Contents lists available at ScienceDirect

EuPA Open Proteomics

journal homepage: www.elsevier.com/locate/euprot

Multi-omics “upstream analysis” of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer

Alexander E. Kel[a,b,c,*], Philip Stegmaier[c], Tagir Valeev[b,d], Jeannette Koschmann[c], Vladimir Poroikov[e], Olga V. Kel-Margoulis[c], Edgar Wingender[c,f]

[a] Institute of Chemical Biology and Fundamental Medicine, SBRAS, Novosibirsk, Russia
[b] Biosoft.ru, Ltd, Novosibirsk, Russia
[c] geneXplain GmbH, D-38302 Wolfenbüttel, Germany
[d] A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia
[e] Institute of Biomedical Chemistry, Moscow, Russia
[f] Institute of Bioinformatics, University Medical Center Göttingen, D-37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 5 September 2016
Accepted 8 September 2016
Available online 9 September 2016

Keywords:
Proteomics data
Microarray data
ChIP-seq
Upstream analysis
Promoter analysis
Pathway analysis

ABSTRACT

We present an “upstream analysis” strategy for causal analysis of multiple “-omics” data. It analyzes promoters using the TRANSFAC database, combines it with an analysis of the upstream signal transduction pathways and identifies master regulators as potential drug targets for a pathological process. We applied this approach to a complex multi-omics data set that contains transcriptomics, proteomics and epigenomics data. We identified the following potential drug targets against induced resistance of cancer cells towards chemotherapy by methotrexate (MTX): TGFalpha, IGFBP7, alpha9-integrin, and the following chemical compounds: zardaverine and divalproex as well as human metabolites such as nicotinamide N-oxide.

© 2016 The Author(s). Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Cancer cells are currently subject of very intense studies of the molecular mechanisms of cancerogenesis. Multiple “-omics” data are generated worldwide measuring expression of proteins, miRNAs and long non-coding RNAs of the cancer cells and, as prerequisite, the epigenomic signatures of DNA methylation and various modifications of chromatin. One of the most important problems is to decipher the mechanisms how cancer cells develop resistance against chemotherapy and search for possible ways to suppress such resistance by interacting with specific molecular targets. One of the important drugs currently widely used in cancer therapy is methotrexate (MTX). Emergence of resistance to MTX of various cancer cells is one of the most important problems in the long-term application of this drug. Several authors compared MTX resistant cells with sensitive cells and generated various sets of “-omics” data [1,2]. We focused our attention on the MTX resistant cells of the colon cancer cell line HT29.

According to the classical view on the mechanism of resistance to the chemotherapy, the resistant clones/lineages are already present in the tumor tissue ab-initio (due to some randomly occurring “favorite” mutations) and get proliferated during the drug treatment while other cells die. However, more recently, a different point of view gets more and more evidences that at least in some cases the cancer cell populations experiencing transitions from a sensitive state to the resistant state during and sometime as a result of the treatment using various chromatin reprogramming mechanisms [3,4]. In this paper we follow this novel point of view and search for such specific reprograming mechanisms in the cancer cells.

Methotrexate (MTX) is a folate antagonist, which kills the proliferating cell by binding tightly to the enzyme dihydrofolate reductase (DHFR). Due to this binding the pathway of de novo DNA synthesis is blocked [1]. But continued administration to patients often results in the emergence of drug-resistance [2]. The analysis of the molecular mechanisms of the resistance can help to identify the most promising targets to combat this resistance. Numerous “-omics” studies on the molecular mechanisms of resistance offer the possibility to mine these high-throughput data by applying

* Corresponding author at: Institute of Chemical Biology and Fundamental Medicine, SBRAS, Novosibirsk, Russia.
E-mail addresses: alexander.kel@genexplain.com, alexander.kel2@googlemail.com (A.E. Kel).

http://dx.doi.org/10.1016/j.euprot.2016.09.002
2212-9685/© 2016 The Author(s). Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

computational tools and analyzing functions and regulation of the involved genes. Such "-omics" data are often deposited in databases such as ArrayExpress [5] or Gene Expression Omnibus (GEO) [6], and derived sets of differentially expressed genes (DEG) (expression signatures) can be found in more specialized databases such as the Expression Atlas [7], the Mouse Expression Database (GXD) [8] and others. These signatures can be used directly for selection of potential drug targets using the mere statistical significance of the expression changes. For a more refined analysis of the molecular mechanisms a conventional approach of mapping the DEG sets to Gene Ontology (GO) categories or to KEGG pathways, for instance by GSEA (gene set enrichment analysis), is usually applied [9,10].

Since such approach provides only a very limited clue to the causes of the observed phenomena, we introduced earlier a novel strategy, the "upstream analysis" approach for causal interpretation of the expression changes [11–13,18]. This strategy comprises two major steps: (1) analysis of promoters and enhancers of identified DEGs to identify transcription factors (TFs) involved in the process under study; (2) reconstruction of signaling pathways that activate these TFs and identification of master-regulators on the top of such pathways. The first step is done with the help of the TRANSFAC database [14] and site identification algorithms, Match [15] and CMA [16]. The second step is done with the help of the TRANSPATH database [17], one of the first signaling pathway databases available, and special graph search algorithms implemented in the geneXplain platform [18].

In this paper, we introduce two enhancements to the upstream analysis approach. First, we add a new graph-weighting schema to the algorithm of master-regulator search that enables to incorporate proteomics data by adding a "context protein" list that pushes the graph search towards those nodes that are expressed in the cell. The second improvement of the approach is an adding the option to analyse TF binding sites in potential enhancer and silencer areas of the genome that are inferred from overlapping transcriptomics and epigenomics ChIP-seq data. These two enhancements of our "upstream analysis" approach at present open the possibility to perform multi-omics studies using the geneXplain platform.

Our study revealed that the novel multi-omics "upstream analysis" approach allows to identify a number of important master regulators of MTX resistance. Among them are some that are known to play essential roles as targets for anti-cancer drug therapy and our results suggest them for the use as anti-resistance targets. These targets were used in the final step of our analysis, i.e. the identification of chemical compounds that have the potential of inhibiting or activating these targets and consequently suppressing the MTX resistance mechanisms.

In silico discovery of chemical compounds that are able to inhibit or activate given molecular targets is one of the most important problems in chemoinformatics. Most often such drug discovery attempts involve the design of molecules that are complementary in shape and charge to the target with which they are supposed to interact. This usually relies on computational molecular modeling techniques. This type of modeling is often referred to as structure-based drug design [19]. In the current work we used an alternative method called ligand-based drug design, or (Q)SAR (Quantitative) Structure-Activity Relationships, which relies on the knowledge of other molecules that bind to the biological target of interest [20]. We are using one of the most powerful instruments in this field, the computer program PASS, which is based on Multilevel Neighborhoods of Atoms (MNA) descriptors to consider the chemical structures of the known ligands of the target of interest and Bayesian approach to estimate the probability that new ligands interact with the same target [21,22]. The PASS program was trained on more the 3500 different molecular targets and can be used now to scan thousands and millions of chemical compounds and find new potential ligands for those targets.

In the current work we applied PASS for the identification of chemical compounds that have the potential to be ligands for the selected targets to combat the MTX resistance mechanisms. Among the promising compounds we found some known drugs, such as zardaverine and divalproex as well as human metabolites such as nicotinamide N-oxide.

As a conclusion, we propose a novel combination of multi-omics bioinformatics analysis with a systems biology approach to the analysis of signaling networks for predicting drug targets and with an advanced chemoinformatics approach for the identification of potentially effective chemical compounds. This approach was successfully applied to the analysis of cancer drug resistance mechanisms.

The workflow of drug target identification is freely accessible online on the geneXplain platform [23].

2. Data and methods

2.1. Microarray data, differential expression analysis

For the analysis of gene expression changes in MTX resistant cells we took publicly available microarray data from Gene Expression Omnibus (NCBI, Bethesda, MD, USA), data entry GSE11440 [24]. The authors analyzed the transcriptome of the colon cancer HT29 cells that were MTX-sensitive and compared them to MTX-resistant cells generated from the same cell line. In total 6 Affymetrix microarray experiments were done, 3 biological replicates for the sensitive cells and 3 replicates for the resistant cells.

Raw microarray data of MTX-resistant and sensitive cells, the latter being used as control in our study, were normalized and background corrected using RMA (Robust Multi-array Average). The Limma (Linear Models for Microarray Data) method was applied to define fold changes of genes and to identify the statistically significantly expressed genes using a Benjamini-Hochberg adjusted *p*-value cutoff ($\leq$0.05) [25].

2.2. Proteomics data

Proteomics data of the HT29 colon cancer cell line were extracted from the PRIDE database (EBI, Hinxton, UK, http://www.ebi.ac.uk/pride), with the project accession number PRD000369 (http://www.ebi.ac.uk/pride/archive/projects/PRD000369). The data were generated and analyzed in the publication [26]. The authors extracted proteins from different regions of multicell tumour spheroids grown from HT29 colon carcinoma cells. They used trypsin digestion iTRAQ 4-plex labeling, 2D separation using OffGel (24 fractions) and RP nanoHPLC, MALDI TOF-TOF MS/MS instruments to determine changes in protein expression across the regions analysed. Authors identified proteins using Mascot software version 2.2 (Matrix Science, U.K.), which compared MS/MS generated data against the Swiss-Prot 2010 human protein database containing 20473 sequences. They set Mascot search parameters for Peptide mass tolerance at 100 ppm (ppm) and MS/MS tolerance at ±0.7 Da. Trypsin proteolysis (cleavage to the C-terminal side of lysine and arginine except when proline is present) was selected allowing for one missed proteolytic cleavage. A 95% confidence threshold ($p < 0.05$) was used for searching the MS/MS data, which corresponded to a Mascot score threshold of $\geq$28. We took the list of proteins (with UniProt accession numbers) from PRIDE (1107 unique accession numbers) and converted them into Ensembl genes (1109 genes). No protein quantitative data were used in our further analysis.

