Published in final edited form as:

Hum Mutat. 2009 January ; 30(1): 39–48. doi:10.1002/humu.20810.

GENOMIC ANALYSIS OF CANCER TISSUE REVEALS THAT SOMATIC MUTATIONS COMMONLY OCCUR IN A SPECIFIC MOTIF

Nick M Makridakis^{1,2}, Lúcio Fábio Caldas Ferraz^{1,3}, and Juergen KV Reichardt⁴

¹Department of Biochemistry and Molecular Biology, Institute for Genetic Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089-9075.

⁴Plunkett Chair of Molecular Biology (Medicine), University of Sydney, Medical Foundation Building (K25), 92 - 94 Parramatta Road, Camperdown, NSW 2006, Australia.

Abstract

Somatic mutations are hallmarks of cancer progression. We sequenced 26 matched human prostate tumor and constitutional DNA samples for somatic alterations in the SRD5A2, HPRT, and HSD3B2 genes, and identified 71 nucleotide substitutions. 79% (56/71) of these substitutions occur within a WKVnRRRnVWK sequence (THEMIS motif; W= A/T, K= G/T, V= G/A/C, R= purine (A/G) and n= any nucleotide), with one mismatch allowed. Literature searches identified this motif with one mismatch allowed in 66% (37/ 56) of the somatic prostate cancer mutations and in 74% (90/ 122) of the somatic breast cancer mutations found in all human genes analyzed. We also found the THEMIS motif with one allowed mismatch in 88% (23/26) of the ras1 gene somatic mutations formed in the SENCAR (SENsitive to skin CARcinogenesis) mouse model, after induction of error-prone DNA repair following mutagenic treatment. The high prevalence of the motif in each of the above mentioned cases cannot be explained by chance ($p < 0.046$). We further identified 27 somatic mutations in the error-prone DNA polymerase genes pol η , pol κ and pol β in these prostate cancer patients. The data suggest that most somatic nucleotide substitutions in human cancer may occur in sites that conform to the THEMIS motif. These mutations may be caused by “mutator” mutations in error-prone DNA polymerase genes.

Keywords

cancer; somatic; mutation; consensus; error prone DNA polymerase

INTRODUCTION

Cancer is thought to evolve through the accumulation of somatic mutations in specific genes, depending on tumor type (Vogelstein and Kinzler, 2004). These mutations are caused by a combination of environmental and heritable factors (Lichtenstein et al., 2000). To date scientists have been unable to identify a common motif at the sites of these somatic mutations suggesting that these somatic events have distinct molecular etiologies, depending, among other factors, on the individual and the type of tumor. We tested the hypothesis that there is a common motif at the sites of somatic mutations in prostate cancer tissue, a tumor type with poorly understood etiology.

Corresponding author: NM Makridakis, Tulane Cancer Center, Tulane University, 1430 Tulane Avenue, SL-68, □New Orleans, LA 70112. E-mail: nmakrida@tulane.edu..

²Current address: Tulane Cancer Center, Tulane University, 1430 Tulane Avenue, SL-68, □New Orleans, LA 70112.

³Current address: Centro de Biologia Molecular e Engenharia Genética, University of Campinas, São Paulo, Brazil.

We screened the HSD3B2, SRD5A2 and HPRT genes for somatic mutations. The SRD5A2 gene (OMIM# 607306) encodes the steroid 5 α -reductase type II enzyme that reduces testosterone (T) to dihydrotestosterone (DHT), the most active androgen in the prostate (Cheng et al, 1993). Thus activating 5 α -reductase mutations may contribute to prostate tumor development. A possible contribution of the SRD5A2 gene to prostate tumor progression was also proposed based on our findings of *de novo* somatic events at the SRD5A2 locus (Akalu et al., 1999a). More recently we reported the biochemical characterization of somatic SRD5A2 mutations in human prostate cancer tissue (Makridakis et al., 2004), including mutations that increase enzyme activity. The HSD3B2 gene (OMIM# 109715) also presents an attractive androgen metabolic candidate gene for prostate cancer risk and progression. The type II 3 β -hydroxysteroid dehydrogenase enzyme encoded by the HSD3B2 gene is mainly expressed in androgenic tissues (Labrie et al, 1992; Lachance et al, 1991), and initiates DHT inactivation (Cheng et al, 1993). However, if a specific mechanism generates somatic mutations at distinct motifs, then it should do so even in genes expected to be unrelated to tumor progression, such as the human HPRT gene. The ubiquitously expressed HPRT gene (OMIM# 308000, 300322) is located on the X chromosome (Pai et al, 1980) and encodes hypoxanthine-guanine phosphoribosyl transferase, an enzyme essential for the purine salvage pathway, responsible for 90% of nucleic acid biosynthesis in normal cells (Zoref and Sperling, 1979).

We report here 71 somatic mutations in these three genes. Analysis of the nucleotide sequence surrounding these mutations revealed that these alterations often occur within a novel motif we call THEMIS (from the ancient Greek goddess of prophecy). In search of a potential mutagenic mechanism we found that the THEMIS motif commonly occurs at the sites of somatic mutations induced by error prone DNA repair in the SENCAR (SENSitive to skin CARcinogenesis) mouse skin model following mutagenic treatment (Chakravarti et al, 2000; Chakravarti et al, 2001).

Error prone (EP) DNA repair (also called translesion synthesis; Goodman, 2002) involves DNA polymerases such as pol η (OMIM# 603968) and pol κ (OMIM# 605650) that are much more accurate when replicating through specific types of DNA damage than undamaged DNA (Pages and Fuchs, 2002). EP DNA polymerases have been proposed to play a role in cancer etiology (e.g. Kunkel, 2003). To date, the only known example of an error prone DNA polymerase causing human cancer comes from pol η (Kunkel, 2003); constitutional DNA mutations that inactivate pol η are associated with a high rate of skin cancers in patients suffering from XPV (*Xeroderma Pigmentosum*-Variant; Johnson et al, 1999). In the absence of pol η , ultraviolet (UV) radiation induced- pyrimidine dimers are bypassed in a manner that generates the mutations which lead to skin cancer (Kunkel, 2003), perhaps by another error prone DNA polymerase (Pages and Fuchs, 2002). Thus error prone polymerase mutations can result in multiple tumor-inducing mutations (mutator phenotype; Loeb et al, 2003) given the right type (and amount) of environmental exposure.

Human pol β (OMIM# 174760) is a DNA polymerase essential for base excision repair (BER) (Sancar et al, 2004). BER is one of the major pathways of DNA repair that removes oxidized and alkylated bases from DNA (Friedberg, 2003). Pol β is not a classic error prone polymerase, yet it causes 67-times more substitution errors than mammalian pol δ (Kunkel 2003). Pol β is also involved in meiotic recombination (Goodman, 2002) and repair of double-stranded DNA breaks through the process of nonhomologous end joining (Wilson and Lieber, 1999). Interestingly, both the pol κ and pol β genes are located in chromosomal regions known to be lost during prostate cancer progression (5q and 8p respectively; e.g. Visakorpi *et al*, 1995). Thus, somatic loss of these polymerases may contribute to prostate tumor progression.

We screened the pol β , η and κ genes for somatic prostate cancer mutations, to test the hypothesis that EP polymerases are commonly mutated in prostate cancer tissue. We report 27

somatic mutations in these EP polymerase genes in the same prostate cancer tissues that have additional somatic mutations in the other analyzed genes reported here (i.e. HSD3B2, HPRT and SRD5A2).

MATERIALS AND METHODS

Tumor specimens

We analyzed 26 patients of Caucasian background with prostatic adenocarcinoma (for further description see Akalu et al, 1999a and Makridakis et al, 2004). These patients underwent radical prostatectomy at the USC Norris Comprehensive Cancer Center. Prostate tissue and blood were collected from each patient. Tumors were staged according to the TNM (Tumor, Nodes, Metastases) staging system (Schroder et al, 1992; see Supplementary Table S1a). Local IRB approval was obtained before study initiation.

