nature genetics

Activating mutations in *FGFR3* and *HRAS* reveal a shared genetic origin for congenital disorders and testicular tumors

Anne Goriely[1], Ruth M S Hansen[1], Indira B Taylor[1], Inge A Olesen[2], Grete Krag Jacobsen[3], Simon J McGowan[4], Susanne P Pfeifer[5], Gilean A T McVean[5], Ewa Rajpert-De Meyts[2] & Andrew O M Wilkie[1]

Genes mutated in congenital malformation syndromes are frequently implicated in oncogenesis[1,2], but the causative germline and somatic mutations occur in separate cells at different times of an organism's life. Here we unify these processes to a single cellular event for mutations arising in male germ cells that show a paternal age effect[3]. Screening of 30 spermatocytic seminomas[4,5] for oncogenic mutations in 17 genes identified 2 mutations in *FGFR3* (both 1948A>G, encoding K650E, which causes thanatophoric dysplasia in the germline)[6] and 5 mutations in *HRAS*. Massively parallel sequencing of sperm DNA showed that levels of the *FGFR3* mutation increase with paternal age and that the mutation spectrum at the Lys650 codon is similar to that observed in bladder cancer[7,8]. Most spermatocytic seminomas show increased immunoreactivity for FGFR3 and/or HRAS. We propose that paternal age-effect mutations activate a common 'selfish' pathway supporting proliferation in the testis, leading to diverse phenotypes in the next generation including fetal lethality, congenital syndromes and cancer predisposition.

Spontaneous germline point mutations in humans occur at average rates of 4–160 × 10^{-9} per nucleotide; tend to be paternal in origin, with a male-to-female mutation ratio of between 2:1 and 7:1; and show only a modest effect of parental age on mutation rate[9–13]. However, a different pattern is evident in some congenital disorders, which arise from specific point mutations that are two to three orders of magnitude more common than expected. The causative mutations nearly always originate from the healthy fathers (male-to-female ratio >10:1), who, on average, are 2–6 years older than the population mean. We term mutations with these collective properties 'paternal age-effect mutations': the best-documented examples occur in the genes *FGFR2*, *FGFR3*, *HRAS*, *PTPN11* and *RET* (**Supplementary Table 1**)[14–19]. In all cases the mutations show dominant inheritance and encode missense substitutions with gain-of-function properties.

The pathological basis of paternal age-effect mutations needs to be explained in the context of normal spermatogenesis, in which progeny of diploid stem cells (spermatogonia) have a choice either to self-renew or to differentiate through a series of mitotic and meiotic divisions, leading to mature sperm[20]. Among paternal age-effect mutations, the properties of the 755C>G transversion in *FGFR2* (a cause of Apert syndrome)[14] have been investigated in most detail. A quantitative assay showed that the mutation level is elevated (10^{-6}–10^{-4}) in the sperm of many healthy men compared to the background germline rate, and increases with age[21]. However, the mutation levels are usually distributed very unequally between the two *FGFR2* alleles, a pattern that is inconsistent with the notion that the mutations originate from many independent replication errors during spermatogenesis. Instead, rare initial mutations could become enriched because of a selective advantage, leading to progressive clonal expansion of the mutant spermatogonia and accounting for the allelic skewing and paternal age effect[21,22]. Putative *FGFR2*-mutant clones have been observed in normal testes, as predicted[23]. Additional analyses of sperm and testes for a different Apert syndrome–associated *FGFR2* mutation[24,25], as well as an *FGFR3* mutation that causes achondroplasia[26–28], support a shared mechanism for the origin of paternal age-effect disorders.

The proposed clonal expansions of spermatogonia are reminiscent of the action of oncogenes in neoplasia; consistent with this, somatic *FGFR2* mutations identical to those causing Apert syndrome are frequent in endometrial carcinoma[29,30]. We therefore proposed that the mutant clones in the testis might themselves progress to testicular tumors[31]. Previous attempts to identify *FGFR2* or *FGFR3* mutations in common testicular germ cell tumors (seminomas and non-seminomas) yielded negative results[31,32]; however, these tumors occur predominantly in young men (aged 25–35 years) and arise from a fetal precursor state[33], which is difficult to reconcile with the proposed origin of paternal age-effect mutations. Here we have investigated spermatocytic seminomas, a rare type of testicular germ cell tumor with a later mean age of onset (~54 years). These tumors present as slow-growing, well-circumscribed swellings that rarely metastasize and are thought to originate from the adult spermatogonial lineage[4,5].

We sequenced mutation hot spots in fibroblast growth factor receptor genes (*FGFR1*, *FGFR2* and *FGFR3*) in 30 spermatocytic seminomas (**Supplementary Table 2**) and found the identical 1948A>G transition

[1]Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [2]Department of Growth & Reproduction and [3]Department of Pathology, Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark. [4]Computational Biology Research Group and [5]Department of Statistics, University of Oxford, Oxford, UK. Correspondence should be addressed to A.O.M.W. (awilkie@hammer.imm.ox.ac.uk).

Received 28 July; accepted 21 September; published online 25 October 2009; doi:10.1038/ng.470

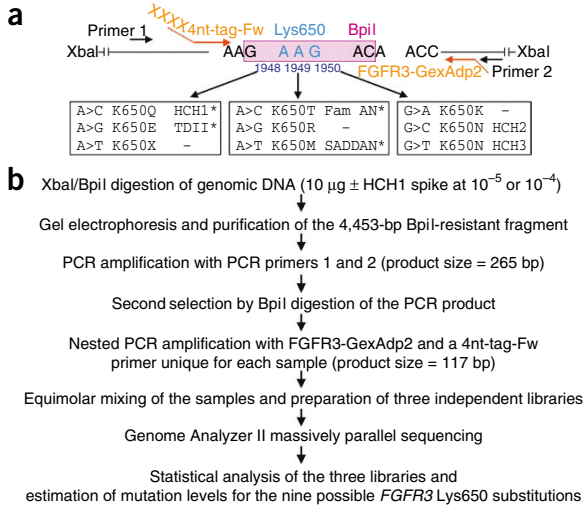

Figure 1 Strategy used to quantify mutation levels at the *FGFR3* Lys650 codon. (**a**) DNA sequence around the Lys650 codon, the relative positions of XbaI and BpiI restriction enzyme sites used for selection, the PCR primers used for amplification (black arrows, first set of PCR primers; orange arrows, nested PCR primers), the nine potential single nucleotide substitutions of the Lys650 codon (located at cDNA position 1948–1950) and the associated germline defects. HCH, hypochondroplasia; TDII, thanatophoric dysplasia type II; Fam AN, familial acanthosis nigricans; SADDAN, severe achondroplasia with developmental delay and acanthosis nigricans; –, not reported as germline mutation. * indicates that the change has also been reported as a somatic mutation in cancer (**Supplementary Table 3**). Note that all substitutions at the Lys650 codon abolish the BpiI recognition site (pink box) and hence are selected by digestion with this enzyme. (**b**) Flow diagram summarizing the strategy adopted for quantification of mutation levels using tagged-oligonucleotide pooled PCR and massively parallel sequencing (see also **Supplementary Fig. 2**).

in *FGFR3* (encoding K650E) in two different tumors (**Supplementary Fig. 1a**). In both cases, histopathologically normal testis adjacent to the tumor did not carry the mutation. This mutation has previously been identified in the germline heterozygous state in the neonatally lethal skeletal disorder thanatophoric dysplasia type II (MIM187601)[6] and as a somatic mutation in bladder tumors[7,8], seborrheic keratoses[34] and multiple myeloma[35] (**Supplementary Table 3**). The FGFR3 K650E substitution is strongly activating, allowing constitutive autophosphorylation of the intracellular tyrosine kinase domain in the absence of ligand (**Supplementary Note**)[36,37]. A paternal age effect in thanatophoric dysplasia was described previously, but genetic studies were not undertaken[38,39].

