Published in final edited form as:

Hum Genet. 2012 September ; 131(9): . doi:10.1007/s00439-012-1183-1.

Identification of germline genomic copy number variation in familial pancreatic cancer

Wigdan Al-Sukhni,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada; Division of General Surgery, Department of Surgery, University of Toronto, Toronto, ON, Canada
wigdan.al.sukhni@utoronto.ca

Sarah Joe,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada

Anath C. Lionel,

Program in Genetics and Genome Biology, The Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada; Department of Molecular Genetics, McLaughlin Centre, University of Toronto, Toronto, ON, Canada

Nora Zwingerman,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada

George Zogopoulos,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada; Division of General Surgery, Department of Surgery, University of Toronto, Toronto, ON, Canada

Christian R. Marshall,

Program in Genetics and Genome Biology, The Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada

Ayelet Borgida,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada

Spring Holter,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada

Aaron Gropper,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada

Sara Moore,

Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada

Melissa Bondy,

MD Anderson Cancer Center, Houston, TX, USA

Alison P. Klein,

The Sol Goldman Pancreatic Research Center, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA

© Springer-Verlag 2012

S. Gallinger: steven.gallinger@uhn.on.ca.

Conflict of interest There are no conflicts of interest to be disclosed by any author of this manuscript.

Ethical approval All experiments described in this study comply with Canadian and US law and were approved by the appropriate ethics review boards at associated institutions.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-012-1183-1) contains supplementary material, which is available to authorized users.

Gloria M. Petersen,
Mayo Clinic College of Medicine, Rochester, MN, USA

Kari G. Rabe,
Mayo Clinic College of Medicine, Rochester, MN, USA

Ann G. Schwartz,
Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA

Sapna Syngal,
Dana-Farber Cancer Institute, Boston, MA, USA

Stephen W. Scherer, and
Program in Genetics and Genome Biology, The Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada; Department of Molecular Genetics, McLaughlin Centre, University of Toronto, Toronto, ON, Canada

Steven Gallinger
Samuel Lunenfeld Research Institute, Mount Sinai Hospital Toronto, ON, Canada; Division of General Surgery, Department of Surgery, University of Toronto, Toronto, ON, Canada; Toronto General Hospital, 10th Floor Rm 10EN206, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada

Abstract

Adenocarcinoma of the pancreas is a significant cause of cancer mortality, and up to 10 % of cases appear to be familial. Heritable genomic copy number variants (CNVs) can modulate gene expression and predispose to disease. Here, we identify candidate predisposition genes for familial pancreatic cancer (FPC) by analyzing germline losses or gains present in one or more high-risk patients and absent in a large control group. A total of 120 FPC cases and 1,194 controls were genotyped on the Affymetrix 500K array, and 36 cases and 2,357 controls were genotyped on the Affymetrix 6.0 array. Detection of CNVs was performed by multiple computational algorithms and partially validated by quantitative PCR. We found no significant difference in the germline CNV profiles of cases and controls. A total of 93 non-redundant FPC-specific CNVs (53 losses and 40 gains) were identified in 50 cases, each CNV present in a single individual. FPC-specific CNVs overlapped the coding region of 88 RefSeq genes. Several of these genes have been reported to be differentially expressed and/or affected by copy number alterations in pancreatic adenocarcinoma. Further investigation in high-risk subjects may elucidate the role of one or more of these genes in genetic predisposition to pancreatic cancer.

Introduction

Adenocarcinoma of the pancreas is the fourth leading cause of cancer death in North America. This disease is usually diagnosed at an advanced inoperable stage and pancreas cancer is generally resistant to chemotherapy and radio-therapy, thus overall 5-year survival rates are poor (<5 %). Even curative-intent surgery fails in about 80 % of resected cases as locoregional and systemic recurrences are common (Vincent et al. 2011). Improving outcome for pancreatic cancer patients depends on a better understanding of its etiology to foster early detection and intervention strategies. Although most pancreatic cancers are not obviously heritable, a significant fraction (5–10 %) of cases appears to be familial. About 20 % of hereditary pancreatic cancers are part of known cancer syndromes [Peutz-Jeghers Syndrome (PJS), *STK11*; familial atypical multiple mole melanoma (FAMMM), *p16/CDKN2A*; hereditary breast and ovarian cancer (HBOC), *BRCA2/BRCA1/PALB2*; and hereditary pancreatitis (HP), *PRSS1*], but most cases of familial pancreatic cancer (FPC) have an unknown genetic etiology (Hruban et al. 2010). Segregation analysis of families with multiple affected members suggests that FPC is caused by heritable alterations in at

least one rare “major gene”, likely in an autosomal dominant manner (Klein et al. 2002). Moreover, multiple case–control and cohort studies have demonstrated that members of FPC families, particularly those with an affected first-degree relative, have a significantly elevated lifetime risk of developing the disease (up to 32- to 56-fold) (Tersmette et al. 2001; Klein et al. 2004; Brune et al. 2010). However, to date traditional methods of linkage analysis for identifying predisposition genes have met with challenges in studying FPC, due in part to probable genetic heterogeneity as well as difficulty in collecting DNA specimens on multiple affected members in a family due to the rapid mortality of the disease.

Recently, it has become clear that submicroscopic copy number variants (CNVs) are prevalent throughout all genomes, accounting for at least 1.2 % of nucleotide variation between any two individuals (Pang et al. 2010). CNVs have been linked to rare genomic disorders (Gu and Lupski 2008) as well as common neurodevelopmental (Morrow 2010), psychiatric (Alaerts and Del-Favero 2009), autoimmune (Schaschl et al. 2009) and metabolic (Lanktree and Hegele 2008) diseases. Some studies have suggested an association between common CNVs and sporadic cancers [e.g. pancreatic cancer (6q13), Huang et al. 2012; neuroblastoma (1q21.1), Diskin et al. 2009; prostate cancer (2p24.3; 20p13; *GSTT1*), Liu et al. 2009; Jin et al. 2011; Nørskov et al. 2011; nasopharyngeal carcinoma (6p21.3), Tse et al. 2011; and endometrial cancer (*GSTT1*), Karageorgi et al. 2011]. The recent paper by Huang et al. (2012) is the first to describe an association of a germline CNV with pancreatic cancer risk: a common 10,379 bp deletion at 6q13 was found to be higher in frequency in sporadic pancreatic cancer patients compared to controls, with an odds ratio of 1.31 for one-copy carriers compared to two-copy carriers. Interestingly, functional analysis of this non-genic deletion suggested that it may be involved in long-range regulation of *CDKN2B*, an established tumor-suppressor gene.

In addition, it is well known that rare germline CNVs contribute to the genetic basis of familial cancer. Indeed, large germline genomic rearrangements cause 15 % of familial adenomatous polyposis (*APC* gene) (Michils et al. 2005), 2 % of breast and ovarian cancer (*BRCA1* gene) (Engert et al. 2008), and 5 % of Lynch Syndrome (*MSH2* and *MLH1* genes) (Taylor et al. 2003) cases. In 1–3 % of Lynch syndrome patients, the causative mutation is a large heritable deletion at the 3' end of the *TACSTD1* gene, which causes transcriptional read-through and epigenetic silencing of the adjacent *MSH2* gene (Kuiper et al. 2011). Furthermore, a report by Shlien et al. (2008) identified an elevated frequency of germline CNVs in individuals with Li Fraumeni syndrome (*TP53* mutation), and suggested that the increased predisposition to cancer in this syndrome may be proportional to the frequency of germline CNVs, many of which overlap known cancer genes.

Since germline CNVs implicated in familial cancers to date are rare with relatively high penetrance, we hypothesized that genes overlapped by rare germline genomic losses or gains identified exclusively in pancreatic cancer patients from high-risk families are candidate FPC genes. Here, we present an analysis of germline CNVs detected in 120 high-risk pancreatic cancer patients and compare them to CNVs in a large cohort of unaffected controls.

