

RESEARCH ARTICLE

# 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

Preethi Sankaranarayanan<sup>1,2‡</sup>, Theodore E. Schomay<sup>1,2‡</sup>, Katherine A. Aiello<sup>1,2</sup>, Orly Alter<sup>1,2,3\*</sup>

**1** Scientific Computing and Imaging (SCI) Institute, University of Utah, Salt Lake City, Utah, United States of America, **2** Department of Bioengineering, University of Utah, Salt Lake City, Utah, United States of America, **3** Department of Human Genetics, University of Utah, Salt Lake City, Utah, United States of America

‡ These authors contributed equally to this work.

\* [orly@sci.utah.edu](mailto:orly@sci.utah.edu)



OPEN ACCESS

**Citation:** 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. doi:10.1371/journal.pone.0121396

**Academic Editor:** Jörg D. Hoheisel, Deutsches Krebsforschungszentrum, GERMANY

**Received:** October 21, 2014

**Accepted:** January 31, 2015

**Published:** April 15, 2015

**Copyright:** © 2015 Sankaranarayanan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was supported by the Utah Science, Technology, and Research (USTAR) Initiative, National Human Genome Research Institute (NHGRI) R01 Grant HG-004302 and National Science Foundation (NSF) CAREER Award DMS-0847173 (to OA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The number of large-scale high-dimensional datasets recording different aspects of a single disease is growing, accompanied by a need for frameworks that can create one coherent model from multiple tensors of matched columns, e.g., patients and platforms, but independent rows, e.g., probes. We define and prove the mathematical properties of a novel tensor generalized singular value decomposition (GSVD), which can simultaneously find the similarities and dissimilarities, i.e., patterns of varying relative significance, between any two such tensors. We demonstrate the tensor GSVD in comparative modeling of patient- and platform-matched but probe-independent ovarian serous cystadenocarcinoma (OV) tumor, mostly high-grade, and normal DNA copy-number profiles, across each chromosome arm, and combination of two arms, separately. The modeling uncovers previously unrecognized patterns of tumor-exclusive platform-consistent co-occurring copy-number alterations (CNAs). We find, first, and validate that each of the patterns across only 7p and Xq, and the combination of 6p+12p, is correlated with a patient's prognosis, is independent of the tumor's stage, the best predictor of OV survival to date, and together with stage makes a better predictor than stage alone. Second, these patterns include most known OV-associated CNAs that map to these chromosome arms, as well as several previously unreported, yet frequent focal CNAs. Third, differential mRNA, microRNA, and protein expression consistently map to the DNA CNAs. A coherent picture emerges for each pattern, suggesting roles for the CNAs in OV pathogenesis and personalized therapy. In 6p+12p, deletion of the

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

p21-encoding *CDKN1A* and p38-encoding *MAPK14* and amplification of *RAD51AP1* and *KRAS* encode for human cell transformation, and are correlated with a cell's immortality, and a patient's shorter survival time. In 7p, *RPA3* deletion and *POLD2* amplification are correlated with DNA stability, and a longer survival. In Xq, *PABPC5* deletion and *BCAP31* amplification are correlated with a cellular immune response, and a longer survival.

## Introduction

The growing number of large-scale high-dimensional datasets recording different aspects of a single disease promise to enhance basic understanding of life on the molecular level as well as medical diagnosis, prognosis, and treatment. This is accompanied by a fundamental need for mathematical frameworks that can create one coherent model from multiple datasets arranged in multiple order-matched, column-matched, and row-independent tensors, i.e., tensors of the same number of dimensions each, with one-to-one mappings among the columns across all but one of the corresponding dimensions among the tensors, but not necessarily among the rows across the one remaining dimension in each tensor. Consider, e.g., the structure of the DNA copy-number datasets in the Cancer Genome Atlas (TCGA) [1, 2]. Profiles of tumor and normal tissues from the same set of patients have the structure of two matrices, i.e., second-order tensors, with a one-to-one mapping between the columns that correspond to the same set of patients, but not necessarily between the rows that correspond to the DNA copy-number probes with valid data in either the tumor or the normal dataset, and may be different. When the tumor and normal profiles are measured in replicates, e.g., by the same set of profiling platforms, then the structure of the tumor and normal datasets is that of two third-order tensors, of matched columns that correspond to the same sets of patients and platforms, and independent rows that correspond to the probes in either the tumor or the normal dataset.

The higher-order generalized singular value decomposition (HO GSVD) is the only simultaneous decomposition to date of more than two such column-matched but row-independent datasets, which is by definition exact, and which mathematical properties allow interpreting its variables and operations in terms of the similar as well as dissimilar, e.g., biomedical reality among the datasets [3, 4]. The HO GSVD generalizes the GSVD [5–12], which was demonstrated in comparative modeling of, e.g., patient-matched but probe-independent glioblastoma (GBM) brain tumor and normal DNA copy-number profiles from TCGA [13]. The modeling uncovered a previously unrecognized genome-wide pattern of tumor-exclusive copy-number alterations (CNAs). Prior to the modeling, DNA copy-number subtypes of GBM predictive of survival and response to chemotherapy were not conclusively identified [14, 15], and the best predictor of GBM survival was the patient's age at diagnosis [16, 17]. Survival analyses [18, 19] showed and validated that the pattern is correlated with a GBM patient's prognosis and response to chemotherapy, is independent of age, and together with age makes a better predictor than age alone. Segmentation [20, 21] of the pattern showed that it includes most known GBM-associated changes in chromosome numbers and focal CNAs, as well as several previously unreported, yet frequent CNAs. This suggested that the pattern is not only correlated, but also possibly causally coordinated with the GBM tumor's pathogenesis. Previously unrecognized targets for personalized GBM drug therapy were also suggested, the tousel-like kinase 2 *TLK2* and the methyltransferase-like 2A *METTL2A* [22–24]. The GSVD comparative modeling, therefore, resulted in new insights into the poorly understood relations between a GBM tumor's genome and a patient's survival phenotype.

The GSVD and HO GSVD, however, are limited to datasets arranged in second-order tensors, i.e., matrices. We define, therefore, a novel tensor GSVD, i.e., an exact simultaneous decomposition of two datasets, arranged in two higher-than-second-order tensors of matched column dimensions but independent row dimensions. The tensor GSVD factors or separates the pair of tensors into corresponding pairs of “subtensors”, i.e., pairs of outer products or combinations of a paired set of patterns each: patterns, one across each of the matched column dimensions, which are identical for both tensors, combined with one pattern across the independent row dimension of either one of the two tensors. The pairs of subtensors are of varying relative mathematical significance, i.e., the significance of one subtensor in a pair in the corresponding tensor relative to the significance of the second subtensor in the second tensor varies among the pairs of subtensors. We prove that the tensor GSVD extends the GSVD and the tensor higher-order singular value decomposition (HOSVD) [25–28] from a decomposition of either two column-matched matrices or one tensor, respectively, to a decomposition of two order-matched, column-matched, and row-independent tensors [29]. We also show that the mathematical properties of the tensor GSVD allow interpreting the subtensors in terms of the biomedical similarities and dissimilarities between the two corresponding high-dimensional datasets.

We demonstrate the tensor GSVD in comparative modeling of patient- and platform-matched but probe-independent ovarian serous cystadenocarcinoma (OV) tumor and normal DNA copy-number profiles from TCGA. Most of the tumors, i.e., >95%, are high-grade tumors [30]. OV accounts for about 90% of all ovarian cancers. Despite recent large-scale profiling efforts, the best predictor of OV survival to date has remained the tumor’s stage at diagnosis, a pathological assessment of the spread of the cancer numbering I to IV [31]. About 25% of primary OV tumors are resistant, and most recurrent OV tumors develop resistance to platinum-based chemotherapy, the first-line treatment for more than 30 years now [32]. Even though there exist drugs for platinum-based chemotherapy-resistant OV tumors, no pathology laboratory diagnostic exists that distinguishes between resistant and sensitive tumors before the treatment [33]. OV tumors exhibit significant CNA variation among them, much more so than, e.g., GBM tumors, and very few frequent CNAs typical of OV have been identified so far. We, therefore, model the profiles across each chromosome arm, and each combination of two chromosome arms, separately. The modeling uncovers previously unrecognized chromosome arm-wide patterns of tumor-exclusive and platform-consistent co-occurring CNAs.

By using survival analyses of the discovery and, separately, validation set of patients, as well as only the platinum-based chemotherapy patients in the discovery and validation sets, we find, first, and validate that each of the patterns across only the chromosome arms 7p and Xq, and across only the combination of the two chromosome arms 6p+12p (but not 6p nor 12p separately), is correlated with an OV patient’s prognosis and response to platinum-based chemotherapy, is independent of stage, and together with stage makes a better predictor than stage alone. By using survival analyses of only the > 95% patients with high-grade tumors, we find and validate that these patterns are also independent of the OV tumor’s grade. We observe three groups of significantly different prognoses among the patients classified by a combination of the 6p+12p, 7p, and Xq tensor GSVD classifications, suggesting a possible implementation of the patterns in a pathology laboratory test. Second, by using segmentation of the 6p+12p, 7p, and Xq patterns, we find that the amplifications and deletions identified by these patterns include most known OV-associated CNAs that map to these chromosome arms [34], as well as several previously unreported, yet frequent focal CNAs [35–38]. Third, by using gene ontology enrichment analyses of the OV tumor mRNA expression profiles of the patients [39, 40], we find that differential mRNA expression between the patients, classified by any one of the three tensor GSVDs, is enriched in ontologies corresponding to one of three hallmarks of cancer [41]: a cell’s immortality in 6p+12p, DNA instability in 7p, and cellular immune response

suppression in Xq. The differential mRNA expression of genes from these enriched ontologies that are located on any one of the chromosome arms is consistent with the CNAs across that arm. Genes that map to amplifications or deletions on any one pattern, are overexpressed or underexpressed, respectively, in the patients which tumor profiles are classified as highly similar to that pattern. The differential expression of all microRNAs and proteins that map to any one of the chromosome arms is also consistent with the CNAs across that arm.

