Accepted Manuscript

Rare germline copy number variations and disease susceptibility in familial melanoma

Jianxin Shi, Weiyin Zhou, Bin Zhu, Paula L. Hyland, Hunter Bennett, Yanzi Xiao, Xijun Zhang, Laura S. Burke, Lei Song, Chih Hao Hsu, Chunhua Yan, Qingrong Chen, Daoud Meerzaman, Casey L. Dagnall, Laurie Burdette, Belynda Hicks, Neal D. Freedman, Stephen J. Chanock, Meredith Yeager, Margaret A. Tucker, Alisa M. Goldstein, Xiaohong R. Yang

PII: S0022-202X(16)32229-1

DOI: 10.1016/j.jid.2016.07.023

Reference: JID 469

To appear in: *The Journal of Investigative Dermatology*

Received Date: 20 January 2016

Revised Date: 2 June 2016

Accepted Date: 5 July 2016

Please cite this article as: Shi J, Zhou W, Zhu B, Hyland PL, Bennett H, Xiao Y, Zhang X, Burke LS, Song L, Hsu CH, Yan C, Chen Q, Meerzaman D, Dagnall CL, Burdette L, Hicks B, Freedman ND, Chanock SJ, Yeager M, Tucker MA, Goldstein AM, Yang XR, Rare germline copy number variations and disease susceptibility in familial melanoma, *The Journal of Investigative Dermatology* (2016), doi: 10.1016/j.jid.2016.07.023.

Rare germline copy number variations and disease susceptibility in familial melanoma

Jianxin Shi[1], Weiyin Zhou[2], Bin Zhu[1], Paula L Hyland[1], Hunter Bennett[1], Yanzi Xiao[1], Xijun Zhang[2], Laura S Burke[1,3], Lei Song[1], Chih Hao Hsu[4], Chunhua Yan[4], Qingrong Chen[4], Daoud Meerzaman[4], Casey L Dagnall[2], Laurie Burdette[2], Belynda Hicks[2], Neal D Freedman[1], Stephen J Chanock[1], Meredith Yeager[2], Margaret A Tucker[1], Alisa M Goldstein[1], Xiaohong R Yang[1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA

[2]Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[3]Statistics Collaborative, Inc., Washington, DC, USA

[4]Computational Genomics Research Group, Center for Biomedical Informatics & Information Technology, National Cancer Institute, Bethesda, MD, USA

Correspondence to: Xiaohong R Yang, Division of Cancer Epidemiology and Genetics, NCI/NIH, Bethesda, MD, USA. Tel: 240-276-7226; Fax: 240-276-7834; Email: royang@mail.nih.gov

Short title: Rare germline CNVs in familial melanoma

Abbreviation list:

CMM – cutaneous malignant melanoma;

CNV – copy number variation;

CGH – comparative genomic hybridization;

DN – dysplastic nevi;

SNP – single nucleotide polymorphism;

LRR – Log R ratio;

BAF – B allele frequency;

qPCR – quantitative PCR;

ddPCR – droplet digital PCR;

CAP – CBL-associated protein

Abstract

Mounting evidence suggests that copy number variations (CNVs) can contribute to cancer susceptibility. The main goal of this study was to evaluate the role of germline CNVs in melanoma predisposition in high-risk melanoma families. We used genome-wide tiling comparative genomic hybridization and SNP arrays to characterize CNVs in 335 individuals (240 melanoma cases) from American melanoma-prone families (22 with germline *CDKN2A* or *CDK4* mutations). We found that the global burden of overall CNVs (or deletions or duplications separately) was not significantly associated with case-control or *CDKN2A/CDK4* mutation status after accounting for the familial dependence. However, we identified several rare CNVs that either involved known melanoma genes (e.g. *PARP1*, *CDKN2A*) or co-segregated with melanoma (duplication on 10q23.23, 3p12.2 and deletions on 8q424.3, 2q22.1) in families without mutations in known melanoma high-risk genes. Some of these CNVs were correlated with expression changes in disrupted genes based on RNASeq data from a subset of melanoma cases included in the CNV study. These results suggest that rare co-segregating CNVs may influence melanoma susceptibility in some melanoma-prone families and genes found in our study warrant further evaluation in future genetic analyses of melanoma.

