Resesarch Paper

Copy-Number Variants in Patients with a Strong Family History of Pancreatic Cancer

Robert Lucito¹
Shubha Suresh²
Kimberly Walter³
Akhilesh Pandey²
B. Lakshmi¹
Alex Krasnitz¹
Jonathan Sebat¹
Michael Wigler¹
Alison P. Klein³⁻⁵
Kieran Brune³
Emily Palmisano³
Anirban Maitra²⁻⁴
Michael Goggins^{3,4,6}
Ralph H. Hruban^{3,4,*}

 $^{\rm 1}\text{Cold}$ Spring Harbor Laboratory; Cold Spring Harbor, New York USA

²Institute for Genetic Medicine; ³Departments of Pathology; ⁴Oncology; ⁵Epidemiology; ⁶Medicine; The Sol Goldman Pancreatic Cancer Research Center; The Johns Hopkins Medical Institutions; Baltimore, Maryland USA

*Correspondence to: Ralph H. Hruban, M.D.; Department of Pathology; The Sol Goldman Pancreatic Cancer Research Center; Johns Hopkins Medical Institutions; 401 North Broadway, Weinberg 2242; Baltimore, Maryland 21231 USA; Tel.: 410.955.9132; Fax: 410.955.0115; Email: rhruban@jhmi.edu

Original manuscript submitted: 06/01/07 Revised manuscript submitted: 07/11/07 Manuscript accepted: 07/12/07

Previously published as a *Cancer Biology & Therapy* E-publication: http://www.landesbioscience.com/journals/cbt/article/4725

KEY WORDS

familial cancer, pancreatic cancer, germline deletions, germline amplifications, variants

ACKNOWLEDGEMENTS

Supported by the NIH SPORE (Specialized Programs of Research Excellence) in Gastrointestinal Cancer Grant CA62924, the Michael Rolfe Pancreatic Cancer Foundation, and the Goldman Family.

ABSTRACT

Copy-number variants such as germ-line deletions and amplifications are associated with inherited genetic disorders including familial cancer. The gene or genes responsible for the majority of familial clustering of pancreatic cancer have not been identified. We used representational oligonucleotide microarray analysis (ROMA) to characterize germ-line copy number variants in 60 cancer patients from 57 familial pancreatic cancer kindreds. Fifty-seven of the 60 patients had pancreatic cancer and three had nonpancreatic cancers (breast, ovary, ovary). A familial pancreatic cancer kindred was defined as a kindred in which at least two first-degree relatives have been diagnosed with pancreatic cancer. Copy-number variants identified in 607 individuals without pancreatic cancer were excluded from further analysis. A total of 56 unique genomic regions with copy-number variants not present in controls were identified, including 31 amplifications and 25 deletions. Two deleted regions were observed in two different patients, and one in three patients. The germ-line amplifications had a mean size of 662 Kb, a median size of 379 Kb (range 8.2 Kb to 2.5 Mb) and included 425 known genes. Examples of genes included in the germ-line amplifications include the MAFK, JunD and BIRC6 genes. The germ-line deletions had a mean size of 375Kb, a median size 151 Kb (range 0.4 Kb to 2.3 Mb) and included 81 known genes. In multivariate analysis controlling for region size, deletions were 90% less likely to involve a gene than were duplications (p < 0.01). Examples of genes included in the germ-line deletions include the FHIT. PDZRN3 and ANKRD3 genes. Selected deletions and amplifications were confirmed using real-time PCR, including a germ-line amplification on chromosome 19. These genetic copy-number variants define potential candidate loci for the familial pancreatic cancer gene.

INTRODUCTION

It has been estimated that 10% of pancreatic cancer has a familial basis.¹⁻⁴ Members of kindreds in which one or more family members have been diagnosed with pancreatic cancer have a significantly increased risk of developing pancreatic cancer, and segregation analyses suggest that this aggregation is caused by the autosomal dominant inheritance of a rare allele.⁵⁻¹³ Several known inherited genetic alterations have already been identified that increase the risk of pancreatic cancer.¹¹ These include inherited mutations in the *BRCA2*, *BRCA1*, *STK11/LKB1*, *PRSS1* and *p16/CDKN2A* genes, and possibly in the *PALLD* gene.¹⁴⁻²⁷ In total, however, these known genes account for <20% of the familial clustering of pancreatic cancer. While linkage analyses would be a natural method to identify additional familial pancreatic cancer genes, the extremely high mortality rate of pancreatic cancer makes it very difficult to obtain germ-line DNA samples from multiple affected members of a kindred.^{28,29} We hypothesized that copy-number variants (CNVs) could be used to localize additional familial pancreatic cancer genes. CNVs have the advantage that they can be identified in single patients and CNVs have been associated with other forms of familial cancer.

CNVs in the human genome have recently been characterized using a variety of assays including representational oligonucleotide microarrays (ROMA), bacterial artificial chromosomes (BAC)-based array comparative genomic hybridization (CGH), single nucleotide polymorphism (SNP) genotyping and paired-end sequences.³⁰⁻³⁶ For example, three recent studies of CNVs using SNP genotyping data from the HapMap Consortium identified over 1000 deletions within the human genome.³⁵⁻³⁷ Most of these deletions involved known genes, and most of the deletions ranged in size from 500 bp to 10.5 kb, but some reported germ-line deletions have been as large as 3 Mb.^{30,35-37} In one of

these analyses, Conrad et al. identified 92 genes that were entirely deleted in at least one of the 60 parent-offspring trios included in their SNP genotyping analyses.³⁵ While these CNVs help define the full spectrum of human variation, they also provide a unique opportunity to identify disease causing genetic loci. For example, germ-line deletions were used by M. Fergusen-Smith to localize the red cell acid phosphatase gene to the short arm of chromosome 2, and by M. Mikkelson et al. to localize the Gc-locus to the long arm of chromosome $4.^{38,39}$