2.3. Epigenomic data on CDK8 co-activator complex in colon cancer

CDK8 is a kinase associating with the mediator complex and is often over-expressed in colorectal cancer [27]. We analyzed data from a study investigating genome-wide localization of CDK8 in human colorectal cancer cell line HT29. The data were extracted from Gene Expression Omnibus (NCBI, Bethesda, MD, USA), data entry GSE53602. In that study Genomic DNA was enriched by chromatin immunoprecipitation (ChIP) and analyzed by Solexa sequencing. ChIP was performed using an antibody against CDK8. We have downloaded the NGS sequences from SRA repository (http://www.ncbi.nlm.nih.gov/sra) and analyzed with the help of the geneXplain platform. Only one biological replica of ChIP-seq data was used here for the further analysis. The ChIP-seq sequence reads were mapped to the human genome build hs19 with the use of the genome mapper Bowtie [28] with default parameters. The peak calling program MACS [29] (without control and with almost all default parameters, except parameter "Enrichment ratio", which was set to value 5 in order to achieve higher number of peaks) was applied then to the obtained alignments, which returned 29,400 peaks of CDK8 complex binding in the whole human genome.

2.4. Analysis of enriched transcription factor binding sites

Transcription factor binding sites in promoters of differentially expressed genes were analyzed using known DNA-binding motifs described in the TRANSFAC® library, release 2014.4 (BIOBASE, Wolfenbüttel, Germany) (http://genexplain.com/transfac). The motifs are specified using position weight matrices (PWMs) that give weights to each nucleotide in each position of the DNA binding motif for a transcription factor or a group of them.

The geneXplain platform provides tools to identify transcription factor binding sites (TFBS) that are enriched in the promoters under study as compared to a background sequence set such as promoters of genes that were not differentially regulated under the condition of the experiment. We denote study and background sets briefly as Yes and No sets. The algorithm for TFBS enrichment analysis, called F-Match, has been described in [11,18] and briefly described in the Supplementary materials (part S1).

In the geneXplain platform, such binding site enrichment analysis is carried out as part of a dedicated workflow. We consider for further analysis only those TFBSs that achieved a Yes/No ratio >1 and a P-value < 0.01. The workflow further maps the matrices to potential transcription factors, and generates visualizations of all results. In the current work we have modified the workflow by considering not only promoter sequences of a standard length of 1100 bp (-1000 to $+100$), but also sequences of potential enhancers and silencers derived from combined transcriptomics and epigenomics data as it is described below. The error rate in this part of the pipeline is controlled by estimating the adjusted *p*-value (using Benjamini-Hochberg procedure) in comparison to TFBS frequency found in randomly selected regions of human genome (adj.p-value < 0.01).

2.5. Finding master regulators in networks

We searched for master regulator molecules in signal transduction pathways upstream of the identified transcription factors using geneXplain platform tools. The master-regulator search uses the TRANSPATH® database (BIOBASE) [17]. A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in TRANSPATH. The main algorithm of master regulator search has been described earlier [11] (see Supplementary material S2.1). The goal of the algorithm is to find nodes in the global signal transduction network that may potentially regulate the activity of the set of transcription factors found at the previous step of analysis. Such nodes are considered as most promising drug targets, since any influence on such a node may switch the transcriptional programs of hundreds of genes that are regulated by the respective TFs. In our analysis we have run the algorithm with the maximum radius of 10 steps upstream of each TF in the input set. Control of the error rate of this algorithm is done by applying it 10000 times to randomly generated sets of input transcription factors of the same size of the sets. Z-score and FDR value of ranks is calculated then for each potential master regulator node on the basis of such random runs (see detailed description in [11]). We control the error rate by the FDR threshold 0.05.

In this paper we are introducing "Context algorithm" that allows incorporation of proteomics data into the analysis of master regulators. A brief description of the "Context algorithm" is done in the Supplementary material (see document S2.3). The algorithm encodes this additional context information as modified edge costs in the signaling network. For instance, the proteomics data gives information about proteins that are expressed in the cell. We call them "context proteins". The idea of the approach is to attract the key node search (e.g. the underlying Dijkstra algortithm for shortest paths) towards context proteins by decreasing the costs of those edges that are close to the context proteins in the network. (see Illustration of the algorithm in Fig. S1).

2.6. Search for chemical compounds targeting master regulators with PASS

The PASS software (www.way2drug.com) aims to predict biological activities of small organic drug-like compounds. The acronym PASS stands for "Prediction of Activity Spectra for Substances". PASS uses 2D structural formulae of organic compounds to simultaneously predict many types of biological activities including such activities as inhibition of a number of important cellular molecular targets. This allows the evaluation of the biological activity profiles for compounds prior to their synthesis and biological testing. The prediction algorithm of PASS is based on Bayesian estimates of probabilities for a compound to belong to the classes of "active" or "inactive", respectively. The mathematical method has been described in several publications, most recently by Filimonov et al. [31]. The predicted activity spectrum is presented in PASS by the list of activities, with probabilities "to be active" ***Pa*** and "to be inactive" ***Pi*** calculated for each activity. In PASS special descriptors, so called Multilevel Neighborhoods of Atoms (MNA), are applied to describe the 2D structural formulae of organic compounds. The molecular structure is represented in PASS by the set of unique MNA descriptors of the 1st and 2nd levels. The details about MNA descriptors are published in [21]. The current release of PASS (2014) is able to predict more than 3800 different biochemical mechanisms of action, such as inhibitors, antagonists or agonists of various protein targets. The PASS program goes together with PharmaExpert – a program for interpretation of PASS results and selecting compounds with the required biological activities on the basis of complex queries.

In this paper, we applied the PASS program to three libraries of chemical compounds in order to find potential ligands for the master regulators found at the previous step. We screened the following three libraries: (1) Top 200 drugs prescribed in the world. Among those 200 drugs, 153 are small organic compounds with known structural formulae; (2) Prestwick chemical library (http://www.prestwickchemical.com/prestwick-chemical-library.html), which is a collection of "1280 small molecules, 100% approved drugs (FDA, EMA and other agencies) selected by medicinal chemists and pharmacists, thus presenting the greatest possible degree of drug-likeness, selected for their high chemical and pharmacological diversity as well as for their known bioavailability and safety in humans". (3) Human metabolites collected in the HMDB, Human Metabolome Database, version 2.5. SDF file with the structural formulae of metabolites is available for download at http://www.hmdb.ca/downloads.

3. Results and discussion

Our strategy of multi-omics "Upstream Analysis" of regulatory genomic regions comprises of two main step (1) a systematic and comprehensive promoter and enhancer analysis on the basis of transcriptomics (differentially regulated genes) and epigenomic data (locations of regions of active chromatin) to identify transcription factors (TFs) involved in regulation of the cellular process under study, and (2) an analysis of the topology of the signal transduction network upstream of transcription factors to identify master regulators, which are signaling proteins in the cell (receptors, their ligands, adapters, kinases, phosphatases, other enzymes involved in signal transduction) that may regulate the activity of transcription factors found in the first step of the analysis. In order to validate this pipeline, previously, we had analyzed a dataset of TNFα-induced genes in human endothelial cells [33] and have demonstrated that our approach detects correctly TNFα as the master regulator and explains activity of other molecules from the TNFα pathway [11,18]. Also, we applied this concept in previous studies and have revealed EGF and IGF2 as regulators during liver tumor development that was experimentally validated [32]. Another experimental validation of this approach was done in our study of varicose vein disease (paper in preparation) where we identified and confirmed experimentally the MFAP5 gene as an important master regulator of the disease process. These and several other currently running studies give us the evidence for the high potential of the approach for the drug target prediction.

3.1. Up- and down-regulated genes in MTX resistant cells

First of all, we identified up- and down-regulated genes from the comparison of transcriptomics data of resistant versus sensitive cells. We analyzed publicly available microarray data [24] and applied Limma (Linear Models for Microarray Data) with a Benjamini-Hochberg adjusted *p*-value cutoff ($\leq$0.05) to retrieve differentially expressed genes (DEG). As result we identified 1951 up-regulated and 2185 down-regulated genes.

The up-regulated genes are enriched by the following GO categories: oxidation-reduction process, lipid metabolic process, purine deoxyribonucleotide metabolic process, dephosphorylation, negative regulation of cell adhesion, cell migration; pathways (TRANSPATH, REACTOME): serotonin degradation, cholesterol metabolism, release of active TGFbeta, metabolism of estrogens, regulation of lipid metabolism by peroxisome proliferator-activated receptor alpha (PPARalpha), extracellular matrix organization.

The down-regulated genes are in turn enriched by the following GO categories: cell cycle, apoptosis, response to virus, protein phosphorylation, organelle fission, response to interferon-alpha, M phase, response to stress; pathways (TRANSPATH, REACTOME): Aurora-B cell cycle regulation, E2F network, cyclosome regulatory network, interferon signaling.

Such GO and pathway analysis gives a general idea of the global processes that changed their activity after establishing the MTX-resistance. They coincided very well with the existing knowledge about the mechanisms of MTX-resistance in cancer cells. According to the results of multiple studies, the most important resistance mechanisms to MTX was found to be connected with an increase of expression of the MTX primary target – enzyme DHFR [1,2]. It is known that this enzyme induction takes place as a result of amplification [34] and enhanced expression [35] of its gene. The increased rate of transcription of this gene is stimulated by enhanced levels of free E2F, not sequestered by hypophosphorylated retinoblastoma protein. The resulting changes in the expression of this important enzyme of nucleotide metabolism is associated, on one side, with the massive changes and re-tuning of the related cellular metabolic pathways that we observed in the respective enrichment of GO terms among the upregulated genes. On the other side, the changes in nucleotide metabolism may lead to changes in the process of cell cycle and apoptosis indicating the slowing down of the processes of cell death. It is interesting to note that the term "protein phosphorylation" was also indicative for the downregulated genes confirming the important role of retinoblastoma hypophosphorylation in developing MTX resistance.

However, the mentioned changes of big functional groups of genes do not provide any key to understand mechanistically how such cellular transformation to the resistant state is achieved and maintained and does not provide molecular targets for possible suppression of the MTX resistance. To answer all these questions we applied our earlier developed concept of "upstream analysis" to the data on MTX resistance.