Introduction

Cutaneous malignant melanoma (CMM) is a heterogeneous disease with multiple factors contributing to its etiology (Tucker and Goldstein, 2003). Approximately 10% of CMM cases have one or more first-degree relatives also affected with melanoma (Goldstein and Tucker, 2001). *CDKN2A* and *CDK4* are the two well-established high-penetrance melanoma susceptibility genes. Recently, *BAP1*, *POT1*, *ACD, TERF2IP*, and *TERT* were identified as potential high-penetrance melanoma susceptibility genes (Read *et al.*, 2016). Together, these genes account for melanoma risk in less than 40% of melanoma-prone families. . In addition, *CDKN2A* mutation carriers demonstrated geographic and phenotypic variations (Bishop *et al.*, 2002). Dysplastic nevi (DN) and genetic variants in *MC1R* have been associated with increased risk of melanoma among *CDKN2A* mutation carriers (Demenais *et al.*, 2010). These findings suggest that the etiology underlying familial CMM is complex and additional high-penetrance and modifier genes remain to be identified.

Recently, copy number variations (CNVs) have been shown to contribute substantially to disease susceptibility in a large number of diseases including cancer (Kuiper *et al.*, 2010; Sebat *et al.*, 2007; Walsh *et al.*, 2008). CNVs can have dramatic phenotypic consequences through altering gene dosage, disrupting coding sequences, or perturbing long-range gene regulation (Kleinjan and van Heyningen, 2005). The overall burden of germline CNVs has been associated with increased risks of Li-Fraumeni syndrome and familial colorectal cancer (Shlien *et al.*, 2008; Yang *et al.*, 2014). Rare CNVs, such as those disrupting known cancer genes, have also been reported as major susceptibility mechanisms in several familial cancers (Yang *et al.*, 2009). For example, large germline deletions of *CDKN2A* involving p16 and p14ARF have been reported in about 2% of melanoma-prone families (Goldstein *et al.*, 2006). In addition, in our previous study, we identified a rare duplication region involving a cluster of CXC chemokine genes in all affected people in a melanoma-prone family (Yang, 2012), further supporting the

role of rare, inherited CNVs in melanoma susceptibility. However, the systematic investigation of germline CNVs in familial melanoma is still lacking. Therefore, the primary goal of this study was to assess whether the overall burden of germline CNVs varied by CMM and/or *CDKN2A*/*CDK4* mutation status in melanoma-prone families with and without known mutations. In addition, we also evaluated co-segregating rare CNVs as susceptibility factors in melanoma-prone families without known mutations.

Results and Discussion

The analysis of CNV burden included 166 CMM cases (88 *CDKN2A/CDK4*-positive and 78 mutation-negative), 44 unaffected family members (5 *CDKN2A/CDK4*-positive and 39 mutation-negative), and 44 unrelated spouses based on Nimblegen 720k array-CGH data. We only included CNVs > 10 kb to reduce false positives. The median numbers of total CNVs, gains, and losses were similar among CMM cases, unaffected family members, and spouses (Table 1). Similar results were obtained for CNVs larger than 100 kb (Table 1) and 500 kb (data not shown). Each test of the global burden of overall CNVs (P=0.46), deletions (P=0.53), or duplications (P=0.64) was not significantly associated with case-control status using the score statistics accounting for familial dependence (Table 1). Adjusting for potential batch effects did not alter the results (data not shown). Further, among CMM cases, the global burden of overall CNVs (P=0.75), deletions (P=0.82), or duplications (P=0.80) did not vary significantly by germline *CDKN2A/CDK4* mutation status. In addition to number of CNVs, we also analyzed number of genes and length of DNA segments affected by CNVs and results were similar (data not shown). Restriction to CNVs that were not reported in the Database of Genomic Variants did not substantially change the results. Previous studies investigating the association between CNV burden and familial

cancer risk have shown inconsistent results (Al-Sukhni *et al.*, 2012; Krepischi *et al.*, 2012; Pylkas *et al.*, 2012). Although various mechanisms may exist by which CNV burden contributes to risk in different cancers, these discrepant results may also reflect the divergent designs of these studies including sample sizes, array platforms, and methods to call CNVs. Our results imply that the vast majority of CNVs in the genome, even those involving large DNA segments, may be polymorphisms that are not related to disease development, at least for some disease traits.

