Personalized disease signatures through information-theoretic compaction of big cancer data

Swetha Vasudevan[1, 2], Efrat Flashner-Abramson[1], F. Remacle[2, 3], R. D. Levine[2, 4, 5], Nataly Kravchenko-Balasha[1,5]

[1] Bio-Medical Sciences Department, The Faculty of Dental Medicine, The Hebrew University of Jerusalem, Jerusalem 9112001, Israel.

[2] Fritz Haber Research Centre, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

[3] Theoretical Physical Chemistry, research Unit Molecular Systems, B6c, University of Liege, B4000 Liege, Belgium

[4] Department of Molecular and Medical Pharmacology, David Geffen School of Medicine and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095

[5] To whom correspondence should be addressed. Email: natalyk@ekmd.huji.ac.il; rafi@fh.huji.ac.il

Abstract

Every individual cancer develops and grows in its own specific way, giving rise to a recognized need for the development of personalized cancer diagnostics. This suggested that the identification of patient-specific oncogene markers would be an effective diagnostics approach. However, tumors that are classified as similar according to the expression levels of certain oncogenes can eventually demonstrate divergent responses to treatment. This implies that the information gained from the identification of tumor-specific biomarkers is still not sufficient.

We present a method to quantitatively transform heterogeneous big cancer data to patient-specific transcription networks. These networks characterize the unbalanced molecular processes that deviate the tissue from the normal state. We study a number of datasets spanning five different cancer types, aiming to capture the extensive inter-patient heterogeneity that exists within a specific cancer type as well as between cancers of different origins. We show that a relatively small number of altered molecular processes suffices to accurately characterize over 500 tumors, demonstrating extreme compaction

of the data. Every patient is characterized by a small specific subset of unbalanced processes. We validate the result by verifying that the processes identified characterize other cancer patients as well.

We show that different patients may display similar oncogene expression levels, albeit carrying biologically distinct tumors that harbor different sets of unbalanced molecular processes. Thus, tumors may be inaccurately classified, and addressed as similar. These findings highlight the need to expand the notion of tumor-specific oncogenic biomarkers to patient-specific, comprehensive transcriptional networks for improved patient-tailored diagnostics.

Significance

Accurate cancer diagnostics is a prerequisite for optimal personalized cancer medicine. We propose an information-theoretic cancer diagnosis that identifies signatures comprising patient-specific oncogenic processes, rather than cancer type-specific biomarkers. Such comprehensive transcriptional signatures should allow for more accurate classification of cancer patients, and better patient-specific diagnostics. The approach we describe herein allows decoding large-scale molecular level information and elucidating patient-specific transcriptional altered network structures. Thereby, we move from cancer type-associated biomarkers to unbiased patient-specific unbalanced oncogenic processes.

/body

Introduction

Cancer results from the acquisition of genetic alterations, which, in turn, lead to significant rewiring of molecular networks. Despite a diverse array of genetic mutations in tumors, there are typically fewer distinct phenotypes than the extent of genetic, epigenetic and transcriptional heterogeneity would suggest (1). Many tumors eventually rely on a limited number of key proteins (oncogenes) that are responsible for cancer growth and survival, a phenomenon known as 'oncogene addiction' (2). This has led to the idea that the identification of the key tumor-specific oncogene biomarkers would be an effective diagnostics strategy. However, extensive molecular variations between patients from the same cancer type, referred to as inter-patient heterogeneity, render it difficult to find common gene markers that correlate well with drug sensitivity and patient survival. Indeed, it was recently proposed that reliable genomic markers should be identified and integrated into the pathology process in order to diagnose and treat each patient optimally (3). Moreover, different molecular processes may give rise to the same list of oncogenic biomarkers (1). Hence, tumors may be classified as similar for the purposes of diagnostics and treatment, despite being biologically different.

