Published in final edited form as:
Cancer Biol Ther. 2008 January ; 7(1): 135–144.

Allele-specific expression in the germline of patients with familial pancreatic cancer:

An unbiased approach to cancer gene discovery

Aik Choon Tan[1], **Jian-Bing Fan**[2], **Collins Karikari**[3], **Marina Bibikova**[2], **Eliza Wickham Garcia**[2], **Lixin Zhou**[2], **David Barker**[2], **David Serre**[4], **Georg Feldmann**[3], **Ralph H. Hruban**[3], **Alison P. Klein**[3], **Michael Goggins**[3], **Fergus J. Couch**[5], **Thomas J. Hudson**[6], **Raimond L. Winslow**[1], **Anirban Maitra**[1,3,7,*], and **Aravinda Chakravarti**[1,7,*]

[1]The Institute for Computational Medicine; Johns Hopkins University; Baltimore, Maryland USA

[2]Illumina Inc.; San Diego, California USA

[3]The Sol Goldman Pancreatic Cancer Research Center

[4]McGill University and Genome Quebec Innovation Center; Montreal, Quebec, Canada

[5]Department of Laboratory Medicine and Pathology; Mayo Clinic College of Medicine; Rochester, Minnesota USA

[6]Ontario Institute for Cancer Research; Toronto, Ontario, Canada

[7]McKusick-Nathans Institute of Genetic Medicine; Johns Hopkins University School of Medicine; Baltimore, Maryland USA

Abstract

Physiologic allele-specific expression (ASE) in germline tissues occurs during random X-chromosome inactivation[1] and in genomic imprinting,[2] wherein the two alleles of a gene in a heterozygous individual are not expressed equally. Recent studies have confirmed the existence of ASE in apparently non-imprinted autosomal genes;[3–14] however, the extent of ASE in the human genome is unknown. We explored ASE in lymphoblastoid cell lines of 145 individuals using an oligonucleotide array based assay. ASE of autosomal genes was found to be a very common phenomenon in ∼20% of heterozygotes at 78% of SNPs at 84% of the genes examined. Comparison of 100 affected individuals from familial pancreatic cancer kindreds and 45 controls revealed three types of changes in the germline: (a) loss of ASE, (b) gain of ASE, and, (c) rare instances of “extreme” (near monoallelic) ASE. The latter changes identified heterozygous deleterious mutations in a subset of these genes. Consequently, an ASE assay efficiently identifies candidate disease genes with altered germline expression properties as compared to controls, and provides insights into mechanisms that confer an inherited disease risk for pancreatic cancer.

[*]Correspondence to: Aravinda Chakravarti; McKusick-Nathans Institute of Genetic Medicine; Johns Hopkins University School of Medicine; Broadway Research Building, Suite 579; 733 N. Broadway; Baltimore, Maryland 21205 USA; Tel.: 410.502.7525; Fax: 410.502.7544; aravinda@jhmi.edu/ Anirban Maitra; The Sol Goldman Pancreatic Cancer Research Center; CRB-2, Room 345; Johns Hopkins University School of Medicine; 1550 Orleans Street; Baltimore, Maryland 21231 USA; Tel.: 410.502.8191; Fax: 410.614.0671; amaitra1@jhmi.edu.

Supplemental materials can be found at: http://www.landesbioscience.com/TanCBT7-1-sup.pdf

Keywords

allele-specific; familial; pancreatic; mutation; microarray; cancer; regulation

Introduction

Transcription is under exquisite genetic control and many of the processes that control this phenomenon are now known.[15] It has been demonstrated that the absolute transcript levels of many genes vary across individuals in numerous species.[4–14,16–19] These variations in transcript levels can result from a number of heritable inter-individual DNA sequence differences, such as single nucleotide polymorphism(s) (SNPs) or copy number variation (CNV)[16] affecting cis- and transacting elements.[20] The impact of these genotypic differences on transcription can range from minimal to profound, with the potential to modulate the corresponding phenotype. Gene expression as assessed by current techniques represents the compendium of transcripts produced by both parental alleles. However, the absolute transcript level fails to account for potential imbalances in relative allelic contribution. This perspective is particularly important for familial cancers, where an individual inherits a germline mutation on one parental allele, followed by a somatic mutation of the second allele in the tumor cells.[21] In these individuals, the transcript pool for the abrogated gene is expected to be contributed predominantly or exclusively from the wild type allele in germline tissues, such as in lymphocytes. Irrespective of the underlying molecular mechanism for allele specific expression (ASE), the ready availability of SNPs within gene transcripts (cSNPs) can be used as convenient “tags” for assessing relative allelic contribution, and identifying instances of profound imbalance that might suggest an underlying genomic abnormality. This “forward genetics” approach is particularly valuable for the many cancers that have a high mortality rate, rendering many of the standard genetic paradigms difficult, if not impossible, to execute. Broadly, this approach could be extended to any disease with a genetic component.

We conducted a large-scale study to examine the extent of ASE in the germline of controls and patients with familial pancreatic cancer, a cancer with an extremely high mortality rate,[22] to test this paradigm. We focused analysis on 2,117 exonic SNPs in 663 genes involved in a variety of cellular processes that likely contribute to tumorigenesis, including regulation of the cell cycle, cell signaling and apoptosis and genes involved in xenobiotic metabolism (Suppl. Table 1). Of these, 23 were known imprinted genes and 16 were X-linked genes, both of which are expected to display different types of ASE and can serve as different positive controls for our assay (Suppl. Table 2). To quantify the differential allelic expression of SNPs we used the Illumina GoldenGate® assay followed by hybridization to universal bead arrays,[23,24] the same technology used for large-scale SNP genotyping[23] and gene expression profiling,[25] except that genomic DNA (gDNA) and cDNA (from mRNA) were independently assessed and then compared to each other.[24] This platform provides a precise measure of ASE as compared to other conventional gene expression profiling platforms since self-normalized allelic ratios, as compared to absolute intensities, can be assessed. We collected and analyzed the genotypes and ASE values from lymphoblastoid cell lines (LCLs) of 100 familial pancreatic cancer patients ascertained through the National

Family Pancreatic Tumor Registry (NFPTR)[26] maintained at the Johns Hopkins University School of Medicine. Among the patients studied, 97 of them are Caucasians, two African-American and one Hispanic. These pancreatic cancer patients come from 98 families. For controls, we used germline samples of 45 individuals from 16 CEPH (Centre d'Etude du Polymorphisme Humain) reference families (Suppl. Table 3).

Results

Detecting ASE in the human genome

To analyze the allele expression data we focused attention on the subset of 413 SNPs in 250 genes that had minor allele frequency (MAF) > 0.1 and were highly informative (Suppl. Fig. 1). To identify ASE in heterozygote germline samples, we used a computational method based on locus-specific linear regression models. Briefly, the distribution of the $\log_2$ dye-intensity ratios for the two SNP alleles were used to predict the ratio and expected range for heterozygotes based on the observed distributions for each homozygote, at the same locus. Observed $\log_2$ dye-intensity ratios for each heterozygote at each SNP were then tested to assess whether they fall in (no ASE) or outside (ASE) these boundaries; these boundaries were calculated based on twice the standard deviation from the mean to account for the relatedness of the CEPH individuals (Fig. 1A). The degree of ASE, for each individual heterozygote, was estimated by the ASE score (θ) that measured the deviation of the observed $\log_2$ dye-intensity ratio from the expected boundaries for each heterozygote. As defined (see Methods), $\theta \geq 1$ corresponds to allele specific expression ratios of 2-fold or greater but does not provide information on which allele at a SNP shows greater or lesser expression since allele designations were arbitrary.

Assessing germline ASE in imprinted and X-linked genes

We first tested this approach using 12 positive control genes. SNPs within known imprinted genes are expected to show extreme ASE as exemplified by the gene *SNRPN* (rs705) (Fig. 1B); overall, 307 of 336 (91%) heterozygotes at 10 SNPs in 6 imprinted genes (*MEST*, *PEG10* and *SNRPN*: paternally imprinted; *ATP10A*, *CPA4* and *KCNQ1*: maternally imprinted) were detected as displaying ASE. The mean ASE score for these imprinting cases is $\theta = 3.1$ corresponding to an average 8.6-fold difference between the expression of the two alleles (Table 1). Heterozygote females for X-linked genes should also demonstrate an element of ASE based on the gene assayed.[1] The blood cell lineage is known to arise from 8–16 precursor cells which are known to undergo random, independent X-inactivation, and, thus, chance can create an observable skew and ASE.[27] As expected, we detect ASE as exemplified by the gene *BIRC4* (rs9856) (Fig. 1C); overall, 202 of 277 (73%) heterozygotes at 11 SNPs in 6 X-linked genes (*BIRC4*, *BTK*, *GUCY2F*, *MECP2*, *IRAK1* and *FHL1*) were detected as displaying ASE. The average ASE score for the X-linked cases is similarly 3.4 corresponding to an average 10.9-fold expression difference between the two alleles (Table 1). The larger threshold and the greater ASE variability for X-linked, as compared to imprinted, genes is expected since, early in development, imprinting is imposed uniformly on all cells whereas X inactivation occurs independently in each precursor cell. Thus, at least for classically imprinted genes, ASE can be efficiently detected using the Illumina BeadArray™ technology with a low (9%) false negative rate. This technology, and our

analysis method, also leads to a low (10%) false positive rate as detailed in the studies of Serre et al.[28] Consequently, the reliability of ASE detected in our study is high and accurate.

ASE rate in the human genome

To estimate the background ASE rate in the human genome we focused on all 17,237 heterozygotes of 45,683 genotypes at 392 SNPs in 238 autosomal nonimprinted genes in the 45 CEPH and 100 NFPTR samples. We observed that ASE is widespread since 19.6% (3,372/17,237) of heterozygotes at 78% (306/392) of SNPs at 84% (200/238) of genes demonstrated ASE in the germline. The population shows a wide distribution in the magnitude of ASE with an average θ of 0.65 (1.6-fold difference); moreover, 3.6% of heterozygotes show an extreme ASE with expression differences 4-fold or greater (Fig. 2). This demonstrates that germline ASE, as assessed using LCLs, is a persistent and widespread feature of the human genome and could be a potent mechanism for phenotypic variation in the population.

Since ASE is widespread, we queried whether germline samples from CEPH and NFPTR individuals differ in any manner with respect to ASE. We first examined the distribution of the fraction of SNPs displaying ASE in the CEPH and NFPTR individuals and these appear identical. Our results show that, 1,062 of 5,363 or 19.8% of heterozygotes at 225 SNPs in 157 genes showed ASE (average θ = 1.56) in the 45 CEPH samples, whereas 2,310 of 11,874 or 19.5% of heterozygotes at 292 SNPs in 198 genes showed ASE (average θ = 1.46) in the 100 NFPTR samples. Consequently, germline ASE is widespread and of equal magnitude in both the control and pancreatic cancer samples. If there are ASE differences in the germline between pancreatic cancer patients and CEPH controls, then they are not apparent at this level. However, as shown in Figure 3, the NFPTR samples, as compared to the CEPH samples, show a definite skew towards the lower end implying that many genes show reduced ASE in pancreatic cancer.

Comparing germline samples from cancer and control patients, however, we find highly significant differences in the behavior of individual SNPs and genes. A total of 211 SNPs demonstrated ASE in both CEPH and NFPTR samples (3,047/9,868 (31%) heterozygotes, average θ = 1.52: Suppl. Table 4); 14 SNPs exhibited ASE exclusively in one or more CEPH samples (19/176 (11%) heterozygotes, average θ = 1.47: Suppl. Table 5); 81 SNPs exhibited ASE exclusively in one or more NFPTR samples (306/2,261 (14%) heterozygotes, average θ = 1.38: Suppl. Table 6); and, 86 SNPs did not display ASE in either sample set (Suppl. Table 7) (Fig. 4). Although it is not unexpected that ASE rates will vary depending on whether SNPs demonstrating ASE are discovered in either CEPH or NFPTR or both or neither, the differences are greater than expected by chance after correction for the NFPTR and CEPH sample size difference. Thus, for the 14 SNPs displaying ASE only in CEPH the probability of not finding any in 350 NFPTR heterozygotes at the expected rate is 1.9×10^{-17}; similarly, for the 81 SNPs displaying ASE only in NFPTR the probability of not finding any at the expected rate in 855 CEPH heterozygotes is 1.0×10^{-52} (Table 2). In turn, the 211 SNPs displaying ASE in both CEPH and NFPTR do so at the near identical rates of 33% and 30%, respectively (Table 2).

These data strongly suggest that there are four classes of SNPs to consider. The most frequent class of SNPs includes those that show ASE in both germline CEPH and NFPTR samples at a high rate (∼30%); these are likely to result from polymorphisms in sequences regulating gene expression. The second class consists of SNPs that do not demonstrate ASE. Two remaining classes of SNPs are those with discordant ASE in the germline of control and cancer samples. These are the most intriguing since they likely represent SNPs in genes that are either silenced or have lost silencing in the germline of pancreatic cancer patients and they represent 95 of the 392 (24%) SNPs we investigated. These genes are particular candidates for an inherited predisposition to pancreatic tumorigenesis, since the probability of both false positives and false negatives is ∼10%.

Extreme ASE in pancreatic cancer genome

However, the most interesting genes may be the ones that showed “extreme” ASE. Genetic changes that profoundly elevate the expression of oncogenes or reduce the expression of tumor suppressor genes result in tumor development.[3] We hypothesize that an “expression threshold” may be required for oncogenesis and that only “extreme” ASE patterns may be significant in affected individuals. Genes that exhibit ASE patterns with scores $\theta \geq 2$ exclusively in individuals with familial pancreatic cancer were selected as pancreatic cancer candidate genes. We observed that “extreme” ASE is more common in the pancreatic cancer germline since 390/625 heterozygotes (62.4%) that demonstrated this pattern are NFPTR samples (Fig. 2). Moreover, affected individuals are at the higher end of the ASE spectrum as 88 NFPTR heterozygotes demonstrated “extreme” ASE exclusively in 52 genes (58 marked SNP transcripts) (Table 3). The fraction of NFPTR individuals exhibiting such variation ranged from 2% to 33% (median = 5%) with allelic expression differences ranging from 4.2 to 55.3-fold. Table 3 lists the 52 candidate genes. These include known pancreatic cancer-related genes (e.g., *BRCA2*, *FANCA*, *FANCD2*, and *PTCH1*)[29–31] and novel candidate genes (e.g., *BARD1*, *CDH1*, *NBN*).

Validating candidate genes

We experimentally validated two candidate genes to demonstrate that the ASE array results were reliable. The type I E-cadherin gene *CDH1*, based on two NFPTR individuals, showed ASE for the SNP rs1801552 (Fig. 5A). DNA sequencing for one patient (NFPTR19) verified the existence of ASE for *CDH1*, showing a heterozygote at the gDNA level (Fig. 5A) and homozygote at the cDNA level (Fig. 5A). Germline mutations of the type I E-cadherin have been shown to be responsible for increased risk in familial gastric cancer.[32,33] This patient, however, has reported no gastric cancer either in self or kindred; instead the family is notable for 6 pancreatic cancers (the proband, both parents and three of his siblings). A similar analysis demonstrated that *CARD15* (also known as *NOD2)* also harbors monoallelic expression in a NFPTR kindred. Mutations of this gene are associated with susceptibility to Crohn’s disease.[34,35] In our study, two *CARD15* SNPs (rs2066842 and rs2066843) in patient NFPTR22 displayed significant preferential expression of allele C (Fig. 5B). DNA sequencing of this individual’s *CARD15* gene showed a heterozygote at the gDNA level but a homozygote at the cDNA level, indicating monoallelic expression for this gene (Fig. 5B).

To assess whether ASE cases represent disease mutations or not we next selected an extreme ASE profile at *BRCA2*. We identified one patient (NFPTR6) where a heterozygous *BRCA2* SNP (rs144848) showed preferential expression of allele C (Fig. 5C) suggesting that this individual carried a germline *BRCA2* mutation. Germline DNA was sequenced from this patient using an independent culture of the implicated LCL and the chromatogram in Figure 4B shows a heterozygous 2041InsA mutation in the *BRCA2* gene resulting in a truncating mutation with nonsense-mediated decay (Fig. 5C). Inherited *BRCA2* gene mutations are known to significantly increase the risk of pancreatic cancer and ∼17% of patients with familial pancreatic cancer harbor germline mutations in this gene.[22,26,36,37] This independent validation confirms the utility of using ASE profiling methods to discover genes responsible for familial pancreatic cancer.

Discussion

Our study reveals three important genetic and genomic lessons. First, ASE is quite widespread in the human genome. However, the magnitude of the inter-allelic expression difference is small since 96.4% of θ scores are less than two and correspond to a four-fold or smaller expression difference. Nevertheless, since so many common polymorphisms are associated with inter-allelic differences in transcript levels, and any individual is likely to harbor multiple SNPs or CNVs at the genes affecting a given trait, most genetic effects on a phenotype are likely polygenic. Second, there are SNPs that show significantly different ASE patterns between control and pancreatic cancer germlines. That is, over and above uncommon instances of “extreme” ASE (see below), the pancreatic cancer germline is unique in removing the ASE effect at some genes and enhancing the effect at yet others. While the underlying bases for these differences are likely to be multifactorial, epigenetic alterations such as allele-specific promoter methylation should be considered as a distinct possibility. Such alterations in promoter methylation could lead to both loss (LOI) and gain (GOI) of imprinting.[38] These changes could be inherited over a few generations and provide evidence of complex multifactorial inheritance. Thus, differences in the germline “epigenome” of familial pancreatic cancer and control patients need to be validated as one of the potential causes for the observed distinctions in the pattern of ASE. Third, “extreme” ASE patterns are more common in the germline of familial pancreatic cancer patients, and affected individuals are at the higher end of the ASE spectrum. As we demonstrate, some of these changes are mutations in candidate pancreatic cancer genes and so an ASE screen can selectively enrich for such genes. With the advent of rapid sequencing technology, DNA sequencing of these genes in a large set of patients is warranted and likely fruitful.

One could argue that these polygenic differences we identified between pancreatic cancer and CEPH germlines are simply an artifact of the in vitro propagation of lymphoblastoid cells. However, we emphasize that most of the cell lines were cultured by one individual (CK) under identical laboratory and media conditions. Furthermore, ASE in the germline appears to be stable over time, as the validation assays for confirming the extreme ASE of *CDH1, CARD15* and *BRCA2* (Fig. 5) were performed several months and cell passages subsequent to the original array experiment. Thus, although further study of different and more relevant tissues is warranted the present results to speak to a biological difference.

The genetics of cancer, including familial cancer, remains an unsolved problem in that its etiology is still largely unknown. This is particularly so for pancreatic cancer. For one, familial pancreatic cancer could be very heterogeneous and our NFPTR patients may each have a highly penetrant mutation in a different gene. Alternatively, familial pancreatic cancer could be due to the multifactorial pattern of ASE at a large number of specific genes. Both possibilities are suggested by the greater compendium of SNPs that that demonstrate ASE (including "extreme" ASE) in one, to at most a few, NFPTR individuals (Fig. 3). An unbiased large-scale ASE analysis of germline samples in familial pancreatic cancer patients would help elucidate polymorphisms, and genes thereof, that might be involved in conferring an inherited predisposition to pancreatic cancer. Indeed, technologies that can use common exonic SNPs to interrogate the expression of each allele of a gene in a quantitative manner, such as in this study, are highly desirable and may provide a useful mechanistic test of complex inheritance.

Materials and Methods

Cell lines

Lymphoblastoid cell lines (LCLs) were used as a source of germline DNA and RNA (cDNA), and unless otherwise specified, the term "germline" refers to LCLs as the source of nucleic acid material. LCLs were established and sampled from 100 individuals from the National Familial Pancreatic Tumor Registry (NFPTR) at the Johns Hopkins University School of Medicine.[2] This study was reviewed and approved by the institutional review board of The Johns Hopkins Medical Institutions (JHMI); written informed consent for genetic studies was obtained from all study participants. As controls, LCLs of 45 individuals from the Centre d'Etude du Polymorphisme Humain (CEPH) reference collection were selected for genotyping and allele-specific gene expression profiling. These individuals were sampled from 16 CEPH families; thirteen of these were trios. The families used were 1340, 1341, 1344, 1345, 1346, 1349, 1350, 1362, 1375, 1413, 1416, 1418, 1420, 1421, 1423 and 1424; further details of each cell line used are provided in Supplementary Table 3.

Illumina allele-specific expression (ASE) assay

We selected 2,117 exonic SNPs in 663 genes for assessing ASE. The genes were selected for their potential involvement in cancer and included those regulating the cell cycle, cell signaling and apoptosis; the detailed list of all genes and specific SNPs tested is given in Supplementary Table 1. Three oligonucleotides for each SNP were designed, synthesized and pooled, as required for the GoldenGate assay[23,24] except that care was taken to design the oligos for the coding DNA strand. The experimental protocols were similar to those used for high-throughput SNP genotyping[23] and gene expression profiling[25] except that DNA and RNA were independently tested on different arrays and compared to each other. RNA was converted into biotinylated cDNA[25] while gDNA was treated according to the GoldenGate SNP genotyping protocol. Biotinylated DNA (corresponding to gDNA or cDNA) was immobilized on paramagnetic beads and pooled SNP-specific assay oligonucleotides were annealed to the DNA. Hybridized oligonucleotides were then extended and ligated to generate amplifiable DNA templates. Subsequently, we performed PCR using universal fluorescently-labeled primers. Finally, single-stranded PCR products were hybridized to a

Sentrix® Array Matrix,[23] and the arrays were imaged using the BeadArray Reader.[39] 96 samples (DNA or RNA) were analyzed simultaneously on each Sentrix Array for all 2,117 SNPs. All experiments were carried out in duplicate. The raw data was deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/projects/geo/) under accession number GSE8054 and GSE8055. As an added measure of data reproducibility using the Bead Array platform, we performed two sets of paired "dye swap" experiments. In one set of experiments, we labeled the "A" allele in germline DNA with Cy5 and the "B" allele with Cy3, and "swapped" the dyes ("A" allele - Cy3, "B" allele - Cy5) for the matching experiment. We then determined the correlation between Cy5 and Cy3 channel intensities for the "A" allele, and similarly for the "B" allele, from the two experiments. In the second set, we correlated Cy5 versus Cy3 intensities for "A" and "B" alleles from the corresponding germline cDNA samples. The correlation coefficients (r^2) for the four independent dye-swap analyses were excellent (0.966 and 0.968 for the DNA dye swaps and 0.949 and 0.968 for the cDNA dye swaps, respectively) (data not shown). These experiments confirmed that the Bead Array platform is reproducible in terms of channel intensities for the two alleles.

Two-stage filtering algorithm

We performed a two-stage filtering algorithm to estimate the expression level of each SNP-specific transcript on the arrays. In the first stage, we estimated the background signal for each SNP allele by averaging the signal intensity from all homozygotes for that allele. Next, we used this background as a threshold to assess whether each SNP transcript in each individual is 'expressed' or 'unexpressed'. In the second stage, for each expressed SNP, we eliminated data from all uninformative SNPs defined as those with less than three heterozygotes or with no homozygotes.

ASE detection algorithm

We constructed a locus-specific SNP linear regression model to determine the extent of allele-specific gene expression for heterozygotes based on the work of Serre et al.[28] Let D_{AA}, D_{AB} and D_{BB} and, correspondingly, R_{AA}, R_{AB} and R_{BB}, denote the $\log_2$ ratio of the fluorescent dye signals for AA, AB and BB individuals (genotypes) at the DNA and RNA levels. Next, for each SNP, we computed the mean (μ) and mean deviation (δ) for each of the homozygote clusters at the DNA and RNA levels, and estimated the maximum and minimum range of variation for each of the homozygote clusters as $\mu \pm 2d$. If one of the homozygote clusters (e.g., AA) contained less than four individuals we assigned the maximum and minimum range of variation equal to those of the alternative allele (e.g., BB).

To estimate the "expected" range of expression variation for heterozygotes, based on the assumption of equal expression of each allele, we used the predicted midpoints of the maximum and minimum ranges of variation of the AA and BB homozygotes. Heterozygous individuals are expected to have their allele expression ratio fall within these expected borders and demonstrate no ASE; ASE is inferred whenever this allele expression ratio falls outside the expected range. Figure 1a graphically describes our procedure.

ASE Score (θ)

We scored each heterozygote individual (*j*) for each informative SNP (*i*) to obtain an ASE score (θ) using the ratio of the SNP transcript expression levels as shown in (Fig. 1a). Let *d* represents the distance of the *j*-th individual's deviate from the heterozygous cluster mean and *r* the distance from the expected borders, the score θ is computed as:

$$\theta(j,SNP_i)=\begin{cases}\frac{|d|}{|r|}=\frac{\left|R_{\mathrm{AB},j}-\max(R_{\mathrm{AB}})\right|}{\left|\overline{R_{\mathrm{AB}}}-\max(R_{\mathrm{AB}})\right|}, \text{if} R_{\mathrm{AB},j}>\max(R_{\mathrm{AB}}),\\ \frac{|d|}{|r|}=\frac{\left|R_{\mathrm{AB},j}-\min(R_{\mathrm{AB}})\right|}{\left|\overline{R_{\mathrm{AB}}}-\min(R_{\mathrm{AB}})\right|}\text{if} R_{\mathrm{AB},j}>\min(R_{\mathrm{AB}}).\end{cases}$$

where $R_{\mathrm{AB},j}$ is the expression value of the *j*-th individual, max(R_{AB}) and min(R_{AB}) are the maximum and minimum expected ranges for the heterozygote cluster, and R_{AB} the mean RNA expression of the heterozygous cluster. Thus, ASE is inferred when θ > 1 (2-fold or greater difference). For this study, we defined "extreme ASE" as θ ≥ 2 corresponding to at least a 4-fold difference in expression between the two alleles. These thresholds are, admittedly, arbitrary and determine the sensitivity and specificity of our ASE detection algorithm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research was funded by grants from the National Institutes of Health [P50CA062924 (A.M.), RO1HD28088 (A.C.), R01MH060007 (A.C.)], the Mayo Clinic Pancreatic Cancer SPORE [P50102701 (F.J.C.)], The Sol Goldman Pancreatic Cancer Research Center (R.H.H.) and the Henry J. Knott professorship (A.C.). The study was designed by J.B.F., D.B., A.M. and A.C.; genes and SNPs was selected by J.B.F. and A.M.; the chip was designed by L.Z.; array experiment and data extraction was performed by M.B. and E.W.G.; the array experiments were supervised by D.B. and J.B.F.; cell culture and DNA and RNA extraction was performed by C.K.; NFPTR samples was acquired by R.H.H., A.P.K. and M.G.; the analysis algorithm was designed by A.C.T., D.S., T.J.H., R.L.W. and A.C.; data analysis and interpretation was conducted by A.C.T., A.M. and A.C.; experimental validation was performed by G.F. and F.J.C.; A.C.T., A.M. and A.C. wrote the paper.

References

1. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature. 2005; 434:400–404. [PubMed: 15772666]
2. Reik W, Walter J. Genomic imprinting: Parental influence on the genome. Nature Rev Genet. 2001; 2:21–32. [PubMed: 11253064]
3. Yan H, Dobbie Z, Gruber SB, Markowitz S, Romans K, Giardiello FM, Kinzler KW, Vogelstein B. Small changes in expression affect predisposition to tumorigenesis. Nature Genet. 2002; 30:25–26. [PubMed: 11743581]
4. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS. Natural variation in human gene expression assessed in lymphoblastoid cells. Nature Genet. 2003; 33:422–425. [PubMed: 12567189]
5. Cheung VG, Spielman RS. The genetics of variation in gene expression. Nature Genetics. 2002; 32:522–525. [PubMed: 12454648]

6. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping deter-minants of human gene expression by regional and genome-wide association. Nature. 2005; 437:1365–1369. [PubMed: 16251966]
7. Morley M, et al. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004; 430:743–747. [PubMed: 15269782]
8. Ge B, Gurd S, Gaudin T, Dore C, Lepage P, Harmsen E, Hudson TJ, Pastinen T. Survey of allelic expression using EST mining. Genome Res. 2005; 15:1584–1591. [PubMed: 16251468]
9. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. Science. 2005; 307:1072–1079. [PubMed: 15718463]
10. o HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. Allelic variation in gene expression is common in the human genome. Genome Res. 2003; 13:1855–1862. [PubMed: 12902379]
11. Monks SA, et al. Genetic inheritance of gene expression in human cell lines. Am J Hum Genet. 2004; 75:1094–1105. [PubMed: 15514893]
12. Pant PV, Tao H, Beilharz EJ, Ballinger DG, Cox DR, Frazer KA. Analysis of allelic differential expression in human white blood cells. Genome Res. 2006; 16:331–339. [PubMed: 16467561]
13. Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brändström H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnett D, Hudson TJ. A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics. 2004; 16:184–193. [PubMed: 14583597]
14. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. Science. 2002; 297:1143. [PubMed: 12183620]
15. Messina DN, Glasscock J, Gish W, Lovett M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. Genome Research. 2004; 14:2041–2047. [PubMed: 15489324]
16. Cheng Q, Yang W, Raimondi SC, Pui CH, Relling MV, Evans WE. Karyotypic abnormalities create discordance of germline genotype and cancer cell phenotypes. Nature Genet. 2005; 37:878–882. [PubMed: 16041371]
17. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science. 2002; 296:752–755. [PubMed: 11923494]
18. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. Genetics of gene expression surveyed in maize, mouse and man. Nature. 2003; 422:297–302. [PubMed: 12646919]
19. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variatns account for differences in gene expression among ethic groups. Nature Genet. 2007; 39:226–231. [PubMed: 17206142]
20. Pastinen T, Ge B, Hudson TJ. Influence of human genome polymorphism on gene expression. Human Molecular Genet. 2006; 15:R9–R16.
21. Knudson AG. Mutation and cancer: Statistical study of retinoblastoma. Proc Nat Acad Sci USA. 1971; 68:820–823. [PubMed: 5279523]
22. Maitra A, Kern S, Hruban RH. Molecular pathogenesis of pancreatic cancer. Best Practice and Research Clinical Gastroenterology. 2006; 20:211–226. [PubMed: 16549325]
23. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS. Highly parallel SNP genotyping. Cold Spring Harbor Symp Quant Biol. 2003; 68:69–78. [PubMed: 15338605]
24. Fan JB, Chee MS, Gunderson KL. Highly parallel genomic assays. Nat Rev Genet. 2006; 7:632–644. [PubMed: 16847463]
25. Fan JB, Yeakley JM, Bibikova M, Chudin E, Wickham E, Chen J, Doucet D, Rigault P, Zhang B, Shen R, McBride C, Li HR, Fu XD, Oliphant A, Barker DL, Chee MS. A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res. 2004; 14:878–885. [PubMed: 15123585]

26. 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]
27. Amos-Landgraf JM, Cottle A, Plenge RM, Friez M, Schwartz CE, Longshore J, Willard HF. X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. Am J Hum Genet. 2006; 79:493–499. [PubMed: 16909387]
28. Serre D, et al. Differential allelic expression in the human genome: Experimental reliability, biological relevance. PLoS Genetics. 2008
29. Petersen GM, Hruban RH. Familial pancreatic cancer: Where are we in 2003? Journal of the National Cancer Institute. 2003; 95:180–181. [PubMed: 12569133]
30. van der Heijden MS, Yeo CJ, Hruban RH, Kern SE. Fanconi Anemia gene mutations in young-onset pancreatic cancer. Cancer Research. 2003; 63:2585–2588. [PubMed: 12750283]
31. Lau J, Kawahira H, Hebrok M. Hedgehog signaling in pancreas development and disease. Cellular and Molecular Life Sciences. 2006; 63:642–652. [PubMed: 16465449]
32. Gayther SA, Gorringe KL, Ramus SJ, Huntsman D, Roviello F, Grehan N, Machado JC, Pinto E, Seruca R, Halling K, MacLeod P, Powell SM, Jackson CE, Ponder BA, Caldas C. Identification of germline E-cadherin mutations in gastric cancer families of european origin. Cancer Res. 1998; 58:4086–4089. [PubMed: 9751616]
33. Richards FM, McKee SA, Rajpar MH, Cole TR, Evans DG, Jankowski JA, McKeown C, Sanders DS, Maher ER. Germline E-cadherin gene (*CDH1*) mutations predispose to familial gastric cancer and colorectal cancer. Hum Mol Genet. 1999; 8:607–610. [PubMed: 10072428]
34. Hugot JP, et al. Associated of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001; 411:599–603. [PubMed: 11385576]
35. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nuñez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001; 411:603–606. [PubMed: 11385577]
36. Goggins M, Schutte M, Lu J, Moskaluk CA, Weinstein CL, Petersen GM, Yeo CJ, Jackson CE, Lynch HT, Hruban RH, Kern SE. Germline *BRCA2* gene mutations in patients with apparently sporadic pancreatic carcinomas. Cancer Res. 1996; 56:5360–5364. [PubMed: 8968085]
37. Murphy KM, Brune KA, Griffin C, Sollenberger JE, Petersen GM, Bansal R, 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; 62:3789–3793. [PubMed: 12097290]
38. Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, Feinberg AP. Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. Science. 2003; 299:1753–1755. [PubMed: 12637750]
39. Barker DL, et al. Self-assembled random arrays: High-performance imaging and genomics applications on a high-density microarray platform. Proc SPIE. 2003; 4966:1–11.

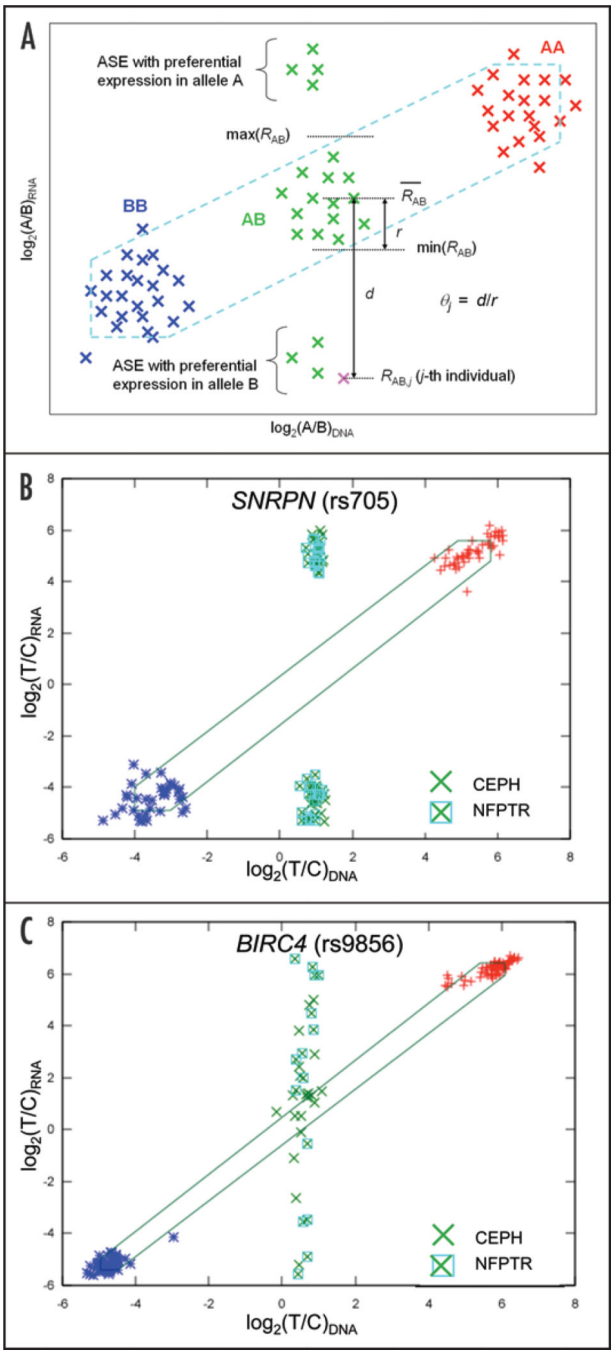


Figure 1.

Allele-specific expression (ASE) detection. (A) Locus-specific ASE detection regression model: Red and blue crosses represent AA and BB homozygotes, respectively; green crosses are heterozygotes. Light blue lines represent the estimated linear models of the expected heterozygote ranges for that particular SNP. Heterozygotes outside the lines represent ASE. Computation of the ASE score (θ): The pink cross represents the *j*-th heterozygous individual with RNA expression ratio $R_{AB,j}$; max(R_{AB}) and min(R_{AB}) are the maximum and minimum "expected" ranges for the heterozygote cluster, and R_{AB} the mean of the

heterozygous cluster. θ is computed as the relative distance of the *j*-th individual deviate from the heterozygous cluster (*d*) as compared to the expected borders (*r*) (See Methods). (B) Data on the rs705 SNP at the imprinted gene *SNRPN* (positive control) in 64 CEPH and NFPTR samples. (C) Data on the rs9856 SNP at the X-linked gene *BIRC4* (positive control) in 35 CEPH and NFPTR female samples. The colors are as in (A).

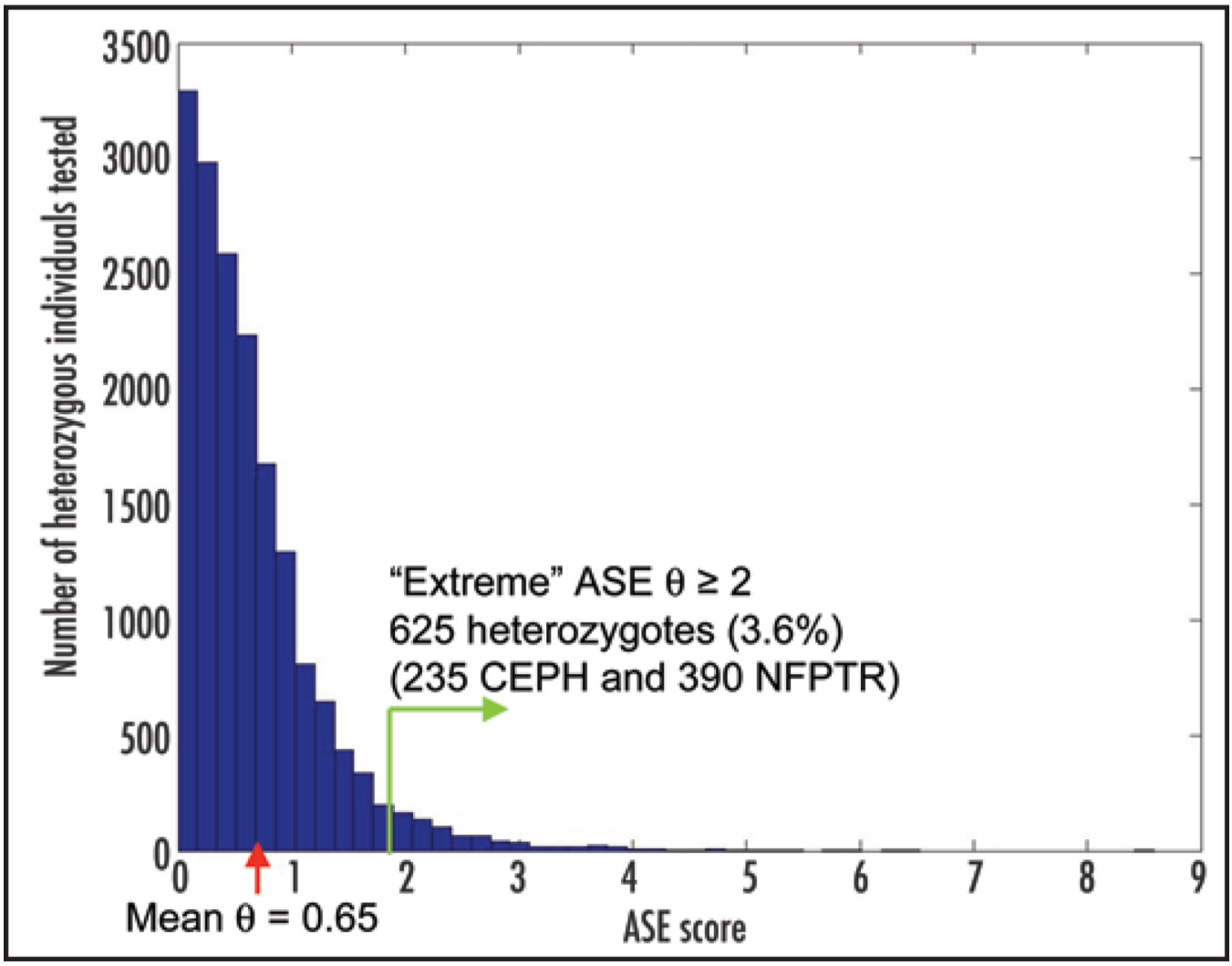


Figure 2.

The distribution of ASE in human samples. The ASE score (θ) distribution for all heterozygous individuals tested, CEPH and NFPTR, is shown. The mean score for 17,237 heterozygotes is 0.65 with 3.6% (625/17,237) showing "extreme" ASE (θ ≥ 2).

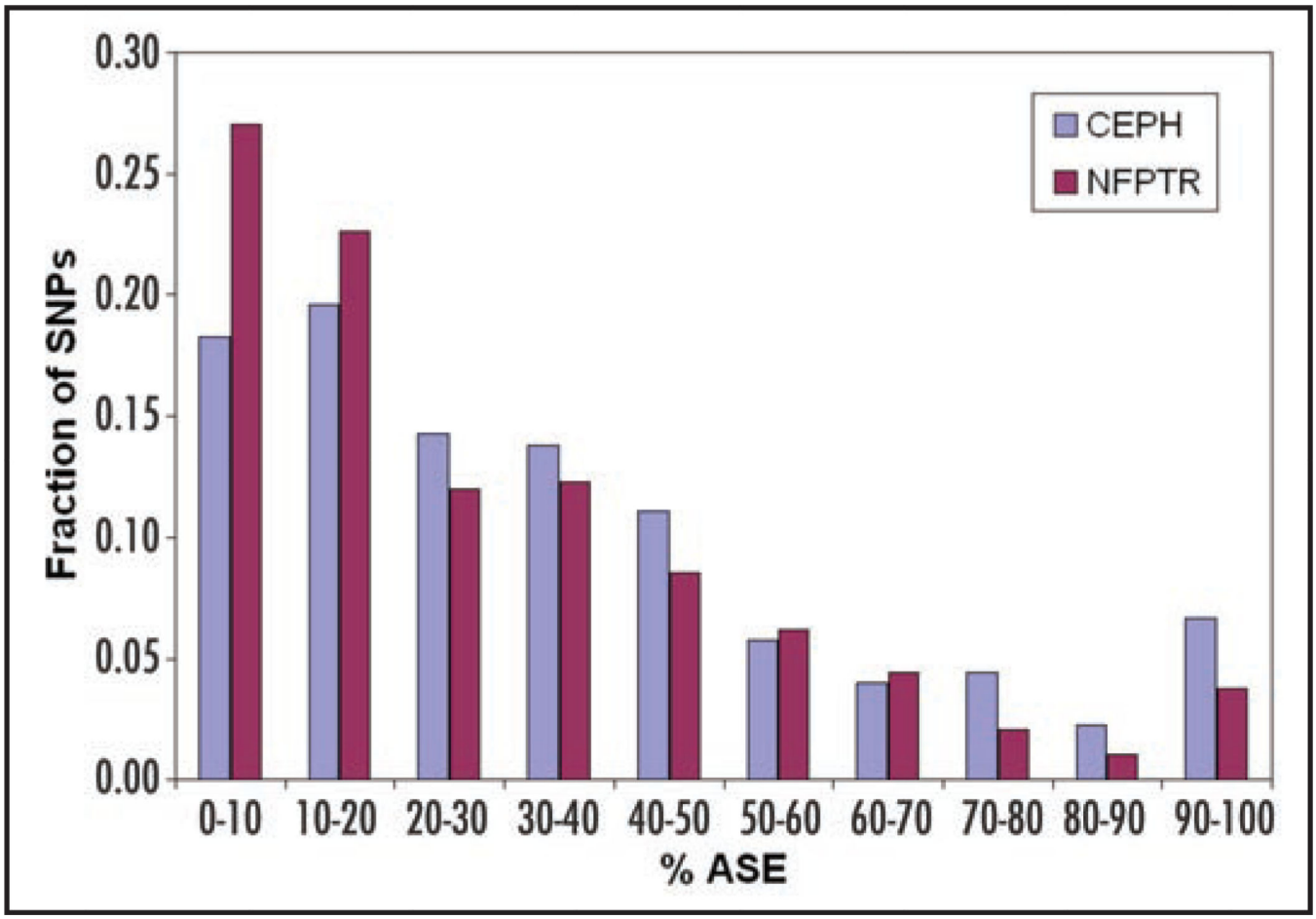
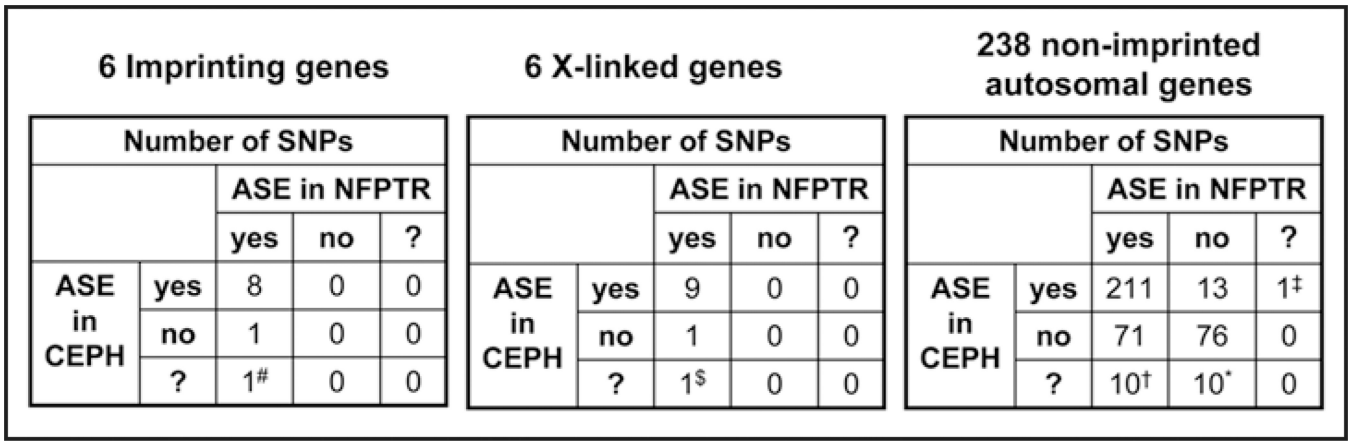


Figure 3.

Global ASE distribution in control and pancreatic cancer samples. A genomic ASE genome index may be estimated as the fraction of heterozygotes displaying ASE (Y-axis represents the percentage of SNPs normalized within control and pancreatic cancer samples). The global distribution of this genomic index is not statistically significant ($p < 0.15$) between CEPH and NFPTR samples, although there is a trend towards lower ASE levels in the pancreatic cancer germline.



6 Imprinting genes

Number of SNPs				
		ASE in NFPTR		
		yes	no	?
ASE in CEPH	yes	8	0	0
	no	1	0	0
	?	1[#]	0	0

6 X-linked genes

Number of SNPs				
		ASE in NFPTR		
		yes	no	?
ASE in CEPH	yes	9	0	0
	no	1	0	0
	?	1[$]	0	0

238 non-imprinted autosomal genes

Number of SNPs				
		ASE in NFPTR		
		yes	no	?
ASE in CEPH	yes	211	13	1[‡]
	no	71	76	0
	?	10[†]	10[*]	0

Figure 4.

The extent of ASE in human samples. A total of 413 SNPs in 250 genes were tested for ASE with distributions in control (CEPH) and familial pancreatic cancer (NFPTR) samples as shown. Also indicated are the distributions of 10 SNPs in 6 known imprinted and 11 SNPs in 6 X-linked genes tested; the results on all other genes are shown separately. The yes/no/? categories refer to heterozygotes showing ASE, heterozygotes not showing ASE or the absence of heterozygotes, respectively. [#]*ATP10A* (rs3816800) had no CEPH heterozygote. [$]*GUCY2F* (rs494589) had no CEPH heterozygote. [‡]Ten genes with ASE in NFPTR had no CEPH heterozygotes in our sample. These genes, and their SNPs, are: *CCR5* (rs1800023), *EPHA1* (rs10952549), *EPHA7* (rs7349683), *FGFR2* (rs1801043), *MAPK4* (rs3288), *MMP1* (rs5854), *MMP10* (rs470168), *RIPK4* (rs3746893), *SLC22A2* (rs3127594) and *TEK* (rs639225). [‡]*CDH17* (SNP rs9417) had no NFPTR heterozygote. *Seven genes with no ASE in NFPTR had no CEPH heterozygotes in our sample. These genes, and their SNPs, are: *AGTR1* (rs5182), *CSF1R* (rs216123), *MMP8* (rs1940475), *NAT2* (rs1208, rs1799929, rs1799930), *ROS1* (rs529038), *THBS1* (rs2228263) and *TNFSF8* (rs3181368).

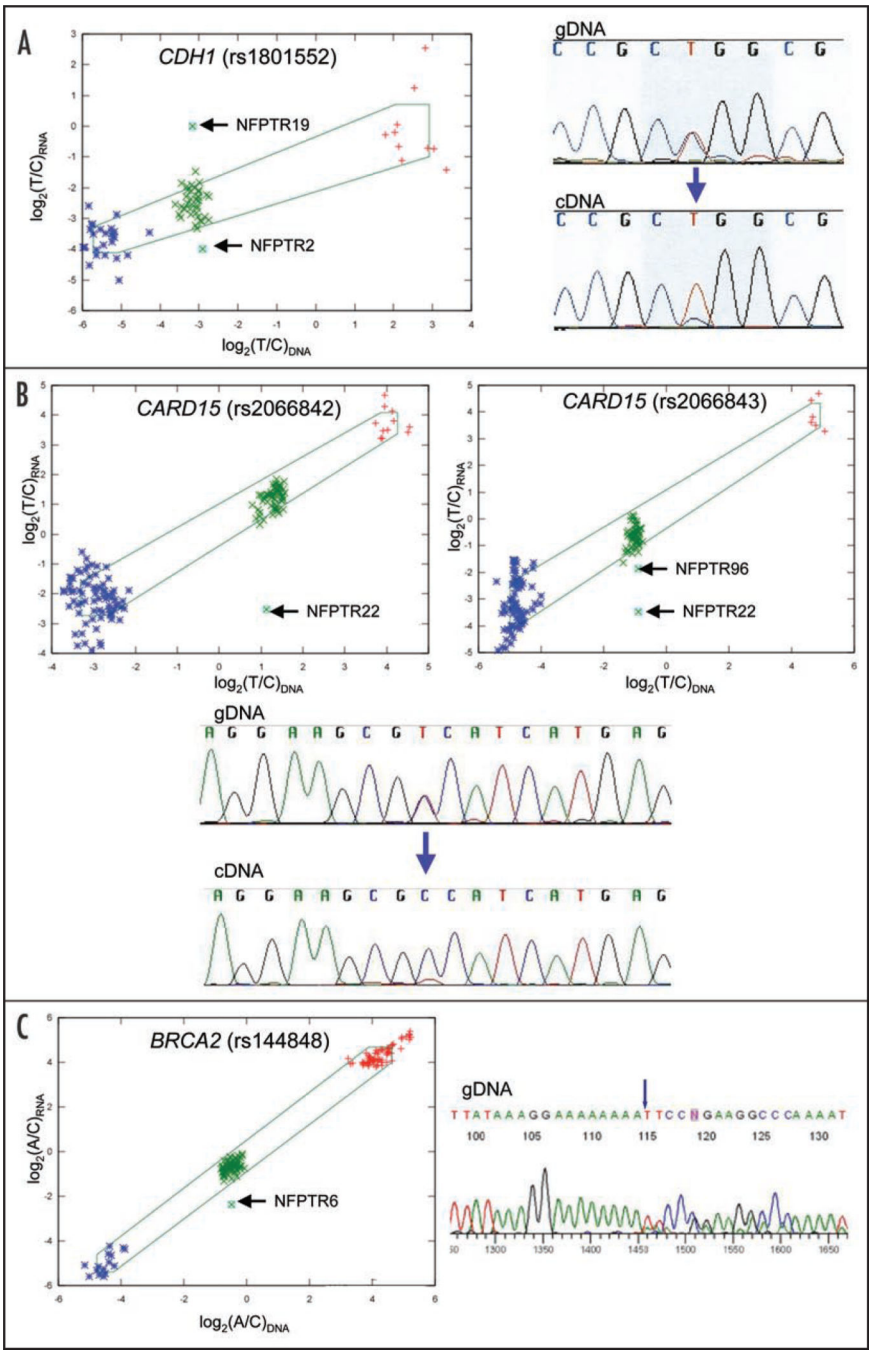


Figure 5.

Three genes displaying “extreme” ASE patterns in pancreatic cancer. (A) *CDH1* shows ASE in NFPTR19 and NFPTR2 (rs1801552). DNA sequencing verifies that NFPTR19 is a heterozygote (CT) at the gDNA level but monoallelic (T) at the cDNA level (blue arrow). (B) *CARD15* (rs2066842, rs2066842) shows ASE in two individuals (NFPTR22, NFPTR96) with NFPTR22 displaying ASE at both SNPs. DNA sequencing confirms that this latter individual is heterozygous (CT) at the gDNA level but monoallelic (C) at the cDNA level (blue arrow). (C) *BRCA2* (rs144848) shows ASE in one individual (NFPTR6) likely arising

from a deleterious 2041insA mutation (right panel). Confirmation of the *BRCA2* deleterious mutation was performed at Myriad Genetics, Salt Lake City, UT.

Table 1

Allele-specific expression at imprinted and X-linked genes in CEPH and NFPTR samples

			Heterozygous Individuals (CEPH + NFPTR)			
Gene	**RefSNP**	**Locus**	**Tested**	**# ASE**	**% ASE**	**Average θ***
(a) Imprinted Genes						
ATP10A	rs1047700	15q11.2	6	3	50	1.44
	rs3816800		7	5	71	2.24
CPA4	rs2171492	7q32	11	11	100	2.22
KCNQ1	rs1057128	11p15.5	13	10	77	2.34
	rs10798		64	60	94	3.72
	rs8234		64	54	84	2.42
MEST	rs10863	7q32	31	24	77	2.34
PEG10	rs13073	7q21	53	53	100	4.53
	rs3750105		23	23	100	4.75
SNRPN	rs705	15q11.2	64 336	64 307	100 91	5.01 3.10 (8.6)‡
(b) X-linked genes						
BIRC4	rs5956583	Xq25	29	17	59	3.72
	rs5958343		35	27	77	3.07
	rs8371		32	24	75	4.57
	rs9856		35	29	83	5.64
BTK	rs1057403	Xq21.33-q22	19	15	79	3.07
	rs700		32	25	78	3.28
FHL1	rs9018	Xq26	14	5	36	1.95
IRAK1	rs1059701	Xq28	34	26	76	2.53
	rs1059703		19	14	74	2.45
MECP2	rs2734647	Xq28	25	20	80	4.23
GUCY2F	rs494589	Xq22	3 277	0 202	0 73	3.45 (10.9)‡

*The average ASE score (θ) is calculated only for heterozygotes showing ASE.

‡Values in parentheses represent average fold change calculated from the average ASE score.

Table 2

ASE analysis of 392 SNPs in 238 autosomal nonimprinted genes in CEPH and NFPTR samples

ASE‡		#	CEPH (n = 45)						NFPTR (n = 100)					
CEPH	NFPTR	SNPs	#Genotypes	#hets	#ASE	%ASE	θ	FC‡	#Genotypes	#hets	#ASE	%ASE	θ	FC‡
+	+	211	7,970	3,162	1,043	33	1.57	3.0	17,442	6,706	2,004	30	1.50	2.8
+	−	14	469	176	19	11	1.47	2.8	938	350	0	0	0.37	1.3
−	+	81	2,513	855	0	0	0.37	1.3	6,199	2,261	306	14	1.36	2.6
−	−	86	3,127	1,170	0	0	0.30	1.2	7,025	2,557	0	0	0.32	1.2

‡The +/− categories refer to heterozygotes showing and not showing ASE, respectively.

‡FC indicates fold change calculated from θ.

Table 3

52 candidate genes (58 SNPs) exhibiting extreme ($\theta \geq 2$) ASE in familial pancreatic cancer (NFPTR) samples

							Heterozygous NFPTR individuals			
Gene	**Locus**	**RefSNP**	**MAF**	**SNP**	**Function**	**Class**[±]	**Tested**	**# extreme ASE**	**% extreme ASE**	**Extreme θ**[†]	**Background θ**[†]
Nonsynonymous SNPs											
MUTYH	1p34.3-p32.1	rs3219484	0.1	198G>A	22Val>Met	a	12	1	8	3	1.58
TNFRSF1B	1p36.3-p36.2	rs1061622	0.26	676T>G	196Met>Arg	b	36	2	6	2.4	1.53
EPHX1	1q42.1	rs1051740	0.32	612T>C	113Tyr>His	b	34	1	3	2.07	1.41
CASP8	2q33-q34	rs1045485	0.13	1192G>C	302Asp>His	b	8	1	13	2.63	1.68
BARD1 (1)[*]	2q34-q35	rs2229571	0.4	1207G>C	378Arg>Ser	b	42	1	2	2.27	1.37
PTCH1 (2)	9q22.3	rs357564	0.33	4132C>T	1315Pro>Leu	b	9	3	33	2.98	2.12
ZWINT	10q21-q22	rs2241666	0.38	596A>G	187Arg>Gly	b	38	1	3	4	1.51
BAG3	10q25.2-q26.2	rs2234962	0.21	757T>C	151Cys>Arg	a	3	1	33	2.26	1.72
GSTP1	11q13	rs947894	0.34	342A>G	105Ile>Val	a	42	1	2	2.44	1.38
MMP7	11q21-q22	rs10502001	0.2	277G>A	77Arg>His	b	31	2	6	2.32	1.51
FLT3	13q12	rs1933437	0.36	738C>T	227Thr>Met	a	24	1	4	2.07	1.61
BRCA2 (3)	13q12.3	rs144848	0.33	1341A>C	372Asn>His	a	46	1	2	2.33	2.33
TEP1	14q11.2	rs1760904	0.49	3624T>C	1195Ser>Pro	b	43	1	2	2.52	1.31
CARD15 (4)	16q12	rs2066842	0.28	907C>T	268Pro>Ser	a	40	1	3	5.79	1.83
FANCA	16q24.3	rs2239359	0.37	1543G>A	501Gly>Ser	b	49	3	6	2.38	1.56
PLAUR	19q13	rs2302524	0.17	889A>G	220Lys>Arg	b	25 482	2 23	8 5	2.32 2.74 (6.7)‡	1.58 1.63 (3.1)[‡]
Synonymous SNPs											
FRAP1 (5)	1p36.2	rs11121705	0.3	1516T>C	479Asp	b	42	4	10	2.57	1.57
FRAP1 (5)	1p36.2	rs1057079	0.27	4810G>A	1577Ala	b	34	3	9	2.57	1.54
DST	6p12-p11	rs2230862	0.5	4176G>A	1358Lys	a	54	1	2	2.37	1.17
DDR1	6p21.3	rs1049623	0.19	2130T>C	599Val	b	55	2	4	2.14	1.42
MET	7q31	rs41736	0.4	4045C>T	1286Asp	a	16	1	6	2.1	1.49
EPHA1	7q34	rs10952549	0.21	1924C>T	613Leu	a	5	1	20	2.39	2.1
PTK2B	8p22-p11.2	rs1030526	0.4	978G>A	110Thr	b	38	1	3	2.19	1.37
TEK	9p21	rs639225	0.48	2110A>G	654Ser	a	17	1	6	2.16	1.36

							Heterozygous NFPTR individuals				
Gene	**Locus**	**RefSNP**	**MAF**	**SNP**	**Function**	**Class±**	**Tested**	**# extreme ASE**	**% extreme ASE**	**Extreme θ†**	**Background θ†**
PTCH1 (2)	9q22.3	rs2066836	0.22	1874C>T	562Ala	b	39	3	8	2.67	1.68
CYP2E1	10q24.3-qter	rs2515641	0.13	1296T>C	421Phe	b	17	1	6	2.33	1.96
PDE1B	12q13	rs1249950	0.45	1642T>C	492Asn	a	46	1	2	2.06	1.44
BRCA2 (3)	13q12.3	rs1799955	0.21	7469A>G	2414Ser	b	32	1	3	2.8	1.57
TCF4	18q21.1	rs6567211	0.39	2123G>A	643Ser	b	42	1	2	4.92	4.92
MAP3K9	14q24.3-q31	rs3829955	0.17	2676C>T	892Asn	a	32	1	3	3.08	1.52
CARD15 (4)	16q12	rs2066843	0.28	1482C>T	459Arg	a	41	2	5	3.16	2
CDH1	16q22.1	rs1801552	0.41	2200T>C	692Ala	b	27	2	7	2.91	1.74
TOB1	17q21	rs4626	0.29	992A>G	319Lys	b	41	1	2	2.06	1.24
RIPK4	21q22.3	rs3746893	0.38	1524G>A	492Ala	b	9 587	2 29	22 5	4.18 2.70 (6.5)‡	2.35 1.80 (3.5)‡
Untranslated region SNPs											
BARD1 (1)	2q34-q35	rs1129804	0.31	44C>G	exon1	b	46	6	13	2.21	1.64
COL4A3	2q36-q37	rs2070735	0.14	5490C>A	exon52	a	16	1	6	2.78	1.7
FANCD2	3p25.3	rs7647987	0.21	4556G>A	intron43	b	29	1	3	2.06	1.37
TNFSF10	3q26	rs1131542	0.34	1297C>A	exon5	b	49	1	2	2.77	1.58
SPARC	5q31.3-q32	rs1059829	0.47	2120T>C	exon10	b	54	2	4	2.14	1.46
FRK	6q21-q22.3	rs495565	0.38	2566G>A	exon8	b	46	1	2	2.56	1.41
SLC22A2	6q26	rs3127594	0.13	2198T>A	exon11	a	9	3	33	2.45	2.02
SERPINE1	7q21.3-q22	rs1050813	0.23	2176G>A	exon9	a	28	2	7	2.11	1.6
Untranslated region SNPs											
SMO	7q32.3	rs1061285	0.13	3660C>A	exon12	b	18	1	6	2.32	1.5
NBN	8q21	rs1063045	0.31	212T>C	exon2	b	42	1	2	3.07	3.07
NOTCH1	9q34.3	rs6563	0.48	9010G>A	exon34	b	46	1	2	2.4	1.55
DNMT2	10p15.1	rs10904889	0.16	2082G>T	exon11	b	19	3	16	2.23	1.64
CCKBR	11p15.4	rs1042048	0.33	1974G>A	exon5	a	6	2	33	2.3	2
LRRC32	11q13.5-q14	rs3781701	0.32	3312T>C	exon34	a	17	1	6	2.26	1.47
MMP1	11q22.3	rs5854	0.33	1750T>C	exon10	a	16	1	6	2.13	1.4
KRAS(6)	12p12.1	rs13096	0.5	3636G>A	exon5	b	47	1	2	2.31	1.4
KRAS(6)	12p12.1	rs1801539	0.5	4534G>A	exon5	b	49	1	2	2.14	2.14

							Heterozygous NFPTR individuals			
Gene	**Locus**	**RefSNP**	**MAF**	**SNP**	**Function**	**Class**[±]	**Tested**	**# extreme ASE**	**% extreme ASE**	**Extreme θ**[†]	**Background θ**[†]
HDAC7A	12q13.1	rs9859	0.21	3798C>A	exon24	b	31	1	3	3.64	1.49
THBS1	15q15	rs1051442	0.15	3771T>C	exon22	b	9	1	11	2.32	1.73
LRRK1	15q26.3	rs1048326	0.14	4704T>C	exon20	b	28	1	4	2.36	1.52
IMPACT	18q11.2-q12.1	rs1053474	0.33	3452G>A	exon11	b	40	1	3	2.41	1.37
GNG7	19p13.3	rs3752174	0.26	808T>C	exon5	b	39	1	3	2.14	1.45
JAG1	20p12	rs7828	0.36	5201T>G	exon26	b	48	1	2	2.22	1.64
TFF2	21q22.3	rs225334	0.4	513G>A	exon4	b	12	1	8	2.13	1.45
							744	36	5	2.39 (5.3)‡	1.65 (3.1)[‡]

[*] Numbers in parentheses identify SNPs marking the same gene.

[±] Class a/b identifies SNPs showing ASE exclusively in NFPTR and those common to CEPH and NFPTR, respectively.

[†] The extreme and background θ are calculated from those showing ASE and all heterozygotes, respectively.

[‡] Values in parentheses represent average fold change calculated from the average θ.