

Frequency of Germline Genomic Homozygosity Associated With Cancer Cases

Guillaume Assié, MD, PhD

Thomas LaFramboise, PhD

Petra Platzer, PhD

Charis Eng, MD, PhD

KNUDSON FIRST INTRODUCED the concept of both a germline predisposition and a somatic event for the inactivation of tumor suppressors in cancer with the “2 hit hypothesis.”¹ This theory applies to heritable cancer syndromes caused by mutations in tumor suppressors with high penetrance. The first hit is an inherited mutation (mutant allele) in the germline while the most simple second hit is somatic loss of the wild-type (normal) allele, leaving a single copy of the mutant allele (FIGURE 1). More commonly, the second hit is a combination of multiple and complex somatic events.² In contrast, most sporadic cancers are complex traits wherein the germline predisposition is likely related to several low-penetrance genomic variants, present in either oncogenes or tumor suppressor genes.

While studying loss of heterozygosity/allelic imbalance (LOH/AI) in cancer tissues using microsatellite markers in germline and in corresponding tumors,³⁻⁷ we anecdotally observed a low frequency of germline heterozygosity in cancer patients compared with controls (C. E., unpublished data, 1993-2006), raising the question whether homozygosity could play a role in cancer predisposition. Further evidence suggesting a role of homozygosity in cancer is brought by studies that showed an increased risk of cancer in inbred populations,⁸⁻¹² by several reports iden-

Context Cancer is a multigenic disease resulting from both germline susceptibility and somatic events. While studying loss of heterozygosity (LOH) in cancer tissues, we anecdotally observed a low frequency of heterozygosity in cancer patients compared with controls, raising the question whether homozygosity could play a role in cancer predisposition.

Objectives To determine the frequency of germline homozygosity in a large series of patients with 3 different types of solid tumors compared with population-based controls.

Design, Setting, and Patients Germline and corresponding tumor DNA isolated from 385 patients with carcinomas (147 breast, 116 prostate, and 122 head and neck carcinomas) were subjected to whole genome (345-microsatellite marker) LOH analysis.

Main Outcome Measures Frequency of homozygosity at microsatellite markers in cancer cases vs controls and frequency of somatic LOH in cancers at loci with the highest homozygosity.

Results We identified 16 loci in common among the 3 cancer types, with significantly increased germline homozygosity frequencies in the cancer patients compared with controls ($P < .001$). In the cases who happened to be germline heterozygous at these 16 loci, we found a mean (SD) LOH frequency of 58% (4.2%) compared with 50% (7.5%) at 197 markers without increased germline homozygosity ($P < .001$). Across the genome, this relationship holds as well ($r = 0.46$; 95% confidence interval, 0.37-0.53; $P < .001$). We validated the association of specific loci with high germline homozygosity frequencies in an independent, single-nucleotide polymorphism-based, public data set of 205 lung carcinomas from white individuals ($P < .05$ to $P < .001$) as well as the correlation between genome-wide germline homozygosity and LOH frequencies ($r = 0.21$; 95% confidence interval, 0.18-0.24; $P < .001$).

Conclusions In our study of 4 different types of solid tumors (our data for 3 types validated in a fourth type), increased germline homozygosity occurred at specific loci. When the germline was heterozygous at these loci, high frequencies of LOH/allelic imbalance occurred at these loci in the corresponding carcinomas.

JAMA. 2008;299(12):1437-1445

www.jama.com

tifying homozygous loci associated with cancer,^{12,13} and from experimental animal inbreeding (eg, backcrossing mice) increasing tumor incidence. Specific situations of homozygosity have also been directly associated with cancer, such as uniparental isodisomy through altered imprinting.¹⁴

Homozygosity is common in humans and extended homozygote tracts have been described in several studies.¹⁵⁻¹⁸ Cancer susceptibility genes are also numerous in the genome. These

facts together increase the likelihood that homozygosity might occur in the loci of cancer susceptibility genes. One

Author Affiliations: Genomic Medicine Institute (Drs Assié, LaFramboise, Platzer, and Eng), Lerner Research Institute (Drs Assié, Platzer, and Eng), and Taussig Cancer Institute (Dr Eng), Cleveland Clinic Foundation, Cleveland, Ohio; Department of Genetics (Drs LaFramboise and Eng) and Case Comprehensive Cancer Center (Dr Eng), Case Western Reserve University School of Medicine, Cleveland.

Corresponding Author: Charis Eng, MD, PhD, Genomic Medicine Institute, Cleveland Clinic Foundation, 9500 Euclid Ave, NE-50, Cleveland, OH 44195 (engc@ccf.org).

can then hypothesize that germline homozygosity at these loci may somehow contribute to cancer predisposition. Indeed, in the case of homozygous tumor suppressor genes, recessive mutant alleles would result in an altered tumor suppressor function with no need of any additional somatic event. In the case of dominant alleles in an oncogene, homozygosity could result in a superoncogenic effect.

Therefore, we hypothesized that germline homozygosity lends low-penetrance susceptibility to cancer by the potential mechanism of affecting the function of numerous genes throughout the genome. We systematically addressed this hypothesis by looking at series of patients with 3 different types of solid tumors for genomic regions with increased germline homozygosity compared with controls and by ex-

ploring somatic LOH in the cancers of patients who happened to be heterozygous at these loci.