Although most familial syndromes appear to be caused by small intragenic mutations in cancer-associated genes, some familial syndromes are caused by larger germ-line deletions or duplications of these same genes. 40-50 Large genomic deletions that inactivate the BRCA2 or BRCA1 genes have been identified in breast cancer families, 10% of APC gene mutation-negative (by the protein truncation test) adenomatous polyposis families have a constitutional deletion of the entire APC gene, approximately 18% of patients with hereditary nonpolyposis colorectal cancer syndrome have large genomic deletions encompassing MLH1 or MSH2, and large germ-line deletions of mitochondrial complex II subunits SDHB and SDHD have been demonstrated in hereditary paraganglioma. 40-43,47 These findings suggest that the characterization of copy-number variants in cohorts of patients with familial cancer could help localize novel familial cancer genes. We therefore characterized the CNVs in a well-defined cohort of 60 cancer patients from familial pancreatic cancer kindreds using ROMA.

METHODS

This study was reviewed and approved by the institutional review board of The Johns Hopkins Medical Institutions (JHMI). Informed consent was obtained from study participants.

The National Familial Pancreas Tumor Registry (NFPTR) was founded at Johns Hopkins University in 1994. The NFPTR is one of the largest registries of familial pancreatic cancer in the world. Patients are recruited into this registry from two sources: (1) patients treated for pancreatic cancer at The Johns Hopkins Hospital are invited (either by an in-person visit or by mail) to participate, and (2) individuals with a personal or family history of pancreatic cancer can refer themselves through the Internet (pathology.jhu. edu/pancreas) or they can be referred by a non-Hopkins health care provider. As of May 1, 2007, 2,339 families had enrolled in the NFPTR. These include 1,489 sporadic kindreds and 850 familial kindreds. Sporadic kindreds are defined as kindreds with at least one pancreatic cancer, but without a pair of first-degree relatives with pancreatic cancer. Familial kindreds are defined as kindreds in which a pair of first-degree relatives has been diagnosed with pancreatic cancer. A total of 191 familial kindreds in which a germ-line blood sample was available from at least one family member were eligible for inclusion in this study. Fifty-seven of these 191 familial kindreds were selected for inclusion in this study based on the availability of DNA samples. An individual with familial pancreatic cancer was included from 56 of the kindreds. A second patient with pancreatic cancer was included from one kindred so that we could test for transmission of the CNVs identified. Three individuals with nonpancreatic cancers (breast, ovary, ovary) who were genetically related (a mother and two were siblings) to a pancreatic cancer patient were also included. A total of 60 cases were therefore analyzed, and almost all were Caucasian. Epstein-Barr virus (EBV) transformed lymphoblastoid cells were used for the ROMA studies of the cases, and nontransformed lymphocytes were used in the confirmations.

In addition to the 60 cases, ROMA data from a series of control individuals was used to identify common CNVs in the population. All individuals with known pancreatic cancer or predisposition to any known cancer were removed from the set leaving 372 individuals. Of these 372 individuals, 75% (282) were EBV transformed, allowing us to identify and control for possible abberations introduced by EBV transformation. Of these 372 individuals, one third were HapMap individuals who have varied origin. We also used normal genetic variation data from two publicly available datasets; one includes 270 (non overlapping CEPH) individuals from four populations with ancestry in Europe, Africa or Asia, and the second utilizes 55 unrelated individuals. 32,51 When overlapping individuals were eliminated, data from a total of 607 controls were included in the analyses. The inclusion of the CEPH individuals creates large diversity in the normal filter set which will be useful in identifying CNVs from a diverse population.

Supplies for ROMA. Enzymes BgIII and T4 DNA ligase were supplied by New England Biolabs. Primers were supplied by Sigma Genosys. Cot1 DNA and tRNA were supplied by Invitrogen. The Megaprime labeling kit, Cy3-conjugated dCTP and Cy5-conjugated dCTP were supplied by Amersham-Bioscience. Taq polymerase [Eppendorf mastermix (2.5X)] was supplied by Eppendorf. Microcon YM-30 filters were supplied by Amicon, and formamide was supplied by Amresco. Phenol:chloroform was supplied by Sigma. NimbleGen photoprint arrays were synthesized by NimbleGen Systems Inc. Design of the ROMA array has been described previously (Ref. 52).

ROMA. Arrays were described previously (Ref. 52). In brief the arrays are based on representational techniques and thus all oligonucleotides map to BgIII fragments that are within the representations size range of 200-1000 bp. The array was designed to the June 2002 build of the genome (NCBI Build 30) and the coordinates have been updated to the May 2004 build (NCBI Build 35). There are roughly 84,000 features on the array and these are scattered across the genome resulting in an average resolution of 30 kb. 500 ng of genomic DNA was used to prepare BglII representations (described previously in ref. 52). Minor changes in the protocol are as follows: as few as four 250-µL tubes were used for each sample for amplification of the representation each with a 100 µl volume reaction. An equal number of tubes were used for reference for each sample that was coamplified. The cycle conditions were 95°C for 1 min, 72°C for 3 min, for 20 cycles, followed by a 10-min extension at 72°C. The contents of the tubes were pooled when completed. Representations were cleaned by phenol:chloroform extraction, precipitated, resuspended and the concentration determined. DNA was labeled as described with minor changes.⁵² Briefly, 2 µg of DNA template was placed (dissolved in TE at pH 8) in a 0.2-mL PCR tube. 5 μL of primers from the Amersham-Pharmacia Megaprime labeling kit were added and pipetted up and down several times. The volume was brought up to 50 µL with dH2O, and mixed. The tubes were placed in Tetrad at 100°C for 5 min, then on ice for 5 min. To this was added 10 µL of labeling buffer from the Amersham-Pharmacia Megaprime labeling Kit, 5 µL of label (Cy3-dCTP or Cy5-dCTP), and 1 µL of NEB Klenow fragment. The tubes were placed in a Tetrad and incubate at 37°C for 2 h. The labeled samples (Cy3 and Cy5) were combined into one Eppendorf tube, to which 50 µL of