3.2. Analysis of promoters and enhancers to identify potentially active TFs

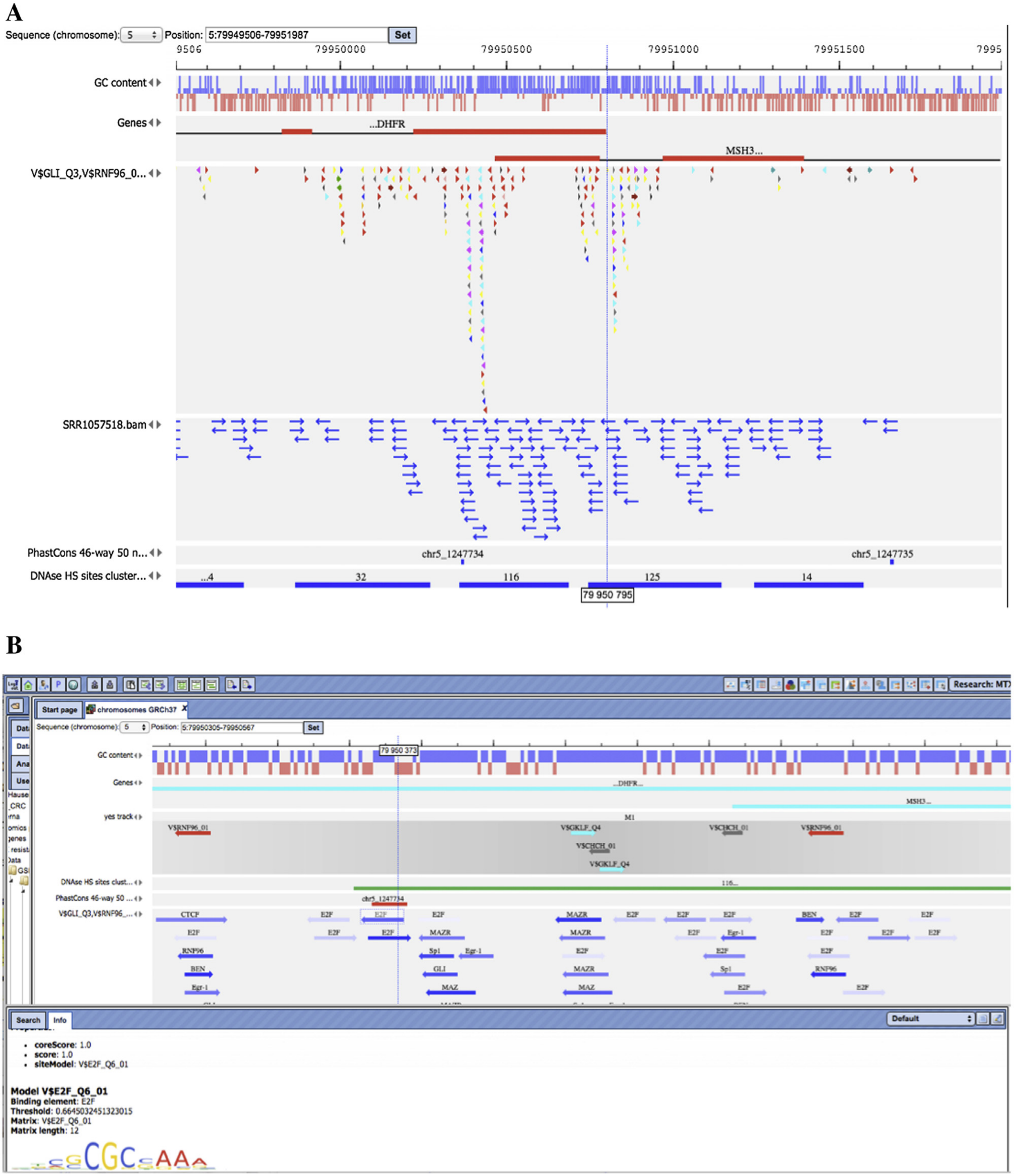
In order to identify transcription factors that may be activated during the transformation of HT29 colon cancer cells into MTX resistant cells we analyzed several important genomic regions of the genes that were differentially regulated during this process. For this, we identified the up- and down-regulated genes using a logFC cut-off (logarithm of the fold change to base 2) higher than 1.5 for up-regulated genes or lower than −1.5 for down-regulated genes ("Yes" sets of genes). As control we used genes expression of which did not change considerably in this experiment ("No" set of genes). From all these genes we extracted the promoter regions from −1000 to +100 bp around TSS (transcription start site). Next, we applied the F-Match algorithm, which searches for TF binding sites in the Yes and No sets of promoter sequences applying the non-redundant set of PWMs from the TRANSFAC library. This program is able to find those PWMs and corresponding transcription factors whose sites are overrepresented in the promoters of Yes set compared to the No set (see Method section). We applied this method separately for the up- and down-regulated genes to identify those specific transcription factors that are involved in activation or inhibition of the expression of these sets of genes. The results of this analysis are presented in Table 1 below. Also, in Fig. 1 we show a map of predicted TF binding sites in the promoter of the *DHFR* gene, the gene encoding the target protein for MTX. Drastic up-regulation of the *DHFR* gene is known as one of the most common mechanisms of the development of MTX resistance [35].

The promoter of this gene has been extensively studied and it was found that expression of the *DHFR* gene is tightly regulated during cell cycle through binding sites for transcription factor E2F [36]. Moreover, it was shown that at least one E2F site is located near an Sp1 site forming a composite element and that E2F and SP1 transcription factors act synergistically in activating *DHFR* transcription [37,38]. It was proposed earlier that the activation of the *DHFR* gene during development of MTX resistance is done through this E2F site [35]. We hypothesize that other transcription factors, such as Sp1 and several other factors, may contribute to the altered activation of *DHFR* and other genes leading to stable up-regulation of such genes, which in turn stabilizes the resistance state of the cells.

Our site frequency analysis indeed revealed sites for E2F and Sp1 factors as overrepresented in the promoters of up-regulated

Table 1
The list of transcription factors identified by site frequency search in promoters and potential enhancers of up-regulated and down-regulated genes. Gene symbol and gene description are given for the genes encoding the respective transcription factors. Expression logFC is the fold change of the expression of these transcription factor genes in the MTX resistant cells. Up-regulated TF genes are marked in red, down-regulated TF genes are marked in blue. PWM is the identifier of the TRANSFAC position weight matrix whose sites are overrepresented in the promoters or enhancers of the genes under study. Yes/No ratio and P-value are the values obtained by the site frequency search in the promoters and enhancers, respectively.

Gene symbol	PWM	Expression logFC	Yes/No ratio Up-regulated promoters	P-value Up-regulated promoters	Yes/No ratio Up-regulated enhancers	P-value Up-regulated enhancers	Yes/No ratio Down-regulated promoters	P-value Down-regulated promoters	Yes/No ratio Down-regulated enhancers	P-value Down-regulated enhancers	Gene description
GLI2	V$GLI_Q3	1.064	1.281	1.30E-16	1.266	2.49E-05	1.095	1.72E-03	2.334	1.51E-05	GLI family zinc finger 2
MAFB	V$MAF_Q4	1.061			1.628	4.46E-05			1.509	4.97E-05	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
CEBPA	V$CEBPA_Q6	0.800			1.365	1.22E-03					CCAAT/enhancer binding protein (C/EBP), alpha
TCF7L2	V$ETS_Q6	0.768			1.424	5.91E-11			1.352	4.18E-08	transcription factor 7-like 2 (T-cell specific, HMG-box)
DBP	V$DBP_Q6	0.540			2.297	4.68E-04					D site of albumin promoter (albumin D-box) binding protein
ATF2	V$CREBP1_01	0.480			1.581	9.32E-04			1.822	1.40E-05	activating transcription factor 2
TRIM28	V$RNF96_01	0.413	1.825	1.21E-18					1.417	8.27E-06	tripartite motif containing 28
FOXA2	V$HNF3B_Q6	0.351			1.697	6.17E-10			1.687	1.49E-11	forkhead box A2
HIF1A	V$HIF1A_Q5	0.326	1.502	5.18E-03							hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
CRX	V$CRX_Q4_01	0.290			1.393	4.06E-03					cone-rod homeobox
HMGA1	V$HMGIY_Q3	0.259									high mobility group AT-hook 1
RFX1	V$RFX1_01	0.252	1.573	2.04E-03					1.325	8.71E-06	regulatory factor X, 1 (influences HLA class II expression)
EGR1	V$EGR1_Q6	0.092	1.604	1.03E-07	1.514	8.52E-05			1.790	1.92E-05	early growth response 1
TFAP2A	V$AP2ALPHA_03	0.092	1.327	3.02E-40	1.153	2.36E-03	1.827	2.61E-04	1.160	1.04E-09	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
SOX2	V$SOX2_Q3_01	0.081			1.530	6.04E-03			1.432	6.63E-03	SRY (sex determining region Y)-box 2
ZNF384	V$CIZ_01	0.060									zinc finger protein 384
CTCF	V$CTCF_01	-0.103	1.298	1.50E-04	19.183	3.53E-16			15.803	2.33E-13	CCCTC-binding factor (zinc finger protein)
HNF1A	V$HNF1A_Q4	-0.146			16.442	3.24E-05			14.502	1.02E-04	HNF1 homeobox A
ZNF263	V$FPM315_01	-0.167	1.523	2.13E-07	1.623	1.35E-03	1.297	3.01E-04	1.689	3.51E-03	zinc finger protein 263
LEF1	V$LEF1_Q5_01	-0.178			1.327	6.13E-04	1.196	1.58E-03	1.330	1.27E-03	lymphoid enhancer-binding factor 1
GTF2IRD1	V$BEN_01	-0.245	1.370	8.53E-25	1.180	2.17E-05			1.334	1.05E-15	GTF2I repeat domain containing 1
TBP	V$TATA_01	-0.309	1.544	9.99E-11							TATA box binding protein
SRY	V$SRY_Q6	NA			1.284	4.65E-04			1.304	1.57E-04	sex determining region Y
MAZ	V$MAZ_Q6_01	-0.328	1.347	3.02E-08	1.709	7.02E-12	1.228	7.21E-03	1.513	3.09E-06	MYC-associated zinc finger protein (purine-binding transcription factor)
KLF6	V$CPBP_Q6	-0.329	1.277	2.02E-13	1.236	4.75E-07			1.238	2.10E-07	Kruppel-like factor 6
CREB1	V$CREB1_Q6	-0.335	1.354	8.03E-03	10.118	2.83E-03			2.882	7.65E-03	cAMP responsive element binding protein 1
E2F1	V$E2F_Q6_01	-0.346	1.302	5.04E-06	1.429	3.29E-05			1.601	1.56E-08	E2F transcription factor 1
NFIA	V$NF1_Q6	-0.425			2.396	2.13E-04	1.873	1.37E-03	1.206	9.60E-04	nuclear factor I/A
E2F7	V$E2F_Q6_01	-0.441	1.302	5.04E-06	1.429	3.29E-05			1.601	1.56E-08	E2F transcription factor 7
SP1	V$SP1_Q6_01	-0.446	1.330	7.13E-10	1.778	4.29E-18			1.767	1.14E-18	Sp1 transcription factor
E2F3	V$E2F_Q6_01	-0.483	1.302	5.04E-06	1.429	3.29E-05			1.601	1.56E-08	E2F transcription factor 3
SRF	V$SRF_Q5_02	-0.502	7.316	5.17E-03			1.859	3.78E-03			serum response factor (c-fos serum response element-binding transcription factor)
ETV6	V$ETS_Q6	-0.526			1.424	5.91E-11			1.352	4.18E-08	ets variant 6
ELK1	V$ETS_Q6	-0.563			1.424	5.91E-11			1.352	4.18E-08	ELK1, member of ETS oncogene family
TCF7	V$ETS_Q6	-0.569			1.424	5.91E-11			1.352	4.18E-08	transcription factor 7 (T-cell specific, HMG-box)
TFDP1	V$E2F_Q6_01	-0.657	1.302	5.04E-06	1.429	3.29E-05			1.601	1.56E-08	transcription factor Dp-1
KLF4	V$GKLF_Q4	-0.669	1.278	3.75E-28	1.263	1.20E-09	1.115	4.71E-05	1.219	1.43E-07	Kruppel-like factor 4 (gut)
TEAD1	V$TEF1_Q6_04	-0.759			3.433	6.64E-04	2.667	8.33E-03	1.379	2.22E-03	TEA domain family member 1 (SV40 transcriptional enhancer factor)
TP53	V$P53_04	-0.805	1.170	1.24E-07			1.279	7.02E-03	1.214	3.54E-03	tumor protein p53
PATZ1	V$MAZR_01	-0.816			1.860	7.70E-04			1.656	6.25E-03	POZ (BTB) and AT hook containing zinc finger 1
ELF4	V$ETS_Q6	-0.873			1.424	5.91E-11			1.352	4.18E-08	E74-like factor 4 (ets domain transcription factor)
GABPB1	V$ETS_Q6	-0.890			1.424	5.91E-11			1.352	4.18E-08	GA binding protein transcription factor, beta subunit 1
STAT1	V$STAT1_Q6	-1.015			1.918	8.48E-04			2.116	9.70E-05	signal transducer and activator of transcription 1, 91kDa
FOXC1	V$FREAC3_01	-1.102			1.215	4.94E-04					forkhead box C1
GATA2	V$GATA_Q6	-1.137			1.239	7.59E-03					GATA binding protein 2
MECP2	V$MECP2_02	-1.317	1.683	9.41E-05	2.616	8.66E-06			2.687	2.82E-06	methyl CpG binding protein 2 (Rett syndrome)
HNF4A	V$HNF4A_Q3	-1.497			8.221	1.87E-04	4.746	4.34E-03	10.876	3.57E-06	hepatocyte nuclear factor 4, alpha
GATA3	V$GATA_Q6	-1.799			1.239	7.59E-03					GATA binding protein 3