Taken together, a coherent picture emerges for each of these previously unrecognized chromosome arm-wide patterns of tumor-exclusive and platform-consistent co-occurring alterations, suggesting roles for the DNA CNAs in OV pathogenesis in addition to personalized diagnosis, prognosis, and treatment. In 6p+12p, loss of the p21-encoding *CDKN1A* and the p38-encoding *MAPK14* on 6p, and gain of *KRAS* on 12p, combined but not separately, can lead to transformation of human normal to tumor cells [42, 43]. These transformation-encoding CNAs, together with deletion of *TNF* on 6p, and amplification of *RAD51AP1* and *ITPR2* on 12p, are correlated with a suppression of cell cycle arrest, senescence, and apoptosis, i.e., a tumor cell’s immortality, and a patient’s shorter survival time [44–55]. Note that there already exist drugs that interact with *CDKN1A*, *MAPK14*, and *RAD51AP1*, even though these genes were not recognized previously as targets for OV drug therapy [56]. In 7p, *RPA3* deletion and *POLD2* amplification are correlated with DNA repair during replication, i.e., DNA stability, and a longer survival time [57, 58]. In Xq, *PABPC5* deletion and *BCAP31* amplification are correlated with a cellular immune response, and a longer survival time [59].

## Mathematical Method: Tensor GSVD

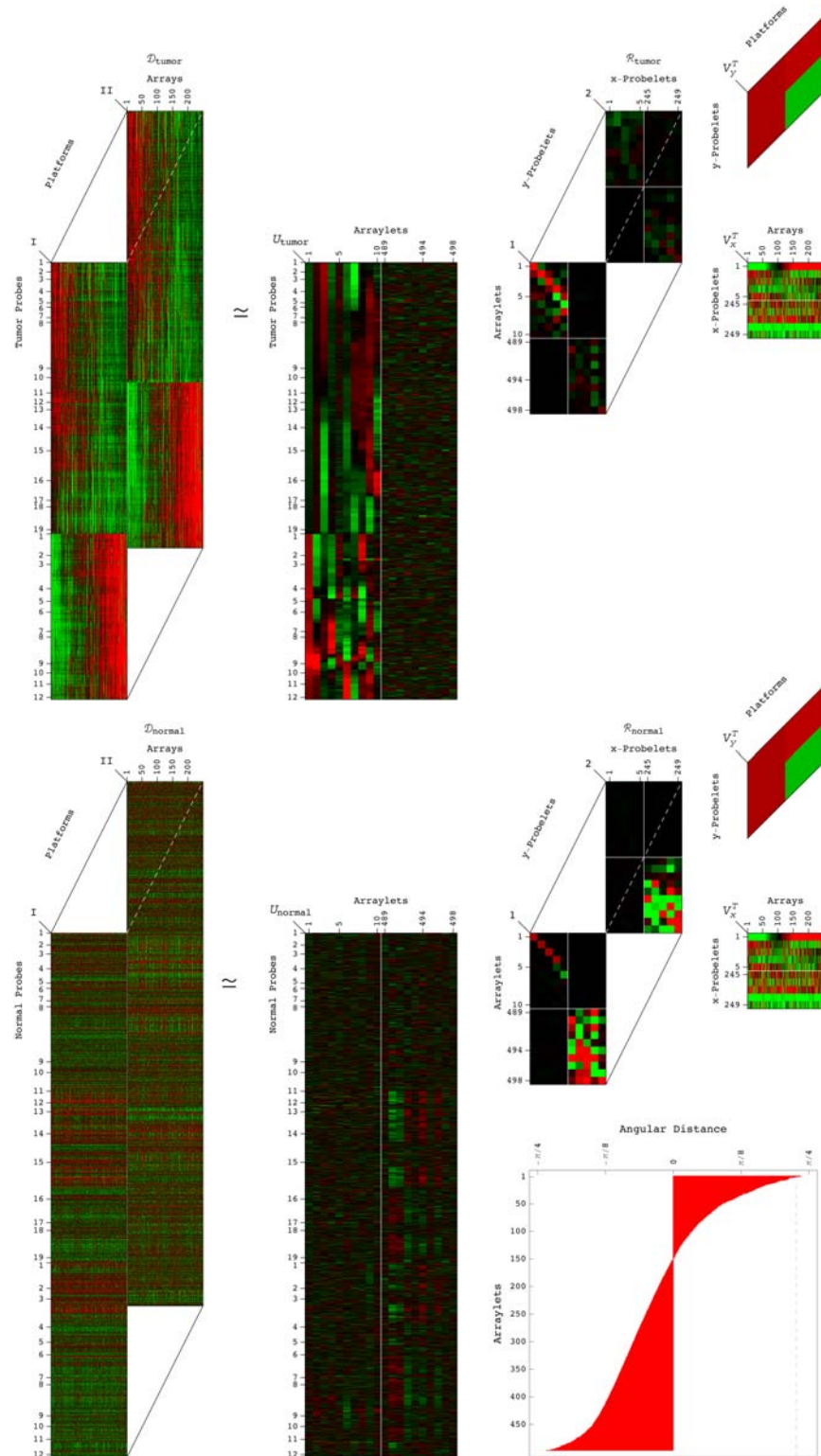
### Discovery Datasets are Pairs of Column-Matched but Row-Independent Tensors

We selected primary OV tumor and normal DNA copy-number profiles of a set of 249 TCGA patients [2] (Sec. 1.1 in [S1 Appendix](#), and [S1 Dataset](#)). Each profile was measured in two replicates by the same set of two DNA microarray platforms. For each chromosome arm or combination of two chromosome arms, the structure of these tumor and normal discovery datasets  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , of  $K_1$ -tumor and  $K_2$ -normal probes  $\times$   $L$ -patients, i.e., arrays  $\times$   $M$ -platforms, is that of two third-order tensors with one-to-one mappings between the column dimensions  $L$  and  $M$ , but different row dimensions  $K_1$  and  $K_2$ , where  $K_1, K_2 \geq LM$ .

### The Tensor GSVD

We define, therefore, a novel tensor GSVD that simultaneously separates the paired datasets into weighted sums of  $LM$  paired “subtensors”, i.e., combinations or outer products of three patterns each: Either one tumor-specific pattern of copy-number variation across the tumor probes, i.e., a “tumor arraylet”  $u_{1,a}$ , or the corresponding normal-specific pattern across the normal probes, i.e., the “normal arraylet”  $u_{2,a}$ , combined with one pattern of copy-number variation across the patients, i.e., an “ $x$ -probelet”  $v_{x,b}^T$  and one pattern across the platforms, i.e., a “ $y$ -probelet”  $v_{y,c}^T$ , which are identical for both the tumor and normal datasets ([Fig. 1](#), and Figs. A and B in [S1 Appendix](#)),

$$\begin{aligned}
 \mathcal{D}_i &= \mathcal{R}_i \times_a U_i \times_b V_x \times_c V_y \\
 &= \sum_{a=1}^{LM} \sum_{b=1}^L \sum_{c=1}^M \mathcal{R}_{i,abc} S_i(a, b, c), \\
 S_i(a, b, c) &= u_{i,a} \otimes v_{x,b}^T \otimes v_{y,c}^T, \quad i = 1, 2,
 \end{aligned} \tag{1}$$



