Title:

Identification of a TSPY co-expression network associated with DNA hypomethylation and tumor gene expression in somatic cancers

Authors:

Tatsuo Kido and Yun-Fai Chris Lau*

Affiliation of both authors:

Division of Cell and Developmental Genetics, Department of Medicine, Veterans Affairs Medical Center, and Institute for Human Genetics, University of California, San Francisco, California 94121, USA

***Corresponding Author:**

Email: Chris.Lau@UCSF.edu

ABSTRACT

Testis specific protein Y-encoded (*TSPY*) is a Y-located proto-oncogene predominantly expressed in normal male germ cells and various types of germ cell tumor. Significantly, *TSPY* is frequently expressed in somatic cancers including liver cancer but not in adjacent normal tissues, suggesting that ectopic *TSPY* expression could be associated with oncogenesis in non-germ cell cancers. Various studies demonstrated that *TSPY* expression promotes growth and proliferation in cancer cells; however, its relationship to other oncogenic events in *TSPY*-positive cancers remains unknown. The present study seeks to correlate *TSPY* expression with other molecular features in clinical cancer samples, by analyses of RNA-Seq transcriptome and DNA methylation data in the Cancer Genome Atlas (TCGA) database. A total of 53 genes, including oncogenic lineage protein 28 homolog B (*LIN28B*) gene and RNA-binding motif protein Y-linked (*RBMY*) gene, are identified to be consistently co-expressed with *TSPY*, and have been collectively designated as the *TSPY* co-expression network (TCN). TCN genes were simultaneously activated in subsets of liver hepatocellular carcinoma (30%) and lung adenocarcinoma (10%) regardless of pathological stage, but only minimally in other cancer types. Further analysis revealed that the DNA methylation level was globally lower in the TCN-active than TCN-silent cancers. The specific expression and methylation patterns of TCN genes suggest that they could be useful as biomarkers for the diagnosis, prognosis and clinical management of cancers, especially those for liver and lung cancers, associated with *TSPY* co-expression network genes.

Keywords: co-expression network; DNA methylation; gene expression signature; cancer subclassification; Y chromosome genes; *TSPY*; cancer/testis antigens

1. INTRODUCTION

The human testis-specific protein Y-encoded (*TSPY*) gene is a tandemly repeated gene on the short arm of the Y-chromosome (Zhang et al., 1992; Lau et al., 2009). It is specifically expressed in testicular germ cells, predominantly in gonocytes in the embryonic testis and proliferating spermatogonia and early spermatocytes in adult testis (Honecker et al., 2004; Lau et al., 2011). Under pathological conditions, *TSPY* is highly expressed in various types of germ cell tumors (GCTs), including carcinoma *in situ*/intratubular germ cell neoplasia unclassified (CIS/ITGCNU, the presumed precursor for all germ cell tumors), seminoma, extragonadal intracranial GCT, and gonadoblastoma in the patients with disorders of sex developments (DSDs) (Lau et al., 2000; Kersemaekers et al., 2005; Li et al., 2007a; Li et al., 2007b). Deletion-mapping studies have localized the *TSPY* repeat units within the critical region harboring the gonadoblastoma locus on the Y-chromosome (GBY) (Page, 1987; Salo et al., 1995; Tsuchiya et al., 1995). *TSPY* is postulated to serve normal functions in germ cells, but could exacerbate the development of germ cell tumors in dysfunctional germ cells (Schnieders et al., 1996; Kersemaekers et al., 2005; Lau et al., 2009). In addition to germ cell tumors, *TSPY* is frequently expressed in various types of somatic cancer such as prostate cancer (Lau et al., 2003; Kido et al., 2010), melanoma (Gallagher et al., 2005), and liver cancer (Yin et al., 2005; Kido et al., 2014). *TSPY* expression in non-gonadal tumors suggests that it could exert a male-specific effect(s) on oncogenic processes, thereby contributing to sexual dimorphisms favoring males in patients of these somatic cancers.

TSPY protein shares a highly conserved SET/NAP-domain initially identified in the nuclear proto-oncogene (SET) protein (also called template-activating factor I, TAF-I) and nucleosome assembly proteins (NAPs) (Lau, 1999; Ozbun et al., 2001). *TSPY* interacts with the

cyclin B/CDK1 complex *via* its SET/NAP-domain and enhances the kinase activity, leading to cell cycle acceleration, especially at G₂/M phase (Li and Lau, 2008), and promotes tumor growth in a mouse xenograft model (Oram et al., 2006). In addition, TSPY interacts with the translation elongation factor eEF1A1/2 and stimulates protein synthesis (Kido and Lau, 2008), involved in the acceleration of cell proliferation (Spilka et al., 2013; Wang et al., 2015). Hence, TSPY could act on the oncogenic processes through multiple pathways in male-specific manners (Lau et al., 2009; Kido and Lau, 2015).

We recently demonstrated that TSPY is expressed in 30-50% of male hepatocellular carcinoma (HCC) cases (Kido et al., 2014), suggesting that TSPY is expressed in an intrinsic subclass of HCC. Currently, beside cell growth and proliferation, the correlation between TSPY expression and other biological parameters remains unclear. Recent genome-wide gene expression studies have revealed that each cancer type can be classified into intrinsic subclasses based on the gene expression profiles (Perou and Borresen-Dale, 2011; Hoadley et al., 2014). The genes in the same co-expression network tend to be enriched with similar functions and/or biological pathways (Stuart et al., 2003; Horvath and Dong, 2008). Indeed, such classification of tumors, based on their expression profiles of functional gene networks, has been associated with various clinical aspects of tumor progression and/or patient survival (Perou and Borresen-Dale, 2011; Wu and Stein, 2012). Further, the sensitivity to drug treatment could also vary among such cancer intrinsic subclasses (Wu et al., 2013; Costello et al., 2014; Geeleher et al., 2014). Hence, gene co-expression network analyses could be important in defining the unique molecular profiles in specific diseases, thereby providing means for disease subclassification, important for development of diagnosis, prognosis and treatment strategies (Gaiteri et al., 2014), and

contributing to the eventual establishment of precision medicine for clinical management of human diseases, including cancers (Collins and Varmus, 2015).

