Trends in Genetics

CellPress
REVIEWS

Review

Germline Variants Impact Somatic Events during Tumorigenesis

Johnny R. Ramroop,[1] Madelyn M. Gerber,[2] and Amanda Ewart Toland [1,3,*]

Cancer is characterized by diverse genetic alterations in both germline and somatic genomes that disrupt normal biology and provide a selective advantage to cells during tumorigenesis. Germline and somatic genomes have been extensively studied independently, leading to numerous biological insights. Analyses integrating data from both genomes have identified genetic variants impacting somatic events in tumors, including hotspot driver mutations. Interactions among specific germline variants and somatic events influence cancer subtypes, treatment response, and clinical outcomes. Investigation of these complex interactions is increasing our understanding of aberrant pathways in tumors that may uncover novel therapeutic targets. Here, we review the literature describing the role of germline genetic variants in promoting the selection and generation of specific mutations during tumorigenesis.

Highlights

The genetic context in which a somatic mutation occurs can impact whether it is likely to be selected for during tumor development.

Germline pathogenic variants in highly penetrant cancer susceptibility genes are associated with specific tumor subtypes as well as with somatic mutations in specific genes and pathways.

A subset of cancer susceptibility alleles identified through genome-wide association studies (GWAS) shows allele-specific copy number gains and losses in tumors.

GWAS have identified genetic variants associated with specific somatic events in cancer, highlighting new biological connections.

Germline variants in immune system genes, such as MHC class 1 genes, enable cells with somatic mutations at specific amino acid residues to evade the immune system.

Germline Variants Impact Tumor Characteristics

Using next-generation sequencing, germline and tumor genomes have been explored independently for mutations that are associated with tumorigenesis, often with the germline genome serving as a reference control for acquired mutations. Analyses integrating data from both genomes are leading to novel associations between **germline variants** (see Glossary) and specific somatic events. Recent investigation of the complex interplay between germline variants and cancer-associated mutations is increasing our understanding of aberrant pathways that may be potential therapeutic targets. In this review, we explore the hypothesis that specific germline variants determine which somatic events and mutations are generated and selected for in cancer cells during tumorigenesis (Figure 1, Key Figure). We also discuss the body of literature supporting a role for germline polymorphic and **pathogenic variants (PVs)** in shaping several key tumor characteristics. Specifically, we address the emerging link between **germline variant by somatic mutation (G×M)** associations. We also focus on tumor characteristics, such as histopathological subtypes, mutational signatures and **microsatellite instability (MSI)**, **allele**-specific copy number changes and **loss of heterozygosity (LOH)**, **somatic mutations** in *cis* and in *trans* with the relevant genetic variant, and the tumor immune response.

Germline Variants and Histopathological Subtypes

Tumors from individuals with hereditary cancer syndromes often exhibit characteristic histopathological features at different frequencies compared with tumors arising in the general population. For example, breast cancers occurring in individuals with *BRCA1* or *BRCA2* germline PVs are associated with different frequencies of cancer subtype, grade, and estrogen receptor (ER) status compared with their germline-negative counterparts [1]. Specifically, basal-like subtype and triple-negative grade 3 breast cancers are highly associated with *BRCA1* PVs. Multiple studies have found associations of lower-**penetrance** common **single nucleotide variants (SNVs)** with specific breast cancer subtypes, particularly ER-positive and ER-negative subtypes [2–7]. The functions of many of these variants, the majority mapping to intergenic regions, remain unknown. A few germline variants associated with ER status map to the *ESR1* gene and lead

[1]Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA
[2]Department of Pathology, The Ohio State University, Columbus, OH, USA
[3]Department of Internal Medicine, The Ohio State University, Columbus, OH, USA

*Correspondence:
Amanda.toland@osumc.edu
(A.E. Toland).

CellPress
REVIEWS

Key Figure

Examples of Germline Variants by Somatic Mutation (G×M) Association

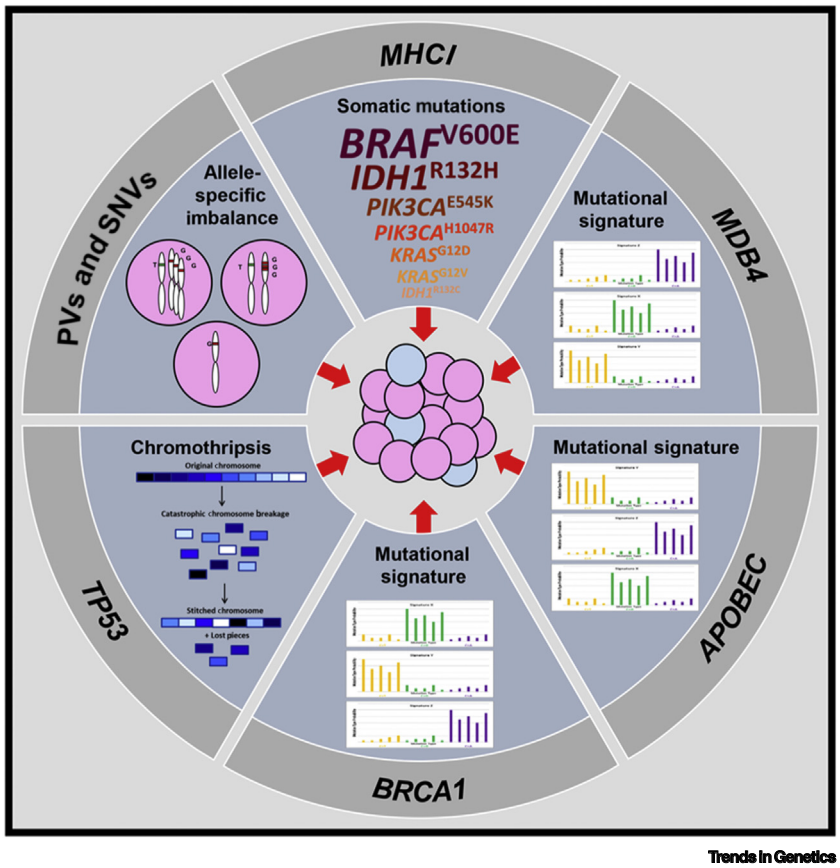

Figure 1. Specific germline polymorphic or pathogenic variants impact which somatic events, mutational signatures, allele-specific copy number changes, or altered signaling pathway are selected for in cancer cells during tumor evolution. The outer circle shows germline variants and the inner wedges represent the associated somatic events for that specific example of a G×M association. *BRCA1* pathogenic variants (PVs) are associated with a mutational signature characterized by small tandem duplications and deletions. *MDB4* PVs are associated with mutational signatures, particularly C>T substitutions in CpG dinucleotides, and *APOBEC* germline variants are associated with the mutational signature C>T or C>G substitution within TCA and TCT motifs. *TP53* PVs are associated with chromothripsis. More generally, germline PVs and single nucleotide variants (SNVs) are associated with preferential gain of the risk/oncogenic allele and/or loss of the nonrisk/tumor-suppressing allele. Specific *HLA* haplotypes have also been linked to higher and lower frequencies of specific pathogenic variants because they result in the immune system being more or less likely to recognize specific somatic mutations, respectively.

to differential *ESR1* expression, providing a mechanistic link between the variant and tumor phenotype [8]. Similarly, PVs in renal cell carcinoma (RCC) genes are associated with specific histological subtypes [9]. Histopathological subtypes often correlate with clinical outcomes. Therefore, knowledge of the associated germline SNVs and PVs can contribute to better predictive power for survival and therapeutic responses. For example, screening for germline *BRCA1/2* PVs can predict risk for ER-negative (triple-negative and basal-like) tumors, which are more aggressive and have poorer prognoses compared with ER-positive (luminal-A and luminal-B) subtypes [10].