Our main goal in this study was to develop a method that classifies patients not only based on their tumor-specific list of oncogenic biomarkers, but rather, based on the molecular context that gave rise to this list of biomarkers. To this end, we investigated gene expression alterations in a cohort of 527 samples consisting of lymphoma, bladder cancer, gastric cancer, colorectal cancer, breast cancer, and normal gastric tissues, and identified the set of ongoing molecular processes that make up each patient-specific transcriptional network. We show how the numerous gene expression alterations that occurred in the large cohort of tumors can be translated to *a few altered molecular processes* (4, 5) that repeat themselves in different combinations in every patient, and accurately characterize the vast inter-patient heterogeneity. These few unbalanced processes, identified by an information theoretic approach, achieved a significant compaction of the big data set. For each patient we identify the tumor-specific *set* of altered, or unbalanced molecular processes.

We show that similar expression levels of certain oncogenes in different patients can be attributed to different combinations of unbalanced processes. This suggests that in order to accurately classify patients, transcriptional networks, instead of lists of biomarkers, should be identified.

The approach described herein provides an important additional step towards accurately decoding cancer information in a patient-specific manner. Our findings highlight the need to incorporate oncogene biomarkers into the context of transcriptional networks in order to accurately characterize patient-specific tumor biology, and to improve patient-tailored diagnostics.

Using information theory to identify patient-specific ongoing cancer processes

Tumors are biological systems in which the balanced homeostatic state has been disturbed due to genomic and environmental factors, or constraints (6–8). These constraints bring about an imbalance in the tissue and result in abnormal gene expression levels reflecting ongoing unbalanced molecular processes. To quantify the imbalance we use a thermodynamically-based information–theoretic strategy. Thermodynamic-based approaches (9–13) and/or information-theoretical approaches, have been successfully applied to the analysis of biological systems in a number of cases (see for example (14–17)). In this study we utilize the thermodynamic-motivated surprisal analysis (6, 7, 18, 19). We have previously applied this analysis to various biological systems (4, 20–22), and also demonstrated its experimental validity (21, 23).

The equation used in the study represents the logarithm of the experimental transcript expression level, $\ln X_i(k)$, of a measured transcript *i*, in every patient *k* as, [7]:

$$\underbrace{\ln X_i(k)}_{\substack{\text{logarithm of}\\ \text{measured intensity}\\ \text{of transcript } i\\ \text{in patient } k}} = \underbrace{\ln X_i^o(k)}_{\substack{\text{logarithm of}\\ \text{intensity of transcript } i\\ \text{in the balanced state}\\ \text{in patient } k}} - \underbrace{\sum_{\alpha=1} G_{i\alpha}\lambda_\alpha(k)}_{\substack{\text{logarithm of deviations in the}\\ \text{intensity of transcipt } i\\ \text{due to the constraints } \alpha=1,2,..\\ \text{as a sum over these processes}}} \qquad [1]$$

where $\ln X_i^0(k)$ is the logarithm of the expression level of the transcript *i* at the balanced state, and the sum, $\sum_{\alpha=1} G_{i\alpha}\lambda_\alpha(k)$, represents the deviations in the logarithm of the expression level of this transcript from the balanced state level due to the environmental/genetic constraints that may operate in the system.

Surprisal analysis identifies which gene transcripts are at their balanced state level for every single tumor. The balanced state term can be represented as $\ln X_i^o(k) = -G_{i0}\lambda_0(k)$ (7), allowing to calculate a weight for the balanced state, $\lambda_0(k)$, for every tumor *k* and the extent

of the participation of each individual transcript i, G_{i0}, in the balanced state process, $\alpha = 0$. We have previously shown that this balanced state is robust, and is common to normal and cancer tissue and even to different organisms, i.e. the transcripts participating in this process do not show any dependence on the patients. (4, 6, 7, 20). The experimental data we wished to analyze in this study originated from several different datasets. We expect that the expression level of transcript i in the balanced state, $X_i^o(k)$, is common to all patients and does not depend on the patient index, k.