1 μg/μL human Cot 1 DNA, 10 μL of 10 mg/mL yeast tRNA and 200 µL of Low TE (3 mM Tris at pH 7.4, 0.2 mM EDTA) were added. This was then loaded into a Microcon Filter and centrifuged for 10 min at 12,600 relative centrifugal force (rcf). The flowthrough was discarded and the sample then washed with 450 μ L of Low TE. The sample was centrifuged at 12,600 rcf and the procedure repeated twice. The labeled sample was collected by inverting the Microcon column into a new tube and centrifuging for two minutes at 12,600 rcf. The labeled sample was transferred to a 200-µL PCR tube and the volume adjusted to 5 µL of Low TE. Slides were prepared as in Lucito et al.⁵² with the following changes. The slides were washed briefly in milliQ H2O and dried with compressed nitrogen. The hybridization solution consisted of 50% formamide, 5X SSC and 0.1%SDS. For each, 30 μl of hybridization solution was added to the 5 µl of labeled sample and mixed. Samples were denatured and hybridized as described earlier (ref. 52). For each sample two hybridizations were performed in color reversal and the ratio from both experiments averaged to minimize dye specific variance. CNVs which appeared in only one experiment were ignored.

Filtering out CNVs seen in normals. The dataset if based on all probes would require a large data frame to represent all samples and would drastically slow down computing time. Much of the data from the genome represents nonevents or vast regions without CNVs. Therefore, the probe-based data was reduced to event-based data. In other words, only the CNVs, location and ratio, are represented in a file, thereby drastically reducing file size without decreasing the information content. This transformation was performed for both the familial pancreatic cancer and the normal filter sample datasets. The familial set was compared to the normal set to remove common CNVs. If a region altered in the familial set was found altered in the normal set, this region was removed from the familial set. In some instances the CNVs found in the familial set were larger than the CNV found in the normal individuals. In these cases the nonoverlapping ends of the CNV were retained in final data set., If the non overlapping CNV was two probes or less the region was removed. We have allowed for some flexibility in the removal of CNVs from the familial set if found in the ROMA normal dataset of 372 individuals which is available as supplemental data. This flexibility allows the retention of CNVs that are found in below algorithmically defined percentages of individuals in the 372 normal filtration set. The data was processed three additional times such that a CNV identified in the familial samples was retained in the data set if it was found in less than 30%, less than 20%, less than 10% or not found in any of the 372 normal individuals (the filtered dataset as well as the unfiltered full dataset are available at lucitolab.cshl.edu/rl_data.html). Python scripts were written to parse out information pertaining to the variations present only in familial pancreatic cancer patients and not in normal samples. There have been a number of papers which have identified CNVs in the normal population. We also used data available from two such publications to supplement our normal variation data. 32,51 The CNVs identified in these studies were compared to the familial set of CNVs identified in the present study. If a CNV in the familial dataset overlapped a CNV reported in one of the control datasets, the CNV was removed from the familial dataset. For the remaining CNVs, regions were defined by the positions of the most upstream and downstream probes. The gene lists were then generated using the UCSC genome browser for the May 2004 Build. Genes that had any exons located within the boundaries specified by the most upstream and downstream probes were considered as candidates and included in the catalog.

Confirmation of CNVs. Six CNV regions were selected for QPCR validation which contained or were relatively close to known oncogenes or tumor suppressors. In addition the regions were found on different chromosomes in different individuals. Two Tagman probe sets were designed for each region using Primer Express version 2.0. The nonfluorescent labeled oligonucleotides were tested on genomic DNA known to have two copies of the regions in question by standard PCR amplification, followed by testing on the same DNA with TaqMan assay and titrating DNA concentrations. TagMan probes used in this study had to be sensitive enough to detect a two-fold difference in either direction which was tested by the above titration studies. After TagMan probeset characterization only three TaqMan pobesets were sensitive enough for further QPCR analysis; (Chromosome 2: 32,537,367-33,153,333, Chromosome 9: 19,426,668-20,442,925 and Chromosome 19: 43,390,659-43,832,483). The TaqMan assay was performed in an ABI Prism 7900HT Sequence Detection System as follows: a master mix was made up of 10X Real-time PCR buffer, 50 mM MgCl2, 5 mM dNTP mix w/ dUTP, Hot Goldstar Taq, Uracil N-Glycosylase (Eurogentec) and 10 mM each of Forward primer, Reverse primer and Taqman probe (Sigma-Genosys). Reactions were performed in a 384-well clear optical reaction plate and each plate was covered with an optical adhesive cover (ABI Prism). The reaction conditions were as follows: 48°C for 2 min, 95°C for 10 min, 95°C for 15 s, 60°C for 1 min, 72°C for 30 s for 40 cycles. Similar analysis was performed for a large region of chr12p, but in this case the data was averaged over four probe sets evenly spaced across the proposed deletion. All other conditions were the same.

After testing the TaqMan probe sets the test DNAs were assayed as follows. The initial DNA assayed was from EBV immortalized lymphoblasts. Four DNAs were tested with the TaqMan probes, where only one was expected to contain a CNP for one of the regions. In addition the same control DNA mentioned above was also used as a template. Three dilutions of the DNA were made so that input template would be 10 ng, 4 ng and 0.4 ng. This increased the number of data points but also allowed the calculation of amplification efficiency. Each dilution was performed in triplicate and each run on the ABI Prism 7900HT Sequence Detection System was performed three times. Overall this yielded 18 data points to average for copy number as well as the two probe sets for each region.