METHODS

Samples and Genomic DNA Preparation

Germline and tumor (somatic) genomic DNA was isolated using standard protocols from 385 patients with 3 types of invasive carcinomas: 147 breast, 116 prostate, and 122 head and neck squamous cell carcinomas. Patients with widely metastatic disease (TXNXM1) were excluded from the study. Tumor DNA was extracted specifically from the carcinomatous epithelium using laser capture microdissection, as previously described.^{3,4} In addition, the corresponding germline DNA for each patient was procured either from peripheral blood (n=32) or from normal tissue using a dif-

ferent tissue block containing only normal tissue (n=353). The use of normal tissue as a reliable source of germline DNA was validated previously.^{3,4,6,19}

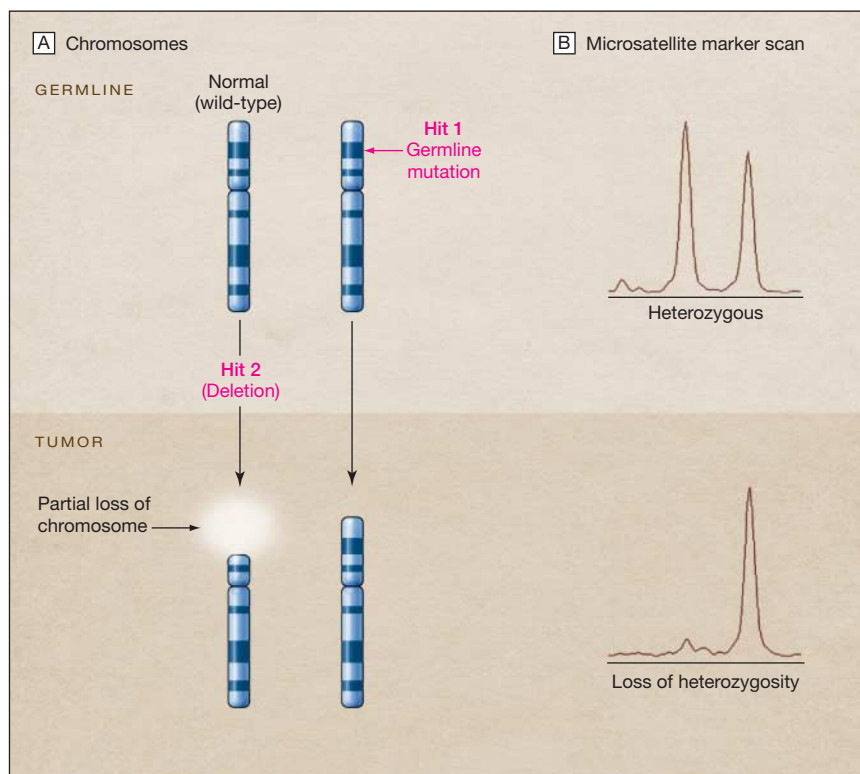
The patients were of northern and predominantly western European ancestry by self-report. Hence, controls were chosen for matched ancestry, which is important for polymorphic marker frequency analysis. The head and neck squamous cell carcinoma and breast cancer series have been previously described in detail.^{3,4,6,19} The study, which used anonymized unlinked samples, was approved under exempt status by the Cleveland Clinic institutional review board for human subjects protection.

Genome-wide Microsatellite Analysis

Polymerase chain reaction was performed as previously described.^{3,4} Briefly, DNA from germline and tumor was amplified using 1 of 72 multiplex primer panels of fluorescent-labeled microsatellite markers (MapPairs genome-wide Human Markers set versions 9 and 10; Invitrogen, Carlsbad, California). A total of 345 autosomal markers were common to all samples and were used for the analysis. Genotyping was performed with the ABI3700 or 3730 xl semiautomated sequencer (Applied Biosystems, Perkin-Elmer Corp, Norwalk, Connecticut). The results were analyzed by automated fluorescence detection using GeneScan collection and analysis software (Applied Biosystems).

Each marker was analyzed in genomic DNA from the germline and tumor from each patient. The germline could be scored as homozygous (single peak in the normal tissue) (FIGURE 2), heterozygous (2 peaks) (Figure 2), or failure (failed polymerase chain reaction). For each heterozygous marker in the germline, somatic LOH/AI in the corresponding tumor (Figure 2) was scored as present when the ratio of peak heights of alleles between germline and tumor DNA was more than 1.5 or less than 0.66, and retention of heterozygosity (Figure 2) was noted when the

Figure 1. Schematic Representation of the Knudson 2-Hit Hypothesis



A germline mutation in a tumor suppressor gene is inherited (first hit) on the right chromosome (A). The paired chromosome is wild-type (normal). A marker near the mutation spot that is heterozygous in the germline is shown at the top (B). When there is a partial loss of the wild-type chromosome during tumorigenesis, the loss-of-heterozygosity (LOH) scan shows only 1 remaining peak.

ratio of peak heights was less than 1.5. A reaction failure was noted when the amplification reaction had failed. In most cases, the ratio of ratios was 1.8 to 2.2 or 0.5 for LOH/AI.