We assessed rare CNVs (i.e., not present in 44 spouse controls, >10 kb) in known melanoma genes in families with and without known mutations. Among the 35 melanoma genes examined (Supplementary table 1), we found a 1.3Mb deletion in the 1q42.12-13 region involving *PARP1* (Figure 1a) in a single CMM case (1001) in a *CDKN2A/CDK4*-negative family with four CMM cases. The deletion region, which included 12 additional genes (Table 2), was not found in two other CMM cases or three unaffected individuals in this family. We further confirmed the deletion in the single case (1001) by qPCR, in which we included all cases and unaffected members in this family with DNA; no one other than 1001 harbored the deletion (Figure 2a). Among 52 samples with RNASeq data, including three cases in this family (1001, 1002, and 3003), the deletion carrier (1001) showed the lowest PARP1 expression level (Figure 3a). PARP1 plays complex roles in multiple cancer-related cell processes and polymorphisms in PARP1 have been associated with melanoma risk (Macgregor *et al.*, 2011). The role of deletions or low expression levels of PARP1 in melanoma development and progression has not been investigated, although aged mice lacking *PARP1* were susceptible to epidermal hyperplasia (Wang *et al.*, 1997).

We also identified a 10 kb deletion on 9p21.3 involving *CDKN2A* exon 1β (Table 2, Figure 1b) in 5 of 6 CMM cases and one obligate gene carrier in a large CMM family. Two additional cases (2003 and 2004) should also carry the deletion based on segregation (Figure 2b). The family was previously

classified as *CDKN2A*-negative since the family did not have mutations from sequencing and one of the three cases examined at the time of the initial analysis did not share the 9p21 haplotype. We validated the deletion in this family by qPCR. p14ARF expression was lower in the two cases with the deletion compared to the case without the deletion, but was not substantially different compared to the average expression level among all analyzed samples. However, the overall expression of p14ARF was very low in blood lymphocytes, the only available sample type for this evaluation (Figure 3b). The identification of the exon 1β deletion in a family that was previously classified as *CDKN2A*-negative based on sequencing analysis alone further highlights the importance of checking for CNVs in genetic screening.

To identify CNVs that might relate to CMM susceptibility, we did an agnostic search in 143 mutation-negative cases for rare co-segregating CNVs that were not present in 44 spouse controls and occurred in <0.1% of 6264 population controls. We identified four CNV regions (Figures 1c-1f) that showed disease co-segregation (occurred in multiple cases) within families; none of them was reported in any examined controls. A 112kb deletion region on chromosome 2q22.1 (Table 3, Figure 1c) showed complete co-segregation with melanoma, occurring in all three cases in a family but not in a spouse or two unaffected family members with DN (Figure 2c). The deletion region, which was initially identified by array-CGH (Figure 1c) and subsequently validated by SNP array and digital PCR, does not contain any coding genes, but does contain a pseudogene (*AC097721.1*) and a long intergenic noncoding RNA (lincRNA) *AC114763.1* (3277bp) as well as part of the lincRNA transcripts for AC097721.2 (transcripts 83,412bp and 67,300bp). Further, the deletion region is enriched for a number of enhancer regulatory elements in penile foreskin cell types including primary melanocytes (NIH Roadmap) and contains numerous DNAse I clusters and transcriptional factor binding interactions (ENCODE) (Supplementary Figure 1). The function of the lincRNAs is not known and we were not able to determine the expression levels of these transcripts in our RNASeq data or in normal skin cells or blood from NIH roadmap data.

We checked expression levels of a number of nearby coding genes including *SPOPL, DARS, CXCR4, NXPH2, LRP1b, HNMT, THSD7B, and YY1P2*, but none of these genes showed altered expression in deletion carriers. It is thus possible that the CNV deletion may spatially disrupt chromatin interactions and the expression of a target gene on a different chromosome in melanocytes. Investigating the expression and function of these lincRNAs in melanocytes and the potential impact of this CNV on chromatin interactions remain interesting biological questions for the future.

Another CNV region, a 480kb duplication on 3p12.2 (Figure 1d) that contains only one gene: *GBE1*, was initially seen in 2/2 cases analyzed in a family with four CMM cases (Table 3). The subsequent digital PCR confirmed the duplication in the two cases but not in the other two cases (Figure 2d). Further, the duplication was inherited from the two cases' unaffected father and not from the affected mother, who also had two affected siblings, which made this region less interesting.

