**Appendix**

**Additional simulation study**

In the simulation presented in Section 4.1, all three datasets have the same error distribution. Here we consider additional simulation where the three datasets have error distributions as , , . That is, the errors have different distributions with different variances. Other settings are the same as in Section 4.1. Summary statistics based on 100 replicates are shown in Table 5 and 6. We can see that although the detailed numerical results are different from Table 1 and 2, the overall patterns are similar. The SGMCP outperforms alternatives with more true positives and fewer false positives.

**Table 5**. Simulation under the homogeneity model. Three datasets have different error distributions. In each cell, the first/second row is the mean number (sd) of true/false positives. When , MCP simplifies to Lasso.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MCP | | | |
| Correlation |  |  |  |  |
| AR | 21.98(4.50) | 20.76(4.68) | 22.24(3.89) | 13.88(4.50) |
| 51.58(67.62) | 51.28(51.80) | 58.74(22.95) | 85.78(47.58) |
| AR | 14.74(2.71) | 16.16(2.29) | 17.96(2.30) | 21.64(1.75) |
| 7.94(10.00) | 16.00(8.02) | 41.26(15.79) | 72.66(25.83) |
| Banded 1 | 21.20(3.98) | 20.82(4.04) | 20.28(3.63) | 15.60(3.60) |
| 42.66(59.47) | 41.98(38.27) | 54.16(25.07) | 70.08(34.88) |
| Banded 2 | 15.80(2.89) | 16.86(2.67) | 17.96(2.12) | 19.64(1.97) |
| 10.10(10.45) | 23.54(22.32) | 46.02(12.15) | 74.40(23.49) |
|  | GMCP | | | |
|  |  |  |  |  |
| AR | 29.58(1.49) | 29.64(2.15) | 29.76(1.70) | 28.92(3.19) |
| 10.38(5.66) | 35.10(12.41) | 92.88(15.31) | 206.04(86.22) |
| AR | 16.92(4.14) | 17.34(5.08) | 19.32(4.78) | 26.22(1.46) |
| 7.68(6.04) | 30.06(17.12) | 44.04(41.16) | 78.30(38.48) |
| Banded 1 | 28.56(3.84) | 29.70(1.09) | 29.82(1.27) | 28.44(2.52) |
| 11.70(6.62) | 39.54(12.04) | 96.90(16.00) | 178.08(82.09) |
| Banded 2 | 17.28(4.44) | 17.88(4.53) | 20.64(3.19) | 25.14(2.18) |
| 8.10(7.89) | 27.54(16.35) | 47.04(35.19) | 98.10(63.39) |
|  | SGMCP | | | |
|  |  |  |  |  |
| AR | 25.86(4.04) | 27.42(2.74) | 28.32(1.56) | 26.52(3.88) |
| 5.22(5.55) | 6.58(6.22) | 22.78(9.94) | 130.62(54.28) |
| AR | 14.72(2.90) | 14.14(2.76) | 15.66(4.08) | 25.86(1.20) |
| 1.78(2.64) | 3.12(3.73) | 18.70(10.19) | 64.58(38.84) |
| Banded 1 | 24.74(4.28) | 24.52(5.98) | 25.70(4.50) | 25.10(3.31) |
| 4.30(4.91) | 8.42(7.15) | 22.80(6.87) | 112.68(64.44) |
| Banded 2 | 15.06(3.36) | 15.18(3.70) | 17.84(4.40) | 24.00(1.95) |
| 1.24(2.54) | 4.70(5.58) | 20.14(13.13) | 72.10(38.83) |