Based on the birth prevalence of thanatophoric dysplasia (which is caused by several different *FGFR3* mutations)[6,27] of 2.4×10^{-5} (refs. 38,39) and by analogy with other *FGFR2* and *FGFR3* mutations[21–28], we expected the 1948A>G mutation to be present at average levels of ~10^{-5} in sperm. Because five other point mutations of the *FGFR3* Lys650 codon (AAG) have been described in germline congenital disorders[36,40,41] and three of these also as somatic mutations in tumors[8,34,35,42] (**Fig. 1a**, **Supplementary Table 3**), we aimed to quantify all nine possible substitutions at *FGFR3* codon 650. We divided each DNA sample into three aliquots, two of which were spiked with different amounts of diluted genomic DNA heterozygous for the *FGFR3* 1948A>C substitution, to provide an internal standard for absolute quantification of mutation levels. Samples were digested with BpiI (recognition sequence GAAGAC) to enrich equally for all *FGFR3* Lys650 codon substitutions. During subsequent PCR amplification, we used primers containing unique 4-nucleotide tags to identify each sample, and then pooled all products for a given spike level to construct three independent libraries for massively parallel sequencing (**Fig. 1b**).

To assess the accuracy and reproducibility of the assay, we estimated mutation levels for two different *FGFR3* Lys650 substitutions (1948A>G (two samples) and 1949A>T) when progressively diluted

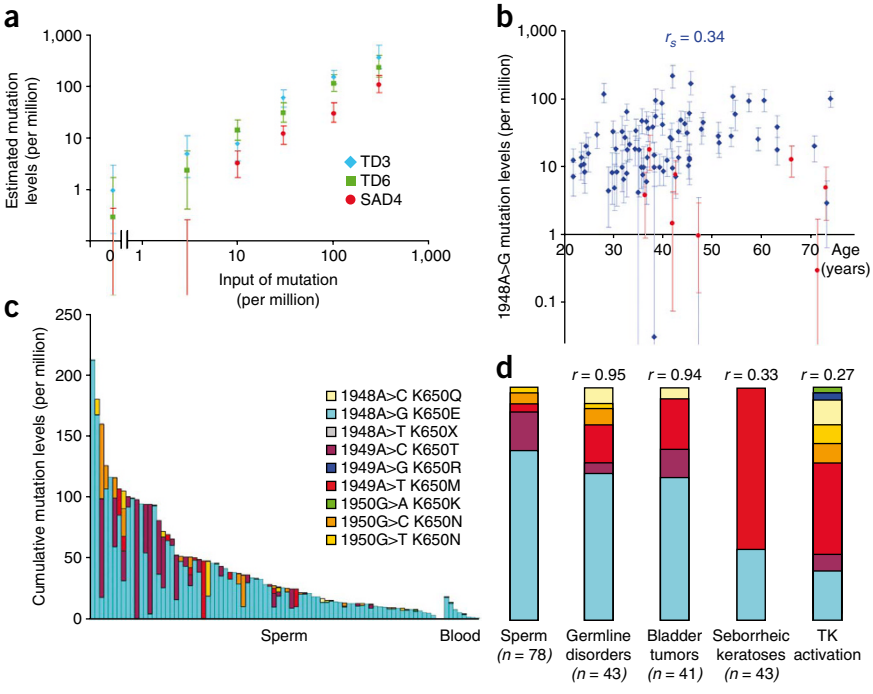

Figure 2 Mutation levels at the *FGFR3* Lys650 codon in sperm and blood quantified by massively parallel sequencing. (**a**) Mutation levels estimated in reconstruction dilution series of three heterozygous control samples (TD3 and TD6, 1948A>G; SAD4, 1949A>T) mixed with normal blood DNA. (**b**) Estimated levels of the 1948A>G (K650E) mutation in blood (red circles, $n = 8$) and sperm (blue diamonds, $n = 78$) in relation to the age of the sample donor. r_s refers to the Spearman's rank correlation coefficient based on the sperm samples. Vertical bars indicate the 95% equal tail probability interval. (**c**) Cumulative levels of each of nine different nucleotide substitutions at the Lys650 codon in sperm (left) and blood (right) samples. (**d**) Correlation of the relative mutation levels in sperm with other pathological features of the *FGFR3* Lys650 codon. From left to right are shown the overall proportions of the nine different nucleotide substitutions at the Lys650 codon in sperm (this work); the relative proportions of mutations in germline disorders, bladder tumors and seborrheic keratoses, each compiled from the literature (**Supplementary Table 3**); and the relative level of tyrosine kinase (TK) activation for each substitution as measured by *in vitro* kinase assay[36]. The Pearson coefficient of correlation r between data for sperm mutations and each of the other four datasets is indicated above the corresponding panel. Color coding of mutations follows key in **c**.

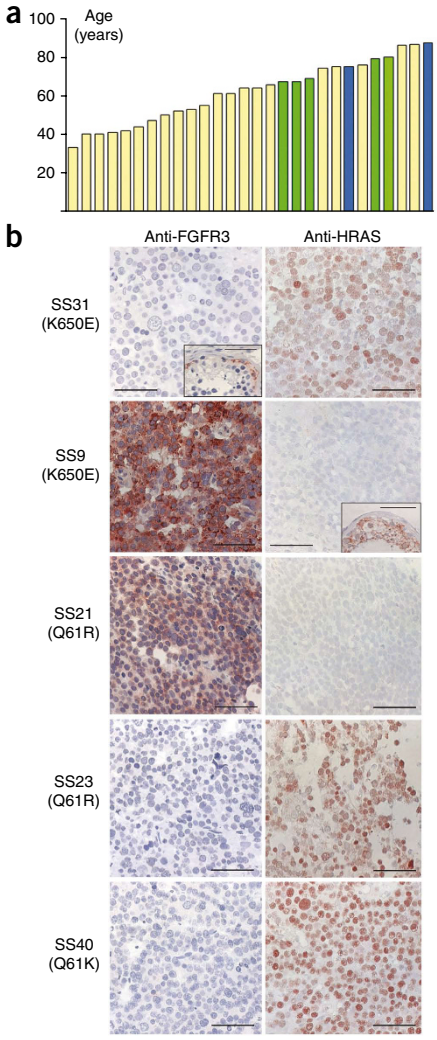

Figure 3 Age distribution in spermatocytic seminomas and immunohistochemical staining of FGFR3 and HRAS. (**a**) Age distribution of subjects at the time of removal of their spermatocytic seminomas. The mutation positive samples are filled in blue (FGFR3 K650E) or green (HRAS Q61R/K); the yellow bars indicate the samples without identified mutation. Age data were unavailable for two tumors. (**b**) Representative immunohistochemical staining for FGFR3 (left) and HRAS (right) antibodies in samples from spermatocytic seminomas mutant for FGFR3 K650E (two top rows) or HRAS Q61R/K mutations (remaining rows). The mutation status of each tumor is indicated. Insets show staining of adjacent normal testicular parenchyma demonstrating ongoing spermatogenesis: note that FGFR3 is normally present in the cell membrane and cytoplasm of spermatogonia at the base of the seminiferous tubule[48], whereas HRAS is found in the nucleus of some basal spermatogonia and in the cytoplasm of primary spermatocytes and round spermatids at the tubular lumen (scale bars, 50 μm).