The two other regions with rare CNVs (10q23.33 and 8q24.3) showed partial disease co-segregation. A duplication of 10q23.33 was identified by array-CGH (Figure 1e, Table 3) and validated by SNP array analyses. It was seen in 2/3 CMM cases and two individuals with DN in a single family (Figure 2e). The two DN individuals who carried the duplication were still young (<35 years old) at exam and below the age at which melanoma generally develops. The third case and an unaffected spouse did not harbor the duplication (Figure 2e). The duplication region contains two genes, *SORBS1* and *PDLIM1,* with *PDLIM1* only partially involved. Gene expression data for SORBS1 and PDLIM1 was available for one case with the duplication (1001) and one case without the duplication (1002) in this family. SORBS1 expression in 1001 was about 18 fold higher compared to 1002 or median expression among 50 other samples (Figure 3c). In contrast, PDLIM1 expression did not vary significantly by duplication status (data not shown). *SORBS1* encodes a CBL-associated protein (CAP) which is a cytoskeletal adaptor protein involved in modulating adhesion-mediated signaling events that lead to cell migration. There is

accumulating evidence suggesting that CAP may shuttle between cell adhesion sites and the nucleus and integrate cell adhesion signals with nuclear functions such as transcriptional control, growth factor signaling, and nuclear transport (Kioka *et al.*, 2002).

A small deletion at 8q24.3 (5-15 kb) (Figure 1f) was seen in 3/4 CMM cases in the same family from which the *PARP1* deletion was observed (Table 3). Although the size of the deletion was uncertain, we included the region in the subsequent analysis since the probe coverage for this region was dense (~10 probes) and the evidence for the deletion was strong. We further validated the deletion using a high-resolution Agilent CGH array platform (0.75 probes/bp) and found that the three CMM cases, two obligate gene carriers, and four unaffected family members harbored the deletion, whereas one CMM case and seven unaffected individuals did not carry the deletion (Figure 2a). The deletion, which was not seen in any other CMM cases or controls examined, partially involved the *ZNF517* gene. The two cases (1001 and 1002) with the deletion had the lowest ZNF517 expression among all 52 samples with RNASeq data (Figure 3d). Another unrelated case who also had similarly low ZNF517 expression did not have the deletion. The function of ZNF517 is not clear, although it may be involved in transcriptional regulation (Lorenz *et al.*, 2010).

In conclusion, several of these CNVs showed only partial segregation with disease, suggesting that they may not act as the major susceptibility mechanism in these families, although the occurrence of phenocopies due to other genes or environmental exposures is not infrequent in families with major gene mutations. Alternatively, these CNVs may function as disease modifiers that affect melanoma risk in concert with other genetic factors (such as rare or common SNPs), epigenetic mechanisms, or environmental/lifestyle exposures. Finally, we cannot rule out the possibility that these CNVs may reflect random variation and not relate to melanoma susceptibility. Compared to SNPs, CNVs influence larger regions of the genome and may have more complex genotype-phenotype relationships that are

not easily validated in experimental models. However, rare, exonic CNVs that segregate in more than one affected individual are infrequent and may potentially be disease-related, especially those that are correlated with gene expression changes. Our results based on high-resolution array-CGH data from a large number of melanoma-prone families suggest that the location of CNVs (i.e. CNVs disrupting important genes or regions), rather than the cumulative frequency of all CNVs, may play a more important role in melanoma susceptibility. Together with our previous findings of rare CNVs as major susceptibility factors in familial cancers, the results from this study further suggest that rare CNVs may contribute to melanoma susceptibility and genes found here warrant further evaluation in future studies of melanoma.

Materials and Methods

Study Population

The details of this family study have been previously described (Goldstein *et al.*, 2005; Goldstein *et al.*, 2000). In brief, melanoma families analyzed in this study were from the United States and included at least two living first-degree relatives with a history of invasive melanoma. All family members who were willing to participate in the study provided written informed consent under an NCI IRB approved protocol. All diagnoses of melanoma were confirmed by histologic review of pathologic material or by pathology reports. All study participants were of European ancestry.

Genome-wide CNV analyses

Array-CGH analysis: Nimblegen whole-genome 720K exon-focused comparative genomic hybridization (CGH) arrays were used to identify germline CNVs in family samples. We analyzed blood-derived DNA

from 139 individuals (74 CMM cases, 65 unaffected) from 26 mutation-negative families and 131 individuals (99 CMM, 32 unaffected) from 22 families with *CDKN2A* (N=20) or *CDK4* (N=2) mutations. The unaffected individuals included 52 unaffected family members and 45 genetically unrelated spouses. The hybridization, normalization, window-averaging, and segmentation analyses were performed by Nimblegen as previously described (Yang *et al.*, 2009).