In the present study, we had explored the characteristic features of TSPY-positive somatic cancers by mining the data sets obtained from the Cancer Genome Atlas (TCGA) (<http://cancergenome.nih.gov/>). Consistent with our previous study (Kido et al., 2014), 30% of male HCC cases were TSPY-positive in the TCGA database (designated as ‘LIHC’ in TCGA database), which was the highest frequency among the somatic cancer types analyzed. A preliminary analysis of the survival duration data demonstrated that the prognosis for TSPY-positive HCC cases was worse than TSPY-negative cases (Fig. S1). Hence, we further performed a comprehensive analysis of gene expression, DNA methylation data and associated clinical features. The gene co-expression network analysis identified a TSPY co-expression network (TCN) in somatic cancers, particularly on HCC and lung adenocarcinoma (LUAD). HCC and LUAD cases could be subclassified into TCN-active and TCN-silent subtypes. Significantly, tumors of TCN-active subtype are associated with global DNA hypomethylation and high-level expression of metastatic miRNAs, features usually associated with poor prognosis of cancers.

2. RESULTS

2.1. Identification of a TSPY gene co-expression network in somatic cancers

To explore the existence of a TSPY gene co-expression network, we initially investigated the datasets of tumor and non-tumor paired samples from male somatic cancers. These samples included 27 cases of HCC, 24 cases of LUAD, 28 cases of head-neck squamous cell carcinoma (HNSC), 10 cases of bladder urothelial carcinoma (BLCA), 16 cases of kidney renal papillary

cell carcinoma (KIRP), 16 cases of thyroid carcinoma (THCA), and 16 cases of colon adenocarcinoma (COAD) (Fig. S2). Hierarchical gene clustering analysis based on the gene expression changes ($\log_2([\text{tumor}]/[\text{non-tumor}])$) demonstrated that 66 genes were closely clustered with *TSPY* (Pearson correlation distance < 0.7) (Fig. 1A, clusters C1-C7). Since *TSPY* and RNA-binding motif protein Y-linked (*RBMY*) genes are multi-copy genes located on Y-chromosome, transcripts from the respective isoforms were collectively grouped as *TSPY* gene and *RBMY* gene, respectively. Verification by using weighted co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) showed that the genes of clusters C1-C6 were classified into the ‘Skyblue’ (co-expressed) module, while genes of cluster C7 were classified to other modules (Figs. 1A and S3). According to these results, 53 genes in the clusters C1-C6 were identified as co-expressed with *TSPY*, and hereto designated as the *TSPY* co-expression network (TCN) (Fig. 1A and Table S1). Forty of 53 TCN genes (75 %) are located on the sex-chromosomes; and 27 genes (51 %) code for known cancer/testis (CT) antigens that are predominantly expressed in normal testis and ectopically in somatic cancers (Table S1) (Simpson et al., 2005; Fratta et al., 2011; Whitehurst, 2014). The full names of the TCN genes are indicated in Table S1.

The expression profile showed that the TCN genes were simultaneously up-regulated in a subset of cases (Fig. 1B), including approximately 35 % of HCC cases and 20 % of LUAD cases, but rarely in other cancer types (Fig. 1B). To further explore if such expression pattern of TCN genes could be associated with any pathological stages, we extended our analysis to the unpaired tumor samples and focused on the 18 TCN genes that were consistently expressed in tumors but not non-tumor liver tissues (designated as TCN18 genes, Fig. 1C). The correlation between the number of up-regulated TCN18 genes and that of TCN53 genes was linear in both HCC (*r*-

squared value= 0.98) and LUAD (r -squared value= 0.92). Each cancer case was scored based on the number of up-regulated TCN18 genes (TCN18-score, i.e., 0 (no gene expression) to 18 (all 18 genes expression)) for classification (Fig. S2). In this analysis, a total of 185 cases of male HCC and 208 cases of male LUAD were investigated for the correlation between TCN18-score and pathological stage. The stage-IV cases were excluded, since the sample sizes were too small for analyses. The results showed that the ratios of high score cases (TCN18-score ≥ 10) were approximately 30% in HCC and 10% in LUAD regardless of pathological stage (Fig. 1D), and that majority of TCN genes were expressed at the early time points of cancer development in a certain group. Tumors with high number of TCN18-expressing genes (i.e., TCN18-score ≥ 10) were classified as the TCN-active tumors, and those with low number of TCN18-expressing genes (i.e., TCN18-score ≤ 4) were classified as the TCN-silent tumors in the following analyses. It is noteworthy that the ratio of extremely high score cases (TCN18-score ≥ 15) increased from stage-I to stage-III of HCC (Fig. 1D). Although additional studies with larger sample size are needed to confirm our findings, the extremely high score (TCN18-score ≥ 15) observed in the present study might be associated with cancer progression and/or late stages in HCC.

The principal component analysis (PCA), based on the gene expression data, showed that TCN-active samples and TCN-silent samples were distributed similarly in the PCA plots (Fig. S4), suggesting that the TCN-active samples were classified independently from potential batch bias(es). Hence, up-regulation of TCN-genes could be specific for a certain group of tumors.