**Fig 1. Tensor generalized singular value decomposition (GSVD) of the patient- and platform-matched DNA copy-number profiles of the 6p+12p chromosome arms.** For each chromosome arm or combination of two chromosome arms, the structure of the tumor and normal discovery datasets ( $\mathcal{D}_1$  and  $\mathcal{D}_2$ ) is that of two third-order tensors with one-to-one mappings between the column dimensions but different row dimensions. The patients, platforms, probes, and tissue types, each represent a degree of freedom. Unfolded into a single

matrix, some of the degrees of freedom are lost and much of the information in the datasets might also be lost. We define a tensor GSVD that simultaneously separates the paired datasets into weighted sums of paired subtensors, i.e., combinations or outer products of three patterns each: Either one tumor-specific pattern of copy-number variation across the tumor probes, i.e., a tumor arraylet (a column basis vector of  $U_1$ ), or the corresponding normal-specific arraylet (a column basis vector of  $U_2$ ), combined with one pattern of variation across the patients, i.e., an  $x$ -probelet (a row basis vector of  $V_x^T$ ), and one pattern across the platforms, i.e., a  $y$ -probelet (a row basis vector of  $V_y^T$ ), which are identical for both the tumor and normal datasets (Equation 1). The tensor GSVD is depicted in a raster display, with relative copy-number gain (red), no change (black), and loss (green), explicitly showing the first through the 5th, and the 245th through the 249th  $6p+12p$   $x$ -probelets, both  $6p+12p$   $y$ -probelets, and the first through the 10th, and the 489th through the 498th  $6p+12p$  tumor and normal arraylets. We prove that the significance of a subtensor in the tumor dataset relative to that of the corresponding subtensor in the normal dataset, i.e., the tensor GSVD angular distance, equals the row mode GSVD angular distance, i.e., the significance of the corresponding tumor arraylet in the tumor dataset relative to that of the normal arraylet in the normal dataset. The tensor GSVD angular distances for the 498 pairs of  $6p+12p$  arraylets are depicted in a bar chart display, where the angular distance corresponding to the first pair of arraylets is  $\sim \pi/4$ . For the  $6p+12p$  combination of two chromosome arms, we find that the most significant subtensor in the tumor dataset (which corresponds to the coefficient of largest magnitude in  $\mathcal{R}_1$ ) is a combination of (i) the first  $y$ -probelet, which is approximately invariant across the platforms, (ii) the first  $x$ -probelet, which classifies the discovery set of patients into two groups of high and low coefficients, of significantly and robustly different prognoses, and (iii) the first, most tumor-exclusive tumor arraylet, which classifies the validation set of patients into two groups of high and low correlations of significantly different prognoses consistent with the  $x$ -probelet's classification of the discovery set.

doi:10.1371/journal.pone.0121396.g001

where  $\times_a U_i$ ,  $\times_b V_x$  and  $\times_c V_y$  denote tensor-matrix multiplications, which contract the  $LM$ -arraylet,  $L$ - $x$ -probelet, and  $M$ - $y$ -probelet dimensions of the "core tensor"  $\mathcal{R}_i$  with those of  $U_i$ ,  $V_x$ , and  $V_y$ , respectively, and where  $\otimes$  denotes an outer product.

**Construction.** Suppose that unfolding (or matricizing) both tensors  $\mathcal{D}_i$  into matrices, each preserving the  $K_i$ -row dimension, e.g., by appending the  $LM$  columns  $\mathcal{D}_{i,lm}$  of the corresponding tensor, gives two full column-rank matrices  $D_i \in \mathbb{R}^{K_i \times LM}$ . We obtain the column bases vectors  $U_i$  from the GSVD of  $D_i$  [5–13], i.e., the "row mode GSVD"

$$D_i = (\dots, \mathcal{D}_{i,lm}, \dots) = U_i \Sigma_i V^T, \quad i = 1, 2. \tag{2}$$

Suppose, similarly, that unfolding both tensors  $\mathcal{D}_i$  into matrices, each preserving the  $L$ - $x$ - (or  $M$ - $y$ -) column dimension, e.g., by appending the  $K_i M$  rows  $\mathcal{D}_{i,k;m}^T$  (or the  $K_i L$  rows  $\mathcal{D}_{i,k;l}^T$ ) of the corresponding tensor, gives two full column-rank matrices  $D_{ix} \in \mathbb{R}^{K_i M \times L}$  (or  $D_{iy} \in \mathbb{R}^{K_i L \times M}$ ). We obtain the  $x$ - (or  $y$ -) row basis vectors  $V_x^T$  (or  $V_y^T$ ), from the GSVD of  $D_{ix}$  (or  $D_{iy}$ ), i.e., the  $x$ - (or  $y$ -) column mode GSVD,

$$\begin{aligned} D_{ix} &= (\dots, \mathcal{D}_{i,k;m}^T, \dots) = U_{ix} \Sigma_{ix} V_x^T, \\ D_{iy} &= (\dots, \mathcal{D}_{i,k;l}^T, \dots) = U_{iy} \Sigma_{iy} V_y^T, \quad i = 1, 2. \end{aligned} \tag{3}$$

Note that the  $x$ - and  $y$ -row bases vectors are, in general, non-orthogonal but normalized, and  $V_x$  and  $V_y$  are invertible. The column bases vectors are normalized and orthogonal, i.e., uncorrelated, such that  $U_i^T U_i = I$ .

The generalized singular values are positive, and are arranged in  $\Sigma_i$ ,  $\Sigma_{ix}$ , and  $\Sigma_{iy}$  in decreasing orders of the corresponding "GSVD angular distances", i.e., decreasing orders of the ratios  $\sigma_{1,a}/\sigma_{2,a}$ ,  $\sigma_{1x,b}/\sigma_{2x,b}$ , and  $\sigma_{1y,c}/\sigma_{2y,c}$  respectively. We then compute the core tensors  $\mathcal{R}_i$  by contracting the row-,  $x$ -, and  $y$ -column dimensions of the tensors  $\mathcal{D}_i$  with those of the matrices  $U_i$ ,  $V_x^{-1}$ , and  $V_y^{-1}$ , respectively. For real tensors, the "tensor generalized singular values"  $\mathcal{R}_{i,abc}$  tabulated in the core tensors are real but not necessarily positive. Our tensor GSVD construction generalizes the GSVD to higher orders in analogy with the generalization of the singular value

decomposition (SVD) by the HOSVD [25–28], and is different from other approaches to the decomposition of two tensors [29].

**Existence, uniqueness and special cases.** We prove that our tensor GSVD exists for two tensors of any order because it is constructed from the GSVDs of the tensors unfolded into full column-rank matrices (Lemma A in S1 Appendix). The tensor GSVD has the same uniqueness properties as the GSVD, where the column bases vectors  $u_{i,a}$  and the row bases vectors  $v_{x,b}^T$  and  $v_{y,c}^T$  are unique, except in degenerate subspaces, defined by subsets of equal generalized singular values  $\sigma_{i,a}$ ,  $\sigma_{ix,b}$ , and  $\sigma_{iy,c}$  respectively, and up to phase factors of  $\pm 1$ , such that each vector captures both parallel and antiparallel patterns (Lemma B in S1 Appendix). The tensor GSVD of two second-order tensors reduces to the GSVD of the corresponding matrices (Corollary A in S1 Appendix). The tensor GSVD of the tensor  $\mathcal{D}_1 \in \mathbb{R}^{LM \times L \times M}$ , which row mode unfolding gives the identity matrix  $D_1 = I \in \mathbb{R}^{LM \times LM}$ , and a tensor  $\mathcal{D}_2$  of the same column dimensions reduces to the HOSVD of  $\mathcal{D}_2$  (Theorem A in S1 Appendix).

**Interpretation.** The significance of the subtensor  $\mathcal{S}_i(a,b,c)$  in the tensor  $\mathcal{D}_i$  is defined proportional to the magnitude of the corresponding tensor generalized singular values  $\mathcal{R}_{i,abc}$  (Fig. C in S1 Appendix), in analogy with the HOSVD,

$$\mathcal{P}_{i,abc} = \mathcal{R}_{i,abc}^2 / \sum_{a=1}^{LM} \sum_{b=1}^L \sum_{c=1}^M \mathcal{R}_{i,abc}^2, \quad i = 1, 2. \tag{4}$$

The significance of  $\mathcal{S}_1(a,b,c)$  in  $\mathcal{D}_1$  relative to that of  $\mathcal{S}_2(a,b,c)$  in  $\mathcal{D}_2$  is defined by the “tensor GSVD angular distance”  $\Theta_{abc}$  as a function of the ratio  $\mathcal{R}_{1,abc}/\mathcal{R}_{2,abc}$ . This is in analogy with, e.g., the row mode GSVD angular distance  $\theta_a$ , which defines the significance of the column basis vector  $u_{1,a}$  in the matrix  $D_1$  of Equation (2) relative to that of  $u_{2,a}$  in  $D_2$  as a function of the ratio  $\sigma_{1,a}/\sigma_{2,a}$ ,

$$\begin{aligned} \Theta_{abc} &= \arctan(\mathcal{R}_{1,abc}/\mathcal{R}_{2,abc}) - \pi/4, \\ \theta_a &= \arctan(\sigma_{1,a}/\sigma_{2,a}) - \pi/4. \end{aligned} \tag{5}$$

Because the ratios of the positive generalized singular values satisfy  $\sigma_{1,a}/\sigma_{2,a} \in [0, \infty)$ , the row mode GSVD angular distances satisfy  $\theta_a \in [-\pi/4, \pi/4]$ . The maximum (or minimum) angular distance, i.e.,  $\theta_a = \pi/4$ , which corresponds to  $\sigma_{1,a}/\sigma_{2,a} > 1$  (or  $-\pi/4$ , which corresponds to  $\sigma_{1,a}/\sigma_{2,a} < 1$ ), indicates that the row basis vector  $v_a^T$  of Equation (2), which corresponds to the column basis vectors  $u_{1,a}$  in  $D_1$  and  $u_{2,a}$  in  $D_2$ , is exclusive to  $D_1$  (or  $D_2$ ). An angular distance of  $\theta_a = 0$ , which corresponds to  $\sigma_{1,a}/\sigma_{2,a} = 1$ , indicates a row basis vector  $v_a^T$  which is of equal significance in, i.e., common to both  $D_1$  and  $D_2$ .