Sequence (chromosome): 5 Position: 5:79949506-79951987 Set

Sequence (chromosome): 5 Position: 5:79950305-79950567 Set

- **coreScore**: 1.0
- **score**: 1.0
- **siteModel**: V$E2F_Q6_01

Model V$E2F_Q6_01
Binding element: E2F
Threshold: 0.6645032451323015
Matrix: V$E2F_Q6_01
Matrix length: 12

Fig. 1. Results of TF binding sites prediction in the overlapping promoters of *DHFR* and *MSH3*. A) Low resolution map of gene structures. Exons are represented by red thick lines, introns by thin black lines. (One can see that the first introns of *DHFR* and *MSH3* genes actually overlap). The dotted vertical line indicates the TSS (transcription start site) for the DHFR gene. Colored triangles show positions of TF binding sites (each color corresponds to one PWM). Clusters of sites can be recognized as peaks of overlapping triangles. The track with blue arrows corresponds to the ChIP-seq reads from CDK8 experiment mapped to this genome region. The peak of the reads indicates the region of high regulatory transcription activity. Similar indicators of the open chromatin are the locations of the DNAse hypersensitivity (from ENCODE) shown in the bottom-most track. Two conserved regions (for 46-way 50% conservation between mammalian genomes) indicate potentially very important regulatory areas in these promoters. B) High resolution map. Each predicted TF binding site is shown as an arrow with the name of PWM (from TRANSFAC) on top of it. The intensity of the blue color corresponds to the score of the binding site. The direction of the arrow shows at which DNA strand the site was recognized by the respective PWM. Known sites for E2F and Sp1 are surrounded by two ovals. The track "yes track" shows composite sites predicted by CMA (see next paragraphs). One can see that predicted TF sites often overlap with each other indicating very complex potential regulatory switches.

genes together with sites for several other TFs. In total we found 29 enriched PWMs in the promoters of upregulated genes and 23 enriched PWMs in the promoters of down-regulated genes. Among them, 22 and 11 PWMs correspond to the transcription factor genes whose expression was significantly up-regulated (see Table 1). Among the TFs whose sites found to be most enriched there are: SRF, POU6F1, RNF96, EGR1, MAZ, E2F1, SP1, KLF. Our analysis correctly identified the known E2F and Sp1 sites in the promoter of the *DHFR* gene and even found a number of clusters of several E2F and Sp1 sites together with sites for the other important transcription factors. These site clusters co-localize with ChIP-seq peaks of the CDK8 mediator complex as well as with regions of DNase I hypersensitive sites (Fig. 1). Also, we found that the region of high homology between 46 mammalian genomes (PhastCons 46-way 50) is also located in the area near the detected site clusters (Fig. 1), which gives additional evidence about the functional importance of this regulatory area of the genome. Interestingly, this regulatory region of the *DHFR* gene also controls the expression of another gene, *MSH3*, which is transcribed in the opposite direction and which is very important for the pathology of colon cancer and also known to be involved in drug resistance mechanisms, since it is involved in DNA repair [39]. As one may see from the gene expression data of the MTX resistant colon cancer cells, both genes *DHFR* and *MSH3* showed significant up-regulation of about 4-fold compared to the MTX sensitive cells.

It is known that regulation of gene expression is controlled not only through promoter sequences but also through enhancers and silencers that can be localized in distal upstream regions as well in introns and in 3′ regions of genes. In order to identify most probable enhancers and silencers acting under the analyzed conditions we chose the ChIP-seq data on the CDK8, which is associated with the mediator complex, a central integrator of transcription proven as a marker of active transcription regulatory regions in colorectal cancer cells (for the HT29 cell line) [40]. The central role of the CDK8 kinase complex in the Wnt pathway, which is very often disregulated in colorectal cancers and contributes to their growth, invasion and survival [41], renders it a suitable marker for active enhancers in colon cancer cells. Identification of the peaks of CDK8 mediator complex binding in the genome of cancer cells was done with the help of the MACS algorithm that analyses the NGS reads from the ChIP-seq experiment and finds the regions most massively covered by the sequence reads, indicating the areas of most active CDK8 binding and pointing to the positions of active enhancers in these cells. The MACS algorithm found 29,400 peaks (see method section) in the whole genome. These peaks overlap with 17,115 genes in the genome and located either in their exons, or introns or in 5′ or 3′ regulatory regions of the genes (2 kb upstream and 2 kb downstream from the gene borders). The length of the detected peaks varies quite a lot between 200 bp and 27,000 bp. For further analysis we have identified summits in each peak (the point in the peak that has the highest number of overlapping sequencing reads, which approximately corresponds to the most intense binding of CDK8 complex and respectively the most intense regulatory activity of the region).

Next, we selected only those CDK8 peaks, whose summits could be found in or near (+/− 2000 bp) the up- or down-regulated genes. This way we predicted the approximate location of the HT29 cell line enhancers and silencers that potentially act to change the regulation of these genes upon development of MTX resistance. We analyzed the regions around the summits of the peaks (+/−200 bp around each summit) for the frequencies of TF sites (predicted by TRANSFAC PWMs), and compared them with the background frequency of the sites in randomly selected genomic regions. The same F-Match algorithm was used here as for the analysis of promoter sequences. Results of the analysis of enhancers and silencers for respective up- and down-regulated genes are summarized in Table 1 below.

As it was mentioned in the introduction, it is important to understand the interactions between transcription factors during their regulation of specific gene activity. We have therefore also applied the CMA algorithm (Composite Module Analyst) for searching composite modules [16] in the promoters of up – and down-regulated genes. The core of CMA is a genetic algorithm that identifies pairs of TF sites that are co-localized on a certain distance to each other in the analyzed promoters and enhancers. We identified a composite module consisting of 6 pairs of TFs (represented by TF PWMs from TRANSFAC) (Table 2) that statistically significantly separates sequences in the Yes and No sets (Wilcoxon *p*-value = 5.41E-24). In Fig. S2 in the Supplementary material we present a screenshot from geneXplain platform with detailed information about the pairs of TF sites that were found in the promoters of up-regulated genes and also the statistical parameters of the constructed composite module.

