Am J Cancer Res 2021;11(11):5299-5318
www.ajcr.us /ISSN:2156-6976/ajcr0133575

Original Article

Patient-matched analysis identifies deregulated networks in prostate cancer to guide personalized therapeutic intervention

Akinchan Kumar[1,2,3,4,5], Yasenya Kasikci[1,2,3,4,5], Alaa Badredine[1,2,3,4,5,8], Karim Azzag[6,9], Marie L Quintyn Ranty[7,10], Falek Zaidi[7], Nathalie Aragou[7], Catherine Mazerolles[7#], Bernard Malavaud[7], Marco A Mendoza-Parra[1,2,3,4,5,11], Laurence Vandel[6,12], Hinrich Gronemeyer[1,2,3,4,5]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Department of Functional Genomics and Cancer, Illkirch, France; [2]Centre National de la Recherche Scientifique, UMR7104, Illkirch, France; [3]Institut National de la Santé et de la Recherche Médicale, U1258, Illkirch, France; [4]Université de Strasbourg, Illkirch, France; [5]Equipe Labellisée Ligue Contre le Cancer; [6]Centre de Biologie du Développement (CBD), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, France; [7]Institut Universitaire du Cancer Toulouse-Oncopole (IUCT-O), Toulouse, France; [8]CNRS UMR8199-EGID Building, Lille University-Faculty of Medicine Henri-Warembourg, Lille, France; [9]Lillehei Heart Institute, Department of Medicine, University of Minnesota, Minneapolis, MN, USA; [10]Pathology Department, CHU, Caen, France; [11]UMR 8030 Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université Evry-val-d'Essonne, University Paris-Saclay, Évry, France; [12]Université Clermont Auvergne, CNRS, Inserm, GReD, Clermont-Ferrand, France. [#]Deceased in 2015.

Received March 26, 2021; Accepted June 12, 2021; Epub November 15, 2021; Published November 30, 2021

Abstract: Prostate cancer (PrCa) is the second most common malignancy in men. More than 50% of advanced prostate cancers display the TMPRSS2-ERG fusion. Despite extensive cancer genome/transcriptome data, little is known about the impact of mutations and altered transcription on regulatory networks in the PrCa of individual patients. Using patient-matched normal and tumor samples, we established somatic variations and differential transcriptome profiles of primary ERG-positive prostate cancers. Integration of protein-protein interaction and gene-regulatory network databases defined highly diverse patient-specific network alterations. Different components of a given regulatory pathway were altered by novel and known mutations and/or aberrant gene expression, including deregulated ERG targets, and were validated by using a novel *in silico* methodology. Consequently, different sets of pathways were altered in each individual PrCa. In a given PrCa, several deregulated pathways share common factors, predicting synergistic effects on cancer progression. Our integrated analysis provides a paradigm to identify druggable key deregulated factors within regulatory networks to guide personalized therapies.

Keywords: Cancer systems biology, prostate cancer, personalized therapy, patient-matched deregulated networks

Introduction

The mutational landscapes of primary and advanced/metastatic PrCa have been extensively analyzed [1-5], as has been the prevalence of the androgen-sensitive TMPRSS2 promoter fusion with ETS transcription factors [6], which endows ETS with responsiveness to the androgen receptor (AR) that is frequently overexpressed in antiandrogen-resistant PrCa [7]. In these studies, recurrent mutations have been found in genes coding for factors regulating a plethora of pathways and key cellular functions, such as the androgen receptor signaling, PI3K/RAS/RAF/WNT pathways, and factors involved in DNA repair and chromatin methylation, or cell cycle control. One of the caveats in all these studies was that, with a few exceptions [2, 8], information was generally compiled from large numbers of tumors from different patients. Thus, while enabling identification of predominant mutations, these studies did not reveal the spectrum of aberrations that existed in individual patients' prostates at diagnosis. All these aberrations may affect different regulatory pathways and their added, possibly synergistic action may be critical for malignancy and tumor progression. Indeed, restoring a normal state would require the correction of a highly complex and dynamically

regulated system of interactive multi-component networks which are deregulated in disease [9]. Towards this goal, the identification of aberrant networks and their inherent hierarchies is essential to design patient-selective therapeutic interventions through generic or key factor-specific modulation of the affected pathways.

Material and methods

Patient sample collection

All samples were collected within 15 minutes after radical prostatectomy to shorten the delay between de-vascularization and freezing, and to ensure preservation of labile molecules. Immediately following prostatectomy, punch biopsies ("carrots") of 8 mm diameter were taken from tumor and adjacent normal tissue, snap-frozen in liquid nitrogen and stored at -80°C. Carrots used for genomic and transcriptomic studies were cut into sequential tissue sections and the tumor cellularity was monitored at regular intervals by histological staining to ensure homogeneity of tumor and normal sections stored in LoBind tubes at -80°C.

Tissue microarrays

Tissues microarrays (TMA) were made from paraffin-embedded tissue cores of histopathology-confirmed prostate cancer and patient-matched tumor-adjacent normal tissue. For each tumor, two representative tumor areas were selected and two cores of 2 mm in diameter were punched and included in paraffin recipient blocs. Two adjacent normal tissues of each selected prostatic sample were arrayed on TMAs and constituted the "normal" counterparts of each tumor sample. The TMAs were performed on the histopathology platform of the Biological Resource Center (CRB) of the Toulouse University Hospital, in a semi-automated way using the EZ-TMA™ Manual Tissue Microarray Kit (IHC World). The slides were examined by HE coloration and immunohistochemical studies were performed on TMA tissues. Immunohistochemistry was done using an automated Dako Autostainer. The following antibodies were used: ERG, EZH2, Androgen Receptor. Slides were digitalized using a Hamamatsu NanoZoomer slide scanner (Japan) at 20× magnification with a resolution of 0.46 microns per pixel. The results were interpreted under an optical microscope by two pathologists (CM and M-LQ), blinded to the clinical data.

Whole exome sequencing (WES) and analysis pipeline

For WES, DNA was isolated from frozen tumor and matched normal tissue (10 sections of 10 µm for each tumor and normal sample) using QIAamp DNA micro kit according to manufacturer's instructions. DNA was processed by GATC Biotech for exome capture, library preparation and sequencing. Briefly, SureSelectXT Human all exon V6 kit was used to capture exons, libraries were prepared using TruSeq DNA library preparation kit according to manufacturer's instruction and Paired-End 125-base sequencing was performed on Illumina HiSeq 2000. FastQ files provided by GATC Biotech were processed for variant discovery with Genome Analysis Toolkit (GATK, 3.7) [10] using default parameters. To assist in WES analysis we developed a WES Analysis Pipeline [written in Python3 with the Snakemake (3.13.3) [11] management tool] and used the Genome Analysis Toolkit (GATK, 3.7) [10] according to the authors' instructions. The following tools were used for each step in the pipeline.

Pre-processing of the samples: FastQ files were first aligned to hg19 using BWA-mem (0.7.17) [12] using standard parameters. The output SAM files were then converted to BAM files using SAM tools (1.6) [13]. BAM files were processed using Picard tools (2.14) to sort by coordinates, remove duplicates and add read group tags (essential to differentiate between Normal and Tumor samples) to samples before indexing them with SAM tools. Then BAM files were recalibrated (BQSR) using GATK, as recommended for enhancing variant calling by providing databases of known polymorphic sites: a set of curated INDEL entries, a Single Nucleotide Polymorphism database dbSNP, the COSMIC database of somatic cancer mutations.

Creating the Panel of Normals (PON): The 'Panel of Normals' is created from the normal samples using GATK. This method is used as a filter to reject artifacts and germline variants that are present in at least two normal samples (-minN 2). It uses as input the hg19, the dbSNP, the COSMIC and the intervals of the genome to analyze only the exons of all genes captured. It generates a new file that will be used when calling the variants.

Variant calling with MuTect2: After pre-processing and alignment, MuTect2 (GATK) was used to call somatic variants [14]. Inputs are the following files: hg19, PON, HumanAllExonV6r2, COSMIC and dbSNP. The normal and tumor samples were compared using the following parameters: pir_mad_threshold: 6; max_alt_alleles_in_normal_count: 5; pir_median_threshold: 35; standard_min_confidence_threshold_for_calling: 30.

Annotation of VCF: The annotations of the VCF files were done using SnpEFF [15] and SnpSift [15] with hg19 as reference genome.

Validation of mutations

Target regions were amplified by PCR. PCR products were purified using Qiagen gel extraction kit and sequenced by Eurofins Genomics using the BigDye Terminator Cycle Sequencing Kit and an ABI 3730xl automated sequencer (Applied Biosystems). The sequencing primers were the same as those used for PCR amplification. Variants were confirmed using SNAP gene viewer.

RNA-sequencing

RNA was isolated from frozen tumor and matched normal tissue using Trizol reagent (Invitrogen). In all cases, two independent sets of adjacent 10 µm sections (N=10) were processed for RNA isolation to generate biological duplicate RNA-seq data. For subsequent interpretations only data that were consistent between the biological duplicates were retained. RNA was further cleaned up using RNeasy MinElute Cleanup Kit. RNA was then sent to GATC Biotech (Konstanz, Germany) for strand-specific, paired-end and Ribo-minus total RNA-seq. Briefly, ribosomal RNA depletion was done using Ribo Zero gold kit (Illumina Inc); libraries were prepared using TruSeq stranded total RNA library prep kit (Illumina Inc.). Paired-end 125 base or 150 base sequencing was performed using Illumina HiSeq 2000. FastQ files received from GATC Biotech were used for further analysis.

RNA-seq analysis pipeline

The analysis pipeline consists of the following steps.

Pre-processing, alignment and counting raw reads: FastQ files were assessed for quality using FastQC. FastQ files were aligned to reference genome (human genome hg19) using the Hisat2 [16] aligner. Aligned SAM files were converted to BAM files and sorted using SAMtools [13]. The R package Summarized Experiment [17] was used for counting raw reads per exon/gene.

Differential gene expression analysis: The patient-specific differential gene expression analysis was done using DESeq2 (1.20.0) [18] according to the general steps described with the parameters given below. The samples have been analyzed by giving the matched raw read counts normal/tumor duplicates as input.

- Removing sum of row counts: 0;

- CooksCutoff: False;

- Alpha: 0.01;

- Subset genes with Adjusted *P*-value ≤0.01;

- Subset genes with Log2FC ≤-1 or Log2FC ≥+1.

The corresponding list of Differentially Expressed Genes (DEGs) for each patient was used for further analysis.

Pathway enrichment analysis

To interpret the gene expression data, the DEG list was loaded into GeneCodis [19] and the Panther pathway analysis function was used to retrieve enriched pathways. Hypergeometric correction of *P*-values was applied and pathways displaying a corrected *P*-value <0.01 were considered enriched. We clustered all pathways of all samples using Plotly in R. For the datasets obtained from TCGA (54 PrCa patient data along with matched normal), HT-seq counts were downloaded for each patient corresponding to tumor and matched normal. DEseq2 was used to identify the DEGs for each patient. Pathway enrichment analysis was performed as described before.

Patient-specific network generation and visualization

To generate the gene networks for individual patients we extracted the list of mutated genes from WES and differentially expressed genes

(DEGs) from RNA-seq of tumor *vs* normal samples for each patient. These lists of genes were queried against two known databases of network interactions, STRING [20], a Protein-Protein Interaction (PPI) database, and CellNet [21], a gene regulatory network (GRN) database. For STRING, we merged the list of genes (DEGs and Mutation, keeping the information whether the gene is a DEG or a mutated gene as attributes), removed any duplicated genes and queried them using an in-house script. As for the parameters, we only chose edge interactions that have been experimentally validated (exp_score ≠ 0). For CellNet, we queried only the differentially expressed genes on the target genes and retrieved along the cognate transcription factors. We chose interactions that had only a z-score ≥5. After obtaining networks from both databases, we proceeded to add the information from WES and RNA-seq whether the genes were mutated, differentially expressed or both, in addition to the information obtained from the databases.

Network visualization and merging using Cytoscape

Individual networks, created by using Cellnet and String for each patient, were visualized using Cytoscape [22]. Finally, CellNet and STRING networks for each patient were merged using the Cytoscape merge function to obtain master networks for each patient. Sub-networks were then extracted for further visualization and analysis.

Identification of putative AR and ERG target genes

A two-step approach was used. First, we collected sequenced read files (bed format) associated to public ChIP-seq assays targeting ERG in TMPRSS2-ERG positive human VCaP prostate cancer (GSM1328978, GSM1328979) and RWPE-1 normal prostate epithelium cells (GSM2195103, GSM2195106). BED Replicate files per cell-type were merged together prior performing peak calling (MACS 1.4; no model, shiftsize =150 nts, *P*-value threshold: 1×10^{-5}), followed by their genomic annotation to the closest transcription start sites (annoPeakR). This analysis allowed to pair the characterized DEGs and mutated genes within the patient-derived networks with genes presenting proximal AR binding sites (<10 kb distance) on VCaP ChIP-seq profiles. This primary analysis has been validated in a second step by comparative visual inspection of ChIP-seq profiles. For this we used the qcGenomics platform, in which the dedicated genome browser NAVi allows to visualize any publicly available ChIP-seq profile. Specifically, we used NAVi to extract all AR and ERG ChIP-seq profiles for TMPRSS2-ERG positive human VCaP prostate cancer and RWPE1 normal prostate epithelium cells. The pre-computed datasets were displayed simultaneously in the NAVi browser for comparative visualization. Only tracks with an apparent high signal-to-noise ratio were retained (VCaP-ERG: GSM2058880, GSM1328978, GSM1378979, GSM1328980, GSM1328981; VCaP-AR: GSM-1410768, RWEP1-ERG: GSM927071, GSM-2195110, GSM2195103; VCaP-GROseq: GSM-2235682). Promoter-proximal ERG binding was scored positive in this visual 'validation' (attributing a yellow color to the respective nodes) only when there was a clearly visible peak above the background at a scale of 30 to 300 (read count intensity; depending on the signal and noise intensities of each profile), provided that there was no other known TSS closer (see Supplementary Figure 6 for examples of gain of ERG binding).

Data availability

The RNA-seq data sets generated in the context of this study from 15 patient-matched tumor and normal prostate tissue are available in the Gene Expression Omnibus (GEO) repository under the accession number GSE133626. The corresponding Exome-seq data sets from the prostates of the same 15 patients are available from the SRA database under the accession number PRJNA555457. Data sets for the pathway analysis shown in Supplementary Figure 3 were downloaded from The Cancer Genome Atlas (TGCA) as described in the methods section. The sequencing statistics of RNA-seq and Exome-seq experiments are specified in Supplementary Table 4.

Results

Overview of the approach

We chose prostate cancer as a solid tumor paradigm to integrate patient-specific differentially expressed (DEGs) and mutated genes, using information from protein-protein interaction

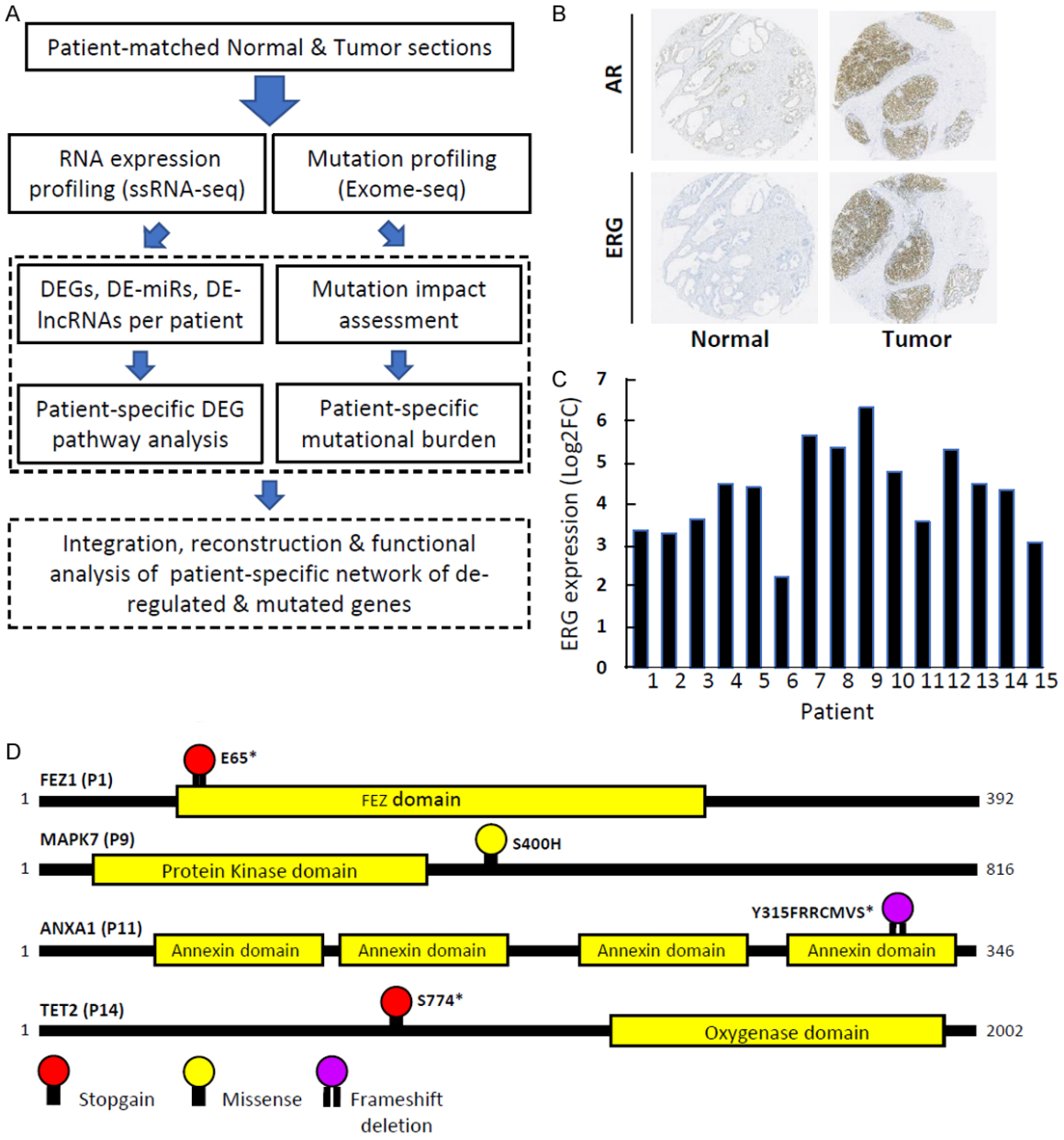


Figure 1. Analysis strategy and characterization of patient-matched samples. A. Sketch of the workflow of this study. B. Representative IHC images of cancer and corresponding matched normal samples from patient 15 (P15) stained with an anti-AR antibody (top panel) or an anti-ERG antibody (bottom panel). C. RNA-seq data of ERG expression in all the patients' tumors relative to their matched normal. Differential ERG expression for all patient-matched duplicate samples was supported by q value $<10^{-69}$ using DEseq (see Methods for details). D. Schematic illustration of 4 novel mutations in P1, P9, P11 and P14, which are predicted to have a high or moderate impact.

and gene-regulatory network databases [21, 23] (**Figure 1A**) to generate patient-specific cancer-modified networks. Extensively characterized normal and tumor frozen punch biopsies from the same prostate were obtained from radical prostatectomy specimens of non-treated patients. 15 primary ERG-positive tumors (T) and matched normal tissue (N) were selected by expert pathologists on the basis that consecutive sections of the same biopsy differed only minimally in tumor cellularity (>80% tumor cells), while the sections of N biopsies from the same prostate had 0% tumor cells. With one exception of patient 14 (P14), the proportion of infiltrating lymphocytes relative to tumor cells was close to 0%, only occasionally rare scattered lymphocytes were observed in the stroma. The tumor sections of

P14 showed up to 25% (area-based) mononuclear immune cells. Immunohistochemistry and RNA-seq (biological duplicates) confirmed ERG overexpression relative to the matched N samples and all samples revealed increased androgen receptor (AR) levels (**Figure 1B**, **1C**).