Microdissection and DNA Extraction

Specimens were formalin fixed, embedded in paraffin, sectioned, and transferred on microscopic slides where they were deparaffinized and stained with hematoxylin and eosin. Selected populations of carcinoma cells were microdissected and tumor DNA was then extracted from the microdissected cells using a method reported by us earlier (Akalu and Reichardt, 1999b). As control, normal (constitutional) DNA was extracted either from microdissected normal cells adjacent to the tumor or from peripheral blood leukocytes (or both).

Molecular analysis

PCR—The entire coding region of the HSD3B2 gene together with the exon-intron splicing junction boundaries, the putative promoter region and the 5' and 3' untranslated regions (UTRs) were amplified by PCR reactions using sets of primers as previously described (Rheaume et al, 1992; Simard et al, 1993). Purified (desalted) oligonucleotides were obtained from IDT (Coralville, IA). Reactions were performed with the polymerases AmpliTaq Gold (Applied Biosystems, Foster City, CA) or HotStart Taq (Qiagen, Valencia, CA) with their corresponding PCR buffer. Reactions carried out with HotStart Taq had an additional reagent (5X Q-solution) provided by the manufacturer. The reaction mixture consisted of 10-20ng of DNA, 1X PCR buffer, 1.5mM MgCl₂, 200μM dNTPs, 0.1-0.2μM of each forward and reverse primer, 1X Q-solution (when necessary), 2.5 units of polymerase and sterile water, in a final volume of 50μl. The reaction was then covered with mineral oil and subjected to thermal cycling in a RoboCycler Gradient 40 (Stratagene, La Jolla, CA) under the following conditions: an initial pre-PCR heat step of 95°C for 2 min (AmpliTaq Gold) or 15 min (HotStart Taq), 50 cycles of denaturation at 95°C for 1.5 min, annealing at 58-70°C for 1.5 min, and elongation at 72°C for 1.5 min. This was followed by a final extension step at 72°C for 10 min. PCR products were purified with the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and then sequenced.

Exons 7-8 of the HPRT gene were PCR amplified as a single 379 bp fragment as described above, except that we used primers described elsewhere (Liu et al, 2003). Sequencing analysis of the HPRT gene was performed as follows.

The original analysis of the SRD5A2 gene is described in Makridakis et al, 2004.

The error prone polymerase genes pol β, pol η and pol κ were PCR amplified as described above, except that we used the primers shown on Supplementary Table S1b. Sequencing analysis was performed as follows.

Sequencing—Sequencing reactions consisted of 20–60 ng of the purified PCR product, 3.2 pmol of each PCR primer, 4 µl of 5X sequencing reaction buffer (ABI), 4 µl of ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Mix (Applied Biosystems, Foster City, CA), and sterile water in a total volume of 20 µl. Reactions were submitted to the following thermal cycle: 96°C for 30 s, 50°C for 15 s, and 60°C for 4 min for a total of 30 cycles. The PCR reactions were then purified according to the manufacturer's recommendations and submitted to electrophoresis. Nucleotide sequences were collected on either an ABI PRISM 377 or a 3100 Automated DNA Sequencer (Applied Biosystems, Foster City, CA). Results were processed with the ABI PRISM Sequence Navigator software (Applied Biosystems, Foster City, CA). Nucleotide substitutions were identified and quantitated by ABI PRISM Factura Feature Identification software with “identify heterozygote limit” set at 10% (Applied Biosystems, Foster City, CA). Each nucleotide substitution was confirmed by at least three independent PCR-sequencing analyses.

Nucleotide changes are numbered according to the appropriate genomic reference sequence as noted, according to journal guidelines (www.hgvs.org/mutnomen).

Microsatellite analysis—The complex dinucleotide repeat in intron 3 of the HSD3B2 gene was amplified by PCR using a fluorescent primer. Tumor DNA and its matched normal DNA were amplified by PCR using the set of primers previously described (Verreault et al, 1994). In some cases, the alternative reverse primer 5'-TGGACCTATGTTTGTGTGTGG-3' was utilized, yielding fragments 96 bp shorter than those described by Verreault et al, 1994. The forward primer (Verreault et al, 1994) was labeled on the 5' end with the fluorescent dye TET™. PCR reactions were performed with HotStart Taq as above described and submitted to the following procedure: initial denaturation step for 15 min at 95°C, 50 cycles with a denaturation step at 95°C for 1 min, annealing at 62°C for 1 min, extension at 72°C for 1 min, followed by a final extension step at 72°C for 10 min. PCR products were loaded on an ABI PRISM 377 Automated DNA Sequencer together with an internal size standard (Genescan-500™ TAMRA; Applied Biosystems, Foster City, CA), according to the manufacturer's recommendations. Genotyping analyzes were performed using the Genescan software. Experiments were performed in triplicate.

Statistics

All p-values were calculated using the chi-square test, with the Yates correction when appropriate (for one degree of freedom).

Determination of the actual target of the THEMIS motif

For each gene we analyzed the sequence that contained the somatic mutations (usually the sequence that was PCR amplified, except for AR and TP53, where we used the cDNA sequence with the addition of short patches of intronic sequence, for the mutations that were in intron-exon boundaries). The analysis was performed using a software program that finds known motifs in DNA sequence imported by the user, called “dna-pattern (strings)” (available at: http://rsat.ulb.ac.be/rsat/dna-pattern_form.cgi) We ran this program online using both 0 and 1 allowed substitutions from the motif, and analyzed the results.

RESULTS

Our prostate cancer samples were previously described (Akalu et al, 1999a). Supplementary Table S1a, summarizes patient information including age and tumor stage as well as all results concerning microsatellite instability and nucleotide substitutions in the HSD3B2 gene. These results are presented in detail in the next sections.

Somatic mutations in the HSD3B2 and HPRT genes in prostate cancer tissue

If there is a common somatic mutation motif in prostate cancer tissue, then it should be ubiquitous. We thus decided to screen both genes containing potential ‘driver’ mutations (such as HSD3B2; Makridakis et al, 2005) and passenger mutations (such as HPRT). Initially the tumor DNA of each patient was screened for somatic alterations, by automated DNA sequencing. Once nucleotide alterations were detected, the constitutional DNA of the same patient was also analyzed to examine if the event was somatic. Somatic events were quite common. In fact, amongst the 26 patients studied, only 5 had no detectable nucleotide HSD3B2 alterations in their tumor (Supplementary Table S1a). We detected 38 single nucleotide somatic mutations total in this gene: one deletion and 37 nucleotide substitutions (Supplementary Table S1a). The HPRT gene was also investigated for somatic alterations in prostate cancer by DNA sequencing: 17 somatic nucleotide substitutions were identified (Table 1). Although only a fraction of the HPRT gene was screened, 9 of the 26 patients analyzed harbor somatic nucleotide substitutions (Table 1).

Nucleotide sequence context and nature of the somatic mutations

The high number of somatic mutations identified in prostate cancer tissue allowed us to determine the sequence context of these nucleotide substitutions. This analysis revealed that most mutations occur within a WKVnRRRnVWK (THEMIS) motif when one mismatch is allowed (Table 2; W= A/T, K= G/T, V= G/A/C, R= purine (A/G); n= any nucleotide; total number of n = 0-2 nucleotides; the underline indicates the position of the mutated base).

30 of the 38 (79%) nucleotide alterations detected in the HSD3B2 gene fit the THEMIS motif with up to one mismatch (Supplementary Table S2). The actual target of the motif (the RRR sequences occurring within the WKVnRRRnVWK context with one allowed mismatch) covers 60% of the HSD3B2 sequence analyzed (see *Methods* for calculation). Thus, 22.8 (60%) of the HSD3B2 mutations are expected to occur in the motif (with up to one mismatch), compared to the 30 (79%) observed ($p=0.026$). We then tested the occurrence of the motif in the HPRT and SRD5A2 somatic mutation sites: 15 out of 17 (88%) of the somatic HPRT mutations and 11 out of 16 (69%) of the previously reported (Makridakis et al, 2004) somatic SRD5A2 mutations fit the motif with one allowed mismatch (Table 1 and Supplementary Table S3). The actual motif target (with one allowed mismatch) covers 68% of the HPRT and 57% of the SRD5A2 sequences analyzed (see *Methods*). Thus 20.6 (out of the 33) HPRT/ SRD5A2 mutations are expected to fit the motif with one allowed mismatch, compared to the 26 (15 +11) observed ($p=0.05$). Comprehensive analysis of all the HPRT, HSD3B2 and SRD5A2 somatic mutations indicates that the number of mutations that fit the motif with up to one mismatch (56) is significantly higher ($p=0.005$) than expected (43.4), based on the number of motifs that exist in these genes.

