

A Minimal Connected Network of Transcription Factors Regulated in Human Tumors and Its Application to the Quest for Universal Cancer Biomarkers

Ahmed Essaghir*, Jean-Baptiste Demoulin

de Duve institute, Université Catholique de Louvain, Brussels, Belgium

Abstract

A universal cancer biomarker candidate for diagnosis is supposed to distinguish, within a broad range of tumors, between healthy and diseased patients. Recently published studies have explored the universal usefulness of some biomarkers in human tumors. In this study, we present an integrative approach to search for potential common cancer biomarkers. Using the TFactS web-tool with a catalogue of experimentally established gene regulations, we could predict transcription factors (TFs) regulated in 305 different human cancer cell lines covering a large panel of tumor types. We also identified chromosomal regions having significant copy number variation (CNV) in these cell lines. Within the scope of TFactS catalogue, 88 TFs whose activity status were explained by their gene expressions and CNVs were identified. Their minimal connected network (MCN) of protein-protein interactions forms a significant module within the human curated TF proteome. Functional analysis of the proteins included in this MCN revealed enrichment in cancer pathways as well as inflammation. The ten most central proteins in MCN are TFs that trans-regulate 157 known genes encoding secreted and transmembrane proteins. In publicly available collections of gene expression data from 8,525 patient tissues, 86 genes were differentially regulated in cancer compared to inflammatory diseases and controls. From TCGA cancer gene expression data sets, 50 genes were significantly associated to patient survival in at least one tumor type. Enrichment analysis shows that these genes mechanistically interact in common cancer pathways. Among these cancer biomarker candidates, TFRC, MET and VEGFA are commonly amplified genes in tumors and their encoded proteins stained positive in more than 80% of malignancies from public databases. They are linked to angiogenesis and hypoxia, which are common in cancer. They could be interesting for further investigations in cancer diagnostic strategies.

Citation: Essaghir A, Demoulin J-B (2012) A Minimal Connected Network of Transcription Factors Regulated in Human Tumors and Its Application to the Quest for Universal Cancer Biomarkers. PLoS ONE 7(6): e39666. doi:10.1371/journal.pone.0039666

Editor: Paolo Provero, University of Turin, Italy

Received: February 6, 2012; **Accepted:** May 25, 2012; **Published:** June 25, 2012

Copyright: © 2012 Essaghir, Demoulin. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by FSR Fellowship from Université Catholique de Louvain. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ahmed.essaghir@uclouvain.be

Introduction

Cancer is a multifactorial disease. Many cancer types and stages have been distinguished. This complexity makes the quest for “universal cancer biomarkers” a challenging task. However, many studies conducted separately on different cancer types have reported common genes with potential biomarker value in treatment or diagnosis [1].

On the basis of literature reviewing or by using high-throughput techniques some authors identified potential biomarkers common to several cancers and tried to develop strategies to identify them from patient biofluids either directly or indirectly. Among these markers, telomerase has been reported as being highly expressed in neoplasms [2]. A platform to capture circulating tumor cells from patient blood and measure their telomerase activity has been proposed as a cancer diagnostic tool [3]. In addition, extra-cellular cAMP-dependent protein kinase A (EC-PKA) has been reported to be a good marker for multiple cancers [4]. Auto-antibodies against EC-PKA measured by ELISA from patients sera have been found to be highly specific to cancer [5]. Follicle-stimulating hormone (FSH) receptor was also reported to be selectively expressed in a variety of tumors [6]. The same observations apply

also to a cytochrome P450 (CYP1B1) [7]. Epigenetic alterations, in addition, could have a diagnostic value in cancer. Indeed, some authors have pointed to cancer-specific DNA methylation patterns as a marker for malignant diseases [8]. They can be detected on cell-free circulating DNA in the blood [9]. Auto-antibodies against leukocyte antigen F (HLA-F) were also detected in patients with various cancer types compared to healthy individuals [10].

Cancer biomarker candidate genes could be identified from literature. Confidence weights can be associated to each gene using its citation frequency [11]. Although initially used to enumerate markers specific to each cancer type, these weighted lists can help selecting common biomarkers in cancer. However, more elaborated strategies have been used to identify common cancer biomarkers, including gene expression meta-analysis across different tumor types [12,13]. They can be associated with function and pathway annotation enrichment filters to select common biomarkers [14].

In this study, we have elaborated an integrative strategy to search for useful biomarkers common to cancer types. Our working hypothesis is based on the assumption that almost all the perturbations that lead to malignancy transformation of normal

cells, although complex and diverse, share common collaborative pathways [15]. In general, these pathways might end by activating and/or repressing some sets of genes. These genes are targets of transcription factors (TFs). Some of these TFs are redundantly modulated between different cell transforming events [16–22]. They could be seen as connections or cross-talk nodes of the cancer leading pathways [23–27]. Thus, there should be a set of minimal connected TFs commonly perturbed in tumors as they share modulated pathways [28]. This set of TFs could be considered as a bottleneck of cancer pathways. If common cancer biomarkers exist, they are more likely to be among the targets of these commonly regulated TFs [29]. In this study, we took advantage of TFactS, a tool that we recently developed to predict TF regulations from high throughput gene expression data [30].

Results

Identification of TFs Regulated in Cancer Cell Lines

Gene expression and SNP data were available for 305 cell lines, from which results were further analyzed. These cell lines represent a broad panel of cancer types covering 28 different histological sites.

We assumed that important TFs would be those for which gene expression and CNV could explain their activity status [31,32]. They could be identified using the regression model shown in Figure 1. To compute all the parameters needed for this model, we identified genes differentially regulated in each cell line compared to the pool of all other cell lines. The median number of regulated genes per cell line is 218 (min: 15 and max: 721), cumulatively involving 4,686 unique known coding genes. Then, each cell line-specific gene list was submitted to TFactS and compared against catalogue of experimentally validated TF target genes using Fisher's test [30]. We have shown that this tool efficiently predicts TF regulation from regulated gene lists [33,34]. On the other hand, the SNP data were normalized and segmented then submitted to the GISTIC algorithm to identify chromosomal regions significantly altered in all these cell lines [35]. Figure S1 shows that significant amplifications and deletions were spread in the whole genome. A restricted analysis of TF-encoding genes revealed that 2,113 of the 2,335 genes known to encode "DNA binding" proteins (GO term) had their loci significantly altered, at least, in one cell line. To select transcription factors relevant to cancer in a more stringent manner, we combined the analysis on expression, activity and CNV (Figure 1).

For each TF, correlation profiles with other TFs were computed based on: regulation (inferred from TFactS analysis), gene expression and genomic alterations (CNV), respectively. The model in Figure 1 uses these correlation scores to find significant TFs, for which gene expression associated to CNV could explain the corresponding inferred regulation. 88 TFs were identified (p -values ≤ 0.05 , Table S1). Supporting our results, CNV affecting some of these TFs in cancer have already been reported, including: TP53, BRCA1, RUNX1 and MYC [36].

