | 1.14 (0.08) | 1.39 (0.07) | 1.39 (0.08) |
| GC bias (%) | 12.5        | 73.3 (65.2–80.7) | 70.2 (62.5–77.8) | 70.2 (63.1–78.6) |
| n crossovers considered | 45,025     | 75                  | 1,624             | 4,114           |
| Complex crossovers, per thousand | 0.16 (0.07–0.28) | 1.0 (0.8–1.2) | 3.1 (0.6–6.0) | 13.3 (8.5–18.2) |

Confidence intervals for CCO gene conversions in paternal transmissions are not reported because of the small number of events. CCO, complex crossover.
marked increase in $G$ with maternal age of 0.58 conversions/Mb per year (95% CI = 0.38–0.78, $P = 1.4 \times 10^{-8}$) that was not observed in fathers (Fig. 2a and Supplementary Note). Based on a very small number of events, the maternal age effect in the sequencing data set was 0.33 conversions/Mb per year (95% CI = −0.18 to 0.83). Although this estimate was not significantly different from 0 ($P = 0.21$), this effect was also not significantly different from the chip data estimate ($P = 0.35$).

Interestingly, the increase in $G$ with maternal age in the chip data set was comparable inside and outside both crossover recombination hotspots and male DSB regions (Fig. 2c,d), resulting in a decrease in odds ratios for co-occurrence with these genome features (Supplementary Fig. 2 and Supplementary Note). The average female crossover recombination rate of maternally transmitted NCO gene conversions further decreased with age (Supplementary Fig. 3). Thus, the NCO gene conversions that accumulate with age seem to be less tied to programmed DSBs than those transmitted by younger mothers.

$G$ increases with local GC content

$G$ increased with local GC content, defined as the GC content of the 100 bp surrounding each MPP (Supplementary Note), a result consistent with GC-biased NCO gene conversions occurring repeatedly at similar locations in evolutionary history. Local GC content was elevated in DSB regions (Supplementary Fig. 4 and Supplementary Table 4), in part because of the GC composition of the PRDM9 motif[31]. We considered the correlation with local GC content inside and outside of DSB regions separately as well as distinguishing between the chip and sequencing data sets and paternal and maternal

---

**Figure 2.** The NCO gene conversion rate, $G$. (a–d) $G$ is shown as a function of age of the parent (a), distance from the telomere (b), age of the mother stratified by presence in a DSB region (c) and age of the mother stratified by presence in a crossover recombination hotspot (d). (e) Percent GC content of the neighboring 100 bases for gene-converted MPPs outside of DSB regions as a function of age of the mother. Blue represents paternal transmission, and red represents maternal transmission. Error bars, 95% confidence intervals. In a, the number of parents in each age bin is presented at the top of the plot. In a and c–e, individuals are grouped into 2-year age bins, with all individuals younger than 20 years grouped together and all individuals 40 years or older grouped together. In b, MPPs are grouped into 5-Mb bins. Crossover recombination rate is computed as the average over all MPPs in a bin of average crossover recombination rate in a 5-kb interval around a marker. The dotted lines represent a linear model. All results are for the chip data set.
transmissions. Even after we had conditioned the data on DSB region status, in seven of eight cases, $G$ was positively correlated with local GC content (Fig. 3 and Supplementary Table 5).

In the chip data set, we observed a decrease ($P = 0.0068$) in local GC content with age in maternally transmitted NCO gene conversions. As this result might be attributed to the high GC content of the PRDM9 binding motif, we restricted the analysis to MPPs outside of male DSB regions, where we also observed a decrease ($P = 0.0037$) (Fig. 2e). The result also remained significant after adjustment of local GC content for the HapMap-based recombination rate$^{32}$ ($P = 0.0029$). Again, this suggests that the NCO gene conversions that accrue in aging mothers are different in mechanism from those found in younger mothers.

We investigated whether $G$ was dependent on the base-pair composition of the SNP’s two alleles and found no such effect (Supplementary Table 6).

