Cell Reports
Report

Linkage of DNA Methylation Quantitative Trait Loci to Human Cancer Risk

Holger Heyn,[1,*] Sergi Sayols,[1] Catia Moutinho,[1] Enrique Vidal,[1] Jose V. Sanchez-Mut,[1] Olafur A. Stefansson,[1] Ernest Nadal,[2] Sebastian Moran,[1] Jorunn E. Eyfjord,[3,4] Eva Gonzalez-Suarez,[1] Miguel Angel Pujana,[5] and Manel Esteller[1,6,7,*]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908 Barcelona, Catalonia, Spain
[2]Thoracic Surgery, Department of Surgery, University of Michigan Medical School, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA
[3]The Icelandic Cancer Society, Molecular and Cell Biology Research Laboratory, 101 Reykjavik, Iceland
[4]Department of Medicine, University of Iceland, 101 Reykjavik, Iceland
[5]Breast Cancer and Systems Biology Unit, Translational Research Laboratory, Catalan Institute of Oncology (ICO), Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, 08908 Barcelona, Catalonia, Spain
[6]Department of Physiological Sciences II, School of Medicine, University of Barcelona, 08036 Barcelona, Catalonia, Spain
[7]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain
*Correspondence: hheyn@idibell.cat (H.H.), mesteller@idibell.cat (M.E.)
http://dx.doi.org/10.1016/j.celrep.2014.03.016

SUMMARY

Epigenetic regulation and, in particular, DNA methylation have been linked to the underlying genetic sequence. DNA methylation quantitative trait loci (meQTL) have been identified through significant associations between the genetic and epigenetic codes in physiological and pathological contexts. We propose that interrogating the interplay between polymorphic alleles and DNA methylation is a powerful method for improving our interpretation of risk alleles identified in genome-wide association studies that otherwise lack mechanistic explanation. We integrated patient cancer risk genotype data and genome-scale DNA methylation profiles of 3,649 primary human tumors, representing 13 solid cancer types. We provide a comprehensive meQTL catalog containing DNA methylation associations for 21% of interrogated cancer risk polymorphisms. Differentially methylated loci harbor previously reported and as-yet-unidentified cancer genes. We suggest that such regulation at the DNA level can provide a considerable amount of new information about the biology of cancer-risk alleles.

INTRODUCTION

The epigenetic code is characterized as being inherited through cell division and even transgenerationally. The exact mechanisms that reprogram the DNA methylation landscape during spermatogenesis and embryogenesis have only been partially explored and are the subject of current investigation (Seisenberger et al., 2013). It is likely that a concerted function of the genetic blueprint and DNA binding factors maintains DNA methylation profiles from mother to daughter cells and from parents to their offspring (Rando, 2012). Integration of DNA methylation profiles and the underlying genotypes reveals a close relationship between regional CpG density and DNA methylation levels, with CpG islands usually lacking modifications and CpG-poor regions being mostly hypermethylated (Weber et al., 2005). Because this simplified model cannot be sustained for many hypomethylated regions (Molaro et al., 2011), including enhancers, insulators, and CpG-poor promoters, additional sources evidently help to shape the DNA methylome.

Screening the genomic sequence and epigenomic modifications at high resolution revealed a direct relationship between the underlying genetic sequence and DNA methylation at specific sites and led to the definition of methylation quantitative trait loci (meQTL) that illustrate the tight interplay between the two layers of information (Gibbs et al., 2010; Zhang et al., 2010; Shoemaker et al., 2010; Heyn et al., 2013). Particularly, variations in the genetic code at SNPs were associated with DNA methylation levels at proximal CpG sites. Because these *cis*-acting associations are likely to appear independently of CpG densities an obvious connection is lacking. Therefore, intermediate mediators, such as DNA binding factors or secondary chromatin structures (McDaniell et al., 2010), are suspected to provide a functional link with the genetic and epigenetic code. This relationship between the genotype and epitype (displaying DNA methylation levels at a given cytosine in the genome) has been determined for natural human variation (Bell et al., 2011; Fraser et al., 2012; Heyn et al., 2013), neurological disorders (Gibbs et al., 2010; Zhang et al., 2010), and rheumatoid arthritis (Liu et al., 2013), highlighting its importance for various phenotypes, including those relating to diseases. Despite the obviously great informative potential of meQTLs in various disease contexts, no comprehensive studies have so far addressed their contribution to the biology of cancer cells.

