Contributions of intrinsic mutation rate and selfish selection to levels of de novo *HRAS* mutations in the paternal germline

Eleni Giannoulatou[a,1], Gilean McVean[b], Indira B. Taylor[a], Simon J. McGowan[a], Geoffrey J. Maher[a], Zamin Iqbal[b], Susanne P. Pfeifer[b,2], Isaac Turner[b], Emma M. M. Burkitt Wright[c], Jennifer Shorto[c], Aysha Itani[d], Karen Turner[d], Lorna Gregory[b], David Buck[b], Ewa Rajpert-De Meyts[e], Leendert H. J. Looijenga[f], Bronwyn Kerr[c], Andrew O. M. Wilkie[a,3], and Anne Goriely[a,1,3]

[a]Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, United Kingdom; [b]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom; [c]Manchester Academic Health Science Centre, University of Manchester, Manchester M13 9WL, United Kingdom; [d]Institute of Reproductive Sciences, Oxford OX4 2HW, United Kingdom; [e]Department of Growth and Reproduction, Copenhagen University Hospital (Rigshospitalet), DK-2100 Copenhagen, Denmark; and [f]Department of Pathology, Erasmus University Medical Centre, 3000 CA Rotterdam, The Netherlands

Edited by Arthur L. Beaudet, Baylor College of Medicine, Houston, TX, and approved October 25, 2013 (received for review June 15, 2013)

The RAS proto-oncogene Harvey rat sarcoma viral oncogene homolog (*HRAS*) encodes a small GTPase that transduces signals from cell surface receptors to intracellular effectors to control cellular behavior. Although somatic *HRAS* mutations have been described in many cancers, germline mutations cause Costello syndrome (CS), a congenital disorder associated with predisposition to malignancy. Based on the epidemiology of CS and the occurrence of *HRAS* mutations in spermatocytic seminoma, we proposed that activating *HRAS* mutations become enriched in sperm through a process akin to tumorigenesis, termed selfish spermatogonial selection. To test this hypothesis, we quantified the levels, in blood and sperm samples, of *HRAS* mutations at the p.G12 codon and compared the results to changes at the p.A11 codon, at which activating mutations do not occur. The data strongly support the role of selection in determining *HRAS* mutation levels in sperm, and hence the occurrence of CS, but we also found differences from the mutation pattern in tumorigenesis. First, the relative prevalence of mutations in sperm correlates weakly with their in vitro activating properties and occurrence in cancers. Second, specific tandem base substitutions (predominantly GC>TT/AA) occur in sperm but not in cancers; genomewide analysis showed that this same mutation is also overrepresented in constitutional pathogenic and polymorphic variants, suggesting a heightened vulnerability to these mutations in the germline. We developed a statistical model to show how both intrinsic mutation rate and selfish selection contribute to the mutational burden borne by the paternal germline.

paternal age effect | male mutation bias | RASopathy | testis

Understanding the factors that influence the apparent rate of de novo mutations in the genome is central to the study of genetic diseases and genome diversity. In humans, germline mutation rates vary by several orders of magnitude, with average rates of 4–160 $\times$ 10^{-9} per nucleotide for different point mutations (1, 2). Mutations also show a parent-of-origin bias that is explained by differences in the biology of germ cells in males and females, with the majority of germline point mutations, small indels, and nonrecurrent copy number variations showing a strong paternal bias, believed to originate during the mitotic replications of spermatogonial stem cells (SSCs) that continue throughout the reproductive life of the male (3). Direct estimates of germline mutation rate, based on whole-genome sequencing (WGS) of two- and three-generation families, concur that among the 30–100 novel point mutations that are acquired in each generation, ~80% originate in the paternal germline (4-7). Two recent studies (6, 7) have further suggested that the major determinant of the total number of de novo germline point mutations is the age of the father at conception, increasing by one to two mutations per year. However, epidemiologically, this rate of increase would be predicted to result in a modest paternal age effect, with the average father of a child with a randomly sampled de novo mutation being ~2.2 y older than the population average (*SI Text*).

We and others have proposed that an additional mechanism promotes the enrichment of de novo pathogenic mutations in the testes of aging men (8–11). This process, which we term selfish spermatogonial selection, accounts for the unusual presentation of a group spontaneous dominant diseases that we collectively call paternal age effect (PAE) disorders, including Apert syndrome, achondroplasia, multiple endocrine neoplasia type 2 (men2b), Costello syndrome (CS), and Noonan syndrome (11). These disorders occur spontaneously with an apparent birth rate that is two to three orders of magnitude above the background rate of mutation (up to 1 in 30,000 for achondroplasia; the estimated birth prevalence of CS in the United Kingdom is ~1:380,000; *SI Text*), show an extreme paternal bias in origin (male-to-female ratio of mutation >10:1) and are associated with

Significance

Harvey rat sarcoma viral oncogene homolog (*HRAS*) occupies an important place in medical history, because it was the first gene in which acquired mutations that led to activation of a normal protein were associated with cancer, making it the prototype of the now canonical oncogene mechanism. Here, we explore what happens when similar *HRAS* mutations occur in male germ cells, an issue of practical importance because the mutations cause a serious congenital disorder, Costello syndrome, if transmitted to offspring. We provide evidence that the mutant germ cells are positively selected, leading to an increased burden of the mutations as men age. Although there are many parallels between this germline process and classical oncogenesis, there are interesting differences of detail, which are explored in this paper.

Author contributions: G.M., A.O.M.W., and A.G. designed research; E.G., I.B.T., G.J.M., and A.G. performed research; A.I., K.T., L.G., D.B., E.R.-D.M., L.H.J.L., and B.K. contributed new reagents/analytic tools; E.G., G.M., S.J.M., Z.I., S.P.P., I.T., E.M.M.B.W., J.S., B.K., A.O.M.W., and A.G. analyzed data; and E.G., G.M., A.O.M.W., and A.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]E.G. and A.G. contributed equally to this work.