Glossary

Allele: alternative forms of DNA located at the same genetic locus on a chromosome, also known as a variant. Diploid organisms have two alleles at each genetic locus, with one allele being inherited from each parent.

Allele-specific imbalance (ASI): somatic DNA copy number alterations in which one allele shows preferential copy number changes (loss or gain) compared with the other allele.

Germline variants: variations in DNA sequence transmitted from parent to offspring via the sperm or egg. These variants are in all cells of the offspring and can be transmitted to future generations.

Genome-wide association studies (GWAS): association studies for a disease (or phenotype) or disease in which genetic variants across the entire genome are tested for association with disease in individuals with the disease (cases) and individuals without the disease (controls).

Germline variant by somatic mutation (G×M) association: when a germline variant is associated with an increased likelihood that a specific somatic mutation will be present in a tumor.

Linkage: the tendency for two or more genes located close together on the same chromosome to be inherited together. 'Linkage study' refers to a family-based method used to map a trait to a genomic location by demonstrating co-segregation of the disease with genetic markers of known chromosomal location.

Loss-of-heterozygosity (LOH): a genetic event whereby one of two different alleles at a locus is lost. When LOH occurs in tumors, the genome is homozygous at that locus but is heterozygous in the corresponding germline DNA.

Microsatellite instability (MSI): the condition of genetic hypermutability resulting from defective DNA mismatch repair.

Oncogene: a gene or a mutant variant of a gene that is associated with tumorigenesis.

Pathogenic variant (PV): a genetic alteration that is associated with or increases an individual's predisposition to a particular disease, also known as a mutation.

Penetrance: the proportion of individuals in a population carrying a

Germline Variants and Tumor Mutational Signatures

Germline variants are associated with mutation patterns in tumors. Breast tumors from individuals with germline *BRCA1* and *BRCA2* PVs have a higher frequency of small tandem duplications at regions of microhomology and deletions of <100 kb [11] (Figure 1). These signatures likely reflect *BRCA1* and *BRCA2* function in homology-directed DNA repair and suggest that poly ADP ribose polymerase (PARP) inhibitors will be beneficial to this patient population [12].

A study that systematically evaluated the association of germline variants with somatic mutation signatures in 2642 individuals across 39 cancer types found that: (i) individuals with *MDB4* PVs showed enrichment of C to T substitutions at CpG dinucleotides (Figure 1); (ii) *BRCA1*-deficient tumors were associated with complex rearrangements, including insertions and tandem duplications; (iii) germline L1/LINE elements were associated with somatic retrotransposition; and (iv) germline *TP53* PVs were associated with chromothripsis [13] (Figure 1). This study also highlighted the relevance of the APOBEC family of cytidine deaminases in cancer. Tumors from individuals with germline variants near *APOBEC3* had reduced levels of APOBEC mutational signatures (Figure 1), characterized by substitutions of C with either T or G within TCA or TCT motifs [14–16]. Specifically, *rs2395185* was associated with *APOBEC* mutation signatures in lung cancer [14] and *rs1014971* in bladder tumors [15] (Table 1). APOBEC signatures in breast tumors also associate with *rs12628403,* a tagging SNV for a polymorphic deletion resulting in a fusion of *APOBEC3A* with noncoding regions of *APOBEC3B* [13,15,16].

The occurrence of MSI in the tumor genome is also influenced by germline PVs. MSI is observed at a high frequency in colorectal cancer (CRC) with germline or somatic PVs in mismatch repair genes [17–19]. Tumors with high MSI are more likely to be of low pathological stage, more likely to carry activating *BRAF* mutations, and less likely to carry activating *KRAS* mutations compared with tumors that are microsatellite stable and are more likely of low pathological stage [20]. Compared with microsatellite stable CRC, MSI is predictive of a favorable outcome and reduced likelihood of metastases [21].

Collectively, these studies demonstrate that germline variants in genes involved in DNA repair or maintenance of genome integrity can result in characteristic mutational patterns reflective of the perturbed function. Additionally, these findings may uncover novel strategies for the prevention of hypermutation or failed DNA repair and, as in the case of MSI-high CRC, can inform clinical outcomes.

Germline Variants and Somatic Copy Number Changes or Mutations in *cis*

In addition to global influences on tumor subtype and mutational signatures, germline variants are also associated with specific somatic events 'in *cis*'; these include phenomena such as LOH**, allele-specific imbalance (ASI)**, and somatic mutations occurring at the same locus as the germline variant. A classic example of the connection between germline variants and somatic events in tumor development is illustrated by Alfred Knudson's 1971 observation, today known as the '**two-hit hypothesis**' [22] (Box 1). Knudson showed that retinoblastoma occurs earlier in individuals who inherited a germline PV in the *RB* gene compared with individuals in the general population. Individuals who inherit a PV in every cell in the retina only need one cell to acquire a 'hit' in the remaining wild-type allele for selective growth advantages leading to cancer, whereas retinal cells from individuals in the general population must acquire two independent events within the same cell to become at risk of developing a tumor. Tumors arising in the context of hereditary cancer syndromes, such as retinoblastoma, frequently show evidence of LOH of the wild-type allele because of the selective advantage that these tumor cells have over cells retaining one wild-type allele.

phenotype or disease-associated genetic variant who manifest the trait. When individuals carry a particular phenotype-associated allele and do not exhibit the phenotype, the gene is said to have reduced or lower penetrance.
Resistance allele: also known as a nonrisk or protective allele. Depending on the context, these alleles are associated with no change in disease risk or are associated with a decreased risk.
Single nucleotide variant (SNV): also known as a single nucleotide polymorphism (SNP). A SNV is a sequence variation in a single nucleotide occurring at a specific genetic location. The term 'SNP' traditionally refers to variation present at a frequency >1% in a population, whereas a SNV can occur at any frequency in a population.
Somatic mutation: a change in DNA that occurs in nongerm cells after conception.
Susceptibility allele: also known as a risk-allele; alleles associated with increased likelihood of developing the disease.
Tumor suppressor genes: genes that suppress tumor development by regulating cell growth and division, stimulating cell death and/or DNA repair.
Two-hit hypothesis: a hypothesis in which two mutations, one in each copy of a tumor suppressor gene, are required for a cell to give rise to a tumor.