We used the Nexus Copy Number™ built-in systematic correction to reduce data waviness and the FASST2 algorithm to identify significant CNVs ($P=1x10^{-6}$; number of probes per segment ≥5; log2 ratio >0.3 for gains and <-0.3 for losses). We only included CNVs > 10 kb, which was determined by the Nexus FASST2 algorithm, to reduce false positives. Sixteen samples with number of CNVs > 2 x standard deviation were removed from subsequent analyses.

SNP array: Genome-wide SNP genotyping (HumanOmniExpress 24v1.0, Illumina) was independently conducted on 120 CMM cases without known mutations (55 also analyzed by array-CGH). Log R ratio (LRR) and B allele frequency (BAF) were used to assess CNVs derived from SNP array data as previously described (Diskin *et al.*, 2008; Peiffer *et al.*, 2006; Staaf *et al.*, 2008). The LRR and BAF values from qualified assays (completion rate >= 90%) were re-normalized (Jacobs *et al.*, 2012) and then analyzed using custom software pipelines that involved BAF Segmentation packages (http://baseplugins.thep.lu.se/wiki/se.lu.onk.BAFsegmentation) to detect CNVs with a minimum of 20 probes per segment to minimize the false discovery. All potential events were plotted. False positive calls were excluded from the analysis based on manual review of each plot.

Technical validation of CNVs

Quantitative PCR/Digital PCR: Quantitative real-time PCR (qPCR) was performed using TaqMan® Copy Number Assays targeting *PARP1* (1q41-42), *CDKN2A* (9p21.3), and *ZNF517* (8q24.3) to confirm CNVs

identified by array-CGH using a standard protocol as previously described (Yang *et al.*, 2009). CNVs for regions 3p12.2 (*GBE1*) and 2q22.1 (upstream of *SPOPL*) were validated by droplet digital PCR (ddPCR) using the RainDrop system (RainDance Technologies, Billerica, MA) with predesigned TaqMan® copy number assays (Life Technologies, Grand Island, NY) multiplexed with a single copy gene (*RNaseP*). Data was analyzed in RainDrop Analyst software, where positive droplets were gated depending on the fluorescent dye detected, FAM (for CNV target) and VIC (for reference gene). The FAM count was divided by the VIC count and multiplied by 2 for diploid copy number.

High-resolution array-CGH: Agilent custom HD-CGH microarray, 8x15K, was used to validate the 8q24.3 region and to assess the co-segregation with disease in germline DNA of four CMM cases, two obligate carriers, eight unaffected family members, and three unrelated spouses in family T. Several control regions were included in the array design. The targeted region (~15 kb) was covered by more than 12,000 probes. Experimental assays and bioinformatics analyses were performed by Oxford Gene Technology (OGT).

RNA sequencing

Total RNA was extracted from Epstein–Barr-virus-transformed lymphoblastoid cell lines using a standard protocol. The RNA Integrity Number (RIN) value was > 9 for all samples. mRNA sequencing of 52 individuals (50 CMM cases and 2 unaffected spouses from mutation-negative families) was performed on the Illumina HiSeq2000/2500 platform with paired-end 100-bp mRNA-seq (~60 million total reads/sample). mRNA reads were mapped by Mapsplice algorithm with the hg19 reference genome database using the Mapsplice algorithm with default parameters and generic annotation file (version TCGA.hg19.June2011.gaf) to annotate genes and exons.

Data analyses

Association between global CNV burden and case-control status: For deletions, we defined G_{ni} as the global burden of deletions for subject i in family n. Let y_{ni} be the case control status with 1 for case and 0 for control. Each hemizygous deletion (referred to as CN1) and homozygous deletion (referred to as CN0) contributes 1 and 2, respectively, to G_{ni}. We assumed that the disease status y_{ni} is determined by G_{ni} according to a logistic model $logit(y_{ni} = 1|G_{ni}) = \alpha + \beta G_{ni}$. The score equation for testing $H_0: \beta = 0$ was derived as $S = \sum_n \sum_{i \in A_n} (y_{ni} - \mu) G_{ni}$, where μ was estimated as the proportion of cases across all families and A_n denotes the set of members of the n^{th} family. Let $\sigma^2 = Var(G_{ni})$ be the variance of the CNV burden. Under H_0 of no association, it is easy to verify that