Identification of large-scale patient-specific genomic changes using exome sequencing analysis

To identify somatic variation, we performed whole-exome sequencing (WES, Supplementary Table 1). Variants were called using MuTect2, which revealed between 49 and 114 mutations in each cancer relative to the corresponding normal prostate tissue; only mutations predicted to have high or moderate impact were considered subsequently (Supplementary File 1). Intriguingly, in addition to classical mutations, for example in *MYC*, *TP53, PTEN* or components of the PI3K and WNT pathways [1, 2, 24], unreported patient-specific mutations were observed in all samples (Supplementary Table 2*;* for validations see Supplementary Figure 1). In P1 three hitherto unreported somatic mutations affected the putative tumor suppressors *BANP* [25], *FEZ1* [26, 27] (**Figure 1D**) and *TINAGL1*, which interferes with both integrin and EGFR signaling [28]. We also found novel mutations in *MAPK7* (R400H) in P9, Annexin A1 (*ANXA1*, frameshift deletion; P11) and a *TET2* mutation that truncates the protein and renders it non-functional (**Figure 1D**; P14). These novel somatic mutations were seen only in single patients. However, the nature of the mutations, often truncating proteins of functional importance, is likely to have a significant impact in the individual case. Moreover, several of those genes were found mutated in other cancer types, supporting their functional impact (Supplementary Figure 2). The ability of TINAGL1 to inhibit progression and metastasis of triple-negative breast cancer [28], provides strong rational for such personalized genomic analysis. Our data underscores the recent notion that "significantly mutated genes" in PrCa may occur at frequencies of only a few percent [29].

RNA-seq analysis revealed patient-specific altered tumor transcriptome

Mutations in regulatory elements (e.g., enhancers) and factors (e.g., transcription factors, epigenetic modulators, enzymes) can affect global gene expression. To integrate these effects in the network analysis, we performed high-throughput strand-specific paired-end total RNA sequencing after ribosomal RNA depletion from matched T and N biopsy sections as biological duplicates. As expected, T *vs*. N analysis of the RNA-seq datasets identified tumor-specific differentially expressed genes (TS-DEGs; Supplementary File 2) with diverse functionalities, comprising (i) cancer-specific deregulated proto-oncogenes like c-*MYC* (all except P6, P11, P13) but also (ii) pleiotropic factors like the serine protease KLK4 (P4, P10, P14, P15), a regulator of AR and the PI3K/AKT/mTOR pathway [30] and of protease-activated receptors [31]. Notably, deletion of *KLK4* impairs PrCa growth [30]. Moreover, (iii) epigenetic modifiers like *JMJD6* (P14), *KDM4B* (P5), *KDM6A* (P2, P8), *KDM6B* (P6, P9), *TET3* (P12, P14), *KAT2A* (P3), *KAT6A* (P2) or *HDAC9* (P2-5, P7-11, P13-15) were differentially expressed in certain tumors. In addition to protein-coding genes, expression of (iv) certain regulatory RNAs was altered in tumors [micro-RNAs (miRs)], as well as long non-coding RNA (lncRNAs); for annotated miRs and lncRNAs, see Supplementary Table 3. Of note, the p53-inducible lncRNA NEAT, a promising therapeutic target whose ablation generates synthetic lethality with chemotherapy and p53 reactivation therapy [32, 33], was over-expressed in 7/15 PrCa samples. A prominent ERG binding site in VCaP and in normal prostate epithelial RWPE-1 cells about 4.6 kb upstream of the NEAT transcriptional start site may account for this deregulation (e.g., GSM2086313, using the qcGenomics browser). The androgen-responsive lncRNA ARLNC1 [34] was up-regulated in 9/15 paired samples but down-regulated in P5. HOTTIP, a component of H3K4 methyltransferase complexes [35] that can act as AR co-activator [36] and was reported as negatively androgen-regulated lncRNA [34] in prostate cancer cells, was downregulated in 9/15 PrCa samples. This included several, but not all of those with up-regulated ARLNC1. A similar divergence was seen with putative tumor suppressor and oncogenic miRNAs that are actively considered for clinical development [37]. For example, the RNA levels of tumor-suppressor miR34a were decreased in three samples (P10, P13 and P15) but increased in P4 and not affected in 11 other samples. MiR222, which displays targe-

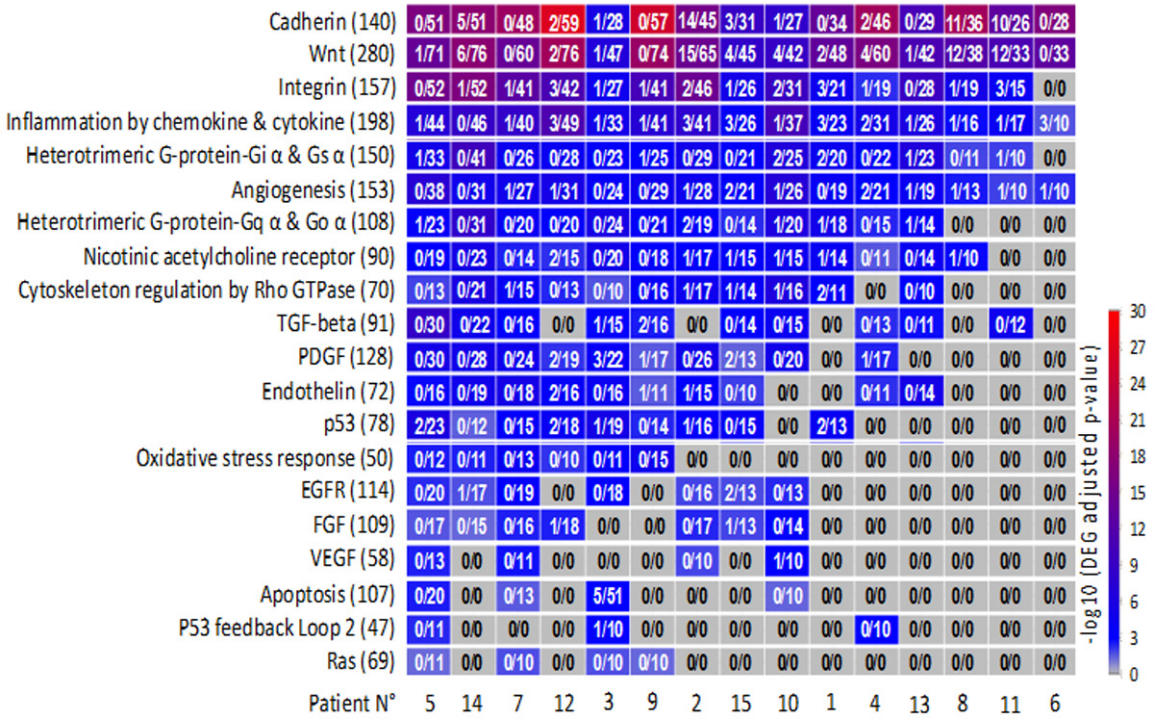


Figure 2. Divergence of pathways and severity of pathway alteration in individual prostate cancer patients. A selection of pathways predicted by Panther to be significantly enriched in patient-specific DEGs are shown on the left of the table while patient identification numbers are given at the bottom. Panther-computed *P*-values for the deregulation of a given pathway are illustrated as blue-to-red color-coded rectangles; grey color indicated no significant alteration. In the rectangles, the numbers of mutated and deregulated genes are given for each pathway and patient (mutated/DEG; see Methods for details).

table oncomiR characteristics in liver, pancreas and lung tumors [12, 37], was unexpectedly down-regulated in 8/15 PrCa samples. Together, these vastly divergent genetic mutations and altered, often counter-intuitive gene expression patterns revealed the need to decipher for each individual patient the complexity of the deregulated systems to identify key targets in critical signaling pathways and/or key nodes in (sub)networks for concomitant intervention at several functionally different levels to generate synergistic effects.

Pathway analysis revealed deregulated pathways in a patient-specific manner

As first step towards the integration of the various deregulated functions within each tumor, we performed patient-centered pathways enrichment analyses for TS-DEGs using Panther in the GenCodis3 environment. While several pathways were commonly deregulated in PrCa of several patients-particularly cadherin, Wnt and integrin signaling-these analyses also demonstrated that in each patient different sets of pathways were deregulated. Indeed, P5 and P6 had, respectively, the most and least severely affected PrCa in terms of numbers of deregulated pathways (**Figure 2**). This finding was further supported by analysis of additional 52 patients from the TCGA repository (Supplementary Figure 3). Moreover, different numbers and components of a commonly deregulated pathway were altered in different patients. As pointed out previously [24], genetic mutations of core Wnt pathway components are rare in PrCa, while abnormal expression of β-catenin is frequent, suggesting that this deregulation occurs indirectly.

Integration of genomic and transcriptomic datasets to generate patient-specific networks

Genes never function in isolation but rather in a highly complex physiological context, which can be illustrated by their communication with other cellular components. To gain a more precise insight into the altered communication by

patient-specific gene deregulations and mutations, we reconstructed master networks from all deregulated genes for each prostate cancer by integrating the connectivity provided by the validated STRING protein-protein and CellNet transcription factor-target gene interaction databases; in addition, we integrated all mutated genes and putative ERG and AR target genes identified by cognate binding sites in the vicinity of the transcriptional start site (TSS) [see Supplementary File 3 (cytoscape master-file for each patient)]. Within these master networks, we studied first the components of the canonical and non-canonical Wnt pathways by merging all 183 deregulated/mutated genes of 15 patients (Supplementary Figure 4A). Displaying the affected components in color in the context of the entire Wnt pathway connectivity revealed an unexpected heterogeneity (Supplementary Figures 4B-L and 5). In P2 (**Figure 3A**) an important signaling factor (phospholipase PLCB1) for the production of second messenger molecules (DAG, IP3) is mutated in the phospholipase domain (V571M) and the expression of multiple other master genes is deregulated, including *PPP3CA*, *GSK3B*, *MYC*, *TP53*, *HDAC1* in addition to several *WNT* and *Frizzled (FZD)* receptor genes. For several of the upregulated DEGs (*GSK3B*, *HDAC1, FZD8*; red arrows) drugs exist which have been approved or are tested in clinical trials [39]. Several druggable genes *(HDAC1, FZD5, FZD8, and MAP3K9)* are also upregulated in P9 but not in P6 while P10 showed overexpressed *FZD8* (**Figure 3B-D**). Notably, ChIP-seq data of TMPRSS2-ERG-positive VCaP and normal RWPE1 prostate epithelial cells indicate that *WNT7B* and *HDAC1* are putative dual AR and ERG target genes, most likely affected by deregulated ERG and possibly, AR signaling (Supplementary Figure 6A, 6B). Even more strikingly, the genes of several key signaling factors (PAK1, CREM) and of the epigenetic modulator SMARCD3 have apparently acquired ERG binding capability in their promoter regions during tumorigenesis (for *SMARCC1,* Supplementary Figure 6C), as it was reported for the ERG-mediated repression of checkpoint kinase 1 [40]. In contrast, P6 showed a very small number of deregulated components of the core Wnt pathway (**Figure 3B**), comprising three upregulated *FZD* receptors along with the cognate *WNT2* ligand acquired ERG binding near the TSS in VCaP cells (Supplementary Figure 6D, 6E). Such a scenario may be addressed with WNT inhibitor-based therapeutics [39]. Patient-specific network alteration was also seen for less frequently affected signaling pathways. The PDGF and EGFR pathways were affected seriously in 10 and 7 patients, respectively (**Figure 4A**, **4E**; merged networks of alterations). However, the scenarios were completely different across individual patients (**Figure 4**; Supplementary Figures 7 and 8). Important changes were seen in P4 and P5 (**Figure 4B**, **4C**) but hardly any in P13 (**Figure 4D**). Note that P13, in contrast to the other patients, did not reveal any upregulated druggable target (red arrow in **Figure 4**). A similar scenario of alterations was found for the EGFR pathway in P5 and P15 (**Figure 4F**, **4G**), while much less nodes were affected in P8 (**Figure 4H**).

Crosstalk among different deregulated pathways in same patient

Finally, given that pathways do not act in isolation, we extracted the affected components of several pathways from the “master networks”. This analysis showed very clearly that, for P2 and P5 several genes of the Wnt, cadherin and integrin pathways are shared between two or even three pathways (**Figure 5**); the same was observed for other combinations of pathways (Supplementary Figure 9). The functional consequence of deregulation/mutation of such genes is predicted to be serious and such nodes may be candidates for therapeutic targeting. It is worth pointing out that also genes at the nexus of several pathways diverged from one patient to another, as shown for P2 and P5 (**Figure 5A**, **5B**). Indeed, hypothetical treatment of these two patients would have to consider different scenarios. In the PrCa of P2, common to two pathways, there is a strong upregulation of the expression of several *WNT* and *FZD* genes, as well as *GSK3B* and *LEF1*. *GSK3B, FZD8, FRK* and *HDAC1* are druggable targets and several compounds have been approved. These genes are functionally connected with important other upregulated genes of the Wnt-pathway, such as *TP53*, *MYC* or *PPP3CA*. For P5, only two *FZD* genes are overexpressed in cancer, including ERG-induced FZD8. Moreover, *HDAC1* is mutated and *MYC* is repressed. On the other hand, *RANBP2* is uniquely overexpressed in P5. Given its multi-functional role in scaffolding for the Ran-GTPase cycle and nuclear pore complex binding, its overexpres-