**Complex NCO events**

We grouped gene-converted MPPs into events, on the basis of their proximity. In NCO gene conversions, most events contained only a small number of gene-converted MPPs, with an average of 1.24 MPPs per event for the chip data set and 1.78 genes per event for the sequencing data set. The smaller number of MPPs per event for the chip data set in comparison to the sequencing data set can in part be attributed to the lower marker density. Interestingly, in NCO gene conversions, maternally transmitted events were tagged by more MPPs than paternally transmitted events: 1.37 versus 1.04 ($P < 0.001$) and 2.34 versus 1.23 ($P < 0.001$) MPPs per event for chip and sequencing data, respectively. However, we did not observe age dependence in the number of MPPs per event (Supplementary Fig. 5).

We partitioned the set of NCO gene conversions into short and long NCO events, on the basis of a distance of 1,000 bp (roughly the size of a DSB region) between the first and last gene-converted MPPs for each event. Supplementary Table 7 shows the length distribution of long NCO events by distance between the first and last markers. Because of the denser marker set, the length of the event could be better estimated in the sequencing data set; hence, some events classified as short NCO events in the chip data would be classified as long NCO events if the denser sequencing data were available (Supplementary Note). Table 1 shows that short NCO events were highly over-represented in male DSB regions and crossover recombination hotspots, while long NCO events were not over-represented in male DSB regions but were over-represented in crossover recombination hotspots.

The tracts of long NCO events contained both gene-converted MPPs and non-gene-converted MPPs (Supplementary Table 8 and Supplementary Note). A similar pattern has previously been observed$^{29}$, and these are referred to as complex NCO gene conversions. Within complex events, both the gene-converted and non-gene-converted MPPs showed a GC bias.

In the chip and sequencing data sets, we estimated that, respectively, at least 46.1% and 65.3% of all long NCO events were complex (Supplementary Table 9). The true rate of complex events is likely to be higher (Supplementary Note), leading us to hypothesize that all long NCO events might be complex. These long, and mostly complex, NCO events were more common in maternal transmissions ($P < 0.001$, for chip and sequencing data sets). A significant increase in $G$ with mother’s age was observed for both short and long events (Supplementary Note).

**Gene conversion of indels**

We estimated $G$ for indels as 5.8 conversion/Mb per generation (95% CI = 4.1–7.9), comparable to that for SNPs. Our results showed a bias of 65.6% (95% CI = 53.3–76.6%, $P = 0.018$; Table 3) toward the shorter allele for indels in allelic gene conversions (Supplementary Note). A direct estimate of gene conversions involving indels has, to our knowledge, not previously been reported. Comparisons between species have yielded conflicting results: a deletion bias has previously been reported for non-allelic gene conversion$^{33}$, whereas a bias toward insertion has previously been reported for allelic gene conversion$^{34}$.

| Table 3 Gene conversion of indels |
|----------------------------------|
| Sequencing data set              |
| Paternal                         |
| Maternal                         |
| Total                            |
| Converted MPPs                   | 23 | 28 | 50 |
| NCO                              | 38 | 28 | 50 |
| CCO                              | 11 | 11 | 22 |
| Gene conversion rate per Mb (95% CI) | 5.2 (3.0–8.0) | 8.8 (5.5–12.1) | 7.0 (5.0–9.0) |
| Deletion bias (%) (95% CI)        | 55.5 (33.3–77.5) | 71.1 (53.2–84.5) | 65.6 (53.2–77.4) |
| OR NCO in DSB (95% CI)            | 35.6 (14.8–86.0) | 4.9 (1.0–11.8) | 13.9 (7.0–24.1) |
| OR NCO in crossover recombination hotspots (95% CI) | 10.4 (2.0–29.0) | 8.4 (0.0–25.4) | 13.9 (7.0–24.1) |
Complex crossover gene conversions

In crossovers, we were only able to detect complex events (Supplementary Fig. 6 and Supplementary Note). The rate reported for CCO gene conversions should be interpreted with caution, as it refers to the fraction of MPPs within a distance of 100 kb from a crossover recombination that show evidence of gene conversion (Supplementary Note) and is not a genome-wide rate. We observed a greater CCO gene conversion rate in the sequencing data set, where more events were detectable (Table 2). Because of our inability to detect all complex events, the true CCO gene conversion rate is likely to be higher than the estimates in both data sets.

The CCO gene conversion rate was greater than the NCO gene conversion rate ($P<0.001$, chip data set; $P<0.001$, sequencing data set). This confirms that, as a group, CCO gene conversions are not independent of crossovers. How CCO gene conversions are related to crossover recombinations remains to be elucidated.