Statistical Analysis of Microsatellite Genotyping

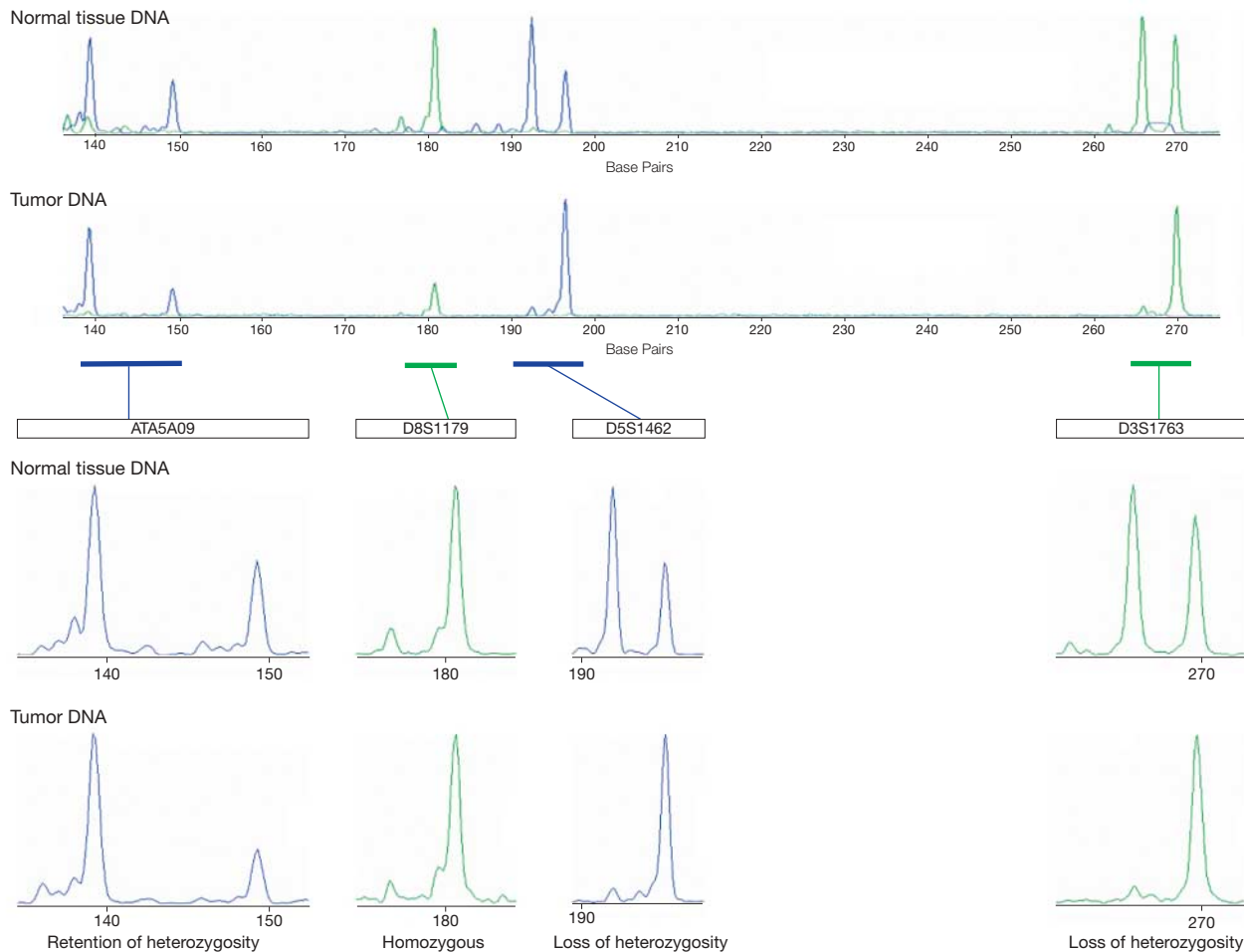
All statistical analyses were performed using R version 2.4.²⁰ A total of 345 autosomal microsatellite markers were analyzed in germline and in tumor DNA samples from 385 patients. For each marker, the frequency of germline homozygosity was deter-

mined as the number of homozygous markers divided by the sum of heterozygous plus homozygous markers. In addition, for each marker, the frequency of LOH/AI was determined as the number of LOH/AI events divided by the sum of the LOH/AI and retention-of-heterozygosity events. For each marker, a median of 105 samples failed either in germline or in tumor because of polymerase chain reaction failures.

We observed that missing information was more abundant for markers with decreased frequency of germline

homozygosity and with increased frequency of LOH in tumor. In statistical terminology, this is known as *nonignorable missingness* and therefore must be addressed. Toward this end, we designed a statistical procedure to estimate the heterozygosity and LOH frequencies in the observed and unobserved (due to genotype failure) data combined. That is, we estimated the true frequencies for all microsatellites as follows. First, we made initial rough estimates of genotype failure rates separately for homozygote, heterozygote, lost (LOH), and retained loci. Such

Figure 2. Loss-of-Heterozygosity Analysis in Tumors



Four examples from genotyping output from 2 patient samples. Each of the top or bottom panel shows the genotyping readout from germline DNA (top) and somatic tumor DNA (bottom). The bottom vertical pairs are enlargements of the top panel genotyping readout, which show microsatellite markers that have retained heterozygosity (ATA5A09, bottom left) and that have undergone loss of heterozygosity (D5S1462 and D3S1763, bottom-right 2 panels). Marker D8S1179 is homozygous in the germline. Note that because the tumor cells were procured by laser capture microdissection, virtually all scoring of loss of heterozygosity and retention of heterozygosity is very clear.

estimates automatically imply, via standard conditional probability, heterozygosity and LOH frequencies for each microsatellite. These individual microsatellite estimates in turn can be used to provide more accurate estimates of the genotype failure rates. This procedure is iterated repeatedly until all estimates converge to optimal solutions (mathematical details available at <http://www.lerner.ccf.org/gmi/igac>). In cancer type-specific investigations such as this, and to minimize ascertainment bias, germline homozygosity frequencies were not determined for markers with fewer than 5 informative patients (2 markers in patients with prostate cancer). Similarly, we did not calculate LOH/AI frequencies for any markers with fewer than 5 informative tumors (1 marker in head and neck squamous cell carcinomas and 4 markers in the prostate cancers, including the 2 germline failures). We chose 5 as a reasonable cutoff under which computing a frequency is not reliable from our experience.

For each marker, the frequency of germline homozygosity was compared with the binomial distribution using a reference germline homozygosity frequency as the null homozygosity probability. The number of homozygotes for a marker was determined as the product of the corrected germline homozygosity frequency for this marker and the number of informative patients. The reference frequencies for a western European-ancestry population were provided for each marker by the Cooperative Human Linkage Center (Marshfield, Wisconsin). These are ancestry-matched controls because our cases are white individuals of western European ancestry. Similarly, LOH/AI frequencies were compared for each marker with the binomial distribution using the average LOH/AI frequencies of the markers on the same chromosome as the null LOH/AI probability. The number of LOH/AI events for a marker was determined as detailed earlier. For testing for both germline homozygosity frequencies and LOH/AI frequencies, a 1-sided exact binomial

test was used (R function `pbinom`, `lower.tail=FALSE`), and *P* values were adjusted for multiple-testing false discovery rate over the 345 markers (R function `p.adjust`, `method="FDR"`). Adjusted *P* values less than .05 were considered significant. The overall relationship between germline homozygosity frequency and somatic LOH/AI frequency across all markers was measured by the Pearson correlation test (R function `cor.test`).

Validation With Single-Nucleotide Polymorphism Microarray Data

A publicly available data set of paired germline and somatic DNA from 205 non-small cell lung carcinomas, which had been subjected to genome-wide single-nucleotide polymorphism (SNP) genotyping with the 250K SNP Chip (Affymetrix, Santa Clara, California),²¹ was used to validate our microsatellite data. All 205 samples came from individuals of European ancestry, confirmed with their SNP genotypes in comparison with the HapMap 3 populations using the program Structure (<http://pritch.bsd.uchicago.edu/software.html>). Both cases and controls were of similar SNP-based ancestry. For each patient, both a tumor sample (fresh/flash frozen) and a normal tissue sample (peripheral blood leukocytes and fresh/flash frozen normal lung tissue) were available. This data set did not have nonignorable missingness. Loss of heterozygosity was identified using dChipSNP.²² dChipSNP was applied to the tumor and germline genotypes using the paired analysis setting. All other parameters were set to default values.