$$Var_0(S) = \sigma^2 \sum_n \left(\sum_{i \in A_i} (y_{ni} - \mu)^2 + 2 \sum_{i<j} (y_{ni} - \mu)(y_{nj} - \mu) \phi_{n,ij} \right),$$

where $\phi_{n,ij}$ is the expected number of alleles identical by decent (IBD) for a relative pair (i, j) in the n^{th} family. Let $|A_n|$ denote the number of subjects in family n. We estimated σ^2 using all samples, i.e., $\hat{\sigma}^2 = \left(\sum_n \sum_{i \in A_n} (G_{ni} - \bar{G})^2 \right) / \sum_n |A_n|$, which provided an unbiased estimate of the variance. The score statistic for testing H_0 was derived as $Z = S/\sqrt{Var_0(S)}$, which followed the standard normal distribution under H_0 asymptotically. The same analysis procedure was applied for testing the association of global burden of duplications, where each duplication with three copies (referred to as CN3) contributed 1 and each duplication with four copies (referred to as CN4) contributed 2 to G_{ni}. When treating duplications and deletions equally, we combined duplications and deletions and used the same analytic framework for testing the association of global CNV burden. We repeated the association analysis to reduce potential batch effects by adjusting for a dummy variable representing two batches.

Next, we used the score statistic to examine whether the CNV burden differed between *CDKN2A/CDK4* mutation carriers and non-carriers in CMM cases. To reduce the potential for batch effects, we estimated the average CNV burden in batch 1 and batch 2 separately using spouse controls. The CNV burden for CMM cases was then subtracted by the estimated average CNV burden in specific batches. The adjusted CNV burden was tested for association between *CDKN2A/CDK4* mutation carriers and non-carriers in CMM cases.

Rare CNV identification: This analysis was restricted to CNVs that were >10 kb and not seen in 44 unrelated spouse controls. We first evaluated CNVs that involved known melanoma genes (Supplementary Table 1), which included established high-, intermediate-, and low-risk melanoma genes based on a literature search (Read *et al.*, 2016). For unknown genes/regions, we focused on CNVs that showed co-segregation with CMM within families (occurred in multiple CMM cases) that are mutation-negative.

Evaluation of CNVs identified among CMM cases in population controls: Population controls were drawn from five published GWASs investigating cancer-susceptibility risk and were analyzed in the Cancer Genomics Research Laboratory of the National Cancer Institute (Supplementary Table 2). Approval by the institutional review board for each study was confirmed, and written informed consent was obtained. LRR values from SNP genotyping data consisting of 6264 cancer-free controls were examined for deviations from expected log2 intensity ratios for each CNV region detected from CMM subjects. For each sample, an unequal variance t-test was applied to identify the mean LRR value for the tested region that was significantly different from that of chromosome 1 (used as the reference). P-value $<1.0\times10^{-6}$ and a mean difference of 0.06 were used to determine statistical significance for CNV calls.

Gene expression: We used an RSEM algorithm with default parameters to estimate the expression abundance at gene level for each sample. For a given gene, overall expression was calculated as Reads Per Kilobase of transcript per Million mapped reads (RPKM), determined as ((raw base counts / read length) × 10^{-9}) / (total reads × gene or transcript length).

Conflict of Interest: None.

Acknowledgement

We are indebted to the participating families, whose generosity and cooperation have made this study possible. We thank Roche Nimblegen, Oxford Gene Technology, and Biodiscovery for their molecular and informatic services. We thank investigators from the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), Cancer Prevention Study-II (CPS-II, American Cancer Society), Linxian Nutrition Intervention Trial (NITC), The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, and Shanxi Upper Gastrointestinal Cancer Genetics Study (SHNX), for allowing us to use control GWAS data. We also acknowledge the contributions to this work that were made by Virginia Pichler, Deborah Zametkin, Mary Fraser, and Barbara Rogers. This research was supported by the Intramural Research Program of the NIH, NCI, DCEG.

References

Al-Sukhni W, Joe S, Lionel AC, Zwingerman N, Zogopoulos G, Marshall CR, et al. Identification of germline genomic copy number variation in familial pancreatic cancer. Hum Genet. 2012;131:1481-94.

Bishop DT, Demenais F, Goldstein AM, Bergman W, Bishop JN, Bressac-de Paillerets B, et al. Geographical variation in the penetrance of CDKN2A mutations for melanoma. J Natl Cancer Inst. 2002;94:894-903.

