A network model for angiogenesis in ovarian cancer

Share Your Story

BMC Bioinformatics

A network model for angiogenesis in ovarian cancer

Kimberly Glass[1,2,3], John Quackenbush[1,2], Dimitrios Spentzos[4], Benjamin Haibe-Kains[5] and Guo-Cheng Yuan[1,2*]

Abstract

Background: We recently identified two robust ovarian cancer subtypes, defined by the expression of genes involved in angiogenesis, with significant differences in clinical outcome. To identify potential regulatory mechanisms that distinguish the subtypes we applied PANDA, a method that uses an integrative approach to model information flow in gene regulatory networks.

Results: We find distinct differences between networks that are active in the angiogenic and non-angiogenic subtypes, largely defined by a set of key transcription factors that, although previously reported to play a role in angiogenesis, are not strongly differentially-expressed between the subtypes. Our network analysis indicates that these factors are involved in the activation (or repression) of different genes in the two subtypes, resulting in differential expression of their network targets. Mechanisms mediating differences between subtypes include a previously unrecognized pro-angiogenic role for increased genome-wide DNA methylation and complex patterns of combinatorial regulation.

Conclusions: The models we develop require a shift in our interpretation of the driving factors in biological networks away from the genes themselves and toward their interactions. The observed regulatory changes between subtypes suggest therapeutic interventions that may help in the treatment of ovarian cancer.

Keywords: Network modeling, Gene regulation, Regulatory networks, Ovarian cancer, Cancer subtypes, Angiogenesis

Background

Ovarian cancer is the fifth leading cause of cancer death for women in the U.S. and the seventh most fatal worldwide [1]. Although ovarian cancer is notable for its initial sensitivity to platinum-based therapies, the vast majority of women eventually recur and succumb to increasingly platinum-resistant disease. Despite significant investment, improvements in patient prognosis have been slow and usually in small increments. The disease generally presents at an advanced stage (III/IV) and the five-year survival rate of advanced disease is less than 30% with median survival only slightly longer than two years [2]. Furthermore, ovarian cancer patients often undergo similar treatment regimens, mainly because the highly suspected multiple subtypes have not yet been well characterized in terms of their biological significance.

In previous work, we analyzed gene expression data from 129 high-grade serous ovarian cancer samples and identified a poor-prognosis subtype characterized by the expression of angiogenic genes [3]. This subtype and the associated differences in patient survival were validated using gene expression data from a collection of 1606 ovarian cancer samples assembled from ten independent published studies. Multiple other subtypes have been proposed [4,5], but the importance of this subtyping, relative to other definitions, is that angiogenesis represents a process that is of potential clinical relevance and it is the only subtyping model which has been shown to be robust and prognostic in multiple independent datasets.

Angiogenesis is one of the well-characterized hallmarks associated with cancer progression [6], playing an important role in maintaining tumor growth [7]. Angiogenesis is facilitated by interactions between cells and the extra-cellular matrix [8,9] and is associated with increased expression among a set of particular genes,

* Correspondence: gcyuan@jimmy.harvard.edu
[1]Dana-Farber Cancer Institute, Boston, MA, USA
[2]Harvard School of Public Health, Boston, MA, USA
Full list of author information is available at the end of the article

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 2 of 17

including, but not limited to, matrix metalloproteinases (MMPs) [10] and VEGF [11-13]. Angiogenesis inhibition is being intensely studied as a possible therapeutic advance in ovarian cancer, but the effects in survival are still modest, suggesting that our understanding of the molecular underpinnings and biological implications of angiogenesis in this disease is still limited [2,14]. A number of clinical trials have tested drugs targeting angiogenic factors and shown these drugs have anti-cancer activity in a subset of ovarian cancer patients [15-18]; however, better understanding of the complex mechanisms driving response to these therapies is crucial to improving their efficacy and patient outcomes [17,19].

Although we have long studied ovarian cancer from the perspective of single genes and their properties, it has become clear that more integrative, systems-level analyses are necessary to better understand how the disease and its subtypes develop and progress, and how it may respond to different therapeutic interventions. The characterization of biological processes can distinguish disease states in cases where single gene biomarkers cannot [20]. The importance of applying network approaches in particular to better understand disease has previously been highlighted [21-24]. Simultaneously, it has become evident that integrative approaches that incorporate multiple sources of data to model biological systems often yield the most informative results [25-27]. Along these lines, we recently described an integrative network inference method, PANDA (**P**assing **A**ttributes between **N**etworks for **D**ata **A**ssimilation), that models information flow in regulatory networks by searching for agreement among various data-types, using information from each to iteratively refine predictions in the others [28].

PANDA models network interactions as communication between "transmitters" and "receivers". In the context of PANDA's regulatory networks, the transmitters are transcription factors and the receivers are their downstream target genes. This approach recognizes that for communication to occur, both the transmitter and the receiver have an essential role – although a transcription factor is *responsible* for regulating a target gene, the gene must also be *available* to be regulated. By constructing a "prior" regulatory network consisting of potential routes for communication (for example, by mapping transcription factor motifs to a reference genome) and integrating with other sources of regulatory information (such as protein interaction and gene expression data), one can estimate the *responsibility* and *availability* of each potential interaction, predict where communication is succeeding and failing, and deduce condition-specific network structures.

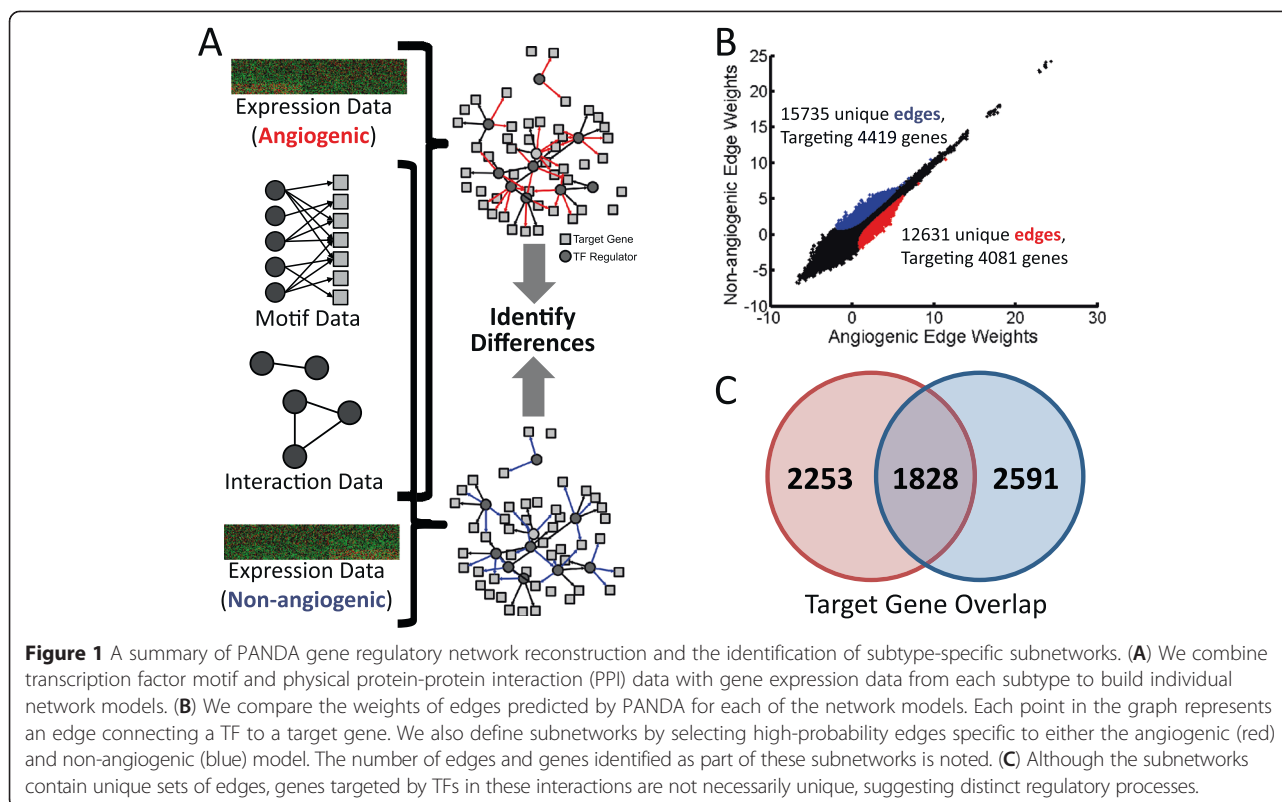Here we describe the application of PANDA to reconstructing subtype-specific regulatory networks in ovarian cancer. We identify differences in network topologies between the angiogenic and non-angiogenic subtypes, and use this information to suggest potential therapies that may be efficacious in treating patients with angiogenic-subtype ovarian tumors.

Results and discussion
Building angiogenic and non-angiogenic specific regulatory network modules

To begin, we downloaded mRNA expression data for 510 ovarian samples profiled by the Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp, [5]), normalized this data using fRMA [29], and mapped probes to Ensembl identifiers using the biomaRt Bioconductor package version 2.8.1 [30]. As reported in [3] we classified samples as belonging to either the angiogenic or non-angiogenic subtype; 188 were classified as part of the angiogenic subtype, and 322 were classified as non-angiogenic. We constructed separate genome-wide regulatory network models for the two subtypes. We began by mapping 111 TFs with known binding motifs to the promoters, here defined as [−750,+250] base-pairs around the transcription start site, of the 12290 genes with expression data in TCGA samples. Because transcriptional regulation involves assembly of protein complexes and allows for combinatorial regulatory processes, we collected information regarding physical protein interaction data between human transcription factors estimated using a mouse-2-hybrid assay [31]. We used PANDA to integrate information from transcription factor binding motifs, protein-protein interaction data, and subtype-specific gene expression, constructing directed transcriptional regulatory networks for the angiogenic and non-angiogenic ovarian cancer subtypes (Figure 1A).

For each edge that connects a transcription factor to its target gene, PANDA assigns a weight, in z-score units, that reflects the confidence level of a potential inferred regulatory relationship. Not surprisingly, we found the edge weights for the angiogenic and non-angiogenic regulatory networks to be highly correlated (Figure 1B), representing common regulatory mechanisms and processes active in both subtypes. However, we also found regulatory edges that were more strongly supported by either the angiogenic or the non-angiogenic subtype. We identified 12631 edges that we assigned to an angiogenic subnetwork, shown in red, and 15735 edges for the non-angiogenic subnetwork, shown in blue. The edges in the angiogenic subnetwork target 4081 genes while the non-angiogenic subnetwork edges target 4419; of these, 1828 genes are targeted in both subnetworks (Figure 1C), although by different upstream transcription factors. This may reflect the fact that in addition to different pathways being activated in each of the subtypes, there may also be a complex "re-wiring" of the networks around commonly targeted genes.

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 3 of 17

Figure 1 A summary of PANDA gene regulatory network reconstruction and the identification of subtype-specific subnetworks. (**A**) We combine transcription factor motif and physical protein-protein interaction (PPI) data with gene expression data from each subtype to build individual network models. (**B**) We compare the weights of edges predicted by PANDA for each of the network models. Each point in the graph represents an edge connecting a TF to a target gene. We also define subnetworks by selecting high-probability edges specific to either the angiogenic (red) and non-angiogenic (blue) model. The number of edges and genes identified as part of these subnetworks is noted. (**C**) Although the subnetworks contain unique sets of edges, genes targeted by TFs in these interactions are not necessarily unique, suggesting distinct regulatory processes.