with normal blood DNA. For both 1948A>G samples we found an excellent correspondence between the amount of input DNA and the estimated mutation level, at least down to an input concentration of 3×10^{-6}. For the 1949A>T sample, mutation levels were underestimated by a factor of ~3.3-fold but were also strongly correlated with the amount of input DNA (**Fig. 2a**).

We estimated levels of the 9 possible single nucleotide substitutions at the *FGFR3* Lys650 codon in 78 sperm and 8 blood samples obtained from healthy donors. Whereas relatively low counts of all mutations were found in the blood samples, levels of the 1948A>G substitution were often much higher (maximum of 2.1×10^{-4}) in sperm samples and were significantly correlated with donor age (Spearman's rank $r_s = 0.34$, $P = 0.002$) (**Fig. 2b**). 1948A>G was the most prevalent substitution in 66/78 sperm donors (**Fig. 2c**; **Supplementary Table 4**) and accounted for 73% of total mutations in these samples. Among the other potential substitutions, 1949A>C (K650T) accounted for 17% of total mutations in sperm and was the most prevalent mutation in 8/78 samples. This change has been described as a constitutional mutation in a few individuals with acanthosis nigricans and mild short stature[41] and as a somatic mutation in bladder tumors[42] (**Supplementary Table 3**). The three other substitutions that accounted for total mutation levels >1%, 1949A>T (K650M) and 1950G>C/T (K650N), have also been observed in constitutional disorders (**Fig. 1a**, **Supplementary Table 3**)[36,40]. By contrast, the three point mutations that encode silent, conservative or stop changes (**Fig. 1a**), which have not been reported as either germline or somatic mutations, were all >1,000-fold less prevalent than 1948A>G.

The variation in prevalence of different mutations at the Lys650 codon, and correspondence with the functional effect of the encoded substitution, suggest differential selection of cells expressing mutant proteins. We compared the average levels in sperm of the nine *FGFR3* Lys650 point mutations with four other measurements for this codon (**Fig. 2d**, **Supplementary Table 3**). There was a strong correlation ($r = 0.95$) with the total number of cases reported of each germline mutation, indicating that the level of mutations in sperm is likely to be the major determinant of the population prevalence of different pathogenic germline mutations. There was also a strong correlation ($r = 0.94$) with the total cases reported of each somatic mutation in bladder tumors, suggesting that the Lys650-mutant sperm are produced by an oncogenic process. The correlations were substantially weaker with the mutation distribution observed in seborrheic keratoses ($r = 0.33$) and with the relative activation potential of the tyrosine kinase measured by *in vitro* assay ($r = 0.27$)[36]: in both contexts, the gain-of-function effect of the K650M mutation appears to surpass that of K650E (**Fig. 2d**). However, heterozygosity for K650M, or for the equivalent K644M mutation in mouse *Fgfr3*, leads to a viable phenotype in both species[40,43], whereas K650E (K644E in mouse) is lethal[6,44], which demonstrates that the *in vitro* kinase measurement does not capture all dimensions of the pathological consequences of these two substitutions (see **Supplementary Note**).

We screened 14 additional genes in the spermatocytic seminomas, including (i) genes mutated in syndromes exhibiting a strong paternal age effect (*HRAS*, *PTPN11*, *RET*); (ii) genes involved in the signal transduction pathways (mitogen-activated protein kinase (MAPK) and the phosphoinositide-3 kinase (PI3K)) of proteins encoded by class (i) genes and for which pathogenic activating mutations have been reported (*AKT1*, *BRAF*, *KRAS*, *MAP2K1*, *MAP2K2*, *NRAS*, *PIK3CA*, *SOS1*); and (iii) genes for which oncogenic mutations are commonly found in tumors (bladder, thyroid and endometrial cancers) in which paternal age-effect mutations have also been described (*CTNNB1*, *EGFR*, *KIT*). Five mutations (all in tumors negative for *FGFR3* mutations) were found in *HRAS* at the Gln61 codon; three were 182A>G (Q61R) transitions and two were 181C>A (Q61K) transversions. All mutations were apparently homozygous, and they were absent in adjacent normal tissue in the four available cases (**Supplementary Fig. 1b**). Q61R and Q61K substitutions are common in human cancers (Catalogue of Somatic Mutations in Cancer: http://www.sanger.ac.uk/perl/genetics/CGP/cosmic) and are both highly activating in a transformation assay[45]. In the germline, heterozygous *HRAS* mutations cause Costello syndrome