2.2. Correlation between the activities of TCN genes and the expression of oncogenic miRNAs

Since miRNAs serve numerous functions, including promoting and suppressing roles, in human oncogenesis, the expression patterns of specific miRNAs could be tightly correlated with the human cancer development and clinical outcomes (Guan et al., 2015; Liu et al., 2015). To explore the molecular biological features of TCN-active tumors, we investigated the differential expression of miRNAs between TCN-active tumors and TCN-silent tumors, using TCC method (Sun et al., 2013). Our results showed that 119 miRNAs and 28 miRNAs were differentially expressed in HCC and LUAD at pathological stage-I, respectively (Fig. 2A and Table S2). Seven miRNAs, including *mir-105-1*, *mir-105-2* and *mir-767*, were commonly expressed at high levels in the TCN-active tumors of both HCC and LUAD (Fig. 2A, blue plots). Mapping of the TCN-related miRNA genes in the UCSC genome browser (hg19/GRCh37) showed that *mir-105-1*, *mir-105-2* and *mir-767* genes are located within the intron 1 of gamma-aminobutyric acid type A receptor alpha 3 subunit (*GABRA3*) gene, a member of TCN genes (Fig. 2B). In addition, 9 TCN genes are located within a 400-kb region of Xq28 (Fig. 2B), suggesting that Xq28 could be a hotspot for TCN genes. It has been reported that over-expression of cancer-secreted *mir-105* disrupts vascular endothelial barriers and can promote cancer metastasis (Zhou et al., 2014). Further, *mir-767* stimulates cell migration in a human hepatic adenocarcinoma cell line (Zha et al., 2014). Hence, we surmise that TCN-genes could potentiate migration and metastasis in TCN-active tumors through the over-expression of the oncogenic *mir-105-1*, *mir-105-2* and *mir-767*.

In addition, all members of the mir-515 family, consisting of 46 miRNAs, were expressed at extremely higher levels (>50 folds) in the TCN-active HCC (Fig. 2A, red plots in HCC). The mir-515 family of miRNAs are coded in the chromosome 19 miRNA cluster (C19MC) that is a primate-specific domain, and transcribed as a single precursor RNA from a CpG island located at ~17 kb upstream of C19MC (Fig. 2C) (Noguer-Dance et al., 2010). While the biological

functions of mir-515 family miRNAs vary in tumor (Table S3), their strong up-regulation could have synergistic impacts on the balance of miRNAs in tumor cells. On the other hand, only a single mir-515 family miRNA was up-regulated in the TCN-active LUAD (Fig. 2A, a red plot in LUAD). These observations suggest that the activities of mir-515 family gene could be differentially regulated between HCC and LUAD.

The expression levels of tumor-suppressive *let-7* miRNAs were slightly but consistently down-regulated in both TCN-active HCC and TCN-active LUAD (Fig. 2A). *let-7* miRNAs are targets of lineage protein 28 homolog B (LIN28B), a member of TCN, and degraded by LIN28B-mediated mechanisms (Zhou et al., 2013). Hence, the tumor suppressive effect by *let-7* is likely repressed by LIN28B in the TCN-active tumors.

2.3. Association between the activities of TCN genes and the methylation status of their genomic DNA

The activities of various CT antigen genes, including *TSPY*, *MAGE* and *SSX* genes, are associated with the DNA methylation status in either their promoter and/or coding regions (Gure et al., 2002; Gallagher et al., 2005; Dyrskjot et al., 2012; Coral et al., 2013). Accordingly, we had examined such correlation between the expression and DNA methylation status of TCN genes in HCC samples in the TCGA database, which were generated with the Illumina Human Methylation450 BeadChip collectively consisting of 191 probes at both 200 bp promoter (TSS200) and exon 1 regions of TCN genes. The β -value indicates the proportion of methylation, ranging from 0 to 1 (0, unmethylated, and 1, totally methylated), at each probed locus. Data for *CSAG2*, *CSAG3*, *CTAG1A/B*, *SSX2*, *TMPRSS15*, and *VCY* genes were not available in the datasets from TCGA database, and were excluded from the analysis. The 191 loci at the

TSS200/exon 1 regions of TCN genes were highly methylated in non-tumor liver samples (β -value= 0.802 ± 0.016 (mean $\pm$ SEM)) (Fig. 3A). Those methylation levels were decreased in the TCN-silent tumors (β -value= 0.668 ± 0.014), and further decreased in the TCN-active tumors (mean β -value= 0.489 ± 0.012) at pathologic stage-I (Fig. 3A). The methylation levels at each locus were compared between pathologic stage-I and stage-III in the TCN-silent tumor group and TCN-active tumor group, respectively. The results showed that the methylation levels of 187 loci (98 %) in TCN-silent tumor group and 186 loci (97%) in TCN-active tumor group showed minimal changes between pathological stage-I and stage-III (Student's t-test $P \geq 0.05$) (Fig. 3B).

To further investigate the differences between TCN-active and TCN-silent tumors, we compared the DNA methylation levels at each locus of the TSS200/exon1 regions of TCN genes between the two groups. The results showed that the methylation levels of 147 loci (77 % of 191 loci) were significantly lower in TCN-active tumors than TCN-silent tumors of HCC at pathologic stage-I, while only 2 loci (1 %) were higher in the TCN-active tumors (Fig. 3C). These observations suggest that the expressions of TCN genes could be associated with the extreme hypomethylation within their promoter and/or exon 1 regions. Together with the results of TCN18 scoring, we surmise that the expression patterns of TCN genes are established epigenetically at early stages of cancer development and persisted throughout the late stages in cancer progression.

The expression of mir-515 family is regulated by DNA methylation of the CpG island at their promoter region (Noguer-Dance et al., 2010). The DNA methylation data of TCGA showed that, similar to observation on the TCN genes, the CpG island at the promoter of mir-515 family was extremely hypomethylated in the TCN-active HCC (Fig. S5). On the other hand, it remained to be highly methylated in the TCN-active LUAD (Fig. S5). Such difference might be related to

the difference in the expression of mir-515 family between HCC and LUAD (Fig. 2A), as noted above.