**Table 6**. Simulation under the heterogeneity model. Three datasets have different error distributions. In each cell, the first/second row is the mean number (sd) of true/false positives. When , MCP simplifies to Lasso.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MCP | | | |
| Correlation |  |  |  |  |
| AR | 21.50(5.30) | 20.66(4.14) | 21.54(3.87) | 14.28(4.33) |
| 54.98(67.00) | 46.30(40.56) | 56.56(20.98) | 78.50(35.85) |
| AR | 14.34(2.73) | 15.80(2.29) | 18.02(1.97) | 22.14(1.55) |
| 8.20(6.79) | 19.28(10.06) | 42.12(12.56) | 70.10(25.23) |
| Banded 1 | 20.36(4.47) | 21.34(3.95) | 20.28(3.78) | 14.84(4.25) |
| 49.74(79.79) | 50.10(43.14) | 54.26(24.47) | 77.32(38.09) |
| Banded 2 | 15.22(2.48) | 17.34(2.68) | 18.90(2.05) | 20.48(2.24) |
| 14.62(37.09) | 23.12(23.66) | 51.40(12.47) | 77.50(26.30) |
|  | GMCP | | | |
|  |  |  |  |  |
| AR | 22.26(6.45) | 22.98(5.53) | 25.14(3.04) | 21.00(6.49) |
| 58.80(79.92) | 84.72(74.51) | 121.86(27.01) | 198.54(110.10) |
| AR | 12.08(3.75) | 12.82(3.55) | 15.46(2.43) | 24.34(1.84) |
| 20.56(9.32) | 41.54(20.30) | 65.24(30.53) | 161.36(77.54) |
| Banded 1 | 21.98(5.67) | 22.54(5.50) | 24.26(3.94) | 18.60(6.97) |
| 53.08(70.92) | 66.98(38.96) | 118.12(18.16) | 136.20(95.57) |
| Banded 2 | 13.62(4.80) | 14.48(4.39) | 15.74(2.72) | 22.40(2.66) |
| 23.88(13.86) | 47.32(19.10) | 87.52(29.54) | 204.88(88.27) |
|  | SGMCP | | | |
|  |  |  |  |  |
| AR | 23.74(3.76) | 22.60(3.14) | 25.14(2.51) | 20.40(6.00) |
| 14.12(9.90) | 19.76(13.12) | 38.14(10.85) | 136.10(85.80) |
| AR | 13.02(3.85) | 13.30(3.94) | 14.10(4.39) | 24.10(1.69) |
| 4.84(4.39) | 9.40(7.94) | 26.00(14.97) | 96.54(46.49) |
| Banded 1 | 20.44(5.42) | 21.72(5.63) | 23.08(4.11) | 20.24(3.66) |
| 11.66(8.28) | 19.46(11.42) | 37.46(12.12) | 120.20(57.42) |
| Banded 2 | 14.42(4.23) | 13.06(4.29) | 14.10(4.06) | 22.02(2.41) |
| 7.02(6.52) | 9.20(7.34) | 27.12(13.69) | 133.22(55.01) |

**Additional data analysis results using SGMCP**

For genes identified using SGMCP (Table 3 and 4), we mine published literature and find that they may have important implications.