Table 1. Germline Variants by Somatic Mutation Associations in *trans*[a]

Cancer type	SNP ID; genomic location[b]	Gene/pathway with somatic mutation; gene locus[c]	Impact	Refs
Glioblastoma	*rs13222385;* chr7:55183900	*LANCL2;* 7p11.2	*P*=0.037	[42]
Multiple	*rs8051518;* chr16:7270514	*SF3B1;* 2q33.1	Eightfold increase incidence	[49]
	rs25673; chr19:2109158	*PTEN;* 10q23.31	Fourfold increase incidence	
Breast	*rs252913;* chr5:56900019	*PIK3CA;* 3q26.32	$P = 0.01$	[51]
	rs331499; chr5:56915096		$P = 0.02$	
Lung	*rs36600*; chr22:29941597	*ARID1A;* 1p36.11	$P = 5.78\times10^{-4}$	[14]
	rs2395185; chr6:32465390	Cell cycle pathway genes	$P = 3.61\times10^{-4}$	
	rs2395185; chr6:32465390	*APOBEC;* genome-wide APOBEC mutational signature	$P = 3.58\times10^{-3}$	
	rs3817963; chr6:32400310	Cell cycle pathway genes	$P = 4.15\times10^{-4}$	
	rs3817963; chr6:32400310	MAPK pathway genes; e.g., *MAPK1* 22q11.22	$P = 8.58\times10–^{-4}$	
Bladder	*rs1014971;* chr22:38936618	*APOBEC ;* Genome-wide APOBEC mutational signature	$P = 1.92\times10^{-5}$	[15]
Gastric	*rs2285947;* chr7:21544470	PDGF pathway; e.g., *PDGFB* 22q13.1	$P = 3.93\times10^{-4}$	[50]
	rs1679709; chr6:28260564	*RCF4* mismatch repair gene; 3q27.3	$P = 1.25\times10^{-2}$	
Lung (mouse)	*rs13459194;* chr19:32049010[d]	*Kras Q61L* ; chr6:77.37 cM[d]	$P = 1.6\times10^{-5}$	[29]
Melanoma	*rs12203592;* chr6:396321	*BRAF* V600E; 7q34	OR 0.59; 95% CI = 0.43–0.79	[53]
		BRAF V600K; 7q34	OR 0.65; 95% CI = 0.41–1.03	
		BRAF (other in exon 15) ; 7q34	OR 1.57; 95% CI = 0.93–2.65	
		NRAS; 1p13.2	OR 0.99; 95% CI = 0.75–1.30	
	rs132985; chr22:38167464	*BRAF* V600E; 7q34	OR 1.32; 95% CI = 1.05–1.67	
		BRAF (other in exon 15) ; 7q34	OR 1.82; 95% CI = 1.11–2.98	
		BRAF V600K; 7q34	OR 1.12; 95% CI = 0.78–1.60	
		NRAS; 1p13.2	OR 0.83; 95% CI = 0.64–1.07	
RCC	*rs10932384;* chr2:211545123	*ERBB4;* 2q34	P = 0.003	[46]

[a]Abbreviations: CI, confidence interval; NSCLC, non-small cell lung cancer; OR, odds ratio.
[b]Genomic locations obtained from NCBI dbSNP, Human assembly GRCh38/hg38.
[c]Somatic mutation locations obtained from NIH, Genetic Home Reference.
[d]Mouse genomic location obtained from UCSC Genome Browser and somatic mutation location obtained from NCBI, GRCm38/mm10.

In tumors, loci harboring **oncogenes** often show copy number gains. **Tumor suppressor gene** loci are more typically lost through somatic mutations, chromosomal rearrangements, or deletions. Copy neutral alteration (in which the DNA is neither lost nor gained) and LOH of the

Box 1. Knudson's Two-Hit Hypothesis

Knudson's two-hit hypothesis, also referred to as the two-mutation hypothesis, was proposed in 1971 by Alfred Knudson from his work on the genetic mechanism behind retinoblastoma [22]. In his study of patients with retinoblastoma, data such as age at diagnosis, gender, family history, unilateral or bilateral disease occurrence, and the approximate number of tumors in each eye were collected and analyzed. At the time of Knudson's study, it was believed that retinoblastoma could be caused by either somatic or germline mutations. Knudson's findings showed that retinoblastoma was caused by a germline mutation in a subset of cases, but for retinoblastoma to occur, a subsequent inactivating mutation in the wild-type allele was required. Hereditary cases, in which individuals inherited one copy of a mutated *RB* gene, were frequently bilateral and occurred at a younger age compared with nonhereditary cases. Knudson concluded that retinoblastoma was caused by two mutations (one in each copy of the gene), hence the two-hit hypothesis. Individuals with no inherited mutation have to somatically accumulate a loss-of-function mutation in each allele for the disease to manifest, and this process takes longer compared with individuals who already have one inherited mutation. Knudson's work illustrates the concept of allele-specific somatic events: for an individual with an inherited pathogenic variant in the *RB* gene to develop retinoblastoma, a somatic loss-of-function mutation in the wild-type allele is required.