Among the TFs whose sites are found in such pairs are: factors of the TCF/LEF family which are involved in the Wnt signaling pathway (often deregulated in colorectal cancers); TRIM28/RNF96 co-repressor that is known to be involved in the inhibition of E2F1 activity by stimulating E2F1-HDAC1 complex formation (http://www.uniprot.org/uniprot/Q13263); Egr1, a known immediate-early response TF, activated by extracellular signals and mediating mitogenic responses [42]; GKLF (KLF4), a transcription factor that regulates proliferation, differentiation, apoptosis and somatic cell reprogramming. Evidence also suggests that KLF4 is a tumor suppressor in certain cancers, including colorectal cancer [43] and several other important transcription factors with known function of regulation processes of cell cycle, differentiation and apoptosis. All these transcription factors were also included into Table 1 for further analysis.

3.3. Find master regulators in networks

The next step of the analysis was the search for potential master regulators that can regulate the activity of the transcription factors identified in the previous step. The master regulator search was done from the list of transcription factors in Table 1 (see above). As

Table 2
Pairs of TFs found by Composite Module Analyst (CMA) in promoters of differentially expressed genes. First and second PWMs are the Position Weight Matrices (PWMs) selected by CMA to be included into the pair. First and second cut-offs are respectively score cut-off for those two PWMs that were optimized by CMA. Distance – is the most frequent distance between sites in the respective pair.

Pair N	First PWM	First cut-off	Second PWM	Secons cut-off	Distance
1	V$GKLF_Q4	0.96	V$ZIC1_05	0.82	55
2	V$RNF96_01	0.9	V$ZFP161_04	0.74	49
3	V$RFX_Q6	0.95	V$LEF1_Q5_01	0.96	51
4	V$CHCH_01	0.99	V$CIZ_01	1	51
5	V$CDPCR1_01	0.91	V$GKLF_Q4	0.96	56
6	V$HMGIY_Q3	0.88	V$NF1A_Q6_01	0.99	50

a set of context proteins we used the list of proteins that were detected by an independent proteomics experiment on the same colorectal cell line HT29. As described in the Methods section this set of expressed proteins contains 1107 unique UniProt accession numbers. We mapped this protein list onto the TRANSPATH database and detected 2092 protein entities (corresponding to various protein isoforms of the initial list of the 1107 UniProt proteins) participating in various signal transduction and metabolic reactions according to the knowledge stored in this database.

The rational of using the proteomics data as the "context protein" list is in the possibility to direct the algorithm of pathway reconstruction and master-regulator search towards those paths through the signal transduction network that go maximally through those proteins that were detected experimentally to be expressed in this type of cells. The algorithm does not exclude completely the other paths through proteins that were not experimentally detected, just because their concentration might be below the detection limit. Therefore they may well be active in the cells and may participate in the transduction of the relevant signals. Nevertheless, the proteins that were detected in the proteomics experiment are considered with higher weights in the algorithm and contribute more in directing the search towards master regulators.

In the current work we set the maximal distance of the search for master regulator equal to 10 steps, which gives a good chance to find regulators that are quite distant in the network and can be at the level of transmembrane receptors or neighboring adaptor proteins, or extracellular molecules, which makes them more accessible for the interactions with the potential drugs.

The next important parameter of the search was the requirement that the master regulator proteins should have an elevated expression in the MTX resistant cells. We checked the fold change of the genes expressing the proteins that were found by the algorithm as potential master-regulators. We require that these genes were statistically significantly up-regulated in the MTX-resistant cells compared to the sensitive cells. In total we identified 220 genes with LogFC >0.5 that encode potential master regulators with a master regulator score >0.3.

We hypothesized that MTX resistance might imply the presence of a **positive feedback loop.** Such loops may constitute when the genes expressing master-regulator proteins stimulate their own expression under the tested conditions and through the signaling cascade including TF activation events at the bottom end. We believe that such positive feedback loops can contribute to the transition of the MTX sensitive to the MTX resistant state of cells. Therefore, we introduce into the algorithm an important requirement that the genes encoding selected master regulators should be up-regulated, that reflects presence of such positive feedback loop in the system. Important remark here is that we assume that change of expression of the genes that encode master-regulator proteins will influence production of these proteins in the cells and finally their activity in the network. Generally, as it was shown before, the correlation between transcriptomics and proteomics data is not always satisfactory [44], especially considering fast processes when level of transcription of many genes is quickly changing whereas the production of the respective proteins is not changing due to various reasons. Obviously quantitative proteomics data measuring the difference of protein level between MTX sensitive and resistant cell lines would be a better source for such identification of potential feedback loops. Since such data are not available (available proteomics data presents proteins in the standard HT29 cell line only, but not in the MTX resistant cells) we use the transcriptome fold changes as the proxy for the possible difference in the protein levels of the master regulator nodes and we also use the available proteomics data as the source for the "context proteins" (see Method section) that are found as multiple nodes in the revealed signal transduction network transferring signal from the master regulators to the transcription factors.

In Fig. 2 below we show the network of the top 10 potential master regulators that were found by the algorithm and which are present in the target list of the PASS (see below). Genes encoding these 10 proteins were also significantly up-regulated in the MTX-resistant cells and therefore can be considered as important drug targets for possible re-sensitization of such cells towards action of MTX. We also show in the figure that several proteins that were experimentally detected in the HT29 by high-throughput proteomics techniques contributed to the detection of these master regulators. On the schema those "context proteins" are shown by gray half-circles decorating these proteins. One can see that these context proteins often connect the identified master regulators with several transcription factors, therefore playing an important role in transducing the signal from the master regulators to these transcription factors, which in turn regulate their target genes upon such signal. The yellow half-circles on the other side show which proteins are encoded by genes that change their expression most significantly in the MTX-resistant cells compared to the sensitive cells. One can see that most of the master-regulators on this schema are up-regulated.

Altogether, we noticed that many of the suggested master regulators are very important proteins that are known to be involved in regulating such process as cell cycle, apoptosis, cell adhesion and metabolism of nucleotides. All those processes that were detected as changed in MTX-resistant cells in our GO analysis above. Also, there are many lines of evidences showing the potential role of some of these proteins in sensitization of anti-cancer drug resistance mechanisms. For instance, it is known that such master regulator as PDE4 (part of the extended network (see full table of master regulator in our paper in Data in Brief [56]), not shown in Fig. 2) is widely expressed in brain tumors and promotes their growth and treatment with the PDE4A inhibitor Rolipram overcomes tumor resistance and mediates tumor regression [45]. TGF-alpha, which is also found in our master-regulator search and which is one of the most highly up-regulated proteins in MTX-resistant cells, has been found potentially responsible for acquired resistance to Trastuzumab in metastatic breast cancer patients [46]. It was also shown that integrin alpha9 (ITGA9), which facilitates accelerated cell migration and regulates cancer cell proliferation and migration, is a target of epigenetic regulation and its overexpression leads to acquired resistance against 5-aza-dC treatment in human breast tumors [47]. Recently, it was shown that inhibition of insulin-like growth factor 1 receptor (IGF1R) leads to sensitization of head and neck cancer cells to cetuximab and methotrexate [48]. Therefore it is extremely interesting that we identified IGFBP7 protein as a potential master regulator, since this protein is a very potent modulator of IGF binding to its receptors. All these facts show that the list of targets selected by the master regulator search algorithm has a very high potential to serve for re-sensitization of colorectal cancer against MTX resistance.

3.4. Prediction of compounds potentially reverting the MTX resistance of cancer cells

To find potential drugs or new chemical compounds that can be used for reverting the MTX resistance we applied the PASS program to three libraries of chemical compounds. We searched for compounds that may serve as inhibitors of master-regulators found in the previous step of the analysis. We analyzed the following libraries: (1) Top 200 drugs prescribed in the world. Among those 200 drugs, 153 are small organic compounds with known structural formulae; (2) Prestwick chemical library, which is a collection of 1280 small drug-like molecules; (3) Human