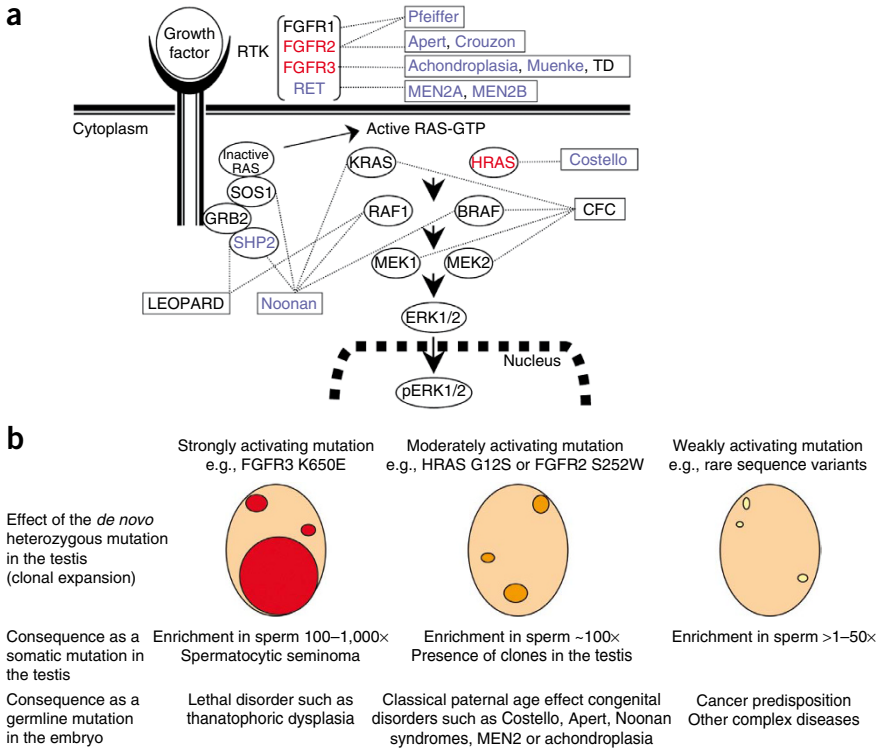

Figure 4 Pathways and phenotypic consequences of selfish mutations in the testis. (**a**) Simplified depiction of components of the growth factor receptor–RAS–MAPK signaling pathway (ovals) and associated congenital disorders (boxes). The disorders and proteins for which a paternal age effect has been documented for mutations in the associated gene are labeled in blue; red highlights those cases for which there is additional experimental evidence for increased levels of causative mutations in testes. CFC, cardio-facio-cutaneous syndrome; MEN, multiple endocrine neoplasia; RTK, receptor tyrosine kinase; TD, thanatophoric dysplasia. (**b**) Proposed consequences of mutations depending on the degree of mutant protein activation. The original mutations are rare events that occur randomly in spermatogonial stem cells or progenitors. Strongly activating mutations (left column), such as FGFR3 K650E, lead over time to the formation of large clones within the testis (red ovals), potentially (with additional mutational events) causing spermatocytic seminomas. The high relative enrichment increases the risk of fertilization by mutant sperm, causing lethal germline phenotypes such as thanatophoric dysplasia. For moderately activating mutations (central column), absence of mutations in spermatocytic seminoma suggests that clonal expansions (orange ovals) occur to a lesser degree and are more likely to be self-limiting. These mutations lead to classical congenital disorders. For weakly activating mutations (right, yellow ovals), the limited clonal expansions would be undetectable using current experimental methods but could contribute to an increased burden of many new, potentially oncogenic, mutations in the next generation.

(MIM218040)[16,17,46], but no substitution at Gln61 has been identified, which is likely to reflect lethality of these mutations[17,45,47] (see **Supplementary Note**). No mutations were identified in the other 13 genes screened (**Supplementary Table 2**). The average age of subjects with spermatocytic seminomas who were positive for *FGFR3* or *HRAS* mutations (74.9 years) was significantly greater than for those who were mutation negative (57.6 years) (Student's *t*-test $P = 0.0009$) (**Fig. 3a**).

To evaluate further the role of FGFR3 and HRAS in spermatocytic seminomas, we examined protein distribution in 26 tumors, including all 7 mutation-positive tumors. FGFR3 (ref. 48) and HRAS are expressed in subpopulations of cells in normal seminiferous tubules (**Fig. 3b** insets). pERK1/2, an activated effector of the MAPK pathway (**Fig. 4a**), is present in the nucleus of a subset of spermatogonia, in primary spermatocytes and in a subset of round spermatids (data not shown). In spermatocytic seminomas, we found strong positive staining for FGFR3 in 5/24 tumors (including 1/2 that were *FGFR3* mutation positive) and for HRAS in 15/26 tumors (including 4/5 that were *HRAS* mutation positive) (**Fig. 3b**); 18/26 (70%) of tumors were strongly positive for one or both proteins. In most tumors (18/24), including 6/7 mutated cases, the presence of FGFR3 and HRAS was mutually exclusive, but this did not show a simple correlation with the gene mutated in the mutation-positive cases (**Fig. 3b**). pERK1/2 was detected in 14/25 (56%) of tumors (**Supplementary Table 2**).

Our results support the proposal[31] that clonal expansion resulting from selective advantage of paternal age-effect mutations can lead to testicular tumors. To our knowledge our work links for the first time in any organism the processes of mutation in the soma (causing neoplasia) and germline (causing heritable disorders in the next generation), which normally occur in different cells, to a mutational event likely happening in the same cell. The clonal expansions presumably involve altered dynamics of stem cell self-renewal, through a proliferative advantage (possibly enhanced by preferential survival)[49] compared to neighboring non-mutant cells, analogous to the role of oncogenes in cancer. Only weak advantage (selection coefficient of 0.002–0.01 per cell generation)[3,23,24] is necessary to account for the observed mutation levels in sperm and the paternal-age excess (2–6 years) observed for the associated germline disorders. Our data favor a premeiotic origin[4,5] for spermatocytic seminoma, which is supported by the observation that transfection of mouse spermatogonial stem cells with mutant *HRAS* (encoding a G12V substitution) causes seminomatous tumors[50]. Men originating tumors containing *FGFR3* and *HRAS* mutations were significantly older than those without such mutations (**Fig. 3a**), suggesting that the mutated tumors represent a distinct pathological subset.

These and previous data[21–28] indicate that activating mutations in *FGFR2*, *FGFR3* and *HRAS* promote clonal expansion in the testis. The encoded proteins are physiologically connected, as HRAS lies downstream of FGFRs in the growth factor/receptor signaling pathway (**Fig. 4a**)[2,46]. By considering the additional genes (*RET* and *PTPN11*) known to be subject to paternal age-effect mutations (**Supplementary Table 1**), these connections can be extended. Thus, RET, like the FGFRs, is a receptor tyrosine kinase (RTK) that signals through RAS and is crucial in spermatogonial renewal[51]. Overexpression of GDNF, the RET ligand, leads to accumulation of undifferentiated spermatogonia and testicular tumors in older mice[52]. *PTPN11* (encoding SHP2) positively regulates RTK-RAS signaling (**Fig. 4a**)[53].

A central role for abnormal FGFR3-RAS signaling in the origin of spermatocytic seminoma is supported by the immunohistochemical analysis (**Fig. 3b**). Both FGFR3 and RAS are present in normal

testis but show markedly elevated staining in most spermatocytic seminomas, including those lacking *FGFR3* and *HRAS* mutations (in which other mechanisms are presumably causing upregulation of the proteins). The reciprocal pattern of increased FGFR3 or HRAS staining in many tumors suggests that elevation of either component is sufficient for pathway activation. Similar observations have been made in low-grade bladder cancer, in which mutations in *FGFR3* and RAS genes are mutually exclusive[54] and FGFR3 mutation status and protein expression do not always correlate[8]. The spectrum of *FGFR3* Lys650 codon mutations in sperm is very similar to that in bladder cancer (**Fig. 2d**), suggesting that similar mutation-selection mechanisms operate in these distinct cellular contexts.

RAS activates multiple pathways, including those typically involved in proliferation (MAPK) and survival (PI3K)[2]. Activating mutations of genes encoding downstream components of the MAPK pathway (BRAF, RAF1, MEK1 and MEK2), cause neuro-cardio-facial-cutaneous syndromes that overlap with those resulting from *PTPN11* and *HRAS* mutations (**Fig. 4a**)[2,46]; inhibition of the MAPK pathway in mice carrying specific *Fgfr2* and *Ptpn11* mutations ameliorates their abnormal phenotypes[55,56]. In 56% of spermatocytic seminomas we observed pERK1/2 staining, indicating pathological activation of the MAPK pathway (**Supplementary Table 2**). However, the PI3K pathway has also been functionally implicated in spermatogonial self-renewal[50].