A large majority of the CCO gene conversions we identified were maternal, demonstrating that complex crossovers are more common in maternal transmissions, as is the case for NCO gene conversions. Another similarity is that we observed an increase in the CCO gene conversion rate with maternal age of 14.0 conversions/Mb per year (95% CI = 0.7–27.3, $P = 0.04$), in the chip data set (Fig. 4a). Moreover, the fraction of crossovers that were complex increased with maternal age ($P = 0.02$ in the chip data set) (Fig. 4b and Supplementary Note). In the sequencing data set, we observed that 0.31% of all paternally transmitted crossover recombinations were complex. This result is in close agreement with a previous estimate of 0.33% obtained using sperm analysis$^{22}$.

GC bias

Like previous studies$^{7,20}$, our results showed a significant GC bias for NCO gene conversions$^4$, where strong base pairs (G or C) preferentially appeared on polymorphic gene-converted base pairs (Table 1): the GC bias was 67.6% (95% CI = 65.7–69.8%) in the chip data set and 69.3% (95% CI = 65.8–72.3%) in the sequencing data set. Short and long NCO events exhibited the same GC bias (Table 1).

Our results (Supplementary Note) indicated that the bias was greater in maternal than paternal transmissions ($P = 0.032$, chip data set; $P = 0.004$, sequencing data set). Further, CpG SNPs showed a greater GC bias than other SNPs ($P = 0.038$, chip data set; $P < 0.001$, sequencing data set).

The GC bias in CCO gene conversions was 70.2% (95% CI = 62.5–77.8%) in the chip data set and 70.1% (95% CI = 63.1–78.8%) in the sequencing data set, which was not significantly different from that observed in NCO gene conversions ($P = 0.56$ and 0.73 for the chip and sequencing data, respectively).

DISCUSSION

In summary, we have used both SNP chip and whole-genome sequencing data sets from three-generation families to search for meiotic gene conversions in humans. Overall, we identified 3,237 MPPs involved in gene conversions. On the basis of the sequencing data, we obtained a sex-averaged estimate of $G$, as 5.9 conversions/Mb per generation. Our results demonstrate that $G$ varies with both age and sex. Thus, the rate for mothers (7.7 conversions/Mb per generation) is 1.91 (95% CI = 1.34–2.58) times greater than that for fathers (4.1 conversions/Mb per generation), in the sequencing data set. Given that the fraction of heterozygous loci (where gene conversions can be detected) in Icelanders$^{35}$ is $6 \times 10^{-4}$, it follows that there is an expectation of 7 detectable NCO gene conversions from fathers and 14 from mothers. These numbers are 12 and 23, respectively, on the basis of worldwide average heterozygosity$^{35}$ of $1 \times 10^{-3}$.

A surprising result was the magnitude of the age-related increase of $G$ in females, where we estimate a 2.42-fold (95% CI = 1.83–3.10) increase from the age of 20 to 40 years. In the case of crossover recombination, the estimated increase between these age points is only 1.042-fold (4.2%), which is thought to be the result of greater viability of eggs with more crossover events$^{19,30}$. Such selection might account for some of the age-related increase of $G$ in females. However, given the more drastic increase in $G$, some other age-related factor must be at work. The increased $G$ in older mothers is less biased toward male DSBs, sex-specific crossover recombination hotspots or regions of high local GC content (Supplementary Note). Although our results are not conclusive, they indicate that a large fraction of the increase in crossover recombinations with maternal age is due to CCO recombinations (Supplementary Note).

Overall, the large fraction of NCO gene conversions in spermatocyte DSB regions and crossover recombination hotspots is consistent with the view that most of them occurred in response to programmed DSBs before the meiotic prophase I arrest of oocytes in the fetal ovary$^{37}$. However, an accumulation of programmed DSBs over subsequent decades does not seem a likely source of the age-related increase of NCO gene conversions in females. Other possible sources may be linked to the age-related deterioration of oocytes across the decades that they are in dictyate arrest, possibly leading to non-disjunction$^{38,39}$. This deterioration may be due to damage-induced DSBs, deficiencies in checkpoint mechanisms$^{37,40}$, failure of cohesins to maintain the cohesion of sister chromatids$^{37,41}$ or the fact that cohesive linkages are not restored at the same rate that they are lost$^{39}$. Further research is needed to determine the sources of these additional NCO gene conversions in older females and whether they are the same as the sources of additional CCO gene conversions. It is interesting to note in this context that risk of aneuploidies increases

Figure 4 Maternal age increase in complex crossovers. (a) CCO gene conversion rate by maternal age in the chip data set. (b) Fraction of maternally transmitted crossovers in the chip data set that are complex, as a function of mother’s age. Error bars, 95% confidence intervals. The dotted lines represent linear models.
drastically with the age of mothers—although there is no direct evidence to link aneuploidies with the age-related increase of gene conversions.