2.4. Association between the activities of TCN genes and the global genomic DNA hypomethylation

Since the activities of major TCN genes were associated with the DNA methylation status of their promoter and/or exon 1 regions, we had extended our analysis to the global DNA methylation status in the HCC samples. Our results showed that the methylation levels were globally lower in TCN-active tumors than TCN-silent tumors at pathologic stage-I, regardless of their chromosomal locations (Figs. 4A and S6). Next, we extended our study to the methylation levels at other functional regions; e.g., 1500 bp blocks upstream of the transcription start sites (TSS1500), 200 bp blocks upstream of the transcription start sites (TSS200), exon 1/5'UTR regions, and whole CpG islands. The results showed that the methylation levels were generally lower in TCN-active tumors than those of TCN-silent tumors at these regions (Fig. 4B). Further, we explored the differences between TCN-related loci and TCN-unrelated loci of the top 3000 loci (based on the *P*-value) of CpG islands, and compared their methylation levels among different sample types. The TCN-related loci were highly methylated in non-tumor tissues, and extremely hypomethylated in the TCN-active tumors at both stage-I and stage-III (Fig. 4C, left side). In contrast, the methylation levels of TCN-unrelated loci were generally low in non-tumor tissues, and their methylation levels were not significantly changed even in the TCN-active tumors (Fig. 4C, right side). These observations suggest that the methylation status of global genomic DNA decreased from TCN-silent to TCN-active tumors. Such changes in methylation

status were not observed in TCN-unrelated tumors, without TCN gene expression subclassification.

Similarly with HCC, data of LUAD also showed that the DNA methylation levels were globally lower in TCN-active tumors than TCN-silent tumors at pathological stage-I (Fig. 4D). The sample size of stage-III TCN-active LUAD was not sufficient for analyses.

3. DISCUSSION

Human cancers are usually heterogeneous, with significant variations with respects to gender, incidence, progression and treatment responses. With the advance of next generation sequencing, intrinsic subtypes of tumors can now be classified based on specific gene expression profiles as in the form of RNA-Seq or transcriptome analysis, which can then be correlated with clinical outcomes and responses to drug treatments (Perou and Borresen-Dale, 2011; Wu and Stein, 2012; Santos et al., 2015). The availability of the TCGA database has greatly facilitated such investigations, focusing on identification of gene expression signatures for specific subtypes of particular cancer of interests, which can then be correlated with the related clinical features of the patients (Gonda et al., 2014; Weisenberger, 2014; Tomczak et al., 2015; Wong et al., 2015). Taking advantage of those important resources, we have identified TCN genes, which could be used to classify HCC and LUAD tumors as TCN-active and TCN-silent groups, based on the expression patterns of its component genes. In the TCN-active tumors, various known proto-oncogenes and oncogenes, including *GABRA3*, *LIN28B*, *MAGEAs*, *RBMV* and *TSPY*, as well as oncogenic miRNAs, were synergistically upregulated (Figs. 1 and 2) (Lau et al., 2000; Monte et al., 2006; Oram et al., 2006; Liu et al., 2008; Lau et al., 2009; Tsuei et al., 2011; Zhou et al., 2013; Zha et al., 2014; Zhou et al., 2014). Hence, we expect that the TCN-active tumors might

express the characteristic pathological properties, although future linkage analysis with detailed pathological and long-term follow-up data will be required. Indeed, our preliminary analysis showed that liver cancer patients positive for *TSPY* had poorer survival rate than those negative for it. It is noteworthy that the involvement of Y-located *RBMY* and *TSPY* genes in the development of HCC has been suggested (Tsuei et al., 2011; Kido et al., 2014). In the cases of liver cancer, there is important sex difference with male to female ratio as high as 6 to 1 among some patient populations (Yeh and Chen, 2010; Nishida et al., 2012). We surmise that a synergistic expression of the Y-located genes, including *TSPY* and *RBMY*, may contribute to such sex differences of liver cancer.

In addition to their oncogenic functions, expression levels of TCN genes are associated with a parallel decreases in methylation levels not only for their respective promoters but also for the global tumor genome as a whole (Figs. 3, 4, and S6). These features were not obvious without the subclassification of tumors according to TCN expression signatures, i.e., TCN-active and TCN-silent categorization. Others showed that the genome-wide DNA hypomethylation is correlated with chromosomal instability leading to loss of heterozygosity (Daskalos et al., 2009; Aoki et al., 2012). Higher rate chromosomal instability increases mutagenic events and is associated with poorer survival of the patients (Carter et al., 2006; Watanabe et al., 2012). Hence, TCN expression signatures and promoter methylation status could be collectively useful biomarkers for subclassification of male patients with liver or lung cancer, which can be correlated to their respective clinical outcomes.

As noted, significant number CT antigens, particularly located at various loci on the X chromosome and their Y-homologues, such as *VCX* and *VCY*, are co-expressed with *TSPY* in human cancers (Fig. 1). CT antigens are usually repeated genes coding for small proteins,

specifically expressed in the testis and various cancers (Fratta et al., 2011; Whitehurst, 2014). Due to their unique cancer-specific expression, they have been proposed as potential immunotherapeutic targets in cancer treatments (Fratta et al., 2011; Taguchi et al., 2014; Hirohashi et al., 2016). Successful translational implementation of TCN-based signatures could be important in establishing the precision medicine for clinical management of these human cancers (Collins and Varmus, 2015).

4. MATERIALS AND METHODS

4.1. Data sets

The level-3 RNAseq version 2 gene expression data with available clinical data were downloaded from the Cancer Genome Atlas (TCGA) portal web site (<http://cancergenome.nih.gov/>). All RNAseq data, used in the present study, were originally deposited by investigators from the University of North Carolina at Chapel Hill. Clinical data included the pathological tumor stage based on the criteria by American Joint Committee on Cancer (AJCC), and higher number indicates more extensive disease (Compton et al., 2012). In addition, DNA methylation data and the level-3 miRNA expression data case-matched with RNAseq data were downloaded similarly. The DNA methylation data have been generated using the Infinium Human Methylation450 BeadChip assay (Illumina, Inc., San Diego, California, USA). The numbers of samples analyzed in the present study were summarized in Fig. S2. Briefly, there were 137 matched pairs of male tumor and corresponding non-tumor samples including 27 cases of HCC (designated as ‘LIHC’ in TCGA database), 24 cases of LUAD, 28 cases of HNSC, 10 cases of BLCA, 16 cases of KIRP, 16 cases of THCA, and 16 cases of

COAD. Unpaired male tumor samples included 158 HCC cases and 184 LUAD cases. All data were downloaded in November 2014.