The Minimal Connected Network of Transcription Factors Regulated in Cancer Cell Lines

We used the Snow web tool to identify the minimal connected network (MCN) of protein-protein interactions involving the 88 TFs associated to cancer from our initial analysis. Snow predicted this MCN by computing the shortest paths linking the input proteins either directly or with one tolerated intermediate protein, based on a built-in database of human protein-protein interactions [37,38]. Restricting our Snow-based analysis to the human protein interactome with at least two

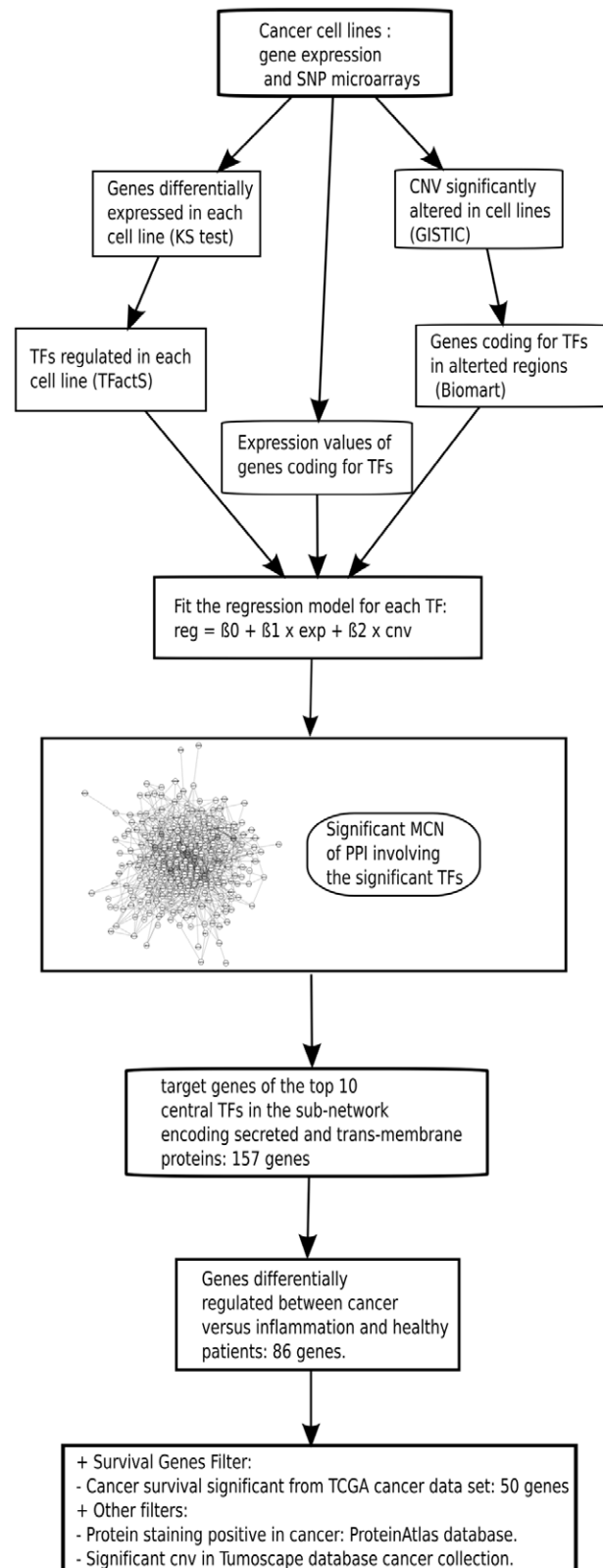


Figure 1. A workflow summarizing the strategy to identify accessible common cancer biomarkers. See text for details. Reg: regulation; Exp: expression; CNV: copy number variation; MCN: minimal connected network; PPI: protein-protein interactions; TF: Transcription factor.
doi:10.1371/journal.pone.0039666.g001

experimental evidences of interaction, we identified a subnetwork connecting 70 out of 88 TFs either directly or with one intermediate. It is remarkable that most of the TFs identified in the first step could be linked in this single protein-protein interaction subnetwork. Eighteen TFs were lost due to our restrictions in the analysis or to their absence in the Snow-annotated interactome. Snow uses Kolmogorov-Smirnov's test to evaluate the significance of the identified subnetwork by comparing its betweenness, connections and clustering coefficient distributions to those generated from 1,000 random networks with the same number of proteins. Our identified subnetwork had significant p-values for all these evaluated parameters (betweenness: $2.06E-37$, connections: $1.68E-47$, clustering coef.: $4.07E-43$). This subnetwork contained two distinct connected components. The first contained almost all interactions of the significant subnetwork and was considered as the cancer cell line-associated TFs MCN for subsequent analysis

(Figure 2). The second connected component, which has only two interactions linking three proteins was discarded.

We then asked whether false positives from TFactS, GISTIC and differential expression analyses could affect the MCN identification. To control these effects, we performed a negative control, in which we analyzed 100 different random lists of 88 TFs from the TFactS catalogue. Each list was submitted to Snow to produce a MCN using the same parameters as above. By comparing the distribution of the betweenness scores from all the random MCNs to the established MCN from our model, we found a significant difference (p-value ~ 0.01 ; KS test). Together with the results discussed above from the built-in comparison with 1,000 random networks performed in Snow, this suggested that our identified MCN constitutes a significant module involving TFs commonly regulated in cancer cell lines.

This MCN might be viewed as a regulatory “round-about” of the majority of regulated pathways in cancer cell lines. Indeed, as depicted in Figure 3, many MCN proteins are involved in many

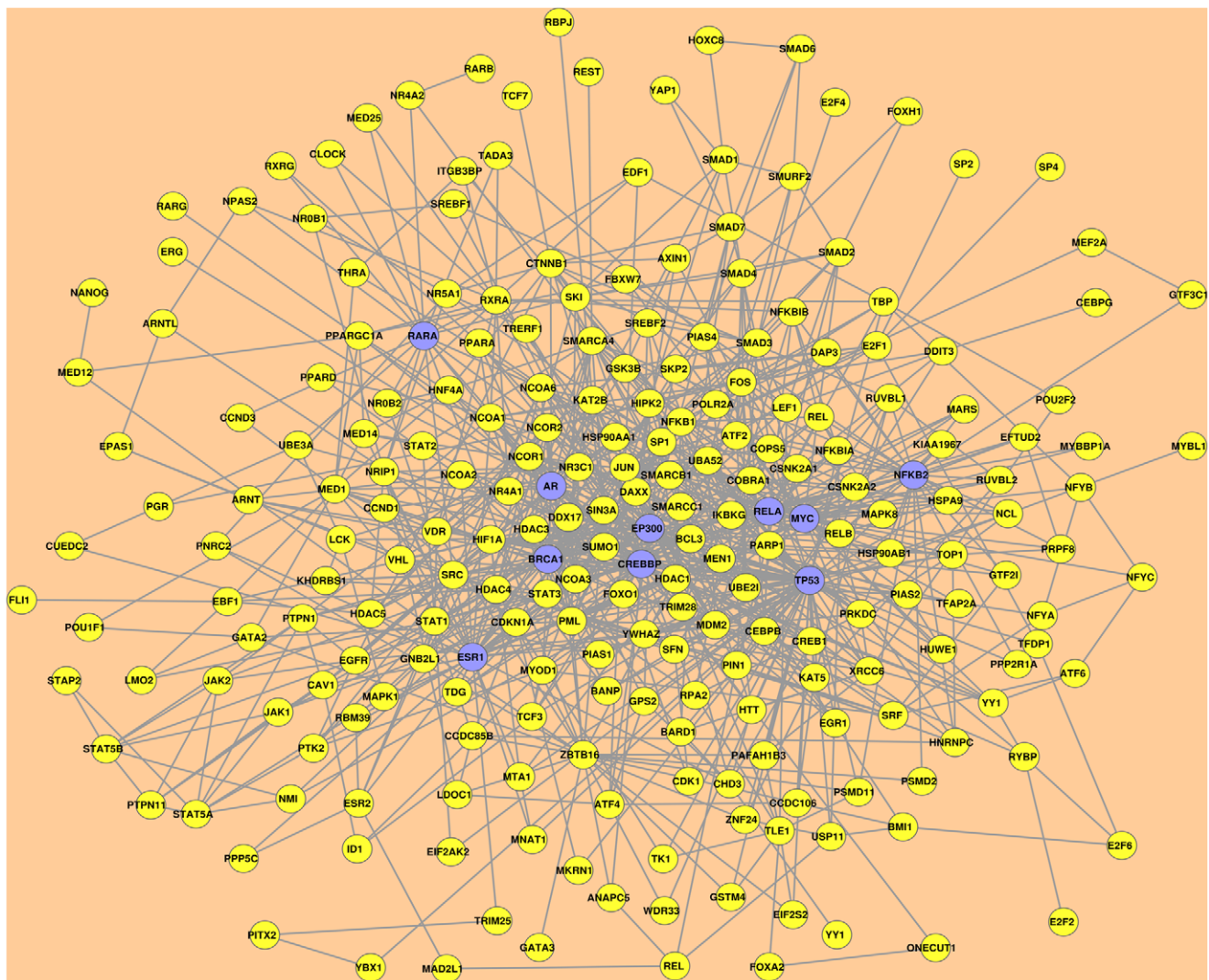


Figure 2. The minimal connected network of TFs regulated in cancer cell lines. The Snow web tool identified a significant human curated protein-protein interaction subnetwork involving 70 out of the 88 TFs correlatively regulated in cancer cell lines. The first connected component as shown here is considered as the minimal connected network (MCN) connecting these TFs. Each node represent a protein. Edges are the protein-protein interactions validated by at least two experimental evidences. Nodes shaded in violet represent the top ten most central TFs in the MCN. Node-ranking was based on the betweenness centrality scores. doi:10.1371/journal.pone.0039666.g002



