PERSPECTIVE OPEN

Functional germline variants as potential co-oncogenes

Divyansh Agarwal [1,2], Christoph Nowak [3,4], Nancy R. Zhang[1,2], Lajos Pusztai[5] and Christos Hatzis[5]

Germline variants that affect the expression or function of proteins contribute to phenotypic variation in humans and likely determine individual characteristics and susceptibility to diseases including cancer. A number of high penetrance germline variants that increase cancer risk have been identified and studied, but germline functional polymorphisms are not typically considered in the context of cancer biology, where the focus is primarily on somatic mutations. Yet, there is evidence from familial cancers indicating that specific cancer subtypes tend to arise in carriers of high-risk germline variants (e.g., triple negative breast cancers in mutated BRCA carriers), which suggests that pre-existing germline variants may determine which complementary somatic driver mutations are needed to drive tumorigenesis. Recent genome sequencing studies of large breast cancer cohorts reported only a handful of highly recurrent driver mutations, suggesting that different oncogenic events drive individual cancers. Here, we propose that germline polymorphisms can function as oncogenic modifiers, or co-oncogenes, and these determine what complementary subsequent somatic events are required for full malignant transformation. Therefore, we propose that germline aberrations should be considered together with somatic mutations to determine what genes drive cancer and how they may be targeted.

INTRODUCTION

The genome of two individuals differs from one another, and from the reference human genome, in 20,000–25,000 non-synonymous single nucleotide polymorphisms (SNPs) in coding regions that affect the functions of hundreds of genes. If noncoding regions are included, up to 3.5–4.3 million SNPs show inter-individual variability.[1,2] Phenotypically normal individuals can also carry several small and large scale structural DNA alterations that affect up to 0.7% of their genome.[2] Collectively, the germline functional variants that individuals carry determine their unique features and influence susceptibility to disease. Cells in self-renewing tissues accumulate somatic mutations over time as a result of natural growth, and can evolve to malignant lesions when acquiring an initiating or driver mutation. These cancer initiating somatic mutations arise in the context of individual germline DNA polymorphisms and thus may be different for different cancers. Most genetic variants detected in any given cancer are germline variants, with solid tumors harboring only a few dozen to a few hundred additional somatic mutations that are considered potentially important in cancer biology.[3,4]

The causal implication of many germline variants in Mendelian diseases has long been recognized (http://omim.org/).[5] Numerous cancer predisposing germline variants have also been identified in tumor suppressor genes (TP53, RB), in genes involved in DNA repair (e.g., BRCA1, PALB2, MLH1, MSH2/6, CHEK2, and ATM), cell proliferation (e.g., PTEN, STK11, RET, and FGFR) and cell adhesion (e.g., CDH1, APC).[6–9] However, only a small proportion of breast cancer risk can be attributed to high penetrance germline mutations, with each individual SNP contributing only marginally to the risk of cancer. This suggests that the impact of any given variant on disease risk depends on co-occurrence of functional variants in different genes possibly acting in the same pathway, and that additional acquired somatic mutations (or epigenetic events) are required for initiation of malignant transformation.[10] This phenomenon can even be observed in individuals who carry high-penetrance germline mutations such as BRCA1/2 carriers,[9,11] in which additional germline variants can substantially influence the cancer risk conferred by a given germline BRCA mutation.[10] The increasing availability of well-powered prospective population studies such as the UK Biobank of >500,000 persons (www.ukbiobank.ac.uk/) is likely to identify additional low-risk variants for cancer.

An unexpected observation from large-scale cancer sequencing efforts was the relative rarity of high frequency, recurrent somatic mutations in common solid tumors.[12] In total, 37% of breast cancers carry somatic mutations in TP53 and 36% in PIK3CA, the two most-frequently mutated genes, while most cancers display a variable assortment of low-frequency mutations in unique combinations. Yet, the majority (80%) of basal-like cancers have a TP53 mutation compared to 9% with a PIK3CA mutation, but only 12% of liminal A cancers have TP53 mutations compared to 45% who have PIK3CA mutations, suggesting a strong association between driver mutations and cancer subtype. A possible explanation could be that the effect of different mutations converges at the biological pathway level. Distinct genes may be affected in different individuals but the net biological impact is the dysregulation of an oncogenic pathway or biological process through the combined action of several mutations.[13] Furthermore, we suggest that functional germline polymorphisms could in effect function as pre-existing driver "hits", which together with complementary somatic mutations acquired at a later time act to