Our results further demonstrate that the control of gene conversions differs between the sexes. Whereas most paternaly transmitted NCO events are short and CCOs are rare in paternal transmissions, maternally transmitted NCO events tend to be long and complex. Our results suggest that there are different biological mechanisms underlying short and long NCO events and, consequently, maternal and paternal transmissions. As the definition of an event is based solely on the proximity of gene-converted MPPs, we cannot discern whether the gene-converted MPPs within the same long event occurred simultaneously in a single process or in several collocated processes. Crossover interference has recently been shown to decrease with maternal age, possibly leading to double-crossover recombination events, which in our study design could be detected as long NCO events. The complex nature of long NCO events and their GC bias make it unlikely that crossover interference explains a large fraction of long NCO events (Supplementary Note). Paternal NCO gene conversions may be enriched for those derived from the SDSa pathway, and maternal NCO gene conversions may be enriched for those resulting from double Holliday junction resolution. The long NCO events, which are complex and mostly maternally transmitted, may also arise from a more complex set of underlying biological mechanisms, including repeated template switching. Analyses of sperm and oocyte promise approaches for obtaining a more complete picture of meiotic gene conversion and its mechanisms.

Our results and those of others show that gene conversions are biased toward G+C base pairs, while mutations are biased toward A+T base pairs and increase in number with age in both sexes, but more strongly with father’s age. Now, it is clear that gene conversions increase in number with mother’s age. On average, the number of gene conversions per generation is comparable to that of mutations. This means that the nucleotide composition of the human genome represents an equilibrium that is maintained by an unwitting battle between the sexes, where male-driven AT-biased mutations are offset by female-driven GC-biased gene conversion events.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported in part by the NIH (NIDA) (R01-DA017932).

AUTHOR CONTRIBUTIONS

B.V.H., D.F.G. and K.S. designed the experiments. B.V.H. wrote the first draft of the manuscript. B.V.H., M.T.H., B.K., U.S., P.S., A.H., A.K., D.F.G. and K.S. reviewed and contributed to subsequent drafts of the manuscript. B.V.H., M.T.H. and A.G. implemented the methodology. B.V.H., M.T.H. and B.K. prepared tables and figures. Aaslauq Ionsadottir and Adalburt Ionsadottir performed the Sanger sequencing. U.T. oversaw the operations of the genotyping facility. B.V.H., M.T.H., E.Z., G.T., A.G. and G.M. processed the data. B.V.H. and M.T.H. analyzed the data. All authors contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.
38. Subramanian, V.V. & Bickel, S.E. Aging predisposes oocytes to meiotic nondisjunction when the cohesin subunit SMC1 is reduced. PLoS Genet. 4, e1000263 (2008).
39. Weng, K.A., Jeffreys, C.A. & Bickel, S.E. Rejuvenation of meiotic cohesion in oocytes during prophase I is required for chiasma maintenance and accurate chromosome segregation. PLoS Genet. 10, e1004607 (2014).
40. Leland, S. et al. Heterozygosity for a Bub1 mutation causes female-specific germ cell aneuploidy in mice. Proc. Natl. Acad. Sci. USA 106, 12776–12781 (2009).
41. Hodges, C.A., Revenkova, E., Jessberger, R., Hassold, T.J. & Hunt, P.A. SMC1β-deficient female mice provide evidence that cohesins are a missing link in age-related nondisjunction. Nat. Genet. 37, 1351–1355 (2005).
42. Nakajo, S.I., Hassold, T.J. & Hunt, P.A. Human aneuploidy: mechanisms and new insights into an age-old problem. Nat. Rev. Genet. 13, 493–504 (2012).
43. Campbell, C.L., Furlotte, N.A., Eriksson, N., Hinds, D. & Auton, A. Escape from crossover interference increases with maternal age. Nat. Commun. 6, 6260 (2015).
44. Martini, E. et al. Genome-wide analysis of heteroduplex DNA in mismatch repair-deficient yeast cells reveals novel properties of meiotic recombination pathways. PLoS Genet. 7, e1002305 (2011).
45. Tsaponina, O. & Haber, J.E. Frequent interchromosomal template switches during gene conversion in S. cerevisiae. Mol. Cell 55, 615–625 (2014).
46. de Boer, E., Jasin, M. & Keeney, S. Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. Genes Dev. 29, 1721–1733 (2015).
47. Wong, W.S.W. et al. New observations on maternal age effect on germline de novo mutations. Nat. Commun. 7, 10486 (2016).
48. Goldmann, J.M. et al. Parent-of-origin-specific signatures of de novo mutations. Nat. Genet. 48, 935–939 (2016).
49. Kong, A. et al. Rate of de novo mutations and the importance of father’s age to disease risk. Nature 488, 471–475 (2012).
Determining anchor MPPs. An anchor is an MPP where one parent is heterozygous and the other parent and siblings meet the accuracy thresholds defined in the Supplementary Note. With respect to the heterozygous parent, the phase of the sibling group (the proband and siblings) is unambiguous at anchors, and the sibling group can be partitioned into two sets, determined by which allele was inherited from the heterozygous parent. Two adjacent anchors, with the same heterozygous parent, induce the same partition in the sibling group unless either anchor was genotyped incorrectly or a gene conversion or crossover recombination occurred between them.