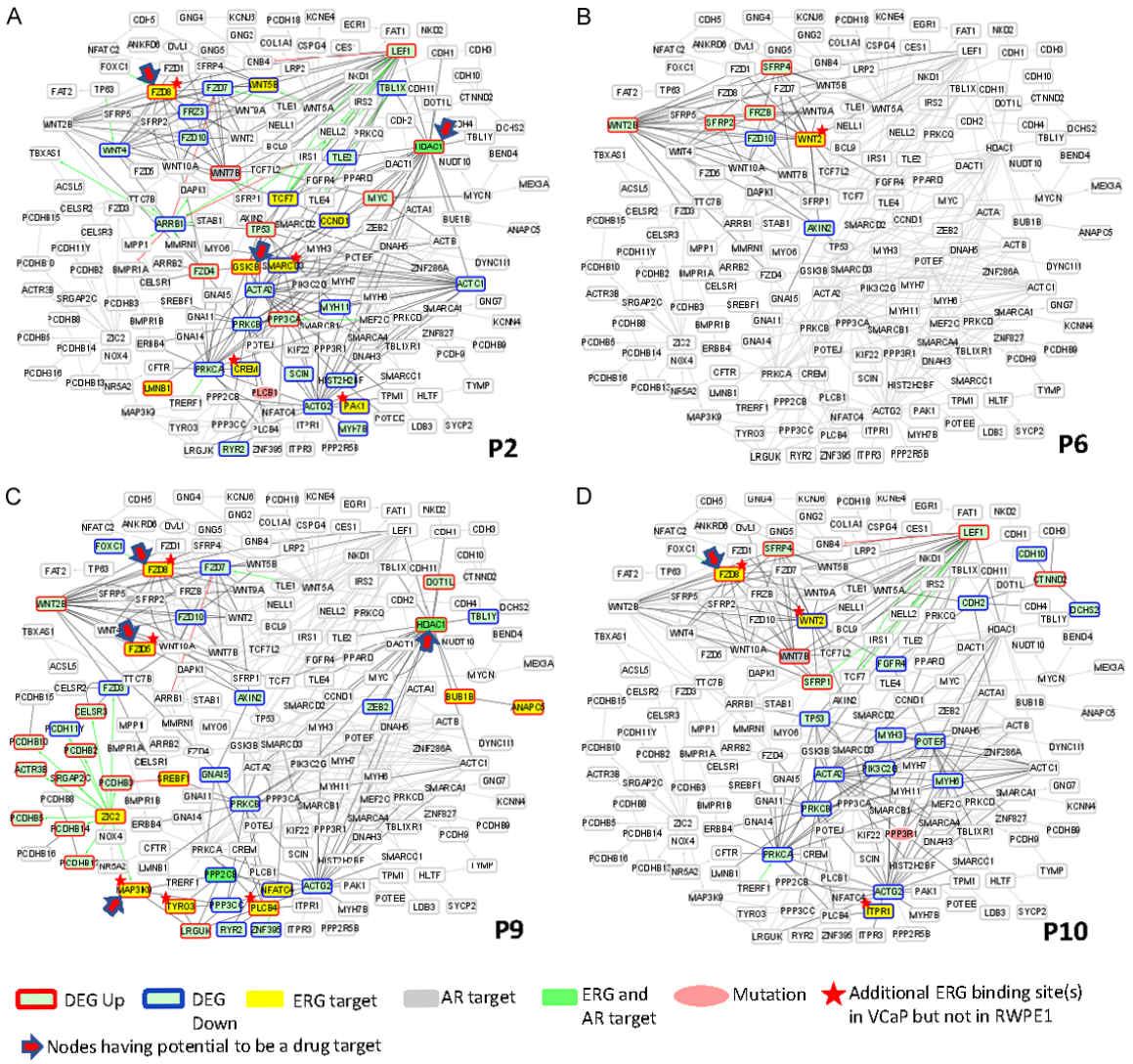


Figure 3. Patient-specific aberrations of the Wnt network are highly divergent. Global networks were established from differentially expressed genes (DEGs, >2-fold) in duplicate patient-matched tumor *vs.* normal samples by integrating the connectivities provided by STRING (protein-protein interaction database) and CellNet (transcription factor-target gene interactions); this integration yielded a 'master network' of DEGs for each patient, revealing the connectivities between deregulated genes. The DEG master network of each patient was complemented by the mutations of predicted high and moderate impact, and the components of the canonical and non-canonical Wnt pathways were extracted. DEGs and mutated genes are depicted in color for (A), P2, (B), P6, (C), P9 and (D), P10 in the background (grey nodes and connectivities) of all merged components of the Wnt pathways that are deregulated or mutated in all 15 patients (Supplementary Figure 4A). Genes for which approved drugs exist or are in clinical trials were identified in drug databases and are indicated by arrows. The corresponding deregulated networks of the other patients are shown in Supplementary Figure 4B-L. When known, connectivities are displayed as green (activation) or red (inhibition) lines; unknown connectivities and protein-protein interactions are displayed as grey lines. DEG specifics and mutations are color-coded as described below the figure.

sion may be an important component of the deregulated network.

Novel approach for in silico validation of deregulated transcription factor cistromes

Even though the above patient-specific DEG networks were derived from true RNA-seq duplicates and several mutations confirmed by Sanger sequencing, we sought additional evidence for validating the network. Given the personalized nature of our study, we excluded cell line or animal studies as proxy. Organoids could not be established either, as all prostates had been deep-frozen. However, we decided to exploit the knowledge existing in databases to

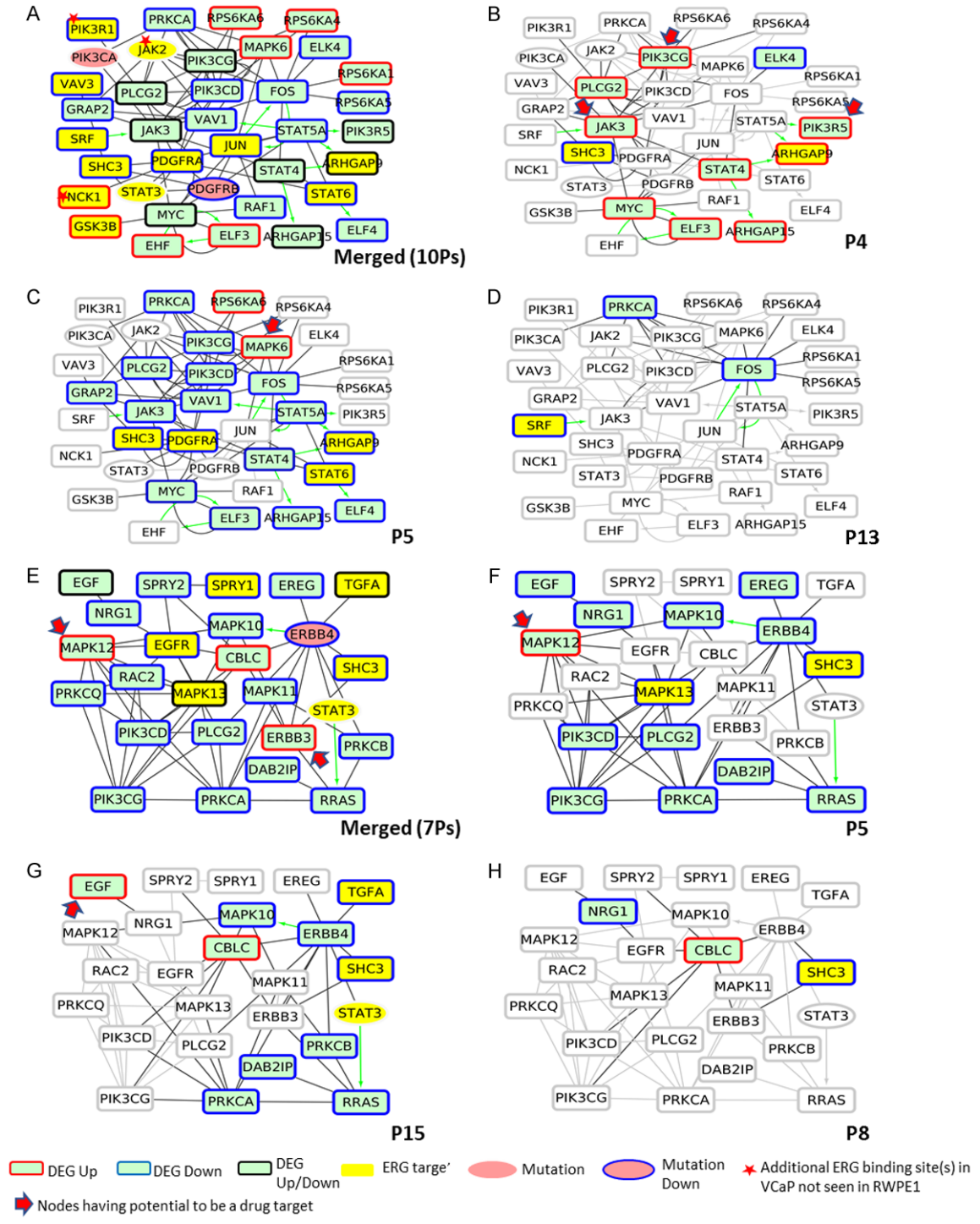


Figure 4. Divergent alterations of the PDGF and EGFR signaling networks in each patient. Ten patients exhibited serious aberrations in the PDGF and 7 patients in the EGFR signaling networks. (A) Merged network of alterations (DEG, mutation) in the PDGF and (E) EGFR networks. Using this merged network as background (grey nodes and connectivities) the aberrations in each individual patient are depicted in color. (B, C) Patients with heavily (P4, P5) or (D), minimally (P13) affected PDGF networks. (F, G) Patients with heavily (P5, P15) or (H) minimally (P8) affected EGFR networks.

provide external supportive evidence for the accuracy of the DEG networks. The rational was as follows: If a network contains upregulated TFs, most if not all of the cognate targets have

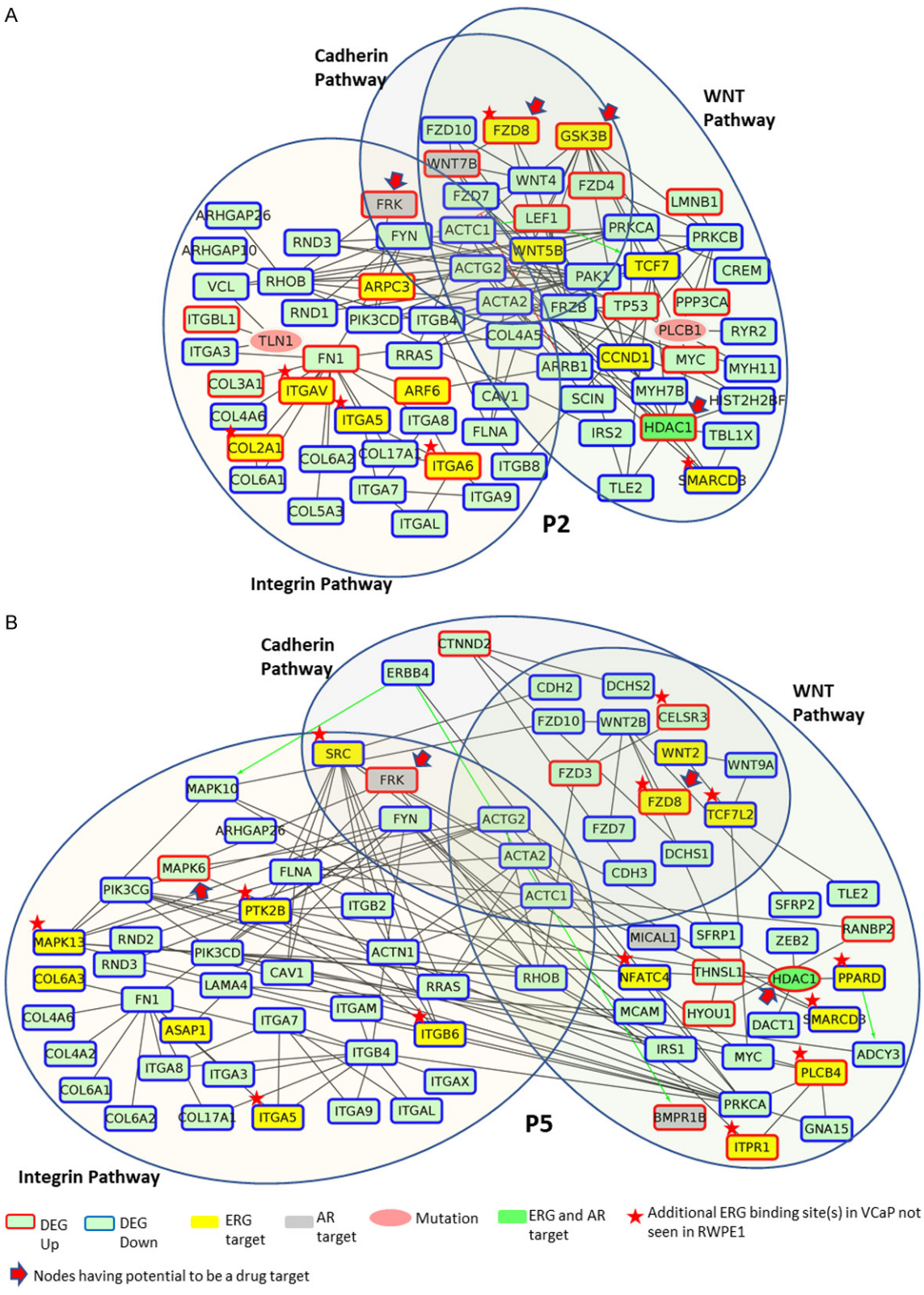


Figure 5. Affected signaling network crosstalk divergently between each other in each patient. (A) Illustration of the merged networks of the affected genes from the Wnt, Cadherin and Integrin signaling pathways in the prostate of P2 revealing that several of the DEGs are common to different pathways. (B) Illustration as in (A) but for P5. Color codes are displayed below the figure.

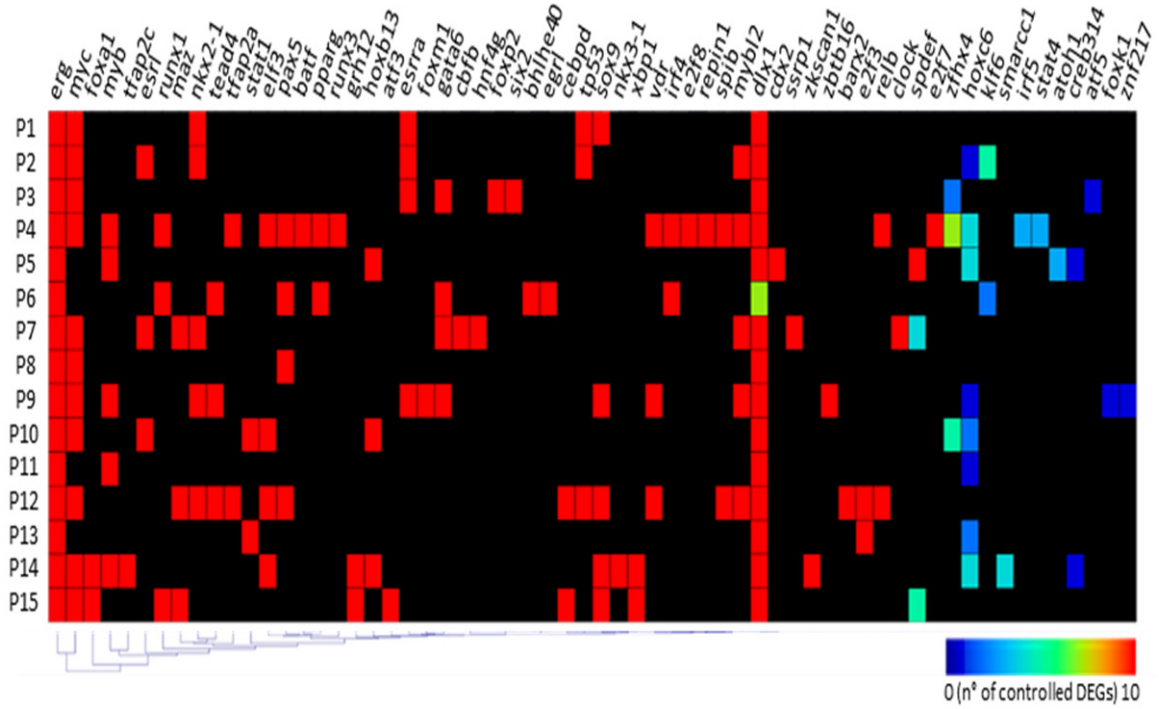


Figure 6. Upregulated TFs associated to DEGs identified for each prostate cancer patient (P1-P15). A large collection of qualified (QC quality A to C) public ChIP-seq datasets for TFs was used for identifying their binding sites (MACS peak calling, pval $<10^{-50}$). Each binding site has been annotated to its most proximal gene promoter (10 kb distance). Only TF-TG (target gene) associations for DEGs retrieved within the 15 patients have been retained. The matrix presents upregulated TFs per patient, the heatmap corresponds to the number of associated DEGs.

been already identified by ChIP-seq. As tumor cells may have TF targets beyond those of the cognate tissue, we collected high confidence targets ($P<10^{-50}$) from all tissues available in the qcGenomics database [41]. We then asked, if the personalized DEG network of a patient contained also the expected cognate TGs of a given TF. In case a significant number of TGs was detected, this provided additional evidence for the accuracy of the network (Supplementary File 4). Indeed, monitoring more than 60 deregulated TFs, the large majority of the corresponding experimentally validated TGs were co-deregulated (**Figure 6**, Supplementary Table 5).

In silico validation of deregulated ERG reveals ribosomal protein genes as frequent targets in addition to patient-specific ones

We tested the validation approach using the TF ERG which is overexpressed in all patient samples due to the TMPRSS2-ERG fusion. In P1, 462 genes were deregulated of which 150 corresponded to ERG TGs. Notably, these ERG target genes contained a strong cluster of ribosomal protein and translation regulatory genes, like *EEF1B2*, *UBA52* or *NCL* (**Figure 7A**). That these genes are *bona fide* ERG target genes was confirmed by qcGenomics profiling of ERG ChIP-seq data sets revealing ERG binding in VCaP (GSM2086309 to GSM2086314) but not in RPWE (GSM2195103, GSM2195106, GSM2195110) cells. While similar results concerning the ERG activation of ribosomal genes were seen for other samples (e.g., P8) with generally less ERG targets being deregulated, several cancers did not reveal such effects despite the overexpression of ERG (e.g., P10, P11, P13, P15; **Figure 7B**, Supplementary Figure 10) but in all cases a network of deregulated ERG target genes was noted. This variability may reflect patient-specific alterations of the chromatin landscape. Together these results show that our novel *in silico* approach is a valuable method to validate patient-specific deregulated TF-TG networks.