For each SNP, the frequency of homozygosity in the germline was obtained by dividing the number of homozygous genotypes, AA and BB, by the total number of genotypes, AA, AB, and BB (ie, homozygous and heterozygous genotypes), in the germline. The frequency of homozygosity was considered for the SNPs within 100 kilobase (kb) of the 12 microsatellite markers shown to have increased homozygosity in cancer patients (the remaining 4 regions could not be assessed because

the relationship of SNPs to these markers remain imprecisely mapped). The frequency of homozygosity in the germline of patients with lung cancer was compared with that of HapMap controls of European ancestry using the distribution function of the binomial distribution, considering 1-sided *P* values (R function `pbinom`). These exact binomial test *P* values were adjusted by false discovery rate (R function `FDR`).

For each SNP, the frequency of LOH in the lung cancer samples was determined as the proportion of SNPs with somatic LOH. Correlations between the frequency of homozygosity in the germline and the frequency of LOH in tumors at each SNP were performed using the Pearson correlation test (`cor.test` in R). For that correlation, frequency of germline homozygosity was determined among SNPs that did not show LOH in tumors, and the frequency of LOH in tumors was determined among SNPs that were germline heterozygous. To take into consideration the low frequency of LOH when using SNPs (mean, 1.7%/SNP), frequencies were computed only for SNPs with at least 100 informative samples.

RESULTS

We explored germline homozygosity by using 345 autosomal microsatellite markers in 385 patients with 3 different types of invasive carcinomas. These markers were selected for their high heterozygosity index and their regular spacing throughout the genome.^{23,24} In the first instance, a frequency of germline homozygosity was determined at each marker (or locus) in the overall (merged) set comprising all 3 groups of patients who had breast, prostate, or head and neck squamous cell carcinomas. This yielded 114 loci with increased frequencies of germline homozygosity compared with those of ancestry-matched controls (adjusted *P* values <.05 to <.001).

We then looked at each group of patients with each cancer type (FIGURE 3). We identified 83 loci in head and neck squamous cell carcinoma, 56 loci in

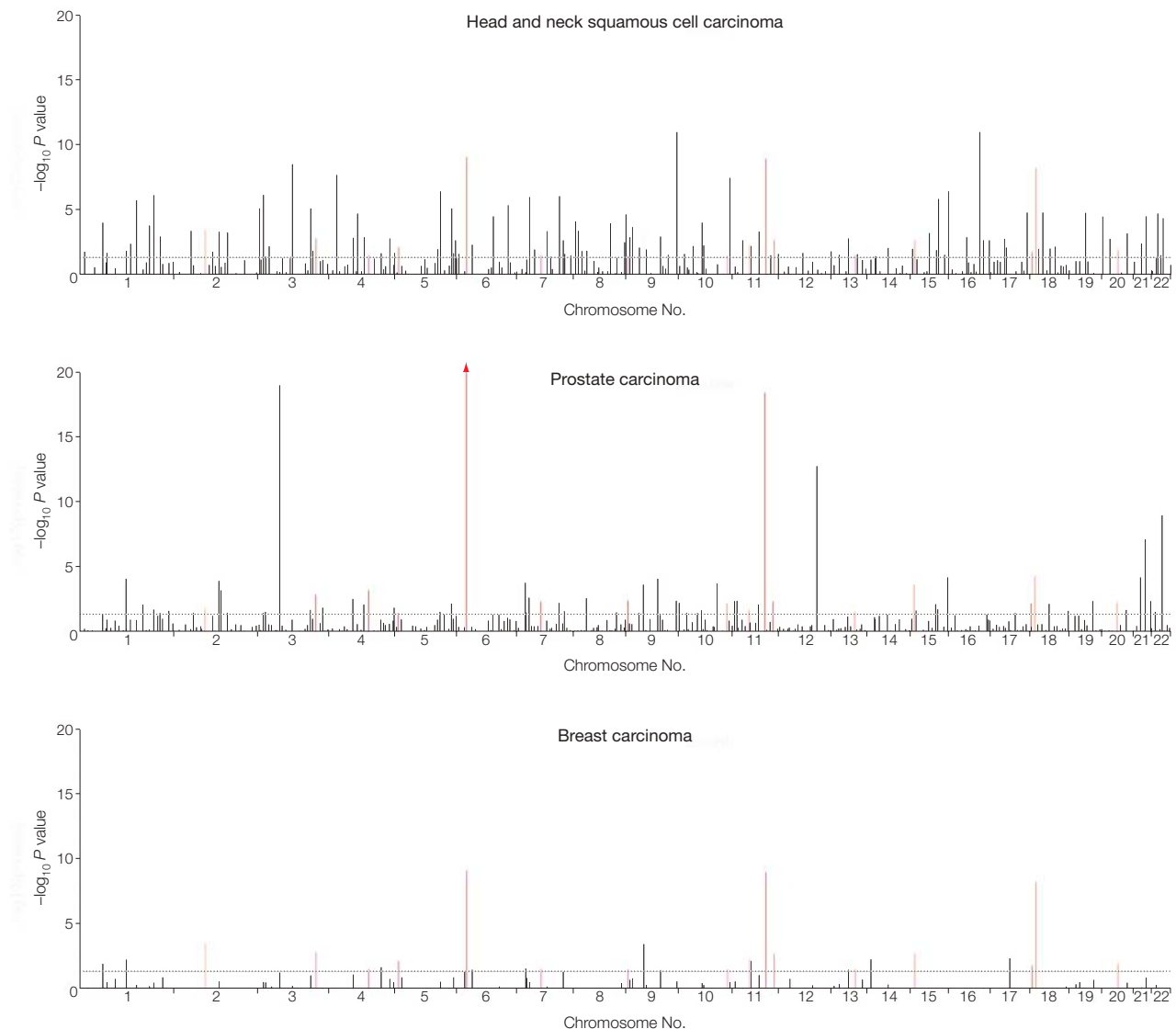
prostate cancer cases, and 26 loci in breast cancer cases with statistically significantly higher frequencies of germline homozygosity compared with those of ancestry-matched controls in each group of patients (adjusted P values $<.05$ to $<.001$) (Figure 3).