We envisage a range of consequences for these selfish[27] mutations occurring in spermatogonial cells (**Fig. 4b**). Heterozygosity for the most highly activating mutations, such as FGFR3 K650E and HRAS Q61R/K, causes severe, lethal phenotypes when transmitted in the germline[6,17]. Additional secondary genetic changes at these loci (**Supplementary Note**) would lead, in combination with other mutations[57], to spermatocytic seminoma. Moderately activating mutations, for example those encoding substitutions at Gly12 or Gly13 in HRAS (Costello syndrome)[16,17,46], in *FGFR2* (Apert syndrome)[14,21–25] and in *FGFR3* (achondroplasia)[15,26,28], lead to clonal expansion that is eventually limited by growth arrest or senescence[58] before overt tumors become apparent. We view this process as analogous to that occurring in skin, where a spectrum of activating *FGFR3* mutations in keratinocytes leads to seborrheic keratoses[34]. Diverse mutations in spermatogonia that confer weaker selective advantage may lead to lower levels of enrichment (>1–50-fold) in sperm. As well as encoding missense substitutions, mutations could confer altered gene expression, which shows a high fraction of positively selected changes in testis[59]. Such subtle effects would be technically challenging to detect, yet could contribute substantially to the burden of human disease, through a common disease–rare variant mechanism[60]. Cancer-predisposing mutations are especially likely to be favored by this process, posing an increased risk to offspring of older fathers.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. GenBank: *FGFR3* genomic DNA, NT_006081; *FGFR3* exon IIIc cDNA, NM_000142; *HRAS* cDNA, NM_005343. cDNA numbering is given relative to the A (= 1) of the ATG initiation codon; note that the *FGFR3* Lys650 (K650) codon is alternatively numbered Lys652 (K652) in the exon IIIb spliceform.

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

ACKNOWLEDGMENTS

We thank the Oxford Fertility Clinic and anonymous sperm donors for help with obtaining samples; L. Andersen, T. Chin-A-Woeng, K. Clark, K. Cook, A. Fenwick, A. Herlihy, H. Kistrup, J. Lim, J.-E. Nielsen, L. Thompson and N. Ward for expert technical support and advice; associates of the Rajpert-De Meyts and Wilkie labs for discussions; staff in numerous pathology departments in Denmark and Lund (Sweden), and G. Turner and I. Roberts, for histopathological support and samples; L. Legeai-Mallet for providing control genomic samples; G. Spagnoli for the MAGE-A4 antibody; M. de Gobbi for information on TaqMan primers for DNA quantification; and S. Robertson for comments on the manuscript. This work was funded by the Danish Cancer Society (DP08147 to E.R.-D.M.) and the Wellcome Trust (078666 to A.O.M.W.).

AUTHOR CONTRIBUTIONS

A.G. designed and performed experiments, analyzed data and wrote the paper; R.M.S.H., I.B.T. and I.A.O. performed experiments; G.K.J. collected tumor samples and constructed tissue arrays; S.J.M., S.P.P. and G.A.T.M. developed analytical tools and analyzed data; E.R.-D.M. collected tumor samples, designed and performed experiments; A.O.M.W. designed experiments, analyzed data and wrote the paper.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Friedman, J.M. Genetics and epidemiology, congenital anomalies and cancer. *Am. J. Hum. Genet.* **60**, 469–473 (1997).
2. Schubbert, S., Shannon, K. & Bollag, G. Hyperactive Ras in developmental disorders and cancer. *Nat. Rev. Cancer* **7**, 295–308 (2007).
3. Crow, J.F. Age and sex effects on human mutation rates: an old problem with new complexities. *J. Radiat. Res. (Tokyo)* **47**(Suppl. B), B75–B82 (2006).
4. Eble, J.N. Spermatocytic seminoma. *Hum. Pathol.* **25**, 1035–1042 (1994).
5. Rajpert-De Meyts, E. *et al.* The immunohistochemical expression pattern of Chk2, p53, $p19^{INK4d}$, MAGE-A4 and other selected antigens provides new evidence for the premeiotic origin of spermatocytic seminoma. *Histopathology* **42**, 217–226 (2003).
6. Tavormina, P.L. *et al.* Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nat. Genet.* **9**, 321–328 (1995).
7. Cappellen, D. *et al.* Frequent activating mutations of *FGFR3* in human bladder and cervix carcinomas. *Nat. Genet.* **23**, 18–20 (1999).
8. Tomlinson, D.C., Baldo, O., Harnden, P. & Knowles, M.A. FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer. *J. Pathol.* **213**, 91–98 (2007).
9. Nachman, M.W. & Crowell, S.L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
10. Kondrashov, A.S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* **21**, 12–27 (2002).
11. Taylor, J., Tyekucheva, S., Zody, M., Chiaromonte, F. & Makova, K.D. Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol. Biol. Evol.* **23**, 565–573 (2006).
12. Böhm, J., Munk-Schulenburg, S., Felscher, S. & Kohlhase, J. *SALL1* mutations in sporadic Townes-Brocks syndrome are of predominantly paternal origin without obvious paternal age effect. *Am. J. Med. Genet.* **140A**, 1904–1908 (2006).
13. Arnheim, N. & Calabrese, P. Understanding what determines the frequency and pattern of human germline mutations. *Nat. Rev. Genet.* **10**, 478–488 (2009).
14. Moloney, D.M. *et al.* Exclusive paternal origin of new mutations in Apert syndrome. *Nat. Genet.* **13**, 48–53 (1996).
15. Wilkin, D.J. *et al.* Mutations in fibroblast growth-factor receptor 3 in sporadic cases of achondroplasia occur exclusively on the paternally derived chromosome. *Am. J. Hum. Genet.* **63**, 711–716 (1998).
16. Sol-Church, K., Stabley, D.L., Nicholson, L., Gonzalez, I.L. & Gripp, K.W. Paternal bias in parental origin of *HRAS* mutations in Costello syndrome. *Hum. Mutat.* **27**, 736–741 (2006).
17. Zampino, G. *et al.* Diversity, parental germline origin, and phenotypic spectrum of de novo *HRAS* missense changes in Costello syndrome. *Hum. Mutat.* **28**, 265–272 (2007).
18. Tartaglia, M. *et al.* Paternal germline origin and sex-ratio distortion in transmission of *PTPN11* mutations in Noonan syndrome. *Am. J. Hum. Genet.* **75**, 492–497 (2004).
19. Carlson, K.M. *et al.* Parent-of-origin effects in multiple endocrine neoplasia type 2B. *Am. J. Hum. Genet.* **55**, 1076–1082 (1994).
20. Dym, M., Kokkinaki, M. & He, Z. Spermatogonial stem cells: mouse and human comparisons. *Birth Defects Res. C* **87**, 27–34 (2009).
21. Goriely, A., McVean, G.A.T., Röjmyr, M., Ingemarsson, B. & Wilkie, A.O.M. Evidence for selective advantage of pathogenic *FGFR2* mutations in the male germ line. *Science* **301**, 643–646 (2003).
22. Goriely, A. *et al.* Gain-of-function amino acid substitutions drive positive selection of *FGFR2* mutations in human spermatogonia. *Proc. Natl. Acad. Sci. USA* **102**, 6051–6056 (2005).
23. Qin, J. *et al.* The molecular anatomy of spontaneous germline mutations in human testes. *PLoS Biol.* **5**, e224 (2007).