4.2. Data analysis

Gene co-expression networks were predicted based on the $\log_2$ value of the fold difference ([tumor]/[non-tumor]) of the RNAseq expression data from paired tumor and non-tumor samples. The data were normalized by iDEGES/edgeR software package before analysis (Sun et al., 2013). To avoid the undefined result of division by zero, the value of 0 in data was converted to 0.001 before calculation of fold changes. Hierarchical clustering, based on the Person's correlation, was performed by using MeV software (version 4.9) (<http://www.tm4.org/mev.html>) (Saeed et al., 2006). WGCNA was performed by using an R software package for WGCNA (Langfelder and Horvath, 2008). Differences in miRNA expression level between two classified groups were analyzed by TCC/edgeR software package (Sun et al., 2013). Analyses of DNA methylation data were performed by using Excel (Microsoft, Redmond, WA, USA). The false discovery rate (FDR) was calculated based on the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995). Overall workflow is presented in Fig. S2.

5. ACKNOWLEDGEMENTS

This study was partially supported by a Merit-Reviewed grant from the Department of Veterans Affairs and a Peer-Reviewed Cancer Research Program grant from the Department of Defense to Y-FCL. Y-FCL is a Research Career Scientist of the Department of Veterans Affairs.

6. REFERENCES

- Aoki, Y., Nojima, M., Suzuki, H., Yasui, H., Maruyama, R., Yamamoto, E., Ashida, M., Itagaki, M., Asaoku, H., Ikeda, H., Hayashi, T., Imai, K., Mori, M., Tokino, T., Ishida, T., Toyota, M., Shinomura, Y., 2012. Genomic vulnerability to LINE-1 hypomethylation is a potential determinant of the clinicogenetic features of multiple myeloma. *Genome Med.* 4, 101.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B (Methodological)* 57, 289-300.
- Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., Szallasi, Z., 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* 38, 1043-1048.
- Collins, F.S., Varmus, H., 2015. A new initiative on precision medicine. *N. Engl. J. Med.* 372, 793-795.
- Compton, C.C., Byrd, D.R., Garcia-Aguilar, J., Kurtzman, S.H., Olawaiye, A., Washington, M.K., 2012. *AJCC Cancer Staging Atlas* (Springer New York).
- Coral, S., Parisi, G., Nicolay, H.J., Colizzi, F., Danielli, R., Fratta, E., Covre, A., Taverna, P., Sigalotti, L., Maio, M., 2013. Immunomodulatory activity of SGI-110, a 5-aza-2'-deoxycytidine-containing demethylating dinucleotide. *Cancer Immunol. Immunother.* 62, 605-614.
- Costello, J.C., Heiser, L.M., Georgii, E., Gonen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., Mpindi J.P., Kallioniemi, O., Honkela, A., Aittokallio, T., Wennerberg, K., NCI DREAM Community, Collins, J.J., Gallahan, D., Singer, D., Saez-Rodriguez, J., Kaski, S., Gray, J.W., Stolovitzky, G., 2014. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* 32, 1202-1212.
- Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., Kotsinas, A., Gorgoulis, V., Field, J.K., Liloglou, T., 2009. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int. J. Cancer* 124, 81-87.
- Dyrskjot, L., Zieger, K., Kissow Lildal, T., Reinert, T., Gruselle, O., Coche, T., Borre, M., Orntoft, T.F., 2012. Expression of MAGE-A3, NY-ESO-1, LAGE-1 and PRAME in urothelial carcinoma. *Br. J. Cancer* 107, 116-122.
- Fratta, E., Coral, S., Covre, A., Parisi, G., Colizzi, F., Danielli, R., Nicolay, H.J., Sigalotti, L., Maio, M., 2011. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Mol. Oncol.* 5, 164-182.
- Gaiteri, C., Ding, Y., French, B., Tseng, G.C., Sibille, E., 2014. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav.* 13, 13-24.
- Gallagher, W.M., Bergin, O.E., Rafferty, M., Kelly, Z.D., Nolan, I.M., Fox, E.J., Culhane, A.C., McArdle, L., Fraga, M.F., Hughes, L., Currid, C.A., O'Mahony, F., Byrne, A., Murphy, A.A., Moss, C., McDonnell, S., Stallings, R.L., Plumb, J.A., Esteller, M., Brown, R., Dervan, P.A.,