Figure 3. Signaling pathway enrichment in the MCN proteins. All proteins (nodes) in the MCN were submitted to DAVID web tool for KEGG pathway enrichment analysis. Significant pathways are shown by categories according to the $-\log_{10}(p\text{-value})$ and the percentage of intersection between the submitted list and queried annotations. doi:10.1371/journal.pone.0039666.g003

cancer types and cancer signaling pathways. Nevertheless, MCN proteins are also significantly involved in immune response pathways. This could reflect an involvement of some MCN TFs like NFkB in both cancer and inflammation [16].

Target Genes of MCN Central Transcription Factors

Transcription factors in the minimal connected network identified above likely represent the main regulatory effectors commonly perturbed in the analyzed cancer cell lines. We focused on the most central TFs in this network. Centrality of nodes in a given network could be estimated using many parameters. Among them, betweenness scores the frequency by which a certain node is within the shortest paths linking any other two nodes. It is thought to be a good estimate of centrality [39]. By top-ranking the 236 MCN nodes according to their betweenness scores, we identified 59 central proteins having scores above the average. These central proteins show the same functional enrichment as the whole MCN.

We arbitrarily selected the top 10 central MCN nodes. Their encoding gene names are: TP53, ESR1, CREBBP, MYC, AR, BRCA1, RELA, RARA, EP300 and NFkB2. These ten TFs concentrate 41% of the total betweenness cumulative scores of the 236 MCN nodes. They could be considered as hubs or collectors of these network interactions. This is in line with the “scale free” model that was suggested to govern the TF protein-protein interactions, in which the hubs were built around TFs associated with malignancies [40]. We argued that common cancer biomarkers are likely to be found among the targets of these most central TFs. 874 unique target genes of these ten TFs are reported in the TFactS catalogue. An enrichment analysis of these genes, using “genetic association db disease” in the DAVID web tool, revealed an over-representation of a large panel of cancer types as well as ontologies related to immune responses and inflammatory diseases (File S1).

Cancer-specific Genes from Targets of MCN Central Transcription Factors

Enrichment analysis performed on MCN proteins as well as the targets of the central TFs showed an association between cancer and inflammation. This association is well documented in literature [41]. Cancer-specific biomarkers have to be differentially expressed in cancer patients compared to healthy individuals and patients with inflammatory diseases [42]. In addition, a universal cancer biomarker should be cancer-specific in a broad panel of tumor types. Since our interest is to identify “accessible” cancer biomarkers, we sought to restrict further analysis only on genes coding for secreted and transmembrane proteins. The SP-PIR annotation keywords database, as used in the DAVID tool, contains 1,689 and 642 genes annotated as encoding secreted and transmembrane proteins, respectively. In the 874 target genes of the ten most central TFs in the MCN, we found 57 genes encoding secreted proteins (p-value: $1.1E-6$) and 110 encoding transmembrane proteins (p-value: $4.3E-5$). This represents a unique set of 157 genes. Thus, identifying the TFs MCN and focusing on target genes of the ten most central TFs allowed us to prioritize a short list of accessible proteins to be analyzed in patient samples for differential expression (Figure 1).

We further filtered this gene list using available patient data. We performed gene expression analysis on an assembled microarray large data set of 8,525 different tissues from patients with cancer or inflammation and healthy individuals (Figure 4, File S2). From the prioritized 157 genes, we could establish a list of 86 cancer-specific transcripts (Figure 4). Among them, 3 genes were approved by FDA for cancer diagnosis, including: EGFR, KLK3 (PSA) and AFP for the diagnosis of colon, prostate and testis cancers, respectively [43]. Moreover, HLA-F in this list has already been reported as detectable in the serum of various cancer patients using indirect ELISA [10].

Potential Biomarkers Common in Cancer

In order to strengthen the likelihood to find potential common biomarkers among the cancer-specific gene list (Figure 5), we filtered these genes based on their significant effect on patient survival in any of the cancer types from TCGA database. The available gene expression data sets from TCGA, covering nine cancer types, were downloaded and analyzed separately for gene-survival association. For each gene, patients were divided in three groups (tertiles) according to the expression levels of the studied gene. Groups of patients with low, intermediate and high expression were then obtained. Making use of the available patient survival data: follow up duration and death status, we fitted Kaplan-Meier curves to these groups. Genes predicting patient survival significantly (log-rank p-value ≤ 0.05), in at least one cancer type, are shown in Table S2. The products of these 50 genes mediate many interacting pathways in cancer, as depicted in Figure S2 (KEGG pathway enrichment, p-value $\sim 4.29E-4$).

For each gene listed in Table S2, we added the following resources: (i) CNV significantly affecting the corresponding gene loci in all tumor types as analyzed in the Tumorscape database [44]; (ii) percentage of immunohistochemical (IHC) positive staining in cancer as detected in ProteinAtlas database [45]. We considered that genes positive for all criteria listed in Table S2 are more likely to be common cancer biomarker candidates. TFRC, VEGFA and MET are the best potential candidates. These genes have been separately associated to many cancer types in literature (Table S3).

Discussion

Cancer types have been screened separately for biomarker identification. Nowadays, there is an emerging effort to search for universal cancer markers. The recently available high-throughput data from cancer patient specimens make this task more affordable in the context of integrative analysis. This study was done within such a framework.

Cancer is a multistage disease, in which normal cells are progressively transformed to malignant ones. This process involves transcription factor (TF) regulation to ensure the transcription of needed genes [46]. We assumed that TFs regulated in cancer would have their activity explained by their coding gene expression level and genomic alterations. We hypothesized that cancer-associated TFs could interact together in a modular manner, such that cancer-triggering events end up perturbing the function of this module. Biomarkers common to many cancer types could be among these TF target genes. We then followed the workflow depicted in Figure 1 to target important genes commonly regulated in cancer that encode accessible proteins. We assumed that focusing on TFs will guide us to find the most valuable part of cancer information, which could be measured by gene expression [47]. Adding CNV data to filter important TFs will strengthen this approach. Whereas, analyzing all regulated genes and significantly altered chromosomal regions without any contextualization in terms of regulators (TFs) will dilute the common cancer biomarker among many false positive outcomes.

As a first step in our quest for common cancer biomarkers, we tried to identify the minimal connected network involving TFs, the activity of which is regulated in tumors. We integrated genomics and transcriptomics data from a panel of cancer cell lines, together with inferred TF regulation from gene expression using TFactS, which has been shown previously to be able to infer accurately TF regulation or activity status from a list of expressed genes [30]. The use of cell lines in this step is justified by the availability of both genomic and expression data. In addition, building meaningful MCN requires data from homogeneous cells, which is not the case of most primary cancer samples, in which genomic alterations and gene expression differ between cancer cells and stromal cells, and even between different cancer cell clones. We identified 88 TFs, which could be the main regulators in cancer cell lines. This step is, however, limited by the TFs represented in TFactS, although they sample the most studied TFs in literature. This step could also be ameliorated by taking into account other genomic alterations, such as mutations. However, whole genome alteration data were not available yet for all the studied cell lines.