Raw QPCR data was analyzed by SDS 2.1 software (ABI Prism) to give cycle threshold (Ct) values for each reaction. Ct values for each concentration of each DNA were averaged and compared to values for control genomic DNA using the equation:

 $^{\Delta\Delta}$ Ct = (Ct, target region - Ct, control region) χ - (Ct, target region - Ct, control region) γ

where χ = the unknown sample (cell line genomic DNA) and γ = the known sample (normal genomic DNA).⁵³

The same QPCR assays were then repeated for nonimmortalized DNA for the regions that passed the first experiments (Chromosome 2:32,537,367–33,153,333, Chromosome 9:19,426,668–20,442,925 and Chromosome 19: 43,390,659–43,832,483). The results for these analyses were evaluated using the equation above.

Statistics. Summary statistics and logistic regression to compare deletions and amplifications was performed using STATA 9 software.

(StataCorp. 2005. Stata Statistical Software: Release 9.0. College Station TSC).

RESULTS

Patients. The 60 cancer patients included in these analyses are summarized in Table 1. Briefly, their mean age was 66.75 years and 35 (61%) were male. Ten patients, including a sibling pair, came from families in which more than four members had been diagnosed with pancreatic cancer. In addition to the 57 patients with pancreatic cancer, three patients had extra-pancreatic malignancies (breast, ovary, ovary).

CNVs found only in cases. Fifty-six of the copy-number variants (CNVs) identified in the patients from the familial pancreatic cancer kindreds were not identified in any of the 607 normal subjects (data for these CNVs, the entire unfiltered dataset as well as data obtained using more flexible constraints for filtering out common CNVs [see methods section] are available as supplemental data from the web lucitolab.cshl. edu/rl_data.html) or the normal CNV data obtained from previous studies.^{32,51} These included 31 amplifications and 25 deletions. Additional data on these loci, including the genes within each CNV, are presented on the web (pathology2.jhu.edu/pancreas/roma.cfm) and as Supplemental Material.

The 31 germ-line amplifications had a mean size of 662 Kb, a median size of 379 Kb (range 8.2 Kb to 2.5 Mb). Twenty-nine (92.3%) of the 31 amplifications involved known genes. A total of 425 genes were found within these amplifications (several examples of possible oncogenes given in Table 3, a complete list on the web: pathology2.jhu.edu/pancreas/roma.cfm) and as Supplemental Material. Some of the larger amplifications presumably represent amplifications that occurred in the process of transforming the lymphoblastoid cell lines. For example, some of the very large CNVs identified are likely not compatible with a normal phenotype and intellectual capacity. Indeed, we observed a large amplification of 12p in a single transformed sample and this amplification could not be confirmed by quantitative PCR analysis of a nontransformed germ-line sample from the same patient (data not shown).

The 25 deletions had a mean size of 375 Kb, a median size 151 Kb (range 0.4 Kb to 2.3 Mb) and included 81 known genes. Several examples of tumor suppressors given in Table 4, a complete list on the web: pathology2.jhu.edu/pancreas/roma.cfm and as Supplemental Material. Sixteen (64 %) of the 25 deletions involved genes. As was true for the amplifications, some of the larger deletions presumably represent deletions that occurred in the process of transforming the lymphoblastoid cell lines.

Transmission of the CNVs. Samples from multiple members of two kindreds were analyzed. Two members of one kindred and three members of a second kindred were included in the analyses. While the CNVs identified demonstrated familial transmission, the CNVs identified were found in the normal individuals as common CNVs in the general population. Thus, in this sampling of multiple family members we were not able to identify the transmission of a pancreatic cancer-specific CNV.

Known genes. In addition to the global approach taken to characterize the CNVs, we also looked at genetic loci known to harbor genes implicated in the development of sporadic and familial

Table 1 Characteristics of the 60 patients

Number of Pancreatic Cancer Patients in Family	Number of Patients with Pancreatic Cancer Evaluated	Age at Diagnoses of Cancer in Years (Standard Deviation)	Gender of Patient Evaluated
2 3 Other	16 pancreatic	67.9 (12.6)) for pancreatic 52 (10) for other	10 Male 9 Female
3	17	65.2 (10.5)	9 Male 8 Female
4	14	67.8 (9.8)	10 Male 4 Female
More than 4	10	66.1 (11.1)	6 Male 4 Female
Total	60	66.75 (10.9)	35 Male 25 Female

pancreatic cancer. These included: *BRCA1*, *BRCA2*, *FHIT*, *STK11*, *p16/CDKN2A*, *TP53*, *SMAD4*, *KRAS* and the recently described familial pancreatic cancer gene locus on chromosome 4q.²⁶ The only change found in these genes was a deletion in the *FHIT* gene, but it was in a noncoding region.

Confirmation of selected CNVs. Six CNV regions detected by ROMA which contained or were relatively close to known oncogenes, tumor suppressors or genes of interest were selected to validate that the CNVs were legitimate alterations in copy number by QPCR. After the QPCR probes were selected, three regions were evaluated. These included the region on chromosome 2 containing copy number gain of the anti apoptotic gene BIRC6,54 the region on chromosome 9, 1.3 Mb away from the INK/ARF locus, and the region on chromosome 19, 475 kb away from the PAK4 gene, 55 a gene which can regulate anchorage independent growth when activated. For each region two sets of probes were designed, to ensure multiple probes within a region were in agreement. DNAs with identified lesions were assayed by QPCR and compared to DNAs which did not have corresponding lesions identified. These assays were first performed with DNA from EBV immortalized lymphoblasts since more material was available. The regions that scored positive were then assayed on nonimmortalized DNA from the same patients. A total of 18 data points were averaged to determine the copy number in reference to two separate samples that did not have a CNP and the results are shown in Table 2. We can see that the QPCR assay accurately identifies that one CNP as having four copies of DNA as compared to two copies in the reference sample.

Gene content of deletions and duplications. In univariate analysis longer deletions/amplification were more likely to involve genes than shorter changes (p = 0.02), and deletions were less likely to involve genes than were amplifications (p < 0.01). In multivariate analysis controlling for region size, deletions were 90% less likely to involve a gene than were duplications (p < 0.01).