The tight connection between the two layers of information might help explain the hitherto unknown connections

Table 1. Cancer Types Interrogated for Genotype-Epitype Associations

Cancer Type	Type ID	Epitype/Genotype Data		GWAS-SNPs	meQTLs
		Cancer	Healthy		
Bladder urothelial carcinoma	BLCA	153	18	6	1
Breast invasive carcinoma	BRCA	576	78	28	5
Colon adenocarcinoma	COAD	255	38	7	1
Glioblastoma multiforme	GBM	110	0	2	1
Kidney renal clear cell carcinoma	KIRC	281	160	4	2
Liver hepatocellular carcinoma	LIHC	98	47	6	1
Lung adenocarcinoma	LUAD	303	32	5	2
Pancreatic adenocarcinoma	PAAD	49	7	16	3
Prostate adenocarcinoma	PRAD	172	49	23	5
Skin cutaneous melanoma	SKCM	46	0	8	2
Stomach adenocarcinoma	STAD	261	2	2	0
Thyroid carcinoma	THCA	435	56	2	1
Uterine corpus endometrioid carcinoma	UCEC	381	42	0	0
Total		3,120	529	109	23[a]

Number of analyzed samples/SNPs GWAS-SNPs is shown. Cancer-related risk polymorphisms present in GWASdb and interrogated on the analyzed genotyping platform. meQTLs, number of risk alleles with significant association to differential CpG methylation.
[a]rs401681 is reported for BLCA and PAAD.

between intergenic and intronic SNPs and the various phenotypes identified in genome-wide association studies (GWAS) (Freedman et al., 2011; Hernandez and Singleton, 2012; Kilpinen and Dermitzakis, 2012). Herein, genotype-epitype connections might also further clarify the causal relationships underlying risk alleles found to be expression quantitative trait loci (Li et al., 2013).

Considering the close interplay between genetic variability and epigenetic gene regulation, we hypothesize that comprehensive analysis of both layers in a context-specific manner could greatly improve our understanding of genetic variability associated with disease biology. Realizing that a major portion of cancer-related SNPs are located in a noncoding context with elusive functional impact, the connection between the genotype and epitype might facilitate GWAS interpretation and allow the identification of novel cancer gene candidates. Consequently, we integrate DNA methylation and genotype data of 3,649 primary samples representing the most frequent solid cancer types and propose that DNA methylation is an important component in cancer risk biology.

RESULTS

Hypothesis-Driven Discovery of Genotype-Epitype Connections at Cancer Risk Alleles

We determined associations between SNPs identified as being cancer related in genome-wide association studies (GWAS-SNPs) and *cis*-acting DNA methylation quantitative trait loci (*cis*-meQTLs). Assuming the genotype-epitype interaction to be predominantly detectable in a cancer-type-specific manner, we separately analyzed the association between GWAS-SNPs and DNA methylation levels for 13 cancer types that rank among the most frequent solid cancer types (SEER Cancer Statistics 2005–2009, Age-Adjusted SEER Incidence Rates and Trends for the Top 15 Cancer Sites) (Table 1). Genotype and epitype data from large cohorts of well-characterized cancer samples were obtained from The Cancer Genome Atlas (TCGA) consortium. In total, using the available genotype data sets, we were able to interrogate 109 GWAS-SNPs related to the respective cancer types and integrated matched DNA methylation and genetic data from 3,649 samples (3,120 primary tumor and 529 matched healthy control samples, Table 1).

GWAS-SNP annotations for the respective cancer type were extracted from related studies present in the GWAS database (Li et al., 2012). In order to determine associations between the genetic and epigenetic codes comprehensively, we combined data from high-resolution genotyping (Affymetrix Genome-Wide Human SNP Array 6.0) and epityping (Illumina HumanMethylation450 BeadChip) array platforms, analyzing 906,600 SNPs and 485,577 CpG sites, respectively. The DNA methylation BeadChip interrogates DNA methylation levels in promoter regions of virtually all protein-coding genes, but also noncoding RNAs (ncRNAs) and regulative loci (Sandoval et al., 2011).

To define *cis*-acting associations between GWAS-SNPs and DNA methylation levels of CpG sites in a 1 Mb region flanking the cancer-related polymorphic site, we applied a multivariate model (random forest selection frequency [RFSF]) for cancer and healthy samples independently and stratified by tissue types. We used the RFSF method, because it is known to perform better in identifying quantitative trait loci compared with other univariate or multivariate approaches (Michaelson et al., 2009, 2010). The method determines direct correlations between polymorphic sites (GWAS-SNPs) and CpG methylation levels enriched over a predefined background.

Differential DNA Methylation Associated with Breast Cancer Risk Alleles Identified Potentially Novel Cancer Genes

To evaluate the merits of the study design and the applied methods, we initially interrogated epigenetic GWAS-SNP associations in the context of breast cancer, which comprises the largest sample cohort interrogated in our study and is the most common tumor type in women throughout the world (SEER Cancer Statistics 2005–2009, Age-Adjusted SEER Incidence Rates and Trends for the Top 15 Cancer Sites Females). In terms of cancer predisposition and risk alleles, breast cancer is one of the best-defined cancer types, having been analyzed in numerous GWA studies. Fifty-seven genetic variants associated with disease risk have been determined in 23 independent studies (Li et al., 2012). Most importantly, the vast majority of the polymorphic sites is located in a noncoding sequence and consequently does not directly affect gene products. However, their proximity or intronic location with respect to cancer-related genes, such as the estrogen receptor 1 (*ESR1*) (Deblois and Giguère, 2013), the fibroblast growth factor receptor 2 (*FGFR2*) (Katoh, 2008), or telomerase reverse transcriptase (*TERT*) (Horn et al., 2013; Huang et al., 2013), suggests that there is a direct regulatory influence of genetic variants, although underlying mechanisms remain difficult to identify. Intriguingly, a number of breast cancer risk polymorphisms were recently reported to be enriched at differentially methylated enhancer sites that are correlated with intertumoral expression variation (Aran and Hellman, 2013). Although they did not analyze direct genotype-epitype associations, the study provides an outlook on the potential of data integration to study cancer risk allele biology.