Materials and methods

This study was approved by the Research Ethics Boards at Mount Sinai Hospital and University Health Network in Toronto, Canada; Office for Human Research Studies at Dana Farber/Harvard Cancer Centre in Boston, Massachusetts; Institutional Review Board at Mayo Clinic in Rochester, Minnesota; Institutional Review Board at MD Anderson Cancer Centre in Houston, Texas; Office of Human Subjects Research at Johns Hopkins University

in Baltimore, Maryland; and Human Investigation Committee at Karmanos Cancer Institute, Wayne State University in Detroit, Michigan.

Cases and controls

Genomic DNA was extracted from peripheral blood or EBV-transformed cell lines of 133 pancreatic cancer patients from 131 high-risk families recruited by PAC-GENE (Pancreatic Cancer Genetic Epidemiology Consortium; PI, G Petersen, Mayo) (Petersen et al. 2006), a six-centre consortium that recruits kindreds containing two or more blood relatives affected with pancreatic cancer for genetic studies. Inclusion criteria in the current study included subjects with two or more affected relatives (“3 + FPC”; $N=79$); subjects with only one affected relative diagnosed at age 49 years or younger (“2 FPC”; $N=22$); and subjects without affected relatives who were diagnosed at age 49 years or younger (“single young”; $N=32$) (some of the families were reassigned based on updated information after analysis; see “Results”). We included young cases with no family history of pancreatic cancer because they may have de novo mutations in the gene(s) of interest, although we acknowledge that the definition of FPC involves more than one affected member in the family. Subjects were excluded if they carried known mutations or were in families with syndromes which predispose to pancreatic cancer (*BRCA2*, *BRCA1*, *p16FAMMM*, *STK11/PJS*, *PRSS1/HP*, Lynch syndrome). The majority of DNA samples were extracted from blood ($N=97$) and the remaining samples were from EBV-transformed lymphoblast cell lines (Online Resource 1). Of the original 128 FPC cases genotyped on the Affymetrix 500K array (see “SNP genotyping and CNV discovery”), eight were subsequently excluded (two subjects had excessively noisy data based on CNV count >40 per analysis run; one subject was discovered to have had chronic lymphocytic leukemia at the time of blood sample donation, making it difficult to distinguish germline from somatic CNVs detected in the sample; and five subjects no longer met inclusion criteria in light of new information that became available after the start of the study), leaving 120 cases in the final analysis with both NspI and StyI chips represented for each sample. Some of the subjects were reassigned to different inclusion criteria after updated information became available, resulting in 68 “3 + FPC” subjects, 28 “2 FPC” subjects, and 24 “single young” subjects contributing to the final set of case CNVs detected on Affymetrix 500K array.

Control samples of matched ancestry ($>95\%$ of cases and controls reported Caucasian ancestry) were obtained from two sources: 45 samples were healthy controls recruited by the Familial Gastrointestinal Cancer Registry (FGICR) (Madlensky et al. 1995) at Mount Sinai Hospital, Toronto, and 1,153 samples were recruited by the Ontario Familial Colon Cancer Registry (OFCCR) (Cotterchio et al. 2005). Almost all control DNA samples were extracted from blood (only 12 OFCCR controls were from lymphoblasts) (Online Resource 2). Two controls were discovered to have a history of sporadic pancreatic cancer (no affected relatives), and two other controls each reported having two relatives with pancreatic cancer, suggesting potential FPC kindreds. After excluding those four samples, 1,194 controls remained in the final analysis.

In addition, we had access to CNV data for 1,234 controls recruited through the Ottawa Heart Institute (OHI) (Stewart et al. 2009) and 1,123 controls of German descent recruited by the POPGEN project (Krawczak et al. 2006). Most of the OHI and POPGEN DNA samples were extracted from blood, and the platform for CNV detection was the Affymetrix 6.0 array.

SNP genotyping and CNV discovery

One hundred and twenty-eight cases and all 1,198 controls were genotyped at approximately 500,000 genome-wide SNPs on the Affymetrix GeneChip Human Mapping 500K Array

(NspI and StyI chips) according to standard protocol as provided by the manufacturer, and genotype calls were made by Affymetrix Genotyping Console (GTC 2.1) software. Copy number losses and gains were estimated using three validated Hidden Markov Model (HMM)-based algorithms (dCHIP, Li and Hung Wong 2001; CNAG, Nannya et al. 2005; and Partek Genomics Suite©). For each sample, all CNV calls from multiple algorithms and/or chips that overlapped by at least 20 % were merged, and CNVs supported by two or more algorithms or chips were considered to be the set of “high-confidence” calls. CNV loci were identified as regions where CNVs from multiple samples overlapped (case and control loci were identified separately) (Online Resource 3 and 4). For 236 controls, only one chip was included in the analysis (137 NspI only; 99 StyI only) due to inadequate hybridization of the second chip.

A subset of the original FPC cohort (33 samples) and five new cases were genotyped on the Affymetrix 6.0 array according to standard protocol as provided by the manufacturer to validate CNVs detected on the Affymetrix 500K array and to detect new CNVs. Arrays meeting Affymetrix quality control guidelines of Contrast QC >0.4 were used for further analysis. The Affymetrix Power Tools platform was used to extract normalized intensities for each array and inter-array intensity correlation was calculated; arrays with average correlation >0.9 were considered suitable for joint analysis. CNV calling was performed using the Birdsuite (Canary + Birdseye) (Korn et al. 2008) and iPattern (Pinto et al. 2011) algorithms, using a reference set that included the 38 FPC cases in addition to 100 other closely correlated Affymetrix 6.0 arrays genotyped at approximately the same time at The Centre for Applied Genomics (TCAG) in Toronto, Canada. (Samples were also analyzed on GTC 4.1, but these data were only used to support calls made on Birdsuite or iPattern.) Detection by either Birdsuite or iPattern was sufficient for the purpose of validating 500K array CNVs. Only “high-confidence” calls (i.e. called by at least two of Birdsuite, iPattern, and/or GTC 4.1 software) were included as novel FPC-specific CNVs.

To verify population ancestry of cases and controls, the program STRUCTURE (Pritchard et al. 2000) was used to estimate ancestry based on clustering of genotypes at 1,089 unlinked autosomal SNPs and assuming three ancestral populations. Genotypes for 270 HapMap samples from CEU, YRI, and CHB/JPT populations were used as reference of known ancestry during clustering. Ancestries were assigned using a co-efficient of ancestry threshold >0.9.

Quantitative PCR validation of CNVs

Quantitative PCR (qPCR) validation of some CNVs was performed using Invitrogen Platinum SYBR Green qPCR Supermix-UDG, with primers designed within the CNV of interest, and *MSH2* as a reference gene. Reactions were performed in replicates of 4–8x. After using a standard curve to confirm that primer efficiency was appropriate (90–110 %), data were analyzed using the ddC_t method (Schmittgen and Livak 2008).

FPC-specific CNVs

We defined “FPC-specific CNVs” as losses or gains detected in FPC cases on the 500K or Affymetrix 6.0 array, and which did not overlap (by 20 % or more) with losses or gains in FCIGR, OFCCR, OHI, or POPGEN controls nor overlapped CNVs reported from non-BAC-based platforms in the Database of Genomic Variants (DGV) (Zhang et al. 2006) (<http://projects.tcag.ca/variation>, updated November 2010). Although we did not control for ancestry in this analysis, we did note which FPC-specific CNVs were detected in non-Caucasian samples.