[2]Present address: Max F. Perutz Laboratories GmbH, Vienna 1030, Austria.

[3]To whom correspondence may be addressed. E-mail: andrew.wilkie@imm.ox.ac.uk or anne.goriely@imm.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1311381110/-/DCSupplemental.

an increase in average paternal age (~2.5–8.7 y excess), relative to the general population. Quantification of specific causative mutations in the *FGFR2*, *FGFR3* (encoding fibroblast growth factor receptors 2 and 3, respectively), *PTPN11,* and *RET* genes in sperm (8, 10, 12) or whole testes (9, 13–16) led to the proposal that spermatogonial cells that have acquired rare spontaneous PAE mutations are positively selected, leading to their progressive clonal expansion over time (8, 11, 17). Immunohistochemical screening of testicular sections from elderly men visualizes likely clonal expansion events within the seminiferous tubules (18).

Supporting the parallels between selfish selection and early events in tumorigenesis, we reported that strongly activating somatic mutations in *FGFR3* and Harvey rat sarcoma viral oncogene homolog (*HRAS*) occur in spermatocytic seminoma (SpS), a rare testicular tumor affecting older men that is thought to represent the extreme outcome of selfish selection. The previous survey (10) identified two tumors with *FGFR3* c.1948A>G (p.K650E) mutations and five tumors with *HRAS* mutations [three samples with c.182A>G (p.Q61R) and two with c.181C>T (p.Q61K)]. The finding of acquired *HRAS* mutations was noteworthy because heterozygous germline mutations cause CS, which exhibits the epidemiological characteristics of a PAE disorder (11, 19, 20). However, whereas all mutations previously identified in SpS affect the p.Q61 codon, 88% of published CS mutations localize to the p.G12 codon (Fig. 1) and none has been described at p.Q61 (Table S1). These codons correspond to two of the three hotspots for mutation in cancer (p.G12, p.G13, and p.Q61), at which missense substitutions act by locking the RAS molecule in a GTP-bound conformation, resulting in a constitutively active state (21).

Although *HRAS* mutations are predicted to be enriched by selfish selection and have been implicated in SSC growth regulation (22), no study has attempted to document their occurrence directly in the sperm of healthy males. To explore further the link between selfish selection and human disease, we quantified the levels in blood and sperm of spontaneous mutations around p.G12 of *HRAS*, the codon most frequently affected both by germline CS mutations and by somatically acquired oncogenic mutations. Our results illustrate both similarities and differences between selfish selection and classical oncogenic processes.

Results

Oncogenic *HRAS* p.G12 Codon Mutations Are Elevated in Sperm. To quantify spontaneous *HRAS* mutations, we developed a protocol (Fig. S1) combining restriction enzyme digestion, PCR amplification, massively parallel sequencing, and statistical analysis (*SI Text*). Observing that every nonsynonymous single nucleotide substitution at the *HRAS* p.G12 codon has been described in cancer and each is associated with a different transforming activity (23, 24), we selected for mutations by digesting genomic DNA with the restriction enzyme MspI (cleaves the WT sequence 5′-CCGG-3′ at c.32_c.35, irrespective of methylation status). This strategy allows equal enrichment of all but 1 of 12 possible single-nucleotide substitutions at the MspI site (Fig. 1, bold) as well as complex mutations. Additional benefits are that the MspI site includes a CpG dinucleotide, enabling comparison of transition and transversion rates in the context of both CpG and non-CpG nucleotides, and encompasses two adjacent codons, so that mutation levels at the p.G12 CS/cancer hotspot can be compared with those at p.A11, at which mutations are anticipated to be selectively neutral. In samples heterozygous for the SNP rs12628, located 46 bp downstream of c.35G, each substitution within the MspI site can be phased, allowing us to establish on which of the two *HRAS* alleles the original contributing mutational events took place (Fig. 1; *SI Text*).

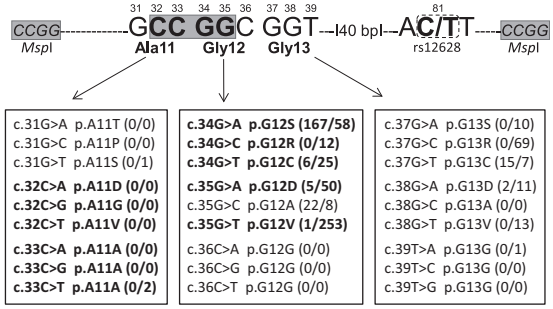

Fig. 1. Genomic context and DNA sequence around the *HRAS* p.G12 codon. The relative positions of the p.A11, p.G12, and p.G13 codons to the rs12628 SNP (dashed box) and the MspI restriction sites used for selection (gray boxes) are indicated. All single-nucleotide substitutions encompassing p.A11-p.G13 codons and corresponding amino acid changes are indicated, with those detected by resistance to MspI digestion in bold. The number of recorded instances of mutation as either a germline (Table S1) or somatic (COSMIC) change is indicated as (germline/somatic). Note that c.35G>C, encoding p.G12A, cannot be assayed by MspI digestion as this mutation creates a new MspI site at position c.35_38.