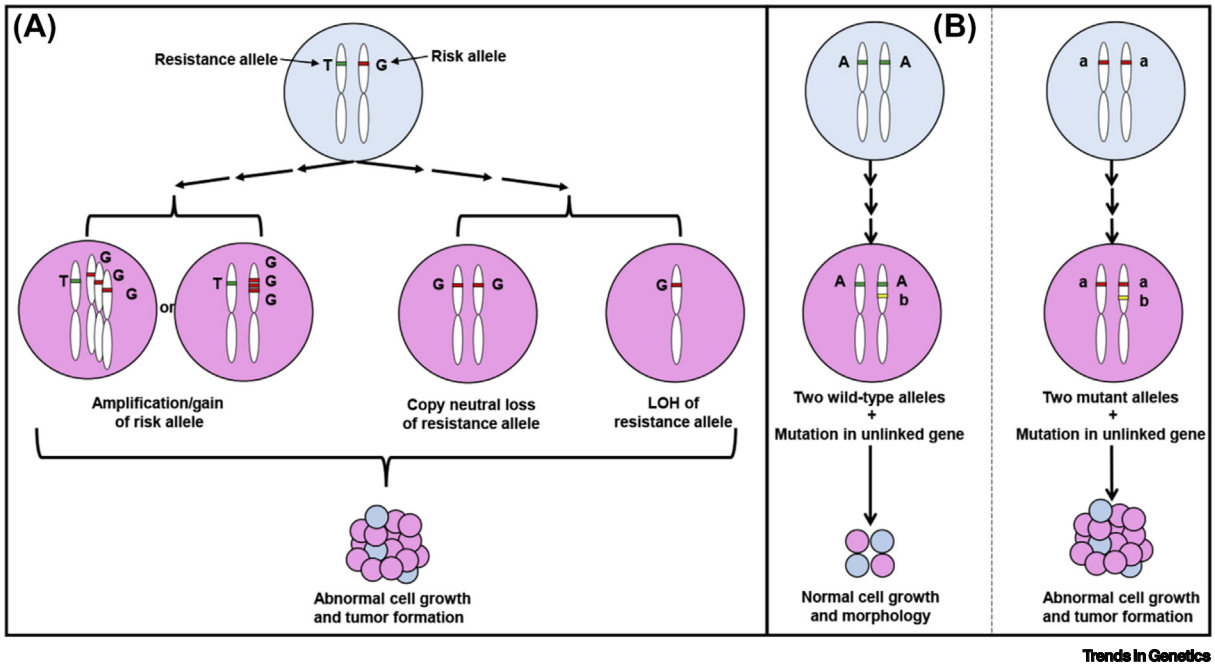

Figure 2. Examples of Germline Alleles Conferring Selective Advantage to Specific Somatic Alterations. (A) Allele-specific imbalance (ASI) in *cis* (affecting the same gene) can arise in at least three ways (from left to right): by amplification or gain of the risk allele ('G'), by copy neutral loss of the resistance allele ('T'), or by loss of heterozygosity (LOH) due to loss of the resistance allele ('T'). ASI can increase the selective advantage of these cells (in pink) via the risk allele promoting proliferation and survival pathways leading to preferential tumor growth of cells with gain of the risk allele over cells that have gain of the protective allele (in blue). (B) A germline genetic variant ('a') may associate with a mutation in an unlinked gene ('b'). However, when a somatic mutation occurs in cells with the nonassociated allele ('A'), there is no selective advantage to the cells. When a somatic mutation occurs in cells with the associated allele ('a'), there is dysregulation of normal pathways and/or functions, resulting in selective advantage and expansion of those cells (in pink).

wild-type (nonmutated) allele are frequent mutations in tumor suppressor genes (Figure 2A). This phenomenon is referred to as ASI (Box 2 and Figure 1). By contrast, activating mutant alleles of oncogenes such as *EGFR, KRAS, PIK3CA*, and *BRAF* are preferentially amplified over nonactivated alleles (Figure 2A), thereby increasing the selective advantage of these cells by promoting their proliferation and survival [23].

Observation of either preferential gain of the mutant allele of an oncogene or LOH of a tumor suppressor gene provides strong evidence for the association of inherited germline variants with specific somatic events in tumors. For example, allele-specific gain of the *KRAS* allele containing a

Box 2. Allele-Specific Imbalance

'ASI' is used to describe nonrandom genetic copy number alterations of one allele relative to another. Typically, this is associated with a relative gain of a highly penetrant oncogenic allele or relative loss of an allele associated with suppression of tumorigenesis. Studies have also found ASI for lower-penetrance alleles associated with increased cancer risk. Quantitative genotyping and copy number analysis are used to identify these events. Normal and tumor DNA from individuals heterozygous for a particular SNV or PV of interest are sequenced, and an ASI score is determined by comparing the allele peak heights between normal and tumor DNA sequencing graphs. In retinoblastoma, the classic tumor suppressor gene *RB* shows ASI, because the wild-type or resistance allele of a tumor suppressor gene is lost and the activating mutated or susceptibility allele is retained and selected for gain or amplification. These allele-specific copy number alterations impact tumorigenesis, progression, and metastasis. Activating mutations in oncogenes increase the selective advantage of cells by affecting proliferation and survival. While an activating somatic mutation in one allele of an oncogene is sufficient for selective growth advantage, malignancies can be promoted by copy number gain, copy neutral alteration, and LOH events. With advances in genome-wide analyses of the tumor genome, ASI is being identified in regions of the genome showing copy number gain, copy neutral alteration, and LOH.

G12D mutation has been observed in 13–55% of CRC cases [24] and is associated with poor overall survival [25]. However, ASI is not restricted to somatic events with large functional impact. Early evidence suggesting that loci harboring low-penetrant cancer **susceptibility** (risk) and **resistance** (protective) **alleles** can also undergo selective copy number changes came from mouse models in which cancer-associated loci were identified by **linkage** [26–29]. In a mouse model for chemically induced skin cancer [26], ~40% of previously identified skin tumor-susceptibility (*Skts*) loci [30] showed evidence of ASI in F1 backcross mice with relative preferential gain of the allele from the susceptible strain or relative preferential loss of the allele from the resistant strain. Notably, the oncogene *Hras* had activating mutations in most of the skin tumors, and 65% of heterozygous tumors showed preferential gain of the chromosome carrying the mutant allele from the susceptible strain [26].

Genome-wide association studies (GWAS) have identified several haplotype-tagging SNVs that associate with cancer risk and somatic mutation frequency (Box 3). One of the first GWAS variants tested for ASI was *rs6983267*, which maps to an intergenic region on 8q24.21 and is associated with increased risk of CRC and prostate cancer [31–33]. When *rs6983267* was genotyped in colon tumor DNA from 466 patients heterozygous for this SNV, 101 tumor DNAs exhibited ASI, with 66% showing relative gain of the risk (G) allele and 34% showing relative gain of the nonrisk (T) allele [34] (Table 2). Similar findings were observed for *rs6983267* in 84 heterozygous cell lines from various malignancies [35] and in an expanded ASI study of 16 CRC GWAS variants in 490 total paired normal and tumor samples [28] (Table 2). However, ASI does not occur at every risk-associated locus, which is consistent with previous mouse studies [26]. Functional studies of *rs6983267* found that gain of the risk allele associated with higher MYC expression and activity, thereby promoting CRC development [35]. Importantly, many variants identified by GWAS are intergenic and can be difficult to functionally assess. Evaluating variants for ASI offers one strategy for prioritizing risk alleles for functional evaluation.

While *rs6983267* maps to a putative long-range enhancer for the MYC oncogene, variants altering protein-coding sequences also exhibit ASI. One example is the variant *rs2273535* (Phe31Ile) in *AURKA*, which was previously identified using data from mouse models, human tumors, and case-control association data as a candidate tumor-susceptibility allele for multiple cancer types [36–40]. In one study, among 48 heterozygous CRC samples with copy number changes at 20q13.2, gain of the risk (A/Ile) allele was favored over the nonrisk (T/Phe) allele (Table 2) and was associated with a higher likelihood of aneuploidy [36]. An independent study found ASI of the *rs2273535* (A/Ile) allele using 125 familial and 110 sporadic CRC cases, with familial heterozygous tumors exhibiting preferential gain of the A allele relative to the T allele [41] (Table 2).