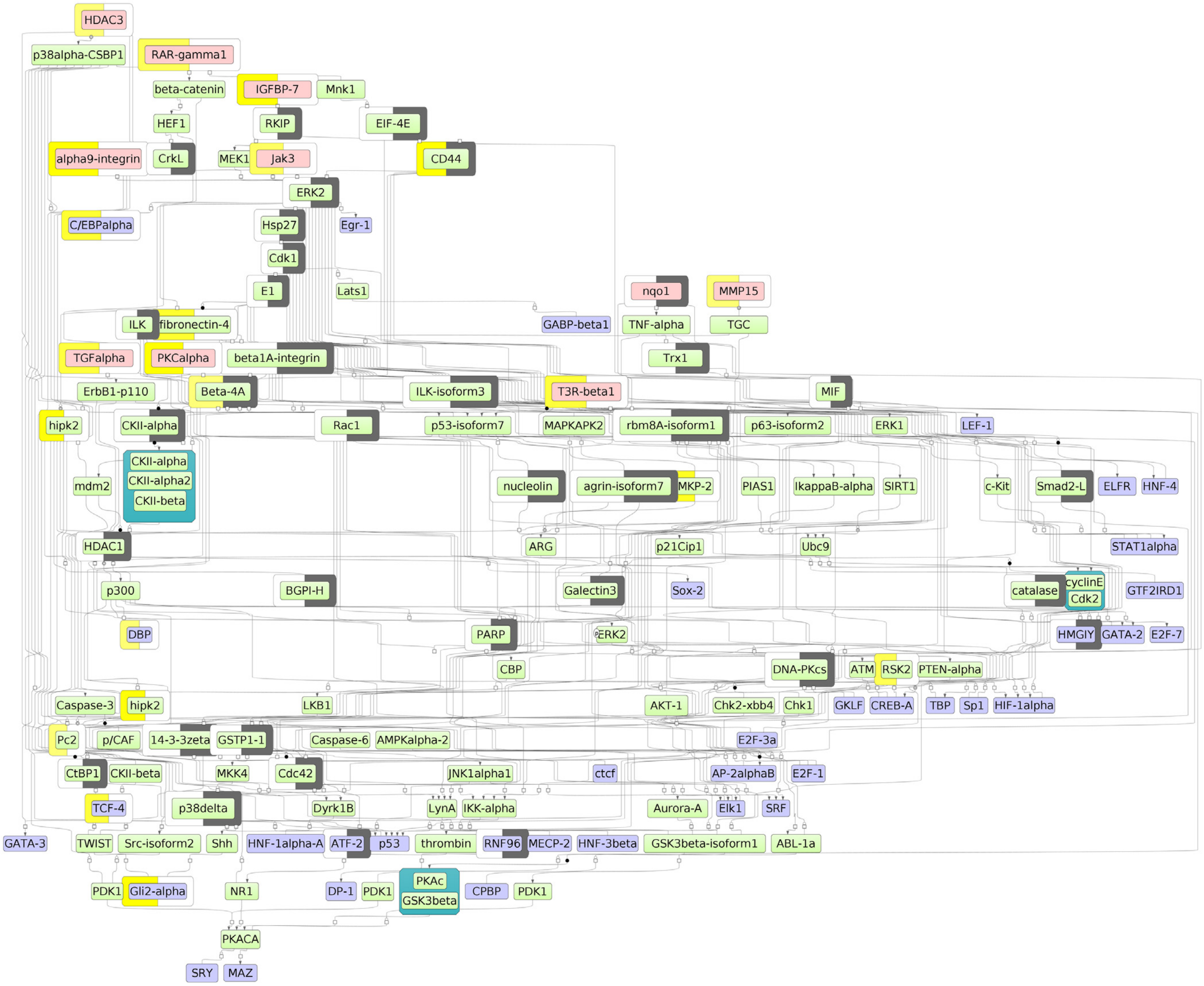


Fig. 2. A part of the predicted signal transduction network of MTX-resistant colorectal cancer cells that is reconstructed with the help of the master-regulator search algorithm implemented in the geneXplain platform. Transcription factors (blue) are shown at the bottom and in the center. Potential master regulators (pink) are shown at the top. The direction of signal flow is from top to bottom. Intermediary molecules are green. Gray half-circles indicate proteins identified by the proteomics experiment in HT29 cell line. Yellow half-circles indicate proteins encoded by genes up-regulated in MTX-resistant cells.

metabolites collected in the HMDB, Human Metabolome Database, version 2.5.

The list of 30 potential targets identified by the master-regulator search that correspond to 19 different PASS activities is shown in Table 3. The PASS activities are represented by inhibitors, agonists and antagonists of the identified targets.

About 14% of the potential master regulators identified at the network analysis step we could associate respective PASS activities (30 out of 220 potential master regulators represented by 19 PASS activities). We considered these 19 PASS activities as an initial set to begin our search for promising compounds.

The results of the scanning of the compound libraries are shown in Table 4. In the library of the top 200 drugs we identified several drugs that fulfilled the criteria of Pa > Pi for 8 activities from the list of 19 activities. In Fig. S3 (see Supplementary material) we show a screenshot of the PharmaExpert program. We identified 5 drugs that all share prediction for two activities – “Integrin antagonist” and “TGF-beta agonist” (which are among most up-regulated targets). For the first drug, **divalproex**, PASS actually predicted in total 8 activities from our list with Pa > Pi (see the full list of predicted activities in the center of the screenshot Fig. S3). Divalproex, which is also known as valproic acid, is an old drug primarily used to treat epilepsy and bipolar disorder and to prevent migraine headaches. Recently a number of clinical trials were performed with this drug and they confirmed its efficacy for treatment of Acute Myeloid Leukaemia [49], Cervical cancer [50] and Breast cancer [51]. So, the use of this drug for potential sensitization of resistant colon cancers towards methotrexate, as we have predicted in our analysis, makes perfect sense.

Another highly potent compound was found by applying PASS to the Prestwick chemical library. Among the best hits we found the known drug zardaverine (see Fig. S4), which is known and highly specific inhibitor of all five subtypes of the enzyme phosphodiesterase (PDE) (as is also predicted by PASS – the Pa = 0.867), which are among our selected targets. PASS also predicted the potential activity of this drug as IGF1 antagonist (this activity was additionally selected by us as possible interfering with one of our targets – IGFBP7 protein) (see Fig. 1). There is a number of recent studies confirming the potential use of zardaverine in cancer therapy, against hepatocellular carcinoma [52] and against Chronic Lymphocytic Leukemia [53].

Table 3
List of 30 potential targets identified by the master-regulator search corresponding to known PASS activities. “Reached from set” – number of transcription factors from the initial set of 49 TFs (see Table 1) that can receive the signal from the master regulator through the signal transduction network with a number of steps less then 10. “Score” – score of the master regulator computed as described in the Methods section. “LogFC” – the logarithm to base 2 of the Fold Change of the expression of the gene encoding the corresponding master-regulator protein in the MTX-resistant versus sensitive cells. “Proteomics” – “yes” means that the respective protein was detected by the proteomics experiment in the HT29 cells.

Proteins: Transpath ID	Master molecule name	ID	Gene description	PASS activity	Reached from set	Score	logFC	Proteomics
MO000034329	alpha9-integrin(h)	ITGA9	alpha 9,integrin	Integrin antagonist	37	0.45	3.14	
MO000057624	PKCalpha(h)	PRKCA	alpha,protein kinase C	Protein kinase C inhibitor	37	0.82	2.36	
MO000133221	DCR2(h)	TNFRSF10D	decoy with truncated death domain,member 10d, tumor necrosis factor receptor superfamily	Tumour necrosis factor agonist	33	0.31	2.07	
MO000002316	cathepsinB (h)	CTSB	cathepsin B	Cathepsin B inhibitor	36	0.37	1.77	
MO000107702	PKAc-beta-isoform1(h)	PRKACB	beta,cAMP-dependent,catalytic,protein kinase	Protein kinase A inhibitor	37	0.75	1.50	
MO000021287	TGFbeta1(h)	TGFB1	beta 1,transforming growth factor	Transforming growth factor agonist	37	0.58	1.45	
MO000126529	MDC9-isoform1(h)	ADAM9	ADAM metallopeptidase domain 9	Metalloproteinase inhibitor	36	0.40	1.12	
MO000043254	DR5-L(h)	TNFRSF10B	member 10b,tumor necrosis factor receptor superfamily	Tumour necrosis factor agonist	36	0.35	0.97	
MO000060291	PKD3-isoform1(h)	PRKD3	protein kinase D3	Protein kinase C inhibitor	36	0.38	0.86	
MO000081115	PDE4A-isoform1(h)	PDE4A	cAMP-specific,phosphodiesterase 4A	Phosphodiesterase IV inhibitor	34	0.31	0.78	
MO000021670	T3R-beta1(h)	THRB	beta,thyroid hormone receptor	Thyroid hormone agonist	36	0.39	0.77	
MO000130575	PI31(h)	PSMF1	macropain) inhibitor subunit 1 (PI31),proteasome (prosome	Proteasome inhibitor	32	0.31	0.76	
MO000080275	TGFalpha-isoform1(h)	TGFA	alpha,transforming growth factor	Transforming growth factor agonist	37	0.62	0.75	
MO000115412	PDGFA-long (h)	PDGFA	platelet-derived growth factor alpha polypeptide	Platelet growth factor antagonist	37	0.51	0.74	
MO000082169	Hic-5-isoform1(h)	TGFB1I1	transforming growth factor beta 1 induced transcript 1	Transforming growth factor agonist	37	0.45	0.72	
MO000079390	HDAC5-isoform1(h)	HDAC5	histone deacetylase 5	Histone deacetylase inhibitor	37	0.42	0.68	
MO000083689	CD26(h)	DPP4	dipeptidyl-peptidase 4	Dipeptidyl peptidase IV inhibitor	36	0.35	0.67	
MO000083701	TGFbeta-2A (h)	TGFB2	beta 2,transforming growth factor	Transforming growth factor agonist	37	0.53	0.66	
MO000025589	RAR-gamma1(h)	RARG	gamma,retinoic acid receptor	Retinoic acid receptor agonist	36	0.37	0.65	
MO000130058	THANK-isoform1(h)	TNFSF13B	member 13b,tumor necrosis factor (ligand) superfamily	Tumour necrosis factor agonist	37	0.39	0.63	
MO000082601	Jak3-isoform2(h)	JAK3	Janus kinase 3	Janus tyrosine kinase 3 inhibitor	37	0.81	0.62	
MO000086979	TUBB2(h)	TUBB2A	beta 2A class IIa,tubulin	Tubulin agonist	37	0.41	0.61	yes
MO000078302	FGFR-2-isoform16(h)	FGFR2	fibroblast growth factor receptor 2	Fibroblast growth factor antagonist	36	0.40	0.60	
MO000025446	T3R-alpha1 (h)	THRA	alpha,thyroid hormone receptor	Thyroid hormone agonist	36	0.37	0.59	
MO000139037	nqo2(h)	NQO2	NAD(P)H dehydrogenase,quinone 2	NAD(P)H dehydrogenase (quinone) inhibitor	36	0.39	0.55	
MO000117489	MMP15(h)	MMP15	matrix metallopeptidase 15 (membrane-inserted)	Metalloproteinase inhibitor	37	0.44	0.54	
MO000079379	HDAC3-isoform1(h)	HDAC3	histone deacetylase 3	Histone deacetylase inhibitor	37	0.68	0.53	
MO000059956	Beta-4C(h)	ITGB4	beta 4,integrin	Integrin antagonist	37	0.58	0.53	yes
MO000059062	CD51-isoform1(h)	ITGAV	alpha V,integrin	Integrin alphaVbeta3 antagonist	37	0.59	0.52	
MO000057416	PKCzeta-isoform1(h)	PRKCZ	protein kinase C,zeta	Protein kinase C inhibitor	37	0.75	0.50	

Finally, the application of PASS to the collection of human metabolites resulted in a number of interesting candidate compounds that can be used in further experimental studies. As one may see in Fig. S5, requiring at least two activities from our list to have Pa > Pi we identified 348 compounds. For the top one, nicotinamide N-oxide, PASS predicted three activities from our list of 19 activities. Again, the activity as an inhibitor of enzyme phosphodiesterase is predicted with very high Pa = 0.707.

Nicotinamide is known to sensitize a number of rodent tumors to single dose of radiation [54]. Its combination with carbogen results in large enhancement of tumor response to certain treatment and it was confirmed in a clinical trials [55]. So, we can assume that this compound can be also a very good candidate for possible sensitization of MTX resistance as we can propose it using analysis of the experiments with the MTX resistant and sensitive cell lines.