94% of the 54 somatic SRD5A2 and HSD3B2 substitutions are transitions (Supplementary Table S1a and Makridakis et al, 2004). In contrast, only 52 % of the 63 constitutional DNA variations found in these genes in controls and in patients with HSD3B2 deficiency (Makridakis et al, 2004; Pang, 2001; and data not shown) are transitions ($p=0.000003$). Moreover, 88 % of the 17 HPRT somatic substitutions are transitions (Table 1). In contrast, only 51% of the 59 HPRT listed SNPs (<http://snpper.chip.org/bio/export-sequence/20797>) are transitions ($p=0.019$). Thus transitions are significantly more common at the somatic mutation sites.

The inclusion of potential driver mutations (such as the missense SRD5A2 mutations; Makridakis et al, 2004) in our analysis may result in a motif that is biased towards variants that survive the tumor selection process rather than the underlying mutation process. To examine this possibility we analyzed the presumed neutral substitutions (i.e. those that do not change the protein structure) separately: this data suggests that neutral substitutions in all three genes

fit the motif well ($p = 0.005$; Supplementary Table S4a) and that in fact there is no significant difference in the motif fitting ratios among presumed drivers and presumed neutral substitutions ($p = 0.3$; Supplementary Table S4b).

Nucleotide substitution rates vary according to sequence context and clearly depend on the nearest neighbor nucleotides (e.g. Lunter and Hein, 2004). We examined whether an established model that predicts nucleotide substitution rates based on sequence context (the dinucleotide substitution model; Lunter and Hein, 2004) fits the observed mutation spectra better than the THEMIS motif. The data, presented in Supplementary Table S5, suggests that the dinucleotide substitution model predicts a much different distribution than the observed ($p < 0.0001$). Thus the observed somatic mutations in all the genes we examined do not fit the expected distribution based on dinucleotide substitution model, but instead frequently fit the THEMIS motif.

The THEMIS motif in the cancer literature

The discovery of the THEMIS motif prompted us to examine the published literature for other kinds of mutations that may fit this motif. This search revealed that the motif is found (with one allowed mismatch) in 88% (23/ 26) of the *ras1* gene mutations detected in the SENCAR (SENSitive to skin CARcinogenesis) mouse skin cancer model after benzo[α]pyrene (Chakravarti et al, 2000) or estradiol-3, 4-quinone (Chakravarti et al, 2001) treatment (Supplementary Table S6 and data not shown). Benzo[α]pyrene is a known etiologic agent of both skin and lung cancer (Chakravarti et al, 2000). These mouse skin H-ras mutations were the result of error-prone DNA repair (Chakravarti et al, 2000, Chakravarti et al, 2001). 24 of the 26 *ras1* gene mutations were also transitions (Supplementary Table S6 and data not shown). The actual motif target covers 62% of the *ras1* sequence analyzed (see *Methods*). Thus, 16.2 (62%) of the *ras1* mutations are expected to occur in the motif (with one mismatch), compared to the 23 (88%) observed ($p = 0.01$). The frequency of the *ras1* mutations that fit the exact motif (without mismatches) is also higher than expected (35% versus 21%) (Supplementary Figure S1), but this trend is not significant.

Additional literature searches identified the THEMIS motif in 66% (37/ 56) of the somatic prostate cancer mutations that we found in TP53, H-ras, K-ras, N-ras, PTEN and the androgen receptor (AR) gene (Table 2). 68% of these mutations are transitions. Examination of the somatic mutations in the AR gene database (Gottlieb et al, 2004) revealed that 68% of the mutations fit this motif (with up to one mismatch) and 80% are transitions. Moreover, half of the most common somatic mutations (in prostate cancer) that activate the H-ras, K-ras, and N-ras oncogenes, at codons 12, 13 and 61, also fit the THEMIS motif with one mismatch allowed. Regarding somatic substitutions in the TP53 gene in prostate cancer (Chi et al, 1994), there is a prevalence of transitions over transversions and 67% (14/ 21) of the mutations fit this motif with up to one mismatch. The most common somatic PTEN mutation, found in one out of 6 prostate cancer tissues (Pesche et al, 1998), also fits this motif. To test the significance of these observations, we calculated the expected number of the AR gene mutations that fit the THEMIS motif based on the actual motif target (see *Methods*) in the AR gene, the most commonly mutated gene in this dataset: 42 out of the 62 (68%) somatic AR mutations (Gottlieb et al, 2004) fit the motif with one mismatch, compared to 32.8 (53%) expected ($p = 0.027$). The frequency of the somatic AR mutations that fit the THEMIS motif exactly (without mismatches) is also higher than expected (16 observed versus 10.4 expected), but this trend does not reach statistical significance ($p = 0.08$). Moreover, 80% of the somatic prostate cancer mutations in the AR gene (Gottlieb et al, 2004) are transitions, while only 57 % of the constitutional AR SNPs (<http://snpper.chip.org/bio/export-sequence/20479>) are transitions ($p = 0.0054$).

After detecting the THEMIS motif in the most common human malignancy in US males (Jemal et al, 2007), we decided to also examine the most common malignancy in females, breast cancer (Jemal et al, 2007). The data, presented in Supplementary Table S7, shows that 74% (90/ 122) of all somatic breast cancer mutations found in the literature, fit this motif with up to one mismatch. The list of analyzed genes includes CHEK2 (Staalesen et al, 2004; Sullivan et al, 2002), CSNK1E (Fuja et al, 2004), GATA3 (Usary et al, 2004), LMO4 (Sutherland et al, 2003), CDH-1 (Becker et al, 1999; Berx et al, 1996), DLG1 (Fuja et al, 2004), EDD/ hHYD (UBR-5) (Fuja et al, 2004), BRCA2 (Lancaster et al, 1996), PTEN (Kurose et al, 2002), TP53 (Sullivan et al, 2002; Kurose et al, 2002; Thorlacius et al, 1993), and PHB (Sato et al 1992; Sato et al, 1993) (Table 2). To test the significance of this finding, we calculated the expected number of somatic mutations that fit the THEMIS motif in the most commonly mutated gene in this dataset (TP53), based on the actual motif target in TP53: 29 out of the 43 (67%) somatic TP53 mutations fit the motif with one mismatch allowed, compared to 23.5 expected (see *Methods*). This higher than expected incidence does not reach statistical significance ($p=0.089$). However, 10 out of the 43 (23%) somatic TP53 mutations fit the motif exactly, compared to 5.6 (13%) expected, and this higher incidence is significant ($p=0.046$). Moreover, 58% of all somatic breast cancer mutations reported in these genes are transitions (Supplementary Table S7). 55 % of the TP53 SNPs (single nucleotide polymorphisms; <http://snpper.chip.org/bio/export-sequence/7966>) are transitions. The difference in the transition frequency between breast cancer and constitutional TP53 substitutions is not significant.

Recently, a GA pattern (or TC in the opposite DNA strand; the mutated base is underlined) was identified at the sites of somatic mutations in the protein kinase gene family in breast cancer (Stevens et al, 2005). Interestingly this pattern emerged mostly from two breast cancer samples thought to display mutator phenotype (Stevens et al, 2005). Since this pattern is similar to the purine core of the WKVnRRRnVWK motif, we decided to analyze these mutations for the presence of the THEMIS motif: we found that 59 out of the 88 (67%) breast cancer mutations fit the THEMIS motif with one allowed mismatch. In contrast, only 8 of the 71 prostate cancer mutations that we report here occur in the GA motif (8.9/ 71 (i.e. 12.5%) expected; $p=0.75$). The GA motif is overrepresented though in the breast cancer mutation dataset (Supplementary Table S7): 25 (of the 122) mutations fit the GA motif, compared to 15.3 expected ($p=0.01$). Thus both breast and prostate cancer mutations occur in the context of a purine-rich core motif, but in breast cancer this core is often GA.