24. Choi, S.-K., Yoon, S.-R., Calabrese, P. & Arnheim, N. 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**, 10143–10148 (2008).
25. Yoon, S.-R. *et al.* The ups and downs of mutation frequencies during aging can account for the Apert syndrome paternal age effect. *PLoS Genet.* **5**, e1000558 (2009).
26. Tiemann-Boege, I. *et al.* The observed human sperm mutation frequency cannot explain the achondroplasia paternal age effect. *Proc. Natl. Acad. Sci. USA* **99**, 14952–14957 (2002).
27. Wilkie, A.O.M. Bad bones, absent smell, selfish testes: the pleiotropic consequences of human FGF receptor mutations. *Cytokine Growth Factor Rev.* **16**, 187–203 (2005).
28. Dakouane Giudicelli, M. *et al.* Increased achondroplasia mutation frequency with advanced age and evidence for G1138A mosaicism in human testis biopsies. *Fertil. Steril.* **89**, 1651–1656 (2008).
29. Pollock, P.M. *et al.* Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes. *Oncogene* **26**, 7158–7162 (2007).
30. Dutt, A. *et al.* Drug-sensitive *FGFR2* mutations in endometrial carcinoma. *Proc. Natl. Acad. Sci. USA* **105**, 8713–8717 (2008).
31. Hansen, R.M.S., Goriely, A., Wall, S.A., Roberts, I.S.D. & Wilkie, A.O.M. Fibroblast growth factor receptor 2, gain-of-function mutations, and tumourigenesis: investigating a potential link. *J. Pathol.* **207**, 27–31 (2005).
32. Bignell, G. *et al.* Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults. *Genes Chromosom. Cancer* **45**, 42–46 (2006).
33. Skakkebæk, N.E., Berthelsen, J.G., Giwercman, A. & Müller, J. Carcinoma-in-situ of the testis: possible origin from gonocytes and precursor of all types of germ cell tumours except spermatocytoma. *Int. J. Androl.* **10**, 19–28 (1987).
34. Hafner, C. *et al.* Spectrum of *FGFR3* mutations in multiple intraindividual seborrheic keratoses. *J. Invest. Dermatol.* **127**, 1883–1885 (2007).
35. Chesi, M. *et al.* Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. *Nat. Genet.* **16**, 260–264 (1997).
36. Bellus, G.A. *et al.* Distinct missense mutations of the FGFR3 Lys650 codon modulate receptor kinase activation and the severity of the skeletal dysplasia phenotype. *Am. J. Hum. Genet.* **67**, 1411–1421 (2000).
37. Chen, H. *et al.* A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Mol. Cell* **27**, 717–730 (2007).
38. Orioli, I.M., Castilla, E.E., Scarano, G. & Mastroiacovo, P. Effect of paternal age in achondroplasia, thanatophoric dysplasia, and osteogenesis imperfecta. *Am. J. Med. Genet.* **59**, 209–217 (1995).
39. Waller, D.K. *et al.* The population-based prevalence of achondroplasia and thanatophoric dysplasia in selected regions of the US. *Am. J. Med. Genet.* **146A**, 2385–2389 (2008).
40. Zankl, A. *et al.* Prenatal and postnatal presentation of severe achondroplasia with developmental delay and acanthosis nigricans (SADDAN) due to the *FGFR3* Lys650Met mutation. *Am. J. Med. Genet.* **146A**, 212–218 (2008).
41. Castro-Feijóo, L. *et al.* Hypochondroplasia and acanthosis nigricans: a new syndrome due to the p.Lys650Thr mutation in the fibroblast growth factor receptor 3 gene? *Eur. J. Endocrinol.* **159**, 243–249 (2008).
42. van Rhijn, B.W.G. *et al.* Novel fibroblast growth factor receptor 3 (*FGFR3*) mutations in bladder cancer previously identified in non-lethal skeletal disorders. *Eur. J. Hum. Genet.* **10**, 819–824 (2002).
43. Iwata, T., Li, C.-L., Deng, C.-X. & Francomano, C.A. Highly activated Fgfr3 with the K644M mutation causes prolonged survival in severe dwarf mice. *Hum. Mol. Genet.* **10**, 1255–1264 (2001).
44. Iwata, T. *et al.* A neonatal lethal mutation in FGFR3 uncouples proliferation and differentiation of growth plate chondrocytes in embryos. *Hum. Mol. Genet.* **9**, 1603–1613 (2000).
45. Der, C.J., Finkel, T. & Cooper, G.M. Biological and biochemical properties of human ras^H genes mutated at codon 61. *Cell* **44**, 167–176 (1986).
46. Aoki, Y., Niihori, T., Narumi, Y., Kure, S. & Matsubara, Y. The RAS/MAPK syndromes: novel roles of the RAS pathway in human genetic disorders. *Hum. Mutat.* **29**, 992–1006 (2008).
47. Srivastava, S.K., Yuasa, Y., Reynolds, S.H. & Aaronson, S.A. Effects of two major activating lesions on the structure and conformation of human ras oncogene products. *Proc. Natl. Acad. Sci. USA* **82**, 38–42 (1985).
48. Juul, A. *et al.* Preserved fertility in a non-mosaic Klinefelter patient with a mutation in the fibroblast growth factor receptor 3 gene: Case Report. *Hum. Reprod.* **22**, 1907–1911 (2007).
49. Nakagawa, T., Nabeshima, Y. & Yoshida, S. Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev. Cell* **12**, 195–206 (2007).
50. Lee, J. *et al.* Genetic reconstruction of mouse spermatogonial stem cell self-renewal in vitro by Ras-cyclin D2 activation. *Cell Stem Cell* **5**, 76–86 (2009).
51. He, Z. *et al.* Gdnf upregulates c-Fos transcription via the Ras/Erk1/2 pathway to promote mouse spermatogonial stem cell proliferation. *Stem Cells* **26**, 266–278 (2008).
52. Meng, X. *et al.* Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489–1493 (2000).
53. Agazie, Y.M., Movilla, N., Ischenko, I. & Hayman, M.J. The phosphotyrosine phosphatase SHP2 is a critical mediator of transformation induced by the oncogenic fibroblast growth factor receptor 3. *Oncogene* **22**, 6909–6918 (2003).
54. Jebar, A.H. *et al.* FGFR3 and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. *Oncogene* **24**, 5218–5225 (2005).
55. Shukla, V., Coumoul, X., Wang, R.-H., Kim, H.-S. & Deng, C.-X. RNA interference and inhibition of MEK-ERK signaling prevent abnormal skeletal phenotypes in a mouse model of craniosynostosis. *Nat. Genet.* **39**, 1145–1150 (2007).
56. Krenz, M. *et al.* Role of ERK1/2 signaling in congenital valve malformations in Noonan syndrome. *Proc. Natl. Acad. Sci. USA* **105**, 18930–18935 (2008).
57. Looijenga, L.H.J. *et al.* Genomic and expression profiling of human spermatocytic seminomas: primary spermatocyte as tumorigenic precursor and *DMRT1* as candidate chromosome 9 gene. *Cancer Res.* **66**, 290–302 (2006).
58. Ota, S., Zhou, Z.Q., Link, J.M. & Hurlin, P.J. The role of senescence and prosurvival signaling in controlling the oncogenic activity of FGFR2 mutants associated with cancer and birth defects. *Hum. Mol. Genet.* **18**, 2609–2621 (2009).
59. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nat. Rev. Genet.* **7**, 693–702 (2006).
60. Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **40**, 695–701 (2008).