In order to determine significant associations between genetic variability and epigenetic variation, we performed an integrative analysis of matched genotype and epitype data from 576 invasive breast carcinomas and 78 normal breast tissues (Table 1). Using the available genotype data sets (Affymetrix Genome-Wide Human SNP Array 6.0), we were able to interrogate the epitype association of 28 out of 57 breast cancer-related GWAS-SNPs. Most strikingly, we observed significant associations between five risk alleles and CpG methylation levels within a 1 Mb region flanking the polymorphism (RFSF, false discovery rate [FDR] $\leq$0.05; Figure 1). Particularly, the five *cis*-acting GWAS-SNPs were associated with DNA methylation levels of seven CpG sites, with one SNP showing multiple associations (Table S1). These genes include reported oncogenes, such as v-myc myelocytomatosis viral oncogene homolog (*MYC*) and collagen, type I, alpha 1 (*COL1A1*), but might also represent novel cancer gene candidates not previously described as being cancer related.

We were able to detect the association between five out of seven GWAS-SNPs and differential CpG methylation in matched healthy breast tissue (RFSF, FDR $\leq$0.05), suggesting these meQTLs are established even before cancer onset (Table S1). Importantly, the associations could be confirmed in an independent set of tumor-adjacent normal breast tissues using Sanger and bisulfite pyrosequencing for geno- and epityping, respectively (Mann-Whitney test, Figure S1). Further, we wondered whether we could detect the epigenetic associations of the breast cancer risk polymorphisms that were present in the healthy cohort even outside the breast tissue context. Importantly, integrating genotype and epitype data of 12 additional tissue types, five out of five breast meQTLs could also be detected in other tissue types (Table S2), suggesting that the epigenetic risk alleles occur independently of tissue types and draw attention to their stable character and potential function as risk epi-polymorphisms.

In particular, the association of the breast risk allele rs2380205 with DNA methylation levels in the promoter of F-box protein, helicase, 18 (*FBXO18*) is an interesting case, because the gene codes for a DNA-dependent ATPase and DNA helicase with ubiquitin ligase activity (Kim et al., 2002) and is involved in the regulation of homologous recombination and stress-induced apoptosis (Figures 2A–2C), processes reported to be aberrant in familial and sporadic breast cancer.

By integrating genotype and epitype data from a large cohort of breast cancer samples, we determined significant associations between breast cancer risk alleles and epigenetic aberrations at already reported and potentially novel cancer genes. Consequently, given the power of the methods applied, we carried out an integrated epitype analysis for the GWAS-SNPs associated with the most frequent solid tumor types.

Comprehensive Profiling for Risk meQTLs in Solid Human Cancer Types

To compile a comprehensive catalog of differential DNA methylation alleles related to cancer risk, we did an analysis of 12 additional cancer types, representing close to 3,000 human samples (2,544 primary tumors and 451 healthy control samples, Table 1). In total, we identified 29 additional genotype-epitype associations related to 18 risk alleles in ten cancer types (RFSF, FDR $\leq$0.05; Figure 1). The entire set of genotype-epitype associations is listed in Table S3. Illustrative examples displaying differential DNA methylation in gene promoters with a previously identified association with tumorigenesis include tumor protein 63 (*TP63*) (Tonon et al., 2005) and growth arrest specific 8 (*GAS8*) (Whitmore et al., 1998) in lung adenocarcinomas and melanomas, respectively.

Similar to the breast cancer analysis, 72% (21 out of 29) of meQTLs could also be detected in a matched healthy context (RFSF, FDR $\leq$0.05; Table S3) and replicated in a technical and biological validation study of tumor-adjacent normal tissues (Figure S2), suggesting these associations to potentially participate in cancer predisposition. Although all risk allele-associated differentially methylated genes are potentially directly associated with cancer formation, we believe it worthwhile to highlight a special case relating to aberrant methylation of the oncogene *TERT*, which was previously reported as being aberrantly regulated in a cancer context (Horn et al., 2013; Huang et al., 2013). Reverse transcriptase is implemented in the maintenance of telomere ends and prevents telomere shortening during cell division and oncogenesis. A genetic variation in the second intron of *TERT* (rs2736100) was previously related to a higher risk of lung adenocarcinomas (Landi et al., 2009), although the functional consequences on the oncogene remain unknown. Integrating an epitype analysis of 303 primary lung adenocarcinomas revealed an association of the risk allele with a differentially methylated CpG site in the gene promoter region