By protein-protein interaction analysis, MCN connecting the majority of the 88 TFs has been identified from the curated human proteome network. The MCN contains both TFs and other proteins. Enrichment analysis revealed that this MCN assembles major known pathways driving multiple cancer types. Strikingly, immune response pathways were also enriched in MCN, which was identified based on cell line data, discarding any tumor micro-environmental effect on these results. This suggests a dual role played by this module of connected TFs in both cancer and inflammation. Results from our negative control procedure suggested that the cancer-associated MCN forms a significant module. This module's most central TFs are susceptible to act as the main “collectors” of marginal perturbations.

In a second step, we arbitrarily limited our analysis to target genes of the top ten most central MCN TFs. Enrichment analysis of these genes revealed a cancer context pathways over-representation, as expected. Since our purpose was to identify

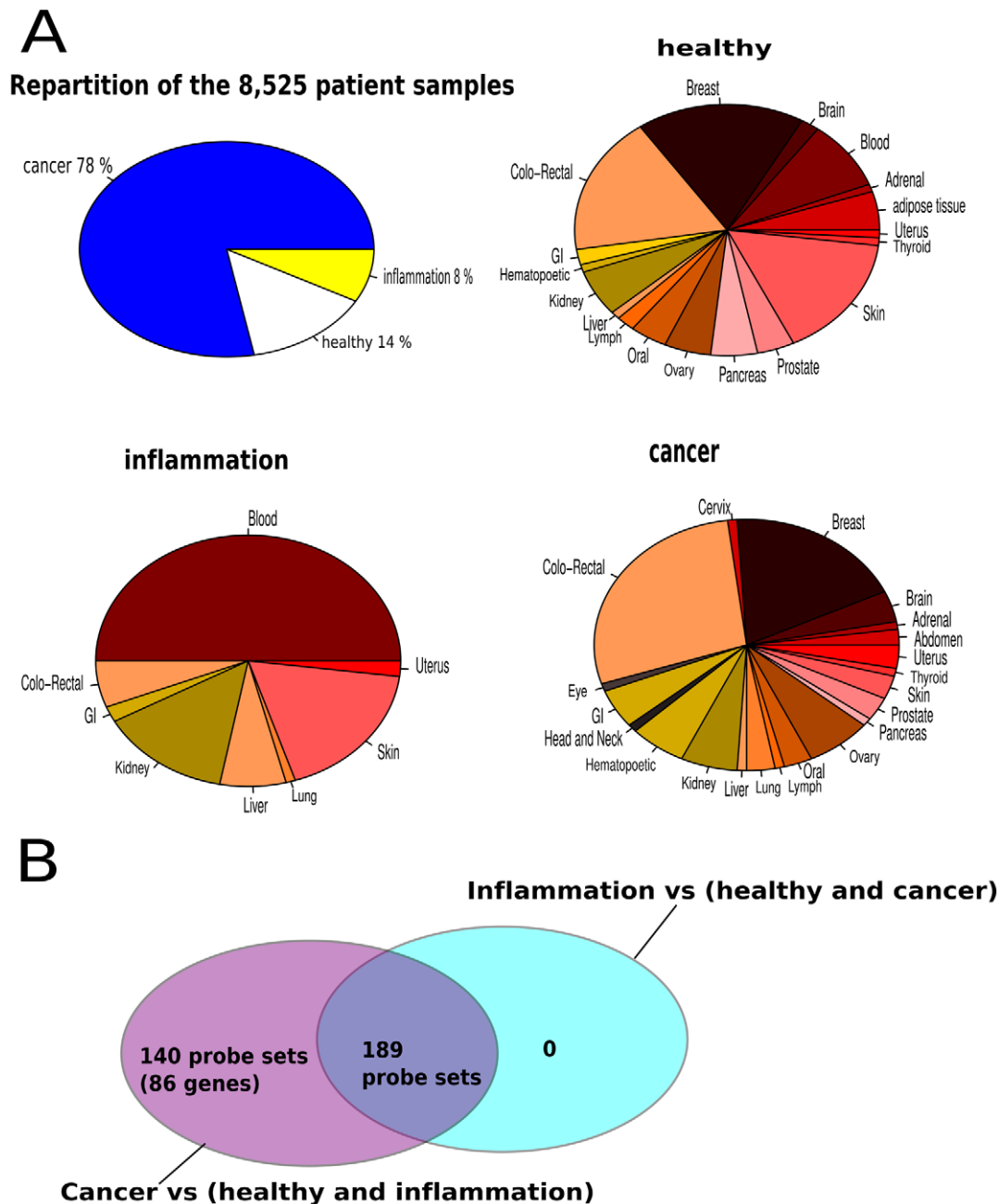


Figure 4. Patient sample repartition and cancer-specific gene expression analysis. Microarray gene expression data representing 8,525 patients samples were downloaded from GEO. A- 78% of patients had different cancer types; 14% are healthy individual and were sampled from different tissues; 8% of patients had inflammation/sepsis and were investigated from the whole blood and other tissues. B- differential expression of the MCN top ten central TFs target gene list coding for secreted and transmembrane proteins were analyzed. Among these genes, as shown in the Venn-diagram, 140 probe sets (86 unique genes) were found to be cancer-specific. GI: Gastro-intestinal.
doi:10.1371/journal.pone.0039666.g004

genes that could be easily probed in patients we filtered this gene list to 157 genes coding for secreted and transmembrane proteins. By comparing their expression in a panel of 8,525 patients, we identified a set of 86 cancer-specific genes differentially expressed in cancer versus normal and inflammation phenotypes. They include three out of six proteins approved by FDA in specific cancer diagnosis: PSA/KLK3, EGFR and AFP. Expression of these three genes could be checked in other cancer types. PSA, prostate specific antigen, for instance, although widely used in prostate cancer diagnosis, it was also reported in kidney, stomach

and breast cancers [48–50]. These results provide an internal validation of our methodology.

We sought to further restrict the analysis by taking into account the potential prognostic value in at least one cancer type. This was performed by associating gene expression to patient survival in TCGA data sets. 50 genes significantly predicted survival in at least one cancer type. Each of these genes could be investigated separately in the corresponding cancer type for prognosis. These genes are significantly involved and interconnected in many cancer pathways (Figure S2). Nevertheless, immunomodulatory

