Biological and clinical significance of sex hormone-related gene expression in testicular cancer

Methods, Tables, Figures

Department of Urology, Medical University of Innsbruck

2024-06-20

# Analysis highlights

Two transcriptome data sets of testicular cancers containing both seminoma and NSGCT (nonseminomatous germ cell tumor) histology cancers, TCGA (1) and GSE99420 (2) were analyzed in the current report. Of note, because of strikingly different distribution of histological subtypes, and in particular significantly lower frequency of seminomas, we were not able to fully reproduce results of the current analysis in another transcriptomic data set, GSE3218 (3). The analysis results for the GSE3218 collective are available in the [R analysis pipeline of the project](https://github.com/PiotrTymoszuk/TesCA-hormones) and may be explored by an interesting reader. The demographic, clinical, and pathological characteristic of the TCGA and GSE99420 cohorts is provided in *Table 1*.

In our previous report, we characterized systemic pathological alterations of pituitary and gonadal hormone levels in a large local cohort of testicular cancer patients. We also delivered initial evidence that genes involved in production of gonadotropins and sex hormones are expressed in testicular cancer tissue at mRNA levels in the publicly available TCGA collective (4). Yet, clinical, prognostic, and biological consequences of presence of the hormone-producing machinery in malignant testicular tissue remained unclear. Our current report pursued two goals. First, we sought to explore cancer tissue expression and co-regulation patterns of 34 genes related to sex hormone regulation, synthesis, and metabolism reported in [‘Metabolism of steroid hormones’](https://reactome.org/content/detail/R-HSA-196071), [‘Estrogen biosynthesis’](https://reactome.org/content/detail/R-HSA-193144), and [‘Androgen biosynthesis’](https://reactome.org/content/detail/R-HSA-193048) Reactome pathways (**Table 2**). Second, we classified testicular cancer samples in respect to varying expression of those sex hormone-related genes by means of semi-supervised clustering, and characterized clinical, prognostic, biological, and pharmacological background of the resulting hormonal clusters.

In a comparison of expression levels of the sex hormone-related genes between NSGCT and seminoma, the *POMC* gene coding for an essential regulator of steroid biosynthesis, and genes involved in metabolism (*HSD17B1*, *HSD17B2*, *SRD5A1*) and transport of sex hormones (*SHBG*) were significantly upregulated in NSGCT in both the TCGA and GSE99420 cohorts. The *SRD5A3* gene coding for a testosterone - DHT (dihydro-testosterone) converting enzyme was the sole transcript upregulated in seminoma in both collectives (**Figure 1**, **Supplementary Table S1**). A more detailed histological subtyping according to ICD-O (international classification of diseases for oncology) was available for the TCGA cohort. In this data set, seminoma caner were characterized by the highest expression of *LHB* (luteinizing hormone subunit beta) and *SRD5A3*. High mRNA levels of genes coding for pituitary gonadotropin subunits *CGA* and *GNRH2* hallmarked mixed germ cell tumors and embryonal carcinomas. Expression of genes involved in general steroid biosynthesis (*STAR*, *POMC*, *CYP11A1*, *FDX1*) and the aromatase-coding gene *CYP19A1* was the highest in embryonal carcinomas. Pituitary hormone-coding *PRL* and *GNRH1* peaked in teratoma cancers, while *SHBG* mRNA levels were the highest in yolk sac tumors (**Supplementary Figure S1**, **Supplementary Table S2**).

A co-expression analysis revealed a highly co-regulated set of hub genes involved in general steroidogenesis (*CYP11A1*, *CYP17A1*, *STAR*, *HSD3B1*, *HSD3B2*) and synthesis of male and female hormones (*CYP19A1*, *HSD17B12*, *HSD17B3*, *SRD5A1*, *SRD5A2*) as well as their transport and catabolism (*SHBG*, \_\_HSD17B2\_) (**Supplementary Figure S3**, **Supplementary Table S3**). Those findings combined with the results of differential gene expression analysis for testicular cancer histologies suggest that testicular cancers in general and NSGCT in particular are equipped with the complete synthetic and regulatory machinery of steroid sex hormone metabolism.

By means of hard-threshold regularized KMEANS clustering (5), we assigned cancer samples of the training TCGA cohort into four clusters defined in respect to differing expression of the sex hormone-related genes (**Table 2**, **Supplementary Figure S3A**). Subsequently, we predicted the hormonal cluster assignment of cancer samples of the GSE99420 cohort with a Random Forest classifier (6,7) fed with cancer tissue levels of the cluster-defining transcripts. There were no significant differences in distribution of sizes of the hormonal clusters between the training and test collective (**Supplementary Figure S3B**). An analysis of pairwise distances between observations of each of the TCGA and GSE99420 cohort and cross-distances between the TCGA and GSE99420 samples indicated sufficient separability of the clusters in each of the cohorts and similarity of the corresponding clusters of the training and test cohort clusters (**Supplementary Figure S3C**). Still, it has to be noted, that in terms of numeric statistics of cluster separability, explanatory performance, and nearest neighborhood preservation, clustering of cancers samples was more efficient in the training TCAG cohort than in the test GSE99420 collective (**Supplementary Table S4**).

Hormonal cluster #1 consisted primarily of NSGCT malignancies (12 to 16% of samples, mixed and teratoma subtypes) characterized by high expression of the pituitary hormones *PRL* and *GNRH1*, *HSD17B2* coding for sex hormone-deactivating enzyme, and *SRD5A1* coding for a testosterone - DHT converting enzyme. The largest hormonal cluster #2 (42 to 50% of samples) included predominantly seminomas and was hallmarked by high expression of the testosterone - DHT converting enzyme *SRD5A3*. Hormonal cluster #3 comprised of NSGCT (8.3 to 18% of samples, mixed histology, embryonal carcinoma, yolk sac tumors) with high expression of *CGA* gonadotropin subunit as well as genes involved in estrogen and estradiol synthesis and inter-conversion (*CYP19A1*, *HSD17B12*, *HSD17B1*), and *SHBG* involved in hormone binding and transport. Finally, 23 to 30% of cancer samples were classified as hormonal cluster #4, which consisted primarily of NSGCT (embryonal carcinoma, mixed histology, yolk sac) with a small fraction of seminomas. The key feature of this cluster was increased expression of genes of general steroid (*STAR*, *POMC*, *CYP11A1*, *CYP17A1*, *HSD3B2*) and testosterone biosynthesis (*HSD17B3*) (**Figure 2**, **Supplementary Figure S4**, **Supplementary Table S5**).

Concerning clinical and pathological features of the hormonal clusters, patients in clusters #1 and #3 were significantly younger as compared with the remaining individuals. Stage I, serum marker-negative cancers were the most frequent in cluster #2, which is consistent with the predominant seminoma histology. By contrast, S2/S3 and stage III cancers were the most common in cluster #3 (**Figure 3A**, **Supplementary Table S6**). Accordingly, cluster #2 and #3 patients exhibited, respectively, the best and poorest progression-free survival (p = 0.042). Analogical yet not significant effects were observed for relapse-free survival (ns (p = 0.056)). Hormonal cluster #1 patients were exposed to intermediate-to-high risk of progression and relapse. Progression and relapse risk in cluster #4 was low-to-intermediate (**Figure 3B**, **Supplementary Figure S5**).

In regularized multi-parameter modeling of progression-free survival by RIDGE Cox proportional hazard regression (8,9), appending of canonical clinical risk factors (age, serum marker status, histological subtype) with the hormonal cluster information had virtually no effect on the model’s accuracy (clinical explanatory factor model: C-index = 0.67, clinical factors and hormonal clusters: C-index = 0.65) but marginally improved its calibration (clinical factors: integrated Brier score [IBS] = 0.22, clinical factors and hormonal clusters: 0.21). In the combined clinical/hormonal cluster model of progression-free survival, each of S2/S3 marker status, hormonal cluster #1 or #3 assignment, and teratoma histology, were independently associated with roughly 25 - 30% increase of progression risk (**Supplementary Figure S5**, \_Supplementary Table S7\_\_).

As predicted by immunedeconvolution of the gene expression data sets by xCell and MCP Counter algorithms (10,11), cluster #1 and, to a lesser extend, cluster #3, displayed significantly higher predicted fibroblast levels as compared with clusters #2 and #4. In turn, cluster #2 cancers were predicted to be abundantly infiltrated by T and B cells. Those results were also corroborated in the TCGA cohort by an analysis of cell fractions predicted by the QuanTIseq procedure (12) (**Figure 4A**, **Supplementary Figure S6**, **Supplementary Table S8**).

Gene set variation analysis, i.e. comparison of scores of Reactome pathway gene signatures revealed profound differences between the hormonal clusters (all signatures: n = 1616, regulated between the clusters in the TCGA and GSE99420 cohorts: n = 719, **Supplementary Table S9**). This points towards distinct biological phenotypes of the hormonal clusters. In more detail, extracellular matrix- (ECM) and growth factor signaling-related gene signature scores were the highest in cluster #1 followed by cluster #3. Gene signatures of antigen processing, inflammatory signaling, apoptosis, and DNA repair demonstrated the highest levels in cluster #2. Cluster #3 was characterized by elevated scores of signatures of FGFR/ERBB/PI3K/AKT, lipid metabolism, steroid hormone and estrogen metabolism. Finally, fatty acid, nucleotide, sugar and androgen metabolism along with cell cycle, mitosis, RNA turnover, apoptosis and autophagy were inferred as the key biological processes primarily for cluster #4 (**Figure 4B**).