[1]Department of Genomics and Computational Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [2]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA; [3]Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala, Sweden; [4]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Sweden and [5]Department of Medicine, Breast Medical Oncology, Yale School of Medicine, Yale University, New Haven, CT, USA
Correspondence: Lajos Pusztai (lajos.pusztai@yale.edu) or Christos Hatzis (christos.hatzis@yale.edu)
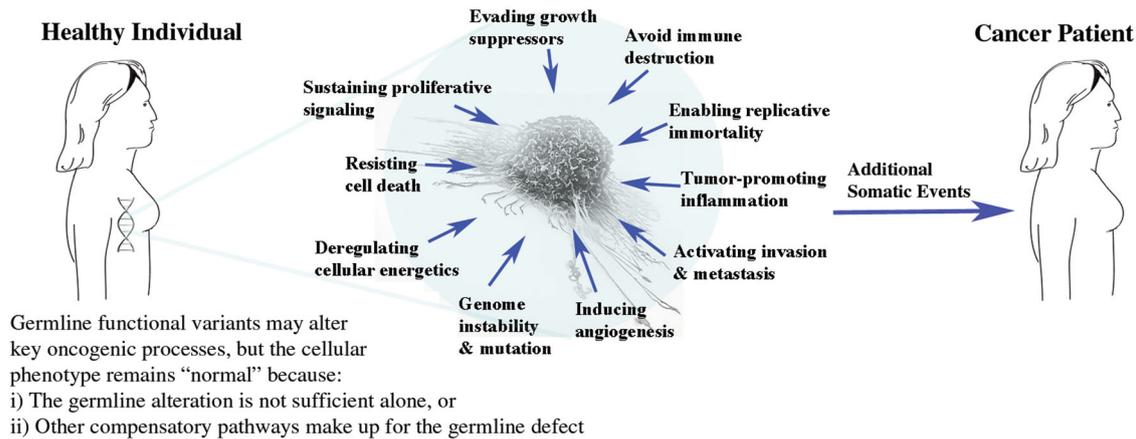
npj nature partner journals   BCRF

Fig. 1 Germline variants may act as co-oncogenes in the dysregulation of biological processes that are the "hallmarks of cancer"

dysregulate key pathways and to trigger full malignant transformation.

This hypothesis is a conceptual extension of Alfred Knudson's two-hit model of a single oncogenic event.[14] Knudson postulated that one germline and one somatic hit are required to disable both alleles of a tumor suppressor gene, which then leads to cancer formation.[14,15] This hypothesis could be extended beyond consideration of only two alleles in the same gene to the dysregulation of oncogenic pathways.[13] Similarly, activation of only one copy of an proto-oncogene may be required for carcinogenesis through a gain-of-function mutation or through amplification of the locus, which can occur potentially in concert with cooperative effects in the germline DNA of the oncogene or of related genes.

We therefore propose that key biological processes that are required for breast carcinogenesis[13] can be altered through effects from the germline DNA, but that these events alone are not sufficient to initiate malignant transformation. Such germline abnormalities may remain silent because of compensatory pathways in normal tissues and lack of dysregulation in other complementary "hallmarks" of cancer until acquired somatic mutations disable these compensatory pathways or activate complementary oncogenic processes (Fig. 1). What constitutes a somatic cancer driver event in a particular patient is conditional on the collection of functional polymorphisms already present in her/his germline DNA. This hypothesis is consistent with the widely observed, but rarely discussed, cell-line-restricted trans-forming effects of many known oncogenes.[16,17]

EXPERIMENTAL DATA SUPPORTING FUNCTIONAL COOPERATION BETWEEN GERMLINE SNPS AND SOMATIC MUTATIONS

There is a growing list of germline variants that have been shown experimentally to impact protein function without having a detectable association with disease risk, with several of these affecting genes implicated in cancer. For example, the M326I variant (rs3730089, variant allele frequency [VAF] of the A-allele 22%) in the p85α regulatory subunit of phosphatidylinositol 3-kinase (PI3K) results in constitutively increased activity of the PI3K pathway.[18] Similarly, the L1016S variant (rs61733127, C-allele VAF 9%) in the PH domain and leucine-rich repeat protein phosphatase 2 (PHLPP2) leads to reduced phosphatase activity and increased signaling through the protein kinase B (AKT) and protein kinase C (PKC).[19]