Demenais F, Mohamdi H, Chaudru V, Goldstein AM, Newton Bishop JA, Bishop DT, et al. Association of MC1R variants and host phenotypes with melanoma risk in CDKN2A mutation carriers: a GenoMEL study. J Natl Cancer Inst. 2010;102:1568-83.

Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008;36:e126.

Goldstein AM, Chan M, Harland M, Gillanders EM, Hayward NK, Avril MF, et al. High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. Cancer Res. 2006;66:9818-28.

Goldstein AM, Landi MT, Tsang S, Fraser MC, Munroe DJ, Tucker MA. Association of MC1R variants and risk of melanoma in melanoma-prone families with CDKN2A mutations. Cancer Epidemiol Biomarkers Prev. 2005;14:2208-12.

Goldstein AM, Struewing JP, Chidambaram A, Fraser MC, Tucker MA. Genotype-phenotype relationships in U.S. melanoma-prone families with CDKN2A and CDK4 mutations. J Natl Cancer Inst. 2000;92:1006-10.

Goldstein AM, Tucker MA. Genetic epidemiology of cutaneous melanoma: a global perspective. Arch Dermatol. 2001;137:1493-6.

Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012;44:651-8.

Kioka N, Ueda K, Amachi T. Vinexin, CAP/ponsin, ArgBP2: a novel adaptor protein family regulating cytoskeletal organization and signal transduction. Cell structure and function. 2002;27:1-7.

Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet. 2005;76:8-32.

Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, et al. Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res. 2012;14:R24.

Kuiper RP, Ligtenberg MJ, Hoogerbrugge N, Geurts van Kessel A. Germline copy number variation and cancer risk. Curr Opin Genet Dev. 2010;20:282-9.

Lorenz P, Dietmann S, Wilhelm T, Koczan D, Autran S, Gad S, et al. The ancient mammalian KRAB zinc finger gene cluster on human chromosome 8q24.3 illustrates principles of C2H2 zinc finger evolution associated with unique expression profiles in human tissues. BMC genomics. 2010;11:206.

Macgregor S, Montgomery GW, Liu JZ, Zhao ZZ, Henders AK, Stark M, et al. Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. Nat Genet. 2011;43:1114-8.

Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res. 2006;16:1136-48.

Pylkas K, Vuorela M, Otsukka M, Kallioniemi A, Jukkola-Vuorinen A, Winqvist R. Rare copy number variants observed in hereditary breast cancer cases disrupt genes in estrogen signaling and TP53 tumor suppression network. PLoS Genet. 2012;8:e1002734.

Read J, Wadt KA, Hayward NK. Melanoma genetics. J Med Genet. 2016;53:1-14.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. Strong association of de novo copy number mutations with autism. Science 2007;316:445-9.

Shlien A, Tabori U, Marshall CR, Pienkowska M, Feuk L, Novokmet A, et al. Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. Proc Natl Acad Sci U S A. 2008;105:11264-9.

Staaf J, Vallon-Christersson J, Lindgren D, Juliusson G, Rosenquist R, Hoglund M, et al. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. BMC Bioinformatics. 2008;9:409.

Tucker MA, Goldstein AM. Melanoma etiology: where are we? Oncogene. 2003;22:3042-52.

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008;320:539-43.

Wang ZQ, Stingl L, Morrison C, Jantsch M, Los M, Schulze-Osthoff K, et al. PARP is important for genomic stability but dispensable in apoptosis. Genes Dev. 1997;11:2347-58.

Yang R, Chen B, Pfutze K, Buch S, Steinke V, Holinski-Feder E, et al. Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. Carcinogenesis. 2014;35:315-23.

Yang XHR, Ng D, Alcorta DA, Liebsch NJ, Sheridan E, Li SF, et al. T (brachyury) gene duplication confers major susceptibility to familial chordoma. Nature Genetics. 2009;41:1176-8.

Yang XR, Brown, K, Landi, MT, Ghiorzo, P, Badenas, C, Xu, M, et al. Duplication of CXC chemokine genes on chromosome 4q13 in a melanoma-prone family. Pigment Cell Melanoma Res. 2012;25:243-7.

Table 1. Median number of copy number variations (CNVs) per genome[1] in melanoma cases (CMM), unaffected family members, and unrelated spouses in high-risk melanoma families with and without *CDKN2A/CDK4* mutations.

CNVs	CMM (N=166)	Unaffected family members (N=44)	Spouses (N=44)
All CNVs[2]			
Total gain	11	12	10.5
Total loss	25	27.5	25
Total CNVs	37	39	34
>100kb CNVs[3]			
Total gain	3	3	3
Total loss	10	10	9
Total CNVs	13	13	12

[1]Based on Nimblegen 720k array-CGH data.