To assess the sensitivity and reproducibility of the assay, we estimated mutation levels in a titration-reconstruction experiment using biological replicates containing 10 μg of control blood DNA (equivalent to ~3.3×10^6 copies of the haploid genome) supplemented with dilution series of genomic DNA from four CS patients heterozygous for *HRAS* mutations [range of input mutant molecules from ~10 (concentration: 3×10^{-6}) to ~1,000 (concentration: 3×10^{-4})]. To quantify mutation levels, samples were spiked with ~100 mutant copies of genomic DNA from a unique CS patient heterozygous for c.35_36GC>AA tandem mutation. We found a good correlation between the amount of input DNA and the mutation levels estimated by massively parallel sequencing (Fig. 2*A*). The levels of the c.34G>A transition were overestimated ~3.6-fold at the lower dilution (3×10^{-6}), but the c.35_36GC>TT tandem mutation exhibited lower mutation levels in blood, and a good correlation between estimates and DNA input was observed down to the 3×10^{-6} level.

We then used the same strategy to quantify five single-nucleotide substitutions at the p.G12 codon and six substitutions at p.A11, in 7 blood and 89 sperm samples from healthy donors (Fig. 2*B*, Fig. S2*A*, and Dataset S1). Estimates of mutation levels from blood varied by mutation, with transitions exhibiting higher levels than transversions, especially within the c.33_34 CpG dinucleotide (Table S2). These levels are likely to reflect a combination of rare endogenous mutations in blood and artifacts during PCR, as this technique generates ~2- to 20-fold more transition than transversion errors (25). Based on these observations, the results of the titration experiment (Fig. 2*A*) and the analysis of skewing with respect to the rs12628 SNP (*SI Text* and Fig. S2*C*), a sample was considered to carry a given substitution if the measured levels were $>3 \times 10^{-6}$, except for transitions, for which the calling threshold was set at 10^{-5}. We next analyzed the levels of individual mutations (relevant statistics and correlation with donor age are summarized in Table S2). The levels for all substitutions involving c.32C and c.33C (in p.A11) did not differ significantly between blood and sperm (Fig. 2*B* and Fig. S2*A*). By contrast, the levels at positions c.34G and c.35G (encoding nonsynonymous changes at p.G12) were frequently higher in sperm than in blood (Fig. 2*B*, *Middle*) and also exhibited positive correlation with sperm donor age (Fig. 2*C* and Fig. S2*A*): in sperm, the c.34G>A (p.G12S) transition was the most widely occurring (55/89 samples had levels $>10^{-5}$) and the most abundant mutation, accounting for 62% of total single-nucleotide substitutions at codon p.G12. It also showed the strongest positive correlation with donor age ($r_s = 0.52$). The level of c.35G>A (p.G12D) (also a transition but not at a CpG) was on average 3.2-fold lower than c.34G>A and was present at $>10^{-5}$ in 17/89 sperm samples. The three other quantifiable single-nucleotide substitutions at p.G12 are transversions that exhibited lower