Table 4
Results of analysis by PASS of three libraries of drugs and chemical compounds. “PASS Activity” is the name of the pharmacological activity that was predicted by the PASS program for a given compound (under condition Pa > Pi). Pa – probability to be active, Pi – probability to be inactive.

Drug/compound name	Library	PASS Activity	Pa	Pi
Divalproex	Top 200 drugs	Integrin antagonist	0.059	0.017
		TGF agonist	0.153	0.04
Zardaverine	Prestwick chemical library	Insulin like growth factor 1 antagonist	0.156	0.05
		Phosphodesterase IV inhibitor	0.867	0.002
Nicotinamide N-oxide	Collection of human metabolites	NAD(P)H dehydrogenase (quinone) inhibitor	0.063	0.057
		Phosphodesterase IV inhibitor	0.707	0.003

The further study will be necessary in order to confirm these findings in vivo and potentially translate them to the clinical applications.

4. Conclusions

In this paper we have applied our earlier developed approach of “upstream analysis,” [11,18] to multi-omics data including transcriptomics microarray data, proteomics data and data on epigenomics (ChIP-seq). All these experimental data were extracted from different publications on experiments that were done by different groups. An important novel part of the approach enabling integration of proteomics data in such analysis is the “Context Algorithm” which is described in this paper. The list of proteins identified with the help of modern methods of proteomics are used in our approach as sets of “context proteins” that help the algorithm to find master regulators in the huge signal transduction networks of the cells. We also introduced a novel way of integrating transcriptomics and epigenomic data, when peaks of active chromatin identified by ChIP-seq experiments are intersected with long 5′ upstream and downstream regions of differentially expressed genes in order to detect the locations of most important “enhancers” and “silencers” of genes driving the MTX-resistance. Frequency analysis of TFBS and analysis of composite regulatory modules in such “enhancer” and “silencer” regions allows to identify more precisely transcription factors involved in the mechanism under study. Our approach gives us a nice possibility to integrate those different types of data helping to achieve our goal of identification of potent drug targets and perspective chemical compounds that can be potentially used to resolve the problem of induced resistance of cancer cells towards chemotherapy by methotrexate (MTX). The considerable part of this analysis has been done with a help of automatic workflows in the geneXplain platform and therefore can be easily reproduced and can be applied to analysis of other similar tasks. The schema of this workflow is shown in Fig. S6 in Supplement.

As a result we identified a number of very promising drug targets, such as, PKC-alpha, TGF-beta, TGF-alpha, cAMP-specific phosphodiesterase 4A, insulin-like growth factor-binding protein 7, alpha9-integrin and several others and reconstructed a potential signal transduction network connecting these targets with the transcription factors triggering activity of the MTX-resistance genes. Many of these proteins are already known as important targets for anti-cancer drug therapy and our results suggest them for the use as anti-resistance targets. Among these targets we also identified very interesting signaling molecules that most probably play an important role in the resistance mechanism. For instance, recently it was shown that integrins (that were suggested by us among the most prominent targets) play a very important role in colon cancer cell resistance to methotrexate by controlling low density of tumor cells [4]. We can speculate that the use of such important new targets as integrins in combination with other predicted targets is a promising way to combat drug resistance in cancer. As the final step of our analysis we applied a chemoinformatics approach (PASS program) for identification of chemical compounds that have a potential of inhibiting or activating the targets predicted at the previous step. This approach demonstrated a very good potential in computational search for such compounds. Among identified compounds that can be potentially used to sensitize the MTX resistance of the studied cell line we suggested known drugs, such as zardaverine and divalproex as well as human metabolites such as nicotinamide N-oxide.

We should emphasize again here that of course all our findings of potential anti-MTX-resistance drug targets and potential compounds should be further validated by extensive in vivo studies in order to think about potential translation of this findings to clinical applications.

Conflicts of interest

AK, PS, JK, OK and EW are employees of geneXplain GmbH, which maintains and distributes the geneXplain platform used in this study.

Author contributions

AK conducted the upstream analysis of all data sets with the geneXplain platform and coordinated the work reported here. PS developed and applied the enriched TFBS finding algorithm. TV contributed to the development of geneXplain platform and algorithms of pathway and promoter analysis. JK contributed to the development of workflows in geneXplain platform. OK contributed to the concept of composite elements and upstream analysis. VP contributed with the PASS program to the methods of prediction of potential active chemical compounds. EW contributed to the classification of transcription factors, to the overall concept of upstream analysis and to the final editing of the manuscript. VP contributed to the application of PASS and PharmaExpert programs for search of active compounds.

Acknowledgments

This work was done with the financial support of Targeted Program “Research and development on priority directions of science and technology in Russia, 2014-2021”, Contract 14.604.21.0101, unique identifier of the applied scientific project: RFMEFI60414X0101. The work was partially supported (VP) in the framework of the Russian State Academies of Sciences Fundamental Research Program for 2013–2020. This work was also supported by the following grants of the EU FP7 program: “SYSCOL”, “SysMedIBD”, “RESOLVE” and “MIMOMICS”. We are also very grateful to our colleague at former Biobase GmbH, Niko Voss, for the ideas of the algorithm on pathway analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.euprot.2016.09.002.

References

[1] M.J. Osborn, M. Freeman, F.M. Huennekens, Inhibition of dihydrofolic reductase by aminopterin and amethopterin, Proc. Soc. Exp. Blot. Med. 97 (1958) 429.

[2] C. Morales, M. Ribas, G. Aiza, M.A. Peinado, Genetic determinants of methotrexate responsiveness and resistance in colon cancer cells, Oncogene 24 (October (45)) (2005) 6842–6847.

[3] J.M. De Anta, C. Mayo de Las Casas, F.X. Real, X. Mayol, Unmasking the mechanisms of colon cancer cell resistance to methotrexate: cell drug sensitivity is dependent on a transiently adaptive mechanism, Gastroentérologie Clinique et Biologique 26 (avril (4)) (2002) 399 (GCB-04-2002-26-4-0399-8320-101019-ART34).

[4] K.R. Fischer, A. Durrans, S. Lee, J. Sheng, F. Li, S.T. Wong, H. Choi, T. El Rayes, S. Ryu, J. Troeger, R.F. Schwabe, L.T. Vahdat, N.K. Altorki, V. Mittal, D. Gao, Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance, Nature 527 (November (7579)) (2015) 472–476, doi:http://dx.doi.org/10.1038/nature15748.

[5] N. Kolesnikov, E. Hastings, M. Keays, O. Melnichuk, Y.A. Tang, E. Williams, M. Dylag, N. Kurbatova, M. Brandizi, T. Burdett, ArrayExpress update—simplifying data submissions, Nucleic Acids Res. 43 (2015) D1113–D1116.

[6] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, et al., NCBI GEO. archive for functional genomics data sets–update, Nucleic Acids Res. 41 (2013) D991–D995.

[7] R. Petryszak, T. Burdett, B. Fiorelli, N.A. Fonseca, M. Gonzalez-Porta, E. Hastings, W. Huber, S. Jupp, M. Keays, N. Kryvych, Expression atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments, Nucleic Acids Res. 42 (2014) D926–D932.

[8] C.M. Smith, J.H. Finger, T.F. Hayamizu, I.J. McCright, J. Xu, J. Berghout, J. Campbell, L.E. Corbani, K.L. Forthofer, P.J. Frost, The mouse gene expression database (GXD): 2014 update, Nucleic Acids Res. 42 (2014) D818–D824.

[9] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.

[10] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets, Nucleic Acids Res. 40 (2012) D109–D114.

[11] A. Kel, N. Voss, R. Jauregui, O. Kel-Margoulis, E. Wingender, Beyond microarrays: find key transcription factors controlling signal transduction pathways, BMC Bioinf. 7 (2006) S13.

[12] H. Michael, J. Hogan, A. Kel, O. Kel-Margoulis, F. Schacherer, N. Voss, E. Wingender, Building a knowledge base for systems pathology, Brief. Bioinform. 9 (2008) 518–531.

[13] P. Stegmaier, N. Voss, T. Meier, A. Kel, E. Wingender, J. Borlak, Advanced computational biology methods identify molecular switches for malignancy in an EGF mouse model of liver cancer, PLoS One 6 (2011) e17738.

[14] E. Wingender, The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation, Brief. Bioinform. 9 (2008) 326–332.

[15] A.E. Kel, E. Gössling, I. Reuter, E. Cheremushkin, O.V. Kel-Margoulis, E. Wingender, MATCH: a tool for searching transcription factor binding sites in DNA sequences, Nucleic Acids Res. 31 (2003) 3576–3579.

[16] T. Waleev, D. Shtokalo, T. Konovalova, N. Voss, E. Cheremushkin, P. Stegmaier, O. Kel-Margoulis, E. Wingender, A. Kel, Composite module analyst: identification of transcription factor binding site combinations using genetic algorithm, Nucleic Acids Res. 34 (July (1)) (2006) W541–W545 (Web Server issue).

[17] M. Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, O. Kel-Margoulis, E. Wingender, TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations, Nucleic Acids Res. 34 (2006) D546–D551.

[18] J. Koschmann, A. Bhar, P. Stegmaier, A.E. Kel, E. Wingender, Upstream analysis: an integrated promoter-pathway analysis approach to causal interpretation of microarray data, Microarrays 4 (2015) 270–286, doi:http://dx.doi.org/10.3390/microarrays4020270.

[19] C.H. Reynolds, K.M. Merz, D. Ringe (Eds.), Drug Design: Structure- and Ligand-Based Approaches, 1st ed., Cambridge University Press, Cambridge, UK, 2010978-0521887236.

[20] A. Tropsha, QSAR in drug discovery, in: C.H. Reynolds, K.M. Merz, D. Ringe (Eds.), Drug Design Structure- and Ligand-Based Approaches, 1st ed., Cambridge University Press, Cambridge, UK, 2010978-0521887236, pp. 151–164.

[21] D. Filimonov, V. Poroikov, Yu. Borodina, T. Gloriozova, Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors, J. Chem. Inf. Comput. Sci 39 (1999) 666–670.

[22] D.A. Filimonov, V.V. Poroikov, in: Alexandre Varnek, Alexander Tropsha (Eds.), Probabilistic Approach in Activity Prediction, RSC Publishing, Cambridge (UK), 2008, pp. 182–216.