Box 3. Genome-wide Association Studies

GWAS are association studies investigating the correlation between genetic variants and phenotype or disease. When an allele is observed more frequently in the genomes of individuals with a disease or phenotype (cases) compared with individuals without the disease or phenotype (controls), the variant is classified as being associated with that disease. Hundreds of GWAS for cancer have been conducted and have identified variants that are associated with increased risk of cancer, as well as other phenotypes, such as tumor subtype, tumor aggressiveness, propensity to metastasize, and response to therapy. Given that the effect size of most of these variants is small, individually they do not have strong predictive value for the phenotype. However, the predictive value increases when variants are combined into a risk score of multiple variants, or a polygenic risk score. Data from GWAS are leading to a better understanding of the biology and mechanisms of disease that may lead to prevention or more effective personalized risk assessments and precision treatment options. G×M association studies, which identify associations between particular germline variants and an increased likelihood of a specific somatic mutation in an individual's tumor, are a new area of genomic exploration. G×M associations may lead to the identification of the context of the cells that is important for selection during tumor development and uncover novel pathways. Such associations may be utilized to predict prognosis, to predict therapeutic response and clinical outcome, and to identify specific pathways that can be targeted for therapeutic intervention.

Table 2. Cancer Susceptibility Variants Showing Allele-Specific Imbalance in *cis*

SNV ID; genomic location[a]	Cancer	Number of heterozygotes with evidence of copy number imbalance	Risk allele (*n*) percentage	Nonrisk allele (*n*) percentage	*P* value[b]	Refs
rs6983267; chr8:127401060	CRC	101	G, loss (34/101) 34%	T, loss (67/101) 66%	$P = 0.0007$	[34]
rs6983267; chr8:127401060	CRC (cell lines)	84	G, loss (33/84) 39%	T, loss (51/84) 61%	$P = 0.05$	[35]
rs6983267; chr8:127401060	CRC	48	G, loss (15/48) 31%	T, loss (33/48) 69%	$P = 0.03$	[28]
rs2273535; chr20:56386485	CRC	23	A, gain (19/23) 83%	T, gain (4/23) 17%	$P = 0.018$	[36]
rs2273535; chr20:56386485	CRC	54	A, gain (38/54) 70%	T, gain (16/54) 30%	$P = 0.03$	[41]
rs13281615; chr8:127343372	cSCC	35	A, gain (28/35) 80%	G, gain (7/35) 20%	$P = 0.012$	[43]
rs6959338; chr7:112077957	Glioblastoma	41	T, gain (33/41) 80%	C, gain (8/41) 20%	$P = 1.1\times10^{-4}$	[42]
rs13222385; chr7:55183900	Glioblastoma	45	G, gain (35/45) 78%	A, gain (10/45) 22%	$P = 2.5\times10^{-4}$	
rs4367471; chr7:104742426	Glioblastoma	24	Minor[c], gain (20/24) 83%	Major[c], gain (4/24) 17%	$P = 0.0015$	
rs4132013; chr7:104762262	Glioblastoma	32	Minor[c], gain (27/32) 84%	Major[c] gain (5/32) 16%	$P = 0.00011$	
rs12343867; chr9:5074189	Myeloproliferative neoplasm	109	C, gain (93/109) 85%	T, gain (16/109) 15%	$P = 5.7\times10^{-6}$	[44]

[a]Genomic locations obtained from NCBI dbSNP, Human assembly GRCh38/hg38.
[b]In all cases, the risk allele showed relative preferential gain or the nonrisk allele showed relative preferential loss.
[c]'Minor' and 'Major' refer to the minor and major allele frequency in Caucasian populations, respectively.

Genomic analyses of glioblastoma have also identified selective amplification of germline variants associated with increased cancer risk. Researchers analyzed 178 glioblastoma tumors from The Cancer Genome Atlas (TCGA) and found selectively amplified SNVs in kinase-encoding genes, including *AGK, DGKB, EGFR, INSR, KIT*, and *RELN* [42]. This group also compared their list of 139 amplified SNVs with 406 SNVs identified in a glioblastoma GWAS, with the rationale that a selectively amplified SNV in glioblastoma may predispose the carrier to tumor initiation and, therefore, occur at higher frequency in cases versus controls. Two SNVs, *rs4367471* and *rs4132013,* demonstrated amplification of the risk alleles over the nonrisk alleles in glioblastomas of germline heterozygotes (Table 2). Two SNVs, *rs6959338* in *DOCK4* at 7q31.1 and *rs13222385* in *EGFR* at 7p11.2, also showed preferential amplification of the risk over the nonrisk allele in the tumor DNA of germline heterozygotes (Table 2). Samples with amplification of the selected-for allele had higher gene expression of *DOCK4* and *EGFR*. Interestingly, *LANCL2*, located near *EGFR*, also exhibited higher expression in *rs13222385* heterozygotes with amplified risk allele (Table 1). Thus, for some variants, preferential allelic gain in tumors may reflect selection based on their role in the regulation of gene expression. This study illustrates the power of having multiple data sets (germline SNV, tumor exomes, and tumor gene expression) for making biological connections.

While the aforementioned studies were inherently limited to testing single tumors from individuals, researchers have utilized solid-organ transplant recipients to better address the role of germline DNA on genetic alterations. Solid-organ transplant recipients are distinctive in that some will develop multiple independent cutaneous squamous cell carcinomas (cSCCs) in the context of an unchanging constitutional genome. This provides a uniquely powerful opportunity to test the

hypothesis that an individual's germline DNA influences the types of genetic alterations that promote cancer initiation and growth. Copy number alterations from individuals with multiple cSCCs revealed higher concordance of chromosomal aberrations within an individual compared with across unrelated individuals, and *rs13281615* showed evidence of ASI in this cohort [43] (Table 2). Thus, genetic background affects the pattern of specific somatic alterations that predisposes individuals to cSCC initiation or progression.

In addition to SNVs being linked to copy number alterations in *cis*, associations between SNVs and somatic mutations within the genes harboring the variants have been detected. A GWAS for myeloproliferative neoplasms associated *rs12343867* in *JAK2* with *JAK2*^{V617F} somatic mutations. Among individuals heterozygous for *rs12343867*, a significant proportion carried the risk (C) allele for *JAK2*^{V617F} mutation [44] (Table 2). Similar findings in a non-small cell lung cancer (NSCLC) study demonstrated that germline variants in *EGFR* were associated with somatic *EGFR* mutations, particularly exon 19 deletions, and that these showed evidence of ASI [45]. In a case-control study of 141 patients with NSCLC, two SNVs in *EGFR*, *rs45559542* and *rs712829*, were observed at a higher frequency in patients with exon 19 deletions. A similar observation was seen in *rs712829* heterozygous cell lines. These data demonstrate that variants can contribute to ASI for oncogenes such as *EGFR*. As another example, analysis of TCGA data for 127 significantly mutated genes across major cancers determined that SNVs may be predictive of RCC risk and predictors of clinical outcomes. Using data from over 650 individuals with RCC, an association was observed between the risk (C) allele of the *ERBB4* intronic variant *rs10932384* and *ERBB4* mutation [46] (Table 1). The same variant was also associated with recurrence and overall survival, suggesting that G×M associations predict recurrence and clinical outcomes.