Among the 345 loci studied, 120 loci had higher-than-predicted frequencies of homozygosity in patients with head and neck squamous cell

carcinoma, 69 loci in patients with prostate cancer, and 27 loci in patients with breast cancer (Figure 3). Sixteen of these loci were in common among the 3 groups of patients with the 3 cancer types (TABLE 1). Copy number variations are numerous throughout the genome²⁵ and can potentially generate chromosomal regions with a single copy of DNA (hemizygous). To explore whether

these 16 loci could be hemizygous due to copy number variations, we screened the Database for Genomic Variants (<http://projects.tcag.ca/variation/project.html>) to identify whether any of these 16 loci identified to have high germline homozygosity frequencies lie within known regions with copy number variations. Only 1 of the 16 loci, D11S1993, is included in such a copy number variation region. As a compari-

Figure 3. Mapping Germline Homozygosity Frequencies in Cancer Patients



For each microsatellite marker, the germline homozygosity frequency in 3 cancer types (head and neck squamous cell carcinoma, prostate carcinoma, and breast carcinoma) was compared with the reference homozygosity frequency. A binomial test was used and P values were adjusted by false discovery rate. The $-\log_{10} P$ value of the comparison is plotted vs the physical position, with the 22 autosomes represented. The dashed line represents the .05 significance limit. In red are loci with increased homozygosity common to all 3 cancer types.

son, among the entire genome-wide set of 345 markers used in this study, 20 are in copy number variation regions, which is proportionally similar (binomial test; $P = .70$). In summary, these data show that germline homozygosity at specific markers is more frequent in cancer patients.

We then analyzed the frequency of somatic events occurring at these 16 loci by looking at the frequency of somatic LOH/AI in heterozygote patients. The mean (SD) LOH/AI frequency at the 16 loci in the 3 cancer types was 58% (4.2%) (range, 48%-65%). As a comparison, we selected 197 microsatel-

lite markers that were not found to have an increased frequency of germline homozygosity in patients with any of the 3 cancer types compared with that of controls. The mean (SD) LOH/AI frequencies at these 197 loci was 50% (7.5%) (range, 34%-70%), which is statistically significantly lower than the mean 58% LOH/AI frequency at the 16 loci ($P < .001$).

Thus far, we have shown that if cancer patients are heterozygous at these 16 loci that have an increased germline homozygosity, then their corresponding tumors have a higher frequency of somatic LOH/AI at these same 16 loci. In light of this observation, we wanted to explore whether this phenomenon could be generalizable. The frequencies of homozygosity or heterozygosity among all markers is linear (and not binary). We hypothesized that if our first observation with the 16 loci was generalizable, then we would also observe that as loci have increasing frequencies of germline homozygosity, those same loci would also have increasing somatic LOH/AI frequencies in the tumors of cancer patients who happened to be heterozygous at these loci. Indeed, we found that this observation could be extended to all 345 microsatellite markers, where we identified a correlation between the frequencies of germline homozygosity and the frequencies of somatic LOH/AI in patients who happened to be heterozygous at the same loci ($r = 0.46$, 95% confidence interval, 0.37-0.53, $P < .001$) (FIGURE 4).

Thus far, we have shown that for loci with increased germline homozygosity among all cancer patients, these same loci that happened to be germline heterozygous in some cancer patients also have an increased LOH/AI frequency in their corresponding tumors (Figure 4). Furthermore, we have shown that this observation can be generalizable; ie, the more frequent the homozygosity at any given locus, the more LOH/AI can be found at those loci in the tumors of cancer patients that happened to be heterozygous in the germline (Figure 4).

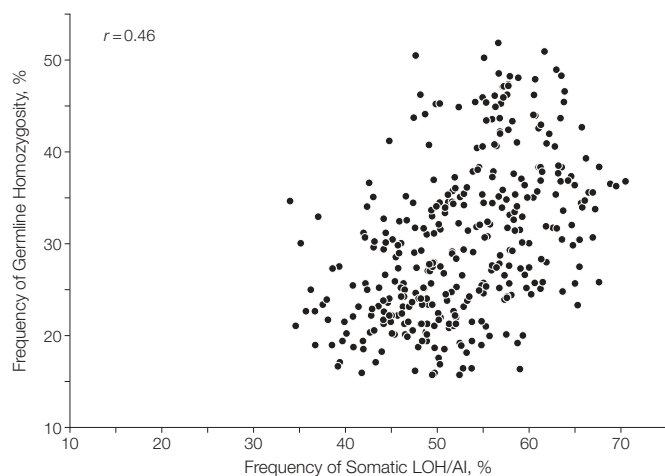
Table 1. Sixteen Microsatellite Markers With Elevated Frequencies of Germline Homozygosity Common to 3 Cancer Types^a

Marker	Locus	P Value			Genes Within 250 kb
		HNSCC	Prostate Cancer	Breast Cancer	
D2S1790	2p11.2	<.01	<.05	<.001	<i>TMSB10, KCMF1, TCF7L1</i>
D3S2427	3q26.31	<.001	<.01	<.01	
D4S2394	4q28.2	<.001	<.001	<.05	
D5S2505	5p15.32	<.001	<.05	<.01	
D6S1959	6p22.3	<.01	<.001	<.001	
D7S3046	7q11.22	<.001	<.01	<.05	<i>DFNB39</i>
D9S1871	9p24.2	<.01	<.01	<.05	<i>GLIS3</i>
D10S1222	10q26.2	<.01	<.01	<.05	<i>DOCK1, FGFR2</i>
D11S1993	11p11.2	<.01	<.05	<.01	<i>API5, TTC17, HSD17B12,</i>
D11S1986	11q23.2	<.001	<.001	<.001	<i>POU2AF1, BTG4, LAYN</i>
D11S4463	11q24.3	<.05	<.01	<.01	<i>BARX2</i>
D13S793	13q31	<.001	<.05	<.05	
D15S822	15q12	<.001	<.001	<.01	<i>GABRA5</i>
GATA178F11	18p11.31	<.001	<.01	<.05	
D18S1376	18p11.31	<.001	<.001	<.001	<i>ZFP161, EPB41L3</i>
D20S477	20p11.21	<.001	<.01	<.05	<i>FOXA2</i>

Abbreviations: HNSCC, head and neck squamous cell carcinoma; kb, kilobase.