Somatic mutations in the pol β , pol η , and pol κ genes in prostate cancer tissue

The significant presence of the THEMIS motif at the sites of somatic ras1 mutations induced by error prone DNA repair suggests that such error prone DNA polymerases may be involved in the etiology of (at least some) of the somatic mutations that fit this motif. We accordingly decided to sequence selected exons from each of the EP DNA polymerase genes pol β , pol η and pol κ for somatic mutations in our prostate cancer tissues, to test the hypothesis that prostate cancer tissue bears common somatic mutations in these genes. This preliminary analysis identified somatic mutations in all three genes, but the gene with most (and more prevalent) mutations was pol β . We, therefore, screened the complete coding sequence of the pol β gene in these patients. The result of these analyses in all polymerase genes is shown on Table 3: We identified 27 somatic mutations in these 26 samples, 14 of which were missense substitutions, 9 were silent or intronic substitutions, two substitutions changed splice acceptor (AG) sites, one was in the promoter region and one in the 5'-UTR. The G31438A intronic substitution is recurrent in patients 5 and 30 (Table 3). The P242R missense substitution was also present in the constitutional DNA of patient 13, but with altered prevalence (40 % average mutant peak in the tumor tissue (Table 3) compared to 60 % average mutant peak in the constitutional DNA (data not shown). Thus an additional somatic event may have occurred in the tumor in this pol

β site. Overall, among 26 patients analyzed, 19 patients (73%) have somatic mutations in an EP polymerase gene and 16 patients (61%) have somatic mutations in pol β .

Somatic instability in prostate cancer tissue

Since the HSD3B2 gene was most commonly “hit” in our sequencing analysis of these prostate cancer tissues, we decided to search for other kinds of genomic instability at this locus. Accordingly, we investigated 20 informative patients for microsatellite instability (MSI) and loss of heterozygosity (LOH) in the complex dinucleotide repeat in intron 3 of the HSD3B2 gene (Devgan et al, 1997), by comparing tumor and matched normal DNA. The results show that 70% of the patients analyzed have LOH/ MSI in the HSD3B2 locus (Supplementary Table S1a). The most common findings were MSI manifested as follows: 5 cases with contraction of the alleles of the tumor DNA, 4 cases with expansion, and 3 cases with a combination of both contraction and expansion (data not shown). Three patients had LOH in the tumor tissue (Supplementary Table S1a).

DISCUSSION

Somatic mutations are commonly found in prostate cancer tissue

We report a high number of somatic prostate cancer mutations in all three “target” genes studied. 38 nucleotide alterations were detected in the HSD3B2 gene in 80% of the 26 patients. We also identified 16 somatic SRD5A2 substitutions, in 60% of the patients (Makridakis et al, 2004) and 17 somatic HPRT substitutions, in 35% of the patients (Table 1). Collectively, we found a total of 71 somatic mutations in these three genes in prostate cancer. The high rate of somatic events identified, even in the HPRT gene, suggests that there may be generalized genomic instability, at least in some of these tumors. As summarized in Table 4, four patients harbor somatic substitutions in all three “target” genes examined, and most of these substitutions fit the THEMIS motif. Significantly, all of these patients have additional somatic mutations in an EP polymerase gene (Table 4).

Nucleotide sequence context of the somatic substitutions

Analysis of the nucleotide sequence that surrounds each HSD3B2, SRD5A2 and HPRT gene mutation showed that 79% of the 71 somatic mutation sites fit the THEMIS motif, with up to one mismatch ($p=0.005$; Table 2). The prevalence of this motif at somatic mutation sites led us to hypothesize that it is common, perhaps even universal. Review of the published literature identified this motif (with up to one mismatch) in 66% of all somatic prostate cancer mutations and 74% of all somatic breast cancer mutations ($p=0.027$ for prostate and $p=0.046$ for breast cancer). Moreover, 67% of the somatic breast cancer mutations thought to result through a mutator phenotype in the protein kinase gene family (Stevens et al, 2004), fit the THEMIS motif with one allowed mismatch. Our analyses, therefore, suggest that most somatic mutations in the most common human malignancies fit the THEMIS motif with up to one mismatch.

Analysis of the type of mutations found in the HSD3B2, SRD5A2 and HPRT genes showed that transitions are significantly more common than expected ($p<0.019$). This finding was confirmed among the other somatic prostate cancer mutations that we found in the published literature. However, analysis of the breast cancer mutations obtained from the literature shows that their transition frequency is not significantly higher compared to constitutional substitutions. Thus, although both prostate and breast cancer mutations fit the THEMIS motif, only in prostate cancer there are more transitions than expected by chance. Moreover, the GA motif is overrepresented only among breast cancer mutation sites, not prostate cancer. These data suggest that similar yet distinct molecular etiologies exist between the generation of somatic mutations in the prostate and breast. For example, both prostate and breast cancer

mutations may be caused by error-prone repair, but the polymerase or the carcinogen involved may be different.

The THEMIS motif includes two “spacer segments” that are of variable length, 0-2 nucleotides. The spacer segments are often in the 1-1 permutation (i.e. one nucleotide gap in each side of the mutation) but the data reported here is only statistically significant when the spacer is variable (0-2). In addition, the degenerate nature of the motif inevitably has the effect that the same mutation site often fits more than one permutation (eg. both 0-1 and 1-1). This fact makes it difficult to calculate exact probabilities for each specific permutation. Variable spacers have been reported before at motif sites in nature. For example, the spacer between the -35 and -10 consensus sequences of the Sigma A protein binding site varies between 16-18 nucleotides (Helmman, 1995) while the spacer between the two boxes of the Sigma B binding site is 13-15 nucleotides long (Petersohn et al, 1999). The biological mechanism that allows this spacer variability is unknown, but we speculate that the wide and flexible active site (Perlow-Poehnelt et al, 2004) of the error-prone polymerases may be responsible: it may accommodate the invariable nine nucleotides of the THEMIS motif in various manners, e.g. with 0-1, 1-1 or 1-0 spacers. Y-family polymerases (such as pol η and κ) actually have two partially overlapping active sites (Chandani and Loechler, 2007), a finding that may also contribute to the spacer variability: the two active sites may have different binding preferences.

Error-prone DNA polymerases: a role in the origin of somatic mutations?

The existence of a frequently mutated somatic motif suggests that many mutations in the most common forms of human cancer may have similar molecular etiology. The THEMIS motif is overrepresented among somatic mutation sites induced by error prone repair in the SENCAR mouse skin model following mutagenic treatment (Supplementary Table S5 and data not shown; $p = 0.01$). These data parallel our findings in prostate/ breast cancer tissue. Thus, we propose that many somatic mutations in many types of human cancer (such as breast, prostate, skin, lung) are caused by error prone DNA repair following carcinogenic exposure.

To test this model we decided to first screen the error prone DNA polymerase genes POLB (pol β), POLH (pol η) and POLK (pol κ), for somatic mutations that may have in turn caused the high number of motif mutations in our prostate cancer samples. We report a total of 27 somatic mutations in the three EP polymerase genes in our 26 samples (Table 3; only the pol β gene was fully sequenced). Four somatic pol β mutations have been previously reported in prostate cancer, but that study was conducted with fewer samples (12 cases) of Asian background (Dobashi et al, 1994). Most (52%) of the somatic mutations reported here are missense, while two change AG splice junctions and another mutation changes a lariat-adenine to guanine (Table 3). The two splice junction mutations are predicted to result in deletions of two to three amino acids, at the minimum. Most (59%) of the 27 somatic mutations are prevalent in the tumor, suggesting that they may be “drivers”.