Among genes identified in the breast cancer study, gene LFNG (Hs.159142) is a member of the fringe gene family which also includes radical and manic fringe genes. Xu et al. (2012) reported that LFNG, which suppresses Jagged/Notch signaling in vivo, is consistently expressed at a low level in basal-like tumors and deletion of this gene in the mouse mammary gland enhances accumulation of activated Notch intracellular domain polypeptides, increases proliferation, and induces basal-like mammary tumors in cooperation with amplification of the Met/Caveolin gene locus. Gene TNPO1 (Hs.168075) encodes the beta subunit of the karyopherin receptor complex which interacts with nuclear localization signals to target nuclear proteins to the nucleus. TNPO1 was significantly down-regulated in non-muscle-invasive tumors with later progression (Dyrskjøt et al. 2009). Gene PRUNE2 (Hs.23311) may play an important role in regulating differentiation, survival and aggressiveness of the tumor cells. Gene AGPAT1 (Hs.240534) encodes an enzyme that converts lysophosphatidic acid (LPA) into phosphatidic acid (PA). The protein encoded by gene RAD50 (Hs.41587) is highly similar to Saccharomyces cerevisiae Rad50, a protein involved in DNA double-strand break repair. Mutations in RAD50 are associated with doubling of breast cancer risks (Walsh and King 2007). Ras-homologous GTPases (RRAGB, Hs.50282) constitute a large family of signal transducers that alternate between an activated, GTP-binding state and an inactivated, GDP-binding state. Three different forms of human pancreatic procarboxypeptidase A (CPA3, Hs.646) have been isolated. This gene encodes a form which is obtained as a binary complex of a procarboxypeptidase A with proproteinase E and functions as a secretory granule metalloexopeptidase. CPA3 is up-regulated in mammospheres of ER-positive breast cancer cells (Kok et al. 2009). Gene NAGA (Hs.75372) encodes the lysosomal enzyme alpha-N-acetylgalactosaminidase, which cleaves alpha-N-acetylgalactosaminyl moieties from glycoconjugates. Immunotherapy by treatment with DBP-MAF has been successfully proposed in several tumors, such as prostate cancer, metastatic breast cancer and colorectal cancer, for which serum NAGA activity has been measured as prognostic index of this kind of therapy (Yamamoto et al. 2008). The protein encoded by Hs.75890 (MBTPS1) has a central role in the regulation of lipid metabolism in cells. The cumulative effect of changes in the expression of MBTPS1 is blockade of estrogen-induced breast cancer cell growth (Liu et al. 2002). Annexin I (ANXA1, Hs.78225) belongs to a family of Ca(2+)-dependent phospholipid binding proteins which have a molecular weight of approximately 35,000 to 40,000 and are preferentially located on the cytosolic face of the plasma membrane. Shen et al. (2006) suggest that suppressed ANXA1 expression in breast tissue is correlated with breast cancer development and progression. The product of gene CLCN7 (Hs.80768) belongs to the CLC chloride channel family of proteins. CLCN7 has been identified to be associated with breast cancer (Ma et al. 2011). Gene BUB1 (Hs.98658) encodes a kinase involved in spindle checkpoint function. Regulation of expression of this gene may be important in cancer (Myrie et al. 2000).

In the lung cancer study, defensins (gene DEFA6) are a family of microbicidal and cytotoxic peptides thought to be involved in host defense. Proteasome PSMB7 is a multicatalytic proteinase complex with a highly ordered ring-shaped 20S core structure. Auto-antibodies against PSMB7 was detected in sera from 8/21 (38.1%) patients with lung cancer (Gao et al. 2005). Gene STAM encodes a member of the signal-transducing adaptor molecule family. Gene CXCL3 has chemotactic activity for neutrophils. It may play a role in inflammation and exert its effects on endothelial cells in an autocrine fashion. Gene LGALS8 encodes a member of the galectin family. Galectins are beta-galactoside-binding animal lectins with conserved carbohydrate recognition domains. It has been identified to be associated with survival in lung cancer (Sun et al. 2008). Gene ADH6 encodes class V alcohol dehydrogenase, which is a member of the alcohol dehydrogenase family. It has been identified to be associated with lung cancer in Lee et al. (2011). The protein encoded by gene PPFIBP1 is a member of the LAR protein-tyrosine phosphatase-interacting protein (liprin) family. Mutations in gene NBN are associated with Nijmegen breakage syndrome, an autosomal recessive chromosomal instability syndrome characterized by microcephaly, growth retardation, immunodeficiency, and cancer predisposition. It has been found that Germ-line variants in the NBS1 gene may play a role in the lung carcinogenesis in cigarette smokers (Medina et al. 2003). PAH encodes the enzyme phenylalanine hydroxylase that is the rate-limiting step in phenylalanine catabolism. Hecht (1999) states that while nicotine itself is not considered to be carcinogenic, each cigarette contains a mixture of carcinogens, including a small dose of polycyclic aromatic hydrocarbons (PAHs) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) among other lung carcinogens, tumor promoters, and co-carcinogens.

**References**

Dyrskjøt L, Ostenfeld MS, Bramsen JB, Silahtaroglu AN, Lamy P, Ramanathan R, Fristrup N, Jensen JL, Andersen CL, Zieger K, Kauppinen S, Ulhøi BP, Kjems J, Borre M, Orntoft TF. (2009) Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. Cancer research. 69 (11): 4851-4860.