Network analysis reveals biological mechanisms associated with regulatory differences

We compared the corresponding subnetworks and identified a subset of transcription factors associated with the strongest regulatory changes. To do this we identified the genes targeted for regulation by a transcription factor in each of the two subnetworks and determined both an "edge enrichment" score as well as a p-value significance for the difference in the number of target genes. Specifically, we define "edge enrichment" for each transcription factor as the out-degree of that transcription factor (number of edges pointing out to a target gene) in the angiogenic subnetwork divided by the out-degree for the same transcription factor in the non-angiogenic subnetwork (multiplied by a normalization factor equal to the total number of edges in the angiogenic subnetwork divided by the total number of edges in the non-angiogenic subnetwork); the statistical significance (quantified as a p-value) of the rewiring is determined by using the hypergeometric distribution model to evaluate the overlap between edges from a transcription factor and edges specific to a particular subnetwork.

On average, a given transcription factor will be associated with around 114 edges in the subtype-specific subnetworks. However some transcription factors are associated with a relatively small number (less than twenty) of edges. These transcription factors were

excluded in further analysis to enhance statistical robustness. We identified ten "key" transcription factors with an edge enrichment greater than 1.5 (or less than 1/1.5) and with $p < 10^{-3}$ (Figure 2A). The identified transcription factors all have established associations with angiogenesis or survival ([32-43], Figure 2B). For example NFKB1 is important for chromatin remodeling during angiogenesis [32], PRRX1 deletion causes vascular anomalies [38], and MZF1 can repress MMP2 [41], a key prognostic factor in ovarian cancer [44].

We tested if these ten transcription factors would also be identified in a simple differential expression analysis using a *t*-test. Only three out of ten TFs are differentially expressed at the $p < 0.01$ cutoff: NFKB1 (more highly expressed in the angiogenic subtype, $p = 7.8 \times 10^{-5}$, FDR = 4.3×10^{-4}), PRRX2 (more highly expressed in the angiogenic subtype, $p = 0.005$, FDR = 0.015), and MZF1 (more highly expressed in the non-angiogenic subtype, $p = 4.1 \times 10^{-7}$, FDR = 4.2×10^{-6}). Thus, PANDA identified transcription factors that are not strongly differentially expressed between the subtypes yet are known to participate in angiogenic processes.

We also tested targets of each transcription factor for differential expression (see Methods) and found significant differences in target expression for six of the seven remaining TFs. For example, ARID3A is not differentially expressed between angiogenic and non-angiogenic
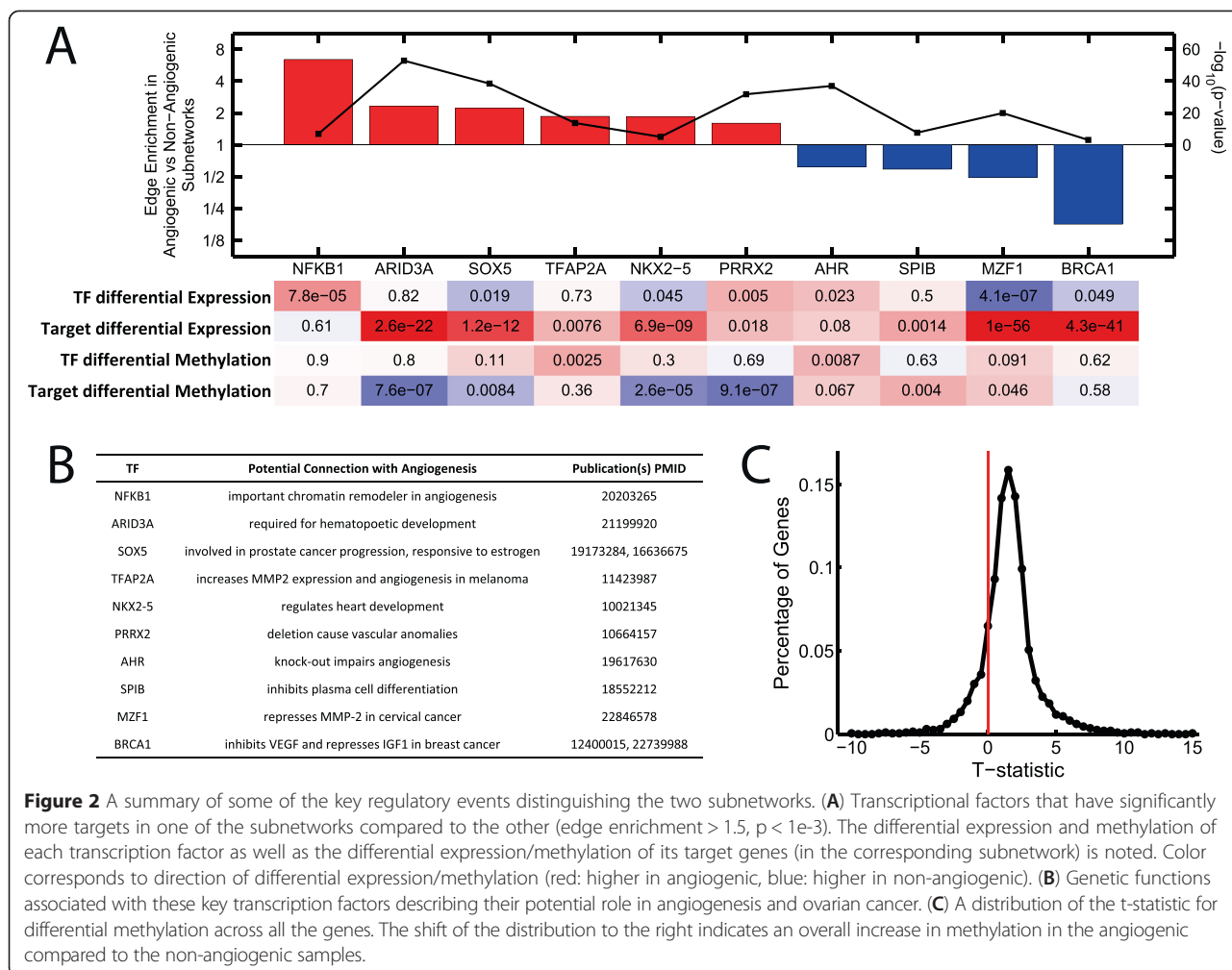
Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 4 of 17

Figure 2 A summary of some of the key regulatory events distinguishing the two subnetworks. (**A**) Transcriptional factors that have significantly more targets in one of the subnetworks compared to the other (edge enrichment > 1.5, p < 1e-3). The differential expression and methylation of each transcription factor as well as the differential expression/methylation of its target genes (in the corresponding subnetwork) is noted. Color corresponds to direction of differential expression/methylation (red: higher in angiogenic, blue: higher in non-angiogenic). (**B**) Genetic functions associated with these key transcription factors describing their potential role in angiogenesis and ovarian cancer. (**C**) A distribution of the t-statistic for differential methylation across all the genes. The shift of the distribution to the right indicates an overall increase in methylation in the angiogenic compared to the non-angiogenic samples.

subtypes, but its targets have significantly ($p = 2.6 \times 10^{-22}$) increased expression in the angiogenic subtype. This suggests that transcription factor activity changes may not be detectable based on their own expression level increases or decreases, but that the expression of their targets can provide information on how they influence phenotype.

There are many factors that could contribute to expression abnormalities in cancer, such as differences in mutations, copy-number variation and epigenetic states. We sought to determine which, if any, of these, might be contributing to the differential expression of the genes targeted by each of our ten key transcription factors. First, we investigated whether copy-number variation might explain the overall change in gene expression (see Methods). Although we find some nominally significant changes for the targets of PRRX2 (p = 0.0029) and ARID3A (p = 0.0257), these genes actually have less overall amplification in the agiogenic subtype compared to the non-angiogenic subtype (t = −2.98 and t = −2.23 respectively). This is despite the fact that their mRNA expression

levels are higher in the angiogenic compared to non-angiogenic subtype. Thus copy-number variation does not appear to be the primary factor driving changes in expression of these target genes.

Epigenetic factors provide an alternative mechanism for differential targeting by these transcription factors. To explore this possibility we mapped DNA methylation data from TCGA to the samples and genes used in our network reconstruction. We used a *t*-test to quantify any potential change in DNA methylation for each gene between the angiogenic and non-angiogenic subtypes and show the distribution of these values in Figure 2C. Overall we found DNA methylation levels to be higher in angiogenic tumor samples (mean value of t-statistic across all genes is 1.52). We next determined the differential methylation of the ten transcription factors and their targets in each of the subnetworks (Figure 2A). Compared to the rest of the genome, genes targeted by four of the ten transcription factors, ARID3A, SOX5, NKX2-5 and PRRX2, are associated with significantly lower

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 5 of 17

methylation in angiogenic samples. It should be noted that this lower level of methylation does not indicate hypomethylation. In fact, the average t-statistic value for the targets of these transcription factors is greater than zero (ARID3A, t = 1.17; SOX5, t = 1.31; NKX2-5, t = 0.70; PRRX2, t = 1.24), indicating that their targets, on average, have higher levels of methylation in the angiogenic compared to the non-angiogenic subtype; however, that increase in methylation is comparatively less than that experienced by genes not targeted by these regulators.

We note that the TCGA methylation data was not used by PANDA to construct the networks, yet is highly concordant with the predicted patterns of gene regulation. Although many of the transcription factors that alter targeting between the aniogenic and non-angiogenic subnetworks are not significantly differentially-expressed between the subnetworks, their targets genes often are. At the same time, these target genes are also differentially-methylated between the subtypes. Overall this analysis provides independent support of the overall network model.

Both transcriptional activation and repression are used to control angiogenic pathways

Transcription factors can either activate or repress gene expression. The target gene expression analysis in Figure 2A provides a preliminary indication about potential regulatory roles for the identified transcription factors. For example, although MZF1 and BRCA1 exhibit an edge-enrichment in the non-angiogenic subnetwork and are themselves also more highly expressed in the non-angiogenic samples, their *targets* show the opposite trend, with significantly higher expression in the angiogenic samples (p = 1.0×10^{-56} and 4.3×10^{-41}, respectively). There are two scenarios consistent with these observations: (1) loss of control by these transcription factors results in the increased expression of their former targets, and (2) increased control by these transcription factors results in the decreased expression of their target genes. Interestingly, BRCA1 is known to negatively regulate IGF1 expression in breast cancer cells [43], which could inhibit angiogenesis as multiple studies have shown that increased levels of IGF1 in cancer calls leads to an increase in cell proliferation [45-47]. Similarly, MZF1 is a repressor of MMP2 [41] and is known to inhibit hematopoietic lineage differentiation in embryonic stem cells [48]. Combined with the analysis presented in Figure 2A, this suggests that MZF1 and BRCA1 act as transcriptional repressors in the non-angiogenic samples.