Annotation of CNVs

Affymetrix 500K and Affymetrix 6.0 array coordinates were aligned to the NCBI hg17 and NCBI hg18 human genome builds, respectively. Genes overlapped by CNVs were identified through the UCSC genome browser (<http://genome.ucsc.edu/>), using the respective human genome build. Information about CNV-overlapped genes was obtained from Entrez Gene (<http://www.ncbi.nlm.nih.gov/gene>) and Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed/>). The Memorial Sloan Kettering Cancer Centre (MSKCC) CancerGenes database (<http://cbio.mskcc.org/CancerGenes/Select.action>) (Higgins et al. 2007) was used to identify genes with reported pathways or functions linked to cancer development. The Wellcome-Trust Sanger Catalogue of Somatic Mutations in Cancer (COSMIC-version 55) database (<http://www.sanger.ac.uk/genetics/CGP/cosmic>) (Shepherd et al. 2011) identified genes that had previously reported point mutations, small insertions/deletions, gene fusions, and copy number alterations in various tumors and cancer cell lines. The Pancreatic Expression Database-version 2.0 (<http://www.pancreasexpression.org>) (Cutts et al. 2011) identified genes which were reported to be differentially expressed or which were overlapped by copy number alterations in pancreatic tumors or pre-invasive lesions.

Comparing CNV profile between cases and controls

Only “high-confidence” CNVs from non-EBV samples were included in the CNV profile comparison to minimize potential cell line artifacts and false calls (WTCCC et al.2010). As well, only controls with data available for both NspI and StyI chips were included in this comparison to minimize bias of undercalling CNVs in single-chip samples. To minimize CNV calling errors for “complex” CNVs (i.e. losses and gains in different samples overlapping the same region), we performed the “rare CNV” analysis only on regions reported as either losses or gains only. CNV loci that are present in fewer than 1 % of the total number of samples (cases + controls) were considered “rare”, excluding EBV samples and the complex CNVs. For losses, 32 cases and 235 controls (total 267 samples) were included in the “rare loss” analysis, so a rare loss was defined as present in fewer than three individuals. For gains, 56 cases and 551 controls (total 607 samples) were included in the “rare gain” analysis, so a rare gain was defined as present in fewer than seven samples.

Statistical analysis

Comparison of medians was performed using the Mann–Whitney *U* test and comparison of mean values was performed using the two-tailed Student’s *t* test with Levene’s test for equal variance. Testing for significant difference in proportions was performed with the two-tailed Fisher’s exact test. A *P* value <0.05 was considered significant. Statistical testing was performed using the SPSS© software package (version 17).

Results

Affymetrix 500K results and validation

STRUCTURE software was used for estimating population ancestry of the 120 FPC cases and 958 controls that had NspI + StyI chips available for analysis: 89.2 % of cases and 94.8 % of controls were Caucasian, 1.7 % of cases and 2.1 % of controls were Asian, and 9.2 % of cases and 3.1 % of controls were of admixed background.

The total number of autosomal CNVs identified in cases and controls was 873 and 10,794, respectively, of which 382 CNVs (123 losses + 259 gains) in cases and 3,115 CNVs (805 losses + 2310 gains) in controls were considered high-confidence calls (corresponding to 66 loss loci + 105 gain loci in cases and 313 loss loci + 467 gain loci in controls). A combination of qPCR and Affymetrix 6.0 genotyping demonstrated a validation rate of 95 % (121/127) for high-confidence CNVs but only 45 % (66/146) for low-confidence CNVs.

Therefore, the remainder of this analysis was limited to high-confidence CNVs in cases and controls. Approximately one-third (121/382) of all high-confidence case CNVs identified on the 500K array, corresponding to half (88/171) of all high-confidence CNV loci in cases, have been confirmed by either the Affymetrix 6.0 array and/or qPCR.

Table 1 compares several key CNV attributes between cases and controls (based on high-confidence CNVs and excluding EBV-derived samples and controls with only one chip in the analysis). Overall, no significant difference was observed in the CNV profile of cases and controls, including such parameters as CNV size, proportion of genic CNVs, proportion of rare CNVs, and average number of CNVs per individual genome. In both groups, gains were larger than losses (median size: cases 228.7 vs. 176.6 kb, $P=0.016$; controls 224.4 vs. 168.0 kb, $P<0.001$) and were more likely to overlap genes (cases 153/190 gains vs. 52/91 losses are genic, $P<0.001$; controls 1,641/2,059 gains vs. 400/731 losses are genic, $P<0.001$).

Affymetrix 6.0 results

In 36 cases genotyped on the Affymetrix 6.0 array (two of the original 38 samples were excluded due to excess noise; see “Methods”), a total of 3,364 autosomal CNVs (2,665 losses and 699 gains) were identified using Birdsuite, and 3,266 autosomal CNVs were identified using iPattern (1,975 losses + 1,291 gains). The high-confidence set comprised 2,187 CNVs (1,656 losses + 531 gains). The median size of high-confidence losses and gains was 12.7 kb (1 kb–1.4 Mb) and 48.9 kb (1 kb–1.6 Mb), respectively, and the average number of losses and gains per genome was 46 and 15, respectively.

FPC-specific CNVs

A total of 64 FPC CNVs (27 losses and 37 gains) detected on the 500K array were not identified in FGICR or OFCCR controls. After further excluding regions that overlapped POPGEN or OHI controls or were reported in the DGV, the number of FPC-specific CNVs identified on the 500K array is 37 CNVs (16 losses and 21 gains). On the Affymetrix 6.0 array, 119 FPC CNVs (71 losses and 48 gains) were not identified in POPGEN or OHI controls, and after further excluding regions which overlapped FGICR and OFCCR controls or were in the DGV, 73 FPC-specific CNVs (45 losses and 28 gains) remained. Combining results from the two arrays (including regions identified on both platforms) yielded a total of 93 non-redundant FPC-specific CNVs (53 losses and 40 gains), each CNV present in a single individual only (a total of 50 FPC cases, including 7 EBV-derived samples); 13 losses and 8 gains were in non-Caucasian individuals (Fig. 1 and Table 2).

Twenty-three FPC-specific losses and 23 FPC-specific gains overlapped introns, exons, and/or untranslated regions of 104 RefSeq genes. Fourteen genes (including one small nuclear RNA) had at least part of their coding regions affected by FPC-specific losses, and 74 genes (including 3 microRNAs) had at least part of their coding regions affected by FPC-specific gains (Online Resource 5). Fifty-five percent of those genes (48/88) have reported non-silent mutations (missense or nonsense variants, insertions/deletions, gene fusions) in different cancers according to the COSMIC v.55 database, whereas only 37 % of genes in all 500K + Affymetrix 6.0 FPC CNVs ($P=0.002$) and only 42 % of genes in all 500K + Affymetrix 6.0 control CNVs ($P=0.022$) had such mutations. None of the genes overlapped by FPC-specific losses were reported to have downregulated expression in pancreatic cancer in the Pancreatic Expression Database, whereas six genes overlapped by gains had reports of upregulation in pancreatic adenocarcinoma and three genes were reported to be upregulated in intraductal papillary mucinous neoplasm, a pre-invasive lesion. Furthermore, four FPC-specific gains overlapped regions reported to have high-level amplification in pancreatic adenocarcinoma in the Pancreatic Expression Database. The four gains overlap eight genes,