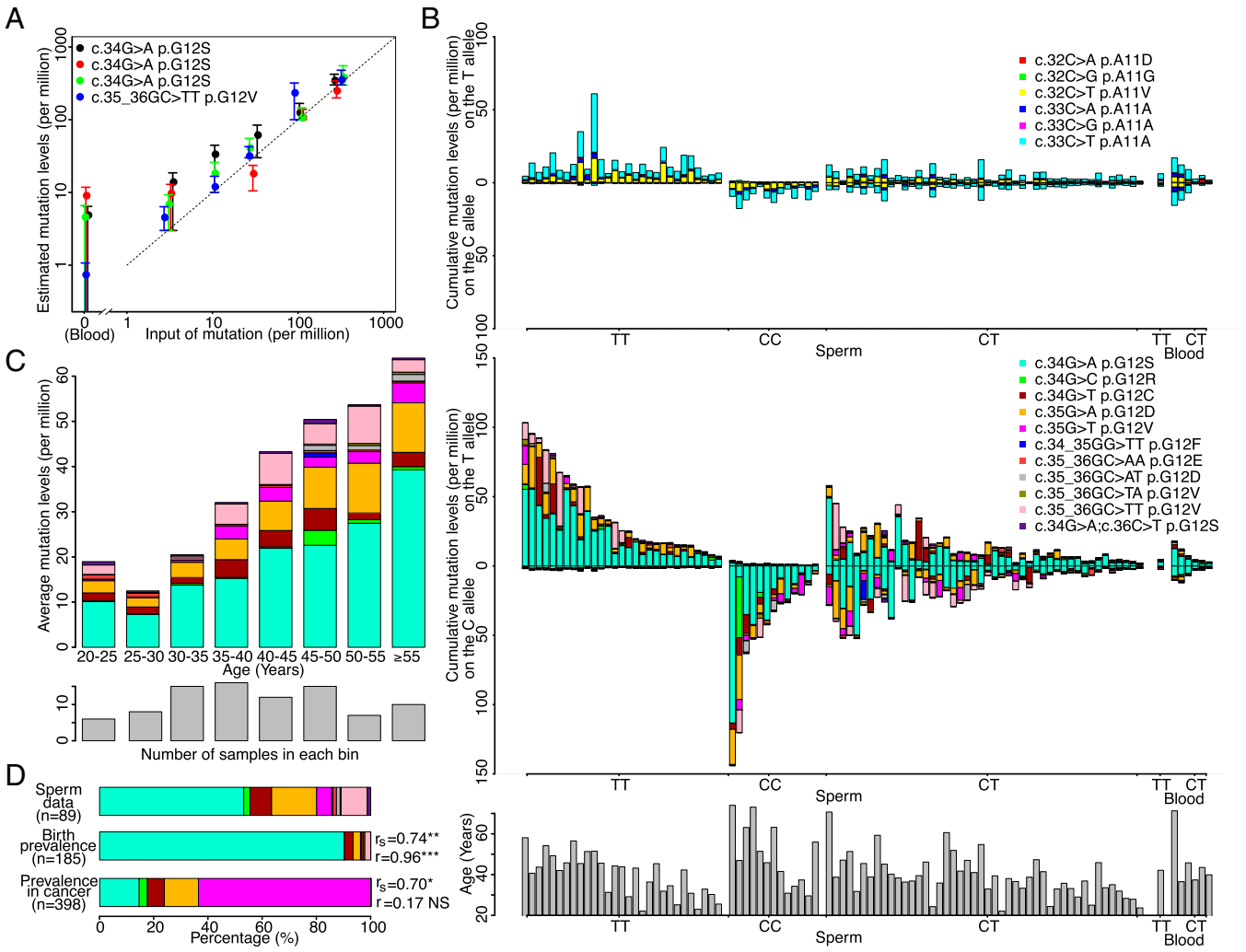

Fig. 2. Estimation of *HRAS* mutation levels within the c.32_35 MspI site (codons p.A11 and p.G12) in sperm and blood samples. (*A*) Mutation levels estimated in a titration-reconstruction experiment with serial dilution of four CS samples mixed with blood carrier DNA and spike DNA. (*B*) Estimation of mutation levels for substitutions at p.A11 (*Top*) and p.G12 (*Middle*) in 89 sperm (*Left*) and 7 blood (*Right*) samples. Color code for each substitution is given on the figure. Samples are organized according to their genotype at the rs12628 SNP (TT to the left, CC in the center, and CT heterozygote to the right). The mutation levels are plotted independently for the two *HRAS* alleles with respect to the SNP, so that the total mutation level for CT samples is the sum of the counts on each allele. The age of the sample donor is given at the *Bottom*. (*C*) Average levels for mutations at codon p.G12 in sperm samples binned by 5-y age group. (*D*) Comparison of levels for different mutations at codon p.G12 in sperm samples (*Top*) with relative prevalence of mutations reported in CS (Table S1) (*Middle*) and in cancer (COSMIC) (*Bottom*). Correlation between sperm data and other measurements are indicated by Spearman (r_s) and Pearson (r) correlation coefficients with statistical significance (NS, not significant; *P = 0.02; **P = 0.009; ***P = 0.000004). The color code used in *C* and *D* is identical to *B*.

background levels and were elevated above 3×10^{-6} in 20/89 samples for c.34G>T (p.G12C), in 18/89 samples for c.35G>T (p.G12V) and in only 3/89 samples for c.34G>C (p.G12R). Levels of these transversions were also correlated with donor age, although more weakly so than for the transitions.

Tandem Base Substitutions Are Overrepresented in the Germline. Given that our protocol would select any substitution resistant to MspI digestion, we asked whether multiple nucleotide substitutions could be identified. Unexpectedly, 31 independent events involving dinucleotide substitutions were observed in sperm samples at levels $>3 \times 10^{-6}$. Aside from c.34G>A;c.36C>T (encoding p.G12S) and c.34_35GG>TT (p.G12F), each observed in single sperm samples, all other dinucleotide mutations were tandem base substitutions (TBS) involving the last two nucleotides of codon p.G12, comprising c.35_36GC>AA (p.G12E) in 1 sample, c.35_36GC>AT (p.G12D) in 4 samples, c.35_36GC>TA (p.G12V) in 3 samples, and c.35_36GC>TT (p.G12V) in 21 samples (Dataset S1, Table S2, and Fig. S2*B*). Surprisingly, given that they both encode the same oncogenic p.G12V change and would therefore be subject to equivalent selection, the average level of the most prevalent TBS, c.35_36GC > TT, was 1.7-fold higher than the level of the c.35G>T single-nucleotide substitution. Levels of this TBS were significantly higher in sperm than blood ($P = 0.00002$) and correlated strongly ($r_s = 0.44$) with donor age (Fig. 2*C* and Table S2).

To assess the implications of the high prevalence of TBS, particularly GC>TT, we asked whether they could be identified in different human genomic datasets (*SI Text*). We first interrogated the Human Gene Mutation Database (HGMD), in which 441 TBS have been cataloged as pathogenic germline mutations. In agreement with a recently published study (26), the most numerous of all 78 possible TBS involve GC>TT (or its reverse complement GC>AA), representing 14.7% of the total, corresponding to a 10.6-fold enrichment over a uniform distribution of TBS ($P < 10^{-16}$; Fig. 3*A* and Table S3). Of the 64 coding GC>TT/AA, 26 encode

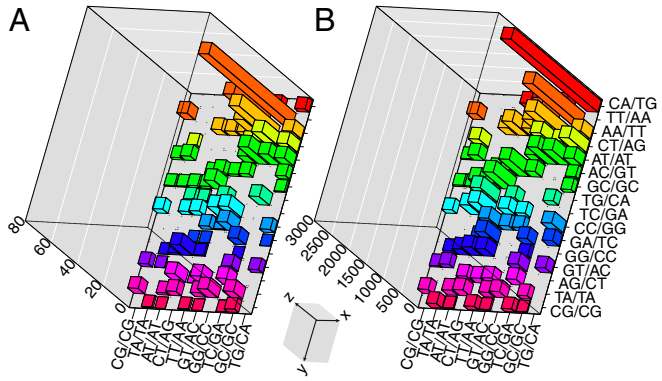

Fig. 3. Lego plots representing the prevalence of TBS in the human genome. (*A*) Data from HGMD. (*B*) Genomewide variation across 85 LWK whole genome sequences (Cortex assembler). The *x* axis represents the original dinucleotide sequences and their reverse complements, whereas the *y* axis indicates the mutated sequence and its reverse complement (the *y* axis legend is the same on both plots and to ease visualization, mutated sequences are shaded in different colors). Plotted on the *z* axis is the total number of events for each TBS (Table S3). Owing to the complementary nature of DNA, only 78 different TBS can occur, and gray areas indicate changes that do not lead to TBS (such as single-nucleotide substitutions) or are identical to their reverse complement.

changes that (due to the specifics of the genetic code) can only arise from a double-nucleotide substitution, including the recurrent *RET* c.2647_2648GC>TT (p.A883F) mutation associated with men2B (11, 27).