The analysis further uncovers the complete set of constraints that operate in the system, including the transcripts that are affected by these constraints and thus deviate from their balanced state levels. A constraint can result from any perturbation in a biological system. Each constraint significantly influences only a subset of transcripts in a similar way, which causes collective deviations of the transcript levels (up or down) from their balanced level. This group of co-varying transcripts represents an altered transcript correlation subnetwork that we name *an unbalanced process*. The unbalanced processes are indexed by $\alpha = 1, 2, 3$. Each unbalanced process can consist of several biological pathways. For example, proteins involved in aerobic glycolysis and MAPK (Mitogen-activated protein kinases) signaling pathways can deviate in a coordinated manner from the balanced state and thus participate in the same unbalanced process (20).

Several unbalanced processes may operate in each tumor, and each transcript can participate in several unbalanced processes due to the non-linearity of biological networks (20).

Singular Value Decomposition, SVD (24–26), is used as a mathematical tool to determine the two sets of parameters that determine the unbalanced processes in surprisal analysis [7]: (a) The $\lambda_a(k)$ values, denoting the weight of unbalanced process, in every tumor k; (b) The G_{ia} values, denoting the extent of the participation of each individual transcript i in the specific unbalanced process, α (7). Transcripts with the highest/lowest G_{ia} values are used to determine the transcript composition of unbalanced processes (**Fig. S1**). To assign a biological meaning for each process transcripts with the most significant G_{ia} values are classified into biological categories according to Gene Ontology (GO) database (**Table S1**). Several biological categories appear in each process (**Table S1**). Note that the weight, G_{ia},

of transcript *i* in a process α is the same for all patients (i.e. is independent of *k*). Hence, the network structure, comprised of co-varying transcripts participating in process α remains constant. The weight, $\lambda_a(k)$, determines whether process α is active in the patient *k*, and to what extent.

Our goal was to utilize surprisal analysis to classify tumors according to the tumor-specific sets of constraints that deviate the cancer tissues from the stable, balanced state. We suggest that such a classification is essential to improve personalized cancer diagnostics.

Integrating biological datasets to study inter-patient heterogeneity

The field of personalized medicine has been accelerating and a massive amount of gene expression data regarding different types of cancer is becoming available. Five different datasets were selected for analysis, each comprising samples from a different type of cancer: lymphoma, bladder cancer, gastric cancer, colorectal cancer and breast cancer (527 in total: 506 tumor samples and 21 normal gastric samples). A concurrent analysis of different datasets will allow identification of the altered biological processes that characterize the inter-patient heterogeneity. Additionally, a large-scale analysis should uncover the patient-specific sets of unbalanced processes with a better signal-to-noise ratio.

As expected, surprisal analysis of the 5 datasets identified a *common* balanced state for each type of cancer, represented by an invariant amplitude of the balanced state $\lambda_0(k)$ for all patients, *k*, of a specific cancer type, including the normal gastric samples (**Fig. 1**, grey color) .This result corresponds to our previous findings demonstrating the robustness of the balanced process (4, 6, 20). The levels of more than 700 transcript probes out of ~20,000 probes were well-fitted by the balanced term alone and were not influenced by any unbalanced process. These transcripts participate in the homeostatic functions of the cell, such as protein and RNA metabolism, and the cell cycle (**Table S1**).

Following determination of the balanced state term separately for each dataset, the intensities of the different sets were normalized and converted to a common scale, such that all 5 datasets shared a common balanced state term (**Fig. 1**). Importantly, the transcript composition of the steady state remained invariant before and after the normalization, suggesting that the intensity differences reflected experimental artifacts and not biological differences (**Fig. S2**). The thermodynamic-based approach underlying surprisal analysis is

what enables to do such a normalization. For the technical details see SI. Importantly, we demonstrate that this normalization does *not* influence the weights of the unbalanced processes nor the weight of individual transcripts in these processes (see SI).

The notion that the balanced state is common to normal and cancerous tissues is highly significant, because it suggests that the search for the tumor gene markers should focus only on the unbalanced processes, greatly reducing the number of possible targets.