Motivated by these observations, we classified the edges in our two subnetworks as either "activating" or "repressing" based on whether changes in the target gene's expression is correlated or anti-correlated with
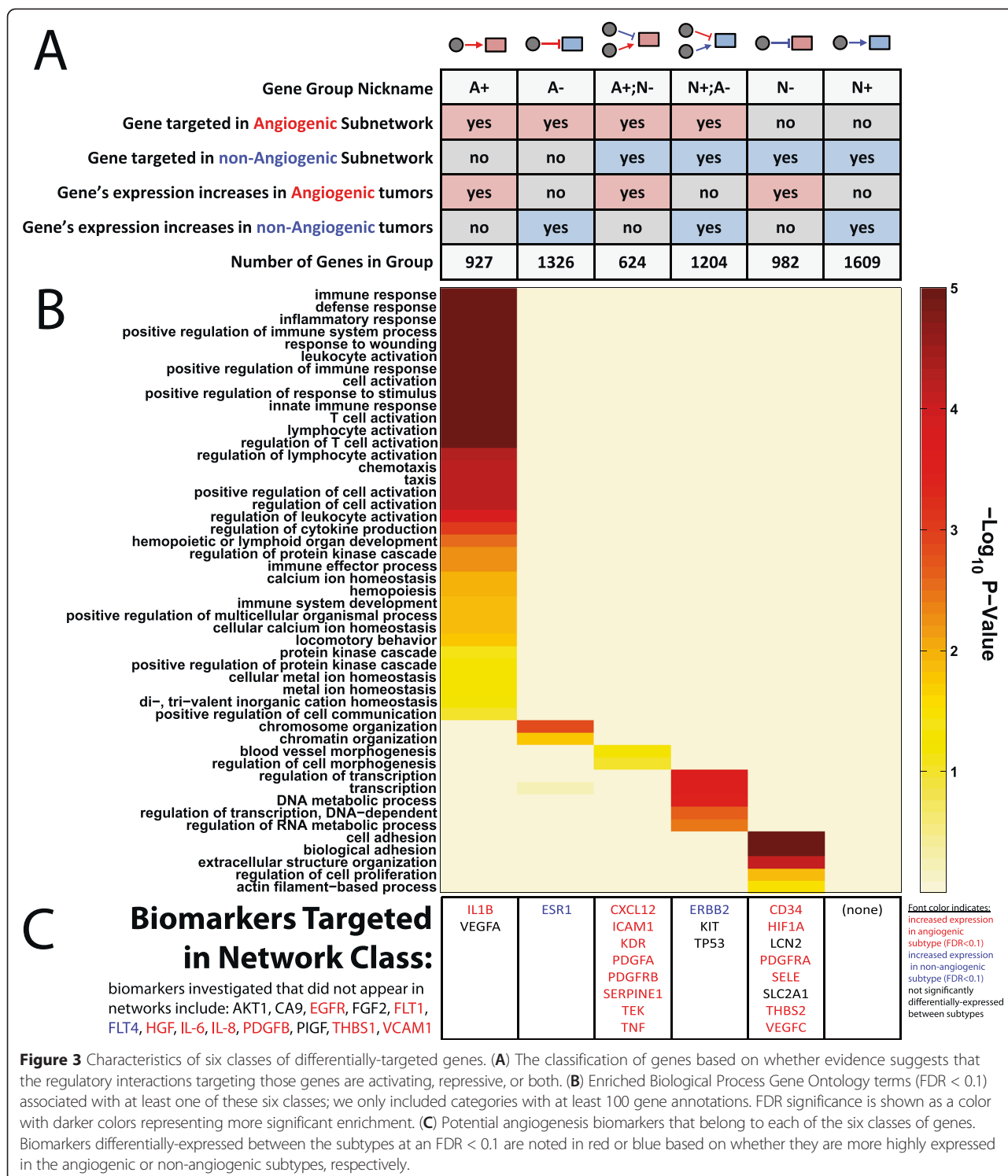
subnetwork assignment. We then assigned each target gene to one of six non-overlapping classes (see Figure 3A):

1) "A+": genes targeted only in the angiogenic subnetwork that are more highly expressed in the angiogenic subtype;
2) "A-": genes targeted only in the angiogenic subnetwork that are more highly expressed in the non-angiogenic subtype;
3) "A+;N-": genes targeted in both subnetworks and more highly expressed in the angiogenic subtype;
4) "N+;A-": genes targeted in both subnetworks and more highly expressed in the non-angiogenic subtype;
5) "N-": genes targeted only in the non-angiogenic subnetwork that are more highly expressed in the angiogenic subtype;
6) "N+": genes targeted only in the non-angiogenic subnetwork that are more highly expressed in the non-angiogenic subtype.

We used DAVID [49] to test for functional enrichment in these six classes of target genes, with the 12290 network genes taken as a background. The FDR p-values for GO "Biological Process" categories with more than one hundred members that have an FDR enrichment of less than 0.1 in at least one of our six classes of genes are illustrated as a heat map in Figure 3B.

The angiogenic-activated class ("A+") has the greatest number of significantly enriched functional categories. Many of these are associated with immune-response; processes associated with angiogenesis are also included, for example "chemotaxis," "hematopoiesis," "positive regulation of cell communication" and "metal ion homeostasis." Some processes found to be enriched for genes repressed in the non-angiogenic subnetwork ("N-" genes), such as "cell adhesion" and "extracellular structure organization," also play a role in angiogenesis. In addition, genes activated in the angiogenic subnetwork but repressed in the non-angiogenic subnetwork ("A+;N-") include those involved in "blood vessel morphogenesis." This suggests angiogenesis involves not only the activation of certain genes in the angiogenic subtype, but also removal of repressive regulatory interactions from the non-angiogenic subtype. In contrast, genes repressed in the angiogenic subnetwork ("A-" genes) are associated with "chromatin organization," consistent with the observed role that epigenetics plays in distinguishing the subtypes (see Figure 2C). Genes activated in the non-angiogenic subnetwork but repressed in the angiogenic-subnetwork ("N+;A-" genes) are involved in functions such as "transcription" and "DNA metabolic process".

We also investigated whether previously identified potential biomarkers for angiogenesis were targeted in our

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 6 of 17

Figure 3 Characteristics of six classes of differentially-targeted genes. (**A**) The classification of genes based on whether evidence suggests that the regulatory interactions targeting those genes are activating, repressive, or both. (**B**) Enriched Biological Process Gene Ontology terms (FDR < 0.1) associated with at least one of these six classes; we only included categories with at least 100 gene annotations. FDR significance is shown as a color with darker colors representing more significant enrichment. (**C**) Potential angiogenesis biomarkers that belong to each of the six classes of genes. Biomarkers differentially-expressed between the subtypes at an FDR < 0.1 are noted in red or blue based on whether they are more highly expressed in the angiogenic or non-angiogenic subtypes, respectively.

networks, and, if so, which "class" of genes those biomarkers belonged to. In particular, we investigated thirty-five biomarkers, described in [50,51], and found twenty-two targeted in our defined subnetworks (Figure 3C). The majority (eighteen) of these biomarkers are targeted

in either the "A+", "A+;N-"or "N-"class, consistent with higher expression in the angiogenic subtype. Interestingly, many of these biomarkers are targeted in the *non-angiogenic* subnetwork ("A+;N-"or "N-"classes). One possible interpretation of these results is that *repressive* regulatory

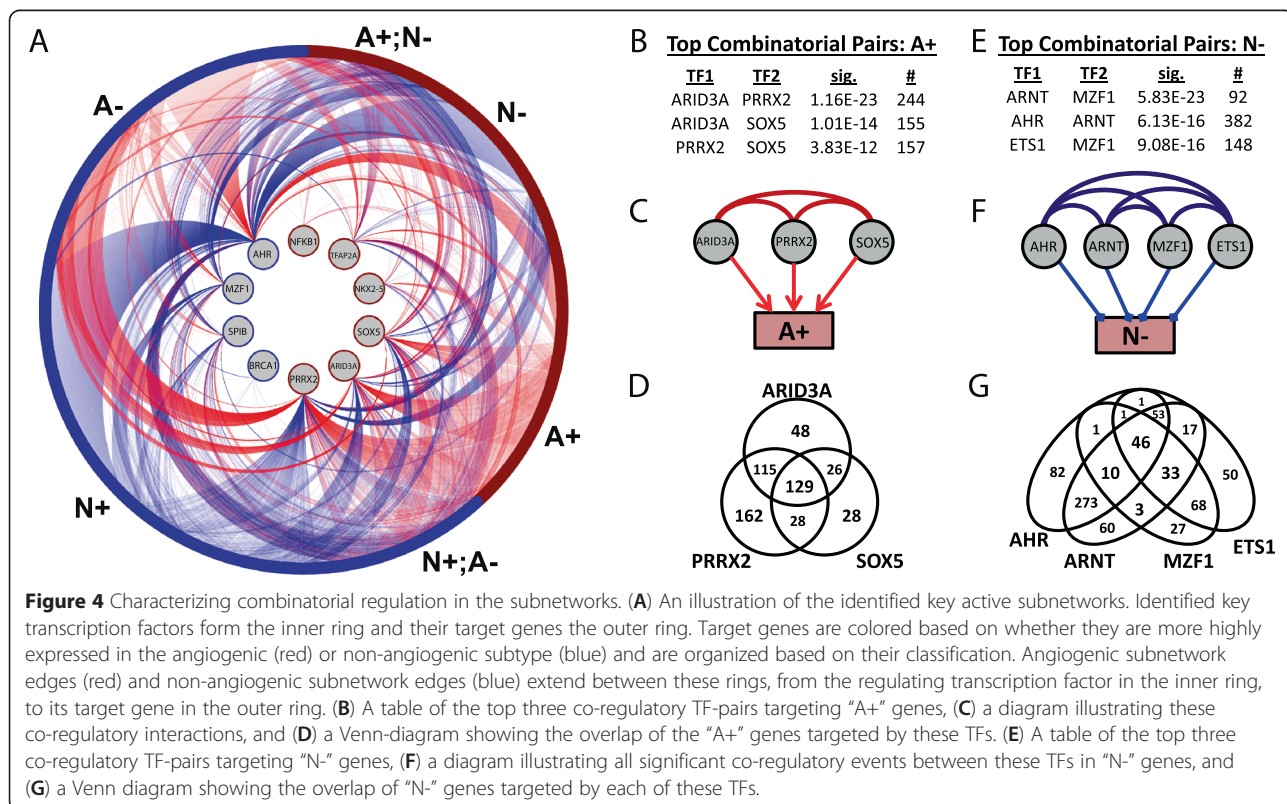Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 7 of 17

features play a role in inhibiting angiogenic progression, in addition to transcriptional activation of these bio-markers in angiogenic tumors. Curiously, three of the four biomarkers not included in the pro-angiogenic net-work classes were identified in only a single study [52] that included twenty patients with inflammatory breast cancer [50]. We note that while many of the network-targeted biomarkers are also significantly differentially-expressed between the subtypes (FDR < 0.1 based on un-paired *t*-test), several are not, including VEGFA (FDR = 0.69), TP53 (FDR = 0.12), LCN2 (FDR = 0.68), KIT (FDR = 0.73) and SLC2A1 (FDR = 0.10). The identi-fication of VEGFA and other biomarkers in our network model despite clear lack of differential-expression may indicate that our network model is able to identify im-portant cellular regulatory alterations even in the ab-sence of distinct changes in downstream target gene expression.

Combinatorial control plays a critical role in potentiating angiogenesis

Regulatory information that pertains to the core of our network can be depicted using a ring diagram represent-ing the union of the two subnetworks (Figure 4A). In this visualization, our ten key regulators form the inner ring, while their targets, colored based on whether they

exhibit higher average expression in the angiogenic (red) or non-angiogenic (blue) samples, form the outer ring. Viewing the two subnetworks, it is clear that there is a high degree of combinatorial gene regulation. To quantify this, we applied the hypergeometric distribution model and, using the union of the genes targeted in both the angiogeneic and non-angiogenic networks as a back-ground, tested for over-representation of genes co-targeted by specific pairs of transcription factors in the various network classifications (either "A+," "A-," "N+" and "N-"). Here, we focus on the three most significant pairs that include at least one of the ten identified key transcription factors. Information for all pairs can be found in the Supplemental Material (Additional file 1: Dataset S1). Note that the genes in the "A-" and "N+" classes had no combinatorial pairs significantly enriched (using a p = 10^{-3} cutoff).