X-ray crystallography studies of human pol β may shed a light on the potential role of the mutated residues: proline-261 forms a hydrogen bond with glutamine-264 (Pelletier et al, 1996), while threonine-292 and methionine-236 are both involved in template binding (Sawaya et al, 1997; Bose-Basu et al, 2004). Disruption of the hydrogen bond between residues 261 and 264 has been proposed to result in the mutator phenotype displayed by the previously reported, prostate cancer-associated I260M pol β mutation (Dalal et al, 2005). Thus the P261L mutation may also disrupt this hydrogen bond and result in prostate tumorigenesis. Moreover, both T292A and M236L substitutions are predicted to destroy the hydrogen bond between the respective residues and the DNA template, affecting DNA synthesis fidelity. Methionine-236 and proline-242 are in the “flexible loop”, a part of pol β that functions to position the primer and has been shown to contain several residues that cause a mutator phenotype when mutated (Dalal et al, 2004). The P242R mutant resulted in similar activity and a four-fold lower mutation

rate than the normal pol β when assayed *in vitro* by a Herpes Simplex Virus thymidine kinase forward mutation assay (Hamid and Eckert, 2005). Thus, the decreased prevalence of the P242R mutant in the tumor compared to the adjacent normal tissue of patient 13 (see *Results* section) may translate into higher mutagenicity in the tumor. Therefore, several of the somatic pol β mutations that we report here may cause a mutator phenotype (Loeb et al, 2003). Future functional studies ought to examine this prediction.

One of the missense pol η mutations is in glycine-263 and both missense pol κ mutations are in threonine-205 (Table 3). The homologous residue of pol κ threonine-205 is pol η threonine-122 (Boudsocq et al, 2002). Missense mutations of both pol η residues 122 and 263 were reported in XPV patients (Broughton et al, 2002). Thus three of the somatic mutations that we identified are in (or are in residues homologous to) pol η residues previously associated with XPV. This finding suggests that these mutations may play a role in carcinogenesis. Yet if some XPV patients are born with a mutation of the same pol η codon that is associated with prostate cancer etiology, then why do these patients get only skin and not prostate cancer? A potential explanation is different environmental exposure: XPV patients are inevitably exposed to sunlight, but not necessarily to prostate cancer inducing mutagens (the prostate is not exposed to sunlight and pol η may be important for repairing other types of damage relevant to the prostate; Table 5). Alternatively, the tumor type specificity may result from the exact nature of the mutation.

Multiple factors determine the mutagenic potential of DNA damage; among them, the choice of the DNA repair machinery evoked to repair the lesions (Pages and Fuchs, 2002) and the nucleotide sequence context wherein a lesion occurs (Beard et al, 2002; Wei et al, 1995). We identified a motif around the sites of human prostate and breast cancer mutations and mouse skin cancer mutations induced by error prone DNA repair. A model for the generation of these mutations (Table 5) may involve the action of environmental mutagens acting in conjunction with (mutant) error-prone polymerase genotypes to cause specific kinds of mutations and/ or at specific sites. A mutant DNA polymerase may result in mutations directly, through decreased fidelity, or indirectly, through the use of a “non-optimal” polymerase for each DNA lesion (as in the XPV paradigm; Pages and Fuchs, 2002). The presence of a motif around the mutation sites can result from either the mutagenic tendencies of specific polymerases or the binding requirements of specific mutagens. The type of error-prone polymerase(s) involved in the etiology of these mutations may be inferred from the known specificities of these enzymes (Table 5), if the type of carcinogenic exposure is known for each patient.

Other mechanisms leading to the generation of the motif mutations are also possible, such as aberrant mismatch repair or base excision repair (BER). In fact, one of the human polymerases that we propose to be involved in this process, pol β (Table 5), is essential for short-patch BER (Sancar et al, 2004). Pol β -deficient mouse fibroblasts do not show the high amounts of A:T to G:C transitions seen in pol β -positive cells following benzo[α]pyrene treatment (Venugopal et al, 2005), suggesting that pol β is responsible for most of these transitions. Moreover, the fact that the THEMIS motif is degenerate may indicate the presence of more than one mechanism (e.g. more than one polymerase or type of damage) with distinct sequence requirements. Alignment of more than one specific sequence motifs may result in a degenerate motif.

In summary, we report here a DNA sequence motif commonly found in prostate cancer mutation sites, termed THEMIS (WKVnRRRnVWK). We extended this finding to human breast and mouse skin cancer and suggest that the THEMIS motif may be the result of error-prone DNA polymerase-induced mutations in these tumors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Dr. Myron Goodman (USC) and Dr. Christine Hackel (São Paulo, Brazil) for their helpful comments, and Dr. Joe Hacia, Dr. Nunzio Bottini, Troy Phipps and Dr. Hooman Allayee (USC) for critically reading this manuscript. This work was supported in part by grants from the NCI P01 CA108964 (project 1) to JKVR and from the ACS-IRG 58-007-45 to NMM. JKVR is a Medical Foundation Fellow at the University of Sydney.

REFERENCES

- Akalu A, Elmajian DE, Highshaw RA, Nichols PW, Reichardt JKV. Somatic mutations at the SRD5A2 locus encoding prostatic steroid 5 α -reductase during prostate cancer progression. *J Urol* 1999a; 161:1355–1358. [PubMed: 10081907]
- Akalu A, Reichardt JKV. A reliable PCR amplification method for microdissected tumor cells obtained from paraffin-embedded tissue. *Genet. Anal* 1999b;15:229–233. [PubMed: 10609759]
- Akiyama N, Alexander D, Aoki Y, Noda M. Characterization of mutations induced by 300 and 320 nm UV radiation in a rat fibroblast cell line. *Mutat. Res* 1996;372:119–131. [PubMed: 9003539]
- Beard WA, Shock DD, Yang XP, DeLauder SF, Wilson SH. Loss of DNA polymerase beta stacking interactions with templating purines, but not pyrimidines, alters catalytic efficiency and fidelity. *J. Biol. Chem* 2002;277:8235–8242. [PubMed: 11756435]
- Becker KF, Reich U, Schott C, Becker I, Berx G, van Roy F, Hofler H. Identification of eleven novel tumor-associated E-cadherin mutations. *Mutations in brief no. 215*. Online. *Hum Mutat* 1999;13:171. [PubMed: 10094558]
- Berx G, Cleton-Jansen AM, Strumane K, de Leeuw WJ, Nollet F, van Roy F, Cornelisse C. E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene* 1996;13:1919–25. [PubMed: 8934538]
- Bose-Basu B, DeRose EF, Kirby TW, Mueller GA, Beard WA, Wilson SH, London RE. Dynamic characterization of a DNA repair enzyme: NMR studies of [methyl-13C]methionine-labeled DNA polymerase beta. *Biochemistry* 2004;43:8911–8922. [PubMed: 15248749]
- Boudsocq F, Ling H, Yang W, Woodgate R. Structure-based interpretation of missense mutations in Y-family DNA polymerases and their implications for polymerase function and lesion bypass. *DNA Repair* 2002;1:343–358. [PubMed: 12509239]
- Broughton BC, Cordonnier A, Kleijer WJ, Jaspers NG, Fawcett H, Raams A, Garritsen VH, Stary A, Avril MF, Boudsocq F, Masutani C, Hanaoka F, Fuchs RP, Sarasin A, Lehmann AR. Molecular analysis of mutations in DNA polymerase eta in xeroderma pigmentosum-variant patients. *Proc. Natl. Acad. Sci. U S A* 2002;99:815–820. [PubMed: 11773631]
- Chakravarti D, Mailander PC, Cavalieri EL, Rogan EG. Evidence that error-prone DNA repair converts dibenzo[a,l]pyrene-induced depurinating lesions into mutations: formation, clonal proliferation and regression of initiated cells carrying H-ras oncogene mutations in early preneoplasia. *Mutat Res* 2000;456:17–32. [PubMed: 11087892]
- Chakravarti D, Mailander PC, Li KM, Higginbotham S, Zhang HL, Gross ML, Meza JL, Cavalieri EL, Rogan EG. Evidence that a burst of DNA depurination in SENCAR mouse skin induces error-prone repair and forms mutations in the H-ras gene. *Oncogene* 2001;20:7945–7953. [PubMed: 11753677]
- Chandani S, Loechler EL. Molecular modeling benzo[a]pyrene N2-dG adducts in the two overlapping active sites of the Y-family DNA polymerase Dpo4. *J Mol Graph Model* 2007;25:658–670. [PubMed: 16782374]
- Cheng, E.; Lee, C.; Aaronson, SA.; Grayhack, J. Endocrinology of the prostate. In: Lepor, H.; Lawson, RE., editors. *Prostate Diseases*. W.B. Saunders Co.; Philadelphia: 1993. p. 57-71.
- Chi SG, deVere White RW, Meyers FJ, Siders DB, Lee F, Gumerlock PH. p53 in prostate cancer: frequent expressed transition mutations. *J Natl Cancer Inst* 1994;86:926–933. [PubMed: 7515114]