The inter-patient heterogeneity among 506 patients is characterized by 12 unbalanced processes

Our next step was to inspect the unbalanced processes that characterized the 506 tumors (not including the 21 normal gastric samples). The analysis revealed that 12 unbalanced processes sufficed to reproduce the deviations from the balanced state across the 506 tumors of 5 types (**Tables S1-S4**). We used three different methods to identify the number of unbalanced processes that characterize the inter-patient variability: (i) calculation of error limits that were based on the fluctuations in the expression levels of the most stable transcripts was used to determine which of the processes possess a weight that exceeds the noise threshold, (ii) error bars for each patient were computed as described previously (27) and (iii) to validate that the number of significant unbalanced processes (only those having amplitude values exciding error limits) is sufficient, we verify that these processes adequately reproduce the experimental data. These three methods are further discussed in the SI, with results shown in **Figures S3-S6**.

To verify the robustness and accuracy of the analysis we randomly picked 50% of the patients from each cancer type (264 patient total, representing about a half of the complete dataset), and found that the unbalanced processes and patient-specific signatures remained the same (**Fig. S9,S10**).

Transcripts can be involved in only one constraint, e.g. GRB2, PTK2B, and CALM3, whereas others participate in 2 or more unbalanced processes, such as EGFR, PD-L1 (CD274), CD44, IRS2, EIF4E, and CDK1 (**Table S3**). Table S1 shows that each unbalanced process can include multiple (sometimes overlapping) biological categories. Importantly, we found that in every cancer type, ***one or more*** *unbalanced processes are shared by all of the patients of this cancer type* (**Fig. 2A**). For example, all of the lymphoma patients were found to harbor the process $\alpha = 1$ with a positive amplitude ($\lambda_1(k) > 0$), which we define as

process 1+ (**Fig. 2A**). Process 2+ was found in all patients of lymphoma as well (**Fig. 2A**). Genes involved in these processes were classified to multiple categories, for example, B-cell signaling, cell proliferation, platelet deregulation and DNA repair (**Table S1**). Recall that the weights, $G_{i\alpha}$, are independent of the patient index, *k*, and that it is the weights, $\lambda_\alpha(k)$, that determines whether a process is active in the specific patient. The sign of $\lambda_\alpha(k)$ determines the direction to which the process deviates the transcripts. Thus, if all lymphoma patients harbor process 1+, it means that process 1 deviates the transcripts in the process in the same manner in all lymphoma patients, i.e. upregulates or downregulates them. Process 1- was found in all patients of bladder cancer (**Fig. 2A)**, and includes genes involved, for example, in intracellular signal transduction and GTPase activation (**Table S1**). Process 3+ appeared in all patients of gastric cancer (**Fig. 2A)**, and includes genes involved in angiogenesis and anti-apoptosis (**Table S1**). Process 2- appeared in all patients of colorectal cancer (**Fig. 2A**), and includes genes involved in IL4 and IL10 production, NFkB signaling (**Table S1**). The breast cancer patients were all found to harbor process 7+ (**Fig. 2A**), which includes VEGFR signaling and glucuronidation (a mechanism of intrinsic drug resistance) (**Table S1**). The finding that certain unbalanced processes are shared by <u>all</u> patients of a particular cancer type is consistent with our earlier findings that there is a dominant process that characterizes a particular type of cancer, as compared to normal samples (4, 6, 7). Note, however, that the same process may also appear in other cancer types, possibly less frequently. For example, process 3- is shared by lymphoma, bladder and colorectal cancers (**Fig. 2A**). This constraint includes transcripts involved, for example, in PGDFR signaling pathway, mRNA processing and splicing (**Table S1**). Process 5- appears in bladder, gastric and breast cancers and comprises transcripts involved in, for example, Wnt signaling, cell-cell adhesion and RNA splicing (**Table S1**). Processes of higher index appear in a smaller number of patients (**Table S2**).

From unbalanced processes to patient-specific signatures

12 unbalanced processes repeat themselves across the 506 tumors. However, not all processes are active in all tumors. Every individual tumor harbors a specific subset, or signature, of active unbalanced processes (**Fig. 2B,C**; see also **Table S4**). Typically, every patient can be accurately represented by a combination of 1 to 5 ongoing processes (**Table S4**

and **Fig. 2B**). Table S4 contains the entire list of 144 patient-specific sets of unbalanced processes that are repeated across 506 tumors.