For "A+" genes, as was seen in Figure 4A, we identify significant co-regulatory associations between ARID3A, PRRX2, and SOX5 (Figure 4B-C) and these three regu-lators share many common targets (Figure 4D). In fact, 58% of the "A+" genes are targeted by at least one of these transcription factors, 32% by at least two, and 14% are targeted by all three, suggesting they may function as a module that coordinately regulates these genes.
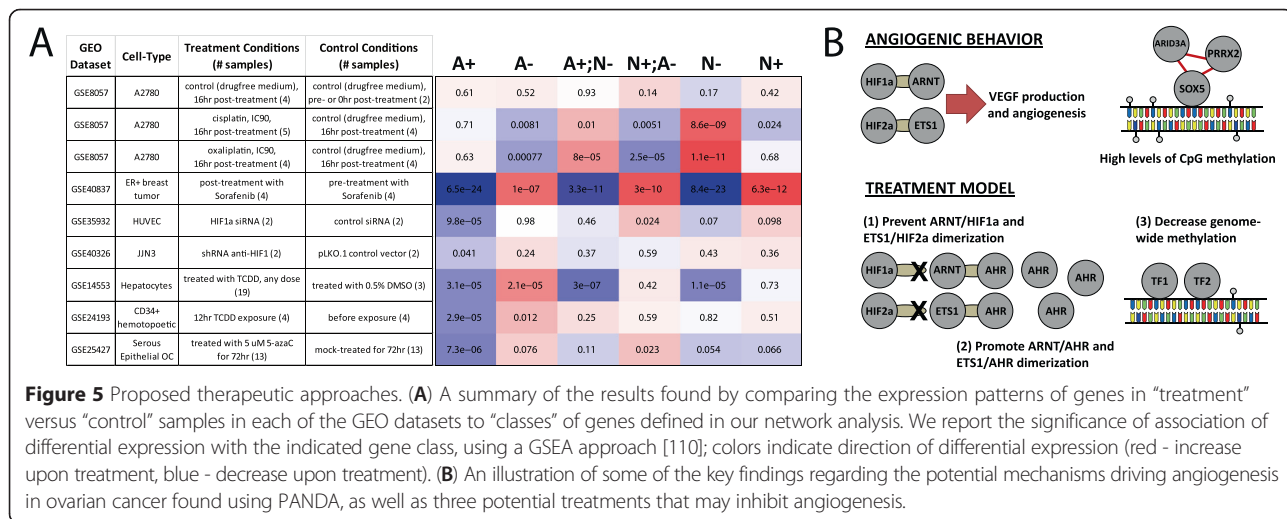

Figure 4 Characterizing combinatorial regulation in the subnetworks. (**A**) An illustration of the identified key active subnetworks. Identified key transcription factors form the inner ring and their target genes the outer ring. Target genes are colored based on whether they are more highly expressed in the angiogenic (red) or non-angiogenic subtype (blue) and are organized based on their classification. Angiogenic subnetwork edges (red) and non-angiogenic subnetwork edges (blue) extend between these rings, from the regulating transcription factor in the inner ring, to its target gene in the outer ring. (**B**) A table of the top three co-regulatory TF-pairs targeting "A+" genes, (**C**) a diagram illustrating these co-regulatory interactions, and (**D**) a Venn-diagram showing the overlap of the "A+" genes targeted by these TFs. (**E**) A table of the top three co-regulatory TF-pairs targeting "N-" genes, (**F**) a diagram illustrating all significant co-regulatory events between these TFs in "N-" genes, and (**G**) a Venn diagram showing the overlap of "N-" genes targeted by each of these TFs.

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 8 of 17

For the "N-" genes, the top three significant co-regulatory transcription factor pairs include combinations of ETS1, ARNT (also known as HIF1β), MZF1 and AHR (Figure 4E). All possible pairs of these four TFs (including those that don't include one of our "key" transcription factors) are enriched in "N-" genes at a statistically significant level (p < 1 × 10^{-6}, Figure 4F), although MZF1 generally only shares targets of AHR or ARNT in combination with ETS1 (Figure 4G). AHR and MZF1 are among our key regulators, and, as noted previously, MZF1 is known to repress MMP2 and reduced cancer invasiveness [41]. However, ETS1 and ARNT were not among our list of key regulators, indicating that combinatorial events might be especially important for these two transcription factors. Previous reports suggest that although various ETS family members can either activate or repress angiogenic pathways [53,54], ETS1, in particular, acts as a mediator of angiogenesis [55], dimerizing with HIF2α to *activate* VEGFR1 and VEGFR2 [56,57]. Similarly, ARNT dimerizes with HIF1α to activate VEGF and angiogenesis [58]. However, the dimerization of *AHR* with ARNT *inhibits* ARNT/HIF1α dimerization, thereby reducing VEGF production and subsequent angiogenesis [59]. Thus, even though ARNT/HIF1α promotes angiogenesis, the fact that *ARNT/AHR* dimerization *inhibits* angiogenesis offers an explanation for our observation that ARNT is associated with the repression of genes. Since both ETS1/HIF2α and ARNT/HIF1α interactions occur through a PAS domain [60,61], it is likely that a similar mechanism underlies our observed combinatorial enrichment of ETS1 with AHR and we hypothesize that ETS1 interaction with AHR prevents dimerization with HIFα proteins, thereby reducing VEGF production and subsequent angiogenesis.

The network model captures the effects of various treatment strategies

We wished to investigate how genes identified using our network model might respond to standard or other treatment protocols. Therefore, we analyzed experimental data (GEO accession numbers GSE8057, GSE40837) measuring gene expression levels in response to several chemotherapy drugs that are commonly used to treat ovarian cancer patients and/or angiogenesis, including cisplatin, oxaliplatin and sorafenib. For each experiment, we used RMA [62] to normalize gene expression CEL files downloaded from the Gene Expression Omnibus and used a custom-CDF to map to Entrez GeneIDs [63]. We selected samples that correspond to either a treatment or control experiment and performed a *t*-test to quantify the differential expression of all genes between these sets of samples. Finally, we computed a summary statistic representing the aggregate differential expression value for the sets of genes within each of the six "classes" defined by our subnetworks (for more details, see Methods). The results are summarized in Figure 5A; intensity of red or blue coloration scales with the significance of increased or decreased expression, respectively, in the treatment compared to the control samples, for the genes belonging to each of our networks "classes".

Platinum-based therapies are widely used in ovarian cancer treatment regimens. Therefore, we began by investigating the effect of the chemotherapy drugs cisplatin and oxaliplatin on the expression levels of genes in A2780 human ovarian carcinoma cells (GSE8057, [64]). As a negative control, we compared expression levels of cells grown in a drug-free medium for 16 hours to their expression at 0 hours. As expected, there is little differential expression and we observe no association with any of our classes of genes (Figure 5A). We next

GEO Dataset	Cell-Type	Treatment Conditions (# samples)	Control Conditions (# samples)	A+	A-	A+;N-	N+;A-	N-	N+
GSE8057	A2780	control (drugfree medium), 16hr post-treatment (4)	control (drugfree medium), pre- or 0hr post-treatment (2)	0.61	0.52	0.93	0.14	0.17	0.42
GSE8057	A2780	cisplatin, IC90, 16hr post-treatment (5)	control (drugfree medium), 16hr post-treatment (4)	0.71	0.0081	0.01	0.0051	8.6e−09	0.024
GSE8057	A2780	oxaliplatin, IC90, 16hr post-treatment (4)	control (drugfree medium), 16hr post-treatment (4)	0.63	0.00077	8e−05	2.5e−05	1.1e−11	0.68
GSE40837	ER+ breast tumor	post-treatment with Sorafenib (4)	pre-treatment with Sorafenib (4)	6.5e−24	1e−07	3.3e−11	3e−10	8.4e−23	6.3e−12
GSE35932	HUVEC	HIF1a siRNA (2)	control siRNA (2)	9.8e−05	0.98	0.46	0.024	0.07	0.098
GSE40326	JJN3	shRNA anti-HIF1 (2)	pLKO.1 control vector (2)	0.041	0.24	0.37	0.59	0.43	0.36
GSE14553	Hepatocytes	treated with TCDD, any dose (19)	treated with 0.5% DMSO (3)	3.1e−05	2.1e−05	3e−07	0.42	1.1e−05	0.73
GSE24193	CD34+ hemotopoetic	12hr TCDD exposure (4)	before exposure (4)	2.9e−05	0.012	0.25	0.59	0.82	0.51
GSE25427	Serous Epithelial OC	treated with 5 uM 5-azaC for 72hr (13)	mock-treated for 72hr (13)	7.3e−06	0.076	0.11	0.023	0.054	0.066

Figure 5 Proposed therapeutic approaches. (**A**) A summary of the results found by comparing the expression patterns of genes in "treatment" versus "control" samples in each of the GEO datasets to "classes" of genes defined in our network analysis. We report the significance of association of differential expression with the indicated gene class, using a GSEA approach [110]; colors indicate direction of differential expression (red - increase upon treatment, blue - decrease upon treatment). (**B**) An illustration of some of the key findings regarding the potential mechanisms driving angiogenesis in ovarian cancer found using PANDA, as well as three potential treatments that may inhibit angiogenesis.

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 9 of 17

compared expression of cells grown for 16 hours following treatment with either cisplatin or oxaliplatin to those grown for 16 hours in drug-free medium. Curiously, in both instances we see that genes normally repressed in the non-angiogenic subnetwork ("N-") actually increase their expression levels following treatment, suggesting that these drugs disrupt regulatory interactions that are important for *repressing* angiogenic activities. This is consistent with the results of a previous study [65] showing that when taken in isolation, cisplatin is not effective for treating angiogenesis in ovarian cancer.

Although decreased expression in "A+" genes does not occur following platinum-based treatment, we reasoned that the effects of a VEGF-inhibiting drug should be reflected in our identified subnetworks. Biopsies of ER+ breast tumors have been collected from patients both prior to and following a clinical trial of sorafenib, and the genome-wide expression levels of genes were measured in these samples (GSE40837). Analysis of this data in the same manner as the platinum-based therapies shows a striking association with the subnetworks (Figure 5A). Genes in the "A+," "A+;N-" and "N-" groups all show a profound decrease in expression post-treatment while genes in the "A-," "A-;N+" and "N+" groups all increase their expression. Although this is breast rather than ovarian cancer, these results are exciting since they rely on patient samples collected from a clinical trial rather than cell-lines, illustrating a patient-level association of an angiogenic-inhibition drug among network-defined genes. This result also serves as a positive control on our network analysis. We note that the six classes of genes we define are not wholey independent of the gene expression data, which we used both to reconstruct the networks as well as to divide target genes into distinct classes. Consequently, we would expect similar results for analyses of other "classes" of genes whose differential-expression is associated with differences between the two subtypes. However, analyzing the networks has the potential to provide additional mechanistic insight into differences between the subtypes and identify other drugs not classically associated with angiogenesis.

Three treatments may synergistically inhibit angiogenic progression

The optimal angiogenesis-based treatment in ovarian cancer is still a matter of ongoing investigation [14]. Commonly-used anti-angiogenic drugs mainly target VEGF, a major contributor to angiogenesis. On the other hand, as described below and illustrated in Figure 5B, several mechanisms highlighted in our network analysis suggest alternate approaches for treatment that, although speculative, could utilize existing compounds to

control, or potentially reverse, angiogenesis in ovarian cancer. For each of these proposed treatments we identified highly-related compounds and ascertained if there is a verifiable effect on gene expression in either ovarian cancer or another human system.