The biological differences between the hormonal clusters were confirmed by an analysis of differential gene expression in particular hormonal clusters as compared with the respective mean gene expression in the TCGA and GSE99420 cohorts (**Supplementary Table S10**). A total of 2000 genes were found to be differentially regulated in cluster #1 as compared with the data set mean in both investigated cohorts; the numbers of such common differentially regulated genes were 2829, 1677, and 468 for clusters #2, #3, and #4, respectively. The top strongest upregulated genes in cluster #1 shared by the TCGA and GSE99420 collective included *SULT1E1*, a gene coding for sex hormone sulfonylating enzyme, genes of keratins (*KRT5*, *KRT17*), retinoid transporter (*TTR*), a transcription factor co-regulating the estrogen-responsive transcriptome (*FOXA1*), and major ECM components (*COL2A1*, *ELN*). Gene coding for testis-specific antigens (*PRAME*, *LUZP4*), testis-specific signaling/transcription mediators (*TCL1B*, *RNF17*, *EZHIP*), RNA binding proteins (*CPEB1*, *LUZP4*) were found among the top activated genes in cluster #2. Genes strongly up-regulated in cluster #3 included *AFP* coding for the testicular cancer marker alpha-fetoprotein, apolipoproteins, lipases and lipoprotein receptors (*APOB*, *APOA2*, *APOA4*, *APOC3*, *PLA2G12B*, *LRP2*), proteases, protease inhibitors and protease receptors (*CST1*, *CCKBR*, *CTSE*), ECM components (*AHSG*, *FGB*, *VTN*, *GPC*), retinoid transporter (*TTR*), and ketone metabolism protein gene (*HMGCS2*). The top regulated features in cluster #4 as compared with the cohort mean comprised genes participating in development (*ARX*, *DLK1*), cell cycle in the testis and spermatogenesis (*PRM2*, *CCNA1*, *FATE1*, *TAF7L*, *CATSPERZ*), post-transcriptional expression regulation (*PIWIL3*), and pituitary hormone secretion (*INHA*) (**Supplementary Figure S7**).

Biological process gene ontology (GO) enrichment analysis was performed for sets of differentially up- or down-regulated genes in particular clusters (13) (**Supplementary Table S11**). Next, GO terms found to be significantly enriched in the TCGA and GSE99420 cohort were grouped by their semantic similarity in respect to Wang distances (14,15). Such analysis revealed regulation of genes and biological processes of RNA turnover, organ development, reproduction, proliferation, ECM and mesenchymal cell phenotype in cluster #1. GO terms related to ECM, lipoprotein metabolism, TGF/BMP/WNT signaling, cell adhesion and motility, transcription, organ and tissue development were found significantly enriched among genes up- and downregulated in cluster #2. GO terms significantly enriched in cluster #3 were associated with organ, sex and mesenchymal differentiation, blood coagulation, growth factor and NOTCH/WNT signaling, wound healing, adhesion and motility (**Supplementary Figure S8**). Of note, we could not identify significantly enriched biological process GO terms in cluster #4 shared by the TCGA and GSE99420 cohort.

A subsequent comparison of activity of transcriptional regulons (16,17) (**Supplementary Table 12**), i.e. set of genes regulated by common transcription factors, in the hormonal clusters as compared with the cohort means revealed activity of SMAD1/3-, CTNNB1-, and HIF1A-responsive transcriptome participating among others in ECM deposition, WNT signaling, and hypoxia response in cluster #1. REST, CIITA, MECP2, and HOXA2 regulons associated with neural and embryonal development and inflammation were found to be significantly activated in both the TCGA and GSE99420 cohort. ESR1, HIF1A, SP1, AP1, JUN, and HNF4A were identified as common activated regulons in cluster #3 with functions in estrogen, hypoxia and stress response, and in control of the sex hormone transporter SHBG expression. POU5F1, CREB1, SP1-responsive transcriptome involved in RNA turnover, teratoma formation, and steroid synthesis were found to activated in cluster #4 caners of the TCGA and GSE99420 cohorts (**Figure 5A**, **Supplementary Table S12**). Concerning signaling pathways included in the PROGENy knowledge model (17,18), high activity of hypoxia and TGF- signaling, and suppression of MAPK and JAK/STAT signaling as compared with the cohort average was predicted for cluster #1 in both the TCGA and GSE99420 cohort based on whole transcriptome expression regulation estimates. Activation of JAK/STAT signaling with concomitant suppression of hypoxia, TGF-, EGFR, and p53 signaling was predicted for cluster #2. Cluster #3 was characterized by significantly higher activity of hypoxia, WNT, TGF-, and EGFR signaling pathways as compared with the cohort mean. Finally, activation of EGFR and estrogen signaling pathways was predicted for cluster #4 in both investigated testicular cancer collectives (**Figure 5B**, **Supplementary Table S13**).

Immune cell infiltration and inflammatory processes were found to differ between the hormonal clusters. In line with increased infiltration of cluster #2 cancers by T cells, clinically relevant immune checkpoint genes *CTLA4*, *TIGIT*, and *PDCD1* were found to be expressed in this cluster at the highest levels (**Supplementary Figure S9**).

Cancer and testis antigens (CTA) represented a highly prominent group of genes significantly differentially regulated between the hormonal clusters. Out of 1170 CTA genes reported in literature (19,20), 282 transcripts were expressed at significantly different levels in the hormonal clusters in both the TCGA and GSE99420 cohort. Interestingly, a vast majority of those common regulated CTA transcript exhibited the highest expression levels in cluster #2 followed by cluster #4. By contrast, only few CTA genes were found to be upregulated in cluster #3 (**Supplementary Figure S10**). This may suggest a normal tissue-like testicular differentiation of cluster #2 seminomas and, in part, cluster #4 malignancies, and extensive de-differentiation of cluster #1 and cluster #3 tumors. This may also reflect high immunogenicity of cluster #2 cancers in line with their T and B cell rich tumor microenvironment (21).

ERBB and FGFR signaling were identified as important biological hallmarks of hormonal clusters #1 and #3. In a detailed analysis of ERBB- (*EGFR*, *ERBB2*, *ERBB3*, *ERBB4*) and FGFR- (*FGFR1*, *FGFR2*, *FGFR3*, *FGFR4*) coding genes and genes coding for their ligands, we could observe peak mRNA levels of *EGFR* and *ERBB2* in cluster #1 followed by cluster #3. *FGFR1* and *FGFR4* genes were upregulated in cluster #3 followed by clusters #4 and #2. *FGFR3* was expression was the highest in cluster #2. Among ERBB receptor ligands, solely *NRG3* was differentially regulated in the hormonal clusters and its expression was significantly elevated in cluster #2. Interestingly, FGF-coding genes, *FGF2*, *FGF4*, *FGF5*, *FGF10*, *FGF17*, and *FGF23* were over-expressed in clusters #1, #3, and #4, suggestive of an auto- or paracrine growth-promoting signaling via FGFR1 and FGFR4 (\_\_Supplementary Figure S11A\_).

Regarding transcripts of receptors of gonadotropins and sex hormones, *FSHR* (FSH receptor), *LHCGR* (LH receptor), *ESR2* (estrogen receptor beta), and *AR* (androgen receptor) were found to be differentially regulated between the hormonal clusters. In more details, *FSHR* was upregulated predominantly in cluster #3 followed by clusters #1 and #4, whereas *LHCGR* was detected at the highest levels in cluster #4. *ESR2* was expressed predominantly abundantly by cluster #2 and #4 cancers. *AR* was expressed significantly higher in clusters #1, #2, and #4 than by cluster #3 tumors (**Supplementary Figure S11B**). Out of 87 published high confidence estrogen responsive genes (22), 42 were found to be significantly differentially regulated between the hormonal clusters in both the TCGA and GSE99420 cohorts. The majority of those common regulated estrogen-responsive genes was expressed in clusters #1 and #3 at the highest levels (**Supplementary Figure S12A**). Among 177 genes (23), mRNA levels of 46 of them differed significantly between the hormonal clusters in both analyzed cohorts. Most of them peaked in clusters #1, #3, and #4, but a prominent group of such common differentially regulated androgen-responsive genes was highly specific for cluster #2 as well (**Supplementary Figure S12B**). Collectively, the results of the analysis of differential gene expression gonadotropin and sex hormone genes, and estrogen- and androgen-regulated transcriptomes suggest that the FSH - estrogen axis may be active in testicular cancers assigned to hormonal clusters #1 and #3. Analogically, co-expression of the major testosterone regulator LH receptor and genes involved in steroidogenesis and testosterone synthesis indicate that the LH - testosterone circuit may be functional in cluster #4 testicular tumors.