- Easty, D.J., 2005. Multiple markers for melanoma progression regulated by DNA methylation: insights from transcriptomic studies. *Carcinogenesis* 26, 1856-1867.
- Geeleher, P., Cox, N.J., Huang, R.S., 2014. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* 15, R47.
- Gonda, D.D., Cheung, V.J., Muller, K.A., Goyal, A., Carter, B.S., Chen, C.C., 2014. The Cancer Genome Atlas expression profiles of low-grade gliomas. *Neurosurg. Focus* 36, E23.
- Guan, Y., Chen, L., Bao, Y., Qiu, B., Pang, C., Cui, R., Wang, Y., 2015. High miR-196a and low miR-367 cooperatively correlate with unfavorable prognosis of high-grade glioma. *Int. J. Clin. Exp. Pathol.* 8, 6576-6588.
- Gure, A.O., Wei, I.J., Old, L.J., Chen, Y.T., 2002. The SSX gene family: characterization of 9 complete genes. *Int. J. Cancer* 101, 448-453.
- Hirohashi, Y., Torigoe, T., Tsukahara, T., Kanaseki, T., Kochin, V., Sato, N., 2016. Immune responses to human cancer stem-like cells/cancer-initiating cells. *Cancer Sci.* 107, 12-17.
- Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V., Zhang, J., Kandoth, C., Akbani, R., Shen, H., Omberg, L., Chu, A., Margolin, A.A., Van't Veer, L.J., Lopez-Bigas, N., Laird, P.W., Raphael, B.J., Ding, L., Robertson, A.G., Byers, L.A., Mills, G.B., Weinstein, J.N., Van Waes, C., Chen, Z., Collisson, E.A., Cancer Genome Atlas Research Network, Benz, C.C., Perou, C.M., Stuart, J.M., 2014. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158, 929-944.
- Honecker, F., Stoop, H., de Krijger, R.R., Chris Lau, Y.F., Bokemeyer, C., Looijenga, L.H., 2004. Pathobiological implications of the expression of markers of testicular carcinoma in situ by fetal germ cells. *J. Pathol.* 203, 849-857.
- Horvath, S., Dong, J., 2008. Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* 4, e1000117.
- Kersemaekers, A.M., Honecker, F., Stoop, H., Cools, M., Molier, M., Wolffenbuttel, K., Bokemeyer, C., Li, Y., Lau, Y.F., Oosterhuis, J.W., Looijenga, L.H., 2005. Identification of germ cells at risk for neoplastic transformation in gonadoblastoma: an immunohistochemical study for OCT3/4 and TSPY. *Hum. Pathol.* 36, 512-521.
- Kido, T., Hatakeyama, S., Ohyama, C., Lau, Y.F., 2010. Expression of the Y-encoded TSPY is associated with progression of prostate cancer. *Genes* 1, 283-293.
- Kido, T., Lau, Y.F., 2008. The human Y-encoded testis-specific protein interacts functionally with eukaryotic translation elongation factor eEF1A, a putative oncoprotein. *Int. J. Cancer* 123, 1573-1585.
- Kido, T., Lau, Y.F., 2015. Roles of the Y chromosome genes in human cancers. *Asian J. Androl.* 17, 373-380.
- Kido, T., Lo, R.C., Li, Y., Lee, J., Tabatabai, Z.L., Ng, I.O., Lau, Y.F., 2014. The potential contributions of a Y-located protooncogene and its X homologue in sexual dimorphisms in hepatocellular carcinoma. *Hum. Pathol.* 45, 1847-1858.

- Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.
- Lau, Y., Chou, P., Iezzoni, J., Alonzo, J., Komuves, L., 2000. Expression of a candidate gene for the gonadoblastoma locus in gonadoblastoma and testicular seminoma. *Cytogenet. Cell Genet.* 91, 160-164.
- Lau, Y.F., 1999. Gonadoblastoma, testicular and prostate cancers, and the TSPY gene. *Am. J. Hum. Genet.* 64, 921-927.
- Lau, Y.F., Lau, H.W., Komuves, L.G., 2003. Expression pattern of a gonadoblastoma candidate gene suggests a role of the Y chromosome in prostate cancer. *Cytogenet. Genome Res.* 101, 250-260.
- Lau, Y.F., Li, Y., Kido, T., 2009. Gonadoblastoma locus and the TSPY gene on the human Y chromosome. *Birth Defects Res. C Embryo Today* 87, 114-122.
- Lau, Y.F., Li, Y., Kido, T., 2011. Role of the Y-located putative gonadoblastoma gene in human spermatogenesis. *Syst. Biol. Reprod. Med.* 57, 27-34.
- Li, Y., Lau, Y.F., 2008. TSPY and its X-encoded homologue interact with cyclin B but exert contrasting functions on cyclin-dependent kinase 1 activities. *Oncogene* 27, 6141-6150.
- Li, Y., Tabatabai, Z.L., Lee, T.L., Hatakeyama, S., Ohyama, C., Chan, W.Y., Looijenga, L.H., Lau, Y.F., 2007a. The Y-encoded TSPY protein: a significant marker potentially plays a role in the pathogenesis of testicular germ cell tumors. *Hum. Pathol.* 38, 1470-1481.
- Li, Y., Vilain, E., Conte, F., Rajpert-De Meyts, E., Lau, Y.F., 2007b. Testis-specific protein Y-encoded gene is expressed in early and late stages of gonadoblastoma and testicular carcinoma in situ. *Urol. Oncol.* 25, 141-146.
- Liu, W., Zhao, Z.Y., Shi, L., Yuan, W.D., 2015. Tissue microRNA-126 expression level predicts outcome in human osteosarcoma. *Diagn. Pathol.* 10, 116.
- Liu, Y., Li, Y.H., Guo, F.J., Wang, J.J., Sun, R.L., Hu, J.Y., Li, G.C., 2008. Gamma-aminobutyric acid promotes human hepatocellular carcinoma growth through overexpressed gamma-aminobutyric acid A receptor alpha 3 subunit. *World J. Gastroenterol.* 14, 7175-7182.
- Monte, M., Simonatto, M., Peche, L.Y., Bublik, D.R., Gobessi, S., Pierotti, M.A., Rodolfo, M., Schneider, C., 2006. MAGE-A tumor antigens target p53 transactivation function through histone deacetylase recruitment and confer resistance to chemotherapeutic agents. *Proc. Natl. Acad. Sci. U. S. A.* 103, 11160-11165.
- Nishida, N., Arizumi, T., Hayaishi, S., Takita, M., Kitai, S., Yada, N., Hagiwara, S., Inoue, T., Minami, Y., Ueshima, K., Sakurai, T., Ikai, I., Kudo, M., 2012. Gender differences in the livers of patients with hepatocellular carcinoma and chronic hepatitis C infection. *Dig. Dis.* 30, 547-553.
- Noguer-Dance, M., Abu-Amero, S., Al-Khtib, M., Lefevre, A., Coullin, P., Moore, G.E., Cavaille, J., 2010. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Hum. Mol. Genet.* 19, 3566-3582.