ARNT and ETS1 dimerization with HIF1α and HIF2α, respectively, must be prevented

As noted above, the dimerization of ETS1 and ARNT with HIF1α and HIF2α, respectively, generally promotes angiogenesis, although AHR may be interfering to repress target gene expression in the non-angiogenic subtype. It is therefore essential that the dimerization of these HIF proteins with ARNT and ETS1 is inhibited. HIF2α dimerizes with ARNT through a PAS-B domain, located on the C-terminus of the ARNT protein. The structural basis for this dimerization has been solved [66,67] and a small molecule ligand has been identified that dimerizes with HIF2α, decreasing affinity of the ARNT/HIF2α heterodimer [60]. Similarly, a compound has been identified that dimerizes with HIF1α, decreasing the affinity of the ARNT/HIF1α heterodimer [68]. Using either or both of these compounds, we believe one could prevent or reverse angiogenic effects driven by ARNT/HIF1α and perhaps also those driven by ETS1/HIF2α dimerization. In lieu of a dimerization inhibitor, we investigated how siRNA depletion of HIF1α affects the expression levels of genes in our identified subnetworks. We observe that in two independent experiments [69,70], "A+" genes exhibit a decrease in expression upon HIFα depletion, as we would expect from our model (Figure 5A).

AHR dimerization with ARNT and ETS1 must be promoted

Preventing the dimerization of ARNT/HIF1α and ETS1/HIF2α may be insufficient; inhibition of angiogenesis is also contingent upon the dimerization of ARNT with AHR (and perhaps also ETS1/AHR). Consequently, we also suggest treatment with an AHR agonist, such as the selective AHR modular (SAhRM) 6-methyl-1,3,8-trichlorodibensofuran (6-MCDF), which has been shown to inhibit carcinogen-induced mammary tumor growth in rats [71]. TRAMP mice fed 6-MCDF in their diet had overall lower levels of serum VEFG and were five times less likely to have metastasis compared to mice on a control diet [72]. Although a known carcinogen, one of the most efficient AHR agonists is the environmental toxin 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD). TCCD has been shown to potentiate ARNT/AHR dimerization thereby inhibiting angiogenesis and preventing vascular remodeling in rat placenta [73]. Curiously, accidental exposure to TCCD was found to potentially decrease incidence of breast and endometrial cancer in a group of women [74]. In our network, "A+" genes show a decrease in expression

Glass et al. BMC Bioinformatics (2015) 16:115

Page 10 of 17

upon treatment with the AHR agonist TCDD in both hepatocytes [75] and CD43+ hematopoetic cells [76] (Figure 5A). The hepatocyte data also shows a significant decrease in expression for "A+;N-" genes and "N-" genes, and an increase in expression for "A-" genes.

Methylation levels across the entire angiogenic genome must be decreased

In patients whose tumors have already become angiogenic, epigenetic alterations may need to be considered. One hallmark of many cancers is alteration of DNA methylation and, indeed, we found higher genome-wide methylation levels in the angiogenic subtype (Figure 2C). Interestingly, SOX5, one of our "key" transcription factors that was also found to play an important combinatorial role with ARID3A and PRRX2, contains an HMG-box which binds the minor groove [77]. Since methylation modifications occupy the major groove of DNA [78], this implies that SOX5 might be mediating key regulatory processes in the angiogenic subtype in a methylation-independent manner. This suggests that it may be necessary to decrease methylation levels across the angiogenic genome, thereby increasing competition with SOX5 binding to gene promoters, and altering their subsequent expression. This could be achieved using, for example, DNA methyltransferase (DNMT) and histone deacetylase (HDAC) inhibitors, which have already shown potential to inhibit angiogenesis in other systems [79,80]; such treatments in ovarian cancer might yield similar results. A hypomethylating agent, such as 5-azacytedine, could also be used to alter the epigenetic landscape and control angiogenic progression. With this in mind, we investigated the expression of ovarian cancer cells both prior to and post treatment with 5-azacytedine [81]; we observe a decrease in expression of the "A+" genes, consistent with our hypotheses (Figure 5A).

Additional potential therapies associated with differentially-targeted genes

We have identified several treatment options that target specific biological mechanisms uncovered when contrasting our network models; however, these are the result of intensive literature mining to determine suitable candidate drugs. We used the connectivity map (CMAP) [82] to determine if a gene classification based on our network analysis could also be used to identify potential drugs for treating angiogenesis in ovarian cancer.

We first used genes assigned to the "A+" and "A-" classification (see Figure 3) to build a "network-signature" suitable for CMAP analysis. As expression information was included in our network model, we also built an "expression-signature" by selecting genes with the most significant changes in expression between in the angiogenic and the non-angiogenic subtypes (based on an unpaired

t-test). A cutoff of $p < 1.6e-6$ (FDR $< 1.45 \times 10^{-5}$) was used to select genes more highly expressed in the angiogenic subtype and $p < 3.05e-4$ (FDR $< 1.4 \times 10^{-3}$) to select genes more highly expressed in the non-angiogenic subtype. We used these criteria to select differentially-expressed genes so that the expression-signature and network-signature had equivalent dimensions (920 "up" genes and 1287 "down" genes in both cases). These two signatures share approximately 20% of their genes, with 225 in common in the "up" direction and 223 in common in the "down" direction. We used CMAP to identify drugs associated with each of these signatures. Comparing results from these two signatures will allow us to distinguish between drug candidates only identified in the network context, and those which would also be identified using a differential-expression analysis.

Table 1 lists drugs significantly associated with the network-signature classification ($p < 0.01$). Most of these drugs are also significantly associated with the expression-signature – not surprising given that these two signatures are non-independent; however, the ranking from the expression-signature is vastly different from that of the network-signature. One significant exception is Prestwick-675, or hippeastrine, which is ranked highly in both analyses. Hippeastrine is an amaryllidaceae alkaloid with potent anti-invasive properties [83] and anticancer activities in cell lines [84]; it is also believed to contribute to the reported anti-cancer activities of the Chinese herb *Lycoris aurea* [85].

One of the most striking results from this analysis is that several drugs are significantly associated with the network-signature but not the expression-signature. The antitussive pentoxyverine is an agonist of the sigma-1 receptor [86], which has been shown to contribute to the induction of cancer-specific apoptosis by interleukin-24, a known inhibitor of angiogenesis [87]. Dicoumarol is an anticoagulant that is often administered to cancer patients. Anticoagulants are believed to be able to interfere with tumor angiogenesis [88] and in clinical trials their overall association with improved patient-survival, while encouraging, may to be limited to a subset of cancers [89]. Dicoumarol, in particular, inhibits furin-like activity by blocking the processing of MMP1 [34] and has been shown to abolish the TNF-induced activation of NFKB1 [90], one of our identified "core" transcription factors.

Another interesting drug identified by the network-signature is harmine. Harmine was recently shown to suppress tumor growth by inhibiting angiogenic activities in endothelial cells [91] and to induce apoptosis by inhibiting the expression of MMP2 in gastric cancer [92]. Interestingly, in light of the network model we propose above, harman, a related alkaloid, stimulates AHR-dependent luciferase activity [93], and harmine is a

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 11 of 17

Table 1 The drugs significantly associated (FDR < 0.1) with the network-signature classification based on CMAP analysis

Drug name	CMAP rank		CMAP P-value		CMAP FDR	
	Network signature	Expression signature	Network signature	Expression signature	Network signature	Expression signature
Harmol	1	61	4.6E-04	1.1×10^{-3}	0.015	4.9×10^{-3}
Harmalol	2	67	4.9E-04	1.8×10^{-3}	0.015	7.1×10^{-3}
Brinzolamide	3	95	5.6E-04	5.8×10^{-3}	0.016	0.020
Edrophonium chloride	4	42	6.0E-04	3.6×10^{-4}	0.017	1.8×10^{-3}
Trimethadione	5	138	6.8E-04	0.035	0.018	0.101
Canavanine	6	39	7.0E-04	2.7×10^{-4}	0.018	1.5×10^{-3}
Oxamniquine	7	49	1.3E-03	6.0×10^{-4}	0.025	2.8×10^{-3}
Metolazone	8	59	1.3E-03	1.0×10^{-3}	0.025	4.4×10^{-3}
Etofylline	9	112	3.2E-03	0.015	0.048	0.048
Harmine**	**10**	**613**	**3.6E-03**	**1.000**	**0.051**	**1.000**
Tetracycline	11	108	4.0E-03	0.013	0.056	0.041
Sotalol	12	136	5.5E-03	0.033	0.069	0.096
5213008	13	55	6.5E-03	8.1×10^{-4}	0.080	3.7×10^{-3}
Prestwick-675	14	1	7.3E-03	6.0×10^{-6}	0.084	5.2×10^{-5}
Pirinixic acid	15	21	8.6E-03	2.4×10^{-5}	0.095	1.8×10^{-4}
Dicoumarol**	**16**	**184**	**8.8E-03**	**0.239**	**0.097**	**0.534**
Pentoxyverine**	**17**	**784**	**9.1E-03**	**1.000**	**0.098**	**1.000**
Adipiodone**	**18**	**689**	**0.010**	**1.000**	**0.100**	**1.000**

Drugs that are not significantly associated with the expression-signature are bolded and asterisked (**).

competing ligand with the well-known AHR-agonist TCDD [94]. The association of these alkaloids with the network-signature is strengthened by the fact that two of harmine's sister compounds, harmol and harmalol, were identified as the top compounds most significantly associated with the network-signature.

Conclusions

Although a wealth of cancer gene expression data has been generated over the last decade, most biological inference has been based on statistical tests at the level of individual genes (with very high rates of spurious associations) followed by functional meta-analysis using gene set enrichment. Our network analysis of angiogenic and non-angiogenic phenotypes in ovarian cancer led us not just to differential expression, but also to the underlying regulatory mechanisms associated with the differential activity of transcriptional programs. By associating differences in regulatory patterns with differences in gene expression we were able to define subsets of genes that are activated or repressed by their regulators. Then, by identifying and exploring relationships between a set of key transcriptional regulators, we were able to identify putative mechanisms by which they might be coordinately working together to activate, or repress, the expression of their target genes. Based on these observations we

propose three therapeutic strategies that may complement or replace currently-used anti-angiogenic treatments. While these proposed strategies remain speculative, and experimental validation will be critical in validating their efficacy, the strategies are supported by our analysis of experiments from independent, published gene expression datasets, where mechanisms closely relevant to those predicted by our models were tested. We anticipate that these treatments could be combined synergistically to better inhibit angiogenesis in ovarian cancer tumors and bypass resistance that develops with the use mono-therapeutic conventional angiogenesis inhibitor regimens.

Until recently no clinical trials have collected the expression data needed to validate the angiogenic subtype classification in ovarian cancer, and consequently there is currently no experimental data supporting increased angiogenesis in the angiogenic subtype. However, there is strong anecdotal evidence supporting the classification, including response rates consistent with what we would have predicted in clinical trials involving angiogenesis inhibitors [95-97]. The use of PANDA in comparing the subtypes led to an independent identification of angiogenic processes, among others. Importantly, PANDA itself does not rely on differential-expression, but rather characterizes the targeting of genes by

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 12 of 17

transcription factors. Indeed, we found that many of the transcription factors with the greatest differential-targeting between the subtype-specific subnetworks are not themselves strongly differentially-expressed between the two subtypes. However, subsequent analysis led us to identify potential therapies that could disrupt the processes that distinguish the subtypes, which are only coincidentally associated with angiogenesis.