Transmembrane

Secreted

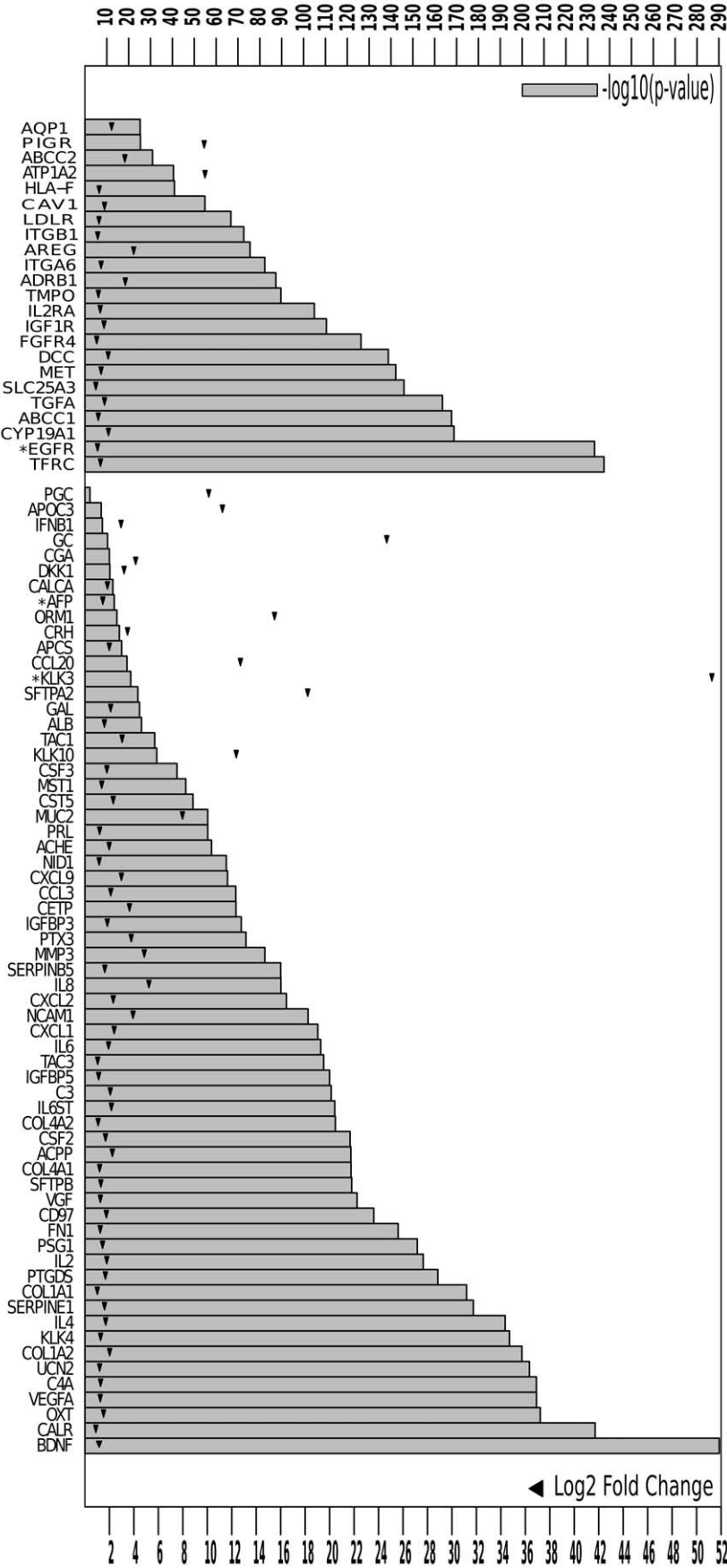


Figure 5. Significant cancer-specific biomarkers from patient gene expression analysis. Cancer-specific gene expression significance and fold change. Significance was attested by B-H p-value correction, and all shown genes have B-H p-value ≤ 0.05 . Bar-plots show the $-\log_{10}$ (uncorrected p-value). Triangles show the logged fold change of the corresponding gene in cancer compared to healthy and inflammation patient phenotypes. FDA approved cancer biomarkers are marked with (*).
doi:10.1371/journal.pone.0039666.g005

cytokines and chemokines were also enriched in this gene list, which might suggest that some of these genes may not fully distinguish patients with cancer from those with inflammatory diseases.

We identified three potential biomarkers common to cancer, i.e. **TFRC**, **VEGFA** and **MET** as evidenced by: (i) gene over-expression in cancer compared to normal and inflammation; (ii) gene expression significantly linked to patient survival in at least two cancer types; (iii) corresponding CNV focally significantly amplified in tumors; (iv) proteins stain positive in more than 80% of cancers. **VEGFA** promotes angiogenesis. Its diagnostic potential was investigated separately in many cancer types (Table S3). **MET**, is a known oncogenic tyrosine kinase receptor for hepatocyte growth factor. It is also associated with many cancer types (Table S3). In addition, it has been reported as a marker for cancer stem cells in: prostate, head and neck, liver, brain and lung cancers [51–56]. **VEGFA** and **MET** synergy in angiogenesis might be targeted for more effective anti-tumoral therapy [57]. **TFRC**, transferrin receptor, is known to be expressed in many tumor types (Table S3). Expression of **VEGFA** and **TFRC** is commonly regulated by **HIF** and **MYC**, which promote angiogenesis and proliferation, respectively [58–60]. The connection between these two TFs via their target genes is known to confer a metabolic advantage to tumors under hypoxia, which is a common condition in malignant diseases [61,62].

In summary, our strategy identified a network of TFs that regulate 50 potential common cancer biomarkers. Currently available data in TCGA, Tumorscape and ProteinAtlas databases pointed to **VEGFA**, **TFRC** and **MET** genes as potential candidates. Literature knowledge associated to these genes corroborates our approach. Taken together, all these observations might suggest to further investigate the usefulness of **VEGFA**, **MET** and **TFRC** as common cancer biomarkers. This could be performed by direct detection of these biomarkers or by checking for the presence of auto-antibodies directed against potential cancer proteins in patient serum, an approach that has gained much interest in the cancer diagnosis field [4,63].

Materials and Methods

Microarray Analysis

The data from 950 microarrays performed by GlaxoSmithKlein laboratories (GSK) on different cancer cell lines were downloaded from arrayExpress (E-MTAB-37). The RMA normalization method was applied using the xps package from R/Bioconductor [64]. Gene expression on each cell line was performed on duplicates or triplicates. Kolmogorov-Smirnov's test was performed to select genes differentially expressed in each cell line compared to others. A Bonferroni correction threshold was applied on p-values. Genes with an e-value ≤ 10 were considered as significantly differentially expressed on the corresponding cell line.

Transcription Factor Regulation Analysis

Each gene list regulated in each cell line was submitted to TFactS to predict regulated TFs [30]. TFactS sign less catalogue (version 2) contains 6,823 regulations linking 345 unique TFs to their 2,650 unique gene targets. For each list of regulated genes,

TFactS predicts the TFs whose targets are enriched in the submitted lists using Fisher's test. In this study, the larger sign-less catalogue was used instead of the restricted sign-sensitive one. TFactS was executed using BatchTFactS default parameters (www.tfacts.org). TFs with a positive e-value score ($-\log_{10}$ (e-value)) were considered as significant. TFs that were not significant in all cell-lines were discarded before the model fitting.

Genomic Copy Number Variation Analysis

The genomics data of the above cell lines were also released by GSK. SNP arrays data sets available on arrayExpress were downloaded (E-MTAB-38). They were analyzed using the aroma-affymetrix package on R/Bioconductor [65]. Briefly we applied a quantile normalization followed by the CRMA summarization and corrected for chip and PCR fragment length effects [66]. Then the GLAD algorithm was applied to raw copy numbers for segmentation [67]. The segmented data were then submitted to GISTIC algorithm to find significantly altered regions in all chromosomes except X and Y. A default q-value threshold of 0.25 was used to select significant regions [35]. Prior to CNV-based correlation matrix computing and the model fitting, CNV values for each gene in the significantly altered chromosomal regions were normalized as follows: (i) for each of the GISTIC-reported significant regions, we determined the median value of the significant CNV peaks; (ii) each gene in a significant chromosomal region has been assigned the value of this median. The values of the CNV were in log2-ratio as outputted by GISTIC. The chromosomal location of genes was obtained using the Ensembl genes 64 database with human "GRCH37.p5" release in Biomart web tool [68].

Identification of the Minimal Connected TF Regulated in Cancer Cell Lines

In order to identify a set of correlated TFs that are commonly regulated in cancer, we considered 305 cell lines, for which both expression and SNP data were available. Each TF has three measurements in each cell line: TF regulation scores estimated by TFactS ($-\log_{10}$ (e-value)), TF-encoding gene expressions (from microarrays) and TF-locus copy number variations (from median normalized GISTIC analysis). Three matrices, with TFs in rows and cell-lines in columns, could be built from these data: a TF regulation matrix, a TF-encoding gene expression matrix and a TF-locus CNV matrix. In each of these matrices, we computed correlations of each TF with the other TFs using Pearson correlation coefficient. These correlations could be represented as TF-TF correlation profiles. Then we fitted the following model for each TF: $R = \beta_0 + \beta_1 * E + \beta_2 * C$, Where:

(R) TF-TF correlation profile based on TFactS scores, only TFs significantly regulated in at least one cell line were used; (E) TF-TF correlation profile based on gene expression; (C) TF-TF correlation profile from significant regions identified by GISTIC algorithm, these correlations were computed using loci copy number variation median-normalized values.