- Oram, S.W., Liu, X.X., Lee, T.L., Chan, W.Y., Lau, Y.F., 2006. TSPY potentiates cell proliferation and tumorigenesis by promoting cell cycle progression in HeLa and NIH3T3 cells. *BMC Cancer* 6, 154.
- Ozbun, L.L., You, L., Kiang, S., Angdisen, J., Martinez, A., Jakowlew, S.B., 2001. Identification of differentially expressed nucleolar TGF-beta1 target (DENTT) in human lung cancer cells that is a new member of the TSPY/SET/NAP-1 superfamily. *Genomics* 73, 179-193.
- Page, D.C., 1987. Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. *Development* 101 (Suppl), 151-155.
- Perou, C.M., Borresen-Dale, A.L., 2011. Systems biology and genomics of breast cancer. *Cold Spring Harb. Perspect. Biol.* 3, a003293.
- Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A., Quackenbush, J., 2006. TM4 microarray software suite. *Methods Enzymol.* 411, 134-193.
- Salo, P., Kaariainen, H., Petrovic, V., Peltomaki, P., Page, D.C., de la Chapelle, A., 1995. Molecular mapping of the putative gonadoblastoma locus on the Y chromosome. *Genes Chromosomes Cancer* 14, 210-214.
- Santos, C., Sanz-Pamplona, R., Nadal, E., Grasselli, J., Pernas, S., Dienstmann, R., Moreno, V., Tabernero, J., Salazar, R., 2015. Intrinsic cancer subtypes--next steps into personalized medicine. *Cell Oncol.* 38, 3-16.
- Schnieders, F., Dork, T., Arnemann, J., Vogel, T., Werner, M., Schmidtke, J., 1996. Testis-specific protein, Y-encoded (TSPY) expression in testicular tissues. *Hum. Mol. Genet.* 5, 1801-1807.
- Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T., Old, L.J., 2005. Cancer/testis antigens, gametogenesis and cancer. *Nat. Rev. Cancer* 5, 615-625.
- Spilka, R., Ernst, C., Mehta, A.K., Haybaeck, J., 2013. Eukaryotic translation initiation factors in cancer development and progression. *Cancer Lett.* 340, 9-21.
- Stuart, J.M., Segal, E., Koller, D., Kim, S.K., 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249-255.
- Sun, J., Nishiyama, T., Shimizu, K., Kadota, K., 2013. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* 14, 219.
- Taguchi, A., Taylor, A.D., Rodriguez, J., Celiktas, M., Liu, H., Ma, X., Zhang, Q., Wong, C.H., Chin, A., Girard, L., Behrens, C., Lam, W.L., Lam, S., Minna, J.D., Wistuba, I.I., Gazdar, A.F., Hanash, S.M., 2014. A search for novel cancer/testis antigens in lung cancer identifies VCX/Y genes, expanding the repertoire of potential immunotherapeutic targets. *Cancer Res.* 74, 4694-4705.
- Tomczak, K., Czerwinska, P., Wiznerowicz, M., 2015. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 19, A68-77.
- Tsuchiya, K., Reijo, R., Page, D.C., Distech, C.M., 1995. Gonadoblastoma: molecular definition of the susceptibility region on the Y chromosome. *Am. J. Hum. Genet.* 57, 1400-1407.

- Tsuei, D.J., Lee, P.H., Peng, H.Y., Lu, H.L., Su, D.S., Jeng, Y.M., Hsu, H.C., Hsu, S.H., Wu, J.F., Ni, Y.H., Chang, M.H., 2011. Male germ cell-specific RNA binding protein RBMY: a new oncogene explaining male predominance in liver cancer. *PLoS One* 6, e26948.
- Wang, W., Nag, S., Zhang, X., Wang, M.H., Wang, H., Zhou, J., Zhang, R., 2015. Ribosomal proteins and human diseases: pathogenesis, molecular mechanisms, and therapeutic implications. *Med. Res. Rev.* 35, 225-285.
- Watanabe, T., Kobunai, T., Yamamoto, Y., Matsuda, K., Ishihara, S., Nozawa, K., Yamada, H., Hayama, T., Inoue, E., Tamura, J., Inuma, H., Akiyoshi, T., Muto, T., 2012. Chromosomal instability (CIN) phenotype, CIN high or CIN low, predicts survival for colorectal cancer. *J. Clin. Oncol.* 30, 2256-2264.
- Weisenberger, D.J., 2014. Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J. Clin. Invest.* 124, 17-23.
- Whitehurst, A.W., 2014. Cause and consequence of cancer/testis antigen activation in cancer. *Annu. Rev. Pharmacol. Toxicol.* 54, 251-272.
- Wong, H.K., Fatimy, R.E., Onodera, C., Wei, Z., Yi, M., Mohan, A., Gowrisankaran, S., Karmali, P., Marcusson, E., Wakimoto, H., Stephens, R., Uhlmann, E.J., Song, J.S., Tannous, B., Krichevsky, A.M., 2015. The cancer genome atlas analysis predicts microRNA for targeting cancer growth and vascularization in glioblastoma. *Mol. Ther.* 23, 1234-1247.
- Wu, D., Pang, Y., Wilkerson, M.D., Wang, D., Hammerman, P.S., Liu, J.S., 2013. Gene-expression data integration to squamous cell lung cancer subtypes reveals drug sensitivity. *Br. J. Cancer* 109, 1599-1608.
- Wu, G., Stein, L., 2012. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* 13, R112.
- Yeh, S.H., Chen, P.J., 2010. Gender disparity of hepatocellular carcinoma: the roles of sex hormones. *Oncology* 78 (Suppl 1), 172-179.
- Yin, Y.H., Li, Y.Y., Qiao, H., Wang, H.C., Yang, X.A., Zhang, H.G., Pang, X.W., Zhang, Y., Chen, W.F., 2005. TSPY is a cancer testis antigen expressed in human hepatocellular carcinoma. *Br. J. Cancer* 93, 458-463.
- Zha, R., Guo, W., Zhang, Z., Qiu, Z., Wang, Q., Ding, J., Huang, S., Chen, T., Gu, J., Yao, M., He, X., 2014. Genome-wide screening identified that miR-134 acts as a metastasis suppressor by targeting integrin beta1 in hepatocellular carcinoma. *PLoS One* 9, e87665.
- Zhang, J.S., Yang-Feng, T.L., Muller, U., Mohandas, T.K., de Jong, P.J., Lau, Y.F., 1992. Molecular isolation and characterization of an expressed gene from the human Y chromosome. *Hum. Mol. Genet.* 1, 717-726.
- Zhou, J., Ng, S.B., Chng, W.J., 2013. LIN28/LIN28B: an emerging oncogenic driver in cancer stem cells. *Int. J. Biochem. Cell Biol.* 45, 973-978.
- Zhou, W., Fong, M.Y., Min, Y., Somlo, G., Liu, L., Palomares, M.R., Yu, Y., Chow, A., O'Connor, S.T., Chin, A.R., Yen, Y., Wang, Y., Marcusson, E.G., Chu, P., Wu, J., Wu, X., Li, A.X., Li, Z., Gao, H., Ren, X., Boldin, M.P., Lin, P.C., Wang, S.E., 2014. Cancer-

secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer Cell* 25, 501-515.