DISCUSSION

The genetic basis for the clustering of pancreatic cancer is poorly understood.¹ A handful of genes have been identified, but they account for only a minority of familial pancreatic cancer kindreds.

Table 2 PCR confirmation of selected copy number variants

CNV Coordinates	Size of CNV (bp)	Type of CNV	QPCR Transformed	QPCR non Transformed	Copy Number
Chromosome 2:					
32,537,367- 33,153,333	615,966	Amplification	-1.2	-1.1	4
Chromosome 9:					
19,426,668- 20,442,925	1,016,257	Deletion	1.2	0.6	1
Chromosome 19:					
43,390,659- 43,832,483	441,824	Amplification	-0.6	-0.5	3

Table 2 shows the results of quantitative PCR (QPCR) for several CNVs identified in the familial samples. The first column defines the chromosome coordinates of the CNVs identified. The second column defines whether the CNV is deleted or duplicated. The third column (QPCR transformed) represents the difference in QPCR ct results for 18 replicates using EBV transformed DNA as a template. The fourth column (QPCR non transformed) represents the difference in QPCR ct results for 18 replicates using non transformed DNA as a template. The fifth column (Copy Number) is an estimate based on normal having two copies of each allele. A negative signifies more copies of the DNA since PCR amplification in the test sample is faster. A positive value signifies less DNA in the test sample since amplification would take longer. A difference close to 0.5 would refer to half of the copies in normal which would be one copy. A value close to 1.0 would signify the same number as normal or two copies. A value of two would indicate two extra copies or four copies total.

Identifying the gene or genes responsible for the majority of familial pancreatic cancer would help guide genetic counseling⁵⁶ and early detection efforts,⁵⁷⁻⁵⁹ and it may form the basis for the

development of future gene-specific therapies. 60,61 Unfortunately, the extremely high mortality rate of pancreatic cancer renders standard approaches for identifying familial cancer genes, such as linkage analysis, extremely difficult. 28,29 Novel approaches to identify the gene or genes responsible for the familial clustering of pancreatic cancer are therefore needed. Ideally, these approaches would be applicable even when the only biospecimen available is a DNA sample from a single affected family member.

A growing body of evidence now makes it clear that copy-number variants (CNVs) are relatively common in the human genome.³⁰ CNVs can predispose to disease through a number of mechanisms.³⁰ They can directly affect gene dosage, or may function in combination with other genetic and environmental factors.³⁰ In specific, germ-line deletions have been reported to cause a number of familial cancer syndromes.⁴⁰⁻⁴⁸

These observations suggest that screening for CNVs should be an efficient way to identify loci that harbor the familial pancreatic cancer gene. We therefore applied ROMA to a well-characterized

Table 3 Selected germ-line duplication
--

Gene	Location	Name	Function
fancd2	chr3:9904280-10967287	Fanconi anemia	In response to DNA damage, localizes to nuclear foci with other proteins (BRCA1 AND BRCA2) involved in homology-directed DNA repair.
mafk	chr7:271085-2158745	v-maf musculoaponeurotic fibrosarcoma oncogene homolog K	transcription factor
rnd 1	chr12:47458533-47775626	Rho family GTPase 1	Member of the Rho GTPase family, possibly, regulates the organization of the actin cytoskeleton in response to extracellular growth factors
wnt10b	chr12:47458533-47775626	wingless-type MMTV integration site family, member 10B	These secreted proteins have been implicated in oncogenesis and in several developmental processes
wnt1	chr12:47458533-47775626	wingless-type MMTV integration site family, member 1	These secreted proteins have been implicated in oncogenesis and in several developmental processes
map2k2	chr19:2984601-5201290	mitogen-activated protein kinase kinase 2	The protein encoded by this gene is a dual specificity protein kinase that belongs to the MAP kinase kinase family. This kinase is known to play a critical role in mitogen growth factor signal transduction. It phosphorylates and thus activates MAPK1/ERK2 and MAPK2/ERK3.
jund	chr19:17785312-19480018	jun D proto-oncogene	A member of the JUN family, and a functional component of the AP1 transcription factor complex. It has been proposed to protect cells from p53-dependent senescence and apoptosis.

Table 4 Selected germ-line deletions

Gene	Location	Name	Function
FHIT	chr3:60219748-60263116	fragile histidine triad gene	Member of the histidine triad gene family involved in purine metabolism. Aberrant transcripts from this gene have been found in about half of all esophageal, stomach, and colon carcinomas.
PDZRN3	chr3:73432926-74029485	PDZ domain containing RING finger 3	This protein is found to be downregulated in gliomas
ANKRD3	chr21:39946855-42268598	ankyrin repeat domain 3	A serine/threonine protein kinase that interacts with protein kinase C-delta. The encoded protein can also activate NFκB.

For a complete listing of all deletions identified: http://pathology2.jhu.edu/pancreas/roma.cfm and as Supplemental Material.

series of 60 cancer patients from 57 familial pancreatic cancer kindreds. Copy-number variants identified in 607 individuals without pancreatic cancer were excluded.

ROMA is an array-based CGH technology in which the complexity of the genetic material analyzed is reduced through a representative sampling of genetic fragments within a certain size-range.^{34,52} The reduced complexity obtained through the use of a representation reduces background noise.^{34,52} The ROMA arrays utilized in this study have >80,000 probes and an average resolution of 30 kb, and a number of phenotypically normal individuals have been analyzed using this technology providing a rich database of CNVs in the general population.^{34,52} ROMA has been used to identify CNVs in cancer as well as normal populations accurately.⁵² The only limitation of ROMA presently is the resolution of the array used for detection and in the future we will be using arrays with up to 4 times the resolution of the current array.