of which three genes (*LOC400643*, *DYNLRB2*, *LRRC30*, and *LAMA1*) are entirely encompassed by their respective gains. *LOC400643* is identified by NCBI Gene database as miscellaneous RNA with no information on its function. There are no reports of differential expression in pancreatic cancer or somatic mutations in *DYNLRB2*, which codes for a light chain component of cytoplasmic dynein 1 complex but this gene is reported to be involved in TGF-beta/SMAD3 signaling (Jin et al. 2009) and reported to be downregulated in hepatocellular carcinoma (Jiang et al. 2001). *LRRC30*, which codes for leucine-rich repeat-containing protein 30, has no reports of differential expression in pancreatic cancer or other association with tumorigenesis, but does have two reported mutations in the COSMIC database (one nonsense mutation in ovarian serous carcinoma and one missense mutation in hepatocellular carcinoma). *LAMA1* codes for laminin, an extracellular matrix component that binds to cells via high-affinity receptors and mediates attachment, migration, and organization of cells into tissues during embryogenesis. The COSMIC v.55 database reports 18 protein-altering or truncating somatic mutations in this gene in tumors of the pancreas, ovary, central nervous system, large intestine, breast, upper aerodigestive tract, and skin. In comparison, for 10,849 COSMIC v.55 database genes that had at least one non-silent/non-intronic mutation, the average number of mutations per gene is 3.7. A similar average number of reported somatic mutations is observed in genes affected by CNVs in our study (determined from the compiled data of 500K and Affymetrix 6.0 arrays): 3.6 mutations per gene for FPC-specific genes ($P = 0.983$), 3.4 mutations per gene for all FPC genes ($P = 0.821$), and 3.7 mutations per gene for all control genes ($P = 0.955$). There is also evidence for differential expression of *LAMA1* in tumors of sites other than the pancreas; one study reported hypermethylation and under-expression of *LAMA1* in colorectal cancer (Kim et al. 2011), while another study reported overexpression of this gene in glioblastoma (Scrideli et al. 2008).

Discussion

Identifying predisposition genes associated with FPC has been challenging due to the rapid lethality of the disease, low rate of tumor resection (resulting in paucity of tissue specimens for analysis), and probable genetic heterogeneity. An estimated 20 % of hereditary cases are linked to cancer syndromes caused by alterations in known genes. However, most families that demonstrate clustering of pancreatic cancer do not meet criteria for known cancer syndromes (Klein et al. 2002). We performed an analysis of germline CNVs in pancreatic cancer patients suspected to have a heritable genetic cause for their disease. These primarily included members of families with three or more affected cases, but also included families with only one or two affected cases if at least one of the cases was under age 50 at diagnosis.

To date, this is the largest study of germline CNVs in unrelated cancer patients from high-risk families. A previous study of 57 pancreatic cancer patients from 56 high-risk kindreds (each containing at least a pair of affected first-degree relatives) used an oligonucleotide-based CGH platform to identify FPC-specific germline CNVs, filtering out losses or gains that were also identified in 607 controls (372 were analyzed in the same study, and 235 were previously reported in two other studies) (Lucito et al. 2007). Twenty-five FPC-specific losses overlapping 81 genes and 31 FPC-specific gains overlapping 425 genes were identified. In our study, we investigated 133 members of 131 high-risk kindreds, of whom 17 subjects were part of the previous CGH study, and we identified 93 FPC-specific CNVs using a combination of Affymetrix 500K and Affymetrix 6.0 arrays. The median size of FPC-specific CNVs in the CGH study was larger than in our FPC-specific CNVs (losses 151 vs. 35.5 kb, gains 379 vs. 73 kb). This may be due, in part, to the lower resolution of the CGH platform (mean inter-marker distance 30 kb) compared to the Affymetrix 500K array (median inter-marker distance 2.5 kb) and Affymetrix 6.0 array (median inter-marker distance 0.7 kb) used in our study. It may also reflect enrichment for somatic CNVs caused

by EBV-transformation, since all FPC DNA samples in the CGH study were extracted from EBV-transformed cells, whereas only 29 samples in our population were EBV lymphoblasts. The size of control populations used to filter CNVs was larger in our study and the number of control CNVs from non-BAC studies currently catalogued in the DGV is greater than was available at the time of publication of the previous FPC CNV study. As a result, some of the CNVs identified as “FPC-specific” in the previous study overlapped CNVs in our controls and/or in the DGV. This may explain the slightly higher (FPC-specific CNVs)-to-sample ratio observed in the CGH study (approximately 1 CNV per sample) compared to our study (0.8 CNV per sample).

It is difficult to estimate concordance in CNV calling between the two studies, as we do not know how many of the 56 FPC-specific CNVs reported in the CGH study were identified in samples that were also used in our study. Only 1/25 loss and 3/31 gain loci reported in the CGH study were also observed in our analysis in samples common to both studies, and all of these overlapped CNVs in our controls and/or in the DGV. Interestingly, multiple reports have demonstrated generally low concordance for CNV calling on different platforms/ algorithms when analyzing the same DNA source (Pinto et al. 2011; Winchester et al. 2009). In addition to CNVs identified in cases common to both studies, there was one FPC-specific loss locus which was identified in two different subjects (one in each study). The region overlapped a gene, *DOCK1* (dedicator of cytokinesis 1), but in our study the loss only encompassed an intronic portion of the gene. This gene may have a role in cellular proliferation and migration (Wang et al. 2010; Sanders et al. 2009), and it has been reported to be over-expressed in high-grade dysplastic lesions (PanIN3), suggesting that it may be important in advancing tumorigenesis (Buchholz et al. 2005).

A number of other genome-wide germline CNV analyses have been reported for various cancers, but only a few have studied familial cancers. In addition to the aforementioned FPC study, microarray-based germline CNV studies have been reported for Li-Fraumeni syndrome (Shlien et al. 2008), young-onset and/or familial colorectal cancer in families without mutations in known predisposition genes (Venkatachalam et al. 2011), and *BRCA1*-associated ovarian cancer (Yoshihara et al. 2011). Shlien et al. described an increased frequency of germline CNVs in 33 Li-Fraumeni family members carrying mutations in the *TP53* gene (of which 23 were affected by cancer), compared to 20 Li-Fraumeni family members with wild-type *TP53* and 70 healthy controls. Since many of the CNVs overlapped or were near important cancer genes, the authors proposed a model whereby baseline genomic instability in these patients progresses over time, leading to more frequent and larger copy number alterations affecting genes that contribute to tumorigenesis. In our study, patients and controls had a similar number of alterations per genome, with similar CNV size, ratios of losses to gains, likelihood of CNVs to overlap genes, and proportion of genic CNVs that were associated with cancer. The lack of significant difference in the germline CNV profile between cases and controls suggests that causative genes for pancreatic cancer do not significantly impact genomic stability in non-tumor cells. Our results are similar to those of Yoshihara et al. who compared 68 Japanese subjects with germline *BRCA1* mutations (of whom 51 had ovarian cancer), 34 sporadic ovarian cancer patients, and 47 healthy controls. They reported no significant difference in the per-genome total number of CNVs between *BRCA1*-mutation carriers and controls, although the number of deletions was higher in the *BRCA1* subjects. Otherwise, they found no evidence for differential clustering of the global CNV data between groups, and no correlation of age at diagnosis with CNV frequency.

We identified 88 genes whose coding regions were partially or completely encompassed by FPC-specific CNVs, and although some are unlikely to be candidate FPC genes (e.g. olfactory receptor genes), many are functionally relevant to carcinogenesis, and are differentially expressed and/or overlap regions that are reported deleted or amplified in

pancreatic adenocarcinoma (Online Resource 5). Moreover, the proportion of genes that were reported in COSMIC v.55 to have protein-altering mutations in tumors or malignant cell lines was significantly higher in FPC-specific genes than in either the full population of cases or in controls. This further suggests that FPC-specific CNVs are enriched for cancer-associated genes. In the report by Yoshihara et al., the primary genetic etiology for the hereditary cancer was already known (*BRCA1*), and the authors presented genes overlapped by *BRCA1*-specific CNVs as potential modifiers to the development of cancer. Alternatively, the study by Venkatachalam et al. identified seven genic CNVs specific to patients with familial colorectal cancer who have no known genetic mutation, each CNV found in a single individual only. In that study, like ours, each gene is considered a potential causative gene for familial colorectal cancer. None of the genes overlapped by cancer patient CNVs reported by Shlien et al., Yoshihara et al., or Venkatachalam et al. were part of our FPC-specific gene list.