Thus, while the ratio  $\sigma_{1,a}/\sigma_{2,a}$  indicates the significance of  $u_{1,a}$  in  $D_1$  relative to the significance of  $u_{2,a}$  in  $D_2$ , this relative significance is defined, as previously described [12, 13], by the angular distance  $\theta_a$ , a function of the ratio  $\sigma_{1,a}/\sigma_{2,a}$ , which is antisymmetric in  $D_1$  and  $D_2$ . Note also that while other functions of the ratio  $\sigma_{1,a}/\sigma_{2,a}$  exist that are antisymmetric in  $D_1$  and  $D_2$ , the angular distance  $\theta_a$ , which is a function of the arctangent of the ratio, i.e.,  $\arctan(\sigma_{1,a}/\sigma_{2,a})$ , is the natural function to use, because the GSVD is related to the cosine-sine (CS) decomposition, as previously described [9], and, thus,  $\sigma_{1,a}$  and  $\sigma_{2,a}$  are related to the sine and the cosine functions of the angle  $\theta_a$ , respectively.

*Theorem 1. The tensor GSVD angular distance equals the row mode GSVD angular distance, i.e.,  $\Theta_{abc} = \theta_a$ .*

*Proof.* The unfolding of  $\mathcal{D}_i$  of Equation (1) into  $D_i$  of Equation (2) unfolds the core tensors  $\mathcal{R}_i$  of Equation (1) into matrices  $R_i$ , which preserve the row dimensions, i.e., the  $LM$ -column

bases dimensions of  $\mathcal{R}_i$ , and gives

$$\begin{aligned} D_i &= U_i R_i (V_x^T \otimes V_y^T), \\ R_i &= \Sigma_i V^T (V_x^{-T} \otimes V_y^{-T}), \quad i = 1, 2, \end{aligned} \tag{6}$$

where  $\otimes$  denotes a Kronecker product. Because  $\Sigma_i$  are positive diagonal matrices, it follows that  $\mathcal{R}_{1,abc}/\mathcal{R}_{2,abc} = R_{1,a}/R_{2,a} = \sigma_{1,a}/\sigma_{2,a}$ . Substituting this in Equation (5) gives  $\Theta_{abc} = \theta_a$ . Note that the proof holds for tensors of higher-than-third order.

From this it follows that the tensor GSVD angular distance  $|\Theta_{abc}| \leq \pi/4$ , and that, therefore, the ratio of the tensor generalized singular values  $\mathcal{R}_{1,abc}/\mathcal{R}_{2,abc} > 0$ , even though  $\mathcal{R}_{1,abc}$  and  $\mathcal{R}_{2,abc}$  are not necessarily positive. It also follows that  $\Theta_{abc} = \pm\pi/4$  indicate a subtensor exclusive to either  $\mathcal{D}_1$  or  $\mathcal{D}_2$ , respectively, and that  $\Theta_{abc} = 0$  indicates a subtensor common to both.

Note that since the generalized singular values are arranged in  $\Sigma_i$  of Equation (2) in a decreasing order of the row mode GSVD angular distances  $\theta_a$ , the most tumor-exclusive tumor subtensors, i.e.,  $\mathcal{S}_1(a,b,c)$  where  $a$  maximizes  $\theta_a$  of Equation (5), correspond to  $a = 1$ , whereas the most normal-exclusive normal subtensors, i.e.,  $\mathcal{S}_2(a,b,c)$  where  $a$  minimizes  $\theta_a$ , correspond to  $a = LM$ .

## Discovery and Validation of CNAs Predicting OV Survival

We compute the tensor GSVD of the tumor and normal discovery datasets for each chromosome arm and each combination of two chromosome arms, separately (S1 Mathematica Notebook). For each arm or arms we examine the most significant subtensor in the tumor dataset, i.e.,  $\mathcal{S}_1(a,b,c)$ , where  $a$ ,  $b$ , and  $c$  maximize  $\mathcal{P}_{1,abc}$  of Equation (4).

We, first, require the subtensor to be tumor-exclusive and platform-consistent: include the tumor arraylet  $u_{1,a}$  that is the most exclusive to the tumor dataset, i.e.,  $u_{1,1}$ , as well as a  $y$ -probelet  $v_{y,c}^T$  of consistent, i.e., approximately equal copy numbers in both platforms. Second, we require the subtensor to be correlated with an OV patient's prognosis in the discovery set of patients, i.e., include an  $x$ -probelet  $v_{x,b}^T$  that classifies the discovery set of patients into two groups of high ( $> 0.5$  standardized median absolute deviation, i.e., sMAD, from the median) and low coefficients, of significantly (log-rank test  $P$ -value  $< 0.05$ ) and robustly (throughout the range of  $\pm 0.1$  sMAD around the cutoff) different prognoses (Fig. 2). Third, we require the subtensor to be correlated with prognosis in the validation set of patients, i.e., include an arraylet that classifies the validation set of patients into two groups of high and low Spearman's rank correlation coefficients of significantly different prognoses, consistent with the  $x$ -probelet's classification of the discovery set of patients (Fig. 3, and Sec. 1.3 in S1 Appendix). Note that the validation set includes 148 TCGA patients, mutually exclusive of the discovery set, with primary OV tumor profiles measured by at least one of the two DNA microarray platforms that were used to measure the discovery datasets (S2 Dataset).

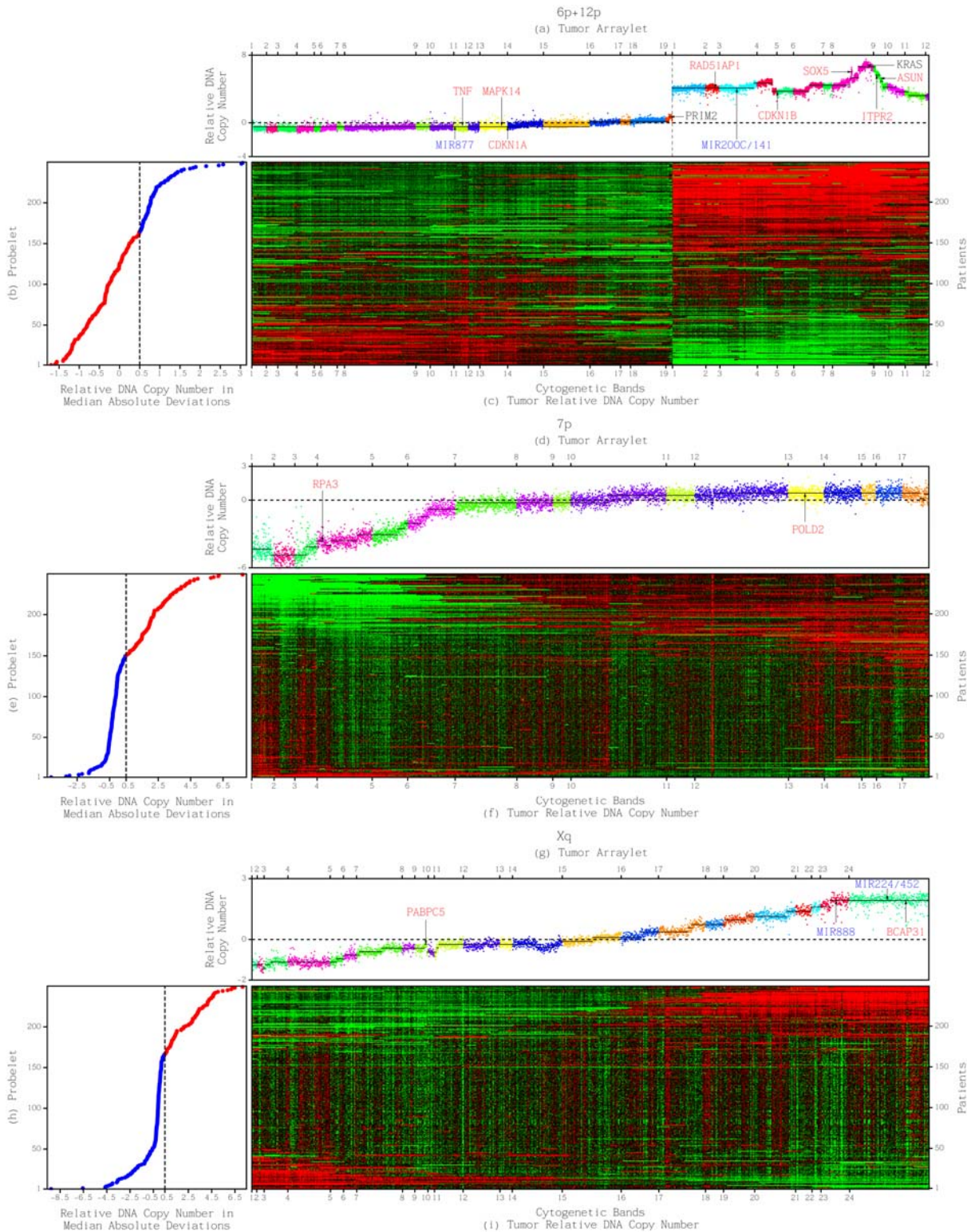
We find that each of the tensor GSVDs of only the chromosome arms 7p and Xq, and only the combination of the two chromosome arms 6p+12p (but not 6p nor 12p separately), uncovers a pattern of tumor-exclusive and platform-consistent co-occurring CNAs that is correlated with an OV patient's prognosis in the discovery and, separately, validation set of patients.