Extensive infiltration by fibroblasts and ECM deposition in cluster #1 and, to a lesser extent, in cluster #3 were inferred from the immunedeconvolution data, gene set variation analysis of Reactome pathways, and GO enrichment analysis. We intended to investigate modulation of matrisome, i.e. genes coding for structural proteins of ECM, ECM regulators, and ECM-associated secreted factors (24,25), between the hormonal clusters. Out of 1001 matrisome genes available for analysis, 333 were found to be significantly differentially regulated between the hormonal clusters in the TCGA and GSE99420 cohort. As shown in **Supplementary Figure S13**, sizable fractions of collagens (e.g. *COL1A1*, *COL3A1*, *COL5A1*), ECM glycoproteins (e.g. laminins and elastin), and proteoglycans (e.g. *DCN*, *OGN*) were found to be strongly upregulated in cluster #1. By contrast, only few matrisome transcripts were found abundantly expressed in cluster #2. Fibrinogen chain-coding transcripts (*FGA*, *FGB*, *FGG*) as well as transcripts of ECM regulators, e.g. proteases and protease inhibitors (serpins, *ADAMTS18*) were expressed in cluster #3 at highest levels followed by cluster #1. Among few matrisome transcripts specific for cluster #4, secreted factors such as *CSH1*, *GDF3*, *FGF2*, *FGF4*, and *IL23* involved in prolactin, growth factor, and inflammatory signaling, constituted the largest gene set (**Supplementary Figure S13**).

Activity of RECON2 model metabolic reactions in the hormonal clusters as compared with the respective cohort averages was assessed by a Monte Carlo algorithm provided with differential expression estimates and errors for all available genes (26,27) (**Supplementary Table S14**). To assess activation or inhibition of metabolic pathways, we analyzed enrichment of RECON metabolic subsystems with significantly activated and inhibited reactions in particular clusters (**Supplementary Table S15**). This enrichment analysis suggested activation of xenobiotic, vitamin D, arachidonic acid and inositol phosphate metabolism, heparan sulfate degradation, and nucleotide inter-conversion in cluster #1. In cluster #2, transport-related subsystems, folate and vitamin E turnover were significantly enriched with activated metabolic reactions. For cluster #3, the enrichment analysis predicted an activation of androgen and estrogen metabolism, fatty acid synthesis, extracellular transport, and methionine and cysteine metabolism. In turn, sugar catabolism and oxidative energy metabolism pathways were predicted to be significantly inhibited. In cluster #4, metabolic subsystems of nucleotide salvage, steroid, pyruvate, and eicosanoid turnover were predicted to be activated, while cytochrome, sphingolipid, and fatty acid oxidation were estimated to be suppressed (**Supplementary Figure S14**). In a detailed analysis of citric acid cycle, we could observe a sustained inhibition of multiple reactions of the pathway in clusters #1 and #3; solely isocitrate dehydrogenase reactions were predicted to be activated in these hormonal clusters. In a similar analysis of oxidative phosphorylation, mitochondrial complexes I, III, and IV were found to be significantly inhibited in hormonal clusters #1, #2, and #3 as compared with the respective cohort average. This inhibition was the largest in cluster #1. In turn, in cluster #4 only complexes I and II were predicted to be inhibited (**Supplementary Figure 15**). Collectively, for clusters #1 and #3, a general suppression of oxidative sugar metabolism via citric acid cycle and oxidative phosphorylation can be inferred.

A total of 50 reactions of the RECON metabolic subsystems ‘Steroid metabolism’ and ‘Androgen and estrogen synthesis and metabolism’ were predicted to be significantly regulated in at least one of the hormonal cluster of the TCGA and GSE99420 cancers as compared with the respective cohort averages (total reactions: n = 104). In cluster #1, multiple reactions of general steroid biosynthesis, corticosteroid and estradiol formation were estimated to be suppressed, while reactions of steroid sulfonylation (DHEA sulfotransferase) and sex hormone inactivation (HSD17B2, HSD17B8) were predicted to be activated. For cluster #2, general suppression of sex hormone synthesis can be postulated. In turn, reactions of estrogen formation (aromatase), conjugation (e.g. DHEA sulfotransferase or glutathione transferase reactions RE2319 and RE2319), and inactivation (HSD17B2, HSD17B8) were predicted to be activated in cluster #3 suggestive of functional regulation of synthesis and catabolism of estrogens in those testicular cancers. In cluster #4, the modeling data revealed an increased activity of general steroid synthesis including the first, rate-limiting steps of the pathway (CYP11A1, CYP17A1). Furthermore, the testosterone synthesis reaction catalyzed by HSD17B3 was found activated, while the sex hormone inactivating reactions HSD17B2 and HSD17B8 were predicted to be inhibited. This suggests functionality of steroid and male hormone production pathways specifically in cluster #4 testicular cancers (**Supplementary Figure S16**).

For the TCGA cohort cancer samples, levels of 194 proteins of relevance for cancer biology were measured by a reverse phase protein array (1). Among them, 125 were found to be significantly differentially regulated in at least one hormonal clusters as compared with the respective whole data set average expression (**Supplementary Table S16**). Specifically in clusters #1 and #3, active phosphorylated forms of SRC and YAP were upregulated. Of interest, this signaling axis was demonstrated to integrate signals from HIPPO, WNT, and growth signaling pathways and drive lung cancer progression (28,29). Other proteins of relevance for pro-oncogenic signaling, EGFR and ERBB2 with their active tyrosine phosphorylated forms were found significantly upregulated in cluster #1 and, to a lesser extend, also in cluster #3. In turn, fibronectin protein levels were the highest in cluster #3. Cluster #2 was characterized by increased levels of the KIT oncogene protein and proteins involved in DNA repair and apoptosis (e.g. PARP1, CHEK2). The negative regulator of hypoxia response VHL, protease inhibitor SEPRINE1, and the adhesion molecule E-cadherin were specifically upregulated at protein level in cluster #4 (**Figure 6**).

Elevated expression of proteins and phosphoproteins of downstream of EGFR/ERBB in cluster #3 and, in particular, in cluster #1 suggests EGFR/ERBB, SRC, c-RAF/MAPK, and PI3K/AKT/mTOR/RICTOR signaling as a major tumor-promoting circuits of these cancers. Those signaling pathways may be also responsible for downregulation of the pro-apoptotic BIM and BAX proteins (30) in clusters #1 and #3. Of note, in cluster #2, B-RAF, C-RAF, and mTOR proteins were upregulated despite suppression of activated forms of ERBB2 and EGFR, which may suggest an alternative mode of activation of mTOR and MAPK signaling, e.g. by KIT (31). Cluster #2 was also characterized by significant upregulation of pro-apoptotic BIM and BAX proteins (**Supplementary Figure S17**). In general, upregulation of of multiple DNA damage sensors and apoptosis activators at protein levels in cluster #2 is congruent with the Reactome pathway gene signature analysis results, and may reflect predisposition of those cancers to cell death inducing signals and anti-cancer treatment.

An analysis of co-expression of proteins in the hormonal clusters #1 and #3 of TCGA testicular cancers underline the relevance of EGF/ERBB, MAPK and AKT/mTOR signaling pathways, whose members were identified as prominent hubs of communities of co-regulated proteins. Interestingly, diverse isoforms of PKC, p38, and JNK were found to be hub proteins in clusters #2 and #4, which may indicate their roles in fueling progression of those testicular cancers. Furthermore, ER-alpha, its phosphorylated form ER-alpha pS118, and AR were also put forward as highly connected proteins i all hormonal clusters (**Supplementary Figures S18** and **S19**). This later finding delivers another piece of evidence of functionality and relevance of sex hormone signaling in testicular cancer.

No significant differences in total mutation burdens, numbers of mutations in protein-coding genes, and gene amplifications were observed in the hormonal subsets of TCGA cancer samples. In turn, significantly higher microsatellite instability scores as well as increased numbers of gene deletions were detected in clusters #1 and #3 as compared with clusters #2 and #4. Effect size of those differences was moderate-to-large (**Supplementary Figure S20**, **Supplementary Table S17**).

Somatic mutations of *KIT*, *KRAS*, and *NRAS* genes were the most frequent in cluster #2, and basically absent in clusters #1 and #3; this effect was significant for the *KIT* gene. Mutations of *PTMA* and *LZTR1* were, in turn, most common in cluster #4, yet the differences in frequencies between the hormonal clusters were not significant. Deletions of *GRID2*, *PDE4D*, *JARID2*, and *DMD* were the most frequent in cluster #1 followed by clusters #3 and #4, and absent in cluster #2. Those differences were, however, not significant. Amplifications of the chromosomal region 12q15 harboring, among others, the *MDM2* proto-oncogene, were found to be highly specific for cluster #3 (significant effect). Amplifications of the chromosomal region 20q11 containing the apoptosis modulator *BCL2L1* were detected predominantly in cluster #1 and, at lower frequencies, in cluster #2. This difference in rates of 20q11 amplification between the hormonal clusters was, however, not significant (**Supplementary Figure S21**, **Supplementary Table S18**).