Several patient-specific overexpressed TFs deregulate their cognate cistromes

To test if the above approach is also valid for TFs other than ERG we established patient-

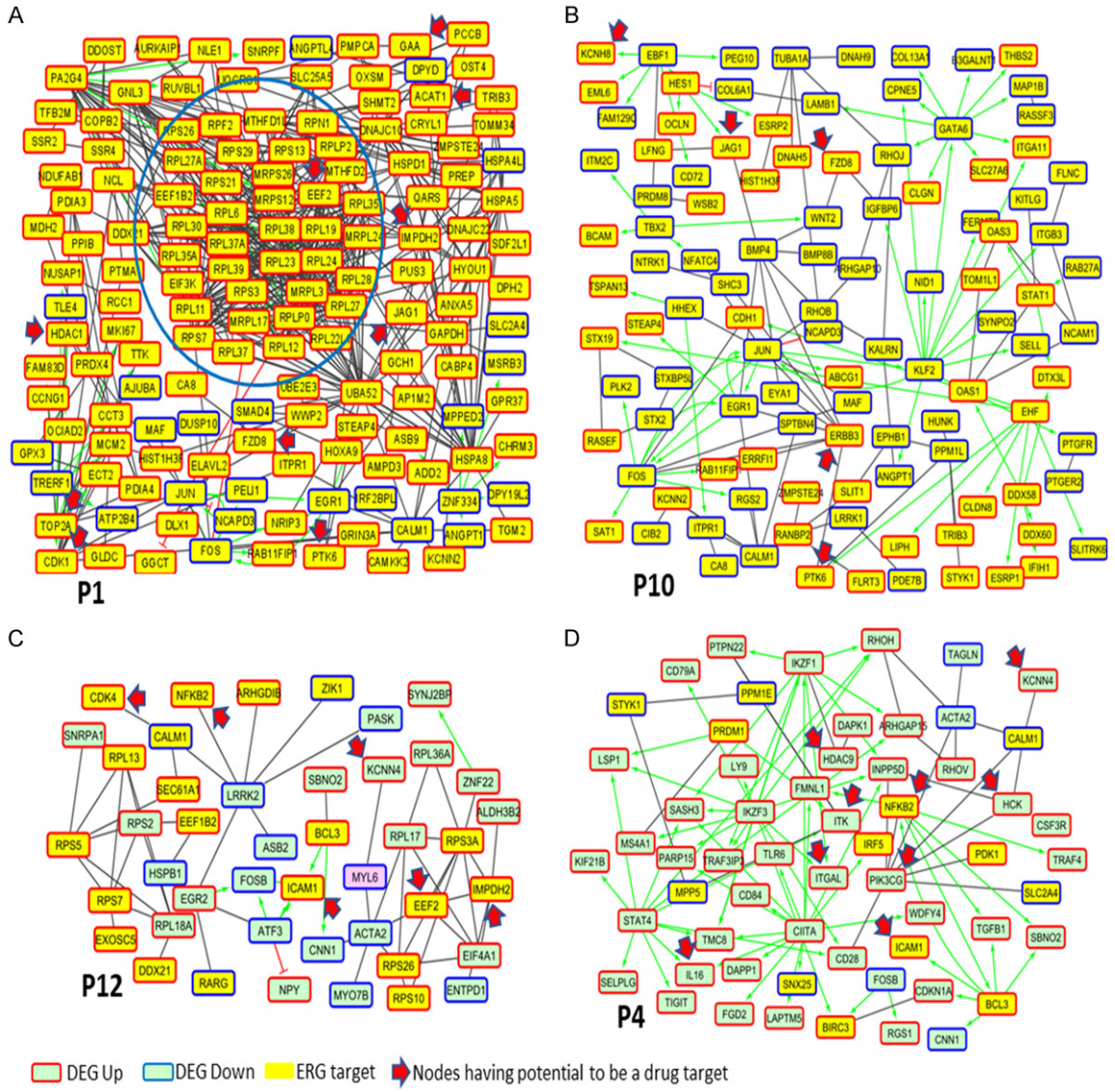


Figure 7. Patient-specific networks of upregulated transcription factor target genes are highly dissimilar. (A) Network of qcGenomics-predicted ERG target genes in P1 and (B) P10. ERG target genes were identified (see text) and the ERG-regulated gene networks were extracted from the master network for each patient. (C) Networks of predicted PAX5 target genes in P12 and (D) P4. Note the strong divergence between the ERG and PAX5 target gene networks between different patient-matched samples. When known, connectivities are displayed as green (activation) or red (inhibition) lines; unknown connectivities and protein-protein interactions are shown as black lines. DEGs and gene mutations are color-coded as described below the figure. The blue circle in (A) reveals tumor-selectively ERG-upregulated ribosomal and translation-associated genes in P1.

specific DEG TG networks for PAX5 (**Figure 7C**, **7D**) and MAZ (Supplementary Figure 11). In all cases the DEG networks are significantly populated by PAX5 and MAZ target genes. Note however, that the repertoire of target genes for a given TF can significantly differ from patient to patient, despite its common overexpression. Interestingly, MAZ can also bind to ribosomal protein genes. Indeed this is fully supported by the corresponding ChIP-seq data using qcGenomics (GSM93-5337, GSM1003613, GSM935272, GSM935-335), even though data are only available for K562, IMR90, Hela and HepG2 cells. The above integration of TF cistrome data further supports our overall notion that in individual tumors very different gene networks can be deregulated with different sets of overexpressed potentially druggable targets.

Discussion

Subsequent to the discoveries of oncogenes and tumor suppressors, and the concept of drivers and passengers of tumorigenesis [42, 43], the enormous progress in genome-wide sequencing together with a plethora of functional genomics applications has raised hopes that cancer genomics will rapidly reveal genome alterations causal to the disease and provide novel targets for therapy. However, like in human genetics, we face a scenario in which the origin of monogenic diseases has been largely deciphered, while the deregulated networks underlying multigenic diseases remain unknown. Indeed, despite several success stories, the anticipated rapid translation from cancer genomics to therapies did not occur. Rather, these studies revealed an ever-increasing complexity of multiple deregulated systems in tumors, intra-tumor and inter-patient heterogeneity and the incomplete understanding of the affected regulatory pathways operating within and between cells and tissues. This complexity is likely the major caveat for translating cancer genomics towards therapy. Moreover, the common approach of comparing hundreds to thousands of patients to identify individual targets may be conceptually problematic, as (i) new single denominators, in addition to those already discovered, may not exist and the deregulation/mutation of multiple interacting genes/pathways may be critically involved in the origin/evolution of the disease. Moreover, (ii) inter-tumoral/inter-individual variation will be disregarded by this approach, as well as (iii) the altered cross-talks between pathways. Finally, (iv) genes may exert distinct functions in different pathways/communication networks and (v) act as functionally divergent paralogues, like the multiple Wnt genes.

Previously, numerous studies were performed to unravel novel potential therapeutic targets using large-scale genomic [2, 44, 45] and integrative transcriptome-genome analyses [1, 8, 46, 47]. Multiple genome-wide association studies led to the identification of more than 160 disease-susceptible loci, most of which have unknown clinical implications [reviewed by 3]. A recent study involving large numbers of PrCa transcriptome profiles from 38 cohorts, developed a classification system based on pathway activation signals and presented a 37-gene signature which can classify PrCa into 3 subtypes [48]. More recently, the analysis of 18 recurrent DNA- and RNA-based genomic alterations, including androgen receptor variant expression and transcriptional output, and neuroendocrine expression signatures revealed *RB1* as the only gene associated with clinical outcome [4]. All these studies focused on common high-frequency targets. Somewhat disturbingly, each study revealed different (sets of) targets and a potential therapeutic application was postulated in several cases. In addition, as pointed out in a recent report that highlighted the role of low-frequency mutations in cancer progression [29], these approaches will miss low-frequency targets. That all the above-mentioned studies do not focus on individual patients builds strong case for a thorough analysis of mutations/aberrations at the individual level. In this respect, a recent network-based integrative study used genomic, transcriptomic and phosphoproteomic datasets to compare treatment-naive and metastatic PrCa and suggested personalized signatures in individual patients [49].

To provide insight into the various aspects of heterogeneity and with the aim of developing a pathway/network-centric rather than a gene-centric approach, we assessed the complexity of alteration occurring during tumorigenesis in the well-defined main class of prostate tumors with the TMPRSS2-ERG fusion [6] by comparing the tumor and its adjacent apparently normal tissue from the same untreated patients. Indeed, a recent report revealed that the use of paired tumor-normal samples improved mutation identification and decreased false-positive rates [50].

This patient-centered network analysis revealed highly divergent patient-specific deregulated and mutated genomic landscapes. All 15 patients except P6 revealed aberrations (mutation, expression) of one or several components of the WNT, Cadherin and Integrin pathways with large differences between deregulated pathway components in each of the patients, thus advocating the need for a patient-centered analysis. As previously pointed out, targeting the WNT pathway is challenging due to its complex nature driving diverse biological processes and cross-talk with multiple other pathways [51]. Therefore, a thorough understanding

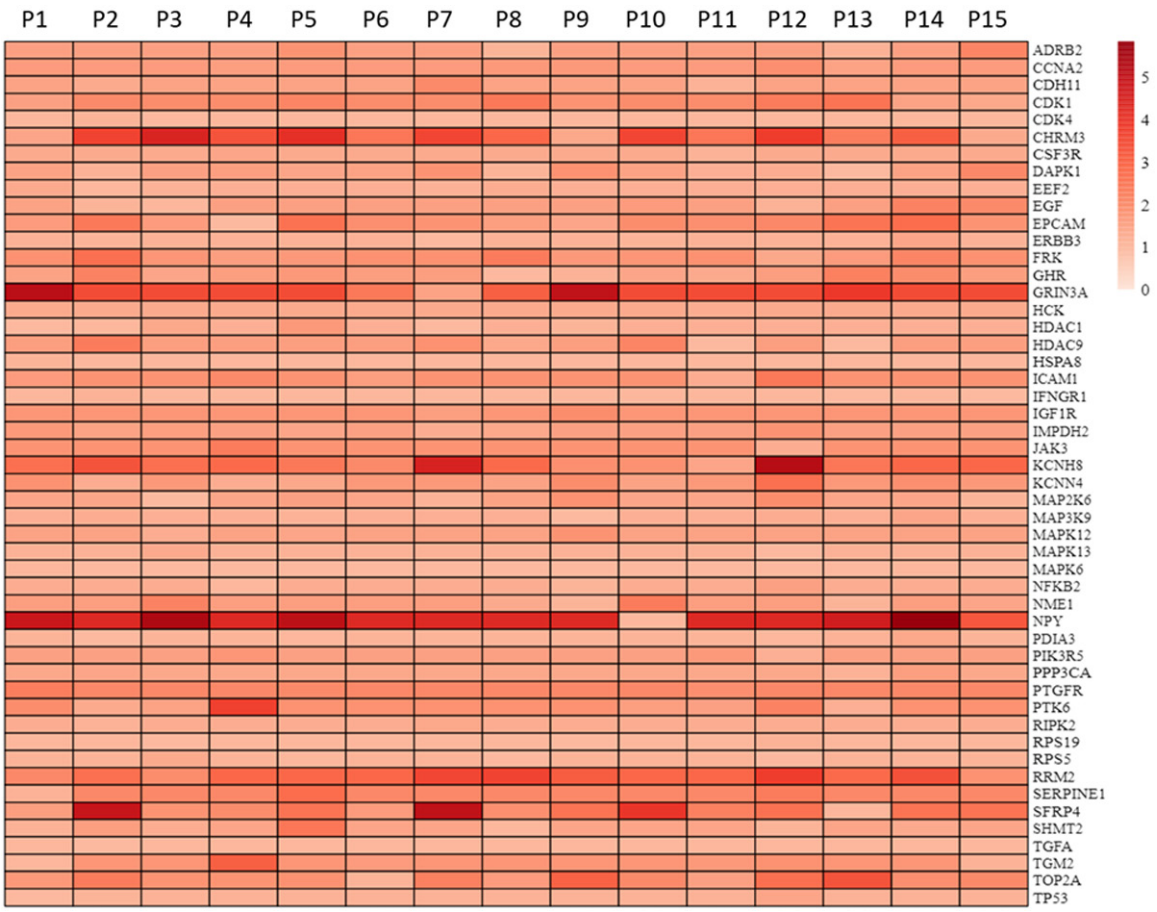


Figure 8. Heat map revealing the extent of upregulation of genes whose products correspond to validated drug targets. As discussed in the text, only gene products for which drugs have been approved or have been/are enrolled in clinical trials are considered, irrespective of the disease for which they have been/are being developed. In the network figures, these genes are marked with a red asterisk. Color codes represent log2-fold changes. P1 to P15, patients 1 to 15. Color codes represent log2 fold change.

about each component of these interacting pathways and key nodes as potential therapeutic target, based on a systems medicine approach [52] with patient-specific networks, is the ideal way to move forward towards targeted therapy. Indeed, patients who received a personalized therapy adapted to their genomic aberrations, as recommended by a multidisciplinary molecular tumor board, had improved oncological outcomes including survival [53]. Our patient-specific network analysis facilitated the identification of key nodes involved in the cross-talk between different deregulated pathways in the same patient. In this respect, we have identified a significant number of genes which correspond to validated drug targets (**Figure 8**). The extent of these deregulated genes varies largely between patients and several highly upregulated genes occur in individual patients (e.g. P1, P7, P9, P12, P14). This information will be useful to decide about combinatorial therapies for individual patients by targeting key nodes of different deregulated pathways. For example, in P2, the upregulated direct ERG targets *ITGAV* and *ITGA6* (Integrin pathway) along with *FZD8* and *GSK3B* (Wnt and Cadherin pathway) could be used as potential drug targets for combinatorial therapy. Similarly, ETS family members ERG and ETV1, overexpressed due to the fusion with TMPRSS2, were reported to directly suppress CHK1 promoting tumorigenesis bypassing DNA damage response [40]. Our patient-specific network analysis in the present study revealed several deregulated genes that acquired additional ERG binding in their respective promoters. These direct targets of ERG, which are involved in different cooperating pathways, correspond

to potential therapeutic targets. Based on the above results and reflections, we propose to develop therapeutic options in the context of a personalized integrative functional genomics analysis rather than trying to identify common single targets from the analysis of large numbers of patients.

Around 97 percent of potential drugs undergoing clinical trials fail to get FDA approval [54]. The off-target toxicity of cancer drugs undergoing clinical trials [55] suggests the need for more robust genetic analysis when predicting the potential drug target. Multilayered patient-specific network analysis will be useful to identify not only putative drug targets but also predict potential off-target effects resulting from pathway cross-talks. Apart from identifying targets of potential therapeutic use, there is strong need for development of novel drugs which can target components of complex pathways like WNT signaling. In this respect, patient-derived organoid cultures [56] which recapitulate the diversity of primary tumors may facilitate screening of novel molecules against these putative therapeutic targets.

In addition, there is a growing importance of single-cell functional genomics done with circulating tumor cells for diagnosis. Recent study involving pan-cancer analysis of chromatin accessibility revealed novel protein-DNA interactions in primary cancer tissues [57]. Integrating chromatin accessibility data from individual patients can identify cancer-specific novel regulatory connections which can be used as potential drug target. Ultimately, additional dimensions like RNA regulators, such as the newly described circular RNAs [58, 59], as well as metabolomics changes, may be integrated in this analysis to reveal what communication networks are at the origin, maintenance and progression of the disease and which regulatory circuits can be modulated for therapeutic purposes, including escape from resistance to therapy.

Acknowledgements

We dedicate this work to the memory of Dr. Catherine Mazerolles who sadly passed away when this study was gaining full speed. We thank all members of our labs for discussions and suggestions. These studies were supported by funds from the Plan Cancer, AVIESAN-ITMO Cancer to HG and LV, the Ligue National Contre le Cancer (HG; Equipe Labellisée); and the Institut National du Cancer (INCa) to HG and LV. Support of the Agence Nationale de la Recherche (ANRT-07-PCVI-0031-01, ANR-10-LABX-0030-INRT and ANR-10-IDEX-0002-02) is acknowledged. Samples from non-treated patients were obtained after written informed consent in accordance with the Declaration of Helsinki and stored at the « CRB Cancer des Hôpitaux de Toulouse (BB-0033-00014) » collection. According to the French law, the CRB Cancer collection has been declared to the Ministry of Higher Education and Research (DC-2008-463) and obtained a transfer agreement (AC-2013-1955) after approbation by ethical committees. Clinical and biological annotations of the samples have been declared to the CNIL (Commission Nationale de l'Informatique et des Libertés).

Disclosure of conflict of interest

None.

Address correspondence to: Hinrich Gronemeyer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Department of Functional Genomics and Cancer, Illkirch, France. E-mail: hg@igbmc.fr; Laurence Vandel, Université Clermont Auvergne, CNRS, Inserm, GReD, Clermont-Ferrand, France. E-mail: laurence.vandel@uca.fr

References

[1] Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, Beltran H, Abida W, Bradley RK, Vinson J, Cao X, Vats P, Kunju LP, Hussain M, Feng FY, Tomlins SA, Cooney KA, Smith DC, Brennan C, Siddiqui J, Mehra R, Chen Y, Rathkopf DE, Morris MJ, Solomon SB, Durack JC, Reuter VE, Gopalan A, Gao J, Loda M, Lis RT, Bowden M, Balk SP, Gaviola G, Sougnez C, Gupta M, Yu EY, Mostaghel EA, Cheng HH, Mulcahy H, True LD, Plymate SR, Dvinge H, Ferraldeschi R, Flohr P, Miranda S, Zafeiriou Z, Tunariu N, Mateo J, Perez-Lopez R, Demichelis F, Robinson BD, Schiffman M, Nanus DM, Tagawa ST, Sigaras A, Eng KW, Elemento O, Sboner A, Heath EI, Scher HI, Pienta KJ, Kantoff P, de Bono JS, Rubin MA, Nelson PS, Garraway LA, Sawyers CL and Chinnaiyan AM. Integrative clinical genomics of advanced prostate cancer. Cell 2015; 161: 1215-1228.