Each TF having significant β_1 (p-value ≤ 0.05) and β_2 (p-value ≤ 0.05) was considered as correlatively regulated in cancer. For these TFs, there is an effect of gene expression and genomic alteration (CNV) on their regulation.

The list of these significant TFs was submitted to SNOW web tool to identify their minimal connected network [38]. The SNOW parameters were set as follows: **curated human protein-protein interactome with at least two experimental evidences, 1,000 random networks for significance, only one extension interaction was allowed to connect the submitted proteins.** A built-in mapping was performed by Snow between submitted gene names and the network protein names.

MCN Protein-protein Interaction Network Analysis

Network analysis was done using Cytoscape [69]. MCN node centrality analysis was computed using CentiScaPe plugin. Proteins were ranked according to highest betweenness [39]. Target genes of the top ten ranked TFs were kept for further analysis.

MCN Negative Control

From the list of all TFs represented in TFactS sign-less catalogue, we generated 100 random lists containing 88 distinct elements. These 100 random lists were submitted to Snow web tool using the same parameters as above. The betweenness distribution of the identified cancer-related MCN was compared to the distribution of the whole random MCNs using Kolmogorov-Smirnov's test in R. The betweenness scores used here are from Snow output.

Functional Enrichment Analysis

The DAVID web tool was used to perform a functional analysis of the selected target genes [70]. A p-value threshold of 0.05 was used for significance.

Cancer Patients Gene Expression Analysis

Data from 8,525 patient samples analyzed with HG-U133A2Plus microarrays were downloaded from GEO database. Patient categories were distributed as follows: 78% cancer, 8% inflammation and 14% healthy. Expression data were log₂-transformed then normalized per gene by subtracting the gene median expression and dividing by the inter-quartile-range of its expression vector. **This data set was used to compare gene expression profiles of genes significantly annotated as coding for secreted or transmembrane proteins using SP-PIR annotation keywords on DAVID from the known targets of the MCN top ten central TFs.** Genes significantly differentially expressed were determined between cancer compared to inflammation and healthy phenotypes on one hand, and between inflammation compared to cancer and healthy phenotypes on the other hand. **Genes specific to cancer were significantly differentially regulated in the former comparison but not in the later. Differentially regulated genes were computed using R/Bioconductor (limma package) and significance was controlled using B-H correction of p-value ≤ 0.05 [64,71].** The cancer-specific gene list identified from this analysis was considered as the accessible cancer-specific biomarker list (genes specific to cancer and coding for secreted and transmembrane proteins).

Survival Analysis

The accessible cancer-specific biomarker list was evaluated for prognosis potential in different TCGA published cancer data sets. **TCGA database offers a set of gene expression from clinically annotated patient samples** (<http://tcga-data.nci.nih.gov/tcga/>). We downloaded (09/27/2011) level2 gene expression data for all the patients from 9 publicly available cancer types (GBM: Glioblastoma multiform, OV: Ovarian serous cystadenocarci-

noma, LAML: Acute Myeloid Leukemia, BRCA: Breast invasive carcinoma, COAD: Colon adenocarcinoma, KIRC: Kidney renal clear cell carcinoma, LUSC: Lung squamous cell carcinoma, UCEC: Uterine Corpus Endometrioid Carcinoma). The number of patients per gene expression data set was distributed as follows: BRCA: 600; COAD: 179; GBM: 536 (agilent) or 555 (affy); KIRC: 72; LAML: 197; LUSC: 161 (agilent) or 134 (affy); OV: 608 (agilent) or 590 (affy); READ: 78; UCEC: 54. **Survival data were among the clinical information available for the majority of these patients.** For each gene, in each cancer type, the expression vector from level2 TCGA data was standardized by median subtraction and inter-quartile range division. **Then each normalized expression vector corresponding to each gene was divided on three groups (tertiles).** Patients in the lower tertile were assigned a gene down-regulation; patients in the intermediate tertile were assigned an intermediate gene expression; patients in the upper tertile were assigned an up-regulation of the gene in interest. The R/Bioconductor (survival package) was used to fit survival curves on categorized patients for each gene [72]. Three Kaplan-Meier survival curves were then fitted for each gene corresponding to up-regulation, intermediate and down-regulation groups of patients according to their death status and follow up duration. Significance of the difference between these curves was estimated based on the log-rank p-value ≤ 0.05 . If data from more than one platform was available (Affymetrix or Agilent), the analysis is done independently on each.

Supporting Information

Figure S1 Copy Number Variation results. Copy Number Variation GISTIC scores and q-values for the whole cancer cell-lines data set. Significance profiles (q-values in the bottom and scores in the top horizontal axes) are shown for chromosomal regions in the left and right vertical axes of the panels for : A-amplified regions, B- deleted regions.

(TIF)

Figure S2 Patient survival-affecting genes are involved in many cross-talking cancer pathways. Genes significantly affecting cancer patient survival are mechanistically interacting to trigger important cancer pathways. These genes are enriched (p-value = $4.29E-4$) in KEGG's cancer pathways analyzed by DAVID web-tool.

(TIF)

File S1 Enrichment analysis of the top 10 MCN TF target genes

(XLS)

File S2 Patient gene expression data sets from GEO database and their accession numbers.

(XLS)

Table S1 Transcription factors regulated in cancer cell lines. For each TF, the TF-TF correlations profiles were determined and the following model was fitted: $\text{reg} = \beta_0 + \beta_1 * \text{exp} + \beta_2 * \text{cnv}$. Where, for each TF versus the others: "reg" is the regulation correlation profile, "exp" is the TF-coding gene expression correlation profile and "cnv" is the TF coding-gene locus copy number variation correlation profile.

(PDF)

Table S2 Filtered cancer-specific genes. Significant cancer-specific genes identified in patient microarray analysis were filtered to those that are significantly associated (p-value ≤ 0.05) with at least one cancer type survival based on TCGA data. The resulting gene list is shown here. Tumorscape database gives q-

values for CNV alterations in a large collection of tumors. Based on the CNV from all tumors pooled together, amplification and deletion q-values corresponding to the chromosomal locations of these genes are shown. Whether the loci is focally affected by the CNV is also shown. ProteinAtlas database contains cancer and normal tissue IHC staining for proteins with available antibodies. The percentage of staining in cancer is shown. N.A: not available, N.S: non-significant; GBM: Glioblastoma mutiforme, OV: Ovarian serous cystadenocarcinoma, LAML: Acute Myeloid Leukemia, BRCA: Breast invasive carcinoma, COAD: Colon adenocarcinoma, KIRC: Kidney renal clear cell carcinoma, LUSC: Lung squamous cell carcinoma, UCEC: Uterine Corpus Endometrioid Carcinoma. K-M: Kaplan-Meier log-rank test.

(PDF)

