Supplementary Document

Y-h. Taguchi, Department of Physics, Chuo University, Tokyo, Japan

Contents

1 Reason for the usage of control/treated instead of treated/control 2

2 Replacing binomial test with shuffle test 2

3 R codes for limma executions 2

4 Biological significance of selected genes: literature searches 3
   4.1 Genes identified by PCA based unsupervised FE when $N' = 1000$ 3
   4.1.1 CCR2 3
   4.1.2 LRRN3 3
   4.1.3 AHR 3
   4.1.4 LOX 3
   4.1.5 PRAMEL1 3
   4.1.6 CD53 3
   4.1.7 ITGAL 4
   4.1.8 SULT1C2 4
   4.1.9 FCGR2B 4
   4.1.10 ELOVL2 4
   4.1.11 PF4 4
   4.1.12 PDHA2 4
   4.1.13 MPO 4
   4.1.14 HAND2 4
   4.1.15 CCL3 4
   4.1.16 HBE2 5
   4.1.17 CMKLR1 5
   4.1.18 DBH 5
   4.1.19 KCNT1 5
   4.1.20 TAAR7B 5
   4.1.21 Fibrinogen beta chain 5
   4.1.22 BMP3 5
   4.1.23 ACTG2 5
   4.1.24 AQP2 5
   4.2 Genes identified by limma based FE 6
   4.2.1 qk 6
   4.2.2 TOP1 6
   4.2.3 Arhgef1 6
   4.2.4 TEAD2 6
   4.2.5 Sirt2 6
   4.2.6 gmfg 6
   4.2.7 alkhh6 6
   4.2.8 MCEE 6
   4.2.9 lbs11 6
   4.2.10 HSPBP1 6
   4.2.11 XRCC1 6
   4.3 Genes identified by SAM based FE, but not by PCA based unsupervised FE, when $N' = 1000$ 7
   4.3.1 MYL1 7
   4.3.2 SLC28A1 7
   4.3.3 PGAM2 7
   4.3.4 ALB 7
   4.3.5 SLC13A3 7
   4.3.6 TTR 7
1 Reason for the usage of control/treated instead of treated/control

One may wonder why the ratio control/treated instead of treated/control was used for gene expression, because the latter is more commonly used. The primary purpose of PCA based unsupervised FE is the identification of outliers. Thus, no matter how the ratio is defined, outliers are expected to remain outliers. If treated/control was employed, 44 genes instead of 48 genes are commonly selected between gene expression and promoter methylation if $N' = 1000$. Thus, the number of commonly selected genes did not change drastically. In addition, among 44 genes, 42 genes are selected when control/treated was employed. Thus, genes commonly selected between gene expression and promoter methylation did not alter dependent upon whether control/treated instead of treated/control was used for gene expression. Because more common genes indicate a more feasible methodology, we used ratio control/treated instead of treated/control.

2 Replacing binomial test with shuffle test

One may wonder if the usage of binomial test to estimate $P$-value for commonly selected genes is reasonable. To validate usage of the binomial test, we shuffled gene order within gene expression/promoter methylation profiles independently one thousand times and repeated the whole analyses with $N' = 1000$. We found that the frequency of commonly selected genes was greater than 48, which is the number of commonly selected genes in the original (unshuffled) data set, and was 36 among one thousand trials, which is fully coincident with a probability of 0.04 estimated by binomial test. This definitely supports the usage of binomial test.

3 R codes for limma executions

For gene expression (eight samples shown in Table 1 of main text), suppose that $x$ includes not a ratio but raw gene expression (rows are probes and columns are samples) with the first column of probe ids,

```r
require(Biobase)
require(limma)
gene_exp <- new("ExpressionSet",expr=data.matrix(log(x[,1])))
fData(gene_exp)[["gene_id"]]<- x[,1]
pData(gene_exp)[["sample_name"]]<- TS
TS <- factor(TS)
design <- model.matrix(~0+TS)
colnames(design) <- levels(TS)
fit <- lmFit(gene_exp, design)
cont.matrix <- makeContrasts(Diff=(E16.VIN-E16.CNTL)-(E13.VIN-E16.CNTL),levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)
TT <- topTable(fit2,number=27342)
```

For promoter methylation (six samples shown in Table 1 of main text), suppose that $x_m$ includes promoter methylation (rows are probes and columns are samples) with the first column of probe ids,

```r
xm[is.na(xm)] <- 1
gene_exp <- new("ExpressionSet",expr=data.matrix(log(xm[,1])))
fData(gene_exp)[["gene_id"]]<- xm[,1]
pData(gene_exp)[["sample_name"]]<- c(rep("E13",3),rep("E16",3))
design <- cbind(WT=1,class=c(0,0,0,1,1,1))
fit <- lmFit(gene_exp, design)
```
4 Biological significance of selected genes: literature searches