^aAdjusted P values for all 3 solid tumor sets combined at these 16 markers $<.001$. All comparisons were with controls.

Figure 4. Frequency of Germline Homozygosity of Microsatellite Markers in All Cancer Patients



Frequency of germline homozygosity of microsatellite markers among all cancer patients as a function of the frequency of loss of heterozygosity/allelic imbalance (LOH/AI) in tumors whose corresponding germline is heterozygous at the same loci, showing a correlation coefficient of 0.46 (Pearson r ; $P < .001$).

To further corroborate this generalized association, we first determined the loci with the highest frequency of somatic LOH/AI events across all the 3 types of cancers. After identifying these “hot-spot” LOH/AI loci, we then determined whether these loci also exhibited (relatively) high frequencies of homozygosity across the germlines of cancer patients. First, loci with high LOH/AI frequencies were identified for each chromosome. Then a list of markers was tabulated that represented those with a significant elevation of both germline homozygosity and LOH/AI frequencies for the entire cohort and for each specific cancer type (TABLE 2). A total of 22 hot spots were identified, including 21 common to the entire se-

ries, 1 in the head and neck squamous cell carcinoma group, 4 in the prostate cancer group, and 0 in the breast cancer group. All 22 of the “overall” hot-spot loci (Table 2) are also found among the regions initially identified as having increased homozygosity frequencies. Again, only 1 of these 22 hot spots (D16S2621) is included in a copy number variation region.

To independently validate our observations, we used an existing published data set comprising SNP genotypes of paired germline and tumor DNA from 205 white patients with lung carcinoma. This publicly available data set provided us with a different solid tumor, which happens to be the most common solid tumor in men,

assessed with a different genotyping technique. The frequencies of homozygosity were compared with HapMap data, and at least 1 SNP (range, 1-16) in each of the 11 of the 12 previously noted evaluable “highly homozygous” regions was found to be significantly more homozygous in lung cancer cases compared with controls ($P < .05$ to $P < .001$) (eTable, available online at <http://www.jama.com>). Thus, we were able to independently replicate our observations of increased germline homozygosity at specific markers in the original 3 solid tumor types. We then used the lung cancer SNP data set to validate our previously noted association between frequency of germline homozygosity and that of somatic LOH. Con-

Table 2. Germline Homozygosity and Somatic LOH/AI Hot Spots^a

Marker	Locus	P Value								Genes at or Within 250 kb
		All		HNSCC		Prostate Cancer		Breast Cancer		
		Homozygous ^b	LOH ^b	Homozygous ^b	LOH ^b	Homozygous ^b	LOH ^b	Homozygous ^b	LOH ^b	
D1S2134	1p33	<.001	<.01	<.05	.61	.14	<.05	.31	.30	
D1S1589	1q25.1	<.01	<.05	.40	.60	<.01	.44	>.99	.38	<i>RABGAP1L, GPR52</i>
D1S3462	1q42.2	<.001	<.05	<.01	.85	<.05	.22	>.99	.33	<i>DISC1, DISC2, TSNAX</i>
D2S1334	2q21.3	<.001	<.05	<.001	.46	<.001	<.01	.28	.68	<i>R3HDM1, MCM6, LCT, ZRANB3, DARS, UBXD2</i>
D2S1776	2q24.3	<.001	<.05	<.001	.61	<.05	.11	>.99	.37	<i>LASS6, ABCB11, G6PC2, SPBC25, NOSTRIN</i>
D3S2409	3p21.31	<.001	<.05	.63	.05	<.001	.30	.06	.75	<i>RHOA, GPX1, USP4, AMT, CCDC71, LAMB2, USP19, NICN1, DAG1, CCDC36</i>
D3S1746	3q25.1	<.001	.14	<.001	.90	<.05	<.05	.09	.56	<i>SUCNR1, AADAC, MBNL1</i>
D3S1763	3q26.1	<.001	<.05	<.05	.81	.12	<.05	>.99	.44	<i>WDR49, PDCD10, SERPIN1</i>
D3S2427	3q26.3	<.001	<.05	<.001	.61	<.01	<.01	<.01	.87	
D4S2623	4q25	<.001	<.01	<.01	.46	<.01	.38	>.99	.06	<i>EGF, ELOVL6, CFI, NOLA1, PLA2G12A, RRH</i>
D8S1128	8q24.21	<.05	<.01	<.05	.32	<.05	.37	>.99	.27	<i>MYC, PVT1</i>
D9S1121	9p21.3	<.01	<.05	<.01	.67	<.05	.27	>.99	<.05	<i>TUSC1</i>
D10S1423	10p12.33	<.05	<.05	<.01	.32	.22	.27	>.99	.38	<i>C10orf112</i>
D11S1986	11q23.2	<.001	<.05	<.001	.52	<.001	.34	<.001	.30	<i>POU2AF1, BTG4, LAYN</i>
D11S4464	11q24.1	<.05	<.05	<.05	.46	.21	.19	>.99	.38	<i>OR6X1, ZNF202, SCN3B, OR6M1, PMP22, OR8D4</i>
D12S2078	12q24.32	<.01	<.001	.55	.05	.35	<.05	.96	.61	
D14S1280	14q12	<.05	<.05	<.05	.42	.09	.64	>.99	.30	<i>NOVA1</i>
D15S655	15q25.3	<.05	<.05	<.05	.26	<.01	.81	>.99	.30	<i>TMEM83</i>
D16S422	16q23.3	<.05	<.01	<.01	.05	>.99	.22	.96	.46	<i>CDH13</i>
D16S2621	16q24.2	<.01	<.05	<.01	.96	.14	<.05	>.99	.30	<i>JPH3, KLHDC4, SLC7A5, CA5A, BANP</i>
D21S2055	21q22.2	<.001	<.05	<.001	<.05	<.001	.28	.14	.75	<i>PCP4, DSCAM, B3GALT5</i>
D22S683	22q12.3	<.001	<.001	<.001	.05	<.001	<.01	>.99	.39	<i>APOL1 through APOL4, MYH9</i>

Abbreviations: HNSCC, head and neck squamous cell carcinoma; kb, kilobase; LOH/AI, loss of heterozygosity/allelic imbalance.