By means of RIDGE linear machine learning models (9,32,33) trained with results of two large in-vitro drug sensitivity screening experiments GDSC and CTRP (34,35), we predicted response to, respectively, 305 and 214 anti-cancer compounds for single cancer samples of the TCGA and GSE99420 cohorts. Next, we compared the drug response predictions in form of log IC50 (i.e. 50% inhibitory concentration, GDSC) or AUC (area under the dose-response curve, CTRP2) between the hormonal clusters by one-way ANOVA with effect size statistic. Differences in the predicted response to a compound between a particular clusters and the respective cohort average was assessed by one-sample T test. P values of those two tests were corrected for multiple testing with the false discovery rate (FDR) method. Differential drug response was considered for pFDR(ANOVA) < 0.05, , and pFDR(T test) < 0.05, which translates to moderate-to-strong significant differences between the clusters and significant differences between the mean response in the clusters and the cohort mean. Among the GDSC-trained data, predicted responses to 124 compounds differed between the hormonal clusters of the TCGA and GSE99420 cohorts. For the CTRP2-trained predictions, differential response to 84 compounds was observed in both collectives (**Supplementary Table S19**). Independently of the training data, cluster #1, and , to a lesser extend cluster #3 cancers, were predicted to be resistant against numerous compounds interfering with DNA synthesis (e.g. folate anti-metabolites: pemetrexed, metothrexate, telomerase inhibitors: MST-312, Telomerase Inhibitor IX), cytoskeleton and cell cycle (e.g. vinblastine, taxane drugs, cell cycle checkpoint inhibitors: rigosertib, BI-2536, MK-1775, PHA-793887, SNS-032, alvocidib), and epigenetics (belinostat, entinostat). Cluster #2 and #4 malignancies were predicted to be, respectively, highly and moderately responsive to those class of drugs. The best response to apoptosis modulators, e.g. obatoclax, venetoclax, PAC-1 or BRD-K35604418, was predicted for cluster #2, while resistance to those substances was predicted for clusters #1, #3, and #4. Of note, this later phenomenon fits well to the increased expression of DNA damage sensors and pro-apoptotic proteins by cluster #2 cancers at both mRNA and protein level. In line with the highly active receptor tyrosine kinase and growth factor signaling, clusters #1 and #3, were estimated to respond well towards tyrosine kinase inhibitors (TKI) such as multi-TKI- and SRC-targeting ponatinib, dasatinib and saracatinib, VEGFR-targeting cabozantinib, foretinib and tivozanib, FGFR inhibitors AZD4547 and PD173074, or EGFR/ERBB inhibitors neratinib and cetuximab. Cluster #2 tumors were in turn predicted to be resistant to such compounds (**Figure 7**, **Supplementary Figure S22**).

# Methods

Bioinformatic and statistics analyses and visualizations were done with R version 4.2.3. For detailed information on methods and software, please refer to Supplementary Material.

## Data sources

Publicly available TCGA (source: [cBioportal](https://www.cbioportal.org/), total cancers: n = 149, seminoma: 43%, nonseminomatous germ cell tumor [NSGCT]: 57% of samples) (1) and GSE99420 (source: [GEO](https://www.ncbi.nlm.nih.gov/geo/), total cancers: n = 60, seminoma: 50%, nonseminomatous germ cell tumor [NSGCT]: 50%) or samples) transcriptomic study collectives were analyzed in the current report. For the TCGA cohort detailed information of demographics, clinical features, histology and tumor pathology, and survival were provided along with RNA sequencing gene expression quantification, protein quantification results by reverse phase protein array, somatic mutations, gene amplifications and deletions (1). For the GSE99420 cohort, information of cancer histology (seminoma/NSGCT) and presence of relapse was provided together with transcriptome measurements with an Illumina microarray platform. Characteristic of the cohorts is provided in **Table 1**.

Normalized, -transformed expression data for genes and proteins provided by authors were used in analyses. Sex hormone-related genes used for characteristic and classification of testicular cancers consisted of genes coding for pituitary gonadotropins, and genes involved in steroid biosynthesis, inter-conversion, transport and catabolism of estrogens and androgens. Those genes were extracted from [‘Metabolism of steroid hormones’](https://reactome.org/content/detail/R-HSA-196071), [‘Estrogen biosynthesis’](https://reactome.org/content/detail/R-HSA-193144), and [‘Androgen biosynthesis’](https://reactome.org/content/detail/R-HSA-193048) Reactome pathways and are listed in **Table 2**. mRNA levels of the sex-hormone genes were adjusted for the cohort/batch effect with the ComBat algorithm (36) prior to differential gene expression, co-regulation, and clustering analyses.

Based on whole-transcriptome gene expression, predictions of anti-cancer drug response in form of log IC50 and AUC (are under the dose-response curve) were made by RIDGE linear regression models (9,32,33) provided by R package [*htGLMNET*](https://github.com/PiotrTymoszuk/htGLMNET) and trained with results of in vitro drug screening experiments GDSC (34) and CTRP2 (35). Estimates of non-malignant cell contents in cancer specimens were computed by *QuanTIseq*, *xCell*, and *MCP Counter* immunedeconvolution algorithms (10–12). Single sample gene set enrichment analysis scores (ssGSEA scores) were computed for Reactome pathway gene signatures (source: [MSig database](https://www.gsea-msigdb.org/gsea/msigdb/index.jsp), version 7.5.1) with the *GSVA* algorithm (37).

Cancer and testis antigen genes, estrogen- and androgen-responsive genes were extracted from [CTpedia database](http://www.cta.lncc.br/) and literature reports (19,20,22,23). Annotation and classification of matrisome genes, i.e. features coding for proteins of extracellular matrix (ECM), was done with *MatrisomeAnalyzeR* R package (24,25). Other genes of interest were manually searched in the [NCBI Gene database](https://www.ncbi.nlm.nih.gov/gene/).

## Statistical analysis

Descriptive statistics presented for numeric variables are medians, interquartile ranges, ranges and numbers of complete observations. For qualitative variables, percentages and counts of observations in the categories are presented. Statistical significance for differences in numeric variables between strata were investigated by two-tailed T tests, Mann-Whitney tests, Kruskal-Wallis tests, and one-way ANOVA with Cohen’s d, r and effect size statistics. Statistical significance for differences in distribution of categories of categorical variables between strata was determined by test with Cramer’s V effect size statistics. P values were adjusted for multiple comparisons with the false discovery rate (FDR) method (38) separately for the analysis step and cohort. If not indicated otherwise, effects with FDR-adjusted p values < 0.05 were considered significant.

## Clustering

Cancer samples in the TCGA training cohort were assigned to ‘hormonal clusters’ in respect to ComBat-adjusted -transformed expression of the sex-hormone related genes by the hard-threshold regularized KMEANS algorithm (5). Choice of the cluster number (k = 4) was motivated by the maximum of mean silhouette width, i.e. a metric of cluster separation (39). Assignment of cancer specimens in the GSE99420 test cohort to the hormonal clusters was accomplished with a Random Forest classifier (6,7) employing ComBat-adjusted -transformed mRNA levels of the cluster-defining sex hormone-related genes.

Quality control of this semi-supervised clustering included comparison of cluster size distribution in the training and test cohort, analysis of pairwise distances between observations in the clusters, and comparison of expression levels of the cluster-defining genes between the hormonal clusters.

## Differential gene and protein expression, predicted drug response in the clusters

Due to strongly non-normal distribution of some sex hormone-related mRNAs, differential expression of the sex hormone-related genes in tumor histologies was assessed by non-parametric Mann-Whitney (effect size: r) and Kruskal-Wallis tests (effect size: ).

Differences of -transformed gene expression (TCGA and GSE99420), -transformed protein levels (TCGA), and predicted drug response metrics (IC50 and AUC, TCGA and GSE99420) between the hormonal clusters were explored with a two step procedure. First, the difference between the hormonal clusters was assessed by one-way ANOVA with effect size statistic (function test\_anova, package [*microViz*](https://github.com/PiotrTymoszuk/microViz)). Second, the difference between the cluster mean of a gene, protein, or sensitivity metric and the respective mean value in the entire cohort was investigated by one-sample T test (function avg\_deviation(), package [*microViz*](https://github.com/PiotrTymoszuk/microViz)). Features with pFDR(ANOVA) < 0.05, 0.14, and pFDR(T test) < 0.05 were considered significantly differentially regulated.

## Demographic, clinical, pathological, and tumor microenvironment characteristic of the hormonal clusters

Differences in demographic and clinical features, and estimates of non-malignant cell content between the hormonal clusters were investigated by Kruskal-Wallis and tests with and Cramer’s V effect size statistics. Differences in relapse- and progression-free survival between the hormonal clusters of the TCGA cohort were assessed by Kaplan-Meier analysis and Peto-Peto tests. Multi-parameter modeling of progression-free survival was accomplished by regularized RIDGE Cox regression (8,9).

## Reactome pathways, GO enrichment, regulons, signaling pathway, and metabolic reaction analysis

ssGSEA scores of Reactome pathway gene signatures were compared between the hormonal clusters by one-way ANOVA with effect size statistic. Gene signatures differing between the clusters with moderate-to-large effect size, i.e. with pFDR < 0.05 and , were considered significant.

Biological process gene ontology (GO) term enrichment for genes found to be differentially regulated between the hormonal clusters was explored by a modified *goana* algorithm (13). For semantic clustering of GO terms found to be significantly enriched in both the TCGA and GSE99420 cohort, a modified implementation of the *GOSemSim* algorithm (14,15) was utilized.

Activity of transcriptional regulons and signaling pathways of the *collecTRI* and *PROGENy* knowledge models was compared between the hormonal clusters by linear modeling tools of the *decoupleR* package (16–18) operating with T statistics of whole-transcriptome differential gene expression. Activity of RECON2 metabolic reactions in the hormonal clusters was assessed by Monte Carlo simulations fed with fold-regulation estimates of differential expression and their standard error obtained for all available genes (R package [*biggrExtra*](https://github.com/PiotrTymoszuk/biggrExtra)) (26,27). Enrichment of RECON metabolic subsystems with significantly activated and inhibited reactions was assessed by comparing the regulated reactions sets with random draws from the entire reaction pool.