4.1 Genes identified by PCA based unsupervised FE when $N' = 1000$

4.1.1 CCR2

CC chemokine receptor type-2 (CCR2) is a member of the G-protein coupled receptor superfamily, and is expressed on the cell surface of monocytes and macrophages. It binds to monocyte chemoattractant protein-1, a CC chemokine, produced at the sites of inflammation and infection. Based on their roles in disease, they are attractive targets for the pharmaceutical industry, and thus, targeting both CCR2 and CCR5 can be a useful strategy [1]. Blockade of CCR2 ameliorates progressive fibrosis in the kidney [2]. TR4 nuclear receptor promotes prostate cancer metastasis via the upregulation of CCL2/CCR2 signaling [3]. CCR2 is related to cancer immunotherapy [4].

4.1.2 LRRN3

3D structures of some LRR protein–ligand complexes show that the concave surface of the LRR domain is ideal for interaction with $\alpha$-helix, indicating that the elongated and curved LRR structure provides a framework for achieving diverse protein–protein interactions [5]. LRRN3 exhibits aberrant expression in neuroblastoma [6]. LRRN3 was suggested to be a novel, clinically relevant biomarker of immune status in HIV-1 infection [7].

4.1.3 AHR

AHR is a ligand-activated transcription factor that controls the expression of a diverse set of genes [8]. One of its known ligands is dioxin, a highly toxic environmental pollutant that is similar to vinclozolin. Vinclozolin was shown to be a weak inducer of AHR [9], which was suggested to be a therapeutic target for chronic kidney disease [10]. AHR has critical roles in sperm development [11]. AHR-mediated genotoxic effects were observed in a prostate cancer cell line [12]. AHR was previously reported to target multiple genes suggested to participate in non-small-cell lung cancer metastasis [13]. AHR also plays critical roles in immunology [14].

4.1.4 LOX

Lysyl oxidase (LOX) is a copper-dependent amine oxidase with a critical role in the biogenesis of connective tissue matrices by crosslinking the extracellular matrix proteins, collagen and elastin. LOX was upregulated in kidney abscesses [15] and affects prostate cancer growth [16]. LOX knockout mice develop sexual development problems including reduced production of sperm [17]. LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis [18], and LOX plays critical roles in mesenchymal stem cell-driven breast cancer malignancy [19]. LOX is generally recognized to be a critical factor of tumor genesis and is targeted for cancer treatment [20].

4.1.5 PRAMEL1

PRAMEL1 is a member of the melanoma antigen preferentially expressed in tumors (PRAME) gene family. The PRAME gene family encodes LRR proteins that function as transcription regulators in cancer cells [21]. The only normal tissues that expressed PRAME proteins were the testes [22]. The aberrant methylation of PRAMEL1 was observed in Dnmt1 overexpressing prostate tissue of mice [23].

4.1.6 CD53

The interaction of mesangial cells with the extracellular matrix plays a major role in kidney biology. CD53 is present in mesangial cells in vivo and in culture [24]. Natural killer (NK) cells are important contributors to the early immune defense against infected or transformed cells. The tetraspanin CD53 modulates responses from activating NK cell receptors, promoting lymphocyte function-associated antigen activation and damping NK cell effector functions [25]. Polymorphonuclear neutrophils in systemic lupus erythematosus, an autoimmune disease, exhibit the enhanced expression of CD53 [26]. CD53, together with carcinoembryonic antigen-related cell adhesion molecule 1, was targeted by early B cell factor (EBF)-1 [27].
4.1.7 ITGAL
ITGAL, also known as CD11A, is one of six genes that is used to successfully discriminate patients with castration-resistant prostate cancer into two risk groups [28]. ITGAL together with CD18 involved in the leukocyte adhesion pathway plays a role in mediating ischemic acute renal failure in rats [29]. CD11A was shown to be absent in testis tumors [30] but was essential for inflammatory and immune responses [31]. ITGAL was previously also reported to be associated with another F3 generation vinclozolin lineage [32].

4.1.8 SULT1C2
SULT1C2 was shown to be abundant in the kidney but was absent in a model of kidney disease [33].

4.1.9 FCGR2B
FCGR2B and FCRLB gene polymorphisms are associated with IgA nephropathy (kidney disease) [34]. FCGR2B expression on B cells and dendritic cells is important for the mucosal induction of antigen-specific immune tolerance [35].

4.1.10 ELOVL2
ELOVL2 controls the level of n-6 28:5 and 30:5 fatty acids in the testes [36]. EVOVL2 was previously reported to be associated with another F3 generation vinclozolin lineage [37].