In our analysis a series of 607 normal controls (372 from our own data and an additional 235 nonoverlapping normal controls from available public data) was used to exclude any CNVs prevalent in the general population.^{32,34,51} Of the 607 normals, 552 were CEPH members of varied heritage, creating a useful normal filter set for the varied heterogeneity that is often found in residents of America. In addition 372 of the control samples were from EBV transformed cells. If the normal filter set were separated into EBV and non EBV transformed samples, and then analyzed for deletions, the EBV transformed controls filtered out twice as many CNVs than did the non transformed counterparts. It appears that there are regions that are commonly altered by the immortalization. Thus, having the EBV transformed samples in the control filter set was useful in filtering out EBV transformation specific CNVs. Interestingly, the average number of copy gain CNVs were equal in the non transformed and EBV transformed samples in both the filter set and in the test set.

Thirty-one distinct loci of amplification were identified as were 25 loci harboring deletions. The majority of these involved known genes (pathology2.jhu.edu/pancreas/roma.cfm and as Supplemental Material). Of interest, in multivariate analyses, after controlling for region size, deletions were 90% less likely to involve a known gene than were amplifications (p < 0.01), suggesting that germ-line deletions of a gene may be more deleterious than amplifications. Examples of a deletion and amplification were confirmed by quantitative PCR.

Functionally interesting candidate genes were selected from the CNV specific genes identified. Duplications are listed in Table 3.

This list contains several genes whose duplication in the germ-line might be expected to contribute to the development of neoplasia. For example, a germ-line duplication of the *JUND* proto-oncogene on chromosome 19 was identified. JunD protects cells from p53-dependent senescence and apoptosis and JunD may protect cells from oxidative stress. 62,63

Selected deletions are listed in Table 4. These deletions include, for example, the *ANKRD3* gene on chromosome 21. The *ANKRD3* gene codes for a protein involved in signal transduction through NFkB, it has been shown to be somatically deleted in sporadic pancreatic cancer, ⁵⁸ its expression is suppressed in prostate, liver, breast, head and neck squamous cell carcinoma, and melanoma. ^{54,55}

There are several limitations to our approach of using CNVs to localize the gene(s) responsible for the familial aggregation of pancreatic cancer. First, we used transformed samples as a source for the germ-line DNA, and the transformation process may introduce occasional genetic changes. We have used a large number of samples in the normal filtration set also transformed by EBV and thus more common alterations introduced in this process should have been removed. However, some of the very large CNVs identified are likely not compatible with a normal phenotype and intellectual capacity. Indeed, we observed a large multiple Mb amplification of 12p in a single transformed sample and this amplification could not be confirmed by quantitative PCR analysis of a nontransformed germ-line sample from the same patient (data not shown). Clearly, CNVs such as these should be confirmed in a nontransformed sample.

Second, CNVs are much more common than was originally thought.^{30,35-37} To date, over 1,000 CNVs have been described (refs. 30 and 35–37). A CNV involving a gene in an individual does not prove that the CNV is the cause of the individual's phenotype.³⁰ CNVs can only help localize candidate genes that then have to be confirmed, usually by sequencing additional affected and nonaffected individuals to determine if intragenic mutations in the gene cosegregate with the disease, or by larger population-based studies.

In summary, we applied ROMA to a series of well-characterized patients with familial pancreatic cancer. A number of copy-number variants were identified that were not seen in a series of 607 controls. Many of these copy-number variants involved genes, and these genes can be prioritized in future attempts to localize the familial pancreatic cancer gene.

References

- Klein AP, Hruban RH, Brune KA, Petersen GM, Goggins M. Familial pancreatic cancer. Cancer J 2001; 7:266-73.
- 2. Lynch HT. Genetics and pancreatic cancer. Arch Surg 1994; 129:266-8.
- Lynch HT, Smyrk T, Kern SE, Hruban RH, Lightdale CJ, Lemon SJ, Lynch JF, Fusaro LR, Fusaro RM, Ghadirian P. Familial pancreatic cancer: A review. Semin Oncol 1996; 23:251-75.
- Lynch HT, Fusaro LR, Lynch JF. Familial pancreatic cancer: A family study. Pancreas 1992; 7:511-5.
- Falk RT, Pickle LW, Fontham ET, Correa P, Fraumeni JF. Life-style risk factors for pancreatic cancer in Louisana: A case-control study. Am J Epidemiol 1988; 128:324-36.
- Fernandez E, La Vecchia C, D'Avanzo B, Negri E, Franceschi S. Family history and the risk of liver, gallbladder, and pancreatic cancer. Cancer Epidemiol Biomarkers Prev 1994; 3:209-12.
- Ghadirian P, Boyle P, Simard A, Baillargeon J, Maisonneuve P, Perret C. Reported family aggregation of pancreatic cancer within a population-based case-control study in the Francophone community in Montreal, Canada. Int J Pancreatol 1991; 10:183-96.
- Ghadirian P, Liu G, Gallinger S, Schmocker B, Paradis AJ, Lal G, Brunet JS, Foulkes WD, Narod SA. Risk of pancreatic cancer among individuals with a family history of cancer of the pancreas. Int J Cancer 2002; 97:807-10.
- Schenk M, Schwartz AG, O'Neal E, Kinnard M, Greenson JK, Fryzek JP, Ying GS, Garabrant DH. Familial risk of pancreatic cancer. J Natl Cancer Inst 2001; 93:640-4.
- Silverman DT, Schiffman M, Everhart J, Goldstein A, Lillemoe KD, Swanson GM, Schwartz AG, Brown LM, Greenberg RS, Schoenberg JB, Pottern LM, Hoover RN, Fraumeni JF. Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. Br J Cancer 1999; 80:1830-7.
- Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJ, Griffin C, Cameron JL, Yeo CJ, Kern SE, Hruban RH. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res 2004; 64:2634-8.
- Hemminki K, Li X. Familial and second primary pancreatic cancers: A nationwide epidemiologic study from Sweden. Int J Cancer 2003; 103:525-30.
- Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, Sulem P, Kristjansson K, Arnason S, Gulcher JR, Bjornsson J, Kong A, Thorsteinsdottir U, Stefansson K. Cancer as a complex phenotype: Pattern of cancer distribution within and beyond the nuclear family. PLoS Med 2004; 1:e65.
- Berman DB, Costalas J, Schultz DC, Grana G, Daly M, Godwin AK. A common mutation in *BRCA2* that predisposes to a variety of cancers is found in both Jewish Ashkenazi and non-Jewish individuals. Cancer Res 1996; 56:3409-14.
- Thorlacius S, Olafsdottir GH, Tryggvadottir L, Neuhausen S, Jonasson JG, Tavitigian SV, Tulinius H, Ogmundsdottir HM, Eyfjord JE. A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. Nat Genet 1996; 13:117-9.
- White K, Held KR, Weber BHF. A BRCA2 germ-line mutation in familial pancreatic carcinoma. Int J Cancer 2001; 91:742-4.
- Murphy KM, Brune KA, Griffin CA, Sollenberger JE, Petersen GM, Bansal RK, Hruban RH, Kern SE. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: Deleterious BRCA2 mutations in 17%. Cancer Res 2002; 62:3789-93.
- Parker JF, Florell SR, Alexander A, DiSario JA, Shami PJ, Leachman SA. Pancreatic carcinoma surveillance in patients with familial melanoma. Arch Dermatol 2003; 139:1019-25.
- Rulyak SJ, Brentnall TA, Lynch HT, Austin MA. Characterization of the neoplastic phenotype in the familial atypical multiple-mole melanoma-pancreatic carcinoma syndrome. Cancer 2003; 98:798-804.
- Bowlby LS. Pancreatic adenocarcinoma in an adolescent male with Peutz-Jeghers syndrome. Hum Pathol 1986; 17:97-99.
- Giardiello FM, Welsh SB, Hamilton SR, Offerhaus GJ, Gittelsohn AM, Booker SV, Krush AJ, Yardley JH, Luk GD. Increased risk of cancer in the Peutz-Jeghers syndrome. N Engl J Med 1987; 316:1511-4.
- Jeghers H, McKusick VA, Katz KH. Generalized intestinal polyposis and melanin spots of the oral mucosa, lips and digits. N Engl J Med 1949; 241:992-1005.
- 23. Su GH, Hruban RH, Bova GS, Goggins M, Bansal RK, Tang DT, Shekher MC, Westerman AM, Entius MM, Yeo CJ, Kern SE. Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers. Am J Pathol 1999; 154:1835-40.
- Lowenfels AB, Maisonneuve EP, Dimagno YE, Gates LK, Perrault J, Whitcomb DC, International Hereditary Pancreatitis Study Group. International Hereditary Pancreatitis Study Group Hereditary pancreatitis and the risk of pancreatic cancer. J Natl Cancer Inst 1997; 89:442-6.
- Thompson D, Easton DF. Cancer Incidence in BRCA1 mutation carriers. J Natl Cancer Inst 2002; 94:1358-65.
- Eberle MA, Pfutzer R, Pogue-Geile KL, Bronner MP, Crispin DA, Kimmey MB, Duerr RH, Kruglyak L, Whitcomb DC, Brentnall TA. A new susceptibility locus for autosomal dominant pancreatic cancer maps to chromosome 4q32-34. Am J Med Genet 2002; 70.
- Pogue-Geile KL, Chen R, Bronner MP, Crnogorac-Jurcevic T, Moyes KW, Dowen S, Otey C, Crispin DA, Whitcomb DC, Brentnall TA. Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism. PLoS Medicine 2006; 3.
- Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ. Cancer statistics, 2005. CA Cancer J Clin 2005; 55:10-30.

1598

- Petersen GM, de Andrade M, Goggins M, Klein AP, Korczak J, Gallinger S, Lynch HT, Syngal S, Rabe K, Seminara D, Hruban RH. Pancreatic Cancer Genetic Epidemiology (PACGENE) Consortium. Cancer Epidemiol Biomarkers Prev 2006; 15:704-10.
- Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006; 7:85-97.
- Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 2005; 77:78-88.
- 32. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet 2004; 36:949-51.
- 33. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. Nat Genet 2005; 37:727-32.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science 2004: 305:525-8.
- Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 2006; 38:75-81.
- Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 2006; 38:82-5.
- 37. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM. Common deletion polymorphisms in the human genome. Nat Genet 2006; 38:86-92.
- Ferguson-Smith MA, Newman BF, Ellis PM, Thomson DM, Riley ID. Assignment by deletion of human red cell acid phosphatase gene locus to the short arm of chromosome 2. Nat New Biol 1973; 243:271-4.
- Mikkelsen M, Jacobsen P, Henningsen K. Possible localization of Gc-System on chromosome 4. Loss of long arm 4 material associated with father-child incompatibility within the Gc-System. Hum Hered 1977; 27:105-7.
- Agata S, Dalla PM, Callegaro M, Scaini MC, Menin C, Ghiotto C, Nicoletto O, Zavagno G, Chieco-Bianchi L, D'Andrea E, Montagna M. Large genomic deletions inactivate the BRCA2 gene in breast cancer families. J Med Genet 2005; 42:e64.
- Renkonen ET, Nieminen P, Abdel-Rahman WM, Moisio AL, Jarvela I, Arte S, Jarvinen HJ, Peltomaki P. Adenomatous polyposis families that screen APC mutation-negative by conventional methods are genetically heterogeneous. J Clin Oncol 2005; 23:5651-9.
- Casey G, Lindor NM, Papadopoulos N, Thibodeau SN, Moskow J, Steelman S, Buzin CH, Sommer SS, Collins CE, Butz M, Aronson M, Gallinger S, Barker MA, Young JP, Jass JR, Hopper JL, Diep A, Bapat B, Salem M, Seminara D, Haile R. Conversion analysis for mutation detection in MLH1 and MSH2 in patients with colorectal cancer. JAMA 2005; 293;799-809.
- 43. McWhinney SR, Pilarski RT, Forrester SR, Schneider MC, Sarquis MM, Dias EP, Eng C. Large germline deletions of mitochondrial complex II subunits SDHB and SDHD in hereditary paraganglioma. J Clin Endocrinol Metab 2004; 89:5694-9.
- 44. Kikuchi M, Ohkura N, Yamaguchi K, Obara T, Tsukada T. Gene dose mapping delineated boundaries of a large germline deletion responsible for multiple endocrine neoplasia type 1. Cancer Lett 2004; 208:81-8.
- Le Meur N, Martin C, Saugier-Veber P, Joly G, Lemoine F, Moirot H, Rossi A, Bachy B, Cabot A, Joly P, Frebourg T. Complete germline deletion of the STK11 gene in a family with Peutz-Jeghers syndrome. Eur J Hum Genet 2004; 12:415-8.
- 46. Maranchie JK, Afonso A, Albert PS, Kalyandrug S, Phillips JL, Zhou S, Peterson J, Ghadimi BM, Hurley K, Riss J, Vasselli JR, Ried T, Zbar B, Choyke P, Walther MM, Klausner RD, Linehan WM. Solid renal tumor severity in von Hippel Lindau disease is related to germline deletion length and location. Hum Mutat 2004; 23:40-6.
- 47. Hogervorst FB, Nederlof PM, Gille JJ, McElgunn CJ, Grippeling M, Pruntel R, Regnerus R, van Welsem T, van Spaendonk R, Menko FH, Kluijt I, Dommering C, Verhoef S, Schouten JP, van't Veer LJ, Pals G. Large genomic deletions and duplications in the *BRCA1* gene identified by a novel quantitative method. Cancer Res 2003; 63:1449-53.
- 48. Sieber OM, Lamlum H, Crabtree MD, Rowan AJ, Barclay E, Lipton L, Hodgson S, Thomas HJ, Neale K, Phillips RK, Farrington SM, Dunlop MG, Mueller HJ, Bisgaard ML, Bulow S, Fidalgo P, Albuquerque C, Scarano MI, Bodmer W, Tomlinson IP, Heinimann K. Whole-gene APC deletions cause classical familial adenomatous polyposis, but not attenuated polyposis or "multiple" colorectal adenomas. Proc Natl Acad Sci USA 2002; 99:2954-8.
- Le Marechal C, Masson E, Chen JM, Morel F, Ruszniewski P, Levy P, Ferec C. Hereditary pancreatitis caused by triplication of the trypsinogen locus. Nat Genet 2006; 38:1372-4.
- 50. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P, Radlwimmer B, Stange EF. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet 2006; 79:439-48.
- 51. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. Nature 2006; 444:444-54.

- 52. Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. Genome Res 2003; 13:2291-2305.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001; 25:402-8.
- Ren J, Shi M, Liu R, Yang QH, Johnson T, Skarnes WC, Du C. The Birc6 (Bruce) gene regulates p53 and the mitochondrial pathway of apoptosis and is essential for mouse embryonic development. Proc Natl Acad Sci USA 2005; 102:565-70.
- Qu J, Cammarano MS, Shi Q, Ha KC, de Lanerolle P, Minden A. Activated PAK4 regulates cell adhesion and anchorage-independent growth. Mol Cell Biol 2001; 21:3523-33.
- Axilbund JE, Brune KA, Canto MI, Brehon BC, Wroblewski LD, Griffin CA. Patient perspective on the values of genetic counseling for familial pancreatic cancer. Hereditary Cancer in Clinical Practice 2005; 3:115-22.
- Canto MI, Goggins M, Hruban RH, Fishman EK, Axilbund JE, Griffin CA, Ali SZ, Richman J, Jagannath S, Kantsevoy SV, Petersen GM, Giardiello FM, Kalloo AN. Screening for early pancreatic neoplasia in high-risk individuals: A prospective controlled study. Clin Gastroenterol Hepatol 2006: 4:766-81.
- Canto MI, Goggins M, Yeo CJ, Griffin C, Axilbund JE, Brune KA, Ali SZ, Jagannath S, Petersen GM, Fishman EK, Piantadosi S, Giardiello FM, Hruban RH. Screening for pancreatic neoplasia in high-risk individuals: An EUS-based approach. Clinical Gastroenterology and Hepatology 2004; 2:606-21.
- Rulyak SJ, Kimmey MB, Veenstra DL, Brentnall TA. Cost-effectiveness of pancreatic cancer screening in familial pancreatic cancer kindreds. Gastrointestinal Endoscopy 2003; 57:23-9.
- 60. Hustinx SR, Hruban RH, Leoni LM, Iacobuzio-Donahue CA, Cameron JL, Yeo CJ, Brown PN, Argani P, Ashfaq R, Fukushima N, Goggins M, Kern SE, Maitra A. Homozygous deletion of the MTAP gene in invasive adenocarcinoma of the pancreas and in periampullary cancer: A potential new target for therapy. Cancer Biol Ther 2005; 4:83-6.
- van der Heijden MS, Brody JR, Dezentje DA, Gallmeier E, Cunningham SC, Swartz MJ, Demarzo AM, Offerhaus GJ, Isacoff WH, Hruban RH, Kern SE. In vivo therapeutic responses contingent on Fanconi Anemia/BRCA2 status of the tumor. Clin Cancer Res 2005; 11:7508-15.
- Gerald D, Berra E, Frapart YM, Chan DA, Giaccia AJ, Mansuy D, Pouyssegur J, Yaniv M, Mechta-Grigoriou F. JunD reduces tumor angiogenesis by protecting cells from oxidative stress. Cell 2004; 118:781-94.
- Weitzman JB, Fiette L, Matsuo K, Yaniv M. JunD protects cells from p53-dependent senescence and apoptosis. Mol Cell 2000; 6:1109-19.