12 unbalanced processes can be assembled into thousands of unique subsets of 1-5 processes. We found varying degrees of inter-tumor heterogeneity in each of the tumor types (**Fig. 2C**): 17 combinations of processes were found in the population of 130 lymphoma patients; 80 combinations were found in the population of 93 bladder cancer patients; 21 combinations in the population of 111 gastric cancer patients; 14 unique combinations in the population of 152 colorectal cancer patients; and 12 combinations of processes were found in the population of 20 breast cancer patients. Thus, some cancer types possess a high degree of heterogeneity (e.g. bladder), whereas others, such as colorectal cancer, are significantly less heterogeneous (**Fig. 2C**).

Similar gene expression levels can result from different combinations of unbalanced processes

One of the main features of surprisal analysis is its ability to assign transcripts to more than one unbalanced process (see above and (7)). For example, epidermal growth factor receptor (EGFR) was found to independently participate in processes 4, 5, 6 and 9; programmed death-ligand 1 (CD 274 (PD-L1), inhibitor of the immune system) participates in processes 5, 7, 8 and 10 (**Table S3**). Therefore, two patients can display similar gene expression levels, while their tumors may harbor different combinations of unbalanced processes. To demonstrate this point, we selected two bladder cancer patients, indexed 164 and 172, and inspected their tumor-specific experimental expression levels of 5 bladder-cancer associated genes: NFkBIA, the inhibitor of NFkB (28, 29), PD-L1 (30), CD44 (31), EGFR (32), and PLAU (33) (**Fig. 3A**). In both tumors, these biomarkers were upregulated relative to their median expression level (**Fig. 3A**). However, surprisal analysis revealed that the tumors are biologically different: patient 164 is characterized by a combination of processes 1, 2, 5 and 7, whereas patient 172 harbors a combination of processes 1, 2, 5, 9, and 10 (**Fig. 3B**, **Table S3**). Figure 3C shows 19 selected genes, and how they were affected by these unbalanced processes in the two patients. In both patients, the induction of NFkBIA is associated with unbalanced process 2 (**Fig. 3B,C**), and the induction of CD44 is associated with processes 2 and 5 (**Fig. 3B,C**). However, the induction of other oncogenes was attributed to <u>different</u> processes. For example, in patient 164, PD-L1 induction was attributed to processes 5 and 7, while in patient 172 PD-L1 induction was attributed to processes 5 and

10 (**Fig. 3B,C**). Similarly, EGFR was induced by process 5 in patient 164, while in patient 172 it was induced due to processes 5 and 9 (**Fig. 3B,C**).

The full lists of G_{ia} values, representing the extent of the participation of each transcript in processes $\alpha = 1, 2, ..., 12$ are presented in **Table S3**.

Patients 164 and 172 serve as an example for two patients carrying tumors of the same type, which may present with similar lists of oncogenic biomarkers, even though their tumors are not the same. Classification of tumors according to similar biomarkers, may lead to significant differences between cancer patients in terms of response to treatment, survival prediction, and more. Deciphering the complete altered transcriptional network in every tumor should enable more accurate diagnosis and classification of patients.

The 12 unbalanced processes identified are active in other cancer patients

Our next step was to verify whether the 12 unbalanced processes that were identified in the 506 tumors are relevant to other cancer patients as well. To answer this, we obtained an additional dataset, which consists of 39 pancreatic tumors. This additional dataset will be referred to as the validation set. The dataset was merged with the previously analyzed 5 datasets (utilizing the normalization method described above), and the combined dataset, comprising 566 patients, was analyzed using surprisal analysis (**Fig. 4**). 13 unbalanced processes were identified in this analysis. Strikingly, the first 12 unbalanced processes appeared to be the same 12 unbalanced processes that were identified in the analysis of the original 527 samples (**Fig. S7,8**) and could fully characterized 36% of the pancreatic patients (**Fig. 4B** ,**Table S5**). The 13th process, which appeared only upon addition of the validation set to the analysis, was essential for the characterization of the remaining ~64% of pancreatic patients (**Fig. 4C, Table S5**). This process did not appear in the original dataset, suggesting that it is a pancreatic cancer-specific constraint. The transcripts involved in unbalanced process 13 were categorized, among others, to the Notch, IL-1, NFkB and EGFR signaling pathways (**Table S5**). These pathways were shown to be involved in pancreatic cancer (34–37).