^aNote adjusted P values <.05 are considered significantly elevated homozygosity frequencies or somatic LOH/AI frequencies. Comparisons of germline homozygosity were against controls. Comparisons of LOH frequencies were against an internal mean for each tumor type.

^bAdjusted P values of significantly elevated germline homozygosity frequency and somatic LOH/AI frequency are provided across the whole data set (All) and for individual cancer types.

firming our data in breast, prostate, and head and neck carcinomas, we found a direct correlation between LOH frequencies and those of germline homozygosity at the corresponding SNPs in the germline that happens to be heterozygous ($r=0.21$, 95% confidence interval, 0.18-0.24, $P<.001$).

COMMENT

Our data suggest that there are specific common loci that are prone to homozygosity in the germline of individuals with breast, prostate, and head and neck squamous cell carcinomas. Importantly, we were able to independently validate our observations in a different type of solid tumor, lung carcinoma, by showing an increased frequency of germline homozygosity in cancer cases compared with ancestry-matched controls. It is of interest to note that the most frequently homozygous SNPs were, for the most part, in proximity to the original microsatellites found to be "most homozygous" in the other 3 solid tumors (eTable). We were also able to confirm the correlation between germline homozygosity and increased somatic LOH at the same marker when the latter's germline happens to be heterozygous.

There are at least 2 hypotheses to explain the increased germline homozygosity in cancer cases compared with controls. First, random mating would generate random homozygosity, and when this homozygosity occurs at enough loci harboring genes relevant to cancer with pathogenic alleles, cancer risk is increased. This could explain the increased homozygosity at specific loci we observed in cancer patients compared to controls. If this were true, then we would expect that homozygosity would favor the frequent alleles. However, this cannot explain the high somatic LOH/AI frequencies observed at these same loci when the germline happens to be heterozygous. Therefore, an alternative hypothesis has to be considered.

Increased germline homozygosity may be postulated to occur at com-

mon fragile sites, which are specific loci with increased sensitivity to DNA damage.²⁶ Common fragile sites, which are scattered across the genome, should be distinguished from "rare" fragile sites, which are also sensitive to DNA damage, because the latter harbor expanded tandem nucleotide repeats that common sites do not have. Hence, the common fragile sites would be good candidate loci to have high frequencies of germline homozygosity in cancer patients and high LOH/AI frequencies in tumors, because they are prone to both homologous recombination as a mechanism of DNA repair after double-strand breaks²⁷ and to rearrangements in tumors, including LOH/AI.²⁶

Remarkably, 1 hot-spot marker we identified (D8S1128) is approximately 250 kb from a common fragile site (*MYC*). Two additional hot spots (D16S2621 and GATA178F11) are 250 kb from 2 rare fragile sites (*HDL2* and *TYMS*) related to triplet nucleotide repeats. However, for these rare fragile sites, the link with somatic LOH/AI is not well established. Nonetheless, our current data may contribute to this link given that our data derive from microsatellite markers, which comprise nucleotide repeats of various repeat lengths. It would, therefore, be rather interesting to distinguish the precise mechanisms of the association of germline homozygosity and cancer susceptibility as well as its link to increased LOH when the germline happens to be heterozygous.

Some may be surprised that the markers associated with traditional high-penetrance susceptibility genes are not prominent among our highly homozygous list. Currently, there are no high-penetrance susceptibility genes for prostate, lung, or head and neck carcinomas. If we focused on the group of patients with breast cancers, we did not see the regions associated with high-penetrance inherited breast cancer syndromes, namely, *BRCA1* (17q12), *BRCA2* (13q14), *PTEN* (10q23.3), or *TP53* (17p13-p15) (Table 1). This non-association may contribute to our overall hypothesis that regions of germline

homozygosity could represent low-penetrance factors predisposing to carcinoma, at least to these 3 cancer types. In contrast, there are at least 3 SNPs identified and replicated as associated with breast cancer cases^{28,29} that are also found in our regions of high germline homozygosity in our patients with breast cancer, namely, 10q26 (SNP in proximity to *FGFR2*), 5q1-3 (*MAP3K1*), and 11p15.5 (*LSP1*).

Similarly, SNPs representing several different loci in the 8q24 region have been shown and validated to be associated with prostate cancer cases,^{30,31} and this 8q24 region is also well represented among our loci with high frequencies of germline homozygosity in our prostate cancer cases. Interestingly, and consistent with our present observations, this 8q24 region also harbors SNPs associated with colorectal and other cancers,^{32,33} perhaps suggesting a more general cancer susceptibility allele or alleles in this region. Our current data reveal several distinct markers in 8q24 with high frequencies of germline homozygosity and LOH for each tumor group as well as in the overall merged series of all patients with these breast, prostate, and head and neck cancers (Table 2).

Cancer susceptibility loci are currently sought by association studies, whereas somatic alterations are most often identified separately. The recent availability of high-density arrays considerably increased the resolution of these studies.^{28,29,31-34} However, currently, somatic and germline cancer genetics are 2 separate fields. In fact, from germline inheritance of a predisposing allele to cancer development is still somewhat of a mechanistic "black box." Our observations link the somatic and germline cancer genetic fields. Our observations are also consistent with cancer as a complex genetic trait if we consider that carcinogenesis occurs when a minimal number of somatic events are required for a cell to become malignant. Continuing this concept, therefore, germline homozygosity at specific loci represents a low penetrance but common etiology for cancer and

serves as a parallel but equivalent pathway to accumulation of somatic LOH.