Germline Variant by Somatic Mutation Association in *trans*

While GWAS have identified hundreds of cancer risk variants [47], few studies have identified G×M associations linking germline variants with distal somatic mutations [48] (Figure 2B). Studies conducting exome sequencing of tumor tissue and genome-wide SNV genotyping of the germline DNA from the same individual can be used to identify some G×M associations but miss intergenic SNVs, where most risk alleles are located. These G×M associations refer to *trans* effects of germline variants, in which the variants influence somatic mutations at distant, and even seemingly unrelated, loci. Significantly, G×M studies have the potential to illuminate unknown biological connections tethering the germline genome to the tumor genome.

One of the first GWAS to systematically test for associations between germline variants and somatic events in human tumors used TCGA data from nearly 6000 cancer cases [49]. The hypothesis was that germline background establishes a context in which a loss- or gain-of-function event in a particular gene may confer a selective growth advantage. G×M analyses of 138 frequently mutated cancer genes identified 62 associations. Validation studies revealed 28 germline loci that were associated with increased somatic alterations in 20 cancer-related genes. For example, the intronic variant *rs8051518* in *RBFOX1* (mapping to 16p13.3) was associated with an eightfold increased incidence of somatic mutations in *SF3B1* (2q33.1) (Table 1). *SF3B1* and *RBFOX1* encode RNA-binding proteins involved in splicing, suggesting a biological rationale for the observed association. Another G×M interaction was observed between *rs25673* (19p13.13) and *PTEN* (10q23.31); individuals with the risk allele of *rs25673* were fourfold more likely to have *PTEN* mutations (Table 1). Two genes at the 19p13.13 locus, *GNA11* and *STK11*, are involved in the PIK3CA/mTOR pathway for which PTEN is a negative regulator. Therefore, *rs25673* could increase the selective advantage of inactivating *PTEN* mutations in cancer progression.

Other studies have identified G×M interactions that provide evidence for risk SNVs influencing somatic events in tumorigenic signaling pathways. In lung cancer, SNVs associate with somatic

driver gene mutations or copy number alterations in *ARID1A*, *CDKN2A*, and genes linked to the cell cycle and MAPK pathways [14] (Table 1). In gastric cancer, several germline SNVs associate with somatic alterations in *SOS1* and other genes in the PDGF and DNA mismatch repair pathways [50] (Table 1). In a cohort of patients with ER-positive breast cancer, G×M analysis found associations of two SNVs, *rs252913* and *rs331499* (5q11.2), with somatic *PIK3CA* variants (3q26.32) [51] (Table 1), which resulted in *MAP3K1* gene overexpression. While *PIK3CA* mutations are frequent in breast cancer, clinical trials of agents targeting this mutation have produced disappointing results [52]. Future studies may capitalize on new insights from G×M interactions to identify altered signaling pathways for which precision therapies will be effective.

While much of the literature has described how germline variants are associated with specific genes in the mutation landscape of cancer, some variants are associated with specific mutations in these genes. A mouse lung cancer susceptibility locus, *Pas1*, which maps near *Kras,* showed ASI of the chromosome from the tumor-susceptible strain [27]. A recent study tested the association of germline variants with the frequency and type of *Kras* mutation in mouse lung tumorigenesis using two chemically inducible lung tumor mouse intercross models. In both models, *Kras* codon 61 was the most frequently mutated codon, and somatic mutations were associated with germline markers at the same chromosome 19 locus. However, two different SNVs were associated with two different *Kras* codon 61 mutations: $Kras^{Q61R}$ with *rs3655407* and $Kras^{Q61L}$ with *rs13483612*. Fine-mapping studies found that $Kras^{Q61L}$ was associated with *rs13459194* in mice homozygous for the allele from the susceptible parent, but this was not observed for the $Kras^{Q61R}$ mutation [29] (Table 1). These studies illustrate that different somatic mutations within the same gene and even affecting the same codon can occur in the context of specific germline variants.

Cutaneous melanomas are categorized by four genomic subtypes: mutant *BRAF*, mutant *RAS* (mainly *NRAS*), mutant *NF1,* and triple-wild-type. In one study, germline DNA from 1223 participants was genotyped for 47 SNVs previously associated with melanoma risk [53]. Melanomas were also interrogated for *BRAF* and *NRAS* mutations. G×M analyses found two SNVs significantly associated with *BRAF* mutations. The *rs12203592* (T) allele of *IRF4* was associated with decreased risk of $BRAF^{V600E}$ and $BRAF^{V600K}$ mutations, but an increased likelihood of other *BRAF* exon 15 mutations (Table 1). The *rs132985* (T) allele of *PLA2G6* was associated with increased likelihood of $BRAF^{V600E}$ and other *BRAF* exon 15 mutations (Table 1) [53].

Germline *MC1R* variants also influence the somatic mutational landscape of melanoma. These variants are associated with red hair, freckling, and sun sensitivity. Melanomas from individuals with *MC1R* variants had a higher somatic mutational burden than from individuals without *MC1R* variants [54]. Furthermore, $BRAF^{V600K}$ somatic variants were less likely to occur in individuals with *MC1R* pigment-related germline variants (D84E, R142H, R151C, R160W and D294H). These same *MC1R* variants were associated with the presence of $BRAF^{V600E}$ PVs, but only in individuals with darker eye and hair color [55]. Collectively, these data suggest that pigment phenotypes, or genes determining these phenotypes, are associated not only with what genes are somatically mutated, but also with the precise mutation ($BRAF^{V600K}$ versus $BRAF^{V600E}$).

Germline Variant by Somatic Mutation Associations and Tumor Immune Response

There are emerging connections between germline variants in immune-related genes and somatic mutations in tumors. Correlations among *MHC-I* (*HLA)* genotypes and a subset of 1000 common driver mutations for cancer were tested in 9000 patients with cancer from TCGA. This study found that patient-specific *HLA* variants were associated with which oncogenic mutations pass undetected by the immune system and allow clonal selection. This powerful study revealed that an individual's *HLA* genotype may predict which driver mutations are likely to occur in the tumors of

specific patients [56]. In general, across individuals with assorted *HLA* variations, *BRAF*^{V600E} mutation is associated with weak immune presentation and higher population mutation frequency, while *IDH1*^{R132C} mutation is associated with strong immune presentation and lower population mutation frequency. Other somatic mutations, such as *PIK3CA*^{E545K}, *PIK3CA*^{H1047R}, *KRAS*^{G12D}, and *KRAS*^{G12V}, are more frequent in individuals with specific *MHCI* genotypes (Figure 1). Patient-specific *MHCII* genotypes influence the somatic mutational landscape and antitumor response similar to *MHCI* [57]. In another G×M study, *rs351855* near *FGFR4* was associated with increased STAT3 signaling, poor prognosis, and increased progression in multiple cancers [58]. Knock-in mice with the *rs351855* risk allele showed a suppressed CD8/CD4 regulatory T cell ratio, decreased tumor infiltration of CD8 T cells, and increased STAT3 signaling *in vivo*. With a better understanding of the role of the innate immune system in the recognition of specific driver mutations, we may be able to predict which somatic mutations an individual's tumor is more likely to contain. Collectively, these data suggest a link between germline variants and immune evasion or suppressed antitumor response during cancer progression.