ONLINE METHODS

Ethical approval. Approval for the study was provided by the Oxfordshire Research Ethics Committee C (OxREC C03.076) and the Joint Research Ethics Committees of the Copenhagen and Frederiksberg Communes (KF 01 265848).

Biological samples. We obtained 43 spermatocytic seminomas collected from tissue archives at hospitals in Denmark, Sweden and Oxford, UK, and included mutation and immunohistochemistry data on 30 samples for which >30% (average 64%; range 30–98%) of exons amplified. The mean age of patients at the time of tumor removal was 61.9 years (range 33–87 years). In addition to having typical morphological features, the tumor type was confirmed by the observation that the cells stained with MAGE-A4 but were negative for PLAP and OCT-3/4[5,57]. None of the tumors was invasive.

Single ejaculates from 78 men (aged 22.1–73.9 years) and eight blood samples from individuals aged 36.6–73 years were donated anonymously. DNA samples from two patients heterozygous for *FGFR3* 1948A>G (TD3 and TD6), one patient heterozygous for 1949A>T (SAD4) and one patient heterozygous for 1948A>C (HCH1-spike, used as the spike DNA) were a kind gift from L. Legeai-Mallet (Hôpital Necker–Enfants Malades, Paris).

Analysis of oncogenic mutations in spermatocytic seminomas. DNA was extracted from paraffin-embedded samples as described[31], and mutation hot spots were analyzed by PCR amplification and DNA sequencing of the following genes: *AKT1*, *BRAF*, *CTNNB1*, *EGFR*, *FGFR1*, *FGFR2*, *FGFR3*, *HRAS*, *KIT*, *KRAS*, *MAP2K1*, *MAP2K2*, *NRAS*, *PIK3CA*, *PTPN11*, *RET* and *SOS1*. All PCRs were set up in a sterile UV hood in a 30-µl reaction volume using 1× FastStart Buffer, 150 µM dNTPs, 0.16 µM primers (forward and reverse), 0.5 U FastStart *Taq* DNA polymerase and 0.05 U *Pwo* DNA polymerase (both from Roche). When available, restriction digests were used to screen for the mutations of interest. The PCR primers, conditions and genotyping methods are given in **Supplementary Table 5** and the results in **Supplementary Table 2**. The products were cleaned (30 min at 37 °C in 0.2× Exo I buffer, 10 U exonuclease I (New England Biolabs) and 2 U shrimp alkaline phosphatase (SAP, United States Biochemical), followed by 15 min at 85 °C), sequenced in both orientations and run on a ABI 3700 DNA sequencer (Applied Biosystems). Positions where germline and/or somatic mutations have been previously reported were specifically examined on chromatograms and scored independently. To evaluate the zygosity status of the tumor samples, we genotyped 4 single nucleotide polymorphisms (SNPs) (rs2071616, rs2659871, rs41279090 and rs2075526).

Immunohistochemistry. Representative cores (2 mm) of tumor and adjacent normal tissue (when available) were punched from paraffin-embedded blocks of spermatocytic seminomas and control tissues (epididymis, prostate, classical seminoma and embryonal carcinoma). Two tissue microarrays containing 40 cores each were constructed and sectioned at 4 µm. Larger sections from some tumors were individually stained. Immunohistochemical staining was performed using diluted monoclonal mouse antibodies to FGFR3 (B-9, 1:50), HRAS (F-235, 1:80), OCT-3/4 (C-10, 1:250) (all from Santa Cruz Biotechnology), MAGE-A4 (1:2,000; a kind gift from G. Spagnoli, Basel, Switzerland) and PLAP (1:50; DAKO), and a monoclonal rabbit antibody to pERK1/2 (20G11, 1:150; Cell Signaling Technology), by means of a standard indirect peroxidase method. Deparaffinized and rehydrated sections were microwaved and incubated first in 0.5% H_2O_2 and then in goat serum (Histostain kit, Zymed) before the addition of the primary antibodies overnight at 4 °C, while control sections of each specimen were incubated in the dilution buffer alone. After washing in Tris buffer and incubation with a peroxidase-conjugated anti-mouse antibody (or anti-rabbit for pERK), the reaction was developed in the presence of 3-amino-9-ethyl carbazole and H_2O_2 (Histostain kit). All sections were counterstained with Mayer's hematoxylin. The sections were evaluated by two independent observers, and semiquantitative scoring was used to assess the relative abundance of stained cells.

Quantification of *FGFR3* Lys650 codon mutation levels in sperm and blood samples. DNA was extracted from blood and whole ejaculates as described[21], and concentrations were precisely estimated at three dilutions against a dilution series of human genomic DNA (Roche) using the TaqMan PCR conditions (**Supplementary Table 6**), designed to quantify a unique 91-bp amplicon located on chromosome 16 (courtesy of M. de Gobbi).

Measurements of mutation levels around the *FGFR3* Lys650 codon (cDNA position 1948–1950) were performed using a strategy similar to that described[21]. Triplicate samples each containing 10 µg of genomic DNA and either 2 ng (spike level 10^{-4}), 0.2 ng (spike level 10^{-5}) or 0 ng (unspiked) of the HCH1-spike DNA were digested with 120 U XbaI and 40 U BpiI (both from Fermentas) for 4 h. The digested DNA samples, flanked by 2 lanes of Lambda DNA/Eco91I marker (Fermentas), were electrophoresed in a 0.9% Tris-borate-EDTA (TBE) agarose gel (without ethidium bromide). This double digestion generates a 4,453-bp XbaI fragment carrying mutant *FGFR3* Lys650 sequences, whereas the normal BpiI-digested Lys650 sequence yields two fragments of 2,446 bp and 2,007 bp (note that the *FGFR3* genomic reference sequence contains a rare G allele at a known A/G SNP rs7688609, located at cDNA position 1953; this nucleotide is shown as A in **Fig. 1a** and **Supplementary Fig. 2**).

To select for mutant sequences, a gel slice corresponding to 4.2–4.6 kb (the marker lanes were cut out of the gel, stained with ethidium bromide and the 4,325 bp and 4,822 bp fragments were labeled with dye before being replaced in their original position in the gel), was excised and gel purified with an E.Z.N.A. MicroElute kit (Omega Bio-Tek). PCR amplification was performed (PCR1, **Supplementary Table 6**) followed by a second round of selection with 30 U BpiI for 4 h in 100 µl volume to yield selected material referred to as 'Pool2'. For each biological sample, an aliquot of 5 µl of the Pool2 material was used as template for a nested amplification (PCR2) in order to prepare the Illumina libraries (see below).