FIGURE LEGENDS

Fig. 1.

Identification and expression analysis of the TSPY co-expression network genes. **A:** A result of hierarchical clustering of the whole-genome gene-expression for seven types of male cancer samples. Sixty-six genes, including multi copy genes of *TSPY* and *RBMY*, were clustered with *TSPY* by MeV program (Pearson correlation distance <0.7). C1-C7 indicate sub-clusters. Genes in sub-clusters C1-C6 were classified into the ‘Skyblue’ module by WGCNA method. Hence, genes of C1-C6 are designated as the TSPY co-expression network (TCN) genes. **B:** A panel of gene expression profile of TCN genes in seven types of male somatic cancer. Red pixels indicate the up-regulation of the indicated gene (row) in respective cancer case (column). The numbers of TSPY-positive (TSPY(+)) cases and ratios per each cancer type are also indicated at the bottom. **C:** List of TCN18 genes that are specifically expressed in tumor samples. **D:** Stacked bar graph summarizing the distribution of TCN18-score at each pathologic stage (I-III). NT indicates non-tumor sample. Number in parenthesis indicates the sample size of each group.

Fig. 2.

Differential expression of miRNAs between TCN-active tumors and TCN-silent tumors. **A:** MA-plot for the differential expression of miRNAs between TCN-active tumors and TCN-silent tumors at pathologic stage-I; HCC (left panel) and LUAD (right panel). Color plots indicate the differentially expressed miRNAs predicted by TCC method ($FDR < 0.05$). Red indicates mir-515 family; blue indicates *mir-105-1/2*, and *mir-767*; green indicates let-7 family. **B:** To scale the representation of the 750 kb genomic region centered on *GABRA3* on human chromosome Xq28.

TCN genes are labeled. **C:** Genomic structure of the chromosome 19 miRNA cluster (C19MC) coding mir-515 family.

Fig. 3.

DNA methylation status of the TCN genes in HCC. **A:** Distributions of the methylation β -value among non-tumor tissues (NT), TCN-silent tumors and TCN-active tumors at the loci within TSS200/exon 1 regions of TCN genes. **B:** Comparison of DNA methylation levels at respective locus within TSS200/exon 1 regions of TCN genes between pathologic stage-I (X-axis) and stage-III (Y-axis) for TCN-silent tumors (left panel) and TCN-active tumors (right panel). Red plots indicate the loci differentially methylated between stage-I and stage-III ($P < 0.05$). Gray plots indicate the loci that were similarly methylated between two stages ($P \geq 0.05$). **C:** In top panel, each plot indicates the difference in DNA methylation β -value ($\Delta\beta$ -value) between tumor and non-tumor samples at respective locus within TSS200/exon 1 regions of TCN genes at pathologic stage-I; red plot, TCN-active tumors; blue plot, TCN-silent tumors. On the bottom panel, orange bars indicate the differences in methylation β -value between TCN-active tumors and TCN-silent tumors at respective locus ($P < 0.05$). C1 to C6 indicate clusters represented in Fig. 1.

Fig. 4.

Differences in global DNA methylation status between TCN-active tumors and TCN-silent tumors. **A:** Top-center panel shows the distribution of the differentially methylated loci between tumor and non-tumor samples in stage-I HCC cases along the X-chromosome cytogenetic map. Each plot indicates the difference of methylation β -value ($\Delta\beta$ -values) ($P < 0.01$) between tumor

and non-tumor samples at respective locus; red plot, TCN-active tumors; blue plot, TCN-silent tumors. Top-right panel shows a histogram of $\Delta\beta$ -values presented in the top-center panel. Each bar indicates the number of loci in respective range of $\Delta\beta$ -value in TCN-active tumors (red) and TCN-silent tumors (blue). Bottom panel shows the distribution of the differentially methylated loci in stage-I HCC along the X-chromosome cytogenetic map. Green bar indicates the difference of β -value between TCN-active tumors and TCN-silent tumors at respective locus. Results of other chromosomes are presented in Fig. S6. **B:** Comparison of DNA methylation levels in stage-I HCC between TCN-active tumors (Y-axis) and TCN-silent tumors (X-axis) per respective functional/characteristic regions: TSS1500, TSS200, exon 1/5'UTR, and whole CpG islands. Only differentially methylated loci were plotted (FDR < 0.05). **C:** Violin-plots consisting of the distributions and probability densities of the methylation levels at the top 3000 TCN-related and TCN-unrelated loci in the indicated sample types; NT, non-tumor; I, stage-I; III, stage-III. **D:** Comparison of DNA methylation levels between TCN-active tumors (Y-axis) and TCN-silent tumors (X-axis) in stage-I LUAD.

C

TCN18 genes

Gene symbol	Chromosome
BAGE	21p11
C2orf51 (TEX37)	2p11
COX7B2	4p12
CT47B1	Xq24
DCAF8L1	Xp21
DCAF8L2	Xp21
DDX53	Xp22
FLJ36000	17p11
LIN28B	6q16
LUZP4	Xq23
MAGEB6	Xp21
RBMY	Yq11
SSX3	Xp11
SSX5	Xp11
TGIF2LX	Xq21
TGIF2LY	Yp11
TSPY	Yp11
VCY	Yq11

A **B** **C**

A

B

C

D