Interestingly, unbalanced processes 1+ and 3- appeared active in all pancreatic patients (**Fig. 4D**). Unbalanced process 11, which was found in 14 patients of bladder cancer (**Fig. 2A**), was active in ~28% of the validation pancreatic set (**Fig. 4D**). Process 12, which represented only bladder cancer patients previously (**Fig.2A**), was found only in 1 pancreatic

patient (**Fig. 4D**). Overall, 16 different combinations of unbalanced processes were found in 39 pancreatic patients, demonstrating a relatively high inter-patient heterogeneity (**Fig. 4E, Table S5**).

Discussion

Personalized medicine aims to sub-divide patients into different categories based on molecular level information. Such a classification often uses biomarker lists to make more informed medical decisions regarding patient diagnosis or treatment. In this study we expand the idea of gene biomarker profiling and show that integration of biomarkers into tumor ongoing unbalanced processes is critical for accurate identification of patient-specific cancer biology.

To obtain exhaustive patient signatures we assembled a large-scale patient dataset of transcript expression levels comprising different cancer types. We showed that the notion of the balanced state allows us to integrate datasets from different experiments into one big dataset thereby increasing the amount of information that can be extracted from the experimental data collected by different groups. The approach can therefore be used in a large spectrum of different studies that require large-scale analysis and integration of various molecular datasets.

Using 506 tumors, from 5 different cancer types, we showed that this diverse collection of tumors can be altogether characterized by only 12 unbalanced processes. The majority of unbalanced processes spanned across different cancer types (**Fig. 2A**), suggesting that unbalanced processes can often be independent of cancer type.

The heterogeneity among the cancer patients was attributed to *different patient-specific combinations of 1-5 unbalanced processes* out of 12, leading eventually to 144 different combinations that represent 144 types of disease. The fact that 144 different diseases can be characterized by only 12 processes makes our approach towards personalized diagnostics quite effective.

Each patient-specific signature is a set of a small number of unbalanced processes thereby offering a considerable compaction of the data. The compaction means that there is a limited set of processes characterizing the entire data set. Furthermore each patient is represented by a subset of those processes.

The result, that each person harbors usually several, and not only one, unbalanced processes may emanate from the co-existence of different intra-tumor cellular subpopulations in each patient (that may vary from patient to patient) or/and due to a parallel operation of the different processes in the same tumor cells. Single cell analysis of a large number of patients will be able to accurately address this topic.

We propose that our? approach can provide a guidance for the patient-specific combined therapies, targeting distinct unbalanced signaling processes in each patient (Flashner-Abramson et al., in preparation).

We validated the approach by adding an independent pancreatic cancer data set. We show that the same set of 12 unbalanced processes remained valid. 36% of the pancreatic patients from the additional dataset were found to harbor different combinations of the previous 12 unbalanced processes. However, the remaining 64% of the pancreatic patients were not fully characterized by the previous 12 unbalanced processes, and were found to harbor an additional, pancreatic cancer-specific unbalanced process, indexed 13. This is consistent with our earlier finding that each cancer has a cancer type-specific dominant process.

The approach we present herein enables extraction of significant signals from large datasets, and gaining in depth, unbiased, patient-specific information. Surprisal analysis efficiently uncovers the altered transcriptional networks in every individual patient, potentially allowing improved classification of cancer patients. Our finding that similar oncogene expression levels in different patients may stem from distinct sets of unbalanced processes underscores the need to extend the initial analysis of tumors and to increase the resolution of cancer patient diagnosis.

