Combining DNA Copy Number and Gene Expression Data to Reveal Sample-Specific Genetic Abnormalities in Pancreatic Cancer

Liviu Badea

AI Lab, National Institute for Research in Informatics
8-10 Averescu Blvd., Bucharest, Romania,
badea@ici.ro

Abstract: DNA copy number and gene expression data are complementary for deciphering the molecular-level mechanisms of cancer. In this paper, we use the Pollack lab pancreatic cancer dataset for determining a set of genes with consistent expression changes in pancreatic cancer and relate these genes with the DNA copy number changes observed in the individual samples. Our rank-covering algorithm can be used to obtain an aggregated view of the candidate genes with DCN alterations that potentially explain the observed phenotype. The algorithm recovered a set of genes with well-known roles in pancreatic cancer, but also suggested an additional promising set of genes, which seem to be involved in cancer, but whose exact role in pancreatic cancer remains to be determined experimentally.

Keywords: bioinformatics, gene expression data analysis, DNA copy number data analysis.

Liviu Badea, a senior researcher at the National Institute for Research and Development in Informatics, Bucharest, graduated with honors in Computer Science from "Politehnica" University Bucharest in 1990. In 1996 he obtained his PhD from the same university under the supervision of Prof. Cristian Giumale with a topic in Artificial Intelligence. Dr. Badea's current research interests are in the fields of Bioinformatics, Artificial Intelligence and the Semantic Web.

1. Introduction

Despite the enormous recent progress in understanding cancer at a molecular level, the precise details are still elusive for many types of carcinomas. Pancreatic cancer is a particularly aggressive disease, with a very poor prognosis, requiring a more precise understanding of its molecular pathogenesis. The technological progress initiated by the introduction of gene expression microarrays about a decade ago has enabled large scale whole genome studies with the aim of identifying disease-specific genes. Although limited by the relatively low number of samples (due to the large costs of the technology), these gene expression studies have revealed a much more complex molecular-level picture than previously expected. Tens to a few hundreds genes were found to be differentially expressed in the samples analyzed, but their precise roles in the (signaling) pathways leading to cancer are only partially known. Even worse, it seems extremely difficult to discern between genetic abnormalities that play a causal role in oncogenesis and those that are merely side-effects. Obviously, the task of identifying new therapeutic targets depends essentially on being able to identify the causal details.

Complementary knowledge about gene mutations, oncogene amplifications, or deletions of tumor-suppressor genes could be very helpful in this respect. The lack of large-scale genome-wide mutation or DNA copy number (DCN) data has been addressed by SNP arrays and microarray-based Comparative Genomic Hybridization (array-CGH) [5]. The latter enables the detection of localized DNA amplifications or deletions with an unprecedented spatial resolution. A nice characteristic of some current approaches (e. g. [5]) consists in using the same microarray platform both for expression *and* DNA copy number (array-CGH) studies. Although in colon cancer only 4% of the amplified genes have increased expression [12], DNA copy numbers are much better correlated with gene expression in at least a few other cancer types, such as breast [5] and pancreatic cancer [4,1].

The results of these studies have mainly emphasized the complexity of the genetic abnormalities involved in pancreatic cancer. There seem to be few, if any, such amplifications or deletions common to all patients thus suggesting a more complex picture of the disease in which perturbations of *distinct* components of certain key pathways are triggered in various different ways, while leading to similar phenotypes.

The fact that our knowledge of the various signaling pathways involved is only partial makes the task of identifying the precise details of oncogenesis even more difficult, requiring a combination of all the available data and knowledge. More specifically, neither the DNA copy number, nor the expression data are sufficient for our purposes. Indeed, certain genes may have changed DCNs without a corresponding change in expression levels[1]. Conversely, many genes suffer significant expression level changes that are

[1] As in the case of genes with very low expression levels in the tissues under study.

not due to DCN changes, but rather to transcriptional regulation, for example. Thus, both DCN and gene expression data are needed – DCN for identifying potential causes of expression changes and expression data for confirming these changes.

Previous studies have mainly concentrated on relating DCN changes with gene expression, e.g. by analyzing the correlation between the two [9]. Moreover, although most these studies have addressed various cancer types, an in-depth analysis of such carcinomas is still lacking.[2] In this paper we aim at a more detailed analysis of the DCN and gene expression data available from the Pollack lab pancreatic cancer study of Bashyam et al [1].

As the different samples seem to have underwent distinct genetic alterations in each sample, we assume that there exists a set of "*common genes*" responsible for the disease phenotype, which are either over- or under-expressed in *all* samples and are directly or indirectly affected by all these different sample-specific DCN changes. Our goal consists in identifying sets of genes with DCN alterations that are potentially responsible for the up-or down-regulation of these "common genes".

The very limited number of samples (23) will most probably not allow us to determine all DCN changes responsible for tumorigenesis in all samples. "Common" genes g_o with a complex combinatorial transcriptional control involving, among many others, a gene g with an altered DCN may be hard (if not impossible) to correlate with g using just 23 samples. However, we argue that such "isolated" DCN genes g are the least interesting as potential therapeutic targets, since they seem to affect g_0 just in a minority of cases. On the other hand, we are interested in altered genes g that are well correlated with the phenotype (and thus with the "common" genes g_0) for the majority of samples. We call such genetic abnormalities "*recurrent*" – these will be the main focus of our *rank covering* algorithm described below.

2. The Data Set

Bashyam et al. [1] have performed simultaneous array Comparative Genomic Hybridization and microarray expression measurements on a set of 23 human pancreatic cell lines (with two additional normal-normal reference array-CGH measurements) using cDNA microarrays containing 39632 human cDNAs (representing about 26000 named human genes). Array-CGH measurements involved co-hybridizing Cy5-labeled genomic DNA from each cell line along with Cy3-labeled sex-matched normal leukocyte DNA. Expression profiling was performed with reference RNA derived from 11 different human cell lines.

We retrieved the normalized intensity ratios from the Stanford Microarray Database [10] and used the CGH-Miner software [6] as described in [1] to identify DNA copy number gains and losses. We considered a DCN gain to be significant if the copy number ratio exceeds the threshold $\theta_{DCN+} = 1.5$, while a DCN loss was deemed significant if the ratio was below $\theta_{DCN-} = 0.5$ (The threshold for DCN amplifications corresponds to the duplication of one of the alleles of a gene in a diploid genome[3], while that for DCN losses is associated to the deletion of an allele).

Chromosomal positions of the arrayed cDNA clones were determined using the 2004 freeze of the UCSC genome browser [2] and clones without a known position or assigned to "random" chromosomes were excluded.

Since the reference used in the expression measurements was a cell line mixture rather than normal pancreatic tissue, we retrieved from public databases two additional samples (Gene Expression Omnibus sample GSM39990 and Stanford Microarray Database sample shao069), measuring the expression levels of genes in normal pancreatic tissue relative to the same 11 cell line mixture used in the Bashyam et al. study. Unfortunately, the large variance of the normal measurements (Pearson correlation of $\log_2$ ratios 0.6) and the lack of more normal samples represent a significant limitation. Moreover, the normal samples contain only 14179 well measured probes (genes), further limiting the analysis to this subset of genes.

We calculated expression ratios relative to the normal samples as follows:

[2] Most published papers in this area simply enumerate and briefly comment short lists of genes selected by human experts (rather than by an automated procedure). This has significant advantages due to the high complexity of expertise required (which is largely non-automatable with current technology), but also the disadvantage of human bias (mainly due to the incompleteness of current knowledge).

[3] Thus leading to a copy number ratio of 3/2.

$$r(Bashyam\ vs\ normal) = \frac{r(Bashyam\ vs\ reference)}{mean(r(normal\ vs\ reference))}.$$

Additionally, we applied for each gene a SNR test (a slight variation of the unpaired t-test) to take into account the variance of the gene in the normal samples when selecting up- or down-regulated genes.

$$SNR = \frac{mean(\log_2 r(Bashyam\ vs\ reference)) - mean(\log_2 r(normal\ vs\ reference))}{std(\log_2 r(Bashyam\ vs\ reference)) + std(\log_2 r(normal\ vs\ reference))}.$$

Expression ratios were called significant if they either exceed the threshold $\theta_{EXPR+} = 2$, or were below $\theta_{EXPR-} = 0.5$.

The following notations will be used throughout the paper:

$DCN+(g,s)$	for $DCN(g,s) \geq \theta_{DCN+}$
$DCN-(g,s)$	for $DCN(g,s) \leq \theta_{DCN-}$
$EXPR+(g,s)$	for $r(g,s) \geq \theta_{EXPR+}$
$EXPR-(g,s)$	for $r(g,s) \leq \theta_{EXPR-}$
$EXPR^{(0)}(g,s)$	for $\theta_{EXPR-} < r(g,s) < \theta_{EXPR+}$

where $DCN(g,s)$ is the DCN ratio for gene g in sample s, while $r(g,s)$ is the corresponding expression ratio (w.r.t. the normal samples).

Since for certain microarray spots expression ratios may be poorly defined (mainly due to low intensities in one of the two channels), we only retained genes whose expression ratios were well measured in more than half of the samples.

Unlike Bashyam et al. who performed mean centering of the (*log*-)expression ratios of the genes (to emphasize their *relative* levels among samples), we avoid mean-centering or variance normalization of the ratios since we are interested in identifying systematically mis-expressed genes, the expression level being important for this purpose.

3. A Simple Model of "Common Genes" and Recurrent DCN Alterations

Since we are interested in finding common molecular mechanisms involved in the studied phenotype as well as causal explanations for all 23 samples, we start with the simplest model and try to confront it with the available data. We first assume that most observed gene expression changes are direct or indirect consequences of DCN changes in various oncogenes and tumor suppressor genes (since no data about point mutations, e.g. SNPs, or promoter hypermethylation are available).

However, not all amplified genes are over-expressed and not all deleted genes are under-expressed. Moreover, even if the expression of a gene is changed as the result of DCN alterations, the gene may not be involved in the phenotype under study.

In the following, we assume that all samples with a given phenotype (pancreatic cancer in our case) are perturbed instances of a given molecular mechanism, with the perturbations affecting possibly distinct genes and pathways. These "*entry pathways*" can only be distinct if they *all* affect a "common mechanism", that is thus systematically perturbed in all samples and explains the observed phenotype. (Note that the "common genes" may include not only the genes directly or indirectly responsible for the disease, but also side-effects. The latter may also be very useful as markers – for example the pancreatic cancer marker CA 19-9.)

In the following we will use the *gene expression data* to identify these "*common genes*" (which are systematically changed in all samples). The *DCN data* will then be employed to determine *gene copy number alterations in the individual samples that might explain the observed changes in the common genes*.

As already mentioned in the Introduction, we are mainly interested in genes affected by recurrent DCN abnormalities, i.e. genes that not only have DCN abnormalities in certain samples, but also have expression levels that are well correlated with the phenotype over the majority of samples, thus representing potential therapeutic targets.

3.1. The "Common" Genes

We constructed two lists of "common" up- and respectively down-regulated genes:

$Common+ = \{ g \mid SNR(g) > 2$ and $EXPR+(g,s)$ over all well-measured samples s, except at most two samples[4] for which $EXPR^{(0)}(g,s) \}$

$Common- = \{ g \mid SNR(g) < -2$ and $EXPR-(g,s)$ over all well-measured samples s, except at most two samples for which $EXPR^{(0)}(g,s) \}$.

The common genes above are systematically up- or down-regulated in virtually all samples and are listed in Annex 1. There were 74 up-regulated and 179 down-regulated clones (corresponding to 48 and respectively 98 genes with names in the NCBI Gene database). Many of the genes with a known role in pancreatic cancer are among these. For example BCL2 is known to be involved in the specific resistance of pancreatic cancer cells to apoptotic stimuli (which constitutes one of the main reasons for its insensitivity to chemo-, radio- and immuno-therapy). Note that the down-regulation of BCL2 in pancreatic cancer is quite unusual for an anti-apoptotic gene, since it is normally over-expressed in other tumor types [7]. TGFB2 and TGFB3 are involved in the Transforming Growth Factor beta pathway and in pancreatic cancer. The ets variant gene 4 ETV4 appears as a key regulator of the differentiation/proliferation balance in pancreatic cancer cells [Pubmed:15461591]. PROX1 (the prospero-related homeobox 1) was observed to be significantly reduced in pancreatic cancer specimens from patients with short survival rates so that loss of Prox1 function may be a driving force behind pancreatic carcinoma progression [Pubmed:16525637].

However, the genes with a known role in pancreatic cancer make up only a minority of the "common" genes. Understanding the precise relationships between the rest, as well as the distinct mechanisms whereby the "common genes" are affected in each sample is a daunting task.

3.2. Genes with Recurrent DCN Alterations

Explaining the sample-specific changes leading to the observed expression changes in the common genes requires first the determination of the genes with DCN alterations that are matched by corresponding expression changes. More precisely, we construct two lists of genes with DCN amplifications and deletions respectively:

$DCN+ = \{ g \mid DCN+(g,s)$ and $EXPR+(g,s)$ in some sample $s \}$
$DCN- = \{ g \mid DCN-(g,s)$ and $EXPR-(g,s)$ in some sample $s \}$

Figure 1 shows the overlaps of the "*Common*±" and "*DCN*±" gene lists:

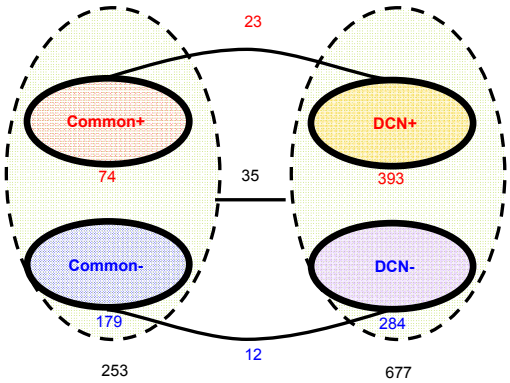

Figure 1. Overlaps of the common and DCN gene lists

The global picture of the genes involved in pancreatic cancer is even more complex if we look at both "common" and "DCN anomaly" genes. A few such genes with DCN anomalies such as SMAD4, BCL2, E2F1

[4] We took the maximum allowed number of samples with small expression changes ($EXPR^{(0)}$) to depend on the number of well measured samples as follows:
$n^{(0)} \le round(1.8 \cdot (n_{samples} - n_{well\ measured}) / n_{samples})$ and is 1 for 14 well-measured samples and respectively 2 for 23.

are well known, but the vast majority is still very poorly characterized. The main contribution of this paper consists in proposing a method for relating "common" genes to DCN alterations in the individual samples.

4. Relating Common Genes to DCN Alterations in the Individual Samples

We start by constructing a matrix of potential causes for the observed expression changes of the common genes in the individual samples. Such potential causes for a common gene G and a sample s are genes g with DCN alterations in s whose expression levels are well-correlated with those of G over the entire set of samples.[5]

More precisely, we recorded – for each common gene G and each sample s – the set of genes g with DCN alterations in s having the best N absolute correlations with G (a number of N=10 such genes were retained for each G and s): $Potential\_causes(G,s) = \{\ g \mid DCN+(g,s)$ and $EXPR+(g,s)$ or $DCN-(g,s)$ and $EXPR-(g,s)$ with the N=10 best absolute correlations $|r(G,g)|\ \}$ where $r(G,g) = \frac{\overline{G \cdot g}}{\sqrt{\overline{G^2} \cdot \overline{g^2}}}$ is the *uncentred* correlation of genes G and g over all samples. (As already discussed in Section 2, we are interested in the correlations of absolute (uncentred) gene expression ratios.)

4.1. The Rank Covering Algorithm

The matrix of potential causes is quite large and an aggregated view for each sample may be useful. More precisely, we may want to determine the main DCN alterations in each sample (responsible for the observed expression changes in the common genes). For this purpose, we use a so-called *rank-covering algorithm*, which we describe below.

In each sample s, the algorithm starts with the best correlated "DCN genes" g for each "common gene" G (g are called genes of rank 1). Then, the common genes G covered by each distinct g are determined and the genes g are sorted according to the number of common genes covered. Such a covering of "common genes" with "DCN genes" may serve as a potential link between DCN alterations and the common phenotype. However, due to the inherent noisy nature of microarray data, the correlations $|r(g_i,G)|$ of the genes g_i covering G may be quite close to one another (typical values are around ±0.9), thus rendering small rank differences meaningless.

We take this into account by considering the covers of "DCN genes" with progressively higher ranks k. More precisely, for an entry in the matrix of potential causes $Potential\_causes(G,s) = \{ g_1, g_2, \ldots, g_N \}$, the rank of DCN gene g_k is by definition k (since the genes g_k are sorted in a descending order w.r.t. their absolute correlations with G, $|r(g_k,G)|$).

Now, for a given k, we study the covers of "DCN genes" g with ranks at most k. Of course, such higher rank covers are larger and (hopefully) more robust to small differences in correlations $|r(g_k,G)|$.

A more detailed description of the algorithm is given below.

rank_covering (Potential_causes)
for ranks $k = 1,2,\ldots,N$
  for all samples s
    let $\Gamma_c \leftarrow Common$ (initially, the list of genes to be covered contains all "common" genes)
    while Γ_c is not empty
      determine the DCN-genes g with rank $\le k$ w.r.t. some common gene $G \in \Gamma_c$:
        $DCN\_genes \leftarrow \{\ g \mid g \in Potential\_causes(G,s)[1..k]$ for some $G \in \Gamma_c\ \}$
      determine the covers of the genes g from DCN_genes:
        $cover_{new}(g) = \{\ G \in \Gamma_c \mid g \in Potential\_causes(G,s)[1..k]\ \}$
        $cover(g) = \{\ G \in Common \mid g \in Potential\_causes(G,s)[1..k]\ \}$
      set $rank\_covering(k,s) \leftarrow \{\ (g, |cover(g)|, |cover_{new}(g)|)\ \mid\ g \in$ DCN_genes, sorted in descending order of $|cover_{new}(g)|\ \}$
    end while

[5] Although a complex combinatorial control of G may involve other genes as well, thus leading to a poor correlation with g over all samples, we argue that such genes g are probably less interesting as potential therapeutic targets, since they seem to affect only a minority of cases. Moreover, the low statistical power due to the very limited number of samples would anyway not allow the inference of such a complex combinatorial control.

end for
end for
return *rank_covering*

Note that for each gene g we return both the total cover $|cover(g)|$, as well as the "new cover" $|cover_{new}(g)|$ (i.e. the number of common genes not covered by previously selected common genes *g'*).

One may ask why a more sophisticated rank covering algorithm is needed instead of simply looking for common genes that have DCN changes in some samples (i.e. considering the intersection *Common* ∩ *DCN*). Actually, considering just the overlap between *Common* and *DCN* would be overly restrictive, since many DCN alterations affect genes that are *not* systematically mis-expressed in *all* samples. (Moreover, many genes with known implication in cancer have well-measured ratios only for a minority of samples, so they would not qualify as "common" genes.)

The main contribution of this paper consists in emphasizing the difference between two distinct classes of genes: the "common" genes (related to the disease phenotype) and respectively the genes with DCN alterations, which play the role of potential causes of the observed changes in the common genes. Although these two classes have a significant overlap, looking at the overlap only would miss a large number of sample-specific DCN alterations.

Note that searching for DCN genes *g* that are correlated to common genes g_0 (using the uncentred correlation) is *not* equivalent to looking for common genes with DCN alterations in some samples. There are DCN genes that do not satisfy the strict fold-change criteria for "common" genes in (virtually) all samples, but which nevertheless are well correlated with other "common" genes (again using the uncentred correlation) – see Figure 2.

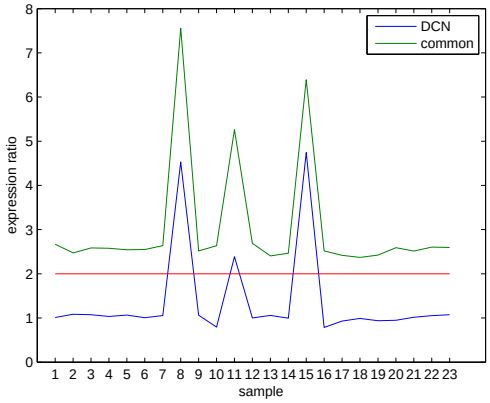

Figure 2. A DCN gene that is not a common gene may be correlated to a common gene (fold change threshold=2)

Furthermore, although all common genes are systematically mis-expressed (in virtually all samples), they are not necessarily highly correlated to each other (see Figure 3).

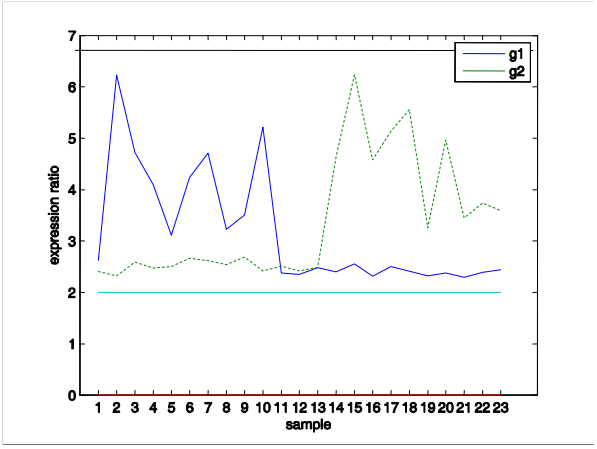

Figure 3. Common genes may not be highly correlated to each other

4.2. Results

Annex 2 presents the results of the rank-covering algorithm run on the pancreatic cancer dataset of Bashyam et al.[6] Note that in many samples the first 3 DCN genes are not "common", thus showing that a simple overlap between DCN and common genes would have missed certain important genes, with known involvement in pancreatic cancer (such as *SMAD4* or ETV4).

The following genes with known involvement in pancreatic cancer stand out: ***BCL2*** [7] (involved in apoptosis, with deletions in cell lines Aspc1, CFPAC1, Panc2.13, Panc8.13, PL4, SW1990, Mpanc-96, Colo-357), ***TCF4*** (deleted in PL4, Su8686, controls the expression of TCF7, a crucial transcription factor in the Wnt pathway), *SMAD4* (a key transcription factor in the TGF-beta pathway, also known under the name DPC4 – 'Deleted in Pancreatic Cancer 4' – is deleted in cell line Panc3.27), TNFSF10 (TRAIL – TNF-related apoptosis-inducing ligand, amplified in cell line Capan2), KRT19 (keratin 19, associated with malignant transformation in pancreatic cancer, was found amplified in CFPAC1, PL45), the ets-family transcription factor ETV4 (key regulator of the differentiation/proliferation balance in pancreatic cancer cells, found amplified in CAPAN1, CFPAC1, Panc2.03, Panc10.05).

The results also suggest potential roles in pancreatic cancer for the following genes: the microtubule associated homolog TPX2 (amplified in Hs766T, Panc1, SW1990 and known to be aberrantly expressed in colorectal cancer), the homeobox gene HOXB6 (amplified in Panc10.05, PL45), the tumor suppressor candidate 3, *TUSC3* (deleted in MIAPACA2, also mentioned in [1]), the topoisomerase II beta *TOP2B* (deleted in PL5, topoisomerase II beta is a substrate for PKC zeta, which significantly influences topoisomerase II inhibitor-induced cytotoxicity by altering its activity through its kinase function – this is meaningful in the context of pancreatic cancer since several PKC genes are altered, e.g. PRKCI and PRKCBP1), TOP2A (amplified in CAPAN1), PRKCI (protein kinase C, iota, an oncogene amplified in Panc1), UBE2C (ubiquitin-conjugating enzyme E2C, amplified in BxPC3, Su8686, is known to play an important role in tumor progression), the N-myristoyltransferase 1, NMT1 (amplified in PL45, is critical for tumor cell proliferation).

The results obtained are encouraging. However, they should be treated with caution due to certain limitations related mostly to the currently available experimental data:

- the small sample size (23)
- the very high noise of aCGH measurements (higher than for gene expression measurements)
- the low significance of the variance information for the expression data (only two normal samples are available)
- the lack of additional potentially relevant high throughput data regarding SNPs, promoter hypermethylation, miRNAs, etc.

5. Conclusions

Gene expression data are typically insufficient for finding the causes of observed gene expression changes in various types of cancer. Therefore, additional data such as those obtained via array-CGH are needed for a more complete picture of oncogenesis. On the other hand, DNA copy number data produced by array-CGH measurements is insufficient for determining potential therapeutic targets, as many genes with DCN alterations do not show significant expression changes. Moreover, combining DNA copy number data with gene expression data is highly non-trivial. Most biologically-oriented papers use human expert evaluations of the data without any automated means of combining the two. On the other hand, the few existing computationally-oriented studies (e.g. [9]) have mainly focused on determining the correlation between DNA copy numbers and gene expression, without a detailed analysis of a cancer dataset and *without linking isolated DCN changes with the phenotype*.

In this paper we address the problem of relating DCN alterations in individual samples with consistent expression changes in "common genes" across all disease samples. Our rank-covering algorithm can be

[6] For lack of space, only the first three "DCN genes" are shown for each rank and each sample. Also due to space limitations, only the total coverings of genes are shown. Note that certain genes do not have standard names in the NCBI Gene database. Furthermore, different clones of the same gene appear under the same name.

[7] In the following, genes in **boldface** are Common genes, while genes in *Italic* font are down-regulated in tumors. Thus, we use the following convention: **Common+**, ***Common–***, DCN+, *DCN–*.

used to obtain an aggregated view of the candidate genes with DCN alterations that potentially explain the observed phenotype. The algorithm recovered a set of genes with well-known roles in pancreatic cancer, but also suggested an additional set of genes, which seem to be involved in cancer, but whose exact role in pancreatic cancer remains to be determined experimentally.

Currently it is believed that pancreatic oncogenesis involves the development of genetic instability followed by clonal expansion of the most favorable genetic aberrations. The wide variety of alterations observed not only in the Bashyam dataset, but also in a parallel study by Heidenblad et al. [3,4] suggest vastly different evolutionary paths for the various samples, thus making the separation between the neutral aberrations and the tumorigenic ones extremely difficult. Combining DCN with gene expression data may be very useful in such a context.

The simple approach presented here is applicable to other datasets as well (such as those produced by SNP arrays, methylation arrays, etc.) and will be very useful for analyzing the increasing number of DCN and SNP datasets which will become available in the near future.

REFERENCES

1. BASHYAM, M.D. et al., **Array-Based Comparative Genomic Hybridization Identifies Localized DNA Amplifications and Homozygous Deletions in Pancreatic Cancer**. Neoplasia. 2005 Jun.; 7(6):556-62.
2. http://genome.ucsc.edu/
3. HEIDENBLAD, M. et al., **Genome-wide Array-based Comparative Genomic Hybridization Reveals Multiple Amplification Targets and Novel Homozygous Deletions in Pancreatic Carcinoma Cell Lines**. Cancer Res. 2004 64(9):3052-9.
4. HEIDENBLAD, M. et al., **Microarray Analyses Reveal Strong Influence of DNA Copy Number Alterations on the Transcriptional Patterns in Pancreatic Cancer: Implications for the Interpretation of Genomic Amplifications.** Oncogene. 2005 Mar. 3;24(10):1794-801.
5. POLLACK, JR. et al., **Microarray Analysis Reveals a Major Direct Role of DNA Copy Number Alteration in the Transcriptional Program of Human Breast Tumors**. Proc. Natl. Acad. Sci. U.S.A. 2002 Oct 1;99(20):12963-8.
6. WANG, P., KIM, Y., POLLACK, J., NARASIMHAN, B., TIBSHIRANI, R., **A Method for Calling Gains and Losses in Array CGH Data.** Biostatistics. 2005 Jan.; 6(1):45-58.
7. WESTPHAL, S., KALTHOFF, H., **Apoptosis: Targets in Pancreatic Cancer.** Mol Cancer. 2003 Jan. 7;2: 6. Review.
8. **Stratagene – Pathway Architect tool**: http://www.stratagene.com/products/showProduct.aspx?pid=733
9. LIPSON, D., et al., **Joint Analysis of DNA Copy Numbers and Gene Expression Levels,** Proc. of Algorithms in Bioinformatics: 4th International Workshop, WABI 2004, Bergen, Norway, September 17-21, 2004, Lecture Notes in Computer Science (LNCS), Vol. 3240/2004, p. 135, Springer 2004.
10. **The Stanford Microarray Database**. http://genome-www5.stanford.edu
11. BHATTACHARJEE et al., **Classification of Human Lung Carcinomas by mRNA Expression Profiling Reveals Distinct Adenocarcinoma Subclasses**. Proc. Natl. Acad. Sci. USA. 2001 Nov. 20;98 (24):13790-5.
12. PLATZER, P. et al., **Silence of Chromosomal Amplifications in Colon Cancer.** Cancer Res. 2002 Feb. 15; 62(4):1134-8.

Annex 1. Common genes

Common+

ABI2	Abl interactor 2
ATAD2	ATPase family, AAA domain containing 2
BUB3	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)
C14orf131	**Chromosome 14 open reading frame 131
CD47	**CD47 antigen (Rh-related antigen, integrin-associated signal transducer)
CDK7	**Cyclin-dependent kinase 7 (MO15 homolog, Xenopus laevis, cdk-activating kinase)
CENPF	**Centromere protein F, 350/400ka (mitosin)
CKS1B	CDC28 protein kinase regulatory subunit 1B
CNAP1	Chromosome condensation-related SMC-associated protein 1
CNIH	Cornichon homolog (Drosophila)
COPB2	**Coatomer protein complex, subunit beta 2 (beta prime)
DHFR	Dihydrofolate reductase
DLG7	Discs, large homolog 7 (Drosophila)
EFNA5	Nuclear RNA-binding protein, putative
ETV4	Ets variant gene 4 (E1A enhancer binding protein, E1AF)
FER1L3	Fer-1-like 3, myoferlin (C. elegans)
FLNB	Filamin B, beta (actin binding protein 278)
GPCR5A	G protein-coupled receptor, family C, group 5, member A
GPR110	G protein-coupled receptor 110
HBEGF	Heparin-binding EGF-like growth factor
HMGA1	High mobility group AT-hook 1
HMMR	**Hyaluronan-mediated motility receptor (RHAMM)
IQGAP3	IQ motif containing GTPase activating protein 3
ITGB4	Integrin, beta 4
KIAA0101	KIAA0101
KRT19	Keratin 19
MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)
MATP	Solute carrier family 45, member 2
MGC5395	AHNAK nucleoprotein (desmoyokin)
MPHOSPH1	M-phase phosphoprotein 1
NQO1	NAD(P)H dehydrogenase, quinone 1
PLCB4	Phospholipase C, beta 4
PLK1	Polo-like kinase 1 (Drosophila)
PPARG	Peroxisome proliferative activated receptor, gamma
PTTG3	**Pituitary tumor-transforming 3
RAD21	RAD21 homolog (S. pombe)
RIF1	Chromosome 1 open reading frame 103
SCD	**Stearoyl-CoA desaturase (delta-9-desaturase)
SLC16A1	AKR7 family pseudogene
SLC2A1	Solute carrier family 2 (facilitated glucose transporter), member 1
STK4	Serine/threonine kinase 4
TGFB2	Transforming growth factor, beta 2
TOP2A	Topoisomerase (DNA) II alpha 170kDa
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)
TRIP13	Thyroid hormone receptor interactor 13
UBE2C	Ubiquitin-conjugating enzyme E2C
YWHAB	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide

Common– [8]

ABCA5	**ATP-binding cassette, sub-family A (ABC1), member 5
ACP2	Acid phosphatase 2, lysosomal
ADH1B	Alcohol dehydrogenase IB (class I), beta polypeptide
AKAP13	A kinase (PRKA) anchor protein 13
ALB	Albumin
APBB1IP	Amyloid beta (A4) precursor protein-binding, family B, member 1 interacting protein
APCS	Amyloid P component, serum
APOC2	Apolipoprotein C-II
APOH	Apolipoprotein H (beta-2-glycoprotein I)
ARHGEF6	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6
BCL2	B-cell CLL/lymphoma 2
C10orf38	Chromosome 10 open reading frame 38
C1QA	Complement component 1, q subcomponent, alpha polypeptide
C1S	Complement component 1, s subcomponent
C2	Complement component 2
C5	Complement component 5
CABC1	Chaperone, ABC1 activity of bc1 complex like (S. pombe)
CALCRL	Calcitonin receptor-like
CCL2	Chemokine (C-C motif) ligand 2
CD200	CD200 antigen
CD3G	CD3G antigen, gamma polypeptide (TiT3 complex)
CD53	CD53 antigen
CG018	Hypothetical gene CG018
COL1A1	Collagen, type I, alpha 1
COL1A2	Collagen, type I, alpha 2
COL3A1	Collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)
COL6A3	Collagen, type VI, alpha 3
COX7B	**Cytochrome c oxidase subunit VIIb
CRYAB	Crystallin, alpha B
CX3CR1	Chemokine (C-X3-C motif) receptor 1
CYBB	Cytochrome b-245, beta polypeptide (chronic granulomatous disease)
DKFZP564J102	DKFZP564J102 protein
DOCK2	Dedicator of cytokinesis 2
DPPA4	Developmental pluripotency associated 4
DZIP1	DAZ interacting protein 1
ELA3B	Elastase 3B, pancreatic
EML1	Echinoderm microtubule associated protein like 1
EPB41L4B	Erythrocyte membrane protein band 4.1 like 4B
FAP	Fibroblast activation protein, alpha
FGFR1	Fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)
FGG	Fibrinogen gamma chain
FGL1	Fibrinogen-like 1
FKBP11	FK506 binding protein 11, 19 kDa
FLJ20152	Hypothetical protein FLJ20152
FLJ20699	Hypothetical protein FLJ20699
FLT1	Fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
FOS	V-fos FBJ murine osteosarcoma viral oncogene homolog
GPNMB	Glycoprotein (transmembrane) nmb
HNRPU	**Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)
IGFBP7	Insulin-like growth factor binding protein 7
IGLC2	Immunoglobulin lambda constant 1 (Mcg marker)
IL1R1	Interleukin 1 receptor, type I
INSR	Insulin receptor
ISG20	Interferon stimulated exonuclease gene 20kDa
ITGB8	Integrin, beta 8
JUN	V-jun sarcoma virus 17 oncogene homolog (avian)
KIAA0146	KIAA0146 protein
KLK2	Kallikrein 2, prostatic
KLK3	Kallikrein 3, (prostate specific antigen)
LOC285513	**Hypothetical protein LOC285513
LPL	Lipoprotein lipase
MAF	V-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)
MAN1C1	Mannosidase, alpha, class 1C, member 1
MGC9850	Polymerase (RNA) I polypeptide D, 16kDa
MKNK1	MAP kinase interacting serine/threonine kinase 1
NPTX2	Neuronal pentraxin II
PAH	Phenylalanine hydroxylase
PDCD4	Programmed cell death 4 (neoplastic transformation inhibitor)
PECAM1	Platelet/endothelial cell adhesion molecule (CD31 antigen)
PGA5	Porin, putative
PHF17	PHD finger protein 17
PLA2G2A	Phospholipase A2, group IIA (platelets, synovial fluid)
POSTN	Periostin, osteoblast specific factor
PROX1	Prospero-related homeobox 1
PTPRC	Protein tyrosine phosphatase, receptor type, C
PTPRN	Protein tyrosine phosphatase, receptor type, N
RAG1	Recombination activating gene 1
REG1A	Regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
RRBP1	Ribosome binding protein 1 homolog 180kDa (dog)
SCG2	Secretogranin II (chromogranin C)
SEMA6A	Sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6A
SERPINA3	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3
SERPINA5	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5
SERPINA7	**Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7
SLC4A3	Solute carrier family 4, anion exchanger, member 3
SNX19	**Sorting nexin 19
SOD2	Superoxide dismutase 2, mitochondrial
TCF4	Transcription factor 4
TF	Transferrin
TGFB3	Transforming growth factor, beta 3
TncRNA	Trophoblast-derived noncoding RNA
TTR	Transthyretin (prealbumin, amyloidosis type I)
UBD	Ubiquitin D
UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4
UGT2B7	**UDP glucuronosyltransferase 2 family, polypeptide B7
WNT5A	**Wingless-type MMTV integration site family, member 5A
XBP1	X-box binding protein 1
ZFHX1B	Zinc finger homeobox 1b

Annex 2. The results of the rank covering algorithm

Cell line \ k	1	2	3	4	5	6	7	8	9
Aspc1	PLEC1 [152] ***BCL2*** [49] *COL6A1* [17]	PLEC1 [170] ***BCL2*** [71] SHFM1 [15]	PLEC1 [174] ***BCL2*** [170] SHFM1 [40]	***BCL2*** [220] PLEC1 [175] PLEC1 [174]	***BCL2*** [245] PLEC1 [191] ***BCL2*** [220]	***BCL2*** [248] PLEC1 [193] ***BCL2*** [245]	***BCL2*** [248] PLEC1 [194] ***BCL2*** [248]	***BCL2*** [251] PLEC1 [195] ***BCL2*** [248]	***BCL2*** [252] PLEC1 [214] ***BCL2*** [251]
BxPC3	STK4 [130] *AI338371* [54] UBE2C [50]	STK4 [188] UBE2C [137] *AI338371* [118]	UBE2C [227] STK4 [216] *AI338371* [197]	STK4 [250] UBE2C [248] *AI338371* [240]	STK4 [252] UBE2C [250] *AI338371* [248]	STK4 [253] UBE2C [251] *AI338371* [249]	STK4 [253] *AI338371* [252] UBE2C [251]	STK4 [253] UBE2C [253] *AI338371* [252]	STK4 [253] UBE2C [253] *AI338371* [252]
CAPAN1	ETV4 [97] TOP2A [76] *WNK3* [28]	ETV4 [165] TOP2A [142] LIPH [9]	ETV4 [193] TOP2A [188] LIPH [16]	TOP2A [215] ETV4 [205] FLNB [91]	TOP2A [234] ETV4 [214] FLNB [120]	TOP2A [240] ETV4 [218] FLNB [130]	TOP2A [243] FLNB [138] ETV4 [226]	TOP2A [247] FLNB [147] ETV4 [231]	TOP2A [249] FLNB [155] ETV4 [238]
Capan2	FNDC3B [234] EVA1 [8] FLJ21827 [7]	FNDC3B [245] EVA1 [111] TNFSF10 [87]	FNDC3B [250] TNFSF10 [133] EVA1 [126]	FNDC3B [250] TNFSF10 [167] EVA1 [176]	FNDC3B [250] AA485418 [8] EVA1 [185]	FNDC3B [253] EVA1 [212] TNFSF10 [206]	FNDC3B [253] EVA1 [219] FLJ21827 [213]	FNDC3B [253] EVA1 [225] FLJ21827 [225]	FNDC3B [253] EVA1 [232] FLJ21827 [232]
CFPAC1	ETV4 [188] ***BCL2*** [17] KRT19 [16]	ETV4 [207] ***BCL2*** [132] KRT19 [21]	ETV4 [212] ***BCL2*** [173] ***BCL2*** [36]	ETV4 [214] ***BCL2*** [192] ***BCL2*** [44]	ETV4 [220] ***BCL2*** [178] ***BCL2*** [46]	ETV4 [222] ***BCL2*** [198] ***BCL2*** [48]	ETV4 [222] ***BCL2*** [207] ***BCL2*** [49]	ETV4 [224] ***BCL2*** [212] ***BCL2*** [50]	ETV4 [234] ***BCL2*** [214] ***BCL2*** [60]
HPAC	MATP [114] ***SERPINA7*** [53] TRIP13 [44]	MATP [179] TRIP13 [71] *DUSP22* [33]	MATP [217] *DUSP22* [56] TRIP13 [106]	MATP [221] *DUSP22* [80] ***SERPINA7*** [197]	MATP [230] *DUSP22* [121] ***SERPINA7*** [201]	MATP [238] *DUSP22* [173] ***SERPINA7*** [203]	MATP [242] *DUSP22* [211] ***SERPINA7*** [210]	MATP [248] *DUSP22* [228] ***SERPINA7*** [218]	MATP [250] *DUSP22* [243] *DUSP22* [228]
HPAFII	***LPL*** [154] ***LPL*** [86] MNAT1 [7]	***LPL*** [244] ***LPL*** [203] MNAT1 [26]	***LPL*** [252] ***LPL*** [212] MNAT1 [140]	***LPL*** [253] ***LPL*** [216] MNAT1 [208]	***LPL*** [253] MNAT1 [238] ***LPL*** [217]	***LPL*** [253] MNAT1 [250] SHFM1 [222]	***LPL*** [253] MNAT1 [251] SHFM1 [237]	***LPL*** [253] MNAT1 [253] SHFM1 [246]	***LPL*** [253] MNAT1 [253] SHFM1 [252]
Hs766T	TPX2 [120] YWHAB [45] YWHAB [21]	TPX2 [146] YWHAB [62] YWHAB [52]	TPX2 [151] YWHAB [91] YWHAB [80]	TPX2 [160] YWHAB [108] YWHAB [92]	TPX2 [170] YWHAB [142] STK4 [73]	TPX2 [174] YWHAB [165] STK4 [84]	YWHAB [189] TPX2 [181] STK4 [89]	YWHAB [205] TPX2 [183] YWHAB [145]	YWHAB [234] TPX2 [192] YWHAB [167]
MIAPACA2	*TUSC3* [137] *TUSC3* [63] *AA878541* [29]	*TUSC3* [205] *TUSC3* [78] *TUSC3* [21]	*TUSC3* [216] *TUSC3* [35] *AA878541* [39]	*TUSC3* [223] *TUSC3* [42] *TUSC3* [216]	*TUSC3* [228] *TUSC3* [73] *TUSC3* [223]	*TUSC3* [239] *TUSC3* [205] *TUSC3* [228]	*TUSC3* [249] *TUSC3* [241] *TUSC3* [239]	*TUSC3* [253] *TUSC3* [251] *TUSC3* [249]	*TUSC3* [253] *TUSC3* [253] *TUSC3* [252]
Panc1	TPX2 [132] ***LPL*** [69] *DEFA1* [23]	TPX2 [208] ***LPL*** [147] *DEFA1* [73]	TPX2 [243] ***LPL*** [225] PRKCI [56]	TPX2 [249] ***LPL*** [244] *DEFA1* [230]	***LPL*** [253] TPX2 [251] *DEFA1* [237]	***LPL*** [253] TPX2 [251] *DEFA1* [245]	***LPL*** [253] TPX2 [253] *DEFA1* [249]	*DEFA1* [253] ***LPL*** [253] TPX2 [253]	*DEFA1* [253] ***LPL*** [253] TPX2 [253]
Panc 2.03	ETV4 [125] FAM49B [21] ATAD2 [19]	ETV4 [173] ATAD2 [30] ***BCL2*** [19]	ETV4 [183] ATAD2 [41] ***BCL2*** [20]	ETV4 [194] ATAD2 [55] ***BCL2*** [31]	ETV4 [205] ATAD2 [58] ***BCL2*** [36]	ETV4 [221] FAM49B [117] ATAD2 [63]	ETV4 [228] FAM49B [138] ATAD2 [75]	ETV4 [233] FAM49B [151] ATAD2 [93]	ETV4 [236] ATAD2 [106] FAM49B [184]
Panc2.13	***BCL2*** [154] ***BCL2*** [40] *DUSP22* [35]	***BCL2*** [163] ***BCL2*** [41] *DUSP22* [39]	***BCL2*** [181] *DUSP22* [46] ***BCL2*** [42]	***BCL2*** [181] *DUSP22* [47] ***BCL2*** [42]	***BCL2*** [207] *DUSP22* [87] ***BCL2*** [187]	***BCL2*** [212] *DUSP22* [90] ***BCL2*** [187]	***BCL2*** [219] *DUSP22* [91] ***BCL2*** [213]	***BCL2*** [222] *DUSP22* [91] ***BCL2*** [219]	***BCL2*** [224] *DUSP22* [139] ***BCL2*** [222]
Panc8.13	RAD21 [72] MGC5395 [69] KTN1 [14]	RAD21 [112] MGC5395 [104] CDKN3 [22]	RAD21 [144] MGC5395 [124] CDKN3 [33]	RAD21 [162] ALKBH [48] ***BCL2*** [61]	RAD21 [171] ***BCL2*** [41] ALKBH [51]	RAD21 [177] ***BCL2*** [65] ALKBH [60]	RAD21 [188] ***BCL2*** [115] ALKBH [67]	RAD21 [193] ***BCL2*** [125] ALKBH [71]	RAD21 [198] ***BCL2*** [133] ***BCL2*** [39]
Panc 10.05	ETV4 [181] HOXB6 [15] *p44S10* [9]	ETV4 [216] *GUCY1A3* [17] HOXB6 [62]	ETV4 [227] KRT19 [24] PVT1 [108]	ETV4 [237] KRT19 [27] *PPAP2B* [31]	ETV4 [238] KRT19 [43] *PPAP2B* [40]	ETV4 [242] KRT19 [66] *PPAP2B* [46]	ETV4 [243] KRT19 [86] *PPAP2B* [60]	ETV4 [247] HOXB6 [197] PVT1 [213]	ETV4 [253] PVT1 [219] HOXB6 [201]
Panc 3.27	LIPH [128] PDLIM7 [44] KIAA0286 [43]	LIPH [154] KIAA0286 [71] *SMAD4* [41]	LIPH [159] KIAA0286 [84] *SMAD4* [67]	LIPH [167] KIAA0286 [86] *SMAD4* [84]	LIPH [172] *SMAD4* [126] PDLIM7 [161]	PDLIM7 [205] *SMAD4* [181] LIPH [180]	PDLIM7 [223] KIAA0286 [160] *SMAD4* [219]	*SMAD4* [236] KIAA0286 [200] PDLIM7 [225]	*SMAD4* [246] KIAA0286 [220] PDLIM7 [226]
PL4	***LPL*** [160] ***BCL2*** [29] ***TCF4*** [25]	***LPL*** [196] ***BCL2*** [88] ***BCL2*** [15]	***LPL*** [201] ***BCL2*** [163] ***BCL2*** [33]	***LPL*** [202] ***BCL2*** [184] ***BCL2*** [42]	***LPL*** [208] ***BCL2*** [169] ***BCL2*** [47]	***LPL*** [211] ***BCL2*** [190] ***BCL2*** [47]	***LPL*** [212] ***BCL2*** [209] ***BCL2*** [48]	***LPL*** [212] ***BCL2*** [211] ***BCL2*** [50]	***LPL*** [229] ***BCL2*** [213] ***BCL2*** [60]
PL5	*MOBKL2B* [226] *RBMS3* [13] *TOP2B* [10]	*MOBKL2B* [231] *RBMS3* [62] *TOP2B* [195]	*TOP2B* [249] *MOBKL2B* [231] *RBMS3* [223]	*HPCL2* [253] *TOP2B* [253] *RBMS3* [253]	*HPCL2* [253] *TOP2B* [253] *RBMS3* [253]	*HPCL2* [253] *TOP2B* [253] *RBMS3* [253]	*HPCL2* [253] *TOP2B* [253] *RBMS3* [253]	*HPCL2* [253] *TOP2B* [253] *RBMS3* [253]	*HPCL2* [253] *TOP2B* [253] *RBMS3* [253]
PL8	FAM49B [139] PVT1 [95] SIAT4A [15]	PVT1 [223] FAM49B [219] SIAT4A [54]	PVT1 [251] FAM49B [249] SIAT4A [237]	PVT1 [253] SIAT4A [252] FAM49B [251]	PVT1 [253] FAM49B [253] SIAT4A [253]	PVT1 [253] FAM49B [253] SIAT4A [253]	PVT1 [253] FAM49B [253] SIAT4A [253]	PVT1 [253] FAM49B [253] SIAT4A [253]	PVT1 [253] FAM49B [253] SIAT4A [253]
PL45	HOXB6 [103] PHB [52] HOXB5 [43]	HOXB6 [136] PHB [97] HOXB5 [85]	HOXB6 [151] PHB [133] HOXB5 [121]	HOXB6 [167] PHB [148] KRT19 [84]	HOXB6 [179] NMT1 [78] KRT19 [98]	HOXB6 [195] NMT1 [105] KRT19 [118]	HOXB6 [201] KRT19 [145] PHB [174]	HOXB5 [206] PHB [177] KRT19 [192]	KRT19 [215] HOXB6 [211] HOXB5 [214]
Su8686	UBE2C [155] ***TCF4*** [73] *SLC35B3* [10]	UBE2C [233] *SLC35B3* [41] ***TCF4*** [154]	UBE2C [248] ***TCF4*** [202] *SLC35B3* [123]	UBE2C [252] ***TCF4*** [237] *SLC35B3* [202]	UBE2C [252] ***TCF4*** [250] *SLC35B3* [226]	UBE2C [253] ***TCF4*** [251] *SLC35B3* [246]	UBE2C [253] *SLC35B3* [252] ***TCF4*** [252]	*SLC35B3* [253] UBE2C [253] ***TCF4*** [252]	*SLC35B3* [253] UBE2C [253] ***TCF4*** [252]
SW1990	TPX2 [153] STK4 [40] ***BCL2*** [17]	TPX2 [171] DPM1 [34] ***BCL2*** [45]	TPX2 [186] STK4 [130] ***BCL2*** [75]	TPX2 [210] ***BCL2*** [85] ***BCL2*** [14]	TPX2 [217] ***BCL2*** [81] ***BCL2*** [16]	TPX2 [219] ***BCL2*** [91] ***BCL2*** [28]	TPX2 [219] ***BCL2*** [150] ***BCL2*** [33]	TPX2 [221] ***BCL2*** [169] ***BCL2*** [40]	TPX2 [230] ***BCL2*** [56] STK4 [192]
Mpanc-96	*MI-ER1* [144] ***BCL2*** [73] ***BCL2*** [11]	***BCL2*** [178] *MI-ER1* [153] ***BCL2*** [36]	***BCL2*** [199] ***BCL2*** [42] ***BCL2*** [6]	***BCL2*** [199] ***BCL2*** [48] ***BCL2*** [6]	***BCL2*** [205] ***BCL2*** [48] ***BCL2*** [201]	***BCL2*** [205] ***BCL2*** [49] ***BCL2*** [205]	***BCL2*** [211] ***BCL2*** [50] ***BCL2*** [207]	***BCL2*** [213] ***BCL2*** [50] ***BCL2*** [211]	*MI-ER1* [218] ***BCL2*** [214] ***BCL2*** [64]
Colo-357	***BCL2*** [83] TRIP13 [45] AHSG [41]	***BCL2*** [112] AHSG [85] ***BCL2*** [31]	***BCL2*** [140] AHSG [93] ***BCL2*** [32]	***BCL2*** [159] AHSG [103] ***BCL2*** [35]	***BCL2*** [177] AHSG [115] ***BCL2*** [38]	***BCL2*** [193] ***BCL2*** [42] AHSG [120]	***BCL2*** [208] ***BCL2*** [46] AHSG [122]	***BCL2*** [210] ***BCL2*** [49] AHSG [124]	***BCL2*** [215] ***BCL2*** [69] AHSG [150]