- Dalal S, Hile S, Eckert KA, Sun KW, Starcevic D, Sweasy JB. Prostate-Cancer-Associated I260M Variant of DNA Polymerase beta Is a Sequence-Specific Mutator. *Biochemistry* 2005;44:15664–15673. [PubMed: 16313169]
- Dalal S, Kosa JL, Sweasy JB. The D246V mutant of DNA polymerase beta misincorporates nucleotides: evidence for a role for the flexible loop in DNA positioning within the active site. *J. Biol. Chem* 2004;279:577–584. [PubMed: 14563842]
- Devgan SA, Henderson BE, Yu MC, Shi C-Y, Pike MC, Ross RK, Reichardt JKV. Genetic variation of 3 β -hydroxysteroid dehydrogenase type II in three racial/ethnic groups: implications for prostate cancer risk. *Prostate* 1997;33:9–12. [PubMed: 9294620]
- Dobashi Y, Shuin T, Tsuruga H, Uemura H, Torigoe S, Kubota Y. DNA polymerase beta gene mutation in human prostate cancer. *Cancer Res* 1994;54:2827–2829. [PubMed: 8187060]
- Friedberg EC. DNA damage and repair. *Nature* 2003;421:436–40. [PubMed: 12540918]
- Fuja TJ, Lin F, Osann KE, Bryant PJ. Somatic mutations and altered expression of the candidate tumor suppressors CSNK1 epsilon, DLG1, and EDD/hHYD in mammary ductal carcinoma. *Cancer Res* 2004;64:942–51. [PubMed: 14871824]
- Goodman MF. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 2002;71:17–50. [PubMed: 12045089]Review
- Gottlieb B, Beitel LK, Wu JH, Trifiro M. The androgen receptor gene mutations database (ARDB): 2004 update. *Hum Mutat* 2004;23:527–533. [PubMed: 15146455]
- Hamid S, Eckert KA. Chemotherapeutic nucleoside analogues and DNA polymerase β . *Proc. Amer. Assoc. Cancer Res* 2005;46:1369.
- Haracska L, Yu SL, Johnson RE, Prakash L, Prakash S. Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase ϵ . *Nat Genet* 2000;25:458–461. [PubMed: 10932195]
- Helmann JD. Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res* 1995;23:2351–2360. [PubMed: 7630711]
- Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics. *CA Cancer J Clin* 2007;57:43–66. [PubMed: 17237035]
- Johnson RE, Kondratik CM, Prakash S, Prakash L. hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* 1999;285:263–265. [PubMed: 10398605]
- Kunkel TA. Considering the cancer consequences of altered DNA polymerase function. *Cancer Cell* 2003;3:105–110. [PubMed: 12620405]Review
- Kurose K, Gilley K, Matsumoto S, Watson PH, Zhou XP, Eng C. Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. *Nat Genet* 2002;32:355–357. [PubMed: 12379854]
- Labrie F, Simard J, Luu-The V, Belanger A, Pelletier G. Structure, function and tissue-specific gene expression of 3 β -Hydroxysteroid Dehydrogenase/5-ene-4-ene isomerase enzymes in classical and peripheral intracrine steroidogenic tissues. *J Steroid Biochem Mol Biol* 1992;43:805–826.
- Lachance Y, Luu-The V, Verreault H, Dumont M, Rheault E, Leblanc G, Labrie F. Structure of the human type II 3 β -hydroxysteroid dehydrogenase/ Δ^5 - Δ^4 isomerase (3 β -HSD) gene: adrenal and gonadal specificity. *DNA Cell Biol* 1991;10:701–711. [PubMed: 1741954]
- Lancaster JM, Wooster R, Mangion J, Phelan CM, Cochran C, Gumbs C, Seal S, Barfoot R, Collins N, Bignell G, Patel S, Hamoudi R, Larsson C, Wiseman RW, Berchuck A, Iglehart JD, Marks JR, Ashworth A, Stratton MR, Futreal PA. BRCA2 mutations in primary breast and ovarian cancers. *Nat Genet* 1996;13:238–240. [PubMed: 8640235]
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytte A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85. [PubMed: 10891514]
- Liu SX, Cao J, An H. Molecular analysis of Tripterygium hypoglaucum (level) Hutch-induced mutations at the HPRT locus in human promyelocytic leukemia cells by multiplex polymerase chain reaction. *Mutagenesis* 2003;18:77–80. [PubMed: 12473739]
- Loeb LA, Loeb KR, Anderson JP. Multiple mutations and cancer. *Proc Natl Acad Sci USA* 2003;100:776–781. [PubMed: 12552134]

- Lunter G, Hein J. A nucleotide substitution model with nearest-neighbour interactions. *Bioinformatics* 2004;20(Suppl 1):i216–223. [PubMed: 15262802]
- Makridakis NM, Akalu A, Reichardt JKV. Identification and characterization of somatic steroid 5 α -reductase (SRD5A2) mutations in human prostate cancer tissue. *Oncogene* 2004;23:7399–7405. [PubMed: 15326487]
- Makridakis NM, Buchanan G, Tilley W, Reichardt JK. Androgen metabolic genes in prostate cancer predisposition and progression. *Front Biosci* 2005;10:2892–2903. [PubMed: 15970543]
- Masutani C, Kusumoto R, Iwai S, Hanaoka F. Mechanisms of accurate translesion synthesis by human DNA polymerase ϵ . *EMBO J* 2000;19:3100–3109. [PubMed: 10856253]
- Meehan T, Straub K, Calvin M. Benzo[α]pyrene diol epoxide covalently binds to deoxyguanosine and deoxyadenosine in DNA. *Nature* 1977;269:725–727. [PubMed: 304179]
- Ogi T, Shinkai Y, Tanaka K, Ohmori H. Polkappa protects mammalian cells against the lethal and mutagenic effects of benzo[a]pyrene. *Proc. Natl. Acad. Sci. U S A* 2002;99:15548–15553. [PubMed: 12432099]
- Pages V, Fuchs RP. How DNA lesions are turned into mutations within cells? *Oncogene* 2002;21:8957–8966. [PubMed: 12483512]Review
- Pai GS, Sprenkle JA, Do TT, Maren CE, Migeon BR. Localization of loci for hypoxanthine phosphate dehydrogenase and glucose-6-phosphate dehydrogenase and biochemical evidence of nonrandom X chromosome expression from studies of a human X-autosome translocation. *Proc. Nat. Acad. Sci* 1980;77:2810–2813. [PubMed: 6930669]
- Pang S. Congenital adrenal hyperplasia owing to 3 β -hydroxysteroid dehydrogenase deficiency. *Endocrinol. Metab. Clin. North Am* 2001;30:81–99. [PubMed: 11344940]
- Pelletier H, Sawaya MR, Wolfle W, Wilson SH, Kraut J. Crystal structures of human DNA polymerase beta complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. *Biochemistry* 1996;35:12742–12761. [PubMed: 8841118]
- Perlow-Poehnell RA, Likhterov I, Scicchitano DA, Geacintov NE, Broyde S. The spacious active site of a Y-family DNA polymerase facilitates promiscuous nucleotide incorporation opposite a bulky carcinogen-DNA adduct: elucidating the structure-function relationship through experimental and computational approaches. *J Biol Chem* 2004;279:36951–36961. [PubMed: 15210693]
- Pesche S, Latil A, Muzeau F, Cussenot O, Fournier G, Longy M, Eng C, Lidereau R. PTEN/MMAC1/TEP1 involvement in primary prostate cancers. *Oncogene* 1998;16:2879–2883. [PubMed: 9671408]
- Petersohn A, Bernhardt J, Gerth U, Höper D, Koburger T, Völker U, Hecker M. Identification of sigma (B)-dependent genes in *Bacillus subtilis* using a promoter consensus-directed search and oligonucleotide hybridization. *J Bacteriol* 1999;181:5718–5724. [PubMed: 10482513]
- Prakash L, Sherman F. Mutagenic specificity: reversion of iso-1-cytochrome c mutants of yeast. *J. Mol. Biol* 1973;79:65–82. [PubMed: 4355598]
- Rheaume E, Simard J, Morel Y, Mebarki F, Zachmann M, Forest MG, New MI, Labrie F. Congenital adrenal hyperplasia due to point mutations in the type II 3 β -hydroxysteroid dehydrogenase gene. *Nat. Genet* 1992;1:239–245. [PubMed: 1363812]
- Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem* 2004;73:39–85. [PubMed: 15189136]
- Sato T, Saito H, Swensen J, Olifant A, Wood C, Danner D, Sakamoto T, Takita K, Kasumi F, Miki Y. The human prohibitin gene located on chromosome 17q21 is mutated in sporadic breast cancer. *Cancer Res* 1992;52:1643–1646. [PubMed: 1540973]
- Sato T, Sakamoto T, Takita K, Saito H, Okui K, Nakamura Y. The human prohibitin (PHB) gene family and its somatic mutations in human tumors. *Genomics* 1993;17:762–764. [PubMed: 8244394]
- Sawaya MR, Prasad R, Wilson SH, Kraut J, Pelletier H. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry* 1997;36:11205–11215. [PubMed: 9287163]
- Schroder FH, Hermanek P, Denis L, Fair DR, Gospodarowicz MK, Pavone-Maculoso M. The TNM classification of prostate cancer. *Prostate* 1992;4:129–138.
- Simard J, Rhéaume E, Sanchez R, Laflamme N, de Launoit Y, Luu-The V, van Seters AP, Gordon RD, Bettendorf M, Heinrich U, Moshang T, New M, Labrie F. Molecular basis of congenital adrenal