[23] Demo workflows. Available online: http://www.genexplain.com/demo-workflows.

[24] E. Selga, C. Morales, V. Noé, M.A. Peinado, et al., Role of caveolin 1, E-cadherin, Enolase 2 and PKCalpha on resistance to methotrexate in human HT29 colon cancer cells, BMC Med. Genomics 1 (August (11)) (2008) 35 (PMID: 18694510).

[25] G.K. Smyth, Limma: linear models for microarray data, in: R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer, New York, 2005, pp. 397–420.

[26] K.M. McMahon, M. Volpato, H.Y. Chi, P. Musiwaro, K. Poterlowicz, Y. Peng, A.J. Scally, L.H. Patterson, R.M. Phillips, C.W. Sutton, Characterization of changes in the proteome in different regions of 3D multicell tumor spheroids, J. Proteome Res. 11 (May (5)) (2012) 2863–2875 PubMed(s): 22416669.

[27] Benjamin L. Allen, Dylan J. Taatjes, The Mediator complex: a central integrator of transcription, Nat. Rev. Mol. Cell Biol. 16 (2015) 155–166.

[28] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.

[29] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based analysis of chIP-Seq (MACS), Genome Biol. 9 (9) (2008) R137, doi:http://dx.doi.org/10.1186/gb-2008-9-9-r137.

[31] D.A. Filimonov, V.V. Poroikov, in: Alexandre Varnek, Alexander Tropsha (Eds.), Probabilistic Approach in Activity Prediction, RSC Publishing, Cambridge (UK), 2008, pp. 182–216.

[32] P. Stegmaier, N. Voss, T. Meier, A. Kel, E. Wingender, J. Borlak, Advanced computational biology methods identify molecular switches for malignancy in an EGF mouse model of liver cancer, PLoS One 6 (2011) e17738.

[33] D. Viemann, M. Goebeler, S. Schmid, K. Klimmek, C. Sorg, S. Ludwig, J. Roth, Transcriptional profiling of IKK2/NF-kappa B- and p38 MAP kinase-dependent gene expression in TNF-alpha-stimulated primary human endothelial cells, Blood 103 (2004) 3365–3373.

[34] R.T. Schimke, R.S. Kaufman, F.W. Alt, R.F. Kellems, Gene amplification and drug resistance in cultured murine cells, Science 202 (1978) 1051.

[35] J.R. Bertino, E. Göker, R. Gorlick, W.W. Li, D. Banerjee, Resistance mechanisms to methotrexate in *Tumors*, Oncologist 1 (4) (1996) 223–226.

[36] L. Good, G.P. Dimri, J. Campisi, K.Y. Chen, Regulation of dihydrofolate reductase gene expression and E2F components in human diploid fibroblasts during growth and senescence, J. Cell. Physiol. 168 (3) (1996) 580–588.

[37] S.Y. Lin, A.R. Black, D. Kostic, S. Pajovic, C.N. Hoover, J.C. Azizkhan, Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction, Mol. Cell. Biol. 16 (April (4)) (1996) 1668–1675.

[38] O.V. Kel-Margoulis, A.E. Kel, I. Reuter, I.V. Deineko, E. Wingender, TRANSCompel: a database on composite regulatory elements in eukaryotic genes, Nucleic Acids Res. 30 (January (1)) (2002) 332–334.

[39] G. Marra, I. Iaccarino, T. Lettieri, G. Roscilli, P. Delmastro, J. Jiricny, Mismatch repair deficiency associated with overexpression of the MSH3 gene, Proc. Natl. Acad. Sci. U. S. A. 95 (15) (1998) 8568–8573, doi:http://dx.doi.org/10.1073/pnas.95.15.8568.PMC21116.PMID9671718.

[40] B.L. Allen, D.J. Taatjes, The Mediator complex: a central integrator of transcription, Nat. Rev. Mol. Cell Biol. 16 (March (3)) (2015) 155–166, doi: http://dx.doi.org/10.1038/nrm3951.

[41] R. Firestein, A.J. Bass, S.Y. Kim, I.F. Dunn, S.J. Silver, I. Guney, E. Freed, A.H. Ligon, N. Vena, S. Ogino, M.G. Chheda, P. Tamayo, S. Finn, Y. Shrestha, J.S. Boehm, S. Jain, E. Bojarski, C. Mermel, J. Barretina, J.A. Chan, J. Baselga, J. Tabernero, D.E. Root, C.S. Fuchs, M. Loda, R.A. Shivdasani, M. Meyerson, W.C. Hahn, CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity, Nature 455 (September (7212)) (2008) 547–551, doi:http://dx.doi.org/10.1038/nature07179.

[42] Y. Zwang, M. Oren, Y. Yarden, Consistency test of the cell cycle: roles for p53 and EGR1, Cancer Res. 72 (2012) 1051–1054.

[43] El-Karim, et al., Krüppel-like factor 4 regulates genetic stability in mouse embryonic fibroblasts, Mol. Cancer (2013), doi:http://dx.doi.org/10.1186/1476-4598-12-89.

[44] G. Chen, T.G. Gharib, C.C. Huang, J.M.G. Taylor, D.E. Misek, S.L.R. Kardia, T.J. Giordano, M.D. Iannettoni, M.B. Orringer, S.M. Hanas, D.G. Beer, Discordant protein and mrna expression in lung adenocarcinomas, Mol. Cell. Proteomics 1 (4) (2002) 304–313.

[45] P. Goldhoff, N.M. Warrington, D.D. Limbrick Jr., A. Hope, B.M. Woerner, E. Jackson, A. Perry, D. Piwnica-Worms, J.B. Rubin, Targeted inhibition of cyclic AMP phosphodiesterase-4 promotes brain tumor regression, Clin. Cancer Res. 14 (December (23)) (2008) 7717–7725.

[46] G. Valabrega, F. Montemurro, I. Sarotto, A. Petrelli, P. Rubini, C. Tacchetti, M. Aglietta, P.M. Comoglio, S. Giordano, TGFalpha expression impairs trastuzumab-induced HER2 downregulation, Oncogene 24 (April (18)) (2005) 3002–3010.

[47] L.A. Mostovich, T.Y. Prudnikova, A.G. Kondratov, D. Loginova, P.V. Vavilov, V.I. Rykova, S.V. Sidorov, T.V. Pavlova, V.I. Kashuba, E.R. Zabarovsky, E.V. Grigorieva, Integrin alpha9 (ITGA9) expression and epigenetic silencing in human breast tumors, Cell Adh. Migr. 5 (September-October (5)) (2011) 395–401, doi:http://dx.doi.org/10.4161/cam.5.5.17949.

[48] H. Hatakeyama, J. Parker, D. Wheeler, P. Harari, S. Levy, C.H. Chung, Effect of insulin-like growth factor 1 receptor inhibitor on sensitization of head and neck cancer cells to cetuximab and methotrexate, J. Clin. Oncol. (2009) ASCO Annual Meeting Proceedings (Post-Meeting Edition).Vol 27, No 15S (May 20 Supplement), 2009: 6079..

[49] G. Bug, M. Ritter, B. Wassmann, C. Schoch, T. Heinzel, K. Schwarz, A. Romanski, O.H. Kramer, M. Kampfmann, D. Hoelzer, A. Neubauer, M. Ruthardt, O.G. Ottmann, Clinical trial of valproic acid and all-trans retinoic acid in patients with poor-risk acute myeloid leukemia, Cancer 104 (12) (2005) 2717–2725, doi:http://dx.doi.org/10.1002/cncr.21589.PMID16294345.

[50] J. Coronel, L. Cetina, I. Pacheco, C. Trejo-Becerril, A. González-Fierro, E. de la Cruz-Hernandez, E. Perez-Cardenas, L. Taja-Chayeb, D. Arias-Bofill, M. Candelaria, S. Vidal, A. Dueñas-González, A double-blind, placebo-controlled, randomized phase III trial of chemotherapy plus epigenetic therapy with hydralazine valproate for advanced cervical cancer. Preliminary results, Med. Oncol. 28 (Suppl. 1) (2011) S540–S546, doi:http://dx.doi.org/10.1007/s12032-010-9700-3 (PMID 20931299).

[51] P. Munster, D. Marchion, E. Bicaku, M. Lacevic, J. Kim, B. Centeno, A. Daud, A. Neuger, S. Minton, D. Sullivan, Clinical and biological effects of valproic acid as a histone deacetylase inhibitor on tumor and surrogate tissues: phase I/II trial of valproic acid and epirubicin/FEC, Clin. Cancer Res. 15 (7) (2009) 2488–2496, doi:http://dx.doi.org/10.1158/1078-0432.CCR-08-1930.PMID19318486.

[52] L. Sun, H. Quan, C. Xie, L. Wang, Y. Hu, L. Lou, Phosphodiesterase 3/4 inhibitor zardaverine exhibits potent and selective antitumor activity against hepatocellular carcinoma both in vitro and in vivo independently of phosphodiesterase inhibition, PLoS One 9 (March (3)) (2014) e90627, doi: http://dx.doi.org/10.1371/journal.pone.0090627 (eCollection 2014).

[53] E. Moon, R. Lee, R. Near, L. Weintraub, S. Wolda, A. Lerner, Inhibition of PDE3B augments PDE4 inhibitor-induced apoptosis in a subset of patients with chronic lymphocytic leukemia, Clin. Cancer Res. 8 (February (2)) (2002) 589–595.

[54] M.R. Horsman, D.J. Chaplin, J.M. Brown, Radiosensitisation by nicotinamide in vivo: a greater enhancement of tumor damage compared to that of normal tissues, Radiat. Res. 109 (1987) 479–489.

[55] E. Kjellen, M.C. Joiner, J.M. Collier, H. Johns, A. Rojas, A therapeutic benefit from combining normobaric carbogen or oxygen with nicotinamide in fractionated X-ray treatments, Radiother. Oncol. 22 (1991) 81–91.

[56] A. Kel, Master regulators and transcriptiption factor binding sites found by upstream analysis of multi-omics data on methotrexate resistance of colon cancer. Data in Brief. submitted.