## Biological Results

### Independent Chromosome Arm-Wide Predictors of OV Survival and Response to Platinum-Based Chemotherapy

To date, the best predictor of OV survival has remained the tumor's stage at diagnosis [31] (Sec. 2.1, and Figs. D and E in S1 Appendix). Additional indicators, such as the residual disease





**Fig 2. Tumor-exclusive and platform-consistent DNA copy-number alterations (CNAs) correlated with ovarian serous cystadenocarcinoma (OV) patients' survival.** (a) Plot of the first 6p+12p tumor arraylet describes a pattern of tumor-exclusive and platform-consistent co-occurring CNAs across the combination of the two chromosome arms 6p+12p. The probes are ordered, and their copy numbers are colored according to each probe's chromosomal band location. Segments (black lines) amplified and deleted include most known OV-associated CNAs that map to 6p+12p (black), including an amplification of *KRAS* and a deletion of *PRIM2*. CNAs previously unrecognized in OV (red) include a deletion of the p38-encoding *MAPK14*, and p21-encoding *CDKN1A*,

and an amplification of *RAD51AP1*, a deletion of *TNF*, and focal amplifications of *ASUN*, *ITPR2*, and the 5' ends of isoforms a and e, and exons 5 and 6 of *SOX5*. A high 6p+12p arraylet correlation is significantly correlated with a patient's shorter survival time. (b) Plot of the first 6p+12p x-probelet describes the classification of the discovery set of patients into two groups of high (blue) and low (red) coefficients. A high 6p+12p x-probelet coefficient is significantly and robustly correlated with a patient's shorter survival time. (c) Raster display of the 6p+12p tumor profiles, where medians of the profiles of the same patient measured by the two platforms were taken, with relative gain (red), no change (black), and loss (green) of DNA copy numbers. (d) Plot of the first 7p tumor arraylet describes a pattern of CNAs across the chromosome arm 7p. CNAs previously unrecognized in OV (red) include a focal deletion of *RPA3* and an amplification of *POLD2*. A high 7p arraylet correlation is significantly correlated with a patient's longer survival time. (e) Plot of the first 7p x-probelet describes the classification of the discovery set of patients into two groups of high (red) and low (blue) coefficients. A high 7p x-probelet coefficient is significantly and robustly correlated with a patient's longer survival time. (f) Raster display of the 7p tumor profiles. (g) Plot of the first Xq tumor arraylet. CNAs previously unrecognized in OV (red) include a focal deletion of *PABPC5* and an amplification of *BCAP31*. A high Xq arraylet correlation is significantly correlated with a patient's longer survival time. (h) Plot of the first Xq x-probelet describes the classification of the discovery set of patients into two groups of high (red) and low (blue) coefficients. A high Xq x-probelet coefficient is significantly and robustly correlated with a patient's longer survival time. (i) Raster display of the Xq tumor profiles.

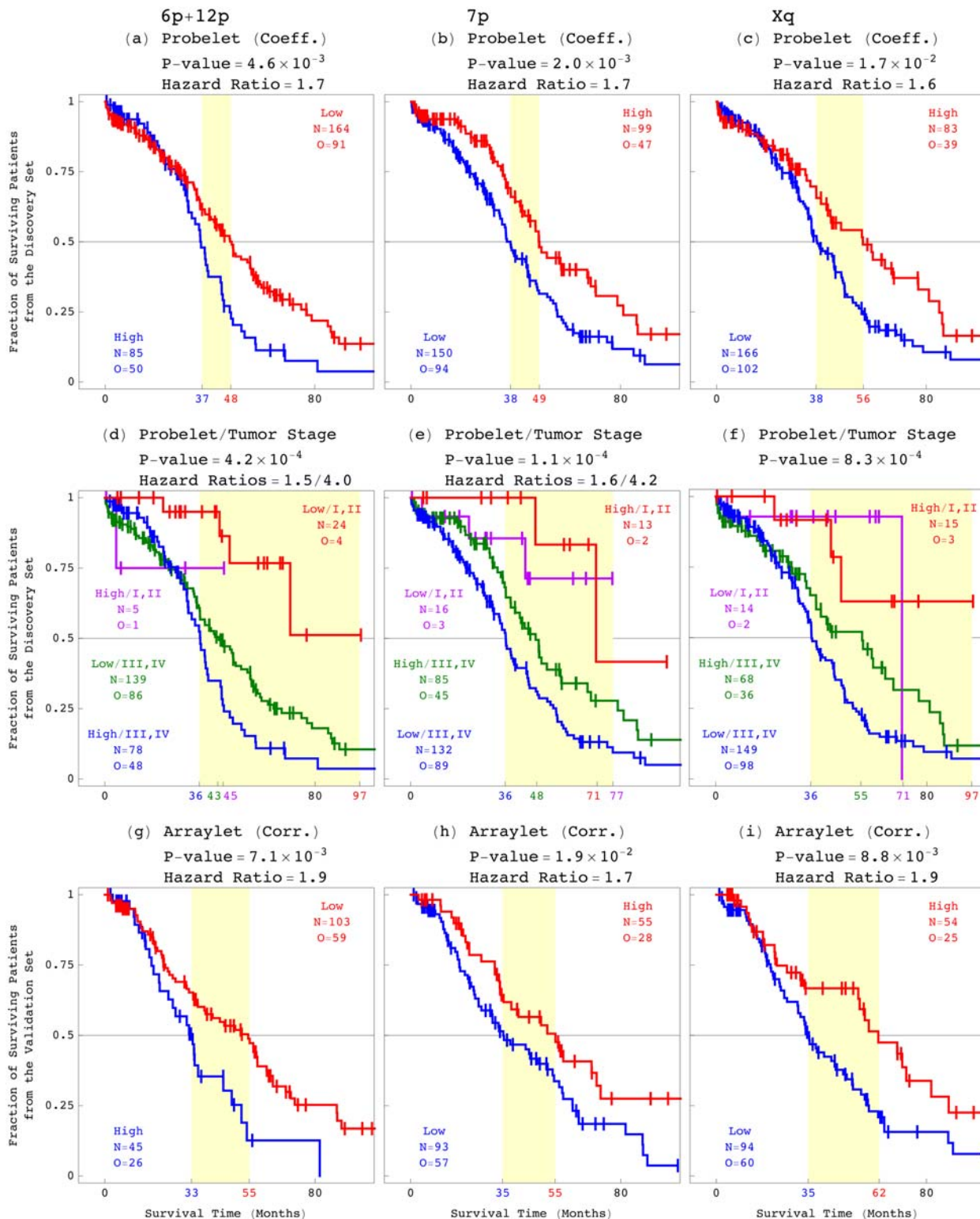
doi:10.1371/journal.pone.0121396.g002

after surgery, the outcome of subsequent therapy, and the neoplasm status, which is the last known status of the disease, are determined during treatment. No diagnostic exists that distinguishes between platinum-based chemotherapy-resistant and -sensitive tumors before the treatment [32, 33].

We find and validate, by using survival analyses of the discovery and, separately, validation set of patients, as well as only the 88% and 95% platinum-based chemotherapy patients in the discovery and validation sets, respectively (Fig. F in [S1 Appendix](#)), that each of the patterns, across 6p+12p, 7p, and Xq, is correlated with an OV patient's prognosis and response to platinum-based chemotherapy, is independent of stage, and together with stage makes a better predictor than stage alone.

We also find and validate that each of these three tensor GSVDs is independent of each of the additional standard indicators (Tables A and B in [S1 Appendix](#)). For example, survival analyses of the discovery set classified by the 6p+12p tensor GSVD into high and low x-probelet coefficients, and by pathology at diagnosis into tumor stages I-II and III-IV, give the bivariate Cox hazard ratios of 1.5 and 4.0, which are similar to the corresponding univariate ratios of 1.7 and 4.4, respectively [18]. Similarly, survival analyses of the validation set classified by the 6p+12p tensor GSVD into high and low arraylet correlation coefficients, and by pathology at diagnosis into tumor stages III and IV, give the bivariate Cox hazard ratios of 1.9 and 1.8, which are the same as the corresponding univariate ratios (Fig. G in [S1 Appendix](#)). This means that the 6p+12p tensor GSVD and stage are independent predictors of survival. Therefore, combined with any one of the standard indicators, each of the three tensor GSVDs makes a better predictor than the standard indicator alone (Figs. H and I in [S1 Appendix](#)). For example, the Kaplan-Meier (KM) median survival time difference of 61 months among the discovery set of patients classified by both the 6p+12p tensor GSVD and stage, is about 85% and more than two years greater than the 33 month difference between the patients classified by stage alone [19]. The KM median survival difference of 34 months among the validation set of patients classified by both the 6p+12p tensor GSVD and stage, is about 62% and more than one year greater than the 21 month difference between the patients classified by stage alone.