Gao WM, Puravs E, Zhao R, Deng B, Wang H, Qiu J, Kuick R, Hinderer R, Misek DE, Omenn GS, Hanash SM. (2005) Identification of autoantibodies to tumor-derived antigens as candidate biomarkers for the early detection of lung cancer. Proc Amer Assoc Cancer Res, Volume 46. Abstract #4363.

Hecht SS. (1999) Tobacco smoke carcinogens and lung cancer. Journal of the National Cancer Institute. 91(14): 1194-1210.

Kok M, Koornstra RH, Margarido TC, Fles R, Armstrong NJ, Linn SC, Van't Veer LJ, Weigelt B. (2009) Mammosphere‐derived gene set predicts outcome in patients with ER‐positive breast cancer. The Journal of Pathology. 218(3): 316-326.

Lee IH, Lushington GH, Visvanathan M. (2011) A filter-based feature selection approach for identifying potential biomarkers for lung cancer. J Clin Bioinforma. 1(1): 11.

Liu Y, Ludes-Meyers J, Zhang Y, Munoz-Medellin D, Kim HT, Lu C, Ge G, Schiff R, Hilsenbeck SG, Osborne CK, Brown PH. (2002) Inhibition of AP-1 transcription factor causes blockade of multiple signal transduction pathways and inhibits breast cancer growth. Oncogene. 21(50): 7680-7689.

Ma S, Huang J, Wei F, Xie Y, Fang K. (2011) Integrative analysis of multiple cancer prognosis studies with gene expression measurements. Statistics in Medicine. 30(28): 3361-3371.

Medina PP, Ahrendt SA, Pollan M, Fernandez P, Sidransky D, Sanchez-Cespedes M. (2003) Screening of homologous recombination gene polymorphisms in lung cancer patients reveals an association of the NBS1-185Gln variant and p53 gene mutations. Cancer Epidemiology Biomarkers & Prevention. 12(8): 699-704.

Myrie KA, Percy MJ, Azim JN, Neeley CK, Petty EM. (2000) Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines. Cancer Letters. 152(2): 193-199.

Shen D, Nooraie F, Elshimali Y, Lonsberry V, He J, Bose S, Chia D, Seligson D, Chang HR, Goodglick L. (2006) Decreased expression of annexin A1 is correlated with breast cancer development and progression as determined by a tissue microarray analysis. Human Pathology. 37(12): 1583-1591.

Sun Z, Wigle DA, Yang P. (2008) Non-overlapping and non-cell-type-specific gene expression signatures predict lung cancer survival. Journal of Clinical Oncology. 26(6): 877-883.

Walsh T. and King MC. (2007) Ten genes for inherited breast cancer. Cancer cell. 11(2): 103-105.

Xu K, Usary J, Kousis PC, Prat A, Wang DY, Adams JR, Wang W, Loch AJ, Deng T, Zhao W, Cardiff RD, Yoon K, Gaiano N, Ling V, Beyene J, Zacksenhaus E, Gridley T, Leong WL, Guidos CJ, Perou CM, Egan SE. (2012) Lunatic fringe deficiency cooperates with the Met/Caveolin gene amplicon to induce basal-like breast cancer. Cancer cell. 21(5): 626-641.

Yamamoto N, Suyama H, Yamamoto N, Ushijima N. (2008) Immunotherapy of metastatic breast cancer patients with vitamin D‐binding protein‐derived macrophage activating factor (GcMAF). International Journal of Cancer. 122(2): 461-467.