In conclusion, we have presented a list of candidate predisposition genes for FPC overlapped by germline CNVs that are specific to the largest cohort of high-risk pancreatic cancer patients published to date. One limitation of our analysis is the coverage and resolution of the platform we used for primary CNV discovery (i.e. Affymetrix 500K array). Since the completion of our study, novel methods of CNV detection have become available, including very high resolution tiling microarrays and next-generation sequencing. We expect future studies using these methods to independently test our findings and detect additional FPC candidate genes. Some of the samples containing FPC-specific CNVs in our study differed in ancestry from the majority of controls, raising the possibility that these CNVs are specific to the respective ancestry group rather than to pancreatic cancer risk. Those CNVs should be investigated further in a larger ethnicity-matched control cohort. Despite these limitations, our list of FPC-specific genes contains several interesting candidates and further screening for mutations in other high-risk pancreatic cancer subjects, along with investigation of the functional role of these genes, would add support to the role of one or more genes in predisposition to FPC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of funding were from National Institute of Health, Pancreatic Cancer Genetic Epidemiology Consortium (PACGENE) Grant#5R01CA097975-08; Lustgarten Foundation; Princess Margaret Hospital Invest-in-Research Fund; Canadian Institute for Health Research (Graduate Research Award; Vanier Doctoral Award); Canadian Society for Surgical Oncology; American Hepato-Pancreato-Biliary Association; Johnson & Johnson Medical Products/University of Toronto Surgeon Scientist Program Fellowship. We thank the following individuals for their contributions: A. Fiebig, A. Franke and S. Schreiber at POPGEN (University of Kiel, Kiel, Germany) and A. Stewart, R. McPherson and R. Roberts of the University of Ottawa Heart Institute (University of Ottawa, Ottawa, Canada) for generously providing population control microarray data; D. Merico and D. Pinto (The Centre for Applied Genomics, Toronto, Canada) for assistance with data review; G. Bader (University of Toronto, Toronto, Canada) and S. Narod (Women's College Hospital, Toronto, Canada) for review of the manuscript; H. Rothenmund (Samuel Lunenfeld Research Institute, Toronto, Canada) for her help with patient selection for the study; T. Selander of the Biospecimen Repository (Samuel Lunenfeld Research Institute, Toronto, Canada) for DNA extraction and storage. PACGENE coordinators (Mayo Clinic: J. Cogswell, T. Hammer, C. Wong; JHU: D. Echavarria; Karmanos: K. Sargent; MDA: P. Adatto, G. Armstrong; Dana Farber: M. Lenn, M. Perez, C. Ukaegbu).

References

Alaerts M, Del-Favero J. Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future. *Hum Mutat.* 2009; 30:1139–1152. [PubMed: 19626716]

- Brune KA, Lau B, Palmisano E, Canto M, Goggins MG, Hruban RH, Klein AP. Importance of age of onset in pancreatic cancer kindreds. *J Natl Cancer Inst.* 2010; 102:119–126. [PubMed: 20068195]
- Buchholz M, Braun M, Heidenblut A, Kestler HA, Klöppel G, Schmiegeler W, Hahn SA, Lüttges J, Gress TM. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. *Oncogene.* 2005; 24:6626–6636. [PubMed: 16103885]
- Cotterchio M, Manno M, Klar N, McLaughlin J, Gallinger S. Colorectal screening is associated with reduced colorectal cancer risk: a case-control study within the population-based Ontario Familial Colorectal Cancer Registry. *Cancer Causes Control.* 2005; 16:865–875. [PubMed: 16132797]
- Cutts RJ, Gadaleta E, Hahn SA, Crnogorac-Jurcovic T, Lemoine NR, Chelala C. The Pancreatic Expression database: 2011 update. *Nucleic Acids Res.* 2011; 39(Database issue):D1023–1028. [PubMed: 20959292]
- Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, Cole K, Mossé YP, Wood A, Lynch JE, Pecor K, Diamond M, Winter C, Wang K, Kim C, Geiger EA, McGrady PW, Blakemore AI, London WB, Shaikh TH, Bradfield J, Grant SF, Li H, Devoto M, Rappaport ER, Hakonarson H, Maris JM. Copy number variation at 1q21.1 associated with neuroblastoma. *Nature.* 2009; 459:987–991. [PubMed: 19536264]
- Engert S, Wappenschmidt B, Betz B, Kast K, Kutsche M, Hellebrand H, Goecke TO, Kiechle M, Niederacher D, Schmutzler RK, Meindl A. MLPA screening in the BRCA1 gene from 1,506 German hereditary breast cancer cases: novel deletions, frequent involvement of exon 17, and occurrence in single early-onset cases. *Hum Mutat.* 2008; 29:948–958. [PubMed: 18431737]
- Gu W, Lupski JR. CNV and nervous system diseases—what's new? *Cytogenet Genome Res.* 2008; 123:54–64. [PubMed: 19287139]
- Higgins ME, Claremont M, Major JE, Sander C, Lash AE. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.* 2007; 35(Database issue):D721–D726. [PubMed: 17088289]
- Hruban RH, Canto MI, Goggins M, Schulick R, Klein AP. Update on familial pancreatic cancer. *Adv Surg.* 2010; 44:293–311. [PubMed: 20919528]
- Huang L, Yu D, Wu C, Zhai K, Jiang G, Cao G, Wang C, Liu Y, Sun M, Li Z, Tan W, Lin D. Copy number variation at 6q13 functions as a long-range regulator and is associated with pancreatic cancer risk. *Carcinogenesis.* 2012; 33(1):94–100. [PubMed: 22016467]
- Jiang J, Yu L, Huang X, Chen X, Li D, Zhang Y, Tang L, Zhao S. Identification of two novel human dynein light chain genes, DNLC2A and DNLC2B, and their expression changes in hepatocellular carcinoma tissues from 68 Chinese patients. *Gene.* 2001; 281:103–113. [PubMed: 11750132]
- Jin Q, Gao G, Mulder KM. Requirement of a dynein light chain in TGFβ₂/Smad3 signaling. *J Cell Physiol.* 2009; 221:707–715. [PubMed: 19711352]
- Jin G, Sun J, Liu W, Zhang Z, Chu LW, Kim ST, Sun J, Feng J, Duggan D, Carpten JD, Wiklund F, Grönberg H, Isaacs WB, Zheng SL, Xu J. Genome-wide copy-number variation analysis identifies common genetic variants at 20p13 associated with aggressiveness of prostate cancer. *Carcinogenesis.* 2011; 32:1057–1062. [PubMed: 21551127]
- Karageorgi S, Prescott J, Wong JY, Lee IM, Buring JE, De Vivo I. GSTM1 and GSTT1 copy number variation in population-based studies of endometrial cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2011; 20:1447–1452. [PubMed: 21558497]
- Kim YH, Lee HC, Kim SY, Yeom YI, Ryu KJ, Min BH, Kim DH, Son HJ, Rhee PL, Kim JJ, Rhee JC, Kim HC, Chun HK, Grady WM, Kim YS. Epigenomic analysis of aberrantly methylated genes in colorectal cancer identifies genes commonly affected by epigenetic alterations. *Ann Surg Oncol.* 2011; 18:2338–2347. [PubMed: 21298349]
- Klein AP, Beaty TH, Bailey-Wilson JE, Brune KA, Hruban RH, Petersen GM. Evidence for a major gene influencing risk of pancreatic cancer. *Genet Epidemiol.* 2002; 23:133–149. [PubMed: 12214307]
- Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJ, Griffin C, Cameron JL, Yeo CJ, Kern S, Hruban RH. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. *Cancer Res.* 2004; 64:2634–2638. [PubMed: 15059921]
- Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. Integrated

genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet.* 2008; 40:1253–1260. [PubMed: 18776909]

- Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* 2006; 9:55–61. [PubMed: 16490960]
- Kuiper RP, Vissers LE, Venkatachalam R, Bodmer D, Hoenselaar E, Goossens M, Haufe A, Kamping E, Niessen RC, Hogervorst FB, Gille JJ, Redeker B, Tops CM, van Gijn ME, van den Ouweland AM, Rahner N, Steinke V, Kahl P, Holinski-Feder E, Morak M, Kloor M, Stemmler S, Betz B, Hutter P, Bunyan DJ, Syngal S, Culver JO, Graham T, Chan TL, Nagtegaal ID, van Krieken JH, Schackert HK, Hoogerbrugge N, van Kessel AG, Ligtenberg MJ. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. *Hum Mutat.* 2011; 32:407–414. [PubMed: 21309036]
- Lanktree M, Hegele RA. Copy number variation in metabolic phenotypes. *Cytogenet Genome Res.* 2008; 123:169–175. [PubMed: 19287152]
- Li C, Hung Wong W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.* 2001; 2(8) RESEARCH0032.
- Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim ST, Sun J, Wiklund F, Wiley K, Isaacs SD, Stattin P, Xu J, Duggan D, Carpten JD, Isaacs WB, Grönberg H, Zheng SL, Chang BL. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res.* 2009; 69:2176–2179. [PubMed: 19258504]
- Lucito R, Suresh S, Walter K, Pandey A, Lakshmi B, Krasnitz A, Sebat J, Wigler M, Klein AP, Brune K, Palmisano E, Maitra A, Goggins M, Hruban RH. Copy-number variants in patients with a strong family history of pancreatic cancer. *Cancer Biol Ther.* 2007; 6:1592–1599. [PubMed: 17912030]
- Madlensky L, Berk TC, Bapat BV, McLeod RS, Couture J, Baron D, Hiruki T, Redston M, Cohen Z, Gallinger S. A preventive registry for hereditary nonpolyposis colorectal cancer. *Can J Oncol.* 1995; 5:355–360. [PubMed: 8853507]
- Michils G, Tejpar S, Thoelen R, van Cutsem E, Vermeesch JR, Fryns JP, Legius E, Matthijs G. Large deletions of the APC gene in 15 % of mutation-negative patients with classical polyposis (FAP): a Belgian study. *Hum Mutat.* 2005; 25:125–134. [PubMed: 15643602]
- Morrow EM. Genomic copy number variation in disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry.* 2010; 49:1091–1104. [PubMed: 20970697]
- Müller A, Holzmann K, Kestler HA. Visualization of genomic aberrations using Affymetrix SNP arrays. *Bioinformatics.* 2007; 23:496–497. [PubMed: 17138589]
- Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, Ogawa S. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* 2005; 65:6071–6079. [PubMed: 16024607]
- Nørskov MS, Frikke-Schmidt R, Bojesen SE, Nordestgaard BG, Loft S, Tybjaerg-Hansen A. Copy number variation in glutathione-S-transferase T1 and M1 predicts incidence and 5-year survival from prostate and bladder cancer, and incidence of corpus uteri cancer in the general population. *Pharmacogenomics J.* 2011; 11:292–299. [PubMed: 20514077]
- Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, Park H, Hurles ME, Lee C, Venter JC, Kirkness EF, Levy S, Feuk L, Scherer SW. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol.* 2010; 11:R52. [PubMed: 20482838]
- Petersen GM, de Andrade M, Goggins M, Hruban RH, Bondy M, Korczak JF, Gallinger S, Lynch HT, Syngal S, Rabe KG, Seminara D, Klein AP. Pancreatic cancer genetic epidemiology consortium. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:704–710. [PubMed: 16614112]
- Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, Macdonald JR, Mills R, Prasad A, Noonan K, Gribble S, Prigmore E, Donahoe PK, Smith RS, Park JH, Hurles ME, Carter NP, Lee C, Scherer SW, Feuk L. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol.* 2011; 29:512–520. [PubMed: 21552272]

- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
- Sanders MA, Ampasala D, Basson MD. DOCK5 and DOCK1 regulate Caco-2 intestinal epithelial cell spreading and migration on collagen IV. *J Biol Chem*. 2009; 284:27–35. [PubMed: 19004829]
- Schaschl H, Aitman TJ, Vyse TJ. Copy number variation in the human genome and its implication in autoimmunity. *Clin Exp Immunol*. 2009; 156:12–16. [PubMed: 19220326]
- Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008; 3:1101–1108. [PubMed: 18546601]
- Scrideli CA, Carlotti CG Jr, Okamoto OK, Andrade VS, Cortez MA, Motta FJ, Lucio-Eterovic AK, Neder L, Rosenberg S, Oba-Shinjo SM, Marie SK, Tone LG. Gene expression profile analysis of primary glioblastomas and non-neoplastic brain tissue: identification of potential target genes by oligonucleotide microarray and real-time quantitative PCR. *J Neurooncol*. 2008; 88:281–291. [PubMed: 18398573]
- Shepherd R, Forbes SA, Beare D, Bamford S, Cole CG, Ward S, Bindal N, Gunasekaran P, Jia M, Kok CY, Leung K, Menzies A, Butler AP, Teague JW, Campbell PJ, Stratton MR, Futreal PA. Data mining using the catalogue of somatic mutations in cancer BioMart. Database (Oxford). 2011 2011. bar018 (print).
- Shlien A, Tabori U, Marshall CR, Pienkowska M, Feuk L, Novokmet A, Nanda S, Druker H, Scherer SW, Malkin D. Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc Natl Acad Sci USA*. 2008; 105:11264–11269. [PubMed: 18685109]
- Stewart AF, Dandona S, Chen L, Assogba O, Belanger M, Ewart G, LaRose R, Doelle H, Williams K, Wells GA, McPherson R, Roberts R. Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa Heart Genomics Study. *J Am Coll Cardiol*. 2009; 53:1471–1472. [PubMed: 19371834]
- Taylor CF, Charlton RS, Burn J, Sheridan E, Taylor GR. Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. *Hum Mutat*. 2003; 22:428–433. [PubMed: 14635101]
- Tersmette AC, Petersen GM, Offerhaus GJ, Falatko FC, Brune KA, Goggins M, Rozenblum E, Wilentz RE, Yeo CJ, Cameron JL, Kern SE, Hruban RH. Increased risk of incident pancreatic cancer among first-degree relatives of patients with familial pancreatic cancer. *Clin Cancer Res*. 2001; 7:738–744. [PubMed: 11297271]
- Tse KP, Su WH, Yang ML, Cheng HY, Tsang NM, Chang KP, Hao SP, Yao Shugart Y, Chang YS. A gender-specific association of CNV at 6p21.3 with NPC susceptibility. *Hum Mol Genet*. 2011; 20:2889–2896. [PubMed: 21536588]
- Venkatachalam R, Verwiel ET, Kamping EJ, Hoenselaar E, Görgens H, Schackert HK, van Krieken JH, Ligtenberg MJ, Hoogerbrugge N, van Kessel AG, Kuiper RP. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. *Int J Cancer*. 2011; 129:1635–1642. [PubMed: 21128281]
- Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. *Lancet*. 2011; 378:607–620. [PubMed: 21620466]
- Wang H, Linghu H, Wang J, Che YL, Xiang TX, Tang WX, Yao ZW. The role of Crk/Dock180/Rac1 pathway in the malignant behavior of human ovarian cancer cell SKOV3. *Tumour Biol*. 2010; 31:59–67. [PubMed: 20237902]
- Wellcome Trust Case Control Consortium. et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*. 2010; 464:713–720. [PubMed: 20360734]
- Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic*. 2009; 8:353–366. [PubMed: 19737800]
- Yoshihara K, Tajima A, Adachi S, Quan J, Sekine M, Kase H, Yahata T, Inoue I, Tanaka K. Germline copy number variations in BRCA1-associated ovarian cancer patients. *Genes Chromosomes Cancer*. 2011; 50:167–177. [PubMed: 21213370]

Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res.* 2006; 115:205–214. [PubMed: 17124402]

Fig. 1. Ideogram indicating chromosomal location of FPC-specific losses and gains; *unfilled circle* loss, *filled box* gain (image generated by IdeogramBrowser online tool <http://www.informatik.uni-ulm.de/ni/staff/HKestler/ideo/>; Müller et al. 2007)

Table 1

High-confidence CNV profile of cases versus controls (excluding EBV-derived samples and excluding controls with data from only one chip)

	FPC cases	Controls	P value
No. of lymphocyte samples	91	950	
No. of high-confidence losses/high-confidence gains	91/190	731/2,059	
Median CNV size (range)	219.5 kb (1.2 kb–6.4 Mb)	219.5 kb (1.2 kb–6.8 Mb)	0.439
Median CNV SNP count (range)	42 (2–417)	40 (2–1,318)	0.578
No. of genic CNVs/all CNVs			
Losses	52/91 (57 %)	400/731 (55 %)	0.738
Gains	153/190 (81 %)	1,646/2,059 (80 %)	0.850
No. of samples with genic CNVs/samples with any CNVs			
Losses	43/59 (73 %)	327/500 (65 %)	0.309
Gains	70/75 (93 %)	765/816 (94 %)	0.805
No. of CNV genes identified as “Cancer Genes” in MSKCC CancerGenes database/all CNV genes recognized by the MSKCC database			
Losses	8/36 (22 %)	35/264 (13 %)	0.201
Gains	53/335 (16 %)	507/2,940 (17 %)	0.541
No. of CNV loci included in rare analysis/all CNV loci			
Losses	36/52 (69 %)	203/290 (70 %)	1.000
Gains	65/83 (78 %)	349/428 (82 %)	0.541
No. of CNVs that are part of rare loci/all CNVs			
Losses	23/91 (25 %)	199/731 (27 %)	0.802
Gains	47/190 (25 %)	461/2,059 (22 %)	0.469
No. of samples with CNVs included in rare analysis/samples with any CNV			
Losses	32/59 (54 %)	235/500 (47 %)	0.335
Gains	56/75 (75 %)	551/816 (68 %)	0.244
No. of samples with rare CNVs/samples with any CNV			
Losses	21/59 (36 %)	169/500 (34 %)	0.773
Gains	37/75 (49 %)	348/816 (43 %)	0.275
No. of genic rare CNVs/all rare CNVs			
Losses	10/23 (43 %)	69/199 (35 %)	0.491
Gains	33/47 (70 %)	330/461 (72 %)	0.866
No. of samples with genic rare CNVs/samples with rare CNVs			
Losses	10/21 (48 %)	63/169 (37 %)	0.476
Gains	27/37 (73 %)	267/348 (77 %)	0.684
Mean CNVs per genome ^a			
Losses	1.5	1.5	0.443
Gains	2.5	2.5	0.956
Mean rare CNVs per genome ^a			
Losses	0.4	0.4	0.919

	FPC cases	Controls	<i>P</i> value
Gains	0.6	0.6	0.498

^aMean and *t* test calculated for losses and gains based only on samples with at least one high-confidence loss or gain, respectively (to avoid the bias of samples which did not get a high-confidence CNV call due to noise)

Table 2

FPC-specific CNVs coordinates and size: list of RefSeq genes overlapping FPC-specific CNVs and list of overlapping amplifications or deletions reported in pancreatic cancer tumor studies (from the Pancreatic Expression Database)

CNV type	CNV Id	Sample ID	Coordinates (hg18)	Size (kb)	RefSeq Genes (includes genes in which only intron is affected)	Overlaps Pancreatic Expression Database CNVs?
Loss	Affy6.0_D_477	97	chr1:62528216–62538049	9.8	KANK4	No
Loss	Affy6.0_D_535	62	chr2:41356049–41390177	34.1	None	No
Loss	Affy500K_D_114 and Affy6_D_74	62	chr2:41474986–41608172	133.2	None	No
Loss	Affy6.0_D_637	204	chr2:55010996–55019655	8.7	EML6	No
Loss	Affy6.0_D_549	99	chr2:78025162–78059816	34.7	SNAR-H	No
Loss	Affy6.0_D_542	61	chr2:148426768–148464448	37.7	ORC4	No
Loss	Affy6.0_D_677	20 (admixed)	chr3:22405124–22481450	76.3	None	No
Loss	Affy500K_D_152	85	chr3:125676839–125815545	138.7	KALRN	No
Loss	Affy6.0_D_671	30	chr3:192351519–192375879	24.4	None	No
Loss	Affy6.0_D_741	28	chr4:53829489–53875712	46.2	SCFD2	No
Loss	Affy6.0_D_739	69	chr4:70867305–70952889	85.6	HTN1, HTN3, STATH	No
Loss	Affy6.0_D_769	28	chr4:123803190–123806840	3.7	None	No
Loss	Affy500K_D_98 and Affy6_D_743	54	chr4:147802903–148190197	387.3	TTC29	No
Loss	Affy6.0_D_844	40	chr4:178997998–179018809	20.8	LOC285501	No
Loss	Affy6.0_D_870	123 (admixed)	chr5:11220277–11229088	8.8	CTNND2	No
Loss	Affy500K_D_234	117 (EBV)	chr5:95640616–96152064	511.4	PCSK1, ERAP1, CAST	High-level amplification
Loss	Affy6.0_D_911	123 (admixed)	chr6:119578774–119604698	25.9	MAN1A1	No
Loss	Affy6.0_D_930	35	chr6:142219243–142324891	105.6	None	No
Loss	Affy6.0_D_925	101	chr6:162342089–162365931	23.8	PARK2	No
Loss	Affy6.0_D_992	64 (admixed)	chr7:23094182–23110722	16.5	None	No
Loss	Affy6.0_D_1029	64 (admixed)	chr8:2578046–2587479	9.4	None	No
Loss	Affy500K_D_220	112 (EBV)	chr8:6371546–6430547	59.0	MCPH1, ANGPT2	No
Loss	Affy500K_D_142	77 (admixed)	chr8:17998784–18145035	146.3	NAT1	No
Loss	Affy6.0_D_1650	20 (admixed)	chr8:58080498–58091757	11.3	None	No
Loss	Affy6.0_D_1065	61	chr8:85558196–85579549	21.4	RALYL	No