Given an MPP and a parent, we defined the left anchor as the closest anchor with a lower numerical coordinate where the parent was heterozygous. The right anchor was defined analogously with a higher numerical coordinate.

Unless a genotyping error occurred, we can be confident in the inheritance pattern for all anchor MPPs. We removed MPPs that seemed to be the result of a genotyping error from the set of anchor markers. When removing these, we might also remove MPPs that are the result of a gene conversion. In a later step, we determined whether each MPP was the result of a gene conversion or a genotyping error.

MPPs whose partition did not agree with neighboring anchors were removed. To formally delineate which markers were removed from the set of anchors, we defined the discrepancy between two anchors as the minimum number of individuals that needed to be moved between the sibling group partitions of the anchors for the partitions to become identical. We computed a local discrepancy score for an anchor A as the sum of the discrepancy between A and its two closest anchors to the left and two closest anchors to the right. Anchor A was removed if, when doing so, the sum of the discrepancy scores for all other anchors was reduced.

Phasing the parents. At a given MPP, a sibling group can be split into four inheritance groups on the basis of which of the two haplotypes each sibling inherits from each parent. The inheritance groups are not known, but when there are no crossover recombinations, gene conversions or genotyping errors, the haplotypes will agree with the haplotypes of both parents inherited at the neighboring anchors. We defined two inheritance groupings, left and right. The left inheritance grouping was determined by the left anchors of both parents, and the right inheritance grouping was based on the right anchors of both parents. Both the left and right inheritance groupings should be identical to the inheritance grouping at the given MPP unless there has been a crossover recombination for either parent in the region or the MPP being examined is gene converted in the proband or one of the siblings. A genotyping error in one of the siblings or either parent may also occur, causing the genotypes not to agree with the left and right inheritance groupings even if they are identical to the true inheritance grouping at the MPP.

Given the genotypes of the individuals in the nuclear family and the inheritance groupings, we assigned alleles to the parents’ haplotypes. For binary MPPs, there are a total of $2^4 = 16$ possible assignments for the two alleles to the four haplotypes. For each such assignment, we inferred the genotypes of both parents and the siblings according to the left inheritance grouping and compared them to observed genotypes. We defined left phasing discrepancies in the nuclear family as the combined number of mismatches between observed and inferred genotypes. Right phasing discrepancies were defined analogously from the right inheritance groups. For each assignment of alleles to haplotypes, we defined the number of phasing discrepancies as the smaller of the left and right phasing discrepancies. A phasing discrepancy can be explained by a crossover recombination, gene conversion or genotyping error.