4.1.11 PF4
Platelet factor 4 (PF4), also known as chemokine (C-X-C motif) ligand 4 (CXCL4), is a small cytokine belonging to the CXC chemokine family. PF4 has a protective role in chronic kidney allografts [38]. Pf4/Cxcl4 was upregulated in the testes of 5-day old rats of 2,3,7,8-tetrachlorodibenzo-p-dioxin-treated dams [39]. PF4 progresses prostate cancer [40] but down-regulates CC chemokine receptors, including CCR2 in human monocytes [41].

4.1.12 PDHA2
PDHA2 is one subunit of the pyruvate dehydrogenase complex [42] and is testis specific gene [43]. The activation of PDHA2 gene expression during spermatogenesis is thought to ensure the continued expression of the protein, thus allowing germ cell viability and functionality [44]. The selection of PDHA2 in this study that integrated gene expression and promoter methylation is convincing because the expression of PDHA2 is mediated by promoter methylation [45].

4.1.13 MPO
Myeloperoxidase (MPO) is a peroxidase enzyme related to renal diseases [46, 47], upregulated in ischemia-reperfusion testes [48], expressed in the prostate gland [49], associated with prostate cancer [50], and related to immunology [51]. MPO was also reported to be associated with another F3 generation vinclozolin lineage [37].

4.1.14 HAND2
The protein encoded by HAND2 belongs to the basic helix-loop-helix family of transcription factors and has critical roles in endometrial cancer [52].

4.1.15 CCL3
CCL3 is a cytokine belonging to the CC chemokine family that is involved in the recruitment and activation of polymorphonuclear leukocytes during acute inflammation. In experimental autoimmune orchitis, the highest content of CCL3 is in the testicular fluid and is associated with the onset of the disease [53]. In autoimmune orchitis, chronic testicular inflammation, chemokines such as CCL2, CCL3, and CCL4 attract immune cells within the testicular interstitium [54]. CCL3 levels were notably increased in human and mouse benign prostatic hyperplasia prostate glands [55]. Interleukin-6 (IL-6) trans-signaling via its soluble receptor sIL-6R governs the influx of innate immune cells to inflammatory foci through the regulation of the chemokine CCL3 [56]. The CCL3-CCR5 axis regulates the intratumoral accumulation of leukocytes and fibroblasts and promotes angiogenesis in murine lung metastasis [57].
4.1.16  HBE2
HBE2 does not have any orthologs to human genes and information regarding its functions is limited. HBE2 is a part of the hemoglobin complex that transfers oxygen in the blood. There have been no reported relationships between HBE2 and disease.

4.1.17  CMKLR1
CMKLR1, also known as CHEMR23, is a G protein-coupled receptor for the chemoattractant adipokine chemerin and the omega-3 fatty acid derived molecule resolvin E1. CMKLR1 is localized specifically in the Leydig cells of human and rat testes [58]. The CHEMR23/Chemerin axis may have a role in the recruitment of dendritic cells within the kidney in patients affected by lupus nephritis [59]. RvE1/ChEMR23-dependent rS6 phosphorylation occurs in macrophages [60]. CMKLR1 plays roles in the regulation of immune responses [61].

4.1.18  DBH
Dopamine beta-hydroxylase (DBH) mediates dopamine related reactions, is related to kidney function and its role is inheritable [62]. DBH is expressed in the testes [63] and in prostate cancer [64]. DBH has a relationship with immunology [65].

4.1.19  KCNT1
KCNT1 is a member of the calcium-activated potassium channel protein family that is important for kidney functions [66].

4.1.20  TAAR7B
TAAR7B is a member of the trace amine-associated receptors, but does not have a human ortholog, while TAAR1, a member of the human trace amine-associated receptors, has a relationship with the kidney [67].

4.1.21  Fibrinogen beta chain
The fibrinogen beta chain is a blood-borne glycoprotein composed of three pairs of nonidentical polypeptide chains. Heterozygosity for fibrinogen results in the efficient resolution of kidney ischemia reperfusion injury [68]. The association of elevated plasma fibrinogen levels with cancer-specific and overall survival was observed in prostate cancer patients [69]. Immune complexes containing citrullinated fibrinogen co-stimulated macrophages via Toll-like receptor 4 and Fcγ receptor [70].

4.1.22  BMP3
BMP3 is a member of the transforming growth factor beta superfamily. Developing human lung and kidney are major sites for the synthesis of BMP3 [71]. BMP3 mRNA was expressed predominantly in the rat prostate adenocarcinoma, PAIII [72].

4.1.23  ACTG2
ACTG2 encodes actin, a gamma-enteric smooth muscle. ACTG2 was overexpressed in a model of recessive polycystic kidney disease [73].