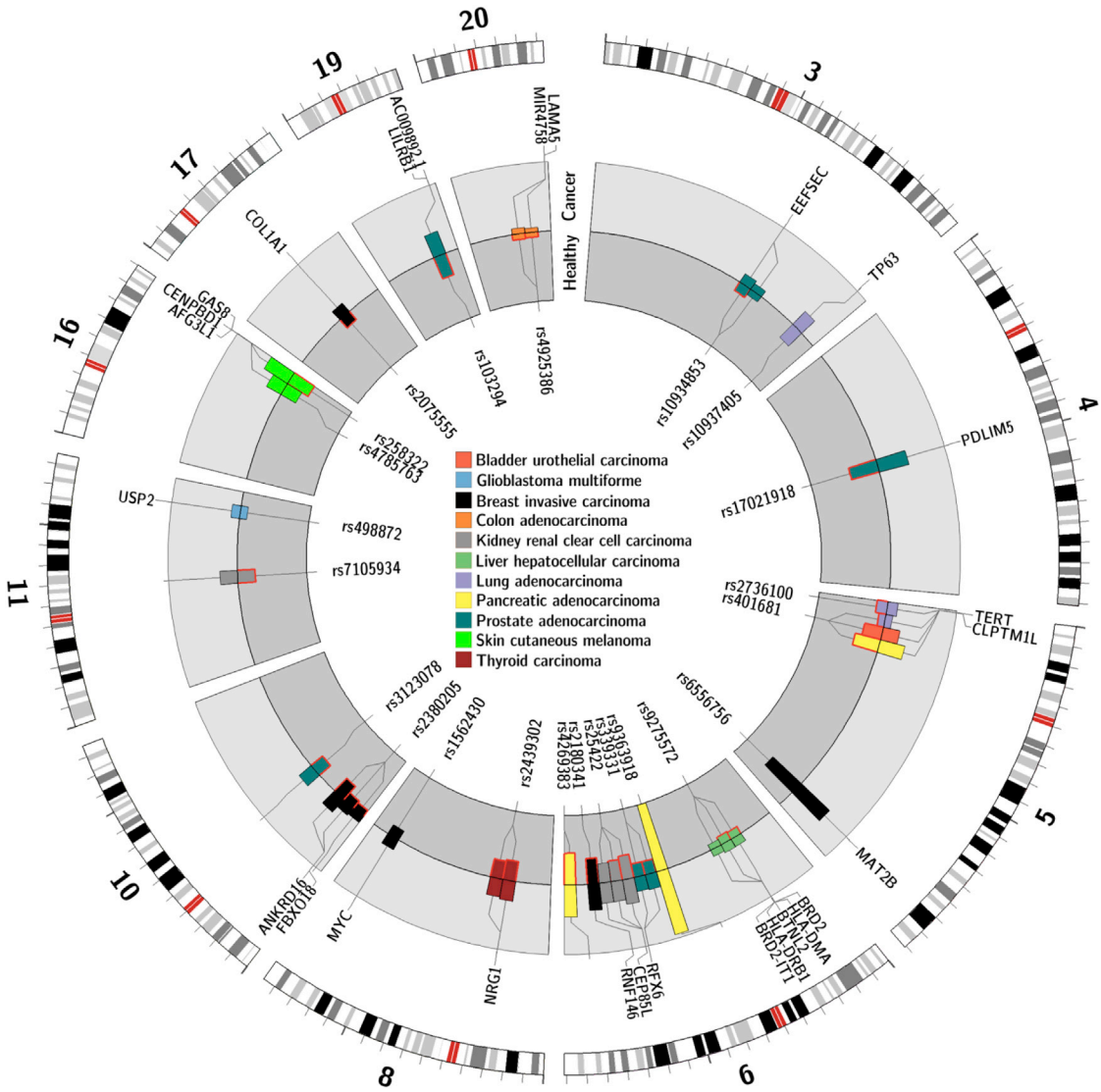

Figure 1. Schematic Overview of the DNA Methylation Quantitative Trait Loci Associated with GWAS Cancer Risk Polymorphisms

A total of 23 GWAS-SNPs (inner ring) had significant associations with DNA methylation levels of 36 CpG sites, representing 27 unique genes (outer ring) (Krzywinski et al., 2009). Random forest selection frequencies for cancer (up) and healthy (down) samples are displayed. Significant associations in healthy cohorts are indicated (red box). The position of the gene linkers (black lines) represents the location on the respective chromosomes. The associated cancer types are color coded.

of *TERT*, approximately 4 kb downstream of the risk SNP (Figures 3A and 3B). Taking advantage of the high resolution of the DNA methylation array, we were able to interrogate further the methylation level of flanking CpG sites in the CpG-dense promoter region of *TERT*. In this way, we identified several flanking loci that were differentially methylated, consistent with the identified meQTL (Figure 3C). Thus, we suggest that the epigenetic modification could participate in deregulating *TERT* in the context of lung tumors. Similarly, the association of cancer risk alleles and epigenetic regulation of *TERT* has been recently described in an ovarian cancer context (Bojesen et al., 2013). Likewise, the additional here listed epigenetic associations might help explain the functional role of risk alleles in their respective tissue types.

In order to underscore the functional impact of differential CpG methylation, we further performed correlation analysis between DNA methylation levels of risk allele associated CpG sites and transcript abundance of assigned target genes (Spearman's correlation test, Table S3). Here, we observed highly significant correlations between CpG methylation levels and the expression of crucial cancer genes, including *MYC*, *TERT*, and *TP63*, further supporting differential DNA methylation to present an important mechanism that might mediate genetic cancer risk (Figure S3).