References

- Mehta S, Shelling A, Muthukaruppan A, Lasham A, Blenkinsop C, et al. (2010) Predictive and prognostic molecular markers for cancer medicine. *Ther Adv Med Oncol* 2: 125–148. doi:http://dx.doi.org/10.1177/1758834009360519.
- Buseman CM, Wright WE, Shay JW (2011) Is telomerase a viable target in cancer? *Mutat Res* 730: 90–7. doi:http://dx.doi.org/10.1016/j.mrfmmm.2011.07.006.
- Xu T, Lu B, Tai Y-C, Goldkorn A (2010) A cancer detection platform which measures telomerase activity from live circulating tumor cells captured on a microfilter. *Cancer Res* 70: 6420–6426. doi:http://dx.doi.org/10.1158/0008-5472.CAN-10-0686.
- Cho-Chung YS (2006) Autoantibody biomarkers in the detection of cancer. *Biochim Biophys Acta* 1762: 587–591. doi:http://dx.doi.org/10.1016/j.bbdis.2006.04.001.
- Nesterova M, Johnson N, Cheadle C, Cho-Chung YS (2006) Autoantibody biomarker opens a new gateway for cancer diagnosis. *Biochim Biophys Acta* 1762: 398–403. doi:http://dx.doi.org/10.1016/j.bbdis.2005.12.010.
- Radu A, Pichon C, Camparo P, Antoine M, Allory Y, et al. (2010) Expression of follicle-stimulating hormone receptor in tumor blood vessels. *N Engl J Med* 363: 1621–1630. doi:http://dx.doi.org/10.1056/NEJMoa1001283.
- Murray GL, Taylor MC, McFadyen MC, McKay JA, Greenlee WF, et al. (1997) Tumor-specific expression of cytochrome P450 CYP1B1. *Cancer Res* 57: 3026–3031.
- Levenson VV (2010) DNA methylation as a universal biomarker. *Expert Rev Mol Diagn* 10: 481–488. doi:http://dx.doi.org/10.1586/erm.10.17.
- Levenson VV, Melnikov AA (2011) The MethDet: a technology for biomarker development. *Expert Rev Mol Diagn* 11: 807–812. doi:http://dx.doi.org/10.1586/erm.11.74.
- Noguchi K, Isogai M, Kuwada E, Noguchi A, Goto S, et al. (2004) Detection of anti-HLA-F antibodies in sera from cancer patients. *Anticancer Res* 24: 3387–3392.
- Polanski M, Anderson NL (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights* 1: 1–48.
- Dawany NB, Dampier WN, Tozeren A (2011) Large-scale integration of microarray data reveals genes and pathways common to multiple cancer types. *Int J Cancer* 128: 2881–2891.
- Basil CF, Zhao Y, Zavaglia K, Jin P, Panelli MC, et al. (2006) Common cancer biomarkers. *Cancer Res* 66: 2953–2961. doi:http://dx.doi.org/10.1158/0008-5472.CAN-05-3433.
- Yang Y, Pospisil P, Iyer LK, Adelstein SJ, Kassai AI (2008) Integrative genomic data mining for discovery of potential blood-borne biomarkers for early diagnosis of cancer. *PLoS One* 3: e3661. doi:http://dx.doi.org/10.1371/journal.pone.0003661.
- Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, et al. (2007) A map of human cancer signaling. *Mol Syst Biol* 3: 152. doi:http://dx.doi.org/10.1038/msb4100200.
- Ben-Neriah Y, Karin M (2011) Inflammation meets cancer, with NF- κ B as the matchmaker. *Nat Immunol* 12: 715–723. doi:http://dx.doi.org/10.1038/ni.2060.
- Javelaud D, Alexaki VI, Denner S, Mohammad KS, Guise TA, et al. (2011) TGF- β /SMAD/GLI2 signaling axis in cancer progression and metastasis. *Cancer Res* 71: 5606–5610. doi:http://dx.doi.org/10.1158/0008-5472.CAN-11-1194.
- Zavoral M, Minarikova P, Zavada F, Salek C, Minarik M (2011) Molecular biology of pancreatic cancer. *World J Gastroenterol* 17: 2897–2908. doi:http://dx.doi.org/10.3748/wjg.v17.i24.2897.
- Steidl C, Gascoyne RD (2011) The molecular pathogenesis of primary mediastinal large B-cell lymphoma. *Blood* 118: 2659–2669. doi:http://dx.doi.org/10.1182/blood-2011-05-326538.
- Abel EV, Aplin AE (2011) Finding the root of the problem: the quest to identify melanoma stem cells. *Front Biosci (Schol Ed)* 3: 937–945.
- Demoulin J-B, Enarsson M, Larsson J, Essaghir A, Heldin C-H, et al. (2006) The gene expression profile of PDGF-treated neural stem cells corresponds to partially differentiated neurons and glia. *Growth Factors* 24: 184–196.
- Laurenti G, Benedetti E, D'Angelo B, Cristiano L, Cinque B, et al. (2011) Hypoxia induces peroxisome proliferator-activated receptor α (PPAR α) and lipid metabolism peroxisomal enzymes in human glioblastoma cells. *J Cell Biochem* 112: 3891–3901. doi:http://dx.doi.org/10.1002/jcb.23323.
- Goodarzi H, Elemento O, Tavazoie S (2009) Revealing global regulatory perturbations across human cancers. *Mol Cell* 36: 900–911. doi:http://dx.doi.org/10.1016/j.molcel.2009.11.016.
- Little JL, Wheeler FB, Koumenis C, Kridel SJ (2008) Disruption of crosstalk between the fatty acid synthesis and proteasome pathways enhances unfolded protein response signaling and cell death. *Mol Cancer Ther* 7: 3816–3824. doi:http://dx.doi.org/10.1158/1535-7163.MCT-08-0558.
- Wilson BJ, Giguère V (2008) Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer* 7: 49. doi:http://dx.doi.org/10.1186/1476-4598-7-49.
- Deshmukh H, Yu J, Shaik J, MacDonald TJ, Perry A, et al. (2011) Identification of transcriptional regulatory networks specific to pilocytic astrocytoma. *BMC Med Genomics* 4: 57. doi:http://dx.doi.org/10.1186/1755-8794-4-57.
- Gioeli D (2011) The Dynamics of the Cell Signaling Network; Implications for Targeted Therapies. *Targeted Therapies*. Humana Press. 33–53.
- Wu G, Feng X, Stein L (2010) A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* 11: R53. doi:http://dx.doi.org/10.1186/gb-2010-11-5-r53.
- Chen L, Xuan J, Wang C, Shih I-M, Wang Y, et al. (2008) Knowledge-guided multi-scale independent component analysis for biomarker identification. *BMC Bioinformatics* 9: 416. doi:http://dx.doi.org/10.1186/1471-2105-9-416.
- Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, et al. (2010) Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. *Nucleic Acids Res* 38: e120. doi:http://dx.doi.org/10.1093/nar/gkq149.
- Xu Y, Duanmu H, Chang Z, Zhang S, Li Z, et al. (2011) The application of gene co-expression network reconstruction based on CNVs and gene expression microarray data in breast cancer. *Mol Biol Rep* 39: 1627–37. doi:http://dx.doi.org/10.1007/s11033-011-0902-3.
- Tran LM, Zhang B, Zhang Z, Zhang C, Xie T, et al. (2011) Inferring causal genomic alterations in breast cancer using gene expression data. *BMC Syst Biol* 5: 121. doi:http://dx.doi.org/10.1186/1752-0509-5-121.
- Pachikian BD, Essaghir A, Demoulin J-B, Neyrinck AM, Catry E, et al. (2011) Hepatic n-3 polyunsaturated fatty acid depletion promotes steatosis and insulin resistance in mice: genomic analysis of cellular targets. *PLoS ONE* 6: e23365. doi:10.1371/journal.pone.0023365.
- Montano-Almendras CP, Essaghir A, Schoemans H, Varis I, Noel LA, et al. (2012) ETV6-PDGFRB and FIP1L1-PDGFR α stimulate human hematopoietic progenitor proliferation and differentiation into eosinophils: role of NF- κ B. *Haematologica*. Available: http://www.ncbi.nlm.nih.gov/pubmed/22271894. Accessed 2012 May 2.
- Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104: 20007–20012. doi:http://dx.doi.org/10.1073/pnas.0710052104.
- Kuiper RP, Ligtenberg MJL, Hoogerbrugge N, Geurts van Kessel A (2010) Germline copy number variation and cancer risk. *Curr Opin Genet Dev* 20: 282–289. doi:10.1016/j.gde.2010.03.005.
- Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, et al. (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res* 38: W210–213. doi:10.1093/nar/gkq388.

Table S3 Evidences of VEGFA, TFRC and MET associations to cancer from a non-exhaustive literature screening.
(PDF)

Acknowledgments

We acknowledge Pr. Valeria Giandominico for her help and advice. Department of Medical Sciences, Endocrine Oncology, Uppsala University, Uppsala, Sweden.

Author Contributions

Conceived and designed the experiments: AE JBD. Performed the experiments: AE. Analyzed the data: AE. Contributed reagents/materials/analysis tools: AE. Wrote the paper: AE JBD. Approved the last version of the paper: JBD.