4.1.24  AQP2
AQP2 is found in the apical cell membranes of kidney collecting duct principal cells and in intracellular vesicles located throughout the cell. AQP2 is expressed in the kidney [74]. AQP2 was localized to Leydig cells, elongated spermatids and round spermatids [75]. Genetic polymorphisms in AQP2 might contribute to chemotherapy responses in lung cancer patients [76].
4.2 Genes identified by limma based FE

4.2.1 qk
MicroRNA-155 promotes the proliferation and invasion abilities of colon cancer cells by targeting quaking (qk) [77]. The tumor suppressing effects of QKI-5 are encoded by the qk gene in prostate cancer [78]. Lipid and fatty acid composition of testes of quaking mice was previously measured [79].

4.2.2 TOP1
Correlation between TOP1 and tyrosyl-DNA phosphodiesterase 1 activities in non-small-cell lung cancer tissues were reported [80]. A novel small molecule hybrid of vorinostat and a topoisomerase inhibitor displays anticancer activity against human hormone-refractory metastatic prostate cancer through dual inhibition of histone deacetylase and TOP1 [81].

4.2.3 Arhgef1
Arhgef1 is known to play roles in cancer [82]. Arhgef1 regulates alpha5beta1 integrin-mediated matrix metalloproteinase expression and is required for homeostatic lung immunity [83].

4.2.4 TEAD2
TEAD2, a Hippo pathway gene, is somatically mutated in gastric and colorectal cancers with high microsatellite instability [84].

4.2.5 Sirt2
A progressive increase in the expression of both SIRT2 and SIRT7 was noted during cancer progression [85]. SIRT2 expression was lower in human prostate cancer [86]. SIRT2 regulates lipopolysaccharide-induced renal tubular CXCL2 and CCL2 expression in kidneys [87].

4.2.6 gmfg
High GMFG expression correlates with poor prognosis and promotes cell migration and invasion in epithelial ovarian cancer [88].

4.2.7 alkbh6
ALKBH6 genes may play important roles in embryonal rhabdomyosarcoma [89].

4.2.8 MCEE
MCEE mutation was observed in gastric and colorectal cancers [90].

4.2.9 hbs1l
Genetic variation at HBS1L-MYB predisposes to myeloproliferative neoplasms [91].

4.2.10 HSPBP1
HSPBP1 expression was reported in leukemia [92].

4.2.11 XRCC1
Polymorphisms of XRCC1 [93] is related to prostate cancer. Transcription of the XRCC1 gene in kidneys of radiosensitive and radioresistant mice following whole-body irradiation [94]. XRCC1 expression is radiobiologically activated in sertoli cells in testis [95].
4.3 Genes identified by SAM based FE, but not by PCA based unsupervised FE, when \( N' = 1000 \)

4.3.1 MYL1
The methylation status of histone H3 lysine 27 in the promoter region of MYL1 was altered by JARID2, which is a direct target of the PAX3-FOXO1 fusion protein, and inhibits myogenic differentiation of rhabdomyosarcoma cells [96].

4.3.2 SLC28A1
SLC28A1 was reported to be downregulated in neoplastic pancreatic tissues [97]. SLC28A1 was expressed in rat glomerulus [98].

4.3.3 PGAM2
Inhibition of PGAM2 by small RNAi or small molecule attenuated tumor growth [99]. PGAM2 was expressed in testis [100].

4.3.4 ALB
Binding of serum albumin on tumor cells was observed [101].

4.3.5 SLC13A3
SLC13A3 plays some roles in prostate cancer [102]. SLC13A3 was expressed in kidneys [103].

4.3.6 TTR
TTR is biomarker of pancreatic ductal adenocarcinoma [104] and prostate cancer [105]. Transthyretin amyloidosis is related to kidney [106]. TTR was observed in fish testis [107]. TTR is functionally related to immunology [108].

4.3.7 ANGPTL1
ANGPTL1 suppresses cancer cell motility [109].

4.3.8 TUBB3
TUBB3 was expressed in non-small-cell lung cancer [110]. TUBB3 was upregulated in prostate cancer [111] and in kidney cancer [112]. TUBB3 is expressed in testis [113].

4.3.9 IL15
IL-15 is a famous cancer therapy target [114]. IL-15 has critical roles in prostate cancer [115]. IL-15 plays critical roles in kidney cancer [116]. An IL-15 isoform is expressed in testis [117]. IL-15 dendritic cells are vaccine candidates for cancer immunotherapy [118].

4.3.10 BACH1
BACH1 plays critical roles in renal cancer [118] and prostate cancer [119].

4.3.11 ZFP36L2
Deletion of ZFP36L2 is related to tumor progression [120]. Functional regulation of ZFP36L2 occurs in response to lipopolysaccharide in mouse RAW264.7 macrophages [121].
References