Methods

Surprisal analysis The analysis was carried out as described before (5–7) and in the main text. Datasets used in the study were obtained from GEO database as indicated in the manuscript. Notion of the stable steady state, which was determined separately for each dataset, allowed us to integrate different datasets in one large matrix. This matrix was analyzed further to determine the unbalanced processes characterizing 5 different cancer types. For more details see the main text and SI.

Calculation of error bars and threshold values To find significant unbalanced processes characterizing each patient and then to calculate a patient specific set of the processes we calculated error bars for the amplitudes of the processes *in each sample* (27) as well as a threshold limit for each type of cancer using stable transcripts, representing baseline fluctuations in the population, as described previously (27) and in SI.

Calculation of patient specific combinations of unbalanced processes Combinations presented in Tables S3, were generated using $\lambda_\alpha(k)$ ($\alpha = 1,2,3,...$) values that exceeded threshold limits and had error bars above 0 (See SI for more details).

Acknowledgements

The funding sources for this work were from Abisch-Frenkel foundation (NKB). FR thanks the Fonds National de la Recherche Scientifique (FNRS -Belgium) for its support.

References

1. Alizadeh AA, et al. (2015) Toward understanding and exploiting tumor heterogeneity. *Nat Med* 21(8):846–853.
2. Weinstein IB (2002) Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science* 297(5578):63–4.
3. Chibon F (2013) Cancer gene expression signatures – The rise and fall? *Eur J Cancer* 49(8):2000–2009.
4. Zadran S, Remacle F, Levine RD (2013) miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. *Proc Natl Acad Sci U S A* 110(47):19160–19165.
5. Kravchenko-Balasha N, et al. (2011) Convergence of logic of cellular regulation in different premalignant cells by an information theoretic approach. *BMC Syst Biol* 5:42.
6. Kravchenko-Balasha N, et al. (2012) On a fundamental structure of gene networks in living cells. *Proc Natl Acad Sci U S A* 109(12):4702–4707.

7. Remacle F, Kravchenko-Balasha N, Levitzki A, Levine RD (2010) Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. *Proc Natl Acad Sci U S A* 107(22):10324–10329.

8. Poovathingal SKK, Kravchenko-Balasha N, Shin YSS, Levine RDD, Heath JRR (2016) Critical Points in Tumorigenesis: A Carcinogen-Initiated Phase Transition Analyzed via Single-Cell Proteomics. *Small* 12(11):1425–31.

9. Sokolovski M, Bhattacherjee A, Kessler N, Levy Y, Horovitz A (2015) Thermodynamic Protein Destabilization by GFP Tagging: A Case of Interdomain Allostery. *Biophys J* 109(6):1157–1161.

10. Lucia U (2013) Thermodynamics and cancer stationary states. *Phys A Stat Mech its Appl* 392(17):3648–3653.

11. Haynie DT (2008) *Biological thermodynamics: Second edition* doi:10.1017/CBO9780511802690.

12. Salamon P, Wootton JC, Konopka AK, Hansen LK (1993) On the robustness of maximum entropy relationships for complexity distributions of nucleotide sequences. *Comput Chem* 17(2):135–148.

13. Renken C, et al. (2002) A thermodynamic model describing the nature of the crista junction: A structural motif in the mitochondrion. *Journal of Structural Biology*, pp 137–144.

14. Nykter M, et al. (2008) Critical networks exhibit maximal information diversity in structure-dynamics relationships. *Phys Rev Lett* 100(5). doi:10.1103/PhysRevLett.100.058702.

15. Waltermann C, Klipp E (2011) Information theory based approaches to cellular signaling. *Biochim Biophys Acta* 1810(10):924–932.

16. Levchenko A, Nemenman I (2014) Cellular noise and information transmission. *Curr Opin Biotechnol* 28:156–164.

17. Atwal GS, et al. (2008) An information-theoretic analysis of genetics, gender and age in cancer patients. *PLoS One* 3(4). doi:10.1371/journal.pone.0001951.

18. Levine RD (2005) *Molecular Reaction Dynamics* (The University Press, Cambridge).

19. Levine RD, Bernstein RB (1974) Energy disposal and energy consumption in elementary chemical reactions. Information theoretic approach. *Acc Chem Res* 7(12):393–400.