- hyperplasia due to 3 β -hydroxysteroid dehydrogenase deficiency. *Mol. Endocrinol* 1993;7:716–728. [PubMed: 8316254]
- Sobol RW, Wilson SH. Mammalian DNA beta-polymerase in base excision repair of alkylation damage. *Prog. Nucleic Acid Res. Mol. Biol* 2001;68:57–74. [PubMed: 11554313]
- Staalesen V, Falck J, Geisler S, Bartkova J, Borresen-Dale AL, Lukas J, Lillehaug JR, Bartek J, Lonning PE. Alternative splicing and mutation status of CHEK2 in stage III breast cancer. *Oncogene* 2004;23:8535–8544. [PubMed: 15361853]
- Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O'Meara S, Parker A, Tarpey P, Avis T, Barthorpe A, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Petty R, Raine K, Shepherd R, Small A, Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Green A, Knowles M, Leung SY, Looijenga LH, Malkowicz B, Pierotti MA, Teh B, Yuen ST, Nicholson AG, Lakhani S, Easton DF, Weber BL, Stratton MR, Futreal PA, Wooster R. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet* 2005;37:590–592. [PubMed: 15908952]
- Sullivan A, Yuille M, Repellin C, Reddy A, Reelfs O, Bell A, Dunne B, Gusterson BA, Osin P, Farrell PJ, Yulug I, Evans A, Ozcelik T, Gasco M, Crook T. Concomitant inactivation of p53 and Chk2 in breast cancer. *Oncogene* 2002;21:1316–1324. [PubMed: 11857075]
- Sutherland KD, Visvader JE, Choong DY, Sum EY, Lindeman GJ, Campbell IG. Mutational analysis of the LMO4 gene, encoding a BRCA1-interacting protein, in breast carcinomas. *Int J Cancer* 2003;107:155–158. [PubMed: 12925972]
- Thorlacius S, Borresen AL, Eyfjord JE. Somatic p53 mutations in human breast carcinomas in an Icelandic population: a prognostic factor. *Cancer Res* 1993;53:1637–1641. [PubMed: 8453635]
- Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, Langerød A, Kåresen R, Oh DS, Dressler LG, Lønning PE, Strausberg RL, Chanock S, Børresen-Dale AL, Perou CM. Mutation of GATA3 in human breast tumors. *Oncogene* 2004;23:7669–7678. [PubMed: 15361840]
- Venugopal V, Mailander PC, Chakravarti D. The role of DNA polymerase β in inducing cytotoxicity and mutations in dibenzo[*a,l*]pyrene-treated mouse fibroblasts. *Proc. Amer. Assoc. Cancer Res* 2005;46:3059.
- Verreault H, Dufort I, Simard J, Labrie F, Luu-The V. Dinucleotide repeat polymorphisms in the *HSD3B2* gene. *Hum. Mol. Genet* 1994;3:384. [PubMed: 8004119]
- Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinänen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi OP. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 1995;9:401–406. [PubMed: 7795646]
- Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004;10:789–799. [PubMed: 15286780]
- Wei D, Maher VM, McCormick JJ. Site-specific rates of excision repair of benzo[*a*]pyrene diol epoxide adducts in the hypoxanthine phosphoribosyltransferase gene of human fibroblasts: correlation with mutation spectra. *Proc. Natl. Acad. Sci. U S A* 1995;92:2204–2208. [PubMed: 7892248]
- Wilson TE, Lieber MR. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway. *J Biol Chem* 1999;274:23599–609. [PubMed: 10438542]
- Yu SL, Johnson RE, Prakash S, Prakash L. Requirement of DNA polymerase eta for error-free bypass of UV-induced CC and TC photoproducts. *Mol. Cell Biol* 2001;21:185–188. [PubMed: 11113193]
- Zoref E, Sperling O. Increased de novo purine synthesis in cultured skin fibroblasts from heterozygotes for the Lesch-Nyhan syndrome: a sensitive marker for carrier detection. *Hum. Hered* 1979;29:64–68. [PubMed: 761926]

Table 1
Somatic nucleotide substitutions in the HPRT gene.

Patient	DNA change	Type of change/ location	Mutant nucleotide peak (%)		Sequence context
			Forward sequence	Reverse sequence	
1	g..39835C>T*	P169S	15	25	<u>gTG GGG <u>TC</u> CTT</u>
	g..39897C>T	Intronic	22	15	AGA <u>T GGt TA</u> AAT
	g..39985T>C	Intronic	34	25	AGA <u>TTTAA</u> AAG
	g..40051C>T*	P184S	19	34	TGt <u>CT</u> GGA ATT
4	g..39835C>T*	P169S	19	29	<u>gTG GGG <u>TC</u> CTT</u>
	g..40051C>T*	P184S	23	34	TGt <u>CT</u> GGA ATT
6	g..39835C>T*	P169S	14	24	<u>gTG GGG <u>TC</u> CTT</u>
	g..40051C>T*	P184S	21	32	TGt <u>CT</u> GGA ATT
11	g..40055A>G	D185G	28	35	<u>TTC C AGA C</u> AAG
	g..40073A>G	Y191C	28	20	gGA <u>T AIG CC</u> CTT
	g..40091A>G	E197G	29	22	<u>ATG AAt A</u> CTT
12	g..39959A>G	Intronic	32	38	<u>TTG T</u> AGA GAG
	g..39890A>G*	Intronic	33	32	<u>TTA A</u> AIG ATT
20	g..40086T>C	Y195Y	44	41	TGA <u>CT</u> AIA ATG
	g..39929G>A	Intronic	18	19	<u>TGA AA</u> tGG CTT
21	g..40000C>T*	Intronic	38	42	<u>ATA AAG</u> AAa
	g..39859G>A	D177N	81	69	<u>AGC C</u> AGA CTG
27	g..39939G>T	Intronic	50	66	<u>ATA A</u> tGG CTT
	g..40011T>C	Intronic	37	30	<u>ATC TA</u> AAt GAT
	g..39907C>T*	Intronic	65	89	<u>TTA Ggt TA</u> AAG
	g..39971G>T*	Intronic	23	33	<u>TGG eAA</u> ATG

The HPRT nucleotide changes are numbered according to the genomic reference sequence (Genbank: M26434.1). In the *Sequence context* column, the nucleotide sequence is displayed using the DNA strand that rendered the best fit to the THEMIS motif. Asterisks indicate mutations that fit the motif only in the non-transcribed strand (the rest of the mutations fit the motif in the transcribed strand). The mutated base is indicated in bold. Lower-case letters indicate mismatches within the consensus motif. The spacer nucleotides are underlined. Sequence contexts that conform to the motif are shaded. The mutant peak intensity is reported as percentage of total peak intensity. In the *Type of change* column, missense mutations are indicated using the single letter code.