The T1010I (rs56391007, T-allele VAF 1%) germline variant in MET (hepatocyte growth factor receptor) increases colony formation, cell migration, invasion, and in vivo tumor growth

when introduced into MCF-10A immortalized breast epithelial cell line.[20] Germline loss-of-function variants in PALB2 (partner and localizer of BRCA2) lead to predisposition to breast and pancreatic cancer. In mouse models, deletion of PALB2 led to mammary tumor formation with long latency. However, co-deletion of TP53 resulted in accelerated tumor formation, providing indirect support for synergistic interactions between a germline and a frequently observed somatic mutation.[21]

A recent analysis of germline and somatic DNA data from nearly 7000 patients from the cancer genome atlas (TCGA) identified almost 400 germline SNPs that were significantly associated with tumors of a particular type compared to all other cancer types.[22] Fifteen of these markers had also been previously identified in GWAS studies, and listed in the National Human Genome Research Institute GWAS catalog, as being associated with significant cancer risk compared to nondiseased controls. This suggests that these germline loci are not only associated with significantly elevated cancer risk, but they also appear to influence the type and biology of the cancer that emerges. Furthermore, 17 germline loci were significantly associated with the somatic alteration rate of known cancer genes. These loci were not proximal to the affected cancer genes but converged to the same biological pathways,[22,23] providing direct evidence for our hypothesis.

POPULATION-BASED INDIRECT EVIDENCE

Past studies have shown that several sporadic cancers have a significant inherited component, and demonstrated genome-wide SNP relatedness in twelve cancer types.[24] Based on the common variants of modest effect that influence breast cancer risk, polygenic risk scores using reported susceptibility loci from population-based studies were developed, and their associations with risks for BRCA1 and BRCA2 carriers were evaluated.[25–27] These polygenic predictors suggest that polygenic risk scores based on common variants could provide clinically useful risk stratification and management of BRCA mutation carriers, and may provide a starting basis for biological interpretation of the confluence of genetic events that drive cancer.

Prognostic and predictive marker research has focused on somatic alterations found in cancer cells and on alterations in the tumor microenvironment. Recent large population studies, and particularly those based on Mendelian randomization (MR) designs to infer causal relationships between correlated variables, have revealed intriguing insights. The MR approach[28] uses common genetic variants associated with an exposure as instruments to test for causal effects on an outcome. Since parental alleles are essentially allocated at random during meiosis, this "quasi-randomization" before birth minimizes bias from

Table 1. Significant associations between germline variants and somatic mutations detected in the TCGA breast cancer data at a false discovery rate ≤12%

Somatic gene	Germline variant					Two-sided Fisher exact test p-value	Direction
	rsID	Gene	SNP type	VAF 1000 genomes	VAF TCGA		
Luminal A ($N = 210$)							
SSPO	rs45551636	PALB2	Missense	0.008	0.028	4×10^{-6}	Co-mutation
MACF1	rs3923647	TLR1	Missense	0.029	0.025	4×10^{-6}	Co-mutation
Luminal B ($N = 115$)							
USH2A	rs8140287	ISX	Missense	0.031	0.045	9×10^{-5}	Co-mutation
ATP10B	rs2273137	NOP56	Missense	0.118	0.084	8×10^{-5}	Co-mutation
PIK3CA	rs12099177	MMP27	Missense	0.168	0.081	8×10^{-5}	Co-exclusion
SYNE1	rs34605667	MPDZ	Missense	0.014	0.037	5×10^{-5}	Co-mutation
Basal ($N = 54$)							
USH2A	rs17848337	SREBF2	Synonymous	0.051	0.051	1×10^{-4}	Co-mutation
TP53	rs55695858	OBP2A	Missense	0.173	0.224	1×10^{-4}	Co-mutation
HER2-enriched ($N = 90$)							
EYS	rs2066518	SMARCAL1	Missense	0.073	0.026	2×10^{-4}	Co-mutation
PIK3CA	rs3774372	ULK4	Missense	0.173	0.176	1×10^{-4}	Co-mutation