We next examined the distribution of TBS in cancer. Strikingly, although 406 single-nucleotide substitutions are recorded at the *HRAS* p.G12 codon in the Catalogue Of Somatic Mutations in Cancer (COSMIC), there is not a single instance of TBS, suggesting that a different pattern of mutations (either caused by distinct mutational mechanisms in somatic and germline cells and/or specific mutagen exposure) is observed in these different cellular contexts. This interpretation is supported by the profile of 3,769 TBS cataloged in COSMIC (Table S3). The most common TBS in somatic tissues are CC/GG>AA/TT (31.4%) and CC/GG>TT/AA (19.2%), which represent mutagen-specific signatures triggered by the action of polycyclic aromatic hydrocarbon components found in cigarette smoke or UV exposure (28), respectively. By contrast, there were only 107 events (2.8% of TBS) of GC>TT/AA, indicating that there is a much less marked enrichment (2.0-fold over random expectation) for this TBS in cancer (Table S3).

To explore further the impact of TBS, we analyzed the prevalence and distribution of TBS contributing to human variation, based on WGS data (*SI Text*). We used Cortex, a de novo assembly-based variant calling algorithm (29) to assess TBS representation in 85 human genome sequences from the Luhya in Webuye, Kenya (LWK) dataset of the 1000 Genomes Project (30) and identified 5,425,856 nucleotide variants, among which 22,898 (0.42%) involved TBS. Strikingly, the GC>TT/AA change was the second most common TBS, observed in 1,417 instances (6.2%), representing a 4.5-fold enrichment (Fig. 3*B* and Table S3). Because the pattern of TBS at the *HRAS* p.G12 codon suggested that the CpG dinucleotide at position c.36_37 (Fig. 1) might influence the apparent c.35_36GC>TT mutation rate, we further characterized the local sequence context in which the 1,417 genomic GC>TT/AA TBS had occurred. Compared with the relative frequency of single substitutions (G>T or C>A) in the same sequence context, the GC>TT/AA TBS is three times as likely to occur as part of a CpG dinucleotide [842 of 103,732 events (0.81%) for the single substitutions and 35 of 1,417 (2.5%) for TBS; $P = 2.2 \times 10^{-8}$; *SI Text*]. These genomewide observations suggest that the sequence context in which the TBS occurs plays an important, although yet uncharacterized, role, and in particular we propose that hypermutability of the C>T transition within the CpG dinucleotide accounts, at least in part, for the high spontaneous GC>TT/AA mutation rate observed in the germline.

Comparison Between Prevalence of *HRAS* Mutations in Sperm and in CS, SpS, and Cancer Datasets. To establish the biological relevance of the measurements of *HRAS* mutation levels in sperm, we compared these data to the distribution of published CS alleles, to experimental data generated in our laboratory on mutations in SpS, and to cancer-associated mutations cataloged in the COSMIC database.

Of the 236 CS cases reported in the literature, 207 (88%) involve mutations at codon p.G12 (Fig. 1 and Table S1). The c.34G>A (p.G12S) mutation, which is associated with a relatively homogenous presentation, is by far the most prevalent (81%). Other p.G12 mutations have also been described, including p.G12A, p.G12C, p.G12D, p.G12V, and p.G12E. These rarer alleles tend to be associated with more severe manifestations, often involving hypertrophic cardiomyopathy and resulting in neonatal mortality (31–34), consistent with biochemical evidence that p.G12S is less activating than any other mutation at this codon (23, 24).

We found a strong correlation between the prevalence of *HRAS* alleles in sperm and the number of cases reported for each CS mutation, indicating that the average level of mutation in sperm is a major determinant of prevalence of different *HRAS* alleles in the CS population (Fig. 2*D*). Comparing the sperm data with observed births of CS, it is apparent that p.G12S is unexpectedly prevalent in CS compared with other p.G12 substitutions, which suggests that these other (more activating) mutations may be associated with a higher risk of demise during the pregnancy (33). In agreement with our finding that TBS are not uncommon in sperm, a total of six CS patients carrying similar mutations have been reported (Table S1). Strikingly, among patients diagnosed with *HRAS* mutations encoding p.G12V, five cases have been associated with TBS (four with c.35_36GC>TT and one with c.35_36GC>TA), whereas only a single patient carried the c.35G>T substitution (31–33). The predominance of the c.35_36GC>TT TBS among observed CS alleles supports the relevance of our sperm data, as this was the majority (21/30) of the TBS observed.

We extended our previous survey of SpS by screening (*SI Text*) a panel of 33 tumors for hotspot mutations in *FGFR3* and *HRAS* (Table S4). No further *FGFR3* mutations were found, but two additional tumor samples harbored apparently homozygous *HRAS* mutations, not observed in matched histologically normal tissue. The mutations were c.37G>C (p.G13R) and c.182A>G (p.Q61R) in tumors from men aged 79 and 81 y, respectively (Fig. S3 *A* and *B*). Although these data confirm that *HRAS* is the most commonly mutated gene in SpS (11%) and mutation positivity is strongly correlated with patient age (Fig. S3*C*), no mutations at codon p.G12 were identified. Currently, all *HRAS* mutations found in SpS are mutually exclusive with CS mutations, which may reflect either embryonic or fetal lethality due to the highly activating nature of the mutations associated with SpS (35) or the inability of mutant SSC to produce differentiating meiotic cells and sperm.

As illustrated in Fig. 2*D*, mutations at the p.G12 codon occur in different relative proportions in sperm compared with cancers. In cancers, HRAS p.G12V, which exhibits the lowest GTPase activity (36) and the highest transformation potential (23), accounts for 64% of mutations at codon p.G12 (Fig. 1), whereas in sperm, HRAS p.G12S (c.34G>A) is most abundant, despite its lower transforming activity. These observations point to a different mechanism of mutation and/or selection in spermatogonia than occurs in most tumors, which we investigated further by statistical modeling.