Concluding Remarks and Future Directions

Integrating germline and somatic tumor genome data provides insight into pathways and molecular mechanisms important for tumorigenesis. Just as tumors develop resistance mutations or have clonal expansions of rare cell populations during therapeutic intervention, an individual's

Outstanding Questions

Are G×M interactions cancer type specific or common across multiple tumor types?

Can identification of G×M interactions aid in distinguishing driver somatic mutations from passenger mutations?

Will characterization of the mechanisms leading to specific G×M interactions provide insight into new therapeutic strategies?

Are histological and pathological subtypes of tumors largely driven by the genetic background of an individual?

EGFR mutations are more common in lung tumors from individuals of Asian ancestry than in lung tumors from individuals of European ancestry. Are differences in the frequency of somatic mutations observed between racial and ethnic groups associated with alleles that show differences in germline variant frequency between these groups?

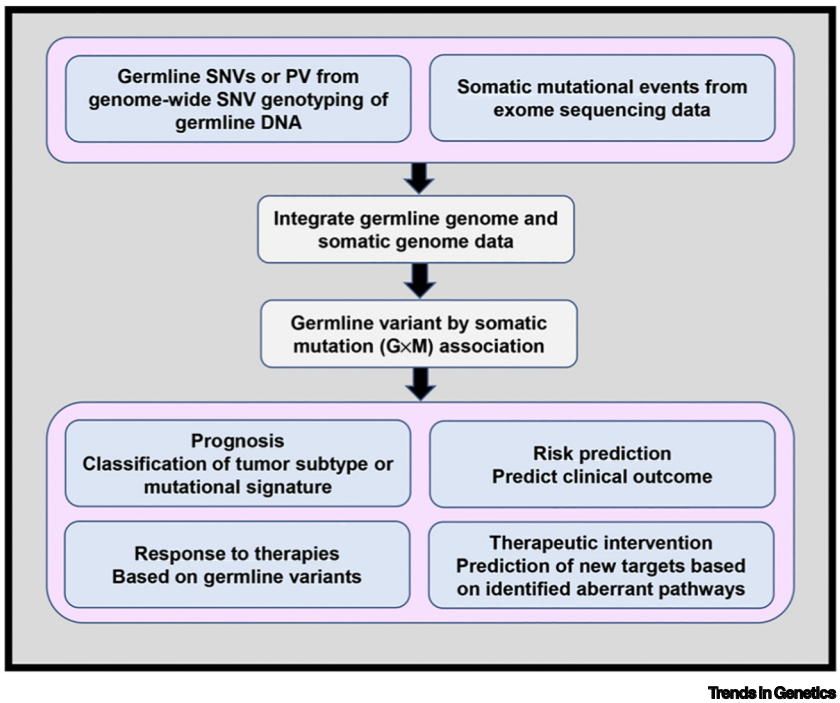

Figure 3. Summary of Germline Variants by Somatic Mutation (G×M) Association Studies and Clinical Application. Integration of data from exome sequencing of tumor DNA and genome-wide single nucleotide variant (SNV) genotyping of germline DNA is revealing associations between specific germline variants and somatic events that influence mutational profiles and cancer subtypes. This information may predict treatment response, clinical outcomes, and identification of novel therapeutic pathways. For example, *BRCA1* pathogenic variants (PVs) are associated with more aggressive breast cancers (triple-negative and basal-like) and poorer prognosis. Tumors with PVs in *BRCA1* are more likely to contain small tandem duplications and deletions, reflecting perturbation of homology-directed DNA repair function. Tumors with disrupted DNA repair are more responsive to chemotherapeutic agents that block other means to repair DNA damage, such as poly ADP ribose polymerase (PARP) inhibitors.

germline genetic background can lead to particular mutational profiles and/or may impart an early selective pressure fostering environments in which the particular somatic events may be more likely to expand or escape normal cell controls. Due to the paucity of studies integrating the germline and somatic genomes, and given that this area of research is still in its infancy, many such associations and mechanisms of these associations remain undiscovered (see Outstanding Questions). It also remains undetermined whether findings from G×M studies will inform individual response to therapies beyond PVs, such as *BRCA1* and *MLH2*. However, given the connections described here, this research direction shows promise. Future studies evaluating G×M associations may also lead to the prediction of tumor subtype or prognosis even before an individual is found to have cancer, which could additionally inform prevention strategies (Figure 3).