The clinical impact of current anti-angiogneic therapies on the outcome of ovarian cancer and other cancers, although real, continues to be modest, despite early highly promising results in mouse models [98]. In seminal ovarian cancer clinical trials only a small amount of improvement in progression free survival (four months) was observed with bevacizumab treatment [99] and angiogenesis inhibitor resistance is frequently seen *de novo*, or over the course of treatment [100]. In addition, several attempts to define biomarkers of clinical response to anti-angiogenesis drugs, have failed to produce a singular strong or consistently predictive biomarker [50,51,101-104]. Such therapeutic and predictive limitations may also reflect our limited understanding of the specific underlying mechanisms driving angiogenic progression. Indeed, we observe that the activity of many previously identified potential biomarkers for angiogenesis may be modulated through complex regulatory features (see Figure 3C) that include an important role both for coordinated transcriptional activation and repression. For instance, our analysis revealed that VEGFA is targeted in the angiogenic network but not differentially expressed between the subtypes. Furthermore, markers such as HIF1 (a proangiogenic factor) and PDGFRA (a kinase contributing to angiogeneisis and frequently targeted by angiogenesis inhibitor drugs), while having increased expression in the angiogenic tumor subtype, were identified as targeted for repression in the non-angiogenic subnetwork. Thus, our proposed network and treatment models begin to address, in more depth, the complex regulatory mechanisms relevant to angiogenesis in ovarian cancer, laying the ground for a network-based subtype categorization that may allow better prediction as well as more rational therapeutic development. Importantly, the methods we use in our analyses are generalizable and could be applied to many other disease settings to suggest new therapeutic approaches.

The specific transcriptional programs activated in angiogenic ovarian cancer, and identified through our use of PANDA, underscore the complex nature of regulatory processes and point to specific interventions that may have an increased likelihood of success. While a great deal of work would be required to validate these drug candidates, and to test whether they are subtype specific, their identification and the plausibility of their specific mode of action suggest that the type of network analysis we performed can identify candidates not found through

the more widely-used gene-by-gene methods for expression analysis. In diseases such as ovarian cancer, where the outcome is poor and there are few viable drug candidates, network-based methods could represent a valuable addition to the existing repertoire of tools for analyzing genomic data.

Methods

Constructing regulatory networks with PANDA

PANDA [28,105] uses three inputs: a motif prior, a set of known protein-protein interactions, and expression data. To create angiogenic and non-angiogenic subtype-specific transcriptional regulatory networks, we ran PANDA twice using the same transcription factor motif prior and protein-protein interaction data, but with gene expression data unique to either the angiogenic or non-angiogenic ovarian cancer subtypes. In both runs the update parameter (α) was set equal to 0.25.

Expression data

Gene expression data were downloaded from TCGA (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp), normalized using fRMA, and individual ovarian cancer samples were assigned to either the angiogenic or non-angiogenic subtype, as described in [3]. Briefly, a Gaussian mixture model was fit to the distribution of these patient scores using Mclust R package version 3.4.10 [106] and the maximum posterior probability was used to classify each sample. Of the 510 samples, 188 were classified as angiogenic, and 322 were classified as non-angiogenic. An R-package containing the data used in this manuscript has been deposited at the URL: http://bcb.dfci.harvard.edu/ovariancancer/.

Motif data

To create our motif prior, we downloaded the position weight matrixes (PWM) of 130 core vertebrate transcription factor binding site motifs from the JASPAR database [107,108] as processed as described in [109]. Namely, to search for motif target candidates, the motif score of each candidate S was defined as motif score = log [P(S|M)/P(S|B)], where P(S|M) is the probability to observe sequence S given the motif M, and P(S|B) is the probability to observe sequence S given the genome background B. To define motif targets, we modeled the motif score distribution by randomly sampling the genome 10^6 times. Targets of motifs were then defined as those with a score at a significance level of $p < 10^{-5}$. We associated genes with these motif targets if that target fell within its promoter region ([−750, 250] base-pairs around a transcriptional start site). It is possible for a motif to correspond to multiple transcription factors; in these cases we included all corresponding transcription factors. This resulted in a transcription factor to target

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 13 of 17

gene mapping. From this mapping we excluded edges connected to transcription factors or genes for which we did not have expression data. This left us with a prior network from 111 transcription factors to 12290 genes.

Protein-protein interactions

Predicted human transcription factor interactions were obtained from [31]. We filtered these interactions to only include those between the 111 transcription factors in our motif prior and used these interactions in constructing the regulatory networks.

Network quality estimation

To evaluate the robustness of our predicted networks we performed a number of variations on the input data used in the reconstruction and determined how it might influence the resulting estimated edge weights. The analysis shown in Additional file 2: Figure S1-S3 demonstrates the predicted networks' robustness to jackknifing the prior edges in the motif data, the protein-protein interaction dataset used, and the samples used to estimate the two subtype networks. See the Additional file 2: Figure S1-S3 legends for more details.

Quantitative network comparison

PANDA estimates a probability that an edge exists in an individual network and reports that estimate in terms of z-score units. We wanted to identify potential regulatory interactions that best characterized each of the subtype-specific networks. Therefore, we selected edges based both on the probability that they are "supported" in the network inference, and on whether they are "different" between the subtypes. To determine the probability that an edge is "supported," we took the value of the inverse cumulative distribution function of a normal distribution to assign a probability value between zero and one for each edge (instead of a z-score). To determine the probability that an edge is "different" between the networks, we first subtracted the z-score weight values estimated by PANDA for the two networks and then determined the value of the inverse cumulative distribution for this difference. The product of these two probabilities represents the probability than an edge is both "supported" and "different." We select edges for which this combined probability is greater than 80%, or:

This 80% cutoff was chosen so that each subnetwork contains roughly 1% of all possible edges. We verified the robustness of our network analysis to this cutoff by varying it systematically between 65% and 95% (see Additional file 2: Figure S4). A file with the edge-enrichment analysis for TFs performed across each of these cutoffs is also supplied in the supplemental material (Additional file 3: Dataset S2).

We recognize that there could be hidden dependencies between the z-scores so this analysis may be overestimating the significance.

Edge enrichment

To identify key transcription factors we calculated two values, an "edge enrichment" score as well as a p-value significance for the difference in the number of target genes. The edge enrichment for a given transcription factor (E_{TF}) can be formulaically defined as: $E_{TF} = (k_A/k_N)/(n_A/n_N)$, where k_A and k_N are the out-degree of the TF in the angiogenic and non-angiogenic subnetwork, respectively, and n_A and n_N are the total number of edges in the angiogenic and non-angiogenic subnetworks, respectively.

The p-value significance in the overlap between edges from a transcription factor and edges specific to the angiogenic subnetwork, as modeled by the hypergeometric distribution, can then be defined as:

$$p_{TF}^{(A)} = \sum_{k_A}^{n_A} \frac{\binom{n_A}{k_A}\binom{n_N}{k_N}}{\binom{n_A + n_N}{k_A + k_N}}$$

An equivalent formula is used to calculate $p_{TF}^{(N)}$. For simplicity we report $p_{TF}^{(A)}$ for $E_{TF} > 1$ and $p_{TF}^{(N)}$ for $E_{TF} < 1$.

In selecting "key" transcription factors we used an edge enrichment of greater than 1.5 (or less than 1/1.5), and a p-value significance less than 10^{-3}. To help account for the fact that these measures are not highly robust for small out-degree values, we also only limited our analysis to transcription factors with twenty or more total edges ($k_A + k_N \geq 20$). This last threshold resulted in excluding one potential TF, ELK4, from being identified as a "key" transcription factor. For a full list of every TF's edge-enrichment, p-value significance and total edge count across a variety of potential subnetwork

$$\text{Edge identified as} \begin{cases} angiogenic-specific & CDF^{-1}\left(W_{ij}^{(A)}\right) \times CDF^{-1}\left(W_{ij}^{(A)} - W_{ij}^{(N)}\right) > 0.8 \\ non-angiogenic-specific & CDF^{-1}\left(W_{ij}^{(N)}\right) \times CDF^{-1}\left(W_{ij}^{(N)} - W_{ij}^{(A)}\right) > 0.8 \\ neither & otherwise \end{cases}$$

Glass et al. BMC Bioinformatics (2015) 16:115

Page 14 of 17

definitions, see supplemental material (Additional file 3: Dataset S2).

Characterizing differential expression/methylation of transcription factors and their target genes

We determined the differential expression between the subtypes for each gene in our network by using the *t*-test. We determined the corresponding significance and adjusted for multiple-hypotheses by applying the Benjamini-Hochberg correction. The differential expression patterns of a set of target genes was determined by comparing the values of the t-statistic for that set of genes to the values of the t-statistic for all other genes [110]. For the same 510 samples for which we have expression data, we also downloaded level-3 methylation data from the TCGA website (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) Of the 14473 genes with methylation data, we limited ourselves to the 10108 included in the expression data, 621 of which had empty values reported across all 510 patient samples analyzed. For the remaining 9487 genes, we compared methylation levels between the two subtypes using a *t*-test. We further tested for the differential methylation patterns for sets of target genes as we did for the gene expression data. Values for the differential-expression and differential methylation of each gene are included in Additional file 4: Dataset S3.

For sets of target genes we also performed a randomization procedure to ensure that the results observed in the above analysis is not coincidental (Additional file 2: Figure S5). See the supplemental figure legend for more details.

Characterizing differential CNV for target genes

To evaluate changes in copy-number for sets of target genes, we downloaded level 3 CNV (SNP Array) data files from TGCA. According to the TCGA documentation, these files contain the results of CBS segmentation of the log R ratio data for each tumor/normal pair. We identified all the segments in which each gene occurs and used a *t*-test to compare the values of the segments identified within subjects classified into the angiogenic subtype to the segments identified within subjects classified into the non-angiogenic subtype. To determine the overall differential-CNV for a set of target genes, we compared the resulting t-statistic value for the set of genes targeted by a particular transcription factor to all other genes [110]. Values for the differential-CNV of each gene are included in Additional file 4: Dataset S3.

Characterizing the association of differential gene expression within classes of network genes

We wished to investigate how genes identified by our network model might respond to standard or other treatment protocols. Therefore, we analyzed publically-

available experimental data measuring gene expression levels in response to various stimuli. For each experiment, we RMA-normalized raw CEL data deposited on the Gene Expression Omnibus using a custom-CDF to map to Entrez GeneIDs [63], selected samples that correspond to either a treatment or control experiment, and performed a *t*-test to quantify the expression differences between the treatment and control samples. Finally, we computed a summary statistic representing the significance of the association of this differential expression with genes in each of the "classes" defined by our subnetworks. Specifically, we calculate a "meta"-t-statistic and associated p-value by comparing the differential-expression t-statistic values for genes in a given network class to the t-statistic values for all other genes [110].

We also performed a randomization procedure to ensure that significant results identified in the above analysis is not accidental, and would not be observed for random "classes" of genes (Additional file 2: Figure S5). See the legend for Additional file 2: Figure S5 for more details.

CMAP analysis

We downloaded the "raw" gene expression CEL files from the Connectivity Map website (http://www.broad institute.org/cmap/cel_file_chunks.jsp) and normalized these using fRMA [29]. This dataset contains the expression of approximately 12,000 genes before and after administration of 1,309 drugs in as many as 5 cell lines. We generated drug perturbation signatures by quantifying the differential gene expression, controlling for tissue type and batch effects using the following model:

$$G_i = \beta_{0,i}^d + \beta_{c,i}^d C^d + \beta_{t,i}^d T^d + \beta_{b,i}^d B^d, \forall i \in M$$

where variables are the same as those used for drug sensitivity signatures except for C^d, representing the concentration of drug d used to treat the cell lines, T^d, representing the tissue of the cell-line treated with drug d, and B^d, representing the batch of the array measuring the effect of drug d. The strength and significance of differential expression of gene i due to perturbation by drug d is given here by the term $\beta_{c,i}^d$ and its associated p-value (Student's *t*-test). We defined the gene signatures for drug perturbations based on estimates for the coefficients of $\beta_{c,i}^d$ and their associated p-values.

Code and materials for repeating the analysis in the paper

The PANDA implementation used to perform this analysis, data input files, output predicted networks, as well as a separate tool to perform edge-enrichment analysis on a pair of PANDA networks is available at http://sourceforge.net/projects/panda-net/.