Note that while the discovery set of patients reflects the general OV patient population, with approximately 5%, 7%, 76%, and 12% of the patients diagnosed at stages I, II, III, and IV, respectively, the validation set reflects the high-stage OV patient population, with approximately 20% and 80% of the patients diagnosed at stages III and IV, respectively. The 6p+12p, 7p, and Xq tensor GSVDs, therefore, predict survival both in the general as well as in the high-stage OV patient population. Note also that the discovery and validation sets each include mostly, i.e., > 95% high-grade, i.e., grades 2 and higher tumors. Tumor grade does not correlate with survival in either the discovery or the validation set of patients. Survival analyses of only



**Fig 3. Survival analyses of the discovery and validation sets of patients classified by tensor GSVD, or tensor GSVD and tumor stage at diagnosis.** (a) Kaplan-Meier (KM) curves of the discovery set of 249 patients classified by the 6p+12p x-probelet coefficient, show a median survival time difference of 11 months, with the corresponding log-rank test  $P$ -value  $< 10^{-2}$ . The univariate Cox proportional hazard ratio is 1.7. (b) Survival analyses of the 249 patients classified by the 7p x-probelet coefficient. (c) The 249 patients classified by the Xq x-probelet coefficient. (d) The 249 patients classified by both the 6p+12p tensor GSVD and tumor stage at diagnosis, show the bivariate Cox hazard ratios of 1.5 and 4.0, which do not differ significantly from the corresponding

univariate hazard ratios of 1.7 and 4.4, respectively. This means that the 6p+12p tensor GSVD is independent of stage, the best predictor of OV survival to date. The 61 months KM median survival time difference is about 85% and more than two years greater than the 33 month difference between the patients classified by stage alone. This means that the tensor GSVD and stage combined make a better predictor than stage alone. (e) The 249 patients classified by both the 7p tensor GSVD and stage. (f) The 249 patients classified by both the Xq tensor GSVD and stage. (g) KM curves of the validation set of 148 stage III-IV patients classified by the 6p+12p arraylet correlation, show a median survival time difference of 22 months, with the corresponding log-rank test  $P$ -value  $< 10^{-2}$ , and the univariate Cox proportional hazard ratio 1.9. This validates the survival analyses of the discovery set of 249 patients. (h) Survival analyses of the 148 patients classified by the 7p arraylet correlation. (i) The 148 patients classified by the Xq arraylet correlation.

doi:10.1371/journal.pone.0121396.g003

the  $> 95\%$  patients with high-grade tumors in the discovery and, separately, validation set give qualitatively the same and quantitatively similar results to those of the analyses of 100% of the patients in each set, respectively. The 6p+12p, 7p, and Xq tensor GSVDs, therefore, predict survival in the high-grade OV patient population, and are independent of the OV tumor's grade as well as the molecular distinctions between high- and low-grade OV tumors [30].

We observe three groups of significantly different prognoses among the discovery and, separately, validation set of patients, as well as only the platinum-based chemotherapy patients, classified by a combination of the three, i.e., 6p+12p, 7p, and Xq, tensor GSVD classifications, each of which is binomial (Fig. 4). In group A, a combination of a low 6p+12p  $x$ -probelet coefficient or arraylet correlation, and high 7p and Xq  $x$ -probelet coefficients or arraylet correlations is indicative of a patient's significantly longer survival time and better response to platinum-based chemotherapy. In group B, the three combinations where just one of the three binomial classifications differs from that of group A, indicate shorter survival time and worse response to chemotherapy than those of group A. In group C, the four combinations where at least two of the three binomial classifications differ from that of group A, indicate shorter survival time and worse response to chemotherapy than those of group B as well as group A. For example, the KM median survival times of the discovery set of patients classified into groups A, B, and C are 86, 52, and 36 months, such that the median survival time of group A is more than four years greater than, and more than twice that of group C.

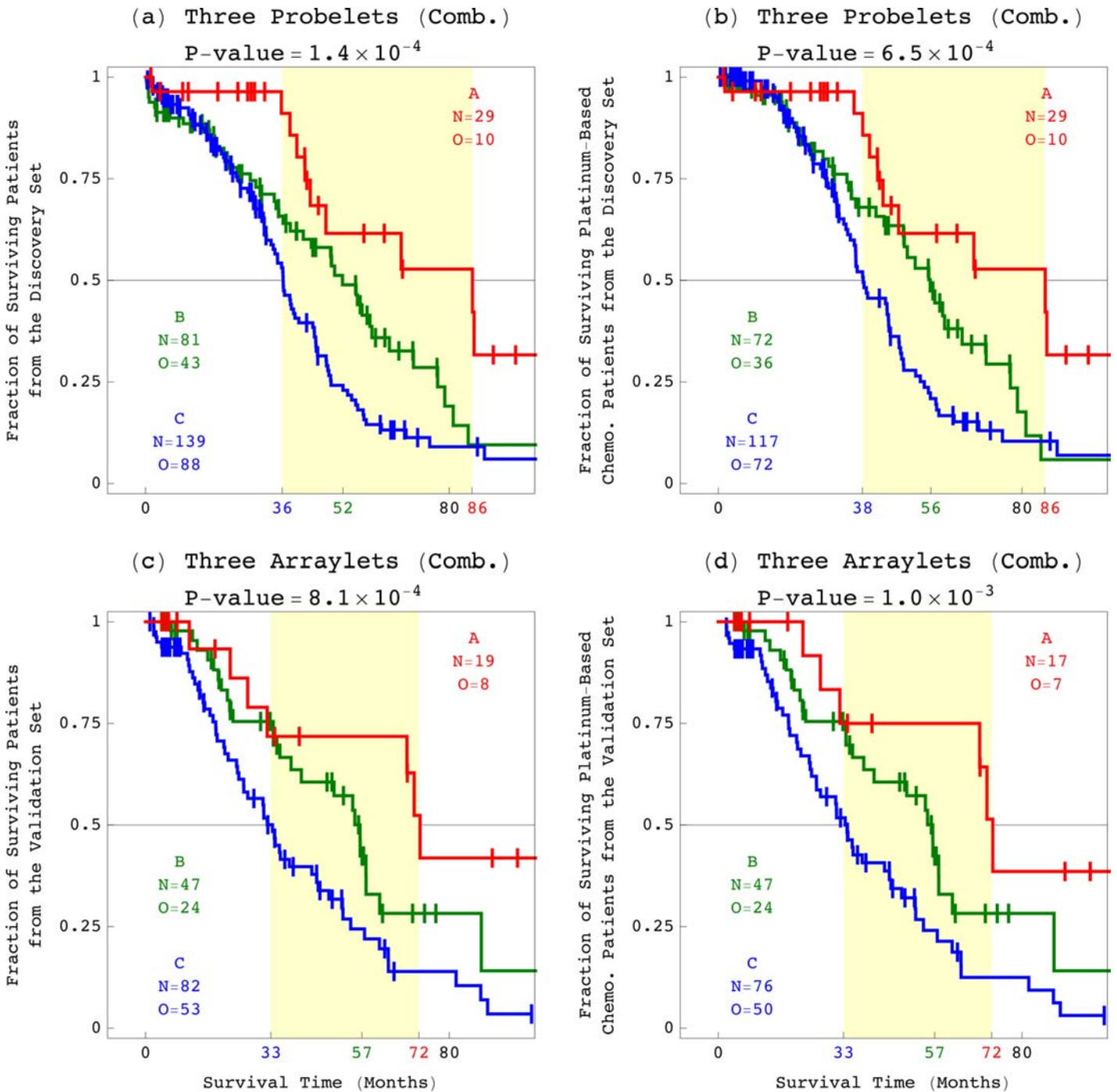
This suggests a possible implementation of the 6p+12p, 7p, and Xq patterns in a pathology laboratory test, where a patient's survival and response to platinum-based chemotherapy is predicted based upon the combination of the correlations of the OV tumor's DNA copy-number profile with the 6p+12p, 7p, and Xq patterns.

## Novel Frequent Focal CNAs Indicating Survival

OV tumors exhibit significant CNA variation among them, much more so than, e.g., GBM brain tumors [2, 13]. Very few frequently occurring OV CNAs have been identified to date.

We find, by using segmentation [20, 21], that the three tensor GSVD arraylets include most known OV-associated CNAs that map to the corresponding chromosome arms, and several previously unreported yet frequent CNAs in  $> 23\%$  of the patients. For example, the 6p+12p arraylet includes two segments corresponding to the only known OV focal CNAs that map to 6p+12p, 7p, or Xq (Sec. 2.2 in S1 Appendix). One, a deletion (6p11.2), overlaps the 3' end unique to isoform a of the DNA primase polypeptide 2-encoding *PRIM2* [2]. The other, an amplification (12p12.1-p11.23), contains several genes, including the Kirsten rat sarcoma viral oncogene homolog *KRAS*, one of three human Ras genes, and the 5' ends of isoforms b and d of the *SRY* (sex determining region Y)-box 5-encoding *SOX5* [34], and is significantly (log-rank test  $P$ -value  $< 0.05$ , and KM median survival time difference  $\geq 12$  months) correlated with OV survival (S3 Dataset).