References

1. Spurdle, A.B. *et al.* (2014) Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large–scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast Cancer Res.* 16, 3419
2. Mulligan, A.M. *et al.* (2011) Common breast cancer susceptibility alleles are associated with tumour subtypes in BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2. *Breast Cancer Res.* 13, R110
3. Stevens, K.N. *et al.* (2011) Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res.* 71, 6240–6249
4. Stevens, K.N. *et al.* (2013) Genetic susceptibility to triple-negative breast cancer. *Cancer Res.* 73, 2025–2030
5. Milne, R.L. *et al.* (2014) A large-scale assessment of two-way SNP interactions in breast cancer susceptibility using 46,450 cases and 42,461 controls from the breast cancer association consortium. *Hum. Mol. Genet.* 23, 1934–1946
6. Couch, F.J. *et al.* (2017) Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol.* 3, 1190–1196
7. Bonifaci, N. *et al.* (2010) Exploring the link between germline and somatic genetic alterations in breast carcinogenesis. *PLoS One* 5, e14078
8. Dunning, A.M. *et al.* (2016) Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nat. Genet.* 48, 374–386
9. Barrisford, G.W. *et al.* (2011) Familial renal cancer: molecular genetics and surgical management. *Int. J. Surg. Oncol.* 2011, 658767
10. Putti, T.C. *et al.* (2005) Estrogen receptor-negative breast carcinomas: a review of morphology and immunophenotypical analysis. *Mod. Pathol.* 18, 26–35
11. Nik-Zainal, S. *et al.* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54
12. Dziadkowiec, K.N. *et al.* (2016) PARP inhibitors: review of mechanisms of action and BRCA1/2 mutation targeting. *Prz Menopauzalny* 15, 215–219
13. Waszak, S.M. *et al.* (2017) Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. bioRxiv 208330
14. Wang, Y. *et al.* (2018) Interaction analysis between germline susceptibility loci and somatic alterations in lung cancer. *Int. J. Cancer* 143, 878–885
15. Middlebrooks, C.D. *et al.* (2016) Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat. Genet.* 48, 1330–1338
16. Nik-Zainal, S. *et al.* (2014) Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* 46, 487–491
17. Planck, M. *et al.* (2002) High frequency of microsatellite instability and loss of mismatch-repair protein expression in patients with double primary tumors of the endometrium and colorectum. *Cancer* 94, 2502–2510
18. Planck, M. *et al.* (2003) Microsatellite instability and expression of MLH1 and MSH2 in carcinomas of the small intestine. *Cancer* 97, 1551–1557
19. Wahlberg, S.S. *et al.* (2002) Evaluation of microsatellite instability and immunohistochemistry for the prediction of germ-line MSH2 and MLH1 mutations in hereditary nonpolyposis colon cancer families. *Cancer Res.* 62, 3485–3492
20. Vilar, E. and Gruber, S.B. (2010) Microsatellite instability in colorectal cancer-the stable evidence. *Nat. Rev. Clin. Oncol.* 7, 153–162
21. Gryfe, R. *et al.* (2000) Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N. Engl. J. Med.* 342, 69–77
22. Knudson Jr., A.G. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* 68, 820–823
23. Yu, C.C. *et al.* (2017) Mutant allele specific imbalance in oncogenes with copy number alterations: Occurrence, mechanisms, and potential clinical implications. *Cancer Lett.* 384, 86–93
24. Malapelle, U. *et al.* (2015) KRAS mutant allele-specific imbalance (MASI) assessment in routine samples of patients with metastatic colorectal cancer. *J. Clin. Pathol.* 68, 265–269
25. Hartman, D.J. *et al.* (2012) Mutant allele-specific imbalance modulates prognostic impact of KRAS mutations in colorectal adenocarcinoma and is associated with worse overall survival. *Int. J. Cancer* 131, 1810–1817
26. Nagase, H. *et al.* (2003) Allele-specific Hras mutations and genetic alterations at tumor susceptibility loci in skin carcinomas from interspecific hybrid mice. *Cancer Res.* 63, 4849–4853
27. Manenti, G. *et al.* (1999) Linkage disequilibrium and physical mapping of Pas1 in mice. *Genome Res.* 9, 639–646
28. Gerber, M.M. *et al.* (2012) Evaluation of allele-specific somatic changes of genome-wide association study susceptibility alleles in human colorectal cancers. *PLoS One* 7, e37672
29. Borrego, A. *et al.* (2018) Germline control of somatic Kras mutations in mouse lung tumors. *Mol. Carcinog.* 57, 745–751
30. Nagase, H. *et al.* (1999) A subset of skin tumor modifier loci determines survival time of tumor-bearing mice. *Proc. Natl. Acad. Sci. U. S. A.* 96, 15032–15037
31. Tomlinson, I. *et al.* (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* 39, 984–988
32. Zanke, B.W. *et al.* (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 39, 989–994
33. Haiman, C.A. *et al.* (2007) Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.* 39, 638–644
34. Tuupanen, S. *et al.* (2008) Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. *Cancer Res.* 68, 14–17
35. Sugimachi, K. *et al.* (2014) Allelic imbalance at an 8q24 oncogenic SNP is involved in activating MYC in human colorectal cancer. *Ann. Surg. Oncol.* 21, S515–S521
36. Ewart-Toland, A. *et al.* (2003) Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat. Genet.* 34, 403–412
37. Ewart-Toland, A. *et al.* (2005) Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. *Carcinogenesis* 26, 1368–1373
38. Sun, H. *et al.* (2011) Lack of an association between AURKA T91A polymorphisms and breast cancer: a meta-analysis involving 32,141 subjects. *Breast Cancer Res. Treat.* 125, 175–179

39. Dai, Z.J. *et al.* (2014) Association between genetic polymorphisms in AURKA (rs2273535 and rs1047972) and breast cancer risk: a meta-analysis involving 37,221 subjects. *Cancer Cell Int.* 14, 91
40. Tang, W. *et al.* (2013) Association between the STK15 F31I polymorphism and cancer susceptibility: a meta-analysis involving 43,626 subjects. *PLoS One* 8, e82790
41. Hienonen, T. *et al.* (2006) Preferential amplification of AURKA 91A (Ile31) in familial colorectal cancers. *Int. J. Cancer* 118, 505–508
42. LaFramboise, T. *et al.* (2010) Allelic selection of amplicons in glioblastoma revealed by combining somatic and germline analysis. *PLoS Genet.* 6, e1001086
43. Dworkin, A.M. *et al.* (2010) Germline variation controls the architecture of somatic alterations in tumors. *PLoS Genet.* 6, e1001136
44. Olcaydu, D. *et al.* (2009) A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat. Genet.* 41, 450–454
45. Liu, W. *et al.* (2011) Functional EGFR germline polymorphisms may confer risk for EGFR somatic mutations in non-small cell lung cancer, with a predominant effect on exon 19 microdeletions. *Cancer Res.* 71, 2423–2427
46. Shu, X. *et al.* (2018) Germline genetic variants in somatically significantly mutated genes in tumors are associated with renal cell carcinoma risk and outcome. *Carcinogenesis* 39, 752–757
47. Chung, C.C. and Chanock, S.J. (2011) Current status of genome-wide association studies in cancer. *Hum. Genet.* 130, 59–78
48. Lu, Y. *et al.* (2014) Most common 'sporadic' cancers have a significant germline genetic component. *Hum. Mol. Genet.* 23, 6112–6118
49. Carter, H. *et al.* (2017) Interaction landscape of inherited polymorphisms with somatic events in cancer. *Cancer Discov.* 7, 410–423
50. Zhang, X. *et al.* (2018) Germline genetic variants were interactively associated with somatic alterations in gastric cancer. *Cancer Med.* 7, 3912–3920
51. Puzone, R. and Pfeffer, U. (2017) SNP variants at the MAP3K1/SETD9 locus 5q11.2 associate with somatic PIK3CA variants in breast cancers. *Eur. J. Hum. Genet.* 25, 384–387
52. Janku, F. *et al.* (2018) Targeting the PI3K pathway in cancer: are we making headway? *Nat. Rev. Clin. Oncol.* 15, 273–291
53. Thomas, N.E. *et al.* (2018) Inherited genetic variants associated with melanoma BRAF/NRAS subtypes. *J. Invest. Dermatol.* 138, 2398–2404
54. Robles-Espinoza, C.D. *et al.* (2016) Germline MC1R status influences somatic mutation burden in melanoma. *Nat. Commun.* 7, 12064
55. Thomas, N.E. *et al.* (2017) Associations of MC1R genotype and patient phenotypes with BRAF and NRAS mutations in melanoma. *J. Invest. Dermatol.* 137, 2588–2598
56. Marty, R. *et al.* (2017) MHC-I genotype restricts the oncogenic mutational landscape. *Cell* 171, 1272–1283
57. Marty Pyke, R. *et al.* (2018) Evolutionary pressure against MHC class II binding cancer mutations. *Cell* 175, 1991
58. Kogan, D. *et al.* (2018) STAT3-enhancing germline mutations contribute to tumor-extrinsic immune evasion. *J. Clin. Invest.* 128, 1867–1872