## Co-regulation networks

Co-regulation of sex hormone-related genes (TCGA and GSE99420) or proteins (TCGA) was investigated by pairwise Spearman’s and Pearson’s r correlation coefficients. Matrices of correlations selected by correlation coefficient cutoffs (genes: 0.3, proteins: r 0.5) were converted to graph objects and analyzed as networks. Graph node importance statistics (degree, betweenness, hub score) and node communities were explored by *igraph* package tools (40). Networks were visualized with *ggnetwork* R package (41).

# Data and code availability

Data from publicly available sources were analyzed. The R analysis pipeline is available as a [GitHub repository](https://github.com/PiotrTymoszuk/TesCA-hormones).

# Tables

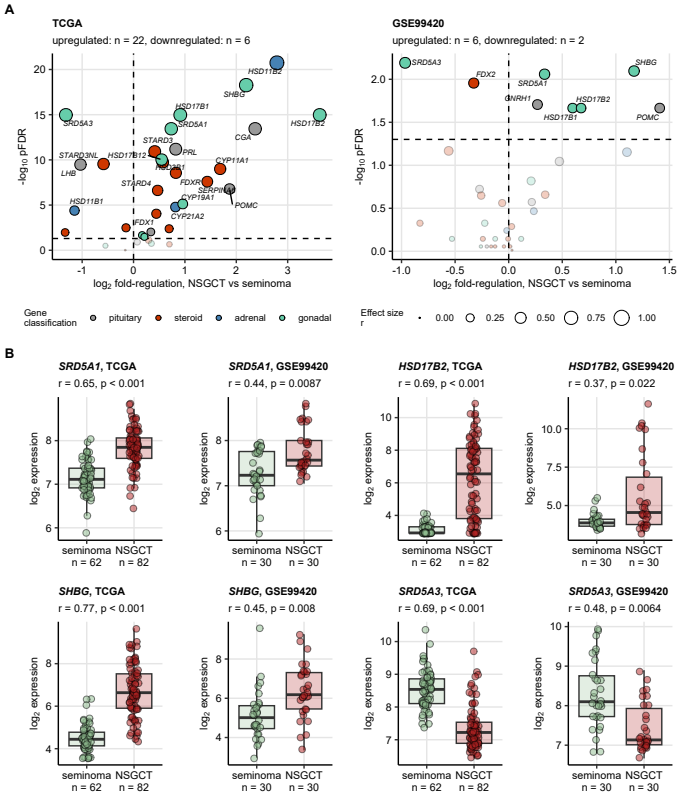
Table 1: Characteristic of the investigated cohorts of testicular cancer patients. Numeric variables are presented as medians with interquartile ranges and ranges. Qualitative variables are shown as percentages and counts of the categories within the complete observation set.

| **Variablea** | **TCGA** | **GSE99420** | **Significanceb** | **Effect sizeb** |
| --- | --- | --- | --- | --- |
| Age | 31 [IQR: 26 - 37] range: 14 - 67 complete: n = 133 |  |  |  |
| Race/Ethnicity | Asian: 3.1% (4) Black or African American: 4.7% (6) White: 92% (118) complete: n = 128 |  |  |  |
| Tumor stage | I: 84% (102) II: 9% (11) III: 7.4% (9) complete: n = 122 |  |  |  |
| Metastasis stage | M0: 97% (114) M1: 3.4% (4) complete: n = 118 |  |  |  |
| Node stage | N0: 79% (46) N1: 17% (10) N2: 3.4% (2) complete: n = 58 |  |  |  |
| IGCCCG | good: 74% (32) intermediate: 21% (9) poor: 4.7% (2) complete: n = 43 |  |  |  |
| Histology | seminoma: 43% (62) NSGCT: 57% (82) complete: n = 144 | seminoma: 50% (30) NSGCT: 50% (30) complete: n = 60 | ns (p = 0.45) | V = 0.064 |
| Histology, ICD-O | seminoma: 49% (65) germinal mixed histology: 20% (27) benign teratoma: 3.8% (5) embryonal carcinoma: 20% (27) malignant teratoma: 2.3% (3) teratocarcinoma: 1.5% (2) yolk sac cancer: 3% (4) complete: n = 133 |  |  |  |
| Marker status | S0: 36% (43) S1: 31% (37) S2: 29% (34) S3: 4.2% (5) complete: n = 119 |  |  |  |
| Neoadjuvant therapy | no: 100% (133) complete: n = 133 |  |  |  |
| Radiation | 16% (21) complete: n = 130 |  |  |  |
| Follow-up, months | 1300 [IQR: 680 - 2700] range: 3 - 7400 complete: n = 133 |  |  |  |
| Death | 3% (4) complete: n = 133 |  |  |  |
| Relapse | 25% (33) complete: n = 133 | 50% (30) complete: n = 60 | p = 0.002 | V = 0.25 |
| Progression | 26% (34) complete: n = 133 |  |  |  |
| aIGCCCG: International Germ Cell Cancer Collaborative Group risk strata; ICD-0: international classification of diseases for oncology, histological subtype. | | | | |
| bNumeric variables: Kruskal-Wallis test with η² effect size statistic. Categorical variables: χ² test with Cramer's V effect size statistic. P values corrected for multiple testing with the false discovery rate method. | | | | |

Table 2: Sex hormone-related genes of interest.

| **Gene classification** | **Gene symbol** | **Entrez ID** |
| --- | --- | --- |
| pituitary | *GNRH1* | 2796 |
| *GNRH2* | 2797 |
| *PRL* | 5617 |
| *CGA* | 1081 |
| *LHB* | 3972 |
| *POMC* | 5443 |
| steroid | *STAR* | 6770 |
| *STARD3* | 10948 |
| *STARD3NL* | 83930 |
| *STARD4* | 134429 |
| *TSPO* | 706 |
| *TSPOAP1* | 9256 |
| *CYP11A1* | 1583 |
| *CYP17A1* | 1586 |
| *FDX1* | 2230 |
| *FDX2* | 112812 |
| *FDXR* | 2232 |
| *HSD3B1* | 3283 |
| *HSD3B2* | 3284 |
| *SERPINA6* | 866 |
| adrenal | *CYP21A2* | 1589 |
| *HSD11B1* | 3290 |
| *HSD11B2* | 3291 |
| gonadal | *HSD17B1* | 3292 |
| *HSD17B2* | 3294 |
| *HSD17B3* | 3293 |
| *HSD17B11* | 51170 |
| *HSD17B12* | 51144 |
| *HSD17B14* | 51171 |
| *CYP19A1* | 1588 |
| *SRD5A1* | 6715 |
| *SRD5A2* | 6716 |
| *SRD5A3* | 79644 |
| *SHBG* | 6462 |

# Figures

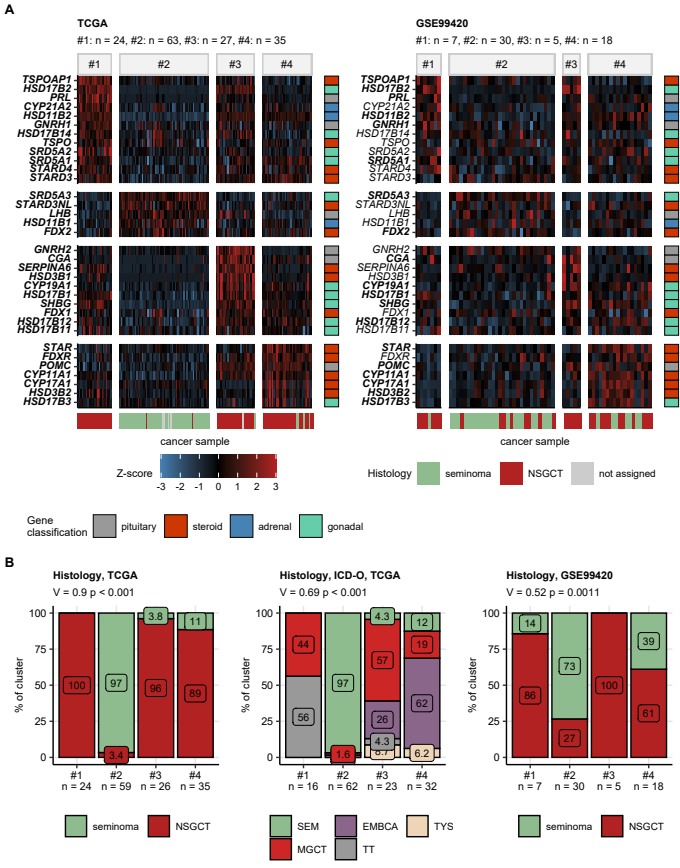


**Figure 1. Differential expression of hormone-related genes in NSGCT and seminoma.**

*-transformed expression of 34 sex hormone-related genes in the cancer tissue was compared between testicular nonseminomatous germ cell tumors (NSGCT) and seminomas in the TCGA and GSE99420 cohorts by Mann-Whitney test with r effect size statistic. P values were corrected for multiple testing with the false discovery rate (FDR) method.*