We also find that the three arraylet patterns include novel frequent focal CNAs (segments  $< 125$  probes). Among these, four amplifications and two deletions are significantly



**Fig 4. Survival analyses of the discovery and validation sets of patients, as well as only the platinum-based chemotherapy patients in the discovery and validation sets, classified by the 6p+12p, 7p, and Xq tensor GSVD combined.** (a) KM curves of the discovery set of 249 patients classified by combination of the 6p+12p, 7p, and Xq x-probelet coefficients, show median survival times of 86, 52, and 36 months for the groups A, B, and C, respectively, with the corresponding log-rank test  $P$ -value  $< 10^{-3}$ . (b) KM survival analysis of only the 218, i.e., ~ 88% platinum-based chemotherapy patients in the discovery set, classified by combination of the three tensor GSVDs, gives qualitatively the same and quantitatively similar results to those of the analyses of 100% of the patients. This means that the combination of the three tensor GSVDs predicts survival in the platinum-based chemotherapy patient population. (c) KM curves of the validation set of 148 stage III-IV patients classified by combination of the 6p+12p, 7p, and Xq arraylet correlation coefficients, show median survival times of 72, 57, and 33 months for the groups A, B, and C, respectively, with the corresponding log-rank test  $P$ -value  $< 10^{-3}$ . This validates the survival analyses of the discovery set of 249 patients. (d) KM survival analysis of only the 140, i.e., ~ 95% platinum-based chemotherapy patients in the validation set, classified by combination of the three tensor GSVDs.

doi:10.1371/journal.pone.0121396.g004

correlated with OV survival (Fig. J in [S1 Appendix](#)). The amplifications flank the segment that contains *KRAS*. Two consecutive segments (12p12.1) contain the 5' ends of isoforms a and e of *SOX5*, and exons 5 and 6, the first exons that are common to isoforms a, b, d, and e of *SOX5* [35]. Two other consecutive segments (12p11.23) contain the inositol 1,4,5-trisphosphate receptor type 2-encoding *ITPR2*, and the asunder spermatogenesis regulator-encoding *ASUN*. *ASUN* was discovered in a screen of expressed sequence tags on 12p11-p12, which DNA amplification correlated with mRNA overexpression in four human testicular seminomas and one ovarian papillary serous adenocarcinoma cell line, exemplifying human germ cell tumors [36]. *ASUN* and its homologs are essential for nuclear division after DNA replication in the HeLa human cervical cancer cell line, the frog, and the fly [37]. One deletion (7p22.1-p21.3) contains the replication protein A3-encoding *RPA3*. The other (Xq21.31) contains the cytoplasmic poly (A)-binding protein 5-encoding *PABPC5*, and the sequence tag site DX214 adjacent to translocation breakpoints observed in premature ovarian failure [38].

### Possible Roles in OV Pathogenesis

We find, by using gene ontology enrichment analyses of the OV tumor mRNA expression profiles of the patients [39, 40], that differential mRNA expression between the patients, classified by any one of the three tensor GSVDs, is enriched in ontologies corresponding to one of three hallmarks of cancer [41]: cell immortality in 6p+12p, DNA instability in 7p, and cellular immune response suppression in Xq.

The differential mRNA expression of genes from these enriched ontologies that are located on any one of the chromosome arms is consistent with the CNAs across that arm (Fig. K in [S1 Appendix](#), and [S4 Dataset](#)). Genes that map to amplifications or deletions on any one arraylet pattern, are overexpressed or underexpressed, respectively, in the patients which tumor profiles are classified, by the corresponding tensor GSVD, as highly similar to that pattern, i.e., patients of high *x*-probelet coefficients or arraylet correlations. The differential expression of all microRNAs and proteins that map to any one of the chromosome arms is also consistent with the CNAs across that arm (Sec. 2.3, and Figs. L and M in [S1 Appendix](#), and [S5](#) and [S6](#) Datasets). A coherent picture emerges for each pattern, suggesting roles for the CNAs in OV pathogenesis in addition to personalized diagnosis, prognosis, and treatment.

**6p+12p. A cell's transformation and immortality are correlated with a patient's shorter survival.** The genes, which are significantly (Mann-Whitney-Wilcoxon *P*-values < 0.05) differentially expressed between the 6p+12p tensor GSVD classes, i.e., in the patient group of high 6p+12p *x*-probelet coefficient or arraylet correlation, relative to the patient group of low coefficient or correlation, are enriched (hypergeometric *P*-values < 10<sup>-3</sup>) in the ontologies of cellular response to ionizing radiation (GO:0071479), and major histocompatibility (MHC) protein complex (GO:0042611). Most of the GO:0071479 genes are underexpressed, including the p21 cyclin-dependent kinase inhibitor-encoding *CDKN1A*, and the p38 mitogen-activated protein kinase-encoding *MAPK14*, which map to a deletion > 45 Mbp on the telomeric part of 6p (6p25.3-p21.1). Also underexpressed is p38, the protein encoded by *MAPK14*. All GO:0042611 genes, including the tumor necrosis factor-encoding *TNF*, are underexpressed, and map to the same deletion. The one microRNA that is significantly differentially expressed between the 6p+12p tensor GSVD classes, and maps to the same deletion, is the splicing-dependent microRNA miR-877\*, which is encoded by the 13th intron of the ATP-binding cassette subfamily F member 1-encoding gene *ABCF1* [44]. Both miR-877\* and *ABCF1* are consistently underexpressed.

One of only two GO:0071479 overexpressed genes is the *RAD51*-associated protein 1-encoding *RAD51API*, which maps to an amplification > 9 Mbp on the telomeric part of 12p

(12p13.33-p13.31) that is significantly correlated with OV survival. All four microRNAs that are differentially expressed between the 6p+12p tensor GSVD classes, and map to the same amplification, miR-200c, miR-200c\*, miR-141, and miR-141\*, are consistently overexpressed. The second protein that is significantly differentially expressed between the 6p+12p tensor GSVD classes is p27. Consistently, the cyclin-dependent kinase inhibitor *CDKN1B*, which encodes p27, maps to a 4.5 Mbp amplification (12p13.2-p12.3) that is significantly correlated with OV survival, and its mRNA is overexpressed. The mRNA encoded by *KRAS* is also overexpressed.

Note that while the 6p+12p pattern of CNAs is correlated with survival in the discovery and, separately, validation sets, neither the 6p nor the 12p pattern alone are correlated with survival. Indeed, experiments studying the conditions for the transformation of human normal to tumor cells indicate that cells, where both p21 and p38 are inactive, are susceptible to Ras-mediated transformation [42, 43]. However, the activation of Ras alone induces tumor-suppressing cellular senescence via the activities of either p21 or p38. The 6p+12p pattern, therefore, which includes the loss of the p21-encoding *CDKN1A* and the p38-encoding *MAPK14* on 6p, and the gain of *KRAS* on 12p, encodes for cellular conditions that combined but not separately can lead to transformation.

In addition, p21 and p38 are necessary for p53-mediated cell cycle arrest [45] and apoptosis [46], respectively, in response to DNA damage. Overexpression of the p21-encoding *CDKN1A* is correlated with a low malignant potential of an ovarian tumor [47]. *RAD51API* overexpression disrupts cell cycle arrest and apoptosis, can lead to cellular resistance to DNA-damaging cancer therapies, such as platinum-based chemotherapy, and may increase DNA instability [48]. *TNF*-induced apoptosis is correlated with downregulation of *ITPR2* [49]. Overexpression of miR-200c, and miR-141, both of which putatively target the *BRCA1* associated protein-1 oncosuppressor-encoding *BAP1*, is correlated with OV tumor growth, dedifferentiation, and invasiveness [50, 51]. Overexpression of the *CDKN1B*-encoded p27, which can promote cellular migration [52] and even proliferation [53], is correlated with a poor OV patient's prognosis [54, 55].

Taken together, previously unrecognized co-occurring deletion of *CDKN1A* and *MAPK14* on 6p and amplification of *KRAS* on 12p, which encode for human cell transformation, together with deletion of *TNF* on 6p, and amplification of *RAD51API* and *ITPR2* on 12p, are correlated with a suppression of cell cycle arrest, senescence, and apoptosis, i.e., a tumor cell's immortality, and a patient's shorter survival time. Note that there already exist drugs that interact with *CDKN1A*, *MAPK14*, and *RAD51API*, even though these genes were not recognized previously as targets for OV drug therapy [56].

**7p. A cell's DNA stability is correlated with a longer survival.** The genes that are significantly differentially expressed between the 7p tensor GSVD classes are enriched (hypergeometric  $P$ -value  $< 10^{-10}$ ) in the ontology of DNA strand elongation involved in DNA replication (GO:0006271). Most of these genes are overexpressed, including the DNA polymerase delta subunit 2-encoding *POLD2* that is essential for DNA replication and repair, which maps to an amplification  $> 17$  Mbp on the centromeric part of 7p (7p14.1-p11.2). Only two genes are underexpressed: *RPA3* on 7p and the DNA ligase IV-encoding *LIG4* on 13q. The interaction of p53 with the *RPA3*-encoded protein mediates suppression of homologous recombination (HR), the preferred cellular mechanism for DNA double-strand break (DSB) repair during replication [57]. *LIG4* is essential for DSB repair via the more error-prone nonhomologous end joining pathway [58]. HR defects are thought to facilitate the significant CNA heterogeneity among OV tumors [2].