[2] Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, Van Allen E, Kryukov GV, Sboner A, Theurillat JP, Soong TD, Nicker-

son E, Auclair D, Tewari A, Beltran H, Onofrio RC, Boysen G, Guiducci C, Barbieri CE, Cibulskis K, Sivachenko A, Carter SL, Saksena G, Voet D, Ramos AH, Winckler W, Cipicchio M, Ardlie K, Kantoff PW, Berger MF, Gabriel SB, Golub TR, Meyerson M, Lander ES, Elemento O, Getz G, Demichelis F, Rubin MA and Garraway LA. Punctuated evolution of prostate cancer genomes. Cell 2013; 153: 666-677.

[3] Farashi S, Kryza T, Clements J and Batra J. Post-GWAS in prostate cancer: from genetic association to biological contribution. Nat Rev Cancer 2019; 19: 46-59.

[4] Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, Cieslik M, Benelli M, Robinson D, Van Allen EM, Sboner A, Fedrizzi T, Mosquera JM, Robinson BD, De Sarkar N, Kunju LP, Tomlins S, Wu YM, Nava Rodrigues D, Loda M, Gopalan A, Reuter VE, Pritchard CC, Mateo J, Bianchini D, Miranda S, Carreira S, Rescigno P, Filipenko J, Vinson J, Montgomery RB, Beltran H, Heath EI, Scher HI, Kantoff PW, Taplin ME, Schultz N, deBono JS, Demichelis F, Nelson PS, Rubin MA, Chinnaiyan AM and Sawyers CL. Genomic correlates of clinical outcome in advanced prostate cancer. Proc Natl Acad Sci U S A 2019; 116: 11428-11436.

[5] Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. Cell 2015; 163: 1011-1025.

[6] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA and Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005; 310: 644-648.

[7] Chen CD, Welsbie DS, Tran C, Baek SH, Chen R, Vessella R, Rosenfeld MG and Sawyers CL. Molecular determinants of resistance to antiandrogen therapy. Nat Med 2004; 10: 33-39.

[8] Haffner MC, Mosbruger T, Esopi DM, Fedor H, Heaphy CM, Walker DA, Adejola N, Gürel M, Hicks J, Meeker AK, Halushka MK, Simons JW, Isaacs WB, De Marzo AM, Nelson WG and Yegnasubramanian S. Tracking the clonal origin of lethal prostate cancer. J Clin Invest 2013; 123: 4918-4922.

[9] Yi S, Lin S, Li Y, Zhao W, Mills GB and Sahni N. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. Nat Rev Genet 2017; 18: 395-410.

[10] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and DePristo MA. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.

[11] Koster J and Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinformatics 2012; 28: 2520-2522.

[12] Garofalo M, Di Leva G, Romano G, Nuovo G, Suh SS, Ngankeu A, Taccioli C, Pichiorri F, Alder H, Secchiero P, Gasparini P, Gonelli A, Costinean S, Acunzo M, Condorelli G and Croce CM. miR-221&222 regulate TRAIL resistance and enhance tumorigenicity through PTEN and TIMP3 downregulation. Cancer Cell 2009; 16: 498-509.

[13] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R; Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25: 2078-2079.

[14] Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES and Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013; 31: 213-219.

[15] Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X and Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6: 80-92.

[16] Kim D, Langmead B and Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015; 12: 357-360.

[17] Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT and Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol 2013; 9: e1003118.

[18] Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.

[19] Tabas-Madrid D, Nogales-Cadenas R and Pascual-Montano A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res 2012; 40: W478-483.

[20] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ and von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017; 45: D362-D368.

[21] Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ and Collins JJ. CellNet: network biology applied to stem cell engineering. Cell 2014; 158: 903-915.

[22] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

[23] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and von Mering C. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.

[24] Kypta RM and Waxman J. Wnt/β-catenin signalling in prostate cancer. Nat Rev Urol 2012; 9: 418-428.

[25] Jalota A, Singh K, Pavithra L, Kaul-Ghanekar R, Jameel S and Chattopadhyay S. Tumor suppressor SMAR1 activates and stabilizes p53 through its arginine-serine-rich motif. J Biol Chem 2005; 280: 16019-16029.

[26] Ishii H, Vecchione A, Murakumo Y, Baldassarre G, Numata S, Trapasso F, Alder H, Baffa R and Croce CM. FEZ1/LZTS1 gene at 8p22 suppresses cancer cell growth and regulates mitosis. Proc Natl Acad Sci U S A 2001; 98: 10374-10379.

[27] Vecchione A, Baldassarre G, Ishii H, Nicoloso MS, Belletti B, Petrocca F, Zanesi N, Fong LY, Battista S, Guarnieri D, Baffa R, Alder H, Farber JL, Donovan PJ and Croce CM. Fez1/Lzts1 absence impairs Cdk1/Cdc25C interaction during mitosis and predisposes mice to cancer development. Cancer Cell 2007; 11: 275-289.

[28] Shen M, Jiang YZ, Wei Y, Ell B, Sheng X, Esposito M, Kang J, Hang X, Zheng H, Rowicki M, Zhang L, Shih WJ, Celià-Terrassa T, Liu Y, Cristea I, Shao ZM and Kang Y. Tinagl1 suppresses triple-negative breast cancer progression and metastasis by simultaneously inhibiting integrin/FAK and EGFR signaling. Cancer Cell 2019; 35: 64-80, e7.

[29] Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, Chatila WK, Chakravarty D, Han GC, Coleman I, Montgomery B, Pritchard C, Morrissey C, Barbieri CE, Beltran H, Sboner A, Zafeiriou Z, Miranda S, Bielski CM, Penson AV, Tolonen C, Huang FW, Robinson D, Wu YM, Lonigro R, Garraway LA, Demichelis F, Kantoff PW, Taplin ME, Abida W, Taylor BS, Scher HI, Nelson PS, de Bono JS, Rubin MA, Sawyers CL and Chinnaiyan AM; PCF/SU2C International Prostate Cancer Dream Team, Schultz N, Van Allen EM. The long tail of oncogenic drivers in prostate cancer. Nat Genet 2018; 50: 645-651.

[30] Jin Y, Qu S, Tesikova M, Wang L, Kristian A, Mælandsmo GM, Kong H, Zhang T, Jerónimo C, Teixeira MR, Yuca E, Tekedereli I, Gorgulu K, Alpay N, Sood AK, Lopez-Berestein G, Danielsen HE, Ozpolat B and Saatcioglu F. Molecular circuit involving KLK4 integrates androgen and mTOR signaling in prostate cancer. Proc Natl Acad Sci U S A 2013; 110: E2572-2581.

[31] Ramsay AJ, Dong Y, Hunt ML, Linn M, Samaratunga H, Clements JA and Hooper JD. Kallikrein-related peptidase 4 (KLK4) initiates intracellular signaling via protease-activated receptors (PARs). KLK4 and PAR-2 are co-expressed during prostate cancer progression. J Biol Chem 2008; 283: 12293-12304.

[32] Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, Boeckx B, Wijnhoven PW, Radaelli E, Vermi W, Leucci E, Lapouge G, Beck B, van den Oord J, Nakagawa S, Hirose T, Sablina AA, Lambrechts D, Aerts S, Blanpain C and Marine JC. p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. Nat Med 2016; 22: 861-868.

[33] Mello SS, Sinow C, Raj N, Mazur PK, Bieging-Rolett K, Broz DK, Imam JFC, Vogel H, Wood LD, Sage J, Hirose T, Nakagawa S, Rinn J and Attardi LD. Neat1 is a p53-inducible lincRNA essential for transformation suppression. Genes Dev 2017; 31: 1095-1108.

[34] Zhang Y, Pitchiaya S, Cieślik M, Niknafs YS, Tien JC, Hosono Y, Iyer MK, Yazdani S, Subramaniam S, Shukla SK, Jiang X, Wang L, Liu TY, Uhl M, Gawronski AR, Qiao Y, Xiao L, Dhanasekaran SM, Juckette KM, Kunju LP, Cao X, Patel U, Batish M, Shukla GC, Paulsen MT, Ljungman M, Jiang H, Mehra R, Backofen R, Sahinalp CS, Freier SM, Watt AT, Guo S, Wei JT, Feng FY, Malik R and Chinnaiyan AM. Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. Nat Genet 2018; 50: 814-824.

[35] Malek R, Gajula RP, Williams RD, Nghiem B, Simons BW, Nugent K, Wang H, Taparra K, Lemtiri-Chlieh G, Yoon AR, True L, An SS, De-Weese TL, Ross AE, Schaeffer EM, Pienta KJ, Hurley PJ, Morrissey C and Tran PT. TWIST1-WDR5-Hottip regulates Hoxa9 chromatin to facilitate prostate cancer metastasis. Cancer Res 2017; 77: 3181-3193.

[36] Malik R, Khan AP, Asangani IA, Cieślik M, Prensner JR, Wang X, Iyer MK, Jiang X, Borkin D, Escara-Wilke J, Stender R, Wu YM, Niknafs YS, Jing X, Qiao Y, Palanisamy N, Kunju LP, Krishnamurthy PM, Yocum AK, Mellacheruvu D, Nesvizhskii AI, Cao X, Dhanasekaran SM, Feng FY, Grembecka J, Cierpicki T and Chinnaiyan

AM. Targeting the MLL complex in castration-resistant prostate cancer. Nat Med 2015; 21: 344-352.
[37] Rupaimoole R and Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov 2017; 16: 203-222.
[38] Mi H, Muruganujan A, Ebert D, Huang X and Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 2019; 47: D419-D426.
[39] Clara JA, Monge C, Yang Y and Takebe N. Targeting signalling pathways and the immune microenvironment of cancer stem cells - a clinical update. Nat Rev Clin Oncol 2020; 17: 204-232.
[40] Lunardi A, Varmeh S, Chen M, Taulli R, Guarnerio J, Ala U, Seitzer N, Ishikawa T, Carver BS, Hobbs RM, Quarantotti V, Ng C, Berger AH, Nardella C, Poliseno L, Montironi R, Castillo-Martin M, Cordon-Cardo C, Signoretti S and Pandolfi PP. Suppression of CHK1 by ETS family members promotes DNA damage response bypass and tumorigenesis. Cancer Discov 2015; 5: 550-563.
[41] Blum M, Cholley PE, Malysheva V, Nicaise S, Moehlin J, Gronemeyer H and Mendoza-Parra MA. A comprehensive resource for retrieving, visualizing, and integrating functional genomics data. Life Science Alliance 2019; 3: e201900546.
[42] Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B and Nowak MA. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 2010; 107: 18545-18550.
[43] Haber DA and Settleman J. Cancer: drivers and passengers. Nature 2007; 446: 145-146.
[44] Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, Nickerson E, Chae SS, Boysen G, Auclair D, Onofrio RC, Park K, Kitabayashi N, MacDonald TY, Sheikh K, Vuong T, Guiducci C, Cibulskis K, Sivachenko A, Carter SL, Saksena G, Voet D, Hussain WM, Ramos AH, Winckler W, Redman MC, Ardlie K, Tewari AK, Mosquera JM, Rupp N, Wild PJ, Moch H, Morrissey C, Nelson PS, Kantoff PW, Gabriel SB, Golub TR, Meyerson M, Lander ES, Getz G, Rubin MA and Garraway LA. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012; 44: 685-689.
[45] Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, Asangani IA, Ateeq B, Chun SY, Siddiqui J, Sam L, Anstett M, Mehra R, Prensner JR, Palanisamy N, Ryslik GA, Vandin F, Raphael BJ, Kunju LP, Rhodes DR, Pienta KJ, Chinnaiyan AM and Tomlins SA. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012; 487: 239-243.
[46] Kron KJ, Murison A, Zhou S, Huang V, Yamaguchi TN, Shiah YJ, Fraser M, van der Kwast T, Boutros PC, Bristow RG and Lupien M. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nat Genet 2017; 49: 1336-1345.
[47] Mancuso N, Gayther S, Gusev A, Zheng W, Penney KL, Kote-Jarai Z, Eeles R, Freedman M, Haiman C, Pasaniuc B and consortium P. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. Nat Commun 2018; 9: 4079.
[48] You S, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Al-Deen Ashab H, Davicioni E, Karnes RJ, Klein EA, Den RB, Ross AE, Schaeffer EM, Garraway IP, Kim J and Freeman MR. Integrated classification of prostate cancer reveals a novel luminal subtype with poor outcome. Cancer Res 2016; 76: 4948-4958.
[49] Drake JM, Paull EO, Graham NA, Lee JK, Smith BA, Titz B, Stoyanova T, Faltermeier CM, Uzunangelov V, Carlin DE, Fleming DT, Wong CK, Newton Y, Sudha S, Vashisht AA, Huang J, Wohlschlegel JA, Graeber TG, Witte ON and Stuart JM. Phosphoproteome integration reveals patient-specific networks in prostate cancer. Cell 2016; 166: 1041-1054.
[50] Beaubier N, Bontrager M, Huether R, Igartua C, Lau D, Tell R, Bobe AM, Bush S, Chang AL, Hoskinson DC, Khan AA, Kudalkar E, Leibowitz BD, Lozachmeur A, Michuda J, Parsons J, Perera JF, Salahudeen A, Shah KP, Taxter T, Zhu W and White KP. Integrated genomic profiling expands clinical options for patients with cancer. Nat Biotechnol 2019; 37: 1351-1360.
[51] Kahn M. Can we safely target the WNT pathway? Nat Rev Drug Discov 2014; 13: 513-532.
[52] Sonawane AR, Weiss ST, Glass K and Sharma A. Network medicine in the age of biomedical big data. Front Genet 2019; 10: 294.
[53] Kato S, Kim KH, Lim HJ, Boichard A, Nikanjam M, Weihe E, Kuo DJ, Eskander RN, Goodman A, Galanina N, Fanta PT, Schwab RB, Shatsky R, Plaxe SC, Sharabi A, Stites E, Adashek JJ, Okamura R, Lee S, Lippman SM, Sicklick JK and Kurzrock R. Real-world data from a molecular tumor board demonstrates improved outcomes with a precision N-of-One strategy. Nat Commun 2020; 11: 4965.
[54] Wong CH, Siah KW and Lo AW. Estimation of clinical trial success rates and related parameters (vol 20, pg 273, 2019). Biostatistics 2019; 20: 366-366.

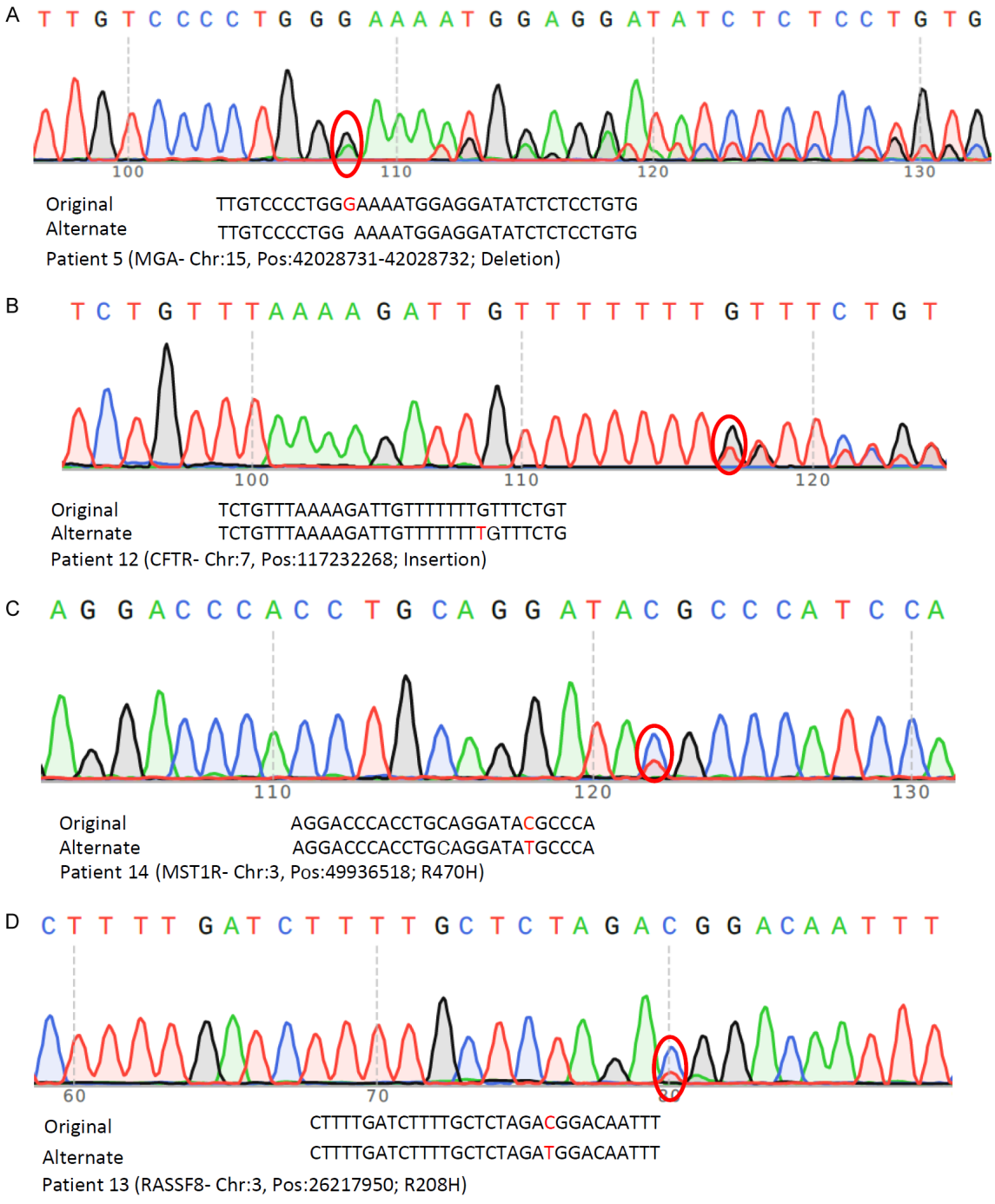
[55] Lin A, Giuliano CJ, Palladino A, John KM, Abramowicz C, Yuan ML, Sausville EL, Lukow DA, Liu L, Chait AR, Galluzzo ZC, Tucker C and Sheltzer JM. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Sci Transl Med 2019; 11: eaaw8412.