The germline variants included in this analysis were the 8598 CADD* predicted-deleterious SNPs (C-score ≥ 20). Only somatic mutations in canonical cancer genes listed in COSMIC and OMIM that had mutations in ≥2% of breast cancer cases were included (1649 mutations in 71 genes). Somatic mutations were aggregated at gene level for statistical comparison with individual SNPs. Molecular subtype annotation was available through the TCGA
(*) Combined annotation-dependent depletion[36]

subsequent environmental factors and reverse causation, and is finding increasing application in polygenic traits.[29] The role of germline variants, whose individually small effects on breast cancer are far from genome-wide significant, is demonstrated, for instance, in evaluation of adult height as a breast cancer risk factor: Zhang and colleagues[30] found a causal effect of genetically predicted increase in height on breast cancer risk both in premenopausal and postmenopausal women, but restricted only to hormone-receptor positive breast cancer. Additionally, recent results from large-scale population studies suggest that prognosis of early-stage breast cancer tends to cluster in families; the survival of first-degree relatives is statistically more similar than expected by chance after adjusting for tumor and treatment characteristics.[31,32] These observations suggest that cancer occurrence and survival are influenced not only by somatic mutations but also by hereditary components whose contribution may be missed in traditional genome-wide association studies.

For the vast majority of SNPs discovered in epidemiologic studies, there is no functional evidence to demonstrate their impact on biological processes and pathways. In principle, by "functional" we refer to any germline variant that is predicted to affect protein function, regardless of the magnitude of altered protein activity, or whether the effect manifests in a clear phenotype or disease. In the illustrative analysis that follows, we focused on genetic variants predicted to have large (i.e., deleterious) effects on protein function. Since most trait-associated SNPs are in noncoding regions of the genome, efforts to identify causal genes through traditional methods such as targeted resequencing and single-gene knock-outs have been challenging. Yet, recent high-throughput, high-specificity gene-editing approaches enabled by CRISPR/Cas9 are expected to dramatically increase the catalog of germline variants with biologically validated function.[33]

We performed an exploratory analysis to examine whether predicted high functional impact germline SNPs co-occur with somatic mutations in different breast cancer subtypes. We filtered the germline SNPs reported in the TCGA breast cancer cohort to retain 8598 germline variants with CADD C-score ≥20,

corresponding to the 1% most deleterious variants in the human genome (details on data processing steps are given in Table 1 notes). To limit the number of comparisons, we tested for co-occurrence of these high functional impact germline variants with somatic mutations occurring only in canonical cancer-causing genes in the catalog of somatic mutations in cancer (COSMIC; http://cancer.sanger.ac.uk)[34] database using a two-sided Fisher exact test. This analysis identified several germline SNP—somatic mutation pairs that co-occurred more frequently than expected by chance, mostly within specific breast cancer subtypes (Table 1). These observations are consistent with the hypothesis of synergistic interactions between functional germline SNPs and rare somatic mutations in oncogenesis.

There are several caveats in this exploratory analysis, including technical issues related to the germline variant calls in the TCGA[35] and the heterogeneity in sequencing depth and quality control filtering between germline and somatic mutation calls. These analyzes will need to be repeated in larger cohorts.

CONCLUSIONS

Most genomic analyzes of sporadic cancers discard the vast number of germline variants that are detected in a cancer as functionally irrelevant. Ignoring the biological contribution of these variants may lead to an overly simplistic view of the cancer genome. Recent data provide evidence suggesting that neoplastic phenotypes can emerge through cooperation between functionally important germline loci and somatic alterations. We propose that this is a wide-spread phenomenon where tolerated germline SNPs can act as co-oncogenes or co-tumor suppressors and functionally synergize with somatic mutations to induce the neoplastic phenotype. Biological pathways can be altered at multiple different levels and alterations in different member genes could cause similar phenotypic effects. Therefore, the possible combination of synergistic germline and somatic mutations is vast. This could explain the relative lack of high frequency recurrent somatic events in most solid tumors and would also suggest that, when studied individually, co-oncogenic SNPs may

not be identified as cancer predisposing variants. This hypothesis is testable in the laboratory; for instance, Table 1 reports germline SNP-somatic mutation pairs whose functional synergy could be evaluated experimentally. More nuanced understanding of interactions between germline and somatic genetic events could fundamentally change our conceptions about breast cancer biology and breast cancer prevention.

AUTHOR CONTRIBUTIONS

L.P., C.H. and D.A. conceived the major ideas of this paper and wrote the paper. D.A. analyzed the data and drafted the manuscript. C.N. and N.Z. edited the paper and acquired content. All authors provided final approval of the version to be published.