38. Minguez P, Götz S, Montaner D, Al-Shahrour F, Dopazo J (2009) SNOW, a web-based tool for the statistical analysis of protein-protein interaction networks. *Nucleic Acids Res* 37: W109–W114. doi:http://dx.doi.org/10.1093/nar/gkp402.
39. Scardoni G, Petterlini M, Laudanna C (2009) Analyzing biological network parameters with CentiScaPe. *Bioinformatics* 25: 2857–2859. doi:http://dx.doi.org/10.1093/bioinformatics/btp517.
40. Rodriguez-Caso C, Medina MA, Solé RV (2005) Topology, tinkering and evolution of the human transcription factor network. *FEBS J* 272: 6423–6434. doi:10.1111/j.1742-4658.2005.05041.x.
41. Aggarwal BB, Sung B (2011) The relationship between inflammation and cancer is analogous to that between fuel and fire. *Oncology (Williston Park)* 25: 414–418.
42. Chechlinska M, Kowalewska M, Nowak R (2010) Systemic inflammation as a confounding factor in cancer biomarker discovery and validation. *Nat Rev Cancer* 10: 2–3. doi:http://dx.doi.org/10.1038/nrc2782.
43. Ludwig JA, Weinstein JN (2005) Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 5: 845–856. doi:http://dx.doi.org/10.1038/nrc1739.
44. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463: 899–905. doi:10.1038/nature08822.
45. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, et al. (2010) Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* 28: 1248–1250. doi:10.1038/nbt1210-1248.
46. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100: 57–70.
47. Lefebvre C, Rieckhof G, Califano A (2012) Reverse-engineering human regulatory networks. *Wiley Interdisciplinary Reviews Systems Biology and Medicine*. Available: http://www.ncbi.nlm.nih.gov/pubmed/22246697. Accessed 30 March 2012.
48. Clements J, Ward G, Kaushal A, Hi SI, Kennett C, et al. (1997) A prostate-specific antigen-like protein associated with renal cell carcinoma in women. *Clin Cancer Res* 3: 1427–1431.
49. Milne ANA, Sitarz R, Carvalho R, Polak MM, Ligtenberg M, et al. (2007) Molecular analysis of primary gastric cancer, corresponding xenografts, and 2 novel gastric carcinoma cell lines reveals novel alterations in gastric carcinogenesis. *Hum Pathol* 38: 903–913. doi:http://dx.doi.org/10.1016/j.humpath.2006.12.010.
50. Chang Y-F, Hung S-H, Lee Y-J, Chen R-C, Su L-C, et al. (2011) Discrimination of breast cancer by measuring prostate-specific antigen levels in women's serum. *Anal Chem* 83: 5324–5328. doi:http://dx.doi.org/10.1021/ac200754x.
51. van Leenders GJLH, Sookhlall R, Teubel WJ, de Ridder CMA, Reneman S, et al. (2011) Activation of c-MET Induces a Stem-Like Phenotype in Human Prostate Cancer. *PLoS One* 6: e26753. doi:http://dx.doi.org/10.1371/journal.pone.0026753.
52. Sun S, Wang Z (2011) Head neck squamous cell carcinoma c-Met⁺ cells display cancer stem cell properties and are responsible for cisplatin-resistance and metastasis. *Int J Cancer* 129: 2337–2348. doi:http://dx.doi.org/10.1002/ijc.25927.
53. You H, Ding W, Dang H, Jiang Y, Rountree CB (2011) c-Met represents a potential therapeutic target for personalized treatment in hepatocellular carcinoma. *Hepatology* 54: 879–89. doi:http://dx.doi.org/10.1002/hep.24450.
54. Li Y, Li A, Glas M, Lal B, Ying M, et al. (2011) c-Met signaling induces a reprogramming network and supports the glioblastoma stem-like phenotype. *Proc Natl Acad Sci U S A* 108: 9951–9956. doi:http://dx.doi.org/10.1073/pnas.1016912108.
55. You W-K, Sennino B, Williamson CW, Falcón B, Hashizume H, et al. (2011) VEGF and c-Met blockade amplify angiogenesis inhibition in pancreatic islet cancer. *Cancer Res* 71: 4758–4768. doi:http://dx.doi.org/10.1158/0008-5472.CAN-10-2527.
56. Basak SK, Veena MS, Oh S, Huang G, Srivatsan E, et al. (2009) The malignant pleural effusion as a model to investigate intratumoral heterogeneity in lung cancer. *PLoS One* 4: e5884.
57. You W-K, Sennino B, Williamson CW, Falcón B, Hashizume H, et al. (2011) VEGF and c-Met blockade amplify angiogenesis inhibition in pancreatic islet cancer. *Cancer Res* 71: 4758–4768. doi:http://dx.doi.org/10.1158/0008-5472.CAN-10-2527.
58. Whan Kim J, Gao P, Liu Y-C, Semenza GL, Dang CV (2007) Hypoxia-inducible factor 1 and dysregulated c-Myc cooperatively induce vascular endothelial growth factor and metabolic switches hexokinase 2 and pyruvate dehydrogenase kinase 1. *Mol Cell Biol* 27: 7381–7393. doi:http://dx.doi.org/10.1128/MCB.00440-07.
59. Tacchini L, Bianchi L, Bernelli-Zazzera A, Cairo G (1999) Transferrin receptor induction by hypoxia. HIF-1-mediated transcriptional activation and cell-specific post-transcriptional regulation. *J Biol Chem* 274: 24142–24146.
60. O'Donnell KA, Yu D, Zeller KI, Kim J-W, Racke F, et al. (2006) Activation of transferrin receptor 1 by c-Myc enhances cellular proliferation and tumorigenesis. *Mol Cell Biol* 26: 2373–2386. doi:http://dx.doi.org/10.1128/MCB.26.6.2373-2386.2006.
61. Dang CV, whan Kim J, Gao P, Yustein J (2008) The interplay between MYC and HIF in cancer. *Nat Rev Cancer* 8: 51–56. doi:http://dx.doi.org/10.1038/nrc2274.
62. Sethi G, Shanmugam MK, Ramachandran L, Kumar AP, Tergaonkar V (2012) Multifaceted link between cancer and inflammation. *Biosci Rep* 32: 1–15. doi:http://dx.doi.org/10.1042/BSR20100136.
63. Cui T, Hurtig M, Elgue G, Li S-C, Veronesi G, et al. (2010) Paraneoplastic antigen Ma2 autoantibodies as specific blood biomarkers for detection of early recurrence of small intestine neuroendocrine tumors. *PLoS ONE* 5: e16010. doi:10.1371/journal.pone.0016010.
64. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80. doi:http://dx.doi.org/10.1186/gb-2004-5-10-r80.
65. Bengtsson H, Simpson K, Bullard J, Hansen K (2008) aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. Available: http://128.32.135.2/tech-reports/745.pdf. Accessed May 2012.
66. Bengtsson H, Irizarry R, Carvalho B, Speed TP (2008) Estimation and assessment of raw copy numbers at the single locus level. *Bioinformatics* 24: 759–767. doi:http://dx.doi.org/10.1093/bioinformatics/btn016.
67. Hupé P, Stransky N, Thiery J-P, Radvanyi F, Barillot E (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* 20: 3413–3422. doi:http://dx.doi.org/10.1093/bioinformatics/bth418.
68. Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, et al. (2011) BioMart Central Portal: an open database network for the biological community. *Database (Oxford)* 2011: bar041. doi:http://dx.doi.org/10.1093/database/bar041.
69. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431–432. doi:http://dx.doi.org/10.1093/bioinformatics/btq675.
70. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57. doi:http://dx.doi.org/10.1038/nprot.2008.211.
71. Smyth GK (2005) Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer. 397–420.
72. Therneau T, original Splus->R port by Thomas Lumley (2011) survival: Survival analysis, including penalised likelihood. Available: http://CRAN.R-project.org/package=survival. Accessed September 2011.