[2]We only included >10kb CNVs in the analysis. The size of the CNVs was determined by the Nexus FASST2 algorithm and small CNVs were excluded using the Nexus size filter option.

[3]These are a subset of all CNVs that are larger than 100kb.

Table 2. Rare[1] copy number variations (CNVs) in known melanoma susceptibility genes

Location[2]	CNV region	CNV type	Length	Subjects with CNV	Freq CMM with CNV	Genes
chr1:226262167-227585111	1q42.12-13[3]	Loss	1.3Mb	1 CMM	1/4	*BC032899, ACBD3, MIXL1, LIN9, PARP1, AK055856, C1orf95, ITPKB, PSEN2, CABC1, ACDK3, CDC42BPA, BC039356*
chr9:21987001-21997614	9p21.3[3]	Loss	10 kb	5 CMM, 1 OC[4], 1 DN	7/8[5]	*CDKN2A (p14Arf)*

[1]Not seen in 44 spouse controls; [2]Based on hg19; [3]Both CNV regions were identified by array-CGH and validated by qPCR; [4]OC: obligate carrier;

[5]CMM cases 2003 and 2004 were not analyzed but can be surmised to carry the deletion based on segregation.

Table 3. Rare[1] copy number variations (CNVs) co-segregating with disease in melanoma families without known mutations

Location[2]	CNV region	CNV type	Length	Subjects with CNV	Freq CMM with CNV	#6264 controls	Discovery	Validation	Genes
chr2:139139643-139249970	2q22.1	Loss	112 kb	3 CMM	3/3	0	Array-CGH SNP array	ddPCR	Near *SPOPL* *CXCR4*
chr3:81243088-81723365	3p12.2	Gain	480 kb	2 CMM	2/2	0	SNP array	ddPCR	*GBE1*
chr8:146022749-146032964	8q24.3	Loss	5-15 kb	3 CMM, 2 OC[3], 2 DN	3/4	ND[4]	Array-CGH	qPCR High-resolution Array-CGH	*ZNF517*
chr10:97014234-97264009	10q23.33	Gain	260kb	2 CMM, 2 DN	2/3	0	Array-CGH SNP array	SNP array	*CLP-36, SORBS1,*

									PDLIM1, *KIAA1296*

[1]Not seen in 44 spouse controls and present in <0.1% 6264 population controls; [2]Based on hg19; [3]OC: obligate carrier; [4]ND: not determined due to the small size of the deletion.

Figure legends:

Figure 1. Rare CNVs identified in melanoma-prone families.

a. A 1.3 MB deletion on 1q42.12-13 by array-CGH (720k exon-focused); b. a 10 kb deletion on 9p21.3 by array-CGH (390k 9p-focused); c. a 112 kb deletion on 2q22.1 by array-CGH (720k exon-focused); d. a 480 kb gain on 3p12.2 by SNP array; e. a 260 kb gain on 10q23.33 by array-CGH (720k exon-focused); f. a 5-15 kb deletion on 8q24.3 by array-CGH (720k exon-focused).

Figure 2. Families with identified copy number variations.

Right half black panel: CMM only; left quarter black panel: DN only; three-quarter black panel: CMM with DN. Number under each person (square as male, circle as female) indicates individual ID, which is unique within a single family.

a. Family T. The melanoma case with the *PARP1* deletion is indicated by an arrow. +: *ZNF517* deletion carriers; -: *ZNF517* deletion non-carriers; b. Family Z. +: *CDKN2A* deletion carriers; -: *CDKN2A* deletion non-carriers; CMM cases 2003 and 2004 can be surmised to carry the deletion based on segregation; c. Family D5. +: 2q22.1 deletion carriers; -: 2q22.1 deletion non-carriers; d. Family F10. +: 3p12.2 deletion carriers; -: 3p12.2 deletion non-carriers; e. Family A4. +: 10q23.33 duplication carriers; -: 10q23.33 duplication non-carriers.

Figure 3. Expression levels of genes affected by copy number variations.

Overall expression level for each gene was calculated as Reads Per Kilobase of transcript per Million mapped reads (RPKM) using RNASeq, which was performed using total RNA extracted from EBV-transformed lymphoblastoid cell lines from 50 familial melanoma cases and 2 unrelated spouse controls.