*(A) Differences in median expression between NSGCT and seminoma and FDR-corrected p values are presented in volcano plots. Each point represents a single gene. Point sizes represent effect size r value. Point color codes for gene classification. Significant genes are highlighted and labelled with their symbols. The dashed horizontal lines represent the pFDR = 0.05 significance cutoffs. Numbers of significantly up- and downregulated genes in NSGCT as compared with seminoma are displayed in the plot captions.*

*(B) Expression of selected significant genes shared by both investigated cohorts. Median expression values with interquartile ranges are presented as boxes with whiskers spanning over 150% of the interquartile ranges. Cancer samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of samples are indicated in the X axes.*

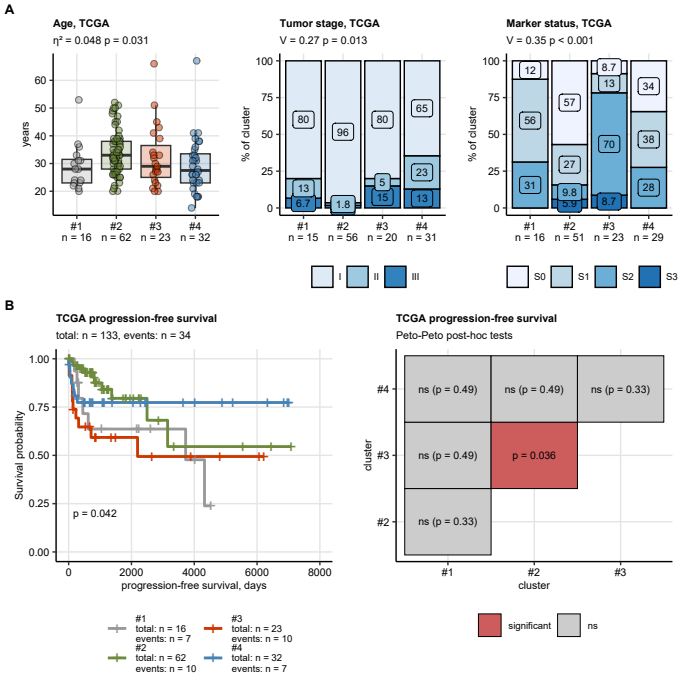


**Figure 2. Hormonal clusters of testicular cancers.**

*Testicular cancer samples in the TCGA training cohort were assigned to the hormonal clusters in respect to expression values of 34 sex hormone-related genes by hard-threshold regularized KMEANS unsupervised clustering. The hormonal cluster assignment was predicted for cancer samples in the GSE99420 training collective by a Random Forest classifier fed with -transformed expression levels of the cluster-defining genes.*

*(A) Normalized -transformed expression values (Z-scores) of the cluster defining genes are presented in a heat map. Genes are arranged by their peak expression in the hormonal clusters of the TCGA cohort. Functional gene classification is color coded in the vertical rug plots. Sample histology (seminoma or nonseminomatous germ cell tumors [NSGCT]) is color coded in the horizontal rug plots. Statistical significance for differences in expression between the hormonal clusters was assessed by Kruskal-Wallis test with effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Significant differences are highlighted by bold font of the gene symbol in the Y axes of the heat maps. Numbers of samples in the clusters are displayed in the plot captions.*

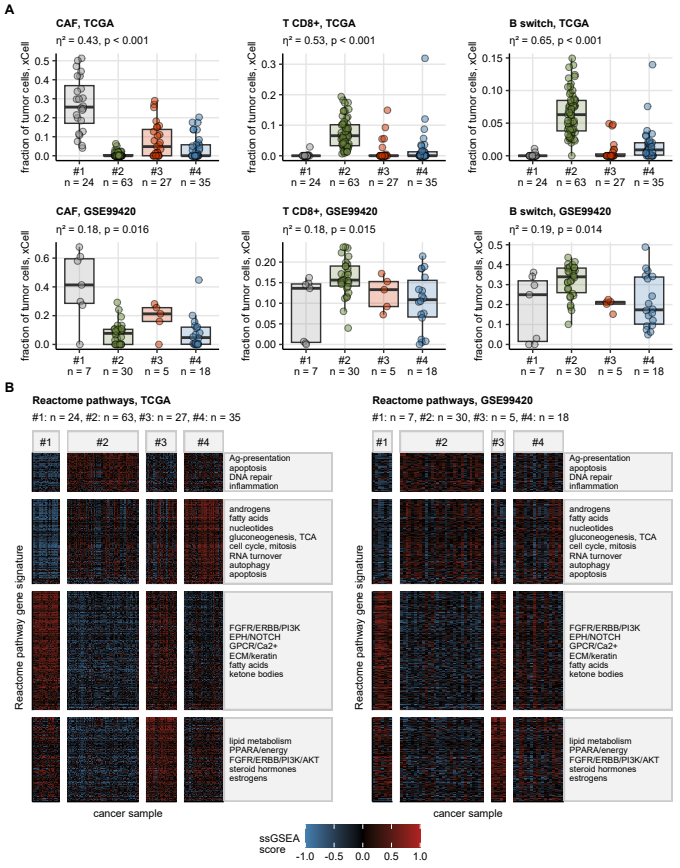
*(B) Differences in distribution of major cancer histologies and histological subtypes (ICD-O: international classification of diseases for oncology) between the hormonal clusters were investigated by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Percentages of histologies within the clusters are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of samples in the clusters are indicated in the X axes.*



**Figure 3. Clinical and prognostic characteristic of the hormonal clusters of testicular cancer.**

*(A) Age at diagnosis (Kruskal-Wallis test, effect size statistic), tumor stage and serum cancer marker status ( test, Cramer’s V effect size statistic) were compared between the clusters. P values were corrected for multiple testing with the false discovery rate method. Median age with interquartile ranges is shown in a box plot with whiskers spanning over 150% of the interquartile ranges and single cancer samples visualized as points. Percentages of tumor stages and patients with cancer marker stages in the hormonal clusters are presented in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of samples in the clusters are indicated in the X axes.*

*(B) General and pairwise differences in progression-free survival between the hormonal clusters were assessed by Peto-Peto tests adjusted for multiple testing with the false discovery rate method. Fractions of surviving patients are visualized in a Kaplan-Meier plot with total numbers of observations and progressions indicated in the plot caption (right panel). Number of observations and progression cases in the hormonal clusters are displayed in the Kaplan-Meier plot legend; p values for the general difference in survival is shown in the plot. P values of the pairwise comparison of survival between the hormonal clusters are displayed in a heat map (left panel).*

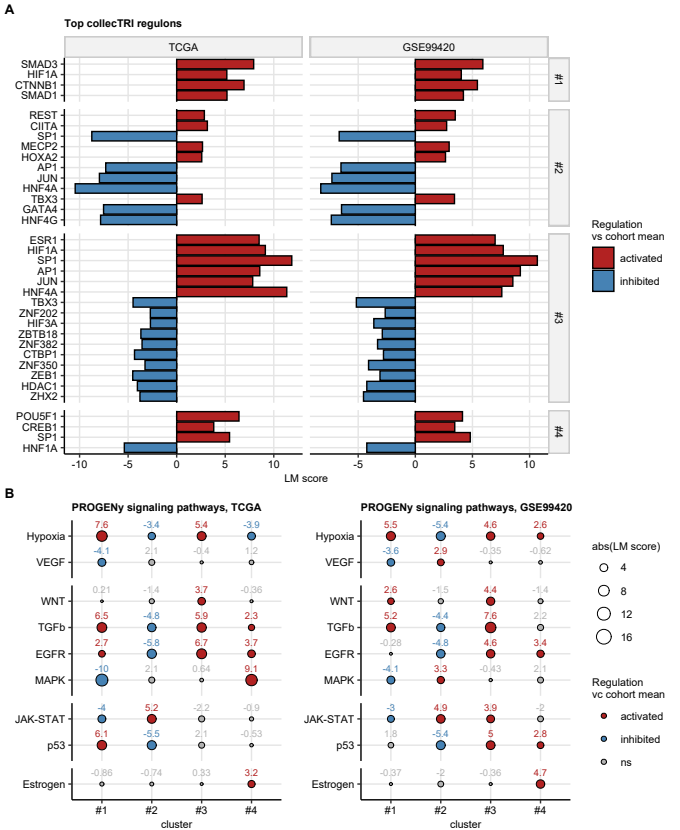


**Figure 4. Tumor microenvironment composition and Reactome pathway gene signatures in the hormonal clusters of testicular carcinoma.**

*(A) Fractions of non-malignant cells in cancer samples were estimated by the xCell immunedeconvolution algorithm and compared between the hormonal clusters with Kruskal-Wallis test with effect size statistic. P values were adjusted for multiple testing with the false discovery rate (FDR) method. Median tumor cell fractions of cancer-associated fibroblasts (CAF), CD8+ T cells, and class-switched B cells with interquartile ranges are visualized as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are represented by points. Effect sizes and p values are displayed in the plot captions. Numbers of cancer samples in the clusters are indicated in the X axes.*