If there was exactly one assignment of the alleles to the parents’ haplotypes with fewer than two phasing discrepancies, the MPP was considered to be phased by the assignment. All other MPPs were removed from further consideration as candidates when searching for gene conversions. When an assignment has no phasing discrepancy, either there is no gene conversion or the MPP is part of a gene conversion tract that includes neighboring anchors. Assignments with a single phasing discrepancy were further candidates for where a gene conversion might have taken place. Not all single phasing discrepancies will represent a gene conversion, as they may also represent a genotyping error or a non-gene-converted MPP in a long gene conversion tract including both neighboring anchors.

When there is more than one assignment that has fewer than two phasing discrepancies, we cannot reliably determine which assignment is the correct one, as all of them can arise from a single genotyping error or gene-converted MPP. An example of when an MPP has multiple assignments with fewer than two phasing discrepancies is when one parent’s haplotypes are not carried by any members of the sibling group. If all individuals are correctly genotyped and there is no gene conversion, then the correct assignment of haplotypes leads to zero phasing discrepancies, while switching the assignment of the haplotype not carried by any members of the sibling group leads to one phasing discrepancy.

When all assignments of alleles to the parents’ haplotypes yield at least two phasing discrepancies, then the MPP genotypes are not consistent with the left and right inheritance groupings, even when allowing for a single individual to either be the carrier of a gene conversion or have a genotyping error. This can occur because of multiple recombinations, a structural variant at the locus, a misplacement of the marker in the assembly or repeated genotyping errors at the marker.

Determining crossover recombinations. For each proband, we located crossover recombinations from each parent separately. We refer to an MPP as informative if the phase of the proband could be determined directly at the MPP without any assumption about the phase inherited from the other parent. Thus, the set of informative MPPs included anchor MPPs as well as MPPs where the parent of interest was heterozygous and either the proband or the other parent was homozygous. In particular, the set included...
MPPs where some of the siblings of the proband did not meet genotyping accuracy thresholds.

To distinguish candidate gene conversions from crossover recombinations, we looked for all inheritance tract changes in the proband. Initially, we assigned inheritance tract changes to all regions between two adjacent informative MPPs where the proband inherited alleles from different haplotypes of the parent of interest. The region of an inheritance tract change could be further narrowed if the parent of interest was heterozygous for additional MPPs between the two informative markers, if the haplotype of the other parent was known and if the two informative MPPs agreed on a haplotype for the other parent. In such a case, we assumed that the proband inherited the same haplotype for the whole region from the other parent and we could determine the proband’s phase at all MPPs in the region where the parent of interest was heterozygous. If the two informative MPPs did not agree on a haplotype for the other parent, the inheritance tract change region was excluded from the search for gene conversions. Once the inheritance tract change region had been narrowed, we assigned an inheritance tract change to the center of the region.

A crossover recombination was assigned to all tract changes where no other tract change occurred within 100 kb. All other tract changes were candidate gene conversions. Additionally, a crossover recombination was assigned if there was an odd number of multiple consecutive tract changes within 100 kb of each other; more precisely, the crossover recombination was assigned to the leftmost or rightmost tract change depending on which induced fewer gene-converted MPPs. In this case, a crossover recombination has occurred along with possible gene conversions.

Determining gene conversion MPPs. Having assigned crossover recombinations and phased the haplotypes of the parents, we searched in the proband for mismatches in observed genotypes in comparison to the genotypes defined by the haplotypes inherited from the parents.

We searched for gene conversion from each parent separately for all MPPs in the phased MPP set after applying the quality filters described in the Supplementary Note.

MPPs passing these filters were counted toward the denominator in our rate computations. If there was a mismatch between the proband’s genotype and the phase-determined haplotypes, we examined whether this mismatch could be due to a gene conversion.

If the mismatch could be explained with the other haplotype of the parent of interest, we attempted to verify the gene conversion in the proband’s children. We verified the genotype in one of two ways. If a child was homozygous, we verified the gene-converted haplotype without requiring the proband’s spouse to be genotyped. Otherwise, if the proband’s spouse was genotyped and homozygous, we used this genotype together with the child’s genotype to verify the gene-converted haplotype. In some cases, we might be unable to verify the haplotype because of inconsistencies such as structural variants, misplacement in the assembly or repeated genotyping errors.