**Data analysis using alternative methods**

**Table 7**. Analysis of breast cancer data using MCP: identified genes and their estimates*.*

|  |  |  |  |
| --- | --- | --- | --- |
| UniGene | D1 | D2 | D3 |
| Hs.104916 | |  | 0.0024 |
| Hs.105751 | |  | 0.0004 |
| Hs.110707 | 0.0031 |  |  |
| Hs.13776 | 0.0018 |  |  |
| Hs.155935 | | 0.0044 |  |
| Hs.159142 | | 0.0135 |  |
| Hs.166204 | 0.0089 |  |  |
| Hs.168075 | 0.0374 |  |  |
| Hs.20013 |  |  | 0.0006 |
| Hs.20478 |  |  | 0.0088 |
| Hs.23311 |  | 0.0051 |  |
| Hs.240534 | |  | -0.0026 |
| Hs.271986 | | 0.0008 |  |
| Hs.281564 | | 0.0049 |  |
| Hs.315177 | | 0.0011 |  |
| Hs.374988 | -0.0004 |  |  |
| Hs.379018 | 0.0030 |  |  |
| Hs.406186 | 0.0034 |  |  |
| Hs.41587 |  |  | 0.0004 |
| Hs.5344 |  |  | -0.0023 |
| Hs.646 |  | 0.0062 |  |
| Hs.75890 |  | 0.0250 |  |
| Hs.78225 | -0.0390 |  |  |
| Hs.80768 |  | 0.0195 |  |
| Hs.81892 |  | -0.0039 |  |
| Hs.89404 |  | 0.0016 |  |
| Hs.94395 |  |  | -0.0168 |
| Hs.979 |  |  | 0.0068 |

**Table 8**. Analysis of breast cancer data using GMCP: identified genes and their estimates*.*

|  |  |  |  |
| --- | --- | --- | --- |
| UniGene | D1 | D2 | D3 |
| Hs.110707 | 0.0086 | 0.0032 | 0.0009 |
| Hs.153687 | 0.0004 | 0.0004 | 1.54E-05 |
| Hs.239189 | -0.0047 | 0.0083 | -0.0065 |
| Hs.25363 | 0.0080 | 0.0041 | 0.0005 |
| Hs.290070 | -0.0036 | 0.0003 | -0.0003 |
| Hs.301654 | 0.0010 | 0.0010 | 0.0002 |
| Hs.355722 | 0.0008 | 0.0002 | -0.0006 |
| Hs.41587 | 0.0016 | 0.0013 | 0.0015 |
| Hs.75890 | -0.0002 | 0.0006 | -0.0001 |
| Hs.78225 | -0.0029 | 0.0005 | 0.0003 |
| Hs.78802 | 0.0081 | -0.0003 | -0.0012 |

**Table 9**. Analysis of breast cancer data using SGLasso: identified genes and their estimates*.*

|  |  |  |  |
| --- | --- | --- | --- |
| UniGene | D1 | D2 | D3 |
| Hs.110707 | 0.0065 | 0.0013 |  |
| Hs.146406 | 0.0021 |  |  |
| Hs.153684 | -0.0015 |  |  |
| Hs.159142 | | 0.0042 | -0.0016 |
| Hs.166204 | 0.004 |  |  |
| Hs.168075 | 0.0086 |  |  |
| Hs.20478 |  | 0.0003 | 0.0004 |
| Hs.23311 |  | 0.0009 |  |
| Hs.239189 | -0.0016 | 0.0049 | -0.0043 |
| Hs.25363 | 0.0043 | 0.0021 |  |
| Hs.281564 | | 0.0025 | 0.0006 |
| Hs.290070 | -0.0053 |  |  |
| Hs.301654 | 0.0014 | 0.0015 |  |
| Hs.355722 | 0.0008 |  | -0.0007 |
| Hs.374988 | -0.0006 | -0.0001 | 0.0002 |
| Hs.379018 | 0.0021 |  |  |
| Hs.381152 | 0.0015 |  |  |
| Hs.406186 | 0.0020 | 0.0011 |  |
| Hs.41587 | 0.0014 | 0.0014 | 0.0022 |
| Hs.50282 | 0.0001 | 0.0001 |  |
| Hs.54697 | 0.0001 | 0.0001 |  |
| Hs.75890 | -0.0002 | 0.0071 |  |
| Hs.78225 | -0.0093 |  |  |
| Hs.78802 | 0.0063 |  |  |
| Hs.80768 |  | 0.0046 |  |
| Hs.82002 | -0.0006 |  | 0.0005 |
| Hs.91093 |  | 0.0003 |  |
| Hs.94395 |  | 0.0001 | -0.0007 |