Modeling Mutation Rate and Selective Advantage. Overall, our findings suggest that the *HRAS* mutation levels in sperm are determined by an interplay between the intrinsic genomic mutation rate of a residue and the selective advantage conferred by the resulting mutant protein on the spermatogonial cell (8, 10, 11). To understand the relative impact of these two factors in shaping the outcome of selfish selection, we developed a statistical model (*SI Text*). We elaborated a simple model (9) in which from the age of puberty (13 y), SSC homeostasis is maintained by regular asymmetrical divisions, i.e., each division generates a daughter spermatogonial cell and a differentiating cell that will ultimately produce sperm. Selfish mutations are predicted to modify the SSC mitotic behavior, allowing occasional symmetric divisions, leading to an exponential enrichment of mutations in sperm over time. To account for the fact that contributing mutations are anticipated to be rare, we model their occurrence as a Poisson random variable with parameter μ (i.e., the mutation rate per cell division). We then define a selection coefficient parameter (s) that corresponds to the probability of the occurrence of such symmetric division at each SSC mitosis. Values of μ and s were then inferred by Monte Carlo simulation for each *HRAS* substitution at codons p.A11 and p.G12 in the 89 sperm samples, with partitioning of the data for the 46 individuals heterozygous for the rs12628 SNP across the two alleles.

The model yields significantly positive values of s for activating *HRAS* mutations at p.G12, whereas for the synonymous or nonactivating mutations at p.A11, s is close to zero (Fig. 4 and Table S5). Although s for different mutations at p.G12 reflects their documented in vitro activating properties (lowest for p.G12S and highest for p.G12V) (23, 24), the narrow range of values of s, within 1.5-fold, means that the relative abundance of a given p.G12 mutation is mainly determined by its mutability μ which varies by a factor sevenfold between c.34G>A (highest; transition at CpG dinucleotide) and c.35G>T (lowest; transversion at non-CpG). The effect of relative mutability is apparent when examining the distribution of mutations on the two *HRAS* alleles in individuals heterozygous for the rs12628 SNP. Whereas levels of rarer mutations (including TBS) generally exhibit a marked skewing preference on one or other allele, indicating that as few as one originating mutation could have contributed to the final levels, this skewing is much less marked for the c.34G>A transition (p.G12S) because of its higher intrinsic mutation rate (*SI Text* and Fig. S2*C*).

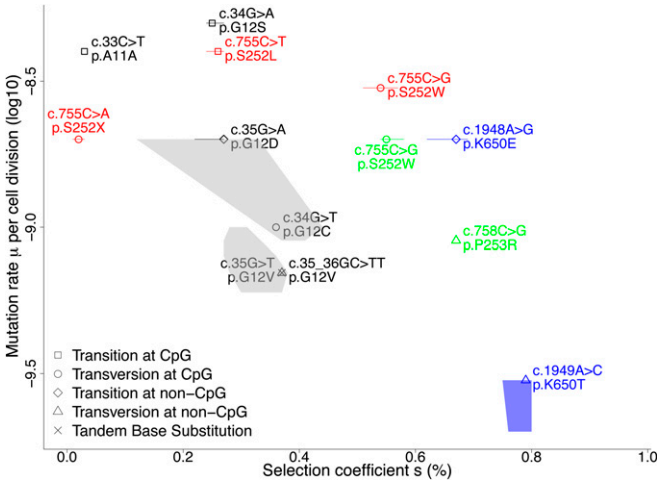

Fig. 4. Contributions of mutation rate (μ) and selection (s) to levels in sperm for mutations in *HRAS*, *FGFR2,* and *FGFR3*. Data for *HRAS* (this work) in black; references for data from previous studies are color-coded as red (8), green (12), and blue (10). Bars and shaded areas represent the 95% confidence intervals (Table S5).

To test further the usefulness of the model, we analyzed three previously published datasets of mutation levels quantified in sperm for substitutions involving *FGFR2* c.755C (8, 12), *FGFR2* c.758C (12), and *FGFR3* c.1948A or c.1949A (10). Although these datasets were obtained using different methodologies, estimates of μ for a given category of substitution broadly agreed both between the datasets and with previously obtained mutation rate estimates (1, 2, 6). Estimates of s and μ for the Apert c.755C>G mutation originating from two independent datasets are also in good agreement. Notably, selection coefficients for the most strongly selected mutations in *FGFR2* and *FGFR3* are 1.5- to 2.1-fold higher than for the most strongly selected mutation in *HRAS*, which is likely to account for the lower birth prevalence of CS (*SI Text*) compared with the disorders associated with the specific *FGFR2* and *FGFR3* mutations (11) (Table S5).

Discussion

The c.35G>T (p.G12V) substitution in *HRAS* is of considerable historical significance, because in 1982 it was the first missense change in a proto-oncogene to be implicated in cancer (37, 38). Three decades later, it is known to be the most frequent oncogenic mutation in *HRAS* (COSMIC), but a rare cause of CS arising through germline mutation (Fig. 1). Here, we have determined the distribution of mutation levels at the p.G12 codon in sperm and used these observations to model their occurrence based on intrinsic mutation rate μ and selection coefficient s. We find that although the levels of mutation in sperm are markedly elevated through a process akin to oncogenesis and are consistent with a mechanism involving selfish selection, there are also differences in the outcome between spermatogenesis and classical oncogenesis. These variations are likely to reflect several underlying biological processes, including differences in intrinsic mutation rates and the specific effects of mutation on proliferation, differentiation, and survival of SSC, acting over many years.