To verify a putative gene conversion in the proband’s child, we first determined whether the child carried the gene-converted haplotype. We searched for the closest MPPs to the left (of lower numerical order) and to the right (of higher numerical order) where the proband was heterozygous and the spouse was homozygous, ignoring MPPs at the ends of chromosomes where there was no such left or right MPP. At these MPPs, we could determine which haplotype the child inherited from the proband. If the child inherited the same haplotype from the proband at both of these MPPs and the distance between these MPPs was less than 1 Mb, we assumed that the child was carrying the corresponding haplotype. Gene conversion could be verified in the child if this was the gene conversion haplotype.

If a marker showed evidence of being part of a structural variant, being misplaced in the assembly or having more than one genotyping error, the marker was flagged as problematic in the family. Specifically, a marker was flagged as problematic if all of the assignments of alleles to the parents’ phases produced two or more phasing discrepancies or if we failed to verify a gene conversion in one of the proband’s children. Markers that were flagged as problematic in more than one family were removed from the set of quality markers input to the algorithm.

Crossover regions and CCO versus NCO gene conversions. For each crossover recombination, we defined a crossover region as the adjacent 100 kb in each direction. If a gene-converted MPP was found within the crossover region, the region was iteratively extended so that it contained all MPPs within 100 kb of the crossover recombination and the gene-converted MPPs found.

MPPs determined to be within and outside a crossover region were used in the computation of CCO gene conversion rate and NCO gene conversion rate, respectively.

Determining gene conversion events. Gene conversion events were determined once all gene-converted MPPs were determined. While searching for gene-converted MPPs, we restricted our search to contiguous tracts of MPPs where the length of the tract was 100 kb. Gene conversion events may, however, contain both gene-converted MPPs and non-gene-converted MPPs.

Within NCOs, we arranged the MPP positions from a parent–proband pair into distinct gene conversion events by traversing the chromosome in numerical order. We considered the first gene-converted MPP found on a chromosome to be part of a new event and iteratively extended the event if a gene-converted MPP was found within 100 kb of the previous gene-converted MPP. Consequently, gene conversion events could be longer than 100 kb.

Within crossovers, MPPs neighboring the same crossover recombination were considered a part of the same event.

Crossover recombination map data. We used a data set of crossover recombinations25 preprocessed as described in the Supplementary Note.

Computation of rates, confidence intervals and P values. We computed the observed gene conversion rate as the number of MPPs where a gene conversion occurred divided by the number of MPPs that were tested for a gene conversion.

For two events, A and B, we computed the odds ratio as \( \frac{n_{11} \times n_{22}}{n_{12} \times n_{21}} \), where \( n_{11} \) represents the number of MPPs that were part of A and B, \( n_{12} \) represents the number of MPPs that were part of A and not B, \( n_{21} \) represents the number of MPPs that were part of B and not A, and \( n_{22} \) represents the number of MPPs that were not part of either event. Odds ratios for crossover recombination hotspots were computed considering only markers where the crossover recombination rate was estimated.

All confidence intervals presented are 95% confidence intervals, and all \( P \) values are two-sided. Confidence intervals for \( G \), odds ratios, the number of MPPs per event, the rate of increase in \( G \) between the age of 20 and 40 years, and GC bias were computed using a bootstrap method50. For the chip data set, 1,000 sets of 7,219 individuals were sampled with replacement from the set of 101 probands. The statistic in question was computed within each set, creating a list of 1,000 statistics. Following sorting of this list, the lower bound of the confidence interval was computed as the mean of entries 25 and 26 and the upper bound was computed as the mean of entries 975 and 976.

Age-of-parent effects were determined using a weighted linear regression with the function \( \text{lm} \) in R (ref. 51). To determine age effect on \( G \), we first computed \( G \) for each proband–parent pair separately. The final model can be expressed as \( \text{lm}(G - \text{ParentAge}, \text{weights} = \text{sqrt}(n)) \), where \( n \) is the number of MPPs considered for the proband. To determine age effect on other statistics, we computed for each proband–parent pair separately the statistic \( S \) in question. The final model can be expressed as \( \text{lm}(S - \text{ParentAge}, \text{weights} = \text{sqrt}(n)) \).

Linear regressions together with their confidence intervals and \( P \) values, for distance to the telomere and chromosome length, were computed using the \( \text{lm} \) function in \( R \), with a matrix containing all MPPs where a gene conversion event could be ascertained. All other linear regressions were implemented using Python52.