**Table 10**. Analysis of lung cancer data using MCP: identified genes and their estimates*.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Probe | Gene | UM | HLM | CAN/DF |
| 201870\_at | TOMM34 | -0.002 |  |  |
| 202184\_s\_at | NUP133 | 0.005 |  |  |
| 203544\_s\_at | STAM |  | -0.012 |  |
| 206799\_at | SCGB1D2 | -0.004 |  |  |
| 207850\_at | CXCL3 |  | -0.008 |  |
| 208933\_s\_at | LGALS8 |  |  | 0.017 |
| 211823\_s\_at | PXN | -0.007 |  |  |
| 214261\_s\_at | ADH6 |  | -0.017 |  |
| 217299\_s\_at | NBN |  | -0.020 |  |
| 217583\_at | PAH |  | -0.018 |  |
| 218809\_at | PANK2 | -0.004 |  |  |
| 220037\_s\_at | LYVE1 | -0.001 |  |  |
| 220266\_s\_at | KLF4 | -0.006 |  |  |
| 222305\_at | HK2 | -0.0003 |  |  |

**Table 11**. Analysis of lung cancer data using GMCP: identified genes and their estimates.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Probe | Gene | UM | HLM | CAN/DF |
| 202758\_s\_at | RFXANK | 0.0004 | -0.002 | -0.0005 |
| 207814\_at | DEFA6 | -0.002 | -0.005 | -0.002 |
| 208933\_s\_at | LGALS8 | 0.008 | 0.008 | 0.011 |
| 208956\_x\_at | DUT | -0.0001 | -0.0004 | -0.0003 |
| 214261\_s\_at | ADH6 | -0.003 | -0.014 | -0.004 |
| 214374\_s\_at | PPFIBP1 | -0.0003 | -0.001 | -0.0003 |
| 217299\_s\_at | NBN | -0.001 | -0.008 | -0.002 |
| 217583\_at | PAH | -0.004 | -0.014 | -0.0007 |
| 219654\_at | PTPLA | -0.001 | -0.002 | -0.0005 |
| 220658\_s\_at | ARNTL2 | -0.002 | -0.003 | -0.003 |

**Table 12**. Analysis of lung cancer data using SGLasso: identified genes and their estimates*.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Probe | Gene | UM | HLM | CAN/DF |
| 200786\_at | PSMB7 |  | -0.0014 | -0.0018 |
| 201174\_s\_at | TERF2IP | 0.0002 | 0.0016 |  |
| 202758\_s\_at | RFXANK |  | -0.0044 |  |
| 203544\_s\_at | STAM |  | -0.0031 |  |
| 206799\_at | SCGB1D2 | -0.0011 |  | -0.0008 |
| 207814\_at | DEFA6 | -0.0005 | -0.0032 | -0.0009 |
| 207850\_at | CXCL3 |  | -0.0054 |  |
| 208025\_s\_at | HMGA2 | -0.0024 |  | -0.0028 |
| 208933\_s\_at | LGALS8 | 0.0039 | 0.0032 | 0.0073 |
| 208956\_x\_at | DUT | -0.0002 | -0.0016 | -0.0015 |
| 209932\_s\_at | DUT | -6.4E-05 | -0.0002 | -0.0002 |
| 211823\_s\_at | PXN | -0.0022 | 0.0008 |  |
| 212986\_s\_at | TLK2 | -0.0006 | -0.0024 |  |
| 213552\_at | GLCE |  | -0.001 |  |
| 214261\_s\_at | ADH6 |  | -0.0113 | -0.0017 |
| 214374\_s\_at | PPFIBP1 |  | -0.0047 |  |
| 217299\_s\_at | NBN |  | -0.0086 |  |
| 217583\_at | PAH | -0.001 | -0.0116 |  |
| 219654\_at | PTPLA | -0.0007 | -0.0035 |  |
| 220266\_s\_at | KLF4 | -3.6E-05 |  |  |
| 221022\_s\_at | PMFBP1 |  | 0.0005 | -0.0001 |
| 221740\_x\_at | LRRC37A2 | 0.0012 |  | 0.0008 |