CONCLUSIONS

Our data derived from 3 different solid tumors, validated in a fourth, demonstrate that high frequencies of germline homozygosity at specific markers are associated with these cancers compared with controls and that at the genome-wide level, there is a direct relationship between frequencies of germline homozygosity and somatic LOH if the germline is heterozygous. This represents, to our knowledge, the first such observation, and although the total number of cancer cases nears 600, the sample size per solid tumor type is still relatively small (116-205). Therefore, our observations here should be validated in these solid tumors and ex-

plored in other malignancies. If our data can be robustly replicated independently, then germline homozygosity at specific loci as low-penetrance alleles predisposing to carcinomas could be taken into account in future cancer risk assessments and management beyond high-penetrance cancer susceptibility genes. Additionally, with further studies and fine structure analyses, it may be possible to use such data to predict the likelihood of LOH in a tumor at specific genomic loci if we knew the relative frequencies of germline homozygosity/heterozygosity at those same loci.

Author Contributions: Dr Eng had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Assié, Eng.

Analysis and interpretation of data: Assié, LaFramboise, Platzer, Eng.

Drafting of the manuscript: Assié, LaFramboise, Platzer, Eng.

Critical revision of the manuscript for important intellectual content: Platzer, Eng.

Statistical analysis: Assié, LaFramboise.

Obtained funding: Eng.

Administrative, technical, or material support: Platzer, Eng.

Study supervision: Eng.

Financial Disclosures: None reported.

Funding/Support: This work was partially funded by grants 1P01CA97189-01A2 and 1P50/U54CA113001-01 from the National Cancer Institute and grant DAMD-02-1-0118 from the Department of Defense Prostate Cancer Research Program (Dr Eng). Dr Assié was supported by clinical research fellowships from the Fondation de France and the Fédération Nationale des Centres de Lutte contre le Cancer. Dr Eng is a recipient of the Doris Duke Distinguished Clinical Scientist Award and is the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic.

Role of the Sponsor: The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation or approval of the manuscript.

Supplementary Information: The eTable is available online at <http://www.jama.com>.

Additional Contributions: We thank Kevin Zbuk, MD, and Mohammed Orloff, PhD, Genomic Medicine Institute, Cleveland Clinic, for critically reviewing drafts of this article on a voluntary basis.

REFERENCES

- Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68(4):820-823.
- Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet*. 2003;33(suppl):238-244.
- Fukino K, Shen L, Matsumoto S, Morrison CD, Mutter GL, Eng C. Combined total genome loss-of-heterozygosity scan of breast cancer stroma and epithelium reveals multiplicity of stromal targets. *Cancer Res*. 2004;64(20):7231-7236.
- Fukino K, Shen L, Patocs A, Mutter GL, Eng C. Genomic instability within tumor stroma and clinicopathologic characteristics of sporadic primary invasive breast carcinomas. *JAMA*. 2007;297(19):2103-2111.
- Kurose K, Hoshaw-Woodard S, Adeyinka A, Lemeshow S, Watson PH, Eng C. Genetic model of multi-step breast carcinogenesis involving the epithelium and stroma: clues to tumour-microenvironment interactions. *Hum Mol Genet*. 2001;10(18):1907-1913.
- Weber F, Xu Y, Zhang L, et al. Microenvironmental genomic alterations correlate with clinicopathologic behavior in head and neck squamous cell carcinomas. *JAMA*. 2007;297(2):187-195.
- Sarquis MS, Weber F, Shen L, et al. High frequency of loss-of-heterozygosity in imprinted compared to non-imprinted genomic regions in follicular thyroid carcinoma and atypical adenomas. *J Clin Endocrinol Metab*. 2006;91(1):262-269.
- Rudan I, Rudan D, Campbell H, et al. Inbreeding and risk of late onset complex disease. *J Med Genet*. 2003;40(12):925-932.
- Shami SA, Qaisar R, Bittles AH. Consanguinity and adult morbidity in Pakistan. *Lancet*. 1991;338(8772):954.
- Rudan I. Inbreeding and cancer incidence in human isolates. *Hum Biol*. 1999;71(2):173-187.
- Simpson JL, Martin AO, Elias S, Sarto GE, Dunn JK. Cancers of the breast and female genital system: search for recessive genetic factors through analysis of human isolate. *Am J Obstet Gynecol*. 1981;141(6):629-636.
- Bhattacharya P, Duttagupta C, Sengupta S. Proline homozygosity in codon 72 of p53: a risk genotype for human papillomavirus related cervical cancer in Indian women. *Cancer Lett*. 2002;188(1-2):207-211.
- Deligezer U, Akisik EE, Dalay N. Homozygosity at the C677T of the MTHFR gene is associated with increased breast cancer risk in the Turkish population. *In Vivo*. 2005;19(5):889-893.
- Henry I, Bonaiti-Pellie C, Chehensse V, et al. Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature*. 1991;351(6328):665-667.
- Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. *Am J Hum Genet*. 1999;65(6):1493-1500.
- Clark AG. The size distribution of homozygous segments in the human genome. *Am J Hum Genet*. 1999;65(6):1489-1492.
- Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*. 2006;15(5):789-795.
- Simon-Sanchez J, Scholz S, Fung HC, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet*. 2007;16(1):1-14.
- Patocs A, Zhang L, Xu Y, et al. Stromal TP53 mutation or 5-locus allelic imbalance and nodal metastases in breast cancer. *N Engl J Med*. 2007;357(25):2543-2551.
- R Development Core Team. *R: a Language and Environment for Statistical Computing*. Vienna, Austria: University of Vienna; 2007.
- Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature*. 2007;450(7171):893-898.
- Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics*. 2004;20(8):1233-1240.
- Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet*. 1989;44(3):388-396.
- Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science*. 2002;298(5602):2381-2385.
- Redon R, Isikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-454.
- Huebner K, Croce CM. FRA3B and other common fragile sites: the weakest links. *Nat Rev Cancer*. 2001;1(3):214-221.
- Jackson SP. Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. 2002;23(5):687-696.
- Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet*. 2007;39(3):352-358.
- Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility genes. *Nature*. 2007;447(7148):1087-1093.
- Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007;39(5):645-649.
- Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39(7):870-874.
- Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*. 2007;39(8):984-988.
- Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet*. 2007;39(8):989-994.
- Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet*. 2007;39(5):631-637.