CNV type	CNV Id	Sample ID	Coordinates (hg18)	Size (kb)	RefSeq Genes (includes genes in which only intron is affected)	Overlaps Pancreatic Expression Database CNVs?
Loss	Affy6.0_D_1069	125	chr8:88575501-88585299	9.8	None	No
Loss	Affy500K_D_93	48	chr8:89782116-89849946	67.8	None	No
Loss	Affy6.0_D_1644	91	chr8:131657747-131683625	25.9	None	No
Loss	Affy6.0_D_1024	16	chr8:138328381-138425832	97.5	None	No
Loss	Affy500K_D_134	74	chr9:2235919-2351848	115.9	None	No
Loss	Affy500K_D_43 and Affy6_D_1112	17	chr9:75525136-75638229	113.1	None	No
Loss	Affy6.0_D_1109	64 (admixed)	chr9:75637796-75657448	19.7	None	No
Loss	Affy6.0_D_1103	27	chr9:95533791-95585819	52.0	None	No
Loss	Affy6.0_D_1108	4	chr9:102517861-102553347	35.5	None	No
Loss	Affy6.0_D_1127	123 (admixed)	chr10:128752241-128780181	27.9	DOCK1	No
Loss	Affy500K_D_40 and Affy6_D_1198	16	chr11:39882017-40010124	128.1	None	No
Loss	Affy500K_D_6	2 (EBV)	chr11:89730130-89888327	158.2	None	No
Loss	Affy6.0_D_1205	204	chr11:104741261-104793318	52.1	None	No
Loss	Affy500K_D_125 and Affy6_D_1246	68	chr12:39394850-39501843	107.0	CNTN1	High-level amplification
Loss	Affy500K_D_83 and Affy6_D_1253	40	chr12:130382166-130686668	304.5	None	No
Loss	Affy6.0_D_1336	101	chr13:39389124-39515818	126.7	None	No
Loss	Affy500K_D_24 and Affy6_D_1342	11 (Asian)	chr13:93544008-93670507	126.5	GPC6	High-level amplification
Loss	Affy6.0_D_1377	54	chr14:42513084-42541303	28.2	None	No
Loss	Affy500K_D_121 and Affy6_D_1383	64 (admixed)	chr14:85216336-85436133	219.8	None	No
Loss	Affy6.0_D_1679	68	chr15:57862260-57891107	28.8	None	No
Loss	Affy6.0_D_1428	35	chr15:60314660-60333770	19.1	None	No
Loss	Affy6.0_D_1467	67	chr16:54046835-54056160	9.3	None	No
Loss	Affy6.0_D_1484	35	chr17:75852813-75870192	17.4	RNF213	No
Loss	Affy6.0_D_1507	11 (Asian)	chr18:3670476-3715553	45.1	DLGAP1	No
Loss	Affy6.0_D_1527	203	chr18:38603464-38605275	1.8	RIT2	No
Loss	Affy6.0_D_1548	61	chr19:61684427-61697318	12.9	LOC100128252	No
Loss	Affy6.0_D_1601	11 (Asian)	chr20:50766640-50780316	13.7	None	No

CNV type	CNV Id	Sample ID	Coordinates (hg18)	Size (kb)	RefSeq Genes (includes genes in which only intron is affected)	Overlaps Pancreatic Expression Database CNVs?
Loss	Affy500K_D_225	114 (EBV)	chr21:23160325-23267106	106.8	None	No
Gain	Affy6.0_G_11	127	chr1:49856085-50089082	233.0	AGBL4	No
Gain	Affy6.0_G_3	123 (admixed)	chr1:157133096-157188413	55.3	PYHIN1	No
Gain	Affy6.0_G_33	69	chr2:21646517-216485115	19.6	None	No
Gain	Affy500K_G_303	65	chr2:230753632-230823051	69.4	SP110, SP140	No
Gain	Affy500K_G_380	82	chr3:143693491-143928895	235.4	ATR, PLS1, TRPC1	No
Gain	Affy6.0_G_39	123 (admixed)	chr3:161448573-161518365	69.8	IFT80	No
Gain	Affy500K_G_88	24	chr4:26691114-26985948	294.8	None	No
Gain	Affy500K_G_369	80	chr4:29195980-29209908	13.9	None	No
Gain	Affy500K_G_602 and Affy6_G_50	123 (admixed)	chr4:72734028-72817447	83.4	None	No
Gain	Affy500K_G_365	79	chr4:93344017-93591992	248.0	GRID2	No
Gain	Affy500K_G_511	107 (EBV)	chr4:105853937-106127766	273.8	None	No
Gain	Affy500K_G_407	86	chr6:48829836-49492706	662.9	None	No
Gain	Affy6.0_G_70	44	chr6:132466247-132479169	12.9	None	No
Gain	Affy500K_G_615 and Affy6_G_77	125	chr7:133223330-133393933	170.6	EXOC4	No
Gain	Affy500K_G_603/604 and Affy6_G_93	123 (admixed)	chr8:39935640-39943638	8.0	IDO2	No
Gain	Affy6.0_G_90	202	chr8:49008716-49049657	40.9	PRKDC, MCM4	No
Gain	Affy6.0_G_95	99	chr8:83294045-83332227	38.2	None	No
Gain	Affy500K_G_69 and Affy6_G_87	18	chr8:108696004-109010881	314.9	RSPO2	High-level amplification
Gain	Affy500K_G_49	12 (admixed)	chr9:81978854-82021829	43.0	None	No
Gain	Affy6.0_G_138	54	chr10:4497158-4555255	58.1	None	No
Gain	Affy500K_G_95	26	chr10:19849680-20589237	739.6	PLXDC2	High-level amplification
Gain	Affy6.0_G_143	17	chr10:71778181-71797516	19.3	LRRC20	No
Gain	Affy6.0_G_152	54	chr11:41420026-41456633	36.6	None	No
Gain	Affy6.0_G_170	20 (admixed)	chr11:65027491-65201466	174.0	LTBP3, PCNXL3, MAP3K11, MIR4489, MALAT1, RELA, SIPA1, SSSCA1, FAM89B, KCNK7, MIR4690, EHBPI1, LOC254100,	No

CNV type	CNV Id	Sample ID	Coordinates (hg18)	Size (kb)	RefSeq Genes (includes genes in which only intron is affected)	Overlaps Pancreatic Expression Database CNVs?
SCYL1						
Gain	Affy500K_G_622 and Affy6_G_158	126	chr11:81521790-81598468	76.7	None	No
Gain	Affy500K_G_502	106 (EBV)	chr12:57378034-57482408	104.4	None	No
Gain	Affy6.0_G_179	11 (Asian)	chr12:81711207-81762121	50.9	TMTC2	No
Gain	Affy6.0_G_194	69	chr13:86091484-86118457	27.0	None	No
Gain	Affy6.0_G_212	67	chr14:73405361-73432688	27.3	ZNF410, PTGR2	No
Gain	Affy6.0_G_235	99	chr15:32814039-32848252	34.2	GJD2	No
Gain	Affy6.0_G_226	44	chr15:70381008-70436843	55.8	HEXA, CELF6	No
Gain	Affy500K_G_136	37 (EBV)	chr16:78810438-79258408	448.0	DYNLRB2, CDYL2, MIR548H4	High-level amplification
Gain	Affy500K_G_105, Affy6_G_283 and Affy6_G_284	28	chr17:2919396-3184579	265.2	OR1D2, OR1G1, OR1A2, OR1A1, OR1D4, OR3A2, OR3A1, OR3A4P	No
Gain	Affy500K_G_176 and Affy6_G_301	44	chr18:2254263-2555103	300.8	METTL4	No
Gain	Affy500K_G_280 and Affy6_G_298	62	chr18:6838462-7291170	452.7	ARHGAP28, LAMA1, LRRC30, LOC400643	High-level amplification
Gain	Affy6.0_G_324	20 (admixed)	chr19:60436319-60696243	259.9	BRSKI, UBE2S, SHISA7, TMEM190, COX6B2, FAM71E2, HSPBP1, TMEM150B, ISOC2, IL11, RPL28, TMEM238, ZNF628, SUV420H2, NAT14, PPP6R1, SSC5D	No
Gain	Affy6.0_G_315	62	chr19:60824299-60923809	99.5	ZNF784, NLRP9, EPN1, CCDC106, ZNF580, U2AF2, ZNF581	No
Gain	Affy6.0_G_326	202	chr20:46926869-46943223	16.4	None	No
Gain	Affy500K_G_225	58	chr21:28431800-28667362	235.6	None	No
Gain	Affy500K_G_226	58	chr21:35973166-36013145	40.0	None	No