[56] Gao D and Chen Y. Organoid development in cancer genome discovery. Curr Opin Genet Dev 2015; 30: 42-48.

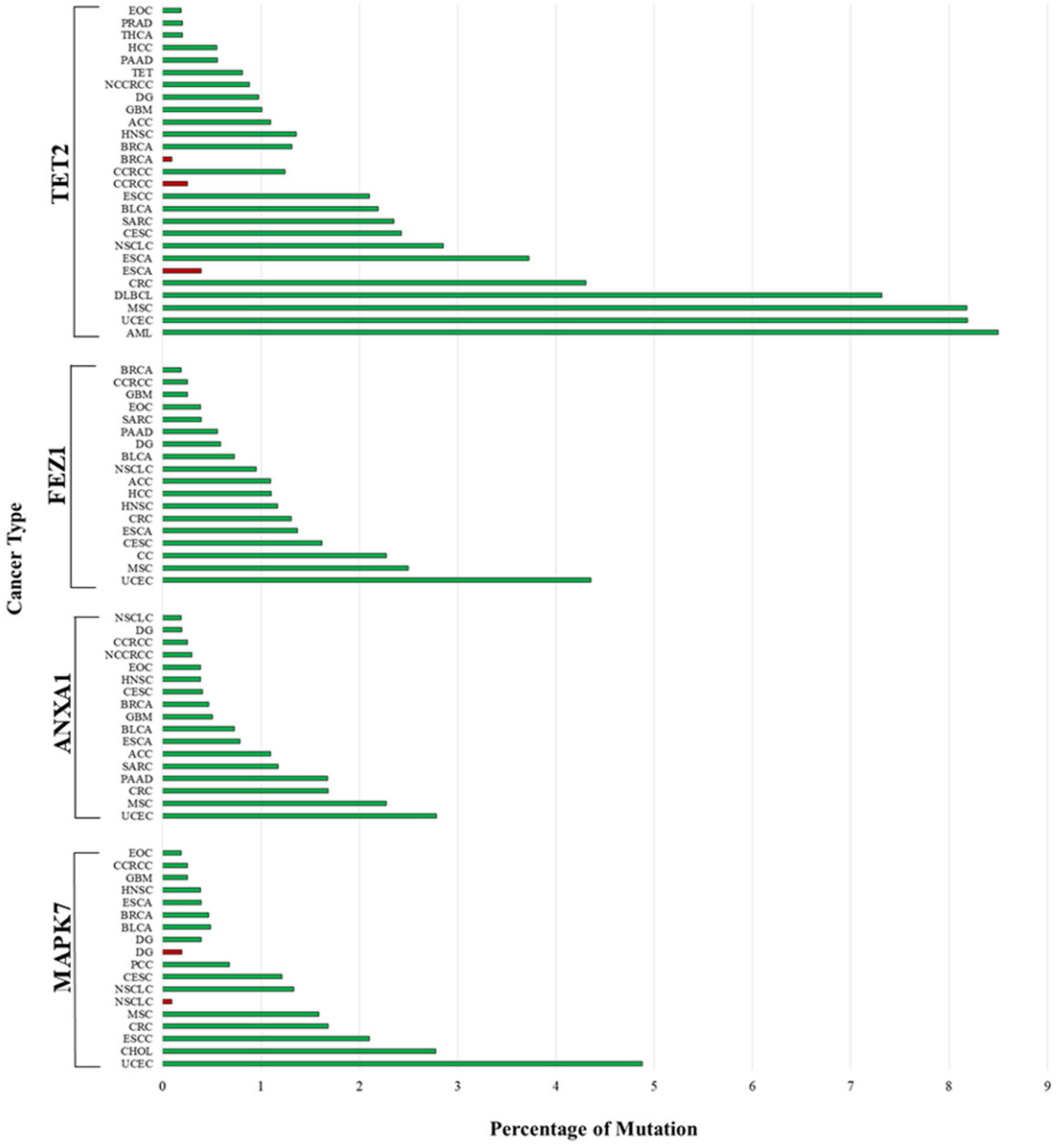
[57] Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW, Satpathy AT, Mumbach MR, Hoadley KA, Robertson AG, Sheffield NC, Felau I, Castro MAA, Berman BP, Staudt LM, Zenklusen JC, Laird PW and Curtis C; Cancer Genome Atlas Analysis Network; Greenleaf WJ and Chang HY. The chromatin accessibility landscape of primary human cancers. Science 2018; 362: eaav1898.

[58] Chen S, Huang V, Xu X, Livingstone J, Soares F, Jeon J, Zeng Y, Hua JT, Petricca J, Guo H, Wang M, Yousif F, Zhang Y, Donmez N, Ahmed M, Volik S, Lapuk A, Chua MLK, Heisler LE, Foucal A, Fox NS, Fraser M, Bhandari V, Shiah YJ, Guan J, Li J, Orain M, Picard V, Hovington H, Bergeron A, Lacombe L, Fradet Y, Tetu B, Liu S, Feng F, Wu X, Shao YW, Komor MA, Sahinalp C, Collins C, Hoogstrate Y, de Jong M, Fijneman RJA, Fei T, Jenster G, van der Kwast T, Bristow RG, Boutros PC and He HH. Widespread and functional RNA circularization in localized prostate cancer. Cell 2019; 176: 831-843, e22.

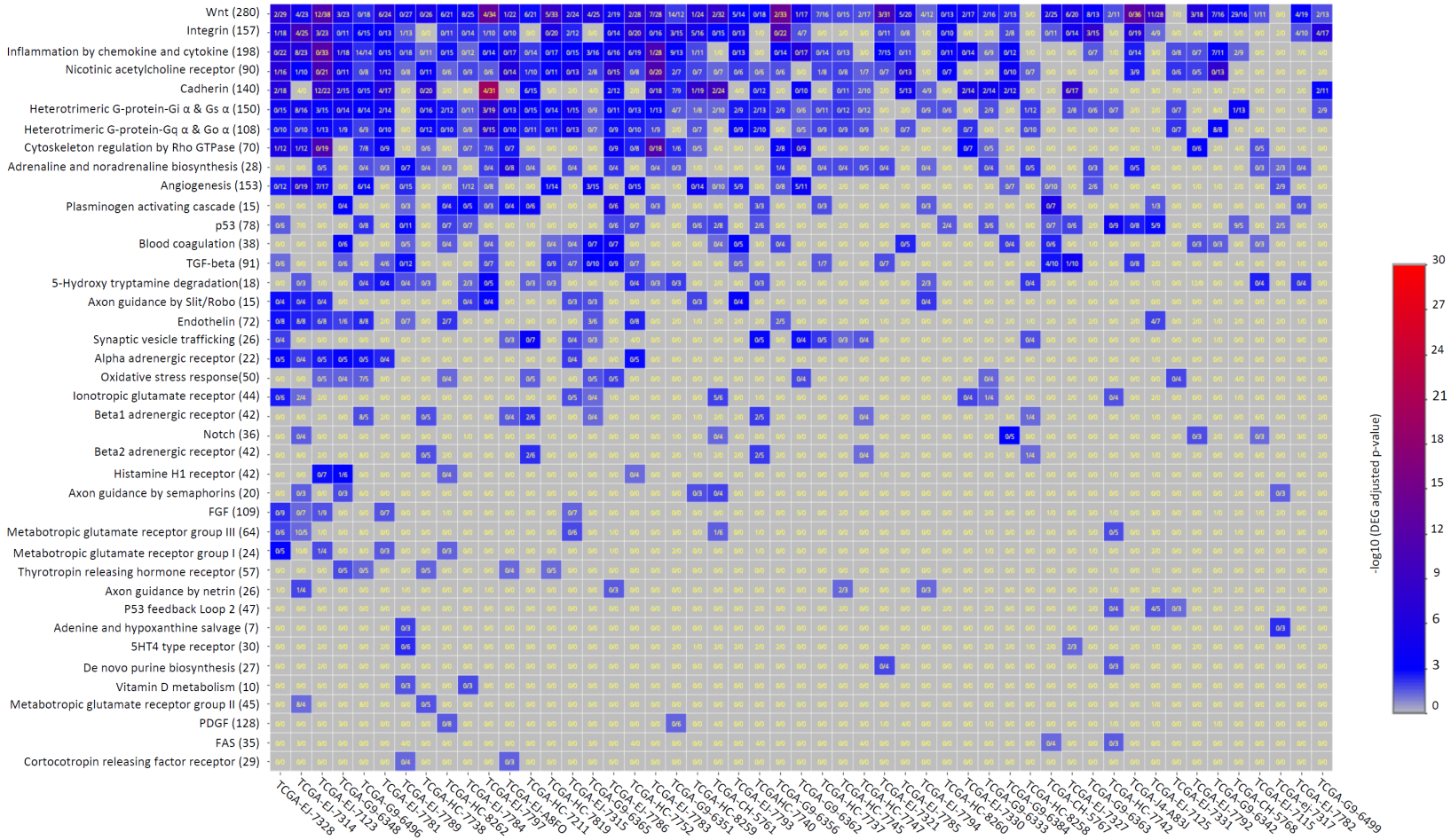
[59] Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, Wu YM, Dhanasekaran SM, Engelke CG, Cao X, Robinson DR, Nesvizhskii AI and Chinnaiyan AM. The landscape of circular RNA in cancer. Cell 2019; 176: 869-881, e13.



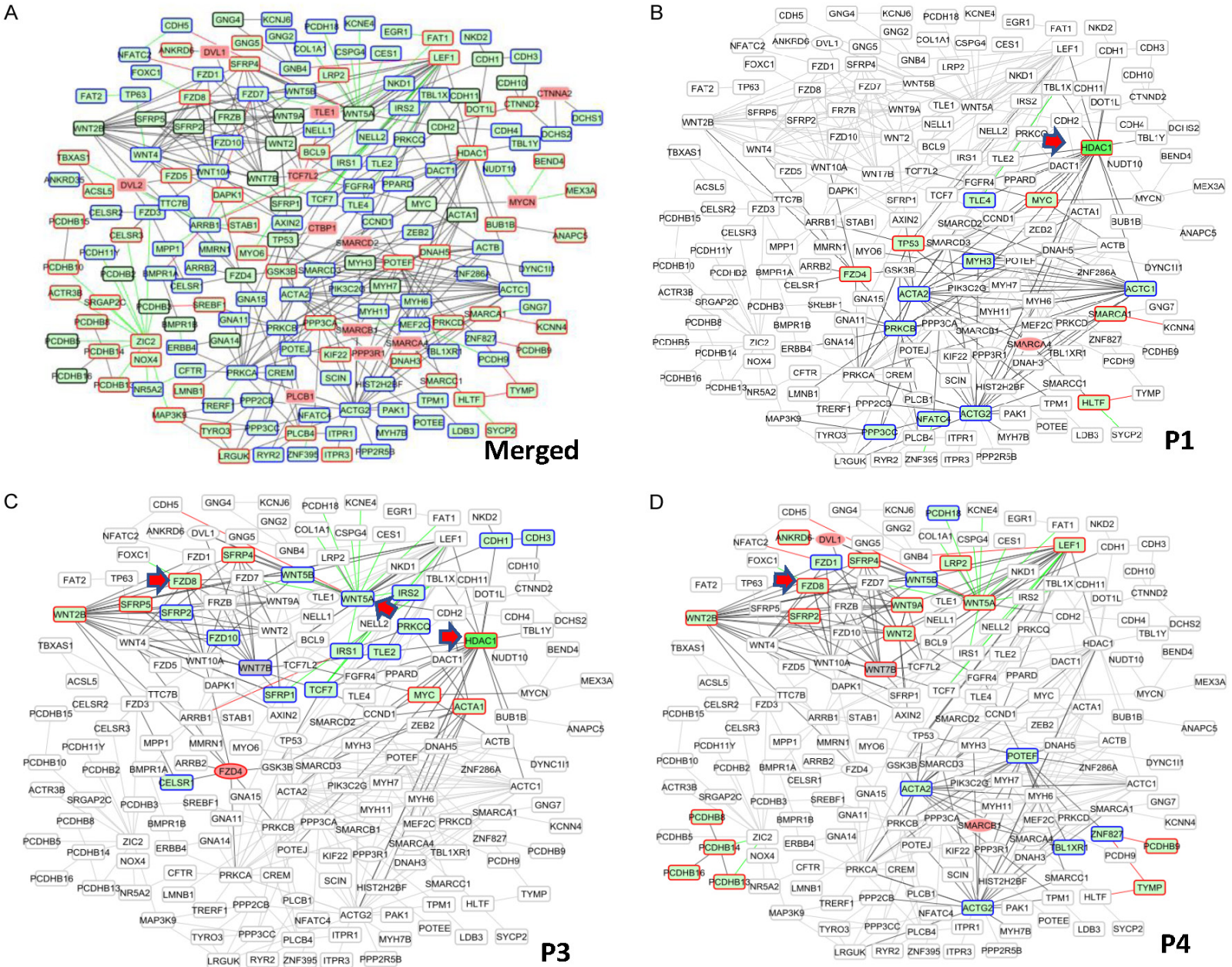
Supplementary Figure 1. Validation of four selected different types of mutations from Exome-seq analysis using PCR coupled to Sanger sequencing. A. Single base deletion in *MGA* gene in patient 5. B. Single base insertion in *CFTR* in patient 12. C. Missense mutation in *MST1R* in patient 14. D. Point mutation in *RASSF8* in patient 13. Red ovals highlight the regions of the mutations. Original (color-coded at the top) and mutated sequences are depicted below each sequence for comparison.