*(B) Single sample gene set enrichment analysis scores (ssGSEA scores) of the Reactome pathway gene signatures were compared between the hormonal clusters by one-way ANOVA with effect size statistic. P values were corrected for multiple testing with the FDR method. moderate-to-large differences with pFDR < 0.05 and were considered significant. ssGSEA scores for gene signatures found to be significant in both the TCGA and GSE99420 cohort are presented in heat maps. Those common significant gene signatures were grouped by unsupervised hierarchical clustering in the TCGA cohort and the signature groups named by their shared biological features. The signature grouping is represented by horizontal facets of the heat maps. Numbers of samples in the hormonal clusters are displayed in the plot captions.*

*Ag: antigen; TCA: tricarboxylic acid cycle/citric acid cycle; GPCR: G protein-coupled receptors; ECM: extracellular matrix; PPARA: peroxisome proliferator activated receptor alpha.*

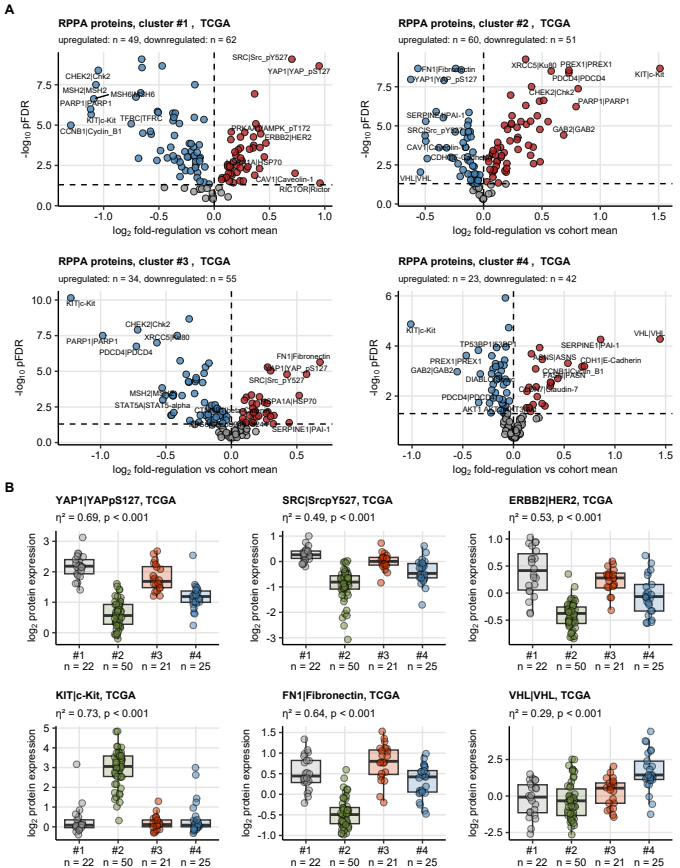


**Figure 5. Modulation of transcriptional regulons and signaling pathways in the hormonal clusters of testicular cancer.**

*Differential modulation of the collecTRI model transcriptional regulons and of the PROGENy signaling pathways in the hormonal clusters as comped with the cohort mean was predicted by decoupleR linear modeling tools fed with T statistics of differential gene expression for all available genes. Linear modelling score (LM score) served as an effect size metric of regulon/signaling pathway modulation (LM score < 0: inhibition, LM score > 0: activation as compared with the cohort mean). P values (LM score 0) were adjusted for multiple testing with the false discovery rate method.*

*(A) LM scores for top strongest modulated transcriptional regulons in the hormonal clusters (horizontal facets) found to be significant in both the TCGA and GSE99420 cohorts presented in bar plots.*

*(B) LM scores for signaling pathways in the hormonal clusters found to be significantly modulated both in the TCGA and GSE99420 cohorts depicted in bubble plots. Point sizes represent absolute values of the LM scores. Point color codes for regulation sign. Points are labelled with the corresponding LM score values.*

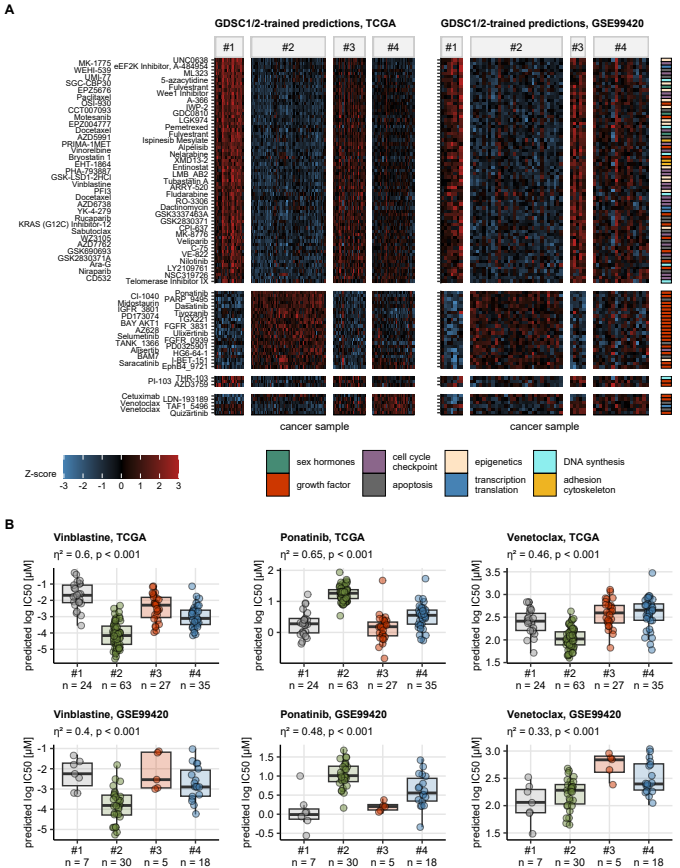


**Figure 6. Differential protein expression in the hormonal clusters of testicular cancers in the TCGA cohort.**

*Expression of 194 cancer biology-relevant proteins was investigated in cancer samples of the TCGA cohort with reverse phase protein array. -transformed expression levels were compared between the hormonal clusters by one-way ANOVA with effect size statistic. Differences in -transformed expression between the cluster and the cohort mean were assessed by one-sample T test. P values were corrected for multiple testing with the false discovery rate (FDR) method. Proteins with pFDR(ANOVA) < 0.05, , and pFDR(T test) < 0.05 were deemed differentially regulated.*

*(A) Differences in expression between the cluster mean and the cohort mean along with FDR-adjusted p values of the T test are presented in Volcano plots. Each point represents a single gene. Point color codes for significance and regulation sign. Top most strongly regulated genes are labeled with their symbols. Numbers of significantly up- and downregulated proteins are displayed in the plot captions:*

*(B) -transformed expression levels of selected proteins discriminating between the hormonal clusters. Median expression values with interquartile ranges are shown as boxes with whiskers spanning over 150% of the interquartile ranges. Cancer samples are visualized as points. Effect sizes and p values for comparison of the clusters with one-way ANOVA are displayed in the plot caption. Numbers of samples in the clusters are indicated in the X axes.*



**Figure 7. Predicted anti-cancer drug response in the hormonal clusters.**

*Anti-cancer drug response in form of log50 (50% inhibitory concentration) in cancer samples was predicted by RIDGE linear models trained with the GDSC drug screening data set. The log IC50 values were compared between the hormonal clusters by one-way ANOVA with effect size statistic. Differences in the log IC50 values between the cluster and the cohort mean were assessed by one-sample T test. P values were corrected for multiple testing with the false discovery rate (FDR) method. Compounds with pFDR(ANOVA) < 0.05, , and pFDR(T test) < 0.05 were considered significant.*

*(A) Normalized log IC50 values (Z-scores) of compounds found to be significant in both the TCGA and GSE99420 cohort are presented in heat maps. The compounds are arranged by their peak expression in the hormonal clusters of the TCGA cohort. Compound classification is color coded in the vertical rug plot.*

*(B) Predicted log IC50 for representative compounds targeting cytoskeleton (vinblastine), growth factor signaling (ponatinib), and apoptosis pathways (venetoclax). Median log IC50 values with interquartile ranges are visualized as boxes with whiskers spanning over 150% of the interquartile ranges. Single cancer samples are depicted as points. Effect sizes and p values for differences between the clusters investigated by one-way ANOVA are displayed in the plot captions. Numbers of samples in the clusters are indicated in the X axes.*

# References

1. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* (2018) 173:400–416.e11. doi: [10.1016/J.CELL.2018.02.052](https://doi.org/10.1016/J.CELL.2018.02.052)

2. Lewin J, Soltan Ghoraie L, Bedard PL, Hamilton RJ, Chung P, Moore M, Jewett MAS, Anson-Cartwright L, Virtanen C, Winegarden N, et al. Gene expression signatures prognostic for relapse in stage I testicular germ cell tumours. *BJU international* (2018) 122:814–822. doi: [10.1111/BJU.14372](https://doi.org/10.1111/BJU.14372)

3. Korkola JE, Houldsworth J, Chadalavada RSV, Olshen AB, Dobrzynski D, Reuter VE, Bosl GJ, Chaganti RSK. Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors. *Cancer research* (2006) 66:820–827. doi: [10.1158/0008-5472.CAN-05-2445](https://doi.org/10.1158/0008-5472.CAN-05-2445)

4. Törzsök P, Oswald D, Dieckmann KP, Angerer M, Scherer LC, Tymoszuk P, Kunz Y, Pinggera GM, Lusuardi L, Horninger W, et al. Subsets of preoperative sex hormones in testicular germ cell cancer: a retrospective multicenter study. *Scientific reports* (2023) 13: doi: [10.1038/S41598-023-41915-7](https://doi.org/10.1038/S41598-023-41915-7)