Taken together, previously unrecognized co-occurring deletion and underexpression of *RPA3*, and amplification and overexpression of *POLD2* on 7p are correlated with DNA DSB repair via HR during replication, i.e., DNA stability, and a longer survival time.

**Xq. Cellular immune response is correlated with a longer survival.** The genes that are differentially expressed between the Xq tensor GSVD classes are enriched (hypergeometric  $P$ -value  $< 10^{-6}$ ) in the ontology of antigen processing and presentation of peptide antigen (GO:0048002). Most of these genes are overexpressed, including the B-cell receptor-associated protein 31-encoding *BCAP31*, which maps to an amplification  $> 11$  Mbp on the telomeric part of Xq (Xq27.3-q28). All three microRNAs that are differentially expressed between the Xq tensor GSVD classes, and map to the same amplification, miR-888, miR-224, and miR-452, together with the gamma-aminobutyric acid (GABA) A receptor epsilon-encoding *GABRE*, which hosts miR-224 and miR-452 in its introns, are consistently overexpressed. Underexpression of miR-224 was implicated in OV pathogenesis [50]. *PABPC5*, which maps to a focal deletion on Xq, is suppressed upon viral infection [59].

Taken together, previously unrecognized co-occurring deletion of *PABPC5*, and amplification and overexpression of *BCAP31* on Xq are correlated with a cellular immune response, and a longer survival time.

## Discussion

We defined a novel tensor GSVD, an exact simultaneous decomposition of two datasets, arranged in two higher-than-second-order tensors of matched column dimensions but independent row dimensions. We showed that the mathematical properties of the tensor GSVD allow interpreting its variables and operations in terms of the similar as well as dissimilar, e.g., biomedical reality between the datasets. We demonstrated the tensor GSVD in comparative modeling of patient- and platform-matched but probe-independent OV tumor and normal DNA copy-number profiles from TCGA. The modeling resulted in new insights into the poorly understood relations between an OV tumor's genome and a patient's survival phenotype. Three previously unrecognized chromosome arm-wide patterns of tumor-exclusive and platform-consistent co-occurring alterations were uncovered, across 6p+12p, 7p, and Xq, that are correlated with an OV patient's survival and response to platinum-based chemotherapy, and are of possible roles in OV pathogenesis, and of a possible implementation in a pathology laboratory test for personalized OV diagnosis, prognosis, and treatment.

Note that unlike previous analyses of the TCGA OV DNA copy-number data, notably by TCGA [2], our analyses were not limited to the 22 human autosomal chromosomes, and include the X chromosome. This is because the tensor GSVD, like the GSVD, comparatively—based upon the structure of the data—separates the matched datasets into uncorrelated, i.e., orthogonal patterns across the tumor and normal probes. Patterns of copy-number variation across the tumor probes that occur in the normal human genome, and are common to the tumor and normal datasets, such as the female-specific X chromosome amplification, are orthogonal to, and, therefore, are separated from the patterns that are exclusive to the tumor dataset. For example, the GSVD comparative modeling of patient-matched GBM tumor and normal copy-number profiles separated the prognosis-correlated GBM tumor-exclusive pattern from the female-specific X chromosome amplification as well as from experimental artifacts (or batch effects) due to experimental variations in, e.g., tissue batch, genomic center, hybridization date, and scanner, without a-priori knowledge of these variations.

Unlike recent approaches to the integrative modeling of different types of large-scale molecular biological profiles from the same set of patients, notably clustering [60, 61], our comparative modeling was not limited to tumor profiles, and included also patient- and platform-matched normal DNA copy-number profiles. This is because the tensor GSVD, like the GSVD, finds not just the similarities but, at the same time also the dissimilarities among the profiles without making any assumptions, except for the structure of the data: two third-order tensors,



of matched columns that correspond to the same sets of patients and platforms, and independent rows that correspond to the probes in either the tumor or the normal dataset. The patients, platforms, tumor and normal probes as well as the tissue types, each represent a degree of freedom. Unfolded into two matrices or appended into a single tensor (or even unfolded and appended into a single matrix), some of the degrees of freedom are lost and much of the information in the datasets might also be lost. For example, SVD of the GBM tumor and normal profiles appended into a single matrix, while it is related to the GSVD of the data, would not separate the tumor dataset into patterns across the tumor probes that are orthogonal.

Additional possible applications of the tensor GSVD in personalized medicine include comparative modeling of two patient- and tissue-matched datasets, each corresponding to (i) a set of large-scale molecular biological profiles, e.g., DNA copy numbers, acquired by a high-throughput technology, e.g., DNA microarrays; (ii) a set of biomedical images or signals; or (iii) a set of cellular pathological observations, e.g., a tumor's stage. Such tensor GSVD comparative models can uncover variations across the patients and tissues that are common to, possibly causally coordinated between the two aspects of the disease. In clinical settings, such tensor GSVD comparative models can determine an individual patient's medical status in relation to all the other patients in a set, and inform the patient's diagnosis, prognosis and treatment.

## Supporting Information

**S1 Appendix. A PDF format file, readable by Adobe Acrobat Reader.**

(PDF)

**S1 Mathematica Notebook. Tensor GSVD of patient- and platform-matched tumor and normal genomic profiles.** A PDF format file, readable by Adobe Acrobat Reader. The corresponding Mathematica 9.0.1 code file, executable by Mathematica and readable by Mathematica Player, is available at [http://www.alterlab.org/OV\\_prognosis/](http://www.alterlab.org/OV_prognosis/).

(PDF)

**S1 Dataset. Discovery Set of Patients.** A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, reproducing TCGA annotations of the discovery set of 249 patients. The tumor and normal profiles of the discovery set of patients measured by each of the two DNA microarray platforms, tabulating relative copy-number variation across the 6p+12p, 7p, and Xq tumor and normal probes, are available in tab-delimited text format files at [http://www.alterlab.org/OV\\_prognosis/](http://www.alterlab.org/OV_prognosis/).

(TXT)

**S2 Dataset. Validation Set of Patients.** A tab-delimited text format file reproducing TCGA annotations of the validation set of 148 patients. The tumor profiles of the validation set of patients, tabulating relative copy-number variation across the 6p+12p, 7p, and Xq tumor probes, are available in tab-delimited text format files at [http://www.alterlab.org/OV\\_prognosis/](http://www.alterlab.org/OV_prognosis/).

(TXT)

**S3 Dataset. First, Most Tumor-Exclusive Tumor Arraylets.** A tab-delimited text format file tabulating the segments of the first, most tumor-exclusive tumor arraylets computed by tensor GSVD of the discovery set of patients across 6p+12p, 7p, or Xq.

(TXT)

**S4 Dataset. Differential mRNA Expression.** A tab-delimited text format file tabulating differential expression of 11,457 autosomal and X chromosome mRNAs in the 6p+12p, 7p, and Xq tensor GSVD classes. The mRNA expression profiles of 394 of the 397 patients in the discovery

and validation sets are available in tab-delimited text format files at [http://www.alterlab.org/OV\\_prognosis/](http://www.alterlab.org/OV_prognosis/).  
(TXT)

**S5 Dataset. Differential microRNA Expression.** A tab-delimited text format file tabulating differential expression of 639 autosomal and X chromosome microRNAs in the 6p+12p, 7p, and Xq tensor GSVD classes. The microRNA expression profiles of 395 patients are available in tab-delimited text format files at [http://www.alterlab.org/OV\\_prognosis/](http://www.alterlab.org/OV_prognosis/).  
(TXT)

**S6 Dataset. Differential Protein Expression.** A tab-delimited text format file tabulating differential expression of 175 antibodies that probe for 136 autosomal and X chromosome proteins in the 6p+12p, 7p, and Xq tensor GSVD classes. The protein expression profiles of 282 patients are available in tab-delimited text format files at [http://www.alterlab.org/OV\\_prognosis/](http://www.alterlab.org/OV_prognosis/).  
(TXT)

## Acknowledgments

We thank RA Horn for thoughtful discussions of matrix analysis in general, and the tensor GSVD in particular. We thank DDL Bowtell and MM Janát-Amsbury for useful notes on OV in general, and the molecular distinctions between high- and low-grade OV tumors in particular. We also thank RA Weinberg for helpful comments on the hallmarks of cancer in general, and the transformation of human normal to tumor cells in particular.

## Author Contributions

Conceived and designed the experiments: OA. Performed the experiments: PS TES KAA OA. Analyzed the data: PS TES KAA OA. Contributed reagents/materials/analysis tools: PS TES KAA OA. Wrote the paper: PS TES KAA OA. Proved mathematical theorems: TES OA.

## References

1. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455: 1061–1068. doi: [10.1038/nature07385](https://doi.org/10.1038/nature07385) PMID: [18772890](https://pubmed.ncbi.nlm.nih.gov/18772890/)
2. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011; 474: 609–615. doi: [10.1038/nature10166](https://doi.org/10.1038/nature10166) PMID: [21720365](https://pubmed.ncbi.nlm.nih.gov/21720365/)
3. Ponnappalli SP, Golub GH, Alter O. A novel higher-order generalized singular value decomposition for comparative analysis of multiple genome-scale datasets. *Stanford University and Yahoo! Research Workshop on Algorithms for Modern Massive Datasets (MMDS)* (Stanford, CA). 2006; June 21–24.
4. Ponnