Supplemental material S3: Integrative mixture of experts to combine clinical factors and gene markers

Biological relevance of Prostate, Breast and CNS data sets

We further analyzed the biological relevance of the selected p^* genes that were integrated with the clinical factors in our integrative ME model (see Supplemental Tables S4-6). The most relevant genes for each data set are presented in Table 1 and are discussed below.

Additional information on prostate data regarding the sPLS gene selection. Furthermore sPLS methods selected interesting genes amongst them STIP1, OGN and ITGB4 could be of relative importance. Stress-induced phosphoprotein 1 (STIP1) is overexpressed in hepatocellular carcinoma (Sun *et al.*, 2007) and involved in pancreatic cancer invasion (Walsh *et al.*, 2009). The osteoglycin (OGN) is underexpressed in colorectal adenoma and cancer and has recently been proposed as a potential biomarker for colorectal cancer detection (Wang *et al.*, 2007). Integrin beta 4 (ITGB4) is a member of integrin proteins mediating cell adhesion. Its downregulation is critical to the ability of tumor cells to disseminate (Howard *et al.*, 2008).

Breast data. The genes selected by the different proposed approaches are related to cell cycle, metabolism, and tumor invasion. Some of these genes have already been integrated in the poor prognosis signature in breast cancer (van't Veer *et al.*, 2002) and other should be carefully considered due to their roles. Amongst them, protein regulator of cytokinesis 1 (PRC1) is one the most frequently selected gene. This gene is necessary for cytokinesis and highly expressed in breast cancer tissues and, on the contrary, barely detectable in any normal human tissues. Downregulation of PRC1 by treatment with siRNA suppresses the growth of breast cancer cells, indicating that it could be a promising molecular target of breast cancer (Shimo *et al.*, 2007). The Insulin Growth factor binding protein 5 (IGFBP5) is a member of the IGFBP family that binds to insulin-like growth factor IGF. IGFBP5 has been shown to either stimulate cell growth (Salih *et al.*, 2004), migration (Abrass *et al.*, 1997) or cell attachment to extracellular matrix (McCaig *et al.*, 2002). However, numerous studies of breast cancer tissues demonstrated that IGFBP5 mRNA level was significantly upregulated and was associated with metastasis (Nishidate *et al.*, 2004) and a poor prognosis (van't Veer *et al.*, 2002; Li *et al.*, 2007). Phosphoglycerate kinase1 (PGK1) is not only an ATP-generating enzyme of the glycolytic pathway, but is also a multifunctional molecule. For example, PGK1 can affect DNA replication and repair in mammalian cell nuclei (Popanda *et al.*, 1998) and is secreted by tumor cells to control the angiogenesis process (Lay *et al.*, 2000). Recently, PGK1 has been found overexpressed in a variety

Table 1: Most relevant selected genes with a potential biomarker status. Expression level in subjects with respect to class 'recurrent' or 'dead' is indicated: overexpressed (+), underexpressed (-).

	Gene Name	Symbol	Level	Gene selection method [rank]	Link to cancer
Breast	Insulin-like growth factor binding protein 5	IGFBP5	+	*t*-test[1,3], RF[5,8,13], sPLS[1,3]	Nishidate *et al.* (2004); van't Veer *et al.* (2002); Li *et al.* (2007); Mita *et al.* (2007)
	Phosphoglycerate mutase 1	PGK1	+	*t*-test[2], RF[11], sPLS[2]	Duan *et al.* (2002); Hwang *et al.* (2006); Zhang *et al.* (2005); Zieker *et al.* (2008)
	Protein regulator of cytokinesis 1	PRC1	+	*t*-test[5], RF[12], sPLS[5]	Shimo *et al.* (2007)
	E2F transcription factor 1	E2F1	+	*t*-test[13]	Zhang *et al.* (2000); Vuaroqueaux *et al.* (2007)
	Adrenomedullin	ADM	+	RF[6]	Oehler *et al.* (2003)
CNS	High mobility group AT-hook 1	HMGA1	+	*t*-test[2], RF[8], sPLS[2]	Liau *et al.* (2008)
	V-myb myeloblastosis viral oncogene homolog (avian)-like 2	MYBL2	+	*t*-test[6]	Raschella *et al.* (1999)
	Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	+	RF[2]	Maraqa *et al.* (2008)
	Ras homolog gene family, member C	RhoC	+	sPLS[3]	Boone *et al.* (2009)
	Heat shock 70kDa protein 9	HSPA9	+	RF[4]	Dundas *et al.* (2005)

of cancers, such as renal, pancreatic carcinoma and multi-drug resistant ovarian cancer (Duan *et al.*, 2002; Hwang *et al.*, 2006). It is also considered as a marker for peritoneal dissemination in gastric cancer (Zieker *et al.*, 2008), as well as a potential target in highly invasive HER-2/neu breast cancer (Zhang *et al.*, 2005). E2F transcription factor 1 (E2F1) plays a crucial role in the control of cell cycle and action of tumor suppressor proteins. Overexpression of E2F-1 mRNA was observed in 40% of gastric carcinomas and in 60% of colorectal adenocarcinomas. Zhang *et al.* (2000) discovered that E2F-1 expression could be involved in breast tumor progression. Recently, E2F-1 transcript level has been identified as a strong predictor of breast cancer outcome (Vuaroqueaux *et al.*, 2007). Finally, adrenomedullin (ADM) is synthesized and is secreted from many mammalian tissues. It has been implicated as a mediator of several diseases such as cardiovascular and renal disorders, sepsis, inflammation, diabetes and cancer. Increased expression of ADM peptide has been reported in numerous tumors and is associated with poor overall outcome in breast cancer (Oehler *et al.*, 2003).

CNS data. Combining genes profiles with clinical data could help identify and define some important markers for medulloblastoma outcome. Numerous genes not yet associated with medulloblastoma outcome have been selected. Those genes are related to metabolism, signaling pathways and oxidative stress.
The high mobility group HMGA1 protein play a role in chromosomal organization and gene transcription regulation. HMGA1 overexpression represents a common feature of human malignant tumors including thyroid, breast, ovary and prostate. Recently, HMGA1 expression has been suggested to promote genomic instability (Takaha *et al.*, 2002). HMGA1 promotes tumorigenicity and is an independent predictor of poor postoperative survival in patients with pancreatic adenocarcinoma (Liau *et al.*, 2008). Carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) is an adhesion molecule, which is overexpressed in a number of human malignancies including pancreatic cancer, gastrointestinal cancer and breast cancers (Scholzel *et al.*, 2000; Blumenthal *et al.*, 2007). Recently, Maraqa *et al.* (2008) proposed CEACAM6 as a powerful predictor of future recurrence in breast cancer. Ras homologous (Rho) family GTPases play a pivotal role in the regulation of numerous cellular functions associated with malignant transformation and metastasis. RhoC, which was selected by sPLS(3b) is overexpressed in carcinoma tissues and high RhoC expression is correlated with tumor progression and metastasis in numerous cancer (Wang *et al.*, 2009; Boone *et al.*, 2009). MYBL2 is a nuclear protein involved in cell cycle progression, cell proliferation, differentiation, and survival. Its overexpression has already been reported in cell lung carcinomas (Hibi *et al.*, 1998). Raschella *et al.* (1999) demonstrated that its expression is a useful prognostic marker in human neuroblastoma. HSPA9 is a member of the heat-shock protein 70 family involved in cell cycle regulation with important roles in cellular senescence and immortalization pathways. HSPA9 expression level has been found to be correlated with a poor survival in colorectal adenocarcinomas (Dundas *et al.*, 2005) and with early liver tumors recurrence (Yi *et al.*, 2008).

References

Abrass, C., Berfield, A., and Andress, D. (1997). Heparin binding domain of insulin-like growth factor binding protein-5 stimulates mesangial cell migration. *AJP-Renal Physiology*, **273**(6), 899.

Blumenthal, R., Leon, E., Hansen, H., and Goldenberg, D. (2007). Expression patterns of CEACAM 5 and CEACAM 6 in primary and metastatic cancers. *BMC cancer*, **7**(1), 2.

Boone, B., Van Gele, M., Lambert, J., Haspeslagh, M., and Brochez, L. (2009). The role of RhoC in growth and metastatic capacity of melanoma. *Journal of Cutaneous Pathology*, **36**(6), 629–636.

Duan, Z., Lamendola, E., Yusuf, R., Penson, R., Preffer, F., and Seiden, M. (2002). Overexpression of human phosphoglycerate kinase 1 (PGK1) induces a multidrug resistance phenotype. *Anticancer research*, **22**(4), 1933–1942.

Dundas, S., Lawrie, L., Rooney, P., and Murray, G. (2005). Mortalin is over-expressed by colorectal adenocarcinomas and correlates with poor survival. *The Journal of Pathology*, **205**(1).

Hibi, K., Liu, Q., Beaudry, G., Madden, S., Westra, W., Wehage, S., Yang, S., Heitmiller, R., Bertelsen, A., Sidransky, D., *et al.* (1998). Serial analysis of gene expression in non-small cell lung cancer. *Cancer Research*, **58**(24), 5690–5694.

Howard, E., Leung, S., Yuen, H., Chua, C., Lee, D., Chan, K., Wang, X., and Wong, Y. (2008). Decreased adhesiveness, resistance to anoikis and suppression of GRP94 are integral to the survival of circulating tumor cells in prostate cancer. *Clinical and Experimental Metastasis*, **25**(5), 497–508.

Hwang, T., Liang, Y., Chien, K., and Yu, J. (2006). Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. *Proteomics*, **6**(7).

Lay, A., Jiang, X., Kisker, O., Flynn, E., Underwood, A., Condron, R., and Hogg, P. (2000). Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature*, **408**(6814), 869–873.

Li, X., Cao, X., Li, X., Zhang, W., and Feng, Y. (2007). Expression level of insulin-like growth factor binding protein 5 mRNA is a prognostic factor for breast cancer. *Cancer Science*, **98**(10), 1592–1596.

Liau, S., Rocha, F., Matros, E., Redston, M., and Whang, E. (2008). High mobility group AT-hook 1 (HMGA1) is an independent prognostic factor and novel therapeutic target in pancreatic adenocarcinoma. *Cancer*, **113**(2), 302.

Maraqa, L., Cummings, M., Peter, M., Shaaban, A., Horgan, K., Hanby, A., and Speirs, V. (2008). Carcinoembryonic antigen cell adhesion molecule 6 predicts breast cancer recurrence following adjuvant tamoxifen. *Clinical Cancer Research*, **14**(2), 405.

McCaig, C., Perks, C., and Holly, J. (2002). Intrinsic actions of IGFBP-3 and IGFBP-5 on Hs578T breast cancer epithelial cells: inhibition or accentuation of attachment and survival is dependent upon the presence of fibronectin. *Journal of cell science*, **115**(22), 4293–4303.

Mita, K., Zhang, Z., Ando, Y., Toyama, T., Hamaguchi, M., Kobayashi, S., Hayashi, S., Fujii, Y., Iwase, H., and Yamashita, H. (2007). Prognostic significance of insulin-like growth factor binding protein (IGFBP)-4 and IGFBP-5 expression in breast cancer. *Japanese Journal of Clinical Oncology*, **37**(8), 575.

Nishidate, T., Katagiri, T., Lin, M., Mano, Y., Miki, Y., Kasumi, F., Yoshimoto, M., Tsunoda, T., Hirata, K., and Nakamura, Y. (2004). Genome-wide gene-expression profiles of breast-cancer cells purified with laser microbeam microdissection: identification of genes associated with progression and metastasis. *International journal of oncology*, **25**(4), 797.

Oehler, M., Fischer, D., Orlowska-Volk, M., Herrle, F., Kieback, D., Rees, M., and Bicknell, R. (2003). Tissue and plasma expression of the angiogenic peptide adrenomedullin in breast cancer. *British journal of cancer*, **89**(10), 1927–1933.

Popanda, O., Fox, G., and Thielmann, H. (1998). Modulation of DNA polymerases α, δ and ε by lactate dehydrogenase and 3-phosphoglycerate kinase. *BBA-Gene Structure and Expression*, **1397**(1), 102–117.

Raschella, G., Cesi, V., Amendola, R., Negroni, A., Tanno, B., Altavista, P., Paolo Tonini, G., De Bernardi, B., and Calabretta, B. (1999). Expression of B-myb in Neuroblastoma Tumors Is a Poor Prognostic Factor Independent from MYCN Amplification 1. *Cancer Research*, **59**(14), 3365–3368.

Salih, D., Tripathi, G., Holding, C., Szestak, T., Gonzalez, M., Carter, E., Cobb, L., Eisemann, J., and Pell, J. (2004). Insulin-like growth factor-binding protein 5 (Igfbp5) compromises survival, growth, muscle development, and fertility in mice. *Proceedings of the National Academy of Sciences*, **101**(12), 4314–4319.

Scholzel, S., Zimmermann, W., Schwarzkopf, G., Grunert, F., Rogaczewski, B., and Thompson, J. (2000). Carcinoembryonic antigen family members CEACAM6 and CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas. *American Journal of Pathology*, **156**(2), 595–605.

Shimo, A., Nishidate, T., Ohta, T., Fukuda, M., Nakamura, Y., and Katagiri, T. (2007). Elevated expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells. *Cancer science*, **98**(2), 174–181.

Sun, W., Xing, B., Sun, Y., Du, X., Lu, M., Hao, C., Lu, Z., Mi, W., Wu, S., Wei, H., *et al.* (2007). Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis: novel protein markers in hepatocellular carcinoma tissues. *Molecular & Cellular Proteomics*, **6**(10), 1798.

Takaha, N., Hawkins, A., Griffin, C., Isaacs, W., and Coffey, D. (2002). High Mobility Group Protein I (Y) A Candidate Architectural Protein for Chromosomal Rearrangements in Prostate Cancer Cells 1. *Cancer Research*, **62**(3), 647–651.

van't Veer, L., Dai, H., Van de Vijver, M., He, Y., Hart, A., Mao, M., Peterse, H., Van der Kooy, K., Marton, M., Witteveen, A., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**(6871), 530.

Vuaroqueaux, V., Urban, P., Labuhn, M., Delorenzi, M., Wirapati, P., Benz, C., Flury, R., Dieterich, H., Spyratos, F., Eppenberger, U., *et al.* (2007). Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome. *Breast Cancer Res*, **9**(3), R33.

Walsh, N., O'Donovan, N., Kennedy, S., Henry, M., Meleady, P., Clynes, M., and Dowling, P. (2009). Identification of pancreatic cancer invasion-related proteins by proteomic analysis. *Proteome Science*, **7**(1), 3.

Wang, H., Liu, X., Liang, J., Yang, K., Sui, A., and Liu, Y. (2009). Expression of RhoA and RhoC in colorectal carcinoma and its relations with clinicopathological parameters. *Clinical Chemistry and Laboratory Medicine*, **47**(7), 811–817.

Wang, Y., Ma, Y., Lu, B., Xu, E., Huang, Q., and Lai, M. (2007). Differential Expression of Mimecan and Thioredoxin Domain Containing Protein 5 in Colorectal Adenoma and Cancer: A Proteomic Study. *Experimental Biology and Medicine*, **232**(9), 1152.

Yi, X., Luk, J., Lee, N., Peng, J., Leng, X., Guan, X., Lau, G., Beretta, L., and Fan, S. (2008). Association of mortalin (HSPA9) with liver cancer metastasis and prediction for early tumor recurrence. *Molecular & Cellular Proteomics*, **7**(2), 315.

Zhang, D., Tai, L., Wong, L., Chiu, L., Sethi, S., and Koay, E. (2005). Proteomic Study Reveals That Proteins Involved in Metabolic and Detoxification Pathways Are Highly Expressed in HER-2/neu-positive Breast Cancer. *Molecular & Cellular Proteomics*, **4**(11), 1686–1696.

Zhang, S., Liu, S., Al-Saleem, L., Holloran, D., Babb, J., Guo, X., and Klein-Szanto, A. (2000). E2F-1: A Proliferative Marker of Breast Neoplasia 1. *Cancer Epidemiology Biomarkers & Prevention*, **9**(4), 395–401.

Zieker, D., Konigsrainer, I., Traub, F., Nieselt, K., Knapp, B., Schillinger, C., Stirnkorb, C., Fend, F., Northoff, H., Kupka, S., *et al.* (2008). PGK1 a potential marker for peritoneal dissemination in gastric cancer. *Cell Physiol Biochem*, **21**(5-6), 429–436.