5. Raymaekers J, Zamar RH. Regularized K-means Through Hard-Thresholding. *Journal of Machine Learning Research* (2022) 23:1–48. <http://jmlr.org/papers/v23/21-0052.html>

6. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* (2017) 77:1–17. doi: [10.18637/JSS.V077.I01](https://doi.org/10.18637/JSS.V077.I01)

7. Breiman L. Random forests. *Machine Learning* (2001) 45:5–32. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)

8. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software* (2011) 39:1–13. doi: [10.18637/JSS.V039.I05](https://doi.org/10.18637/JSS.V039.I05)

9. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* (2010) 33:1–22. doi: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01)

10. Aran D, Hu Z, Butte AJ. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology* (2017) 18:220. doi: [10.1186/s13059-017-1349-1](https://doi.org/10.1186/s13059-017-1349-1)

11. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology* (2016) 17:218. doi: [10.1186/s13059-016-1070-5](https://doi.org/10.1186/s13059-016-1070-5)

12. Finotello F, Mayer C, Plattner C, Laschober G, Rieder Di, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Medicine* (2019) 11:34. doi: [10.1186/s13073-019-0638-6](https://doi.org/10.1186/s13073-019-0638-6)

13. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* (2010) 11:R14. doi: [10.1186/gb-2010-11-2-r14](https://doi.org/10.1186/gb-2010-11-2-r14)

14. Yu G. Gene Ontology Semantic Similarity Analysis Using GOSemSim. *Methods in Molecular Biology* (2020) 2117:207–215. doi: [10.1007/978-1-0716-0301-7\_11](https://doi.org/10.1007/978-1-0716-0301-7_11)

15. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* (2010) 26:976–978. doi: [10.1093/BIOINFORMATICS/BTQ064](https://doi.org/10.1093/BIOINFORMATICS/BTQ064)

16. Müller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia-i-Mompel P, Fallegger R, Türei D, Lægreid A, Saez-Rodriguez J. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic acids research* (2023) 51:10934–10949. doi: [10.1093/NAR/GKAD841](https://doi.org/10.1093/NAR/GKAD841)

17. Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, Taus P, Dugourd A, Holland CH, Ramirez Flores RO, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* (2022) 2: doi: [10.1093/BIOADV/VBAC016](https://doi.org/10.1093/BIOADV/VBAC016)

18. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, Garnett MJ, Blüthgen N, Saez-Rodriguez J. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications 2017 9:1* (2018) 9:1–11. doi: [10.1038/s41467-017-02391-6](https://doi.org/10.1038/s41467-017-02391-6)

19. Wang C, Gu Y, Zhang K, Xie K, Zhu M, Dai N, Jiang Y, Guo X, Liu M, Dai J, et al. Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. *Nature Communications 2016 7:1* (2016) 7:1–12. doi: [10.1038/ncomms10499](https://doi.org/10.1038/ncomms10499)

20. CTpedia. CTpedia. (2024) <http://www.cta.lncc.br/> [Accessed June 12, 2024]

21. Raza A, Merhi M, Inchakalody VP, Krishnankutty R, Relecom A, Uddin S, Dermime S. Unleashing the immune response to NY-ESO-1 cancer testis antigen as a potential target for cancer immunotherapy. *Journal of Translational Medicine* (2020) 18:1–11. doi: [10.1186/S12967-020-02306-Y/FIGURES/2](https://doi.org/10.1186/S12967-020-02306-Y/FIGURES/2)

22. Li Z, Li T, Yates ME, Wu Y, Ferber A, Chen L, Brown DD, Carroll JS, Sikora MJ, Tseng GC, et al. The EstroGene Database Reveals Diverse Temporal, Context-Dependent, and Bidirectional Estrogen Receptor Regulomes in Breast Cancer. *Cancer Research* (2023) 83:2656–2674. doi: [10.1158/0008-5472.CAN-23-0539/727161/AM/THE-ESTROGENE-DATABASE-REVEALS-DIVERSE-TEMPORAL](https://doi.org/10.1158/0008-5472.CAN-23-0539/727161/AM/THE-ESTROGENE-DATABASE-REVEALS-DIVERSE-TEMPORAL)

23. Massie CE, Lynch A, Ramos-Montoya A, Boren J, Stark R, Fazli L, Warren A, Scott H, Madhu B, Sharma N, et al. The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *The EMBO Journal* (2011) 30:2719. doi: [10.1038/EMBOJ.2011.158](https://doi.org/10.1038/EMBOJ.2011.158)

24. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular and Cellular Proteomics* (2012) 11:M111.014647. doi: [10.1074/mcp.M111.014647](https://doi.org/10.1074/mcp.M111.014647)

25. Petrov PB, Considine JM, Izzi V, Naba A. Matrisome AnalyzeR – a suite of tools to annotate and quantify ECM molecules in big datasets across organisms. *Journal of Cell Science* (2023) 136: doi: [10.1242/JCS.261255/325836/AM/MATRISOME-ANALYZER-A-SUITE-OF-TOOLS-TO-ANNOTATE](https://doi.org/10.1242/JCS.261255/325836/AM/MATRISOME-ANALYZER-A-SUITE-OF-TOOLS-TO-ANNOTATE)

26. Gavai AK, Supandi F, Hettling H, Murrell P, Leunissen JAM, Van Beek JHGM. Using Bioconductor Package BiGGR for Metabolic Flux Estimation Based on Gene Expression Changes in Brain. *PLOS ONE* (2015) 10:e0119016. doi: [10.1371/JOURNAL.PONE.0119016](https://doi.org/10.1371/JOURNAL.PONE.0119016)

27. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* (2016) 44:D515–D522. doi: [10.1093/NAR/GKV1049](https://doi.org/10.1093/NAR/GKV1049)

28. Lamar JM, Xiao Y, Norton E, Jiang ZG, Gerhard GM, Kooner S, Warren JSA, Hynes RO. SRC tyrosine kinase activates the YAP/TAZ axis and thereby drives tumor growth and metastasis. *Journal of Biological Chemistry* (2019) 294:2302–2317. doi: [10.1074/JBC.RA118.004364/ATTACHMENT/356A6DC7-0635-4AA9-92EE-0033311C7338/MMC1.ZIP](https://doi.org/10.1074/JBC.RA118.004364/ATTACHMENT/356A6DC7-0635-4AA9-92EE-0033311C7338/MMC1.ZIP)

29. Hsu PC, Yang CT, Jablons DM, You L. The Crosstalk between Src and Hippo/YAP Signaling Pathways in Non-Small Cell Lung Cancer (NSCLC). *Cancers* (2020) 12: doi: [10.3390/CANCERS12061361](https://doi.org/10.3390/CANCERS12061361)

30. Alam M, Alam S, Shamsi A, Adnan M, Elasbali AM, Al-Soud WA, Alreshidi M, Hawsawi YMR, Tippana A, Pasupuleti VR, et al. Bax/Bcl-2 Cascade Is Regulated by the EGFR Pathway: Therapeutic Targeting of Non-Small Cell Lung Cancer. *Frontiers in Oncology* (2022) 12:1. doi: [10.3389/FONC.2022.869672](https://doi.org/10.3389/FONC.2022.869672)

31. Cardoso HJ, Figueira MI, Socorro S. The stem cell factor (SCF)/c-KIT signalling in testis and prostate cancer. *Journal of Cell Communication and Signaling* (2017) 11:297. doi: [10.1007/S12079-017-0399-1](https://doi.org/10.1007/S12079-017-0399-1)

32. Geeleher P, Cox N, Stephanie Huang R. pRRophetic: An R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLOS ONE* (2014) 9:e107468. doi: [10.1371/JOURNAL.PONE.0107468](https://doi.org/10.1371/JOURNAL.PONE.0107468)

33. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Briefings in Bioinformatics* (2021) 22:1–7. doi: [10.1093/BIB/BBAB260](https://doi.org/10.1093/BIB/BBAB260)

34. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research* (2013) 41:D955–D961. doi: [10.1093/NAR/GKS1111](https://doi.org/10.1093/NAR/GKS1111)

35. Seashore-Ludlow B, Rees MG, Cheah JH, Coko M, Price EV, Coletti ME, Jones V, Bodycombe NE, Soule CK, Gould J, et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer discovery* (2015) 5:1210–1223. doi: [10.1158/2159-8290.CD-15-0235](https://doi.org/10.1158/2159-8290.CD-15-0235)

36. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28:882. doi: [10.1093/BIOINFORMATICS/BTS034](https://doi.org/10.1093/BIOINFORMATICS/BTS034)

37. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* (2013) 14:7. doi: [10.1186/1471-2105-14-7](https://doi.org/10.1186/1471-2105-14-7)

38. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* (1995) 57:289–300. doi: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)

39. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* (1987) 20:53–65. doi: [10.1016/0377-0427(87)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

40. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* (2006) Complex Sy:1695. [https://igraph.org](https://igraph.org/)

41. Briatte F, Bojanowski M, Canouil M, Charlop-Powers Z, Fisher JC, Johnson K, Rinker T. ggnetwork: Geometries to Plot Networks with ’ggplot2’. (2021) <https://cran.r-project.org/package=ggnetwork>