All other \( P \) values not previously discussed were computed using bootstrapping. One thousand simulations were used analogously to the description above, and a \( P \) value was computed by counting the number of times the single-sided event of interest occurred and dividing by the number of simulations. The single-sided \( P \) value was then multiplied by 2 to obtain a double-sided \( P \) value. In cases where the event of interest did not occur in 1,000 simulations, the \( P \) value is reported as <0.001.
To compute $G$ corrected for crossover recombination, linear regression was performed with gene conversion as a response and the local sex-specific crossover recombination rate as an explanatory variable. All MPPs where an NCO gene conversion could be ascertained and the crossover recombination rate had been determined were used. Corrected $G$ rates were computed by inserting the genomic average crossover recombination rate, consisting of 1.572 cM/Mb for maternal transmissions and 0.772 cM/Mb for paternal transmissions, into the regression formula. Confidence intervals were computed using the `predict.lm` function in R.

50. Efron, B. & Tibshirani, R.J. *An Introduction to the Bootstrap* (CRC Press, 1994).
51. Ihaka, R. & Gentleman, R.R. A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314 (1996).
52. van Rossum, G. & Drake, F.L. *PYTHON Reference Manual* (Centrum voor Wiskunde en Informatica, 1995).
Author Correction: The rate of meiotic gene conversion varies by sex and age

Bjarni V Halldorsson, Marteinn T Hardarson, Birte Kehr, Unnur Styktasdottir, Arnaldur Gylfason, Gudmar Thorleifsson, Florian Zink, Adalbjorg Jonasdottir, Aslaug Jonasdottir, Patrick Sulem, Gisli Masson, Unnur Thorsteinsdottir, Agnar Helgason, Augustine Kong, Daniel F Gudbjartsson and Kari Stefansson

Correction to: Nature Genetics https://doi.org/10.1038/ng.3669, published online 19 September 2016.

In the version of this article published, statements about the impact of insertions and deletions on gene conversions were incorrect. We reported a bias toward deletions, whereas in fact the bias was toward insertions. We are deeply indebted to Laurent Duret and Brice Letcher for noticing this mistake in our manuscript. The following statements are incorrect in the published manuscript.

Abstract:
Published text: “For indels, we demonstrate a 65.6% preference for the shorter allele.”
Correct text: “For indels, we demonstrate a 65.6% preference for the longer allele.”

Results:
Published text: “Our results showed a bias of 65.6% (95% CI = 53.3–76.6%, P = 0.018; Table 3) toward the shorter allele for indels in allelic gene conversions.”
Correct text: “Our results showed a bias of 65.6% (95% CI = 53.3–76.6%, P = 0.018; Table 3) toward the longer allele for indels in allelic gene conversions.”

Table 3:
Published text: “Deletion bias (%)
Correct text: “Insertion bias (%)

Supplementary information:
Published text: “We observe a bias towards the shorter alleles of 65.6% (95% CI 53.2–77.4, p: 0.018) in the set of gene converted indels (cf. Table 3). Of the 61 gene converted indels, 40 were deletions from the reference genome, while 21 were insertions into the reference genome. Deletion bias was estimated as 65% (95% CI 50.0–78.9) for indels that are deletions from the reference genome and 66.7% (95% CI 42.1–85.7), for indels that are insertions into the reference genome. No difference was found in deletion bias between those markers that are deletions vs. those that are insertions (p-value: 0.95). Thirty of the 61 gene converted indels represented the deletion or insertion of one base-pair. No significant difference could be found in deletion bias by indel length.”
Correct text: “We observe a bias towards the longer alleles of 65.6% (95% CI 53.2–77.4, p: 0.018) in the set of gene converted indels (cf. Table 3). Of the 61 gene converted indels, 40 were deletions from the reference genome, while 21 were insertions into the reference genome. Insertion bias was estimated as 65% (95% CI 50.0–78.9) for indels that are deletions from the reference genome and 66.7% (95% CI 42.1–85.7), for indels that are insertions into the reference genome. No difference was found in insertion bias between those markers that are deletions vs. those that are insertions (p-value: 0.95). Thirty of the 61 gene converted indels represented the deletion or insertion of one base-pair. No significant difference could be found in insertion bias by indel length.”