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 15 of 17

Additional files

Additional file 1: Dataset S1. (TranscriptionFactor_Combinatorial Enrichment.txt): The significance of co-targeting by all pairs of transcription factors in the various network classifications.

Additional file 2: Figure S1. Analysis demonstrating that network estimates are robust to the exact expression samples used. **Figure S2.** Analysis examining the effect of the prior motif structure on network estimates. **Figure S3.** Evaluation of how different protein-interaction databases affect PANDA's predicted networks. **Figure S4.** Analysis examining the consequences of varying the threshold used to define the angiogenic and non-angiogenic subnetworks. **Figure S5.** Analysis examining whether the differential methylation and expression of target genes and gene classes might be observed by chance.

Additional file 3: Dataset S2. (TF_Statistics_Across_P-cutoffs.txt): A file containing the edge-enrichment and p-values for TF-differential-targeting between subnetworks identified across various P-cutoff values. The results included in this file were used to generate Additional file 2: Figure S4. The P-cutoff value of 0.8 was used to define the subnetworks used in the analysis shown in the main text (see Figure 2A).

Additional file 4: Dataset S3. (AllGenes_SubtypeInformation.txt): File containing various characteristics of the genes included in our network reconstruction. This includes (1) whether each gene was identified as a key TF or is a previously-identified biomarker for angiogenesis; (2) the network and expression classification of each gene (used for functional and CMAP analysis); and (3) the differential-expression, -methylation and –CNV of each gene between the subtypes.

Author details

[1]Dana-Farber Cancer Institute, Boston, MA, USA. [2]Harvard School of Public Health, Boston, MA, USA. [3]Brigham and Women's Hospital, Boston, MA, USA. [4]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [5]Princess Margaret Cancer Center, University Health Network, Toronto, ON M5G 2M9, Canada.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin. 2012;62(1):10–29.
2. Cannistra SA. Cancer of the ovary. N Engl J Med. 2004;351(24):2519–29.
3. Bentink S, Haibe-Kains B, Risch T, Fan JB, Hirsch MS, Holton K, et al. Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. PLoS One. 2012;7(2):e30269.
4. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res. 2008;14(16):5198–208.
5. TCGA. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15.
6. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57–70.
7. Folkman J. Tumor angiogenesis: therapeutic implications. N Engl J Med. 1971;285(21):1182–6.
8. Ingber DE, Folkman J. Mechanochemical switching between growth and differentiation during fibroblast growth factor-stimulated angiogenesis in vitro: role of extracellular matrix. J Cell Biol. 1989;109(1):317–30.
9. Dike LE, Chen CS, Mrksich M, Tien J, Whitesides GM, Ingber DE. Geometric control of switching between growth, apoptosis, and differentiation during angiogenesis using micropatterned substrates. In Vitro Cell Dev Biol Anim. 1999;35(8):441–8.
10. Moses MA, Wiederschain D, Loughlin KR, Zurakowski D, Lamb CC, Freeman MR. Increased incidence of matrix metalloproteinases in urine of cancer patients. Cancer Res. 1998;58(7):1395–9.
11. Veikkola T, Karkkainen M, Claesson-Welsh L, Alitalo K. Regulation of angiogenesis via vascular endothelial growth factor receptors. Cancer Res. 2000;60(2):203–12.
12. Korpelainen EI, Alitalo K. Signaling angiogenesis and lymphangiogenesis. Curr Opin Cell Biol. 1998;10(2):159–64.
13. Ferrara N. Role of vascular endothelial growth factor in the regulation of angiogenesis. Kidney Int. 1999;56(3):794–814.
14. Burger RA. Overview of anti-angiogenic agents in development for ovarian cancer. Gynecol Oncol. 2011;121(1):230–8.
15. Burger RA, Sill MW, Monk BJ, Greer BE, Sorosky JI. Phase II trial of bevacizumab in persistent or recurrent epithelial ovarian cancer or primary peritoneal cancer: a Gynecologic Oncology Group Study. J Clin Oncol. 2007;25(33):5165–71.
16. Cannistra SA, Matulonis UA, Penson RT, Hambleton J, Dupont J, Mackey H, et al. Phase II study of bevacizumab in patients with platinum-resistant ovarian cancer or peritoneal serous cancer. J Clin Oncol. 2007;25(33):5180–6.
17. Matulonis UA, Berlin S, Ivy P, Tyburski K, Krasner C, Zarwan C, et al. Cediranib, an oral inhibitor of vascular endothelial growth factor receptor kinases, is an active drug in recurrent epithelial ovarian, fallopian tube, and peritoneal cancer. J Clinical Oncol. 2009;27(33):5601–6.
18. Aghajanian C, Blank SV, Goff BA, Judson PL, Teneriello MG, Husain A, et al. OCEANS: a randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. J Clinical Oncol. 2012;30(17):2039–45.
19. Liu J, Matulonis UA. Anti-angiogenic agents in ovarian cancer: dawn of a new era? Curr Oncol Rep. 2011;13(6):450–8.
20. Emmert-Streib F. The chronic fatigue syndrome: a comparative pathway analysis. J Comput Biol. 2007;14(7):961–72.
21. Barabasi AL. Network medicine–from obesity to the "diseasome". N Engl J Med. 2007;357(4):404–7.
22. Silverman EK, Loscalzo J. Developing new drug treatments in the era of network medicine. Clin Pharmacol Ther. 2013;93(1):26–8.
23. Silverman EK, Loscalzo J. Network medicine approaches to the genetics of complex diseases. Discov Med. 2012;14(75):143–52.
24. Papin JA, Reed JL, Palsson BO. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. Trends Biochem Sci. 2004;29(12):641–7.
25. Nariai N, Tamada Y, Imoto S, Miyano S. Estimating gene regulatory networks and protein-protein interactions of Saccharomyces cerevisiae from multiple genome-wide data. Bioinformatics. 2005;21 Suppl 2:ii206–12.
26. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA. Combining location and expression data for principled discovery of genetic regulatory network models. Pac Symp Biocomput. 2002;7:437–49.
27. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models-a review. Biosystems. 2009;96(1):86–103.
28. Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing messages between networks to refine predicted interactions. PLoS One. 2013;8(5), e64832.
29. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics (Oxford, England). 2010;11(2):242–53.
30. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart – biological queries made easy. BMC Genomics. 2009;10(1):22.
31. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell. 2010;140(5):744–52.
32. Aurora AB, Biyashev D, Mirochnik Y, Zaichuk TA, Sanchez-Martinez C, Renault MA, et al. NF-kappaB balances vascular regression and angiogenesis via chromatin remodeling and NFAT displacement. Blood. 2010;116(3):475–84.

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 16 of 17

33. Webb CF, Bryant J, Popowski M, Allred L, Kim D, Harriss J, et al. The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. Mol Cell Biol. 2011;31(5):1041–53.

34. Komiyama T, Coppola JM, Larsen MJ, van Dort ME, Ross BD, Day R, et al. Inhibition of furin/proprotein convertase-catalyzed surface and intracellular processing by small molecules. J Biol Chem. 2009;284(23):15729–38.

35. Stevens TA, Meech R. BARX2 and estrogen receptor-alpha (ESR1) coordinately regulate the production of alternatively spliced ESR1 isoforms and control breast cancer cell growth and invasion. Oncogene. 2006;25(39):5426–35.

36. Gershenwald JE, Sumner W, Calderone T, Wang Z, Huang S, Bar-Eli M. Dominant-negative transcription factor AP-2 augments SB-2 melanoma tumor growth in vivo. Oncogene. 2001;20(26):3363–75.

37. Tanaka M, Chen Z, Bartunkova S, Yamasaki N, Izumo S. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. Development. 1999;126(6):1269–80.

38. Bergwerff M, Gittenberger-de Groot AC, Wisse LJ, DeRuiter MC, Wessels A, Martin JF, et al. Loss of function of the Prx1 and Prx2 homeobox genes alters architecture of the great elastic arteries and ductus arteriosus. Virchows Archiv. 2000;436(1):12–9.

39. Roman AC, Carvajal-Gonzalez JM, Rico-Leo EM, Fernandez-Salguero PM. Dioxin receptor deficiency impairs angiogenesis by a mechanism involving VEGF-A depletion in the endothelium and transforming growth factor-beta overexpression in the stroma. J Biol Chem. 2009;284(37):25135–48.

40. Schmidlin H, Diehl SA, Nagasawa M, Scheeren FA, Schotte R, Uittenbogaart CH, et al. Spi-B inhibits human plasma cell differentiation by repressing BLIMP1 and XBP-1 expression. Blood. 2008;112(5):1804–12.

41. Tsai SJ, Hwang JM, Hsieh SC, Ying TH, Hsieh YH. Overexpression of myeloid zinc finger 1 suppresses matrix metalloproteinase-2 expression and reduces invasiveness of SiHa human cervical cancer cells. Biochem Biophys Res Commun. 2012;425(2):462–7.

42. Kawai H, Li H, Chun P, Avraham S, Avraham HK. Direct interaction between BRCA1 and the estrogen receptor regulates vascular endothelial growth factor (VEGF) transcription and secretion in breast cancer cells. Oncogene. 2002;21(50):7730–9.

43. Kang HJ, Yi YW, Kim HJ, Hong YB, Seong YS, Bae I. BRCA1 negatively regulates IGF-1 expression through an estrogen-responsive element-like site. Cell Death Dis. 2012;3:e336.

44. Huang KJ, Sui LH. The relevance and role of vascular endothelial growth factor C, matrix metalloproteinase-2 and E-cadherin in epithelial ovarian cancer. Med Oncol. 2012;29(1):318–23.

45. Arnaldez FI, Helman LJ. Targeting the insulin growth factor receptor 1. Hematol Oncol Clin North Am. 2012;26(3):527–42. vii-viii.

46. Yu H, Rohan T. Role of the insulin-like growth factor family in cancer development and progression. J Natl Cancer Inst. 2000;92(18):1472–89.

47. Davison Z, de Blacquiere GE, Westley BR, May FE. Insulin-like growth factor-dependent proliferation and survival of triple-negative breast cancer cells: implications for therapy. Neoplasia. 2011;13(6):504–15.

48. Perrotti D, Melotti P, Skorski T, Casella I, Peschle C, Calabretta B. Overexpression of the zinc finger protein MZF1 inhibits hematopoietic development from embryonic stem cells: correlation with negative regulation of CD34 and c-myb promoter activity. Mol Cell Biol. 1995;15(11):6075–87.

49. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

50. Gerger A, LaBonte M, Lenz HJ. Molecular predictors of response to antiangiogenesis therapies. Cancer J. 2011;17(2):134–41.

51. Murukesh N, Dive C, Jayson GC. Biomarkers of angiogenesis and their role in the development of VEGF inhibitors. Br J Cancer. 2010;102(1):8–18.

52. Yang SX, Steinberg SM, Nguyen D, Wu TD, Modrusan Z, Swain SM. Gene expression profile and angiogenic marker correlates with response to neoadjuvant bevacizumab followed by bevacizumab plus chemotherapy in breast cancer. Clin Cancer Res. 2008;14(18):5893–9.

53. Lelievre E, Lionneton F, Soncin F, Vandenbunder B. The Ets family contains transcriptional activators and repressors involved in angiogenesis. Int J Biochem Cell Biol. 2001;33(4):391–407.

54. Sato Y. Role of ETS family transcription factors in vascular development and angiogenesis. Cell Struct Funct. 2001;26(1):19–24.

55. Khatun S, Fujimoto J, Toyoki H, Tamaya T. Clinical implications of expression of ETS-1 in relation to angiogenesis in ovarian cancers. Cancer Sci. 2003;94(9):769–73.