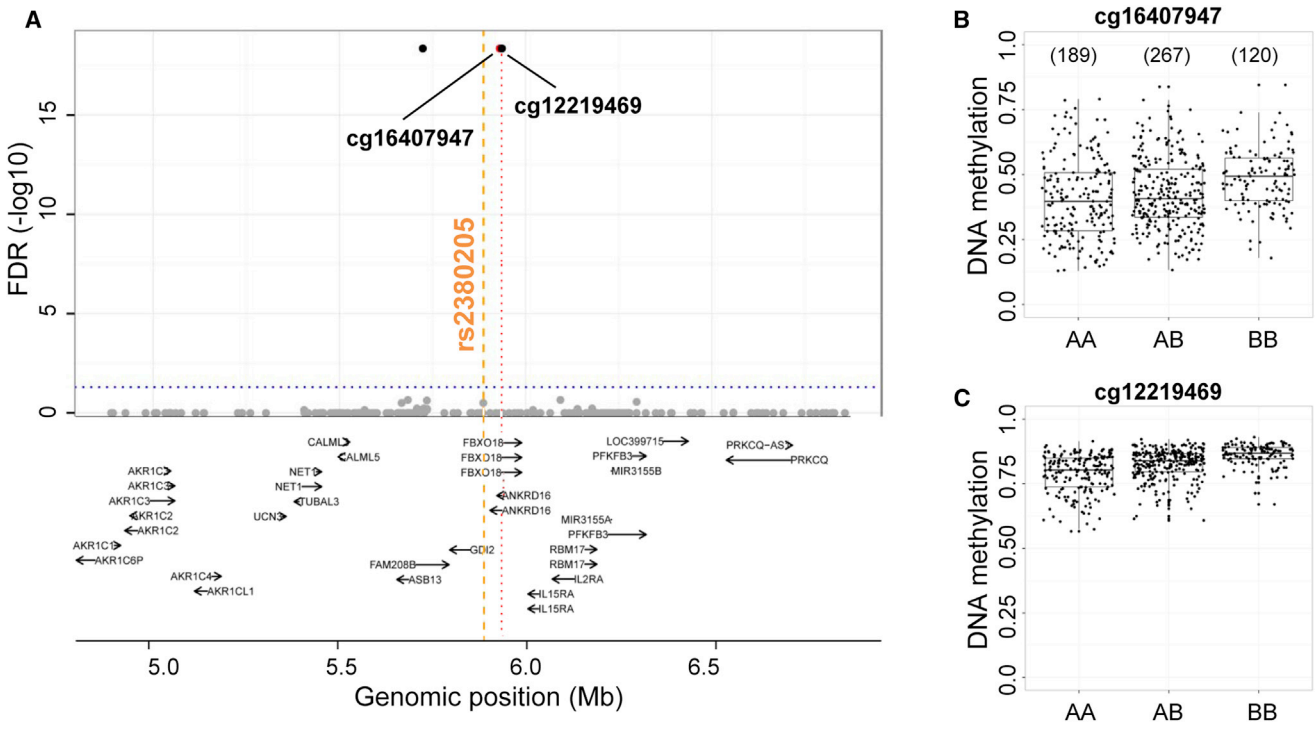

Figure 2. The Breast Cancer Risk SNP rs2380205 Is Associated with Differential CpG Methylation in the *FBXO18* Promoter

(A) Epigenetic association studies revealed a significant (RFSF, FDR $\leq$0.05) connection between rs2380205 and two CpG sites (cg12219469 and cg16407947) in the promoter region of *FBXO18*. Significance values are shown of the CpG sites and gene locations in the interrogated 1 Mb window flanking the breast-cancer-related SNP.

(B and C) Relationship between the breast cancer risk genotype (rs2380205) and DNA methylation level at cg16407947 (B) and cg12219469 (C) in the promoter of *FBXO18*. Allele frequencies are displayed.

DISCUSSION

Comprehensive profiling of the genome with respect to variable loci between individuals discordant for certain phenotypes revealed a plethora of SNP sites with a significant association with diverse phenotypes, including cancer. However, the functional consequence of risk polymorphisms was often hard to determine because their gene products were mainly unaffected by the variation. Consequently, we hoped to explain the lack of any direct connection by adding another layer of information. Considering epigenetic gene regulation as a potential intermediate event connecting genotype-phenotype associations, we integrated high-resolution epitype data with well-known cancer risk polymorphisms. The resulting catalog of genotype-epitype associations is a rich source of information with which we may interpret GWAS-defined cancer risk markers.