Regarding the primary mutations that fuel the eventual supply of mutant sperm, the most surprising observation was of multiple TBS, particularly c.35_36GC>TT, which had an estimated μ indistinguishable from the c.35G>T transversion encoding the same amino acid change, p.G12V (Fig. 4). Considerable confidence that these events are real and not experimental artifacts is provided by the observation of multiple TBS in CS (Table S1). Although changes in two or more nucleotides arising through independent mutational events are expected to be extremely rare ($\sim 10^{-11}$) (2), several studies suggest that these events are more common than expected by chance (26, 39, 40). TBS can either result from a single concomitant mutational event involving adjacent nucleotides or have arisen through two hits of increasing selectivity, a mechanism that has been demonstrated in individuals heterozygous for rare *FGFR2* mutations (17). In the case of TBS in sperm at the *HRAS* p.G12 codon, a mechanism involving a single concomitant mutational event is suggested by three observations. First, nearly all complex changes (30/31) involved bases adjacent to one another within the p.G12 codon (c.34_35 or c.35_36). Second, 28/30 TBS encode amino acid changes that are also observed as single-nucleotide substitutions (encoding p.G12V and p.G12D) and among these, 13/28 occurred in samples that have no detectable levels of the corresponding single-nucleotide substitution (c.35G>A or c.35G>T). Third, most TBS (25/30) involved a C>T transition at position c.36 (encoding a synonymous change as single substitution and not a known polymorphism); as c.36C is a part of a CpG dinucleotide (c.36_37), it raises the possibility that hypermutability at this site could influence the apparent mutation rate of the adjacent nucleotide. This proposal is supported by the threefold enrichment of 3′-adjacent guanine nucleotides in the case of GC>TT/AA TBS in humans, compared with GC>TC SNP. Hypermutability associated with methylated CpG sequence context has been described in UV-induced CC>TT dypirimidine changes observed in sun-exposed skin lesions (28). In nucleotide excision repair-deficient cells, methylated CpG sequences frequently undergo

CG>TT tandem mutations in response to oxidative DNA damage (41). Our work highlights a predisposition to specific TBS that seems largely restricted to germ cells, and should stimulate efforts to investigate its biochemical nature.

Our statistical model incorporated a selection parameter *s*, defined as the probability of symmetric division at each SSC mitosis. Reassuringly, estimates of *s* for mutations at the neutral p.A11 control codon were close to zero, whereas we found positive values of *s* for all mutations at the p.G12 codon, consistent with clonal expansion and selfish selection. Moreover *s* was highest for p.G12V (c.35G>T) and lowest for p.G12S (c.34G>A) (Fig. 4 and Table S5), consistent with their relative in vitro transforming potential (23, 24). Of note, *s* may encompass a number of biological processes other than the balance between symmetric and asymmetric division, including differential survival of cells undergoing stochastic divisions (42) and cell competition (43). In this context, survival simply implies the production of mature sperm, so this could be impaired by several pathologies such as spermatogenic arrest, senescence, or apoptosis (44). In any case, the net result of the narrow range of *s* for different mutations at the p.G12 codon of *HRAS* is that μ outweighs *s* in determining that the most prevalent mutation, both in sperm and in CS, is c.34G>A (p.G12S).

Finally, it will be of interest to consider the present results when analyzing de novo mutation load on a genomewide scale. Although direct estimates of mutation rate based on WGS of family trios have singled out the importance of paternal age as the major determinant of the total number of de novo mutations (6, 7), it is apparent that the vast majority of reported mutations occur in noncoding parts of the genome and are likely to be neutral. Therefore, characterization of the influence of paternal age, not only on the total mutational load, but specifically for different functional classes of mutations, might provide a means to estimate what fraction of these newly acquired mutations is likely to be attributable to mechanisms such as selfish selection, and hence the overall role that this process plays in genome diversity and disease.

Materials and Methods

Single ejaculates from 89 healthy men (aged 22–74 y) were donated anonymously, and the age of the donor was recorded. Blood samples were obtained from seven individuals aged 36–71 y. Written informed consent was obtained from all donors, and samples were collected with the permission of the Oxfordshire Research Ethics Committee (OxREC C03.076). For SpS analysis, 33 samples were collected from tissue archives. For a detailed description of the methods, see *SI Text*.

ACKNOWLEDGMENTS. We thank Steve Twigg and Oliver Venn for helpful discussions, and the High-Throughput Genomics Group for the generation of the sequencing data. Financial support was provided by Wellcome Trust Programme Grants 091182 (to A.G., G.M., and A.O.M.W.) and 086084 (to G.M.), Research Training Fellowship 090120 (to E.M.M.B.W.), and Core Award 090532 to Wellcome Trust Centre for Human Genetics; Medical Research Council Hub Grant G0900747; and Danish Cancer Society Grant A2127 (to E.R.-D.M.). I.T. is a recipient of an Engineering and Physical Sciences Research Council studentship.

1. Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156(1):297–304.
2. Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 21(1):12–27.
3. Campbell CD, Eichler EE (2013) Properties and rates of germline mutations in humans. *Trends Genet* 29(10):575–584.
4. Roach JC, et al. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978):636–639.
5. Conrad DF, et al.; 1000 Genomes Project (2011) Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43(7):712–714.
6. Kong A, et al. (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–475.
7. Michaelson JJ, et al. (2012) Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151(7):1431–1442.
8. Goriely A, McVean GA, Röjmyr M, Ingemarsson B, Wilkie AOM (2003) Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* 301(5633):643–646.
9. Qin J, et al. (2007) The molecular anatomy of spontaneous germline mutations in human testes. *PLoS Biol* 5(9):e224.
10. Goriely A, et al. (2009) Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors. *Nat Genet* 41(11):1247–1252.
11. Goriely A, Wilkie AOM (2012) Paternal age effect mutations and selfish spermatogonial selection: Causes and consequences for human disease. *Am J Hum Genet* 90(2):175–200.
12. Yoon SR, et al. (2009) The ups and downs of mutation frequencies during aging can account for the Apert syndrome paternal age effect. *PLoS Genet* 5(7):e1000558.
13. Choi SK, Yoon SR, Calabrese P, Arnheim N (2008) A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common human disease mutations. *Proc Natl Acad Sci USA* 105(29):10143–10148.
14. Choi SK, Yoon SR, Calabrese P, Arnheim N (2012) Positive selection for new disease mutations in the human germline: Evidence from the heritable cancer syndrome multiple endocrine neoplasia type 2B. *PLoS Genet* 8(2):e1002420.
15. Shinde DN, et al. (2013) New evidence for positive selection helps explain the paternal age effect observed in achondroplasia. *Hum Mol Genet* 22(20):4117–4126.
16. Yoon SR, et al. (2013) Age-dependent germline mosaicism of the most common Noonan syndrome mutation shows the signature of germline selection. *Am J Hum Genet* 92:917–926.
17. Goriely A, et al. (2005) Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. *Proc Natl Acad Sci USA* 102(17):6051–6056.
18. Lim J, et al. (2012) Selfish spermatogonial selection: Evidence from an immunohistochemical screen in testes of elderly men. *PLoS ONE* 7(8):e42382.
19. Sol-Church K, Stabley DL, Nicholson L, Gonzalez IL, Gripp KW (2006) Paternal bias in parental origin of HRAS mutations in Costello syndrome. *Hum Mutat* 27(8):736–741.
20. Zampino G, et al. (2007) Diversity, parental germline origin, and phenotypic spectrum of de novo HRAS missense changes in Costello syndrome. *Hum Mutat* 28(3):265–272.
21. Scheffzek K, et al. (1997) The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* 277(5324):333–338.
22. Lee J, et al. (2009) Genetic reconstruction of mouse spermatogonial stem cell self-renewal in vitro by Ras-cyclin D2 activation. *Cell Stem Cell* 5(1):76–86.
23. Fasano O, et al. (1984) Analysis of the transforming potential of the human H-ras gene by random mutagenesis. *Proc Natl Acad Sci USA* 81(13):4008–4012.
24. Seeburg PH, Colby WW, Capon DJ, Goeddel DV, Levinson AD (1984) Biological properties of human c-Ha-ras1 genes mutated at codon 12. *Nature* 312(5989):71–75.
25. Bracho MA, Moya A, Barrio E (1998) Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. *J Gen Virol* 79(Pt 12):2921–2928.
26. Chen JM, Férec C, Cooper DN (2013) Patterns and mutational signatures of tandem base substitutions causing human inherited disease. *Hum Mutat* 34(8):1119–1130.
27. Gimm O, et al. (1997) Germline dinucleotide mutation in codon 883 of the RET proto-oncogene in multiple endocrine neoplasia type 2B without codon 918 mutation. *J Clin Endocrinol Metab* 82(11):3902–3904.
28. Alexandrov LB, et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415–421.
29. Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G (2012) De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat Genet* 44(2):226–232.
30. Abecasis GR, et al.; 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.
31. Aoki Y, et al. (2005) Germline mutations in HRAS proto-oncogene cause Costello syndrome. *Nat Genet* 37(10):1038–1040.
32. van der Burgt I, et al. (2007) Myopathy caused by HRAS germline mutations: Implications for disturbed myogenic differentiation in the presence of constitutive HRas activation. *J Med Genet* 44(7):459–462.
33. Burkitt-Wright EM, et al. (2012) Neonatal lethal Costello syndrome and unusual dinucleotide deletion/insertion mutations in HRAS predicting p.Gly12Val. *Am J Med Genet A* 158A(5):1102–1110.
34. Lorenz S, et al. (2012) Two cases with severe lethal course of Costello syndrome associated with HRAS p.G12C and p.G12D. *Eur J Med Genet* 55(11):615–619.
35. Der CJ, Finkel T, Cooper GM (1986) Biological and biochemical properties of human rasH genes mutated at codon 61. *Cell* 44(1):167–176.
36. Colby WW, Hayflick JS, Clark SG, Levinson AD (1986) Biochemical characterization of polypeptides encoded by mutated human Ha-ras1 genes. *Mol Cell Biol* 6(2):730–734.
37. Reddy EP, Reynolds RK, Santos E, Barbacid M (1982) A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* 300(5888):149–152.
38. Tabin CJ, et al. (1982) Mechanism of activation of a human oncogene. *Nature* 300(5888):143–149.
39. Averof M, Rokas A, Wolfe KH, Sharp PM (2000) Evidence for a high frequency of simultaneous double-nucleotide substitutions. *Science* 287(5456):1283–1286.
40. Schrider DR, Hourmozdi JN, Hahn MW (2011) Pervasive multinucleotide mutational events in eukaryotes. *Curr Biol* 21(12):1051–1054.
41. Lee DH, O'Connor TR, Pfeifer GP (2002) Oxidative DNA damage induced by copper and hydrogen peroxide promotes CG—>TT tandem mutations at methylated CpG dinucleotides in nucleotide excision repair-deficient cells. *Nucleic Acids Res* 30(16):3566–3573.
42. Klein AM, Nakagawa T, Ichikawa R, Yoshida S, Simons BD (2010) Mouse germ line stem cells undergo rapid and stochastic turnover. *Cell Stem Cell* 7(2):214–224.
43. Moreno E (2008) Is cell competition relevant to cancer? *Nat Rev Cancer* 8(2):141–147.
44. Paul C, Robaire B (2013) Ageing of the male germ line. *Nat Rev Urol* 10(4):227–234.