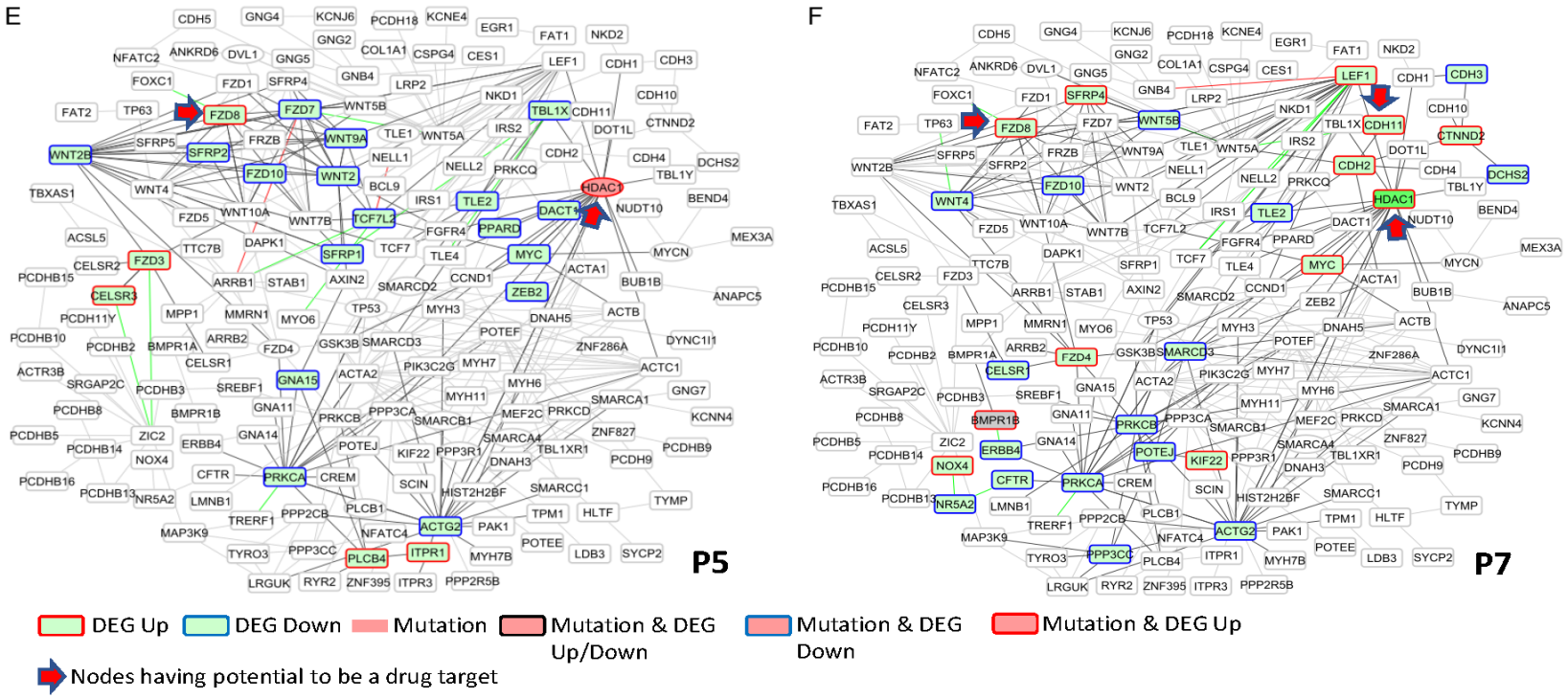


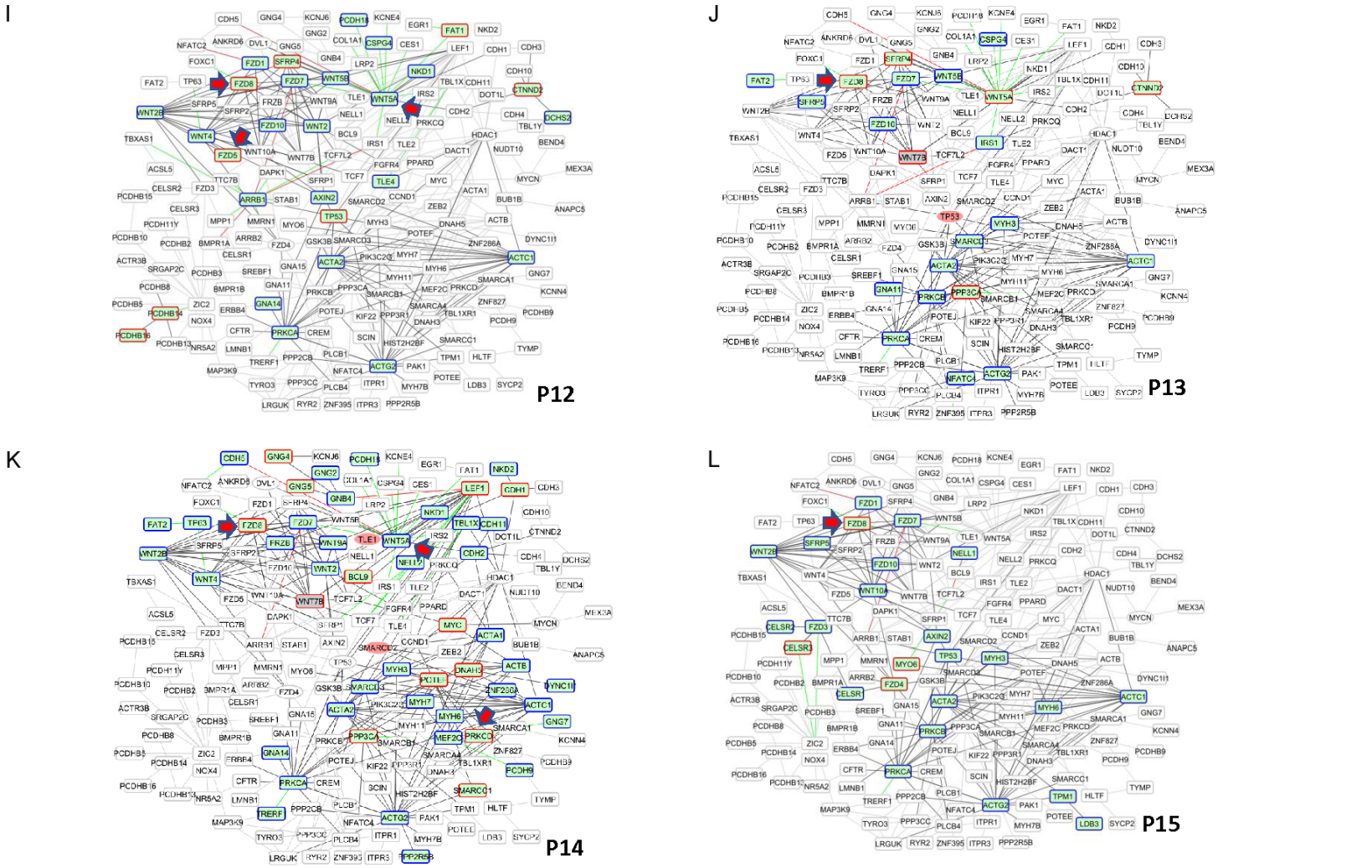
Supplementary Figure 2. Frequency of mutations in the 4 genes of **Figure 1D** in other cancer types (ACC, Adrenocortical Carcinoma; AML, Acute Myeloid Leukemia; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast Invasive Carcinoma; CC, Cervical Cancer; CCRCC, Clear Cell Renal Cell Carcinoma; CESC, Cervical Squamous Cell Carcinoma; CHOL, Cholangiocarcinoma; CRC, Colorectal Cancer; DG, Diffuse Glioma; DLBCL, Diffuse Large B-Cell Lymphoma; EOC, Epithelial Ovarian Cancer; ESCA, Esophagogastric Adenocarcinoma; ESCC, Esophageal Squamous Cell Carcinoma; GBM, Glioblastoma; HCC, Hepatocellular carcinoma; HNSC, Head and Neck Squamous Cell Carcinoma; LGG, Low Grade Glioma; LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma; MSC, Melanoma Skin Cancer; NCCRCC, Non-clear Cell Renal Cell Carcinoma; NSCLC, Non-Small Cell Lung Cancer; PAAD, Pancreatic Adenocarcinoma; PCC, Pheochromocytoma; PRAD, Prostate Adenocarcinoma; SARC, Sarcoma; TET, Thymic Epithelial Tumor; THCA, Thyroid Carcinoma; UCEC, Uterine Corpus Endometrial Carcinoma; Green, mutation; Red, fusion).



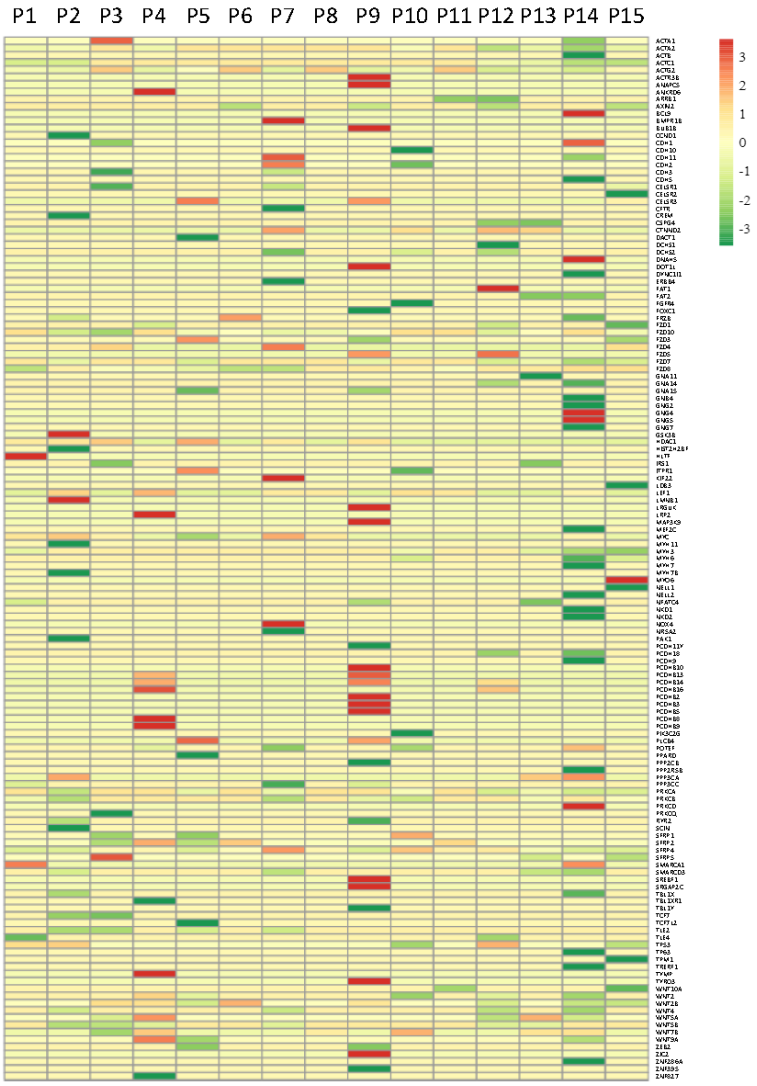
Supplementary Figure 3. Heatmap showing enriched pathways in 51 individual patients using gene expression data from TCGA for prostate cancer. Note that we selected only datasets for which patient-matched normal and prostate cancer tissue was available. The Y-axis specifies the pathways predicted by PANTHER to be enriched in the DEGs of each individual patient; numbers in parentheses indicate the number of total genes known to constitute the pathways. In the X-axis, TCGA codes for patients are given. Each box provides the following information: total number of mutated genes/total number of DEGs affected in each patient for the concerned pathway. *P*-values corresponding to the PANTHER-predicted alteration of a given pathway are shown in color code (scale is shown on the right of the table).


A
Merged
B
P1
C
P3
D
P4

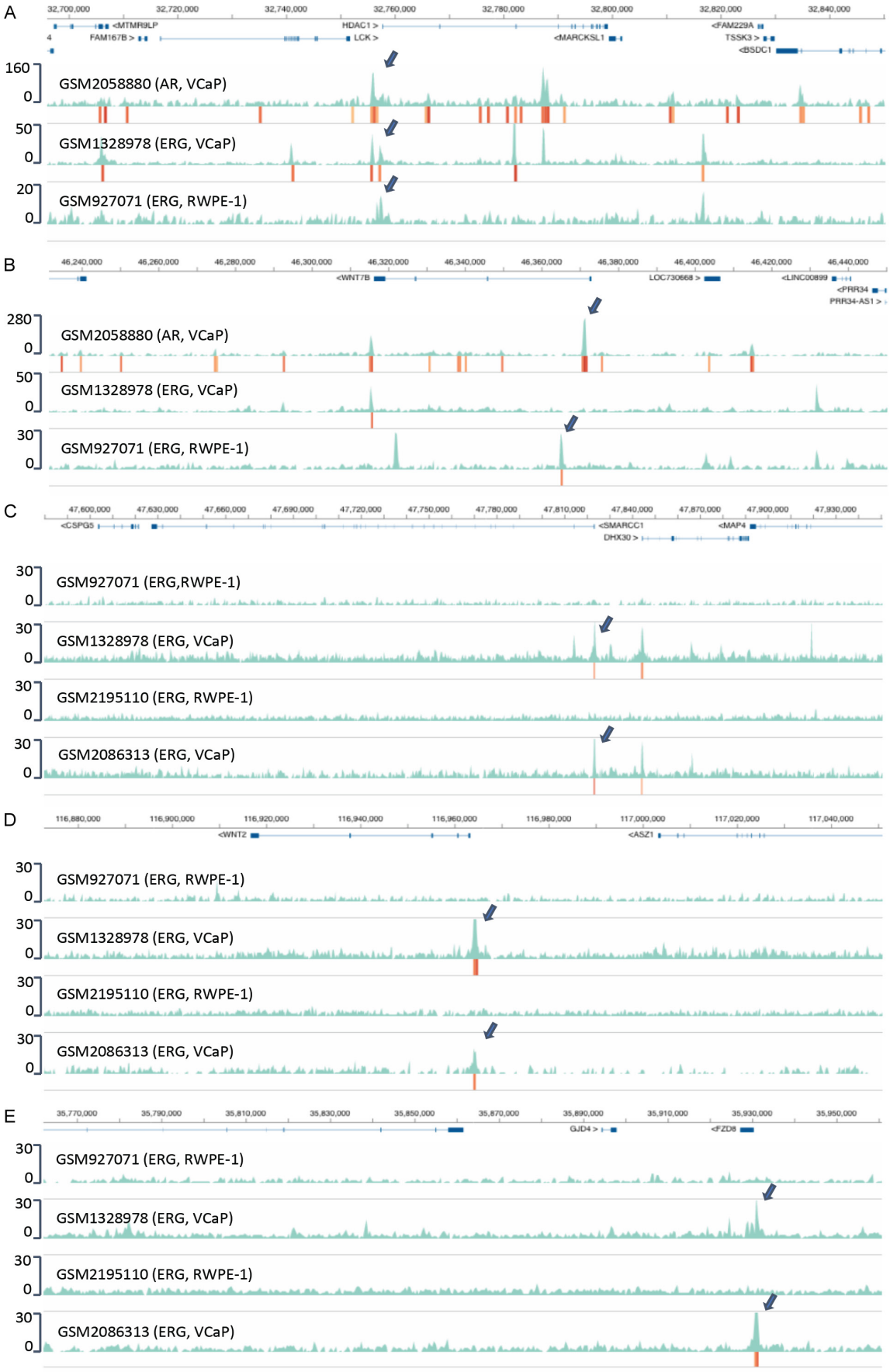

E
F
P5
P7
DEG Up
DEG Down
Mutation
Mutation & DEG Up/Down
Mutation & DEG Down
Mutation & DEG Up
Nodes having potential to be a drug target
G
H
P8
P11



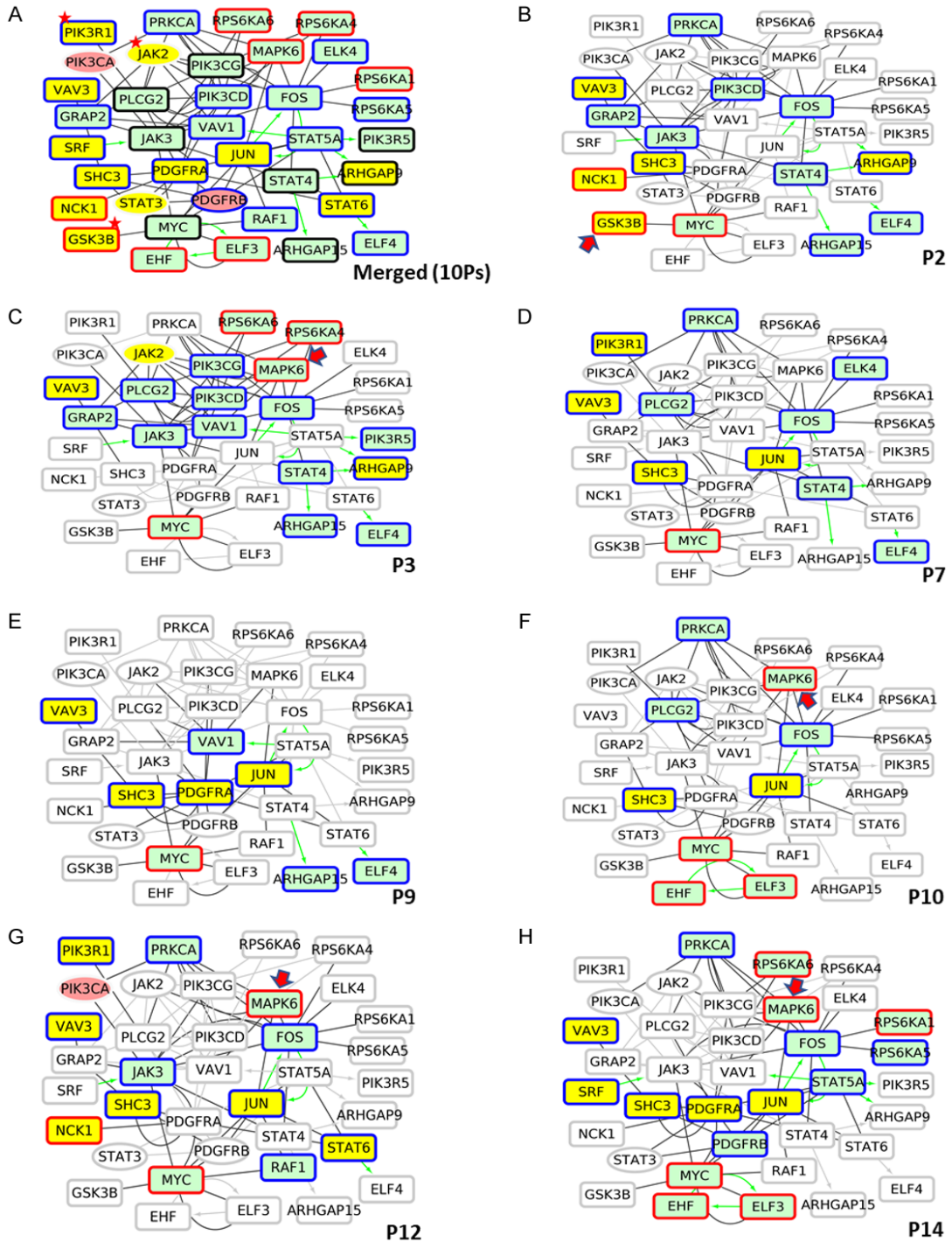
Supplementary Figure 4. Patient-specific deregulations of the Wnt pathway components for the indicated patients (P). (A) Merged network of all 15 patients used a background in (B-L). Frame and color codes are shown below P5. (B-L) Deregulated and/or mutated components of the Wnt pathway in each patient. Note that for the Wnt pathways components both inhibitory and activating drugs are in clinical trails such as Foxy-5 (WNT5A-mimicking peptide that activating Fzd2 and Fzd5) and Ipafricept (Fzd8 antibody blocking). (Clara et al (2020) Nat Rev Clin Onc 17,204).