ADDITIONAL INFORMATION

Competing interests: The authors declare that they have no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
2. Khurana, E. et al. Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).
3. Kan, Z. et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
4. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
5. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM. org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–798 (2015).
6. Agarwal, D. et al. FGF receptor genes and breast cancer susceptibility: results from the Breast Cancer Association Consortium. *Br. J. Cancer* **110**, 1088–1100 (2014).
7. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361, 361e351-352 (2013).
8. Antoniou, A. C., Foulkes, W. D. & Tischkowitz, M. Breast-cancer risk in families with mutations in PALB2. *N. Engl. J. Med.* **371**, 1651–1652 (2014).
9. Hall, J. M. et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
10. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
11. Milne, R. L. & Antoniou, A. C. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. *Ann. Oncol.* **22**, i11–17 (2011).
12. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumors. *Nature* **490**, 61–70 (2012).
13. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
14. Knudson, A. G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* **68**, 820–823 (1971).
15. Friend, S. H. et al. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643–646 (1986).
16. Dimri, G., Band, H. & Band, V. Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res.* **7**, 171–179 (2005).
17. Wazer, D. E., Liu, X. L., Chu, Q., Gao, Q. & Band, V. Immortalization of distinct human mammary epithelial cell types by human papilloma virus 16 E6 or E7. *Proc. Natl Acad. Sci. USA* **92**, 3687–3691 (1995).
18. Almind, K. et al. Characterization of the Met326Ile variant of phosphatidylinositol 3-kinase p85alpha. *Proc. Natl Acad. Sci. USA* **99**, 2124–2128 (2002).
19. Brognard, J., Niederst, M., Reyes, G., Warfel, N. & Newton, A. C. Common polymorphism in the phosphatase PHLPP2 results in reduced regulation of Akt and protein kinase C. *J. Biol. Chem.* **284**, 15215–15223 (2009).
20. Liu, S. et al. Functional consequence of the MET-T1010I polymorphism in breast cancer. *Oncotarget* **6**, 2604–2614 (2015).
21. Bowman-Colin, C. et al. Palb2 synergizes with Trp53 to suppress mammary tumor formation in a model of inherited breast cancer. *Proc. Natl Acad. Sci. USA* **110**, 8632–8637 (2013).
22. Carter, H. et al. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. *Cancer Discov.* **7**, 410–423 (2017).
23. Carter, H. & Ideker, T. Common genetic variation in the germline influences where and how tumors develop. *Mol. Cell Oncol.* **4**, e1302905 (2017).
24. Lu, Y. et al. Most common 'sporadic' cancers have a significant germline genetic component. *Hum. Mol. Genet.* **23**, 6112–6118 (2014).
25. Kuchenbaecker, K. B. et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J. Natl Cancer Inst.* https://doi.org/10.1093/jnci/djw302 (2017).
26. Chatterjee, N. et al. Projecting the performance of risk prediction based on polygenic analyzes of genome-wide association studies. *Nat. Genet.* **45**, 400–405 (2013).
27. Wacholder, S. et al. Performance of common genetic variants in breast-cancer risk models. *N. Engl. J. Med.* **362**, 986–993 (2010).
28. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
29. Nowak, C. et al. Effect of insulin resistance on monounsaturated fatty acid levels: a multi-cohort non-targeted metabolomics and Mendelian randomization study. *PLoS Genet.* **12**, e1006379 (2016).
30. Zhang, B. et al. Height and breast cancer risk: evidence from prospective studies and mendelian randomization. *J. Natl Cancer Inst.* https://doi.org/10.1093/jnci/djv219 (2015).
31. Verkooijen, H. M. et al. Breast cancer prognosis is inherited independently of patient, tumor and treatment characteristics. *Int. J. Cancer* **130**, 2103–2110 (2012).
32. Goode, E. L. et al. Effect of germ-line genetic variation on breast cancer survival in a population-based study. *Cancer Res.* **62**, 3052–3057 (2002).
33. Donovan, K. F. et al. Creation of novel protein variants with CRISPR/Cas9-mediated mutagenesis: turning a screening by-product into a discovery tool. *PLoS One* **12**, e0170445 (2017).
34. Forbes, S. A. et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–811 (2015).
35. Buckley, A. R. et al. Pan-cancer analysis reveals technical artifacts in TCGA germline variant calls. *BMC Genom.* **18**, 458 (2017).
36. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).