56. Elvert G, Kappel A, Heidenreich R, Englmeier U, Lanz S, Acker T, et al. Cooperative interaction of hypoxia-inducible factor-2alpha (HIF-2alpha) and Ets-1 in the transcriptional activation of vascular endothelial growth factor receptor-2 (Flk-1). J Biol Chem. 2003;278(9):7520–30.

57. Dutta D, Ray S, Vivian JL, Paul S. Activation of the VEGFR1 chromatin domain: an angiogenic signal-ETS1/HIF-2alpha regulatory axis. J Biol Chem. 2008;283(37):25404–13.

58. Beischlag TV, Taylor RT, Rose DW, Yoon D, Chen Y, Lee WH, et al. Recruitment of thyroid hormone receptor/retinoblastoma-interacting protein 230 by the aryl hydrocarbon receptor nuclear translocator is required for the transcriptional response to both dioxin and hypoxia. J Biol Chem. 2004;279(52):54620–8.

59. Fritz WA, Lin TM, Peterson RE. The aryl hydrocarbon receptor (AhR) inhibits vanadate-induced vascular endothelial growth factor (VEGF) production in TRAMP prostates. Carcinogenesis. 2008;29(5):1077–82.

60. Scheuermann TH, Tomchick DR, Machius M, Guo Y, Bruick RK, Gardner KH. Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor. Proc Natl Acad Sci U S A. 2009;106(2):450–5.

61. Hao N, Whitelaw ML, Shearwin KE, Dodd IB, Chapman-Smith A. Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR. Nucleic Acids Res. 2011;39(9):3695–709.

62. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64.

63. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005;33(20):e175.

64. Brun YF, Varma R, Hector SM, Pendyala L, Tummala R, Greco WR. Simultaneous modeling of concentration-effect and time-course patterns in gene expression data from microarrays. Cancer Genomics Proteomics. 2008;5(1):43–53.

65. Li D, Williams JI, Pietras RJ. Squalamine and cisplatin block angiogenesis and growth of human ovarian cancer cells with or without HER-2 gene overexpression. Oncogene. 2002;21(18):2805–14.

66. Card PB, Erbel PJ, Gardner KH. Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. J Mol Biol. 2005;353(3):664–77.

67. 1x0o: http://www.rcsb.org/pdb/explore.do?structureId=1x0o.

68. Cardoso R, Love R, Nilsson CL, Bergqvist S, Nowlin D, Yan J, et al. Identification of Cys255 in HIF-1alpha as a novel site for development of covalent inhibitors of HIF-1alpha/ARNT PasB domain protein-protein interaction. Protein Sci. 2012;21(12):1885–96.

69. Mimura I, Nangaku M, Kanki Y, Tsutsumi S, Inoue T, Kohro T, et al. Dynamic change of chromatin conformation in response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative interaction of hypoxia-inducible factor 1 and KDM3A. Mol Cell Biol. 2012;32(15):3018–32.

70. Storti P, Bolzoni M, Donofrio G, Airoldi I, Guasco D, Toscani D, et al. Hypoxia-inducible factor (HIF)-1alpha suppression in myeloma cells blocks tumoral growth in vivo inhibiting angiogenesis and bone destruction. Leukemia. 2013;27(8):1697–706.

71. Safe S. Molecular biology of the Ah receptor and its role in carcinogenesis. Toxicol Lett. 2001;120(1–3):1–7.

72. Fritz WA, Lin TM, Safe S, Moore RW, Peterson RE. The selective aryl hydrocarbon receptor modulator 6-methyl-1,3,8-trichlorodibenzofuran inhibits prostate tumor metastasis in TRAMP mice. Biochem Pharmacol. 2009;77(7):1151–60.

73. Ishimura R, Kawakami T, Ohsako S, Tohyama C. Dioxin-induced toxicity on vascular remodeling of the placenta. Biochem Pharmacol. 2009;77(4):660–9.

74. Bertazzi A, Pesatori AC, Consonni D, Tironi A, Landi MT, Zocchetti C. Cancer incidence in a population accidentally exposed to 2,3,7,8-tetrachlorodibenzo-para-dioxin. Epidemiology. 1993;4(5):398–406.

75. Carlson EA, McCulloch C, Koganti A, Goodwin SB, Sutter TR, Silkworth JB. Divergent transcriptomic responses to aryl hydrocarbon receptor agonists between rat and human primary hepatocytes. Toxicol Sci. 2009;112(1):257–72.

76. Fracchiolla NS, Todoerti K, Bertazzi PA, Servida F, Corradini P, Carniti C, et al. Dioxin exposure of human CD34+ hemopoietic cells induces gene expression modulation that recapitulates its in vivo clinical and biological effects. Toxicology. 2011;283(1):18–23.

77. Privalov PL, Jelesarov I, Read CM, Dragan AI, Crane-Robinson C. The energetics of HMG box interactions with DNA: thermodynamics of the DNA binding of the HMG box from mouse sox-5. J Mol Biol. 1999;294(4):997–1013.

Glass *et al. BMC Bioinformatics* (2015) 16:115

Page 17 of 17

78. Prokhortchouk E, Defossez PA. The cell biology of DNA methylation in mammals. Biochim Biophys Acta. 2008;1783(11):2167–73.

79. Hellebrekers DM, Griffioen AW, van Engeland M. Dual targeting of epigenetic therapy in cancer. Biochim Biophys Acta. 2007;1775(1):76–91.

80. Yoo CB, Jones PA. Epigenetic therapy of cancer: past, present and future. Nat Rev Drug Discov. 2006;5(1):37–50.

81. Matsumura N, Huang Z, Mori S, Baba T, Fujii S, Konishi I, et al. Epigenetic suppression of the TGF-beta pathway revealed by transcriptome profiling in ovarian cancer. Genome Res. 2011;21(1):74–82.

82. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929–35.

83. Evidente A, Kireev AS, Jenkins AR, Romero AE, Steelant WF, Van Slambrouck S, et al. Biological evaluation of structurally diverse amaryllidaceae alkaloids and their synthetic derivatives: discovery of novel leads for anticancer drug design. Planta Med. 2009;75(5):501–7.

84. Antoun MD, Mendoza NT, Rios YR, Proctor GR, Wickramaratne DB, Pezzuto JM, et al. Cytotoxicity of Hymenocallis expansa alkaloids. J Nat Prod. 1993;56(8):1423–5.

85. Liao N, Ao M, Zhang P, Yu L. Extracts of Lycoris aurea induce apoptosis in murine sarcoma S180 cells. Molecules. 2012;17(4):3723–35.

86. Brown C, Fezoui M, Selig WM, Schwartz CE, Ellis JL. Antitussive activity of sigma-1 receptor agonists in the guinea-pig. Br J Pharmacol. 2004;141(2):233–40.

87. Do W, Herrera C, Mighty J, Shumskaya M, Redenti SM, Sauane M. Sigma 1 Receptor plays a prominent role in IL-24-induced cancer-specific apoptosis. Biochem Biophys Res Commun. 2013;439(2):215–20.

88. Bobek V, Kovarik J. Antitumor and antimetastatic effect of warfarin and heparins. Biomed Pharmacother. 2004;58(4):213–9.

89. Gerotziafas GT, Papageorgiou C, Hatmi M, Samama MM, Elalamy I. Clinical studies with anticoagulants to improve survival in cancer patients. Pathophysiol Haemost Thromb. 2008;36(3–4):204–11.

90. Ahn KS, Sethi G, Jain AK, Jaiswal AK, Aggarwal BB. Genetic deletion of NAD (P)H:quinone oxidoreductase 1 abrogates activation of nuclear factor-kappaB, IkappaBalpha kinase, c-Jun N-terminal kinase, Akt, p38, and p44/42 mitogen-activated protein kinases and potentiates apoptosis. J Biol Chem. 2006;281(29):19798–808.

91. Dai F, Chen Y, Song Y, Huang L, Zhai D, Dong Y, et al. A natural small molecule harmine inhibits angiogenesis and suppresses tumour growth through activation of p53 in endothelial cells. PLoS One. 2012;7(12):e52162.

92. Zhang H, Sun K, Ding J, Xu H, Zhu L, Zhang K, et al. Harmine induces apoptosis and inhibits tumor cell proliferation, migration and invasion through down-regulation of cyclooxygenase-2 expression in gastric cancer. Phytomedicine. 2014;21(3):348–55.

93. El Gendy MA, El-Kadi AO. Harman induces CYP1A1 enzyme through an aryl hydrocarbon receptor mechanism. Toxicol Appl Pharmacol. 2010;249(1):55–64.

94. El Gendy MA, Soshilov AA, Denison MS, El-Kadi AO. Transcriptional and posttranslational inhibition of dioxin-mediated induction of CYP1A1 by harmine and harmol. Toxicol Lett. 2012;208(1):51–61.

95. Kumaran GC, Jayson GC, Clamp AR. Antiangiogenic drugs in ovarian cancer. Br J Cancer. 2009;100(1):1–7.

96. Spannuth WA, Sood AK, Coleman RL. Angiogenesis as a strategic target for ovarian cancer therapy. Nat Clin Pract Oncol. 2008;5(4):194–204.

97. Teoh D, Secord AA. Antiangiogenic agents in combination with chemotherapy for the treatment of epithelial ovarian cancer. Int J Gynecol Cancer. 2012;22(3):348–59.

98. O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, et al. Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell. 1997;88(2):277–85.

99. Burger RA, Brady MF, Bookman MA, Fleming GF, Monk BJ, Huang H, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. N Engl J Med. 2011;365(26):2473–83.

100. Moreno Garcia V, Basu B, Molife LR, Kaye SB. Combining antiangiogenics to overcome resistance: rationale and clinical experience. Clin Cancer Res. 2012;18(14):3750–61.

101. Collinson F, Hutchinson M, Craven RA, Cairns DA, Zougman A, Wind TC, et al. Predicting response to bevacizumab in ovarian cancer: a panel of potential biomarkers informing treatment selection. Clin Cancer Res. 2013;19(18):5227–39.

102. Madsen CV, Steffensen KD, Olsen DA, Waldstrom M, Smerdel M, Adimi P, et al. Serial measurements of serum PDGF-AA, PDGF-BB, FGF2, and VEGF in multiresistant ovarian cancer patients treated with bevacizumab. J Ovarian Res. 2012;5(1):23.

103. Horowitz NS, Penson RT, Duda DG, di Tomaso E, Boucher Y, Ancukiewicz M, et al. Safety, Efficacy, and Biomarker Exploration in a Phase II Study of Bevacizumab, Oxaliplatin, and Gemcitabine in Recurrent Mullerian Carcinoma. Clin Ovarian Cancer Gynecol Malign. 2011;4(1):26–33.

104. Sessa C, Guibal A, Del Conte G, Ruegg C. Biomarkers of angiogenesis for the development of antiangiogenic therapies in oncology: tools or decorations? Nat Clin Pract Oncol. 2008;5(7):378–91.

105. Glass K: PANDA Implementation (http://sourceforge.net/projects/panda-net/). 2013.

106. Fraley C, Raftery AE. Model-Based Clustering, Discriminant Analysis, and Density Estimation. J Am Stat Assoc. 2002;97(458):611–31.

107. Stormo GD. DNA binding sites: representation and discovery. Bioinformatics. 2000;16(1):16–23.

108. Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet. 2004;5(4):276–87.

109. Xu J, Shao Z, Glass K, Bauer DE, Pinello L, Van Handel B, et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev Cell. 2012;23(4):796–811.

110. Oron AP, Jiang Z, Gentleman R. Gene set enrichment analysis using linear models and diagnostics. Bioinformatics. 2008;24(22):2586–91.