Reconstruction dilution experiment. Genomic samples heterozygous for *FGFR3* 1948A>G (TD3 and TD6) and 1949A>T (SAD4) mutations were used in reconstruction experiments in which 10 µg of blood genomic DNA was mixed with a diluted series of mutant DNA corresponding to final mutation concentrations of 0 (no added mutant DNA), 3×10^{-6} (0.06 ng), 10^{-5} (0.2 ng), 3×10^{-5} (0.6 ng), 10^{-4} (2 ng) and 3×10^{-4} (6 ng) and taken through the same protocol as the sperm and blood samples (that is, each dilution sample was mixed with three dilutions of the HCH1-spike DNA). The dilution samples were analyzed together with the blood and sperm samples.

Control DNAs. Mutant genomic samples TD3, TD6 and SAD4 and four normal (wild-type) genomic DNA samples were included as controls in the analysis. These samples were taken through the same protocol of amplification with the exception that the BpiI enzyme was omitted from all the incubation steps, and hence there was no selection imposed on these samples.

Preparation of the Illumina libraries for massively parallel sequencing. Three independent libraries were prepared for massively parallel sequencing using a modified version of the Illumina protocol "Digital Gene Expression-Tag Profiling with DpnII". Each library contained a mixture of 112 DNA species and was characterized by a specific amount of the HCH1-spike DNA (library 1 was unspiked, library 2 contained the spike at 10^{-5} and library 3 contained the spike at 10^{-4}). Primer sequences and reaction conditions are provided in **Supplementary Table 6**.

Five microliters of each Pool2 sample was used for PCR amplification (PCR2) using 112 different forward 4nt-tag-Fw primers (containing a common *FGFR3* sequence preceded by a unique 4-nt tag and a DpnII cloning site; **Supplementary Fig. 2**). The tag primers were synthesized, HPLC-purified and checked for synthesis error by mass spectroscopy (Thermo Electron Corporation). After confirming efficient amplification, equal volumes of PCR products were mixed so as to be represented in a roughly equimolar ratio in the library. In parallel, the reconstruction-dilution samples and control DNAs were also amplified with unique 4nt-tag-Fw primers and added to the PCR mix. For each library, pooled PCR products were purified (E.Z.N.A. PCR purification kit), digested with 100 U of DpnII (New England Biolabs) for 1 h 30 min, dephosphorylated using 4 U of SAP for 1 h at 37 °C and heat inactivated for 15 min at 80 °C; this was followed by purification (E.Z.N.A. MicroElute PCR purification kit) and resuspension in 15 µl sterile water. The purified fragments (10 µl) were ligated to the Illumina Gex Adapter 1 (annealed Adapter1a and Adapter1b sequences) using the Adapter ligation conditions. The ligation reaction was electrophoresed in a 3% Tris-acetate-EDTA (TAE) agarose gel and a

slice of the expected size (137 bp) was excised, purified (E.Z.N.A. gel purification kit) and resuspended in 25 µl sterile water. Each library was then enriched by a final amplification (Gex PCR) using 1.5 µl of purified ligated products. The three libraries were prepared independently to avoid cross-contamination and were quantified using a fluorometer (Qubit) before being validated by direct sequencing, pyrosequencing (Biotage) and Bioanalyzer (Agilent).

Massively parallel sequencing of the libraries was performed on three different channels of an Illumina GAII sequencer by ServiceXS (Leiden, The Netherlands) for 36 cycles using the Gex-DpnII sequencing primer (**Supplementary Fig. 2b**).

Estimation of mutation levels from Illumina GAII sequencing data. To estimate the levels of each single nucleotide mutation at the *FGFR3* Lys650 codon (corresponding to cycles 23–25 of the Illumina sequencing scheme; **Supplementary Fig. 2b**), we used a Bayesian approach to fit a model to the observed counts at each nucleotide in the sequencing data that allows for bias in frequencies derived from sequencing error and noise introduced by the rounds of PCR and digestion. The model is similar to that used previously for pyrosequencing[21], but allows for a different error structure arising from the GAII Illumina technology. We discarded reads where the minimum Phred quality score for any of the four bases of the tag (cycles 1–4), or three bases of the Lys650 codon (cycles 23–25), was under 10. We also eliminated reads with apparent frameshift errors or mutations outside the BpiI restriction site (that is, in the invariant 16-nt primer sequence). Let z_i be the $\log_{10}$ frequency of mutation i in the sample. Our prior for z_i is normal $(-8, 2)$ for all non–wild-type species. After addition of the spike at concentration s_j in experiment j ($s_0 = 0, s_1 = 10^{-5}, s_2 = 10^{-4}$), the normalized frequencies of the non–wild-type mutations are given by

$$f_{ij} = \frac{10^{z_i} + I_{i=\text{spike}}\, s_j}{s_j + \sum_i 10^{z_i}},$$

where I is an indicator function that takes the value 0 or 1 depending on whether the mutation in question is the same as the HCH1-spike (1948A>C). We allow each experiment (defined by the level of spike and the lane of the machine on which sequencing was carried out) to have separate bias and signal-to-noise ratio parameters. Specifically, we assume that the counts for the mutations in the sequencing data are multinomial with frequency vector

$$\mathbf{x}_j = \text{Dirichlet}(B_j f_{1j} + w_{1j}, B_j f_{2j} + w_{2j}, \ldots),$$

where B_j represents the signal-to-noise ratio for experiment j and w_{ij} represents the background for mutation i in experiment j. We assume a uniform $(0, 1{,}000)$ prior for each B_j and a uniform$(0, 100)$ prior for each w_{ij}.

Through use of the multinomial-Dirichlet model, we can integrate over $\mathbf{x}$ to give the contribution to the total likelihood from experiment j in a given sample

$$L_j = \frac{\Gamma\left(B_j + W_j\right)}{\prod_i \Gamma\left(B_{ij} f_{ij} + w_{ij}\right)} \times \frac{\prod_i \Gamma\left(B_{ij} f_{ij} + w_{ij} + C_{ij}\right)}{\Gamma\left(B_j + W_j + C_j\right)},$$

where $W_j = \sum_i w_{ij}$, C_{ij} is the number of reads observed with mutation i in experiment j, and $C_j = \sum_i C_{ij}$. Likelihoods are combined across samples and experiments to give a total likelihood.

To estimate parameters, we used Metropolis-Hastings MCMC with sequential update of each parameter, a burn-in period of 3,000 complete parameter updates and a subsequent 10,000 samples. Multiple runs from different starting points were conducted to check for convergence, and visual inspection was used to compare observed to expected values as a means of checking model adequacy. Titration data (**Fig. 2a**), normal wild-type control DNA and biological samples were analyzed together. Samples from the chain are summarized by the posterior mean value and the 95% equal-tailed probability interval (ETPI). Estimated levels of all *FGFR3* Lys650 codon substitutions for sperm and blood samples are given in **Supplementary Table 4**.

URLs. Digital Gene Expression-Tag Profiling with DpnII, http://illumina.ucr.edu/ht/documentation/molbiol-docs/DGE-DpnII-Sample-Prep.pdf/view.

doi:10.1038/ng.470