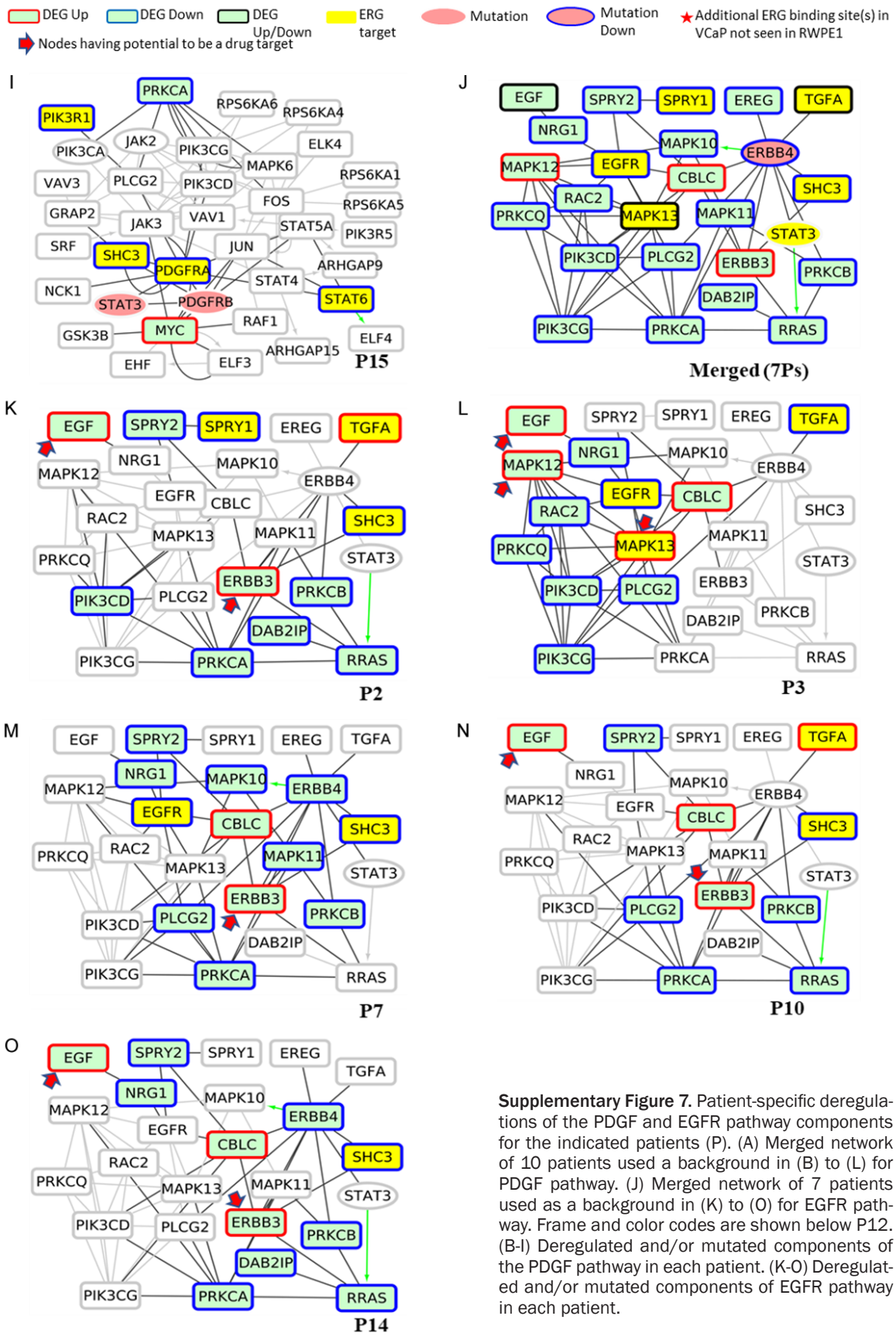


Supplementary Figure 5. Heat map showing differentially expressed genes of the WNT pathway. Color codes represent log2 fold changes. P1 to P15: Patients 1 to 15.

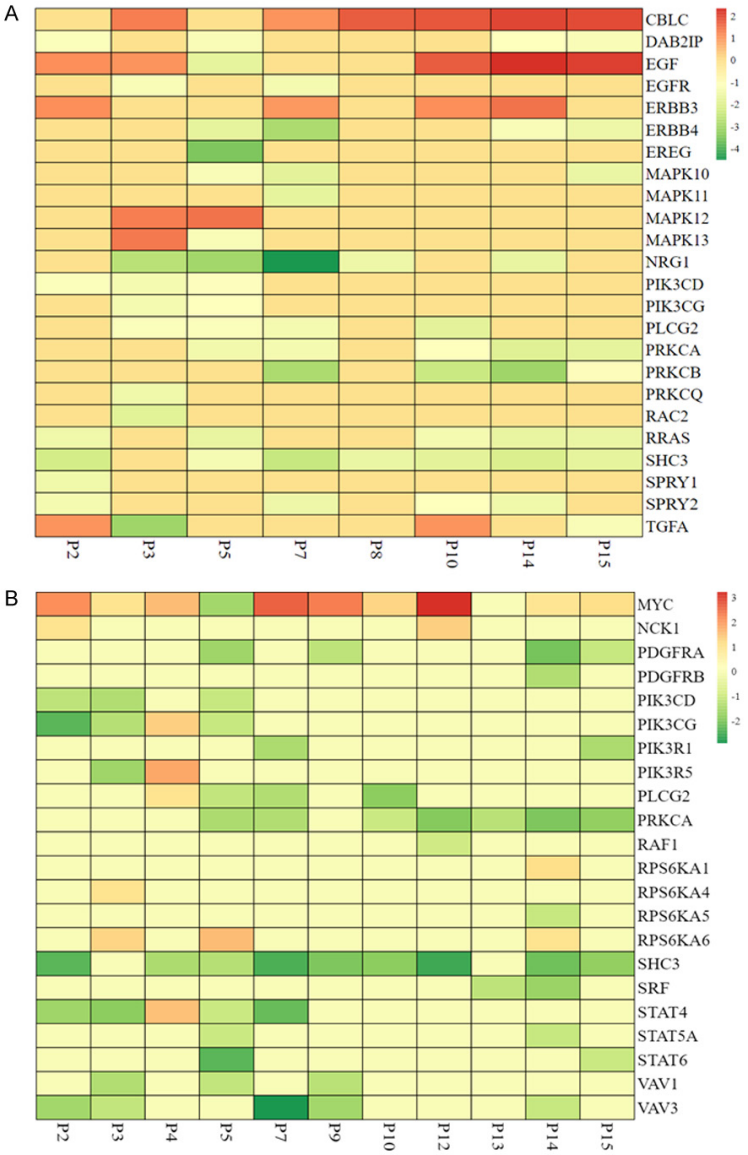

A
32,700,000
32,720,000
32,740,000
32,760,000
32,780,000
32,800,000
32,820,000
32,840,000
<MTMR9LP
HDAC1 >
<FAM229A
FAM167B >
LCK >
<MARCKSL1
TSSK3 >
<BSDC1
160
0
GSM2058880 (AR, VCaP)
50
0
GSM1328978 (ERG, VCaP)
20
0
GSM927071 (ERG, RWPE-1)
B
46,240,000
46,260,000
46,280,000
46,300,000
46,320,000
46,340,000
46,360,000
46,380,000
46,400,000
46,420,000
46,440,000
<WNT7B
LOC730668 >
<LINC00899
<PRR34
PRR34-AS1 >
280
0
GSM2058880 (AR, VCaP)
50
0
GSM1328978 (ERG, VCaP)
30
0
GSM927071 (ERG, RWPE-1)
C
47,600,000
47,630,000
47,660,000
47,690,000
47,720,000
47,750,000
47,780,000
47,810,000
47,840,000
47,870,000
47,900,000
47,930,000
<CSPG5
<SMARCC1
<MAP4
DHX30 >
30
0
GSM927071 (ERG,RWPE-1)
30
0
GSM1328978 (ERG, VCaP)
30
0
GSM2195110 (ERG, RWPE-1)
30
0
GSM2086313 (ERG, VCaP)
D
116,880,000
116,900,000
116,920,000
116,940,000
116,960,000
116,980,000
117,000,000
117,020,000
117,040,000
<WNT2
<ASZ1
30
0
GSM927071 (ERG, RWPE-1)
30
0
GSM1328978 (ERG, VCaP)
30
0
GSM2195110 (ERG, RWPE-1)
30
0
GSM2086313 (ERG, VCaP)
E
35,770,000
35,790,000
35,810,000
35,830,000
35,850,000
35,870,000
35,890,000
35,910,000
35,930,000
35,950,000
GJD4 >
<FZD8
30
0
GSM927071 (ERG, RWPE-1)
30
0
GSM1328978 (ERG, VCaP)
30
0
GSM2195110 (ERG, RWPE-1)
30
0
GSM2086313 (ERG, VCaP)

Supplementary Figure 6. AR binding and acquisition of ERG binding sites in VCaP prostate cancer cells relative to RWPE-1 prostate epithelial cells. Screenshots of qcGenomics browser NAVi displaying genes that show AR and/or ERG binding in their promoter regions. (A), *HDAC1*; (B), *WNT7B; (*C), *SMARCCl*; (D), *WNT2*; (E), *FDZ8*. ChIP-seq data sets in (A and B) are from GEO accession numbers (from top to bottom) GSM2058880 (AR, VCaP), GSM1328978 (ERG, VCaP) and GSM927071 (ERG, RWPE-1), as specified. Note that in (A) ERG binding at the *HDAC1* promoter is seen in VCaP and RWPE-1 cells, while in (B) for *WNT7B* a promoter-proximal ERG binding is seen in 'normal' RWPE-1 but not in VCaP cells; this ERG binding site is distant from the AR binding site. The ERG ChIP-seq data sets in (C-E) are from GEO accession numbers (from top to bottom) GSM927071 for RWPE-1, GSM1328978 for VCaP (both use anti-ERG antibody Epitomics 2805-1), GSM2195110 for RWPE-1 and GSM2086313 for VCaP. GSM2195110 was done by using Anti-ERG Clone 9FY Biocare #CM421 C, GSM2195110 used an anti-ERG antibody but did not provide the source. Note the consistency between corresponding experiments with different antibodies.

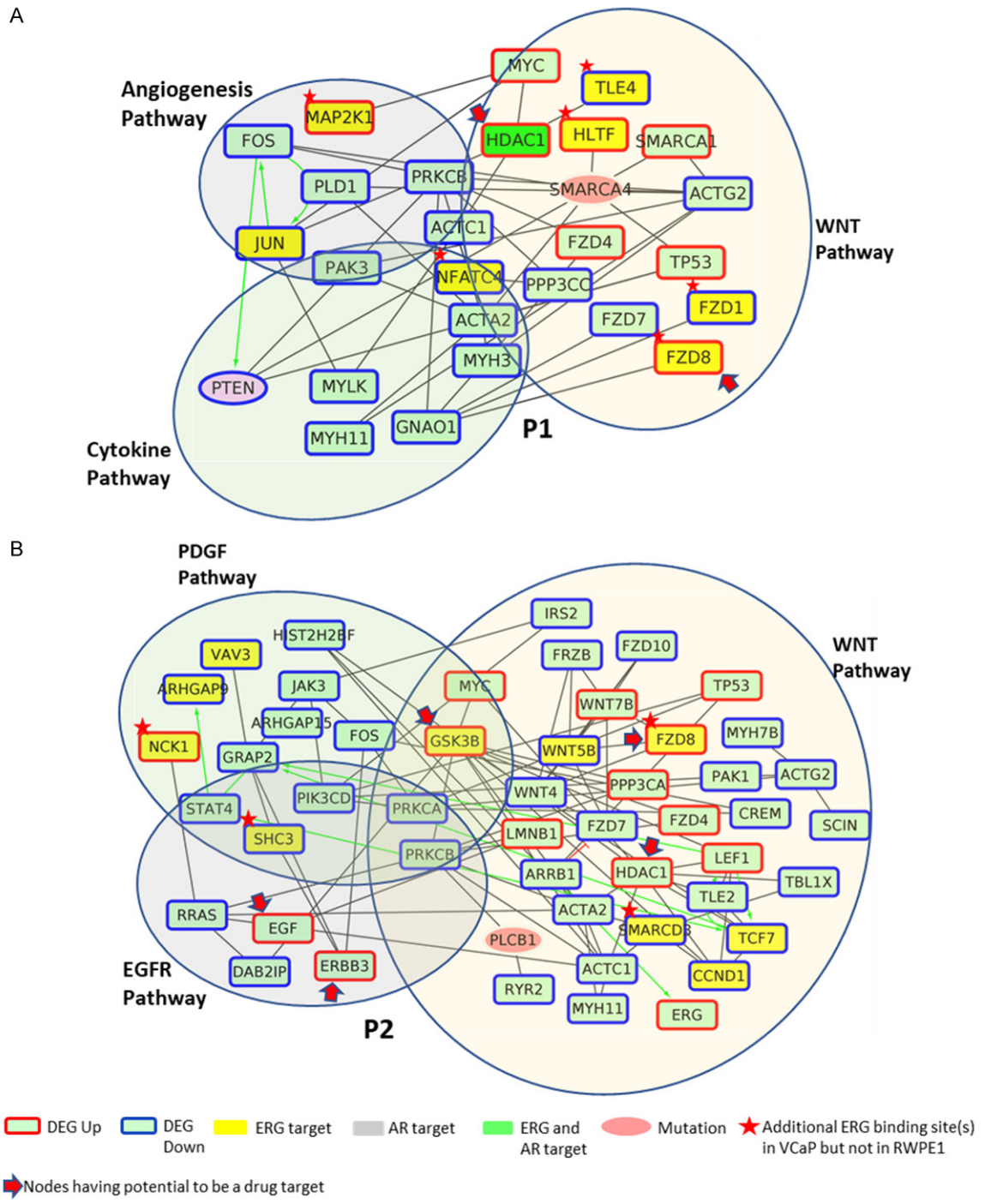




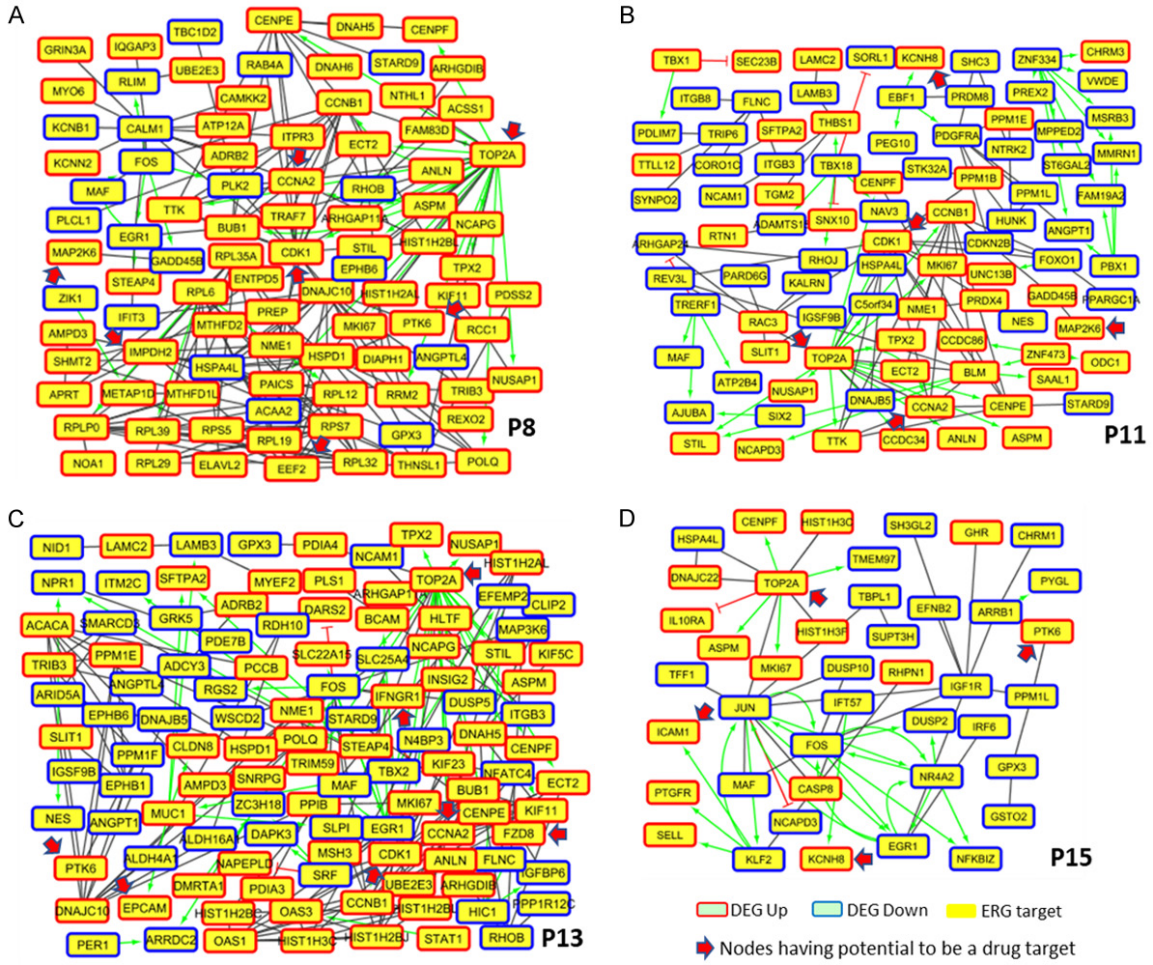
Supplementary Figure 7. Patient-specific deregulations of the PDGF and EGFR pathway components for the indicated patients (P). (A) Merged network of 10 patients used a background in (B) to (L) for PDGF pathway. (J) Merged network of 7 patients used as a background in (K) to (O) for EGFR pathway. Frame and color codes are shown below P12. (B-I) Deregulated and/or mutated components of the PDGF pathway in each patient. (K-O) Deregulated and/or mutated components of EGFR pathway in each patient.



Supplementary Figure 8. Heat maps showing differentially expressed genes of the EGFR (A) and PDGF (B) pathway. Color codes represent log2 fold changes. Px, patient number.



Supplementary Figure 9. Patient-specific deregulations of the PDGF and EGFR pathway components for the indicated patients (P). A. Network display of deregulated and mutated factors of patient P1 in the WNT, Angiogenesis and Cytokine Pathways to reveal connectivities between the different pathways. B. Similar representation of the de-regulated and mutated factors in P2 for the WNT, PDGF and EGFR pathways.



Supplementary Figure 10. Patient-specific networks comprising ERG transcription factor regulated DEGs. Majority of ERG-regulated DEGs are not common among (A) patient 8, (B) patient 11, (C) patient 13 and (D) patient 15 suggesting the need for a patient-centric approach.

Supplementary Figure 11. Patient-specific networks comprising MAZ transcription factor regulated DEGs. Majority of MAZ-regulated DEGs are not common among (A) patient 7 and (B) patient 12 suggesting the need for a patient-centric approach.