Table 2

Most somatic mutations in cancer fit the THEMIS motif.

BREAST		PROSTATE	
GENE	% FIT	GENE	% FIT
CHEK2	90 %	H-RAS, K-RAS, N-RAS	50 %
CSNK1E	82 %	TP53	67 %
GATA3	67 %	PTEN	100 %
LMO4	100 %	AR	68 %
CDH-1	84 %	HSD3B2	79 %
DLG1	67 %	SRD5A2	69 %
UBR-5	80 %	HPRT	88 %
BRCA2	100 %		
PHB	60 %		
PTEN	64 %		
TP53	67 %		

Notes: The % of somatic mutations that fit the WKVnRRRnVWK motif (with one mismatch allowed) is indicated by gene and type of cancer. Bold format shows data reported in this manuscript. The rest of the data was analyzed from the published literature. For details and references see text.

Table 3
Somatic mutations detected in polymerase genes in prostate cancer patients.

Gene (accession #)	Patient	DNA change	Type of Change/ location	Mutant peak (%)	Predicted/ known effect
POLB (AF491812.1)	1	g.23912G>A	AG-splice junction	59 %	Deletion of amino acids 184-185
	2	g.25254G>A	E216K	52 %	
	4	g.960C>T	5'-UTR	52 %	
		g.18145T>C	N128N	54%	
	5	g.32481C>T	P261L	100 %	Altered fidelity
		g.32573A>G	T292A	100 %	Altered fidelity
		g.32592T>C	I298T	100 %	
		g.32439G>A*	Intron 12	79 %	
	8	g.11628T>C	Intron 3	27 %	
	10	g.25314A>T	M236L	38 %	Altered fidelity
	11	g.15180G>A	E123K	100%	
	12	g.1444G>C	K27N	60 %	
	13	g.25236C>T	L210L	42 %	
		g.31911C>G**	P242R	40 %	Altered fidelity
	15	g.921C>T	Promoter	100 %	
	20	g.25302G>A	E232K	100 %	
	23	g.12622A>G	AG-splice junction	25 %	Deletion of amino acids 88-90
	24	g.11630T>C	Intron 3	30 %	
	29	g.32521G>T	G274G	35 %	
	30	g.32467T>C	Intron 12	51 %	
POLH (AY388614.1)		g.32439G>A*	Intron 12	100 %	
	1	g.28891G>A	G259R	23 %	
	6	g.29021A>T	Intron 7	46 %	
	20	g.28901G>C	G263A	60 %	XPV
POLK (AY273797.1)		g.28967C>T	S284F	66 %	
	12	g.67088C>A	T205K	28 %	

Gene (accession #)	Patient	DNA change	Type of Change/ location	Mutant peak (%)	Predicted/known effect
	21	g.66992A>G	Intron 5	68 %	Changes Lariat-A
	27	g.67088C>T	T205I	30 %	

Notes: For both mutations that change the invariant splice junction (AG), an in-frame AG exists shortly downstream. Utilization of the alternative AG is predicted to result in the deletions shown in the last column. XPV denotes a pol η residue mutated in an XPV patient. For further details, see text.

* Creates a new branch (Lariat-A) site

** Also in constitutional DNA (rs3136797)

Table 4

Patients with somatic nucleotide substitutions in all “target” genes analyzed (HSD3B2, SRD5A2 and HPRT genes).

Patient	Gene	DNA change	Sequence context
1	HPRT	g.39835C>T*	<u>g</u> TG GGG <u>TC</u> CTT
		g.39897C>T	AGA <u>T</u> GGt <u>TA</u> AAT
		g.39985T>C	AGA <u>TT</u> tAA AAG
		g.40051C>T*	TGt <u>CT</u> GGA ATT
	SRD5A2	g.888G>A	TGC <u>C</u> AGc CcG
		g.1890G>A	AcA AGG <u>TGG</u> CTT
	HSD3B2	g.8774C>T*	TGG <u>GT</u> GGA GTT
	POLB	g.23912G>A	ATC <u>A</u> cAG GTG
	POLH	g.28891G>A	GTC <u>TT</u> GGA <u>G</u> GAA
6	HPRT	g.39835C>T*	<u>g</u> TG GGG <u>TC</u> CTT
		g.40051C>T*	TGt <u>CT</u> GGA ATT
	SRD5A2	g.888G>A	TGC <u>C</u> AGc CcG
		g.1294T>C*	AGA AGG CAG
	HSD3B2	g.1551A>G	AGC AGG AgG
		g.1571G>A	TcA GAG GAT
		g.1622T>C	TcC AAG <u>GCC</u> CTG
		g.1671A>G	AGt AAA CTT
	POLH	g.29021A>T	TTT <u>TT</u> AAA ATC
11	HPRT	g.40055A>G	TTC <u>C</u> AGA <u>C</u> AAG
		g.40073A>G	gGA <u>T</u> AtG <u>CC</u> CTT
		g.40091A>G	ATG AAt <u>A</u> CTT
	SRD5A2	g.2019T>C	cGC AGc <u>CC</u> AAG
	HSD3B2	g.8089G>A	AGA AGG CTG
		g.8174G>A	TGG GGA AG GAG
	POLB	g.15180G>A	cTA GAA <u>G</u> GTG
21	HPRT	g.39859G>A	AGC <u>C</u> AGA CTG
		g.39939T>G	AGC AAt <u>TAT</u> AAG
		g.40011T>C	ATC <u>TA</u> AAt GAT
	SRD5A2	g.888G>A	TGC <u>C</u> AGc CcG
		g.1914G>A	cTG GAG <u>CC</u> AAT
		g.1927C>T*	AcC GAG <u>GA</u> AAT
	HSD3B2	g.8006A>G	AGG AAA <u>T</u> CAT
		g.8577T>C*	AGG <u>T</u> GAA CAc
	POLK	g.66992A>G	ATT <u>T</u> AAA CTT

Refer to Table 1 for legends. Polymerase mutations are shown in different font. Sequence contexts that conform to the THEMIS motif are shaded. Reference Genbank sequences: HPRT: M26434.1, SRD5A2: L03843.1, HSD3B2: M77144.1, POLB: AF491812.1, POLH: AY388614.1, POLK: AY273797.1.

Table 5

Models for the potential etiology of the observed somatic mutation motifs.

DNA sequence motif/ Type of mutation		Potential mutagen involved in DNA damage	Potential polymerase involved in DNA damage repair and/ or mutation generation
WKV <u>RRR</u> VWK	RR on the strand that fits the motif	PAH ¹ Cisplatin ⁴ AAF-dG 8-oxo-dG	* pol β ² / pol κ ³ pol η ⁴ pol η ⁵
	YY on the other strand	UV radiation ⁶	pol η ⁷
Transition		Alkylating agent ⁸	EP polymerase (e.g. pol β ⁹)

Notes: These models involve the action of a specific environmental mutagen acting in conjunction with an error prone polymerase to cause specific kinds of mutations (or at specific sites). Pol β is not a “classic” error prone polymerase, yet it causes 67-times more substitution errors than mammalian pol δ (Kunkel, 2003). The models presented in this table are not mutually exclusive (e.g. a transition may also occur in the WKV**RRR**VWK motif). Bold letters indicate the potential sites of the altered nucleotide. For further details see text. Key: PAH= polyaromatic hydrocarbons (e.g. benzo[*a*]pyrene), AAF-dG= acetylaminofluorene-deoxyguanosine, 8-oxo-dG= 8-oxo-deoxyguanosine, EP= error prone, UV= ultraviolet. 1: Meehan et al, 1977; 2: Venugopal et al, 2005; 3: Ogi et al, 2002; 4: Masutani et al, 2000; 5: Haracska et al, 2000; 6: Akiyama et al, 1996; 7: Yu et al, 2001; 8: Prakash and Sherman, 1973; 9: Sobol and Wilson, 2001.

* The choice of the polymerase involved may depend on the type of the PAH adduct that is present (depurinating or stable).