20. Kravchenko-Balasha N, Johnson H, White FMFM, Heath JRJR, Levine RDD (2016) A Thermodynamic-Based Interpretation of Protein Expression Heterogeneity in Different Glioblastoma Multiforme Tumors Identifies Tumor-Specific Unbalanced Processes. *J Phys Chem B* 120(26):5990–7.

21. Kravchenko-Balasha N, Shin YS, Sutherland A, Levine RD, Heath JR (2016) Intercellular signaling through secreted proteins induces free-energy gradient-directed cell movement. *Proc Natl Acad Sci* 113(20):5520–5.

22. Facciotti MT (2013) Thermodynamically inspired classifier for molecular phenotypes of health and disease. *Proc Natl Acad Sci U S A* 110(48):19181–19182.

23. Kravchenko-Balasha N, Wang J, Remacle F, Levine RDD, Heath JRR (2014) Glioblastoma cellular architectures are predicted through the characterization of two-cell interactions. *Proc Natl Acad Sci U S A* 111(17):6521–6526.

24. Alter O, Brown PO, Botstein D (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A* 97(18):10101–10106.

25. Aiello KA, Alter O (2016) Platform-independent genome-wide pattern of DNA copy-number alterations predicting astrocytoma survival and response to treatment revealed by the GSVD formulated as a comparative spectral decomposition. *PLoS One* 11(10). doi:10.1371/journal.pone.0164546.

26. Sankaranarayanan P, Schomay TE, Aiello KA, Alter O (2015) Tensor GSVD of Patient- and Platform-Matched Tumor and Normal DNA Copy-Number Profiles Uncovers Chromosome Arm-Wide Patterns of Tumor-Exclusive Platform-Consistent Alterations Encoding for Cell Transformation and Predicting Ovarian Cancer Survival. *PLoS One* 10(4):e0121396.

27. Gross A, Levine RD (2013) Surprisal analysis of transcripts expression levels in the presence of noise: a reliable determination of the onset of a tumor phenotype. *PLoS One* 8(4):e61554.

28. Zhu J, et al. (2017) NF-κB p65 Overexpression Promotes Bladder Cancer Cell Migration via FBW7-Mediated Degradation of RhoGDIα Protein. *Neoplasia* 19(9):672–683.

29. Ito Y, et al. (2015) Down-regulation of NF kappa B activation is an effective therapeutic modality in acquired platinum-resistant bladder cancer. *BMC Cancer* 15:324.

30. Bellmunt J, Powles T, Vogelzang NJ (2017) A review on the evolution of PD-1/PD-L1 immunotherapy for bladder cancer: The future is now. *Cancer Treat Rev* 54:58–67.

31. Wu C-T, Lin W-Y, Chang Y-H, Chen W-C, Chen M-F (2017) Impact of CD44 expression on radiation response for bladder cancer. *J Cancer* 8(7):1137–1144.

32. Colquhoun AJ, Mellon JK (2002) Epidermal growth factor receptor and bladder cancer. *Postgrad Med J* 78(924):584–589.

33. Hasui Y, Nishi S, Kitada S, Osada Y (1993) UROKINASE-TYPE PLASMINOGEN

ACTIVATOR ANTIGEN AS A PROGNOSTIC FACTOR IN BLADDER CANCER. *Japanese J Urol* 84(9):1624–1628.

34. Avila JL, Kissil JL (2013) Notch signaling in pancreatic cancer: oncogene or tumor suppressor? *Trends Mol Med* 19(5):320–327.

35. Prabhu L, Mundade R, Korc M, Loehrer PJ, Lu T (2014) Critical role of NF-κB in pancreatic cancer. *Oncotarget* 5(22):10969–75.

36. Zhuang Z, et al. (2016) IL1 Receptor Antagonist Inhibits Pancreatic Cancer Growth by Abrogating NF- B Activation. *Clin Cancer Res* 22(6):1432–1444.

37. Troiani T, et al. (2012) Targeting EGFR in pancreatic cancer treatment. *Curr Drug Targets* 13(6):802–810.