A Catalog of Epigenetic Associations to Cancer Risk Alleles

Taken together, massively integrating genotype data from cancer-related polymorphic sites with potentially *cis*-affected CpG methylation levels and analyzing cancer types that rank among the most abundant solid tumor types (SEER Cancer Statistics 2005-2009, Age-Adjusted SEER Incidence Rates and Trends for the Top 15 Cancer Sites Both Sexes), the risk meQTL catalog described here features significant association of 21% (23 out of 109) of the cancer-related polymorphisms interrogated. As the risk polymorphisms are predominantly located in a noncoding context, their epigenetic association might aid functional interpretation and further clarify their role in cancer biology. The majority of the genes reported (apart from *MYC* and *COL1A1*) have not previously been defined as cancer genes on the basis of their genetic alterations (Cancer Gene Consensus, Sanger Institute). Among those, the risk allele association to *FBXO18* represents a particularly interesting case as the gene product actively participates in the formation of double-strand breaks and the activation of tumor protein p53 (TP53) -dependent apoptosis following DNA replication stress (Fugger et al., 2013; Jeong et al., 2013). *FBXO18*-deficient cells have an impaired ability to activate the cytotoxic-stress-induced cascade, resulting in increased cell survival. *FBXO18*-deficient cells are hypersensitive to topoisomerase inhibitors, due to their involvement in mitotic progression (Laulier et al., 2010). FBXO18 is involved in processes reported to be aberrant in familial and sporadic breast cancer, and although the gene itself has not yet been reported in the context of cancer, its functional implication in DNA repair draws attention to its possible importance for disrupting DNA integrity and thereby cancer formation. According to its regulatory association to genetic breast cancer risk, *FBXO18* might represent an epigenetically regulated breast

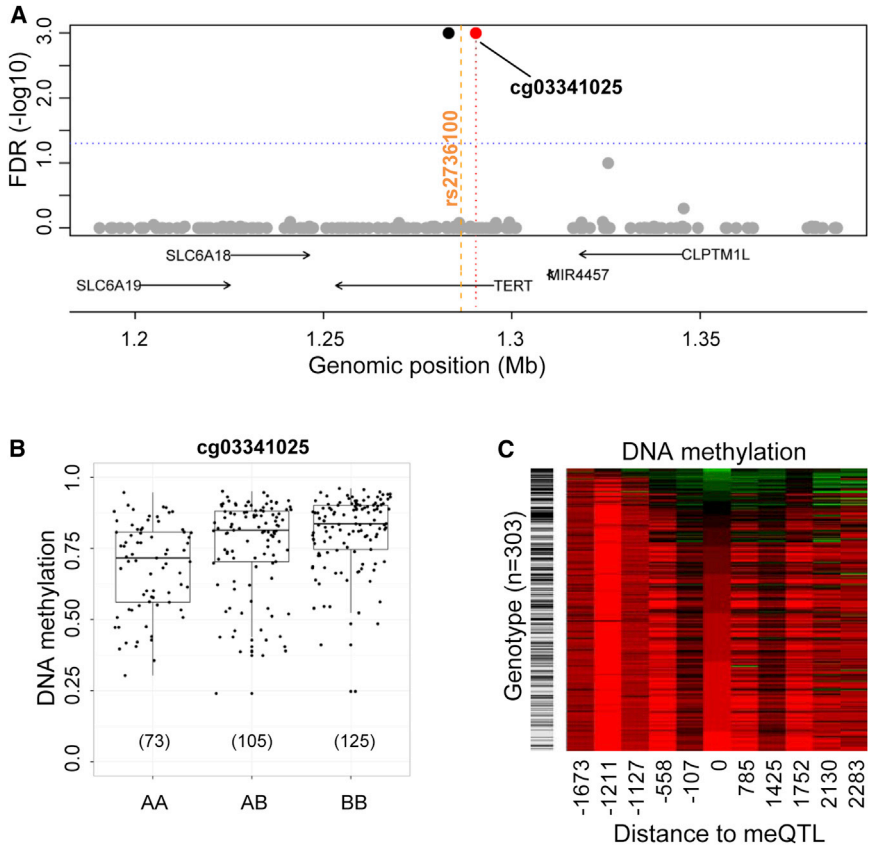

Figure 3. Association of the Lung Cancer Risk SNP rs2736100 and Differential Promoter Methylation of *TERT*

(A) Epigenetic association studies revealed a significant (RFSF, FDR $\leq$ 0.05) connection between rs2736100 and a CpG (cg03341025) in the promoter region of *TERT*. Significance values are presented of the CpG sites and gene locations in a 100 kb window flanking the lung cancer-related SNP.

(B) Relationship between the lung cancer risk polymorphism (rs2736100) and DNA methylation level at cg03341025 in the *TERT* promoter. Allele frequencies are displayed.

(C) Differential DNA methylation in the *TERT* promoter region. The identified meQTL (cg03341025) and flanking CpG sites display altered DNA methylation associated with the lung cancer risk polymorphism. One horizontal line represents DNA methylation levels of a single lung cancer patient with the indicated genotypes (black: AA; dark gray: AB; light gray: BB). Distances of neighboring CpG sites to the meQTL are indicated.

cancer candidate gene. However, further studies need to clarify the functional implication of *FBXO18* in breast tumorigenesis.

Detecting the risk allele associations in a cancer and healthy context supports the potential implementation of the meQTLs as a risk factor for cancer and potential integral components of cancer risk allele biology, wherein aberrant DNA methylation might function as a mediator for the respective risk alleles. Likewise, genetic risk polymorphisms display germline variants present in the diseased tissues and the matched normal counterparts. Breast cancer 1, early onset (*BRCA1*), the most commonly mutated gene in inherited breast cancer, is a well-studied example of breast cancer susceptibility. Although, germline mutations are present in all healthy tissue types, they mainly exhibit increased predisposition for breast and ovarian cancer, being responsible for 40% of inherited breast cancer. In line, we detected breast cancer risk meQTLs in cancer samples and normal tissues, including nonrelated tissue types.

However, as the underlying data sets (cancer samples and adjacent normal tissues) did not include control subjects without disease, future studies have to elucidate if GWAS risk alleles exhibit similar relationships in cancer-unrelated donors, further supporting meQTLs as mechanistic player in cancer predisposition.

By presenting previously undescribed genotype-epitype-phenotype connections in human cancers, we make it easier to interpret risk alleles. Given the significance of the epigenetic association described in this study, we suggest that their annotation in genotype databases, such as the SNP database (SNPdb), should be taken into consideration. In this context, the annotation of expression QTLs already provides important extra information with which to evaluate the biological consequences of genetic variation. The addition of meQTL information could further support an accurate interpretation.

EXPERIMENTAL PROCEDURES

The entire set of experimental procedures is available in Supplemental Information.

Methylation Quantitative Trait Loci Identification

Methylation quantitative trait loci associated with the entire set of GWAS-SNP on the genotyping platform (Affymetrix Genome-Wide Human SNP Array 6.0) were identified by interrogating CpG sites (represented on the Infinium HumanMethylation450) located in a ±1 Mb window flanking the polymorphic sites. The window was reduced by 100 Kb steps if it contained more than 1,000 CpGs. We used the multivariate random forest selection frequency (RFSF) method, as described in Michaelson et al. (2010), to identify associations between a SNP and its neighboring CpG sites.

The Random Forest algorithm is implemented in R in the randomForest package (Liaw and Wiener, 2002). First, we set the random forests algorithm to generate 2,000 trees for classification, and calculated the selection frequency (SF) of the variables (CpG sites) used in building the classification model. Bias correction was then applied to the frequencies by subtracting the deviation between the SF of the variable under the null hypothesis (no association between the SNP and the methylation value) and the average SF of all variables under the null hypothesis; we used 1,000 forests of ten trees to derive the SF under the null hypothesis, randomly sampling the three possible genotypes with the same frequency, and applying the correction to the original SF. Finally, in order to measure how closely the genotype was associated with the SF of a CpG site, we constructed an empirical null distribution from the SFs of ten forests of 2,000 trees by permuting the genotypes

of our samples, and deriving a q value for every CpG by comparing its SF with those under the null hypothesis. Here, the null distribution was created with the assumption that the selection frequencies were not linked to any relationship between the SNP and the CpG methylation values. Thus, we shuffled the genotypes of the samples and calculated ten forests of 2,000 trees, each to create a null distribution of SF following our null hypothesis. Then, we calculated the empirical cumulative distribution of the null SF values, which was used to test our initial hypothesis that the SF is linked to the relationship between CpG methylation and genotype. For selected associations, our initial hypothesis stated that the SF was higher than what would be expected under the null hypothesis. The number of tests depended on the number of CpG probes on the HumanMethylation450 BeadChip located within the analyzed window flanking the respective risk SNPs (range: 0–978, mean: 391). Accordingly, the p value was adjusted by FDR.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.03.016.

ACKNOWLEDGMENTS

This study was supported by the grant HEALTH-F5-2011-282510-BLUEPRINT, the European Research Council (ERC) under grant agreement no. 268626 - EPINORC project, Ministry of Economy and Competitiveness grant (MINECO Project - SAF2011-22803), and the Health and Science Departments of the Catalan Government (Generalitat de Catalunya). H.H. is a Juan de la Cierva Postdoctoral Fellow. M.E. is an Institucio Catalana de Recerca i Estudis Avançats (ICREA) Research Professor. We thank J. Méndez-González for critically reading the manuscript.

Received: November 15, 2013
Revised: February 10, 2014
Accepted: March 5, 2014
Published: April 3, 2014

REFERENCES

Aran, D., and Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. Cell *154*, 11–13.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. *12*, R10.

Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., et al.; Australian Cancer Study; Australian Ovarian Cancer Study; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab); Gene Environment Interaction and Breast Cancer (GENICA); Swedish Breast Cancer Study (SWE-BRCA); Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); Epidemiological study of BRCA1 & BRCA2 Mutation Carriers (EMBRACE); Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO) (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet. *45*, 371–384, e1–e2.

Deblois, G., and Giguère, V. (2013). Oestrogen-related receptors in breast cancer: control of cellular metabolism and beyond. Nat. Rev. Cancer *13*, 27–36.

Fraser, H.B., Lam, L.L., Neumann, S.M., and Kobor, M.S. (2012). Population-specificity of human DNA methylation. Genome Biol. *13*, R8.

Freedman, M.L., Monteiro, A.N.A., Gayther, S.A., Coetzee, G.A., Risch, A., Plass, C., Casey, G., De Biasi, M., Carlson, C., Duggan, D., et al. (2011). Principles for the post-GWAS functional characterization of cancer risk loci. Nat. Genet. *43*, 513–518.

Fugger, K., Chu, W.K., Haahr, P., Kousholt, A.N., Beck, H., Payne, M.J., Hanada, K., Hickson, I.D., and Sørensen, C.S. (2013). FBH1 co-operates with MUS81 in inducing DNA double-strand breaks and cell death following replication stress. Nat. Commun. *4*, 1423.

Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.-L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J., et al. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. *6*, e1000952.

Hernandez, D.G., and Singleton, A.B. (2012). Using DNA methylation to understand biological consequences of genetic variability. Neurodegener. Dis. *9*, 53–59.

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., Monk, D., Hata, K., Marques-Bonet, T., Wang, L., and Esteller, M. (2013). DNA methylation contributes to natural human variation. Genome Res. *23*, 1363–1372.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., et al. (2013). TERT promoter mutations in familial and sporadic melanoma. Science *339*, 959–961.

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. Science *339*, 957–959.

Jeong, Y.-T., Rossi, M., Cermak, L., Saraf, A., Florens, L., Washburn, M.P., Sung, P., Schildkraut, C.L., and Pagano, M. (2013). FBH1 promotes DNA double-strand breakage and apoptosis in response to DNA replication stress. J. Cell Biol. *200*, 141–149.

Katoh, M. (2008). Cancer genomics and genetics of FGFR2 (Review). Int. J. Oncol. *33*, 233–237.

Kilpinen, H., and Dermitzakis, E.T. (2012). Genetic and epigenetic contribution to complex traits. Hum. Mol. Genet. *21* (R1), R24–R28.

Kim, J., Kim, J.-H., Lee, S.-H., Kim, D.-H., Kang, H.-Y., Bae, S.-H., Pan, Z.-Q., and Seo, Y.-S. (2002). The novel human DNA helicase hFBH1 is an F-box protein. J. Biol. Chem. *277*, 24530–24537.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Landi, M.T., Chatterjee, N., Yu, K., Goldin, L.R., Goldstein, A.M., Rotunno, M., Mirabello, L., Jacobs, K., Wheeler, W., Yeager, M., et al. (2009). A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am. J. Hum. Genet. *85*, 679–691.

Laulier, C., Cheng, A., Huang, N., and Stark, J.M. (2010). Mammalian Fbh1 is important to restore normal mitotic progression following decatenation stress. DNA Repair (Amst.) *9*, 708–717.

Li, M.J., Wang, P., Liu, X., Lim, E.L., Wang, Z., Yeager, M., Wong, M.P., Sham, P.C., Chanock, S.J., and Wang, J. (2012). GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res. *40* (Database issue), D1047–D1054.

Li, Q., Seo, J.-H., Stranger, B., McKenna, A., Pe'er, I., Laframboise, T., Brown, M., Tyekucheva, S., and Freedman, M.L. (2013). Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell *152*, 633–641.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News *2*, 18–22.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. *31*, 142–147.

McDaniell, R., Lee, B.-K., Song, L., Liu, Z., Boyle, A.P., Erdos, M.R., Scott, L.J., Morken, M.A., Kucera, K.S., Battenhouse, A., et al. (2010). Heritable individual-specific and allele-specific chromatin signatures in humans. Science *328*, 235–239.

Michaelson, J.J., Loguercio, S., and Beyer, A. (2009). Detection and interpretation of expression quantitative trait loci (eQTL). Methods *48*, 265–276.

Michaelson, J.J., Alberts, R., Schughart, K., and Beyer, A. (2010). Data-driven assessment of eQTL mapping methods. BMC Genomics *11*, 502.

Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W.R., Hannon, G.J., and Smith, A.D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell *146*, 1029–1041.

Rando, O.J. (2012). Daddy issues: paternal effects on phenotype. Cell *151*, 702–708.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics *6*, 692–702.

Seisenberger, S., Peat, J.R., and Reik, W. (2013). Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. Curr. Opin. Cell Biol. *25*, 281–288.

Shoemaker, R., Deng, J., Wang, W., and Zhang, K. (2010). Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res. *20*, 883–889.

Tonon, G., Wong, K.-K., Maulik, G., Brennan, C., Feng, B., Zhang, Y., Khatry, D.B., Protopopov, A., You, M.J., Aguirre, A.J., et al. (2005). High-resolution genomic profiles of human lung cancer. Proc. Natl. Acad. Sci. USA *102*, 9625–9630.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. *37*, 853–862.

Whitmore, S.A., Settasatian, C., Crawford, J., Lower, K.M., McCallum, B., Seshadri, R., Cornelisse, C.J., Moerland, E.W., Cleton-Jansen, A.M., Tipping, A.J., et al. (1998). Characterization and screening for mutations of the growth arrest-specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer. Genomics *52*, 325–331.

Zhang, D., Cheng, L., Badner, J.A., Chen, C., Chen, Q., Luo, W., Craig, D.W., Redman, M., Gershon, E.S., and Liu, C. (2010). Genetic control of individual differences in gene-specific methylation in human brain. Am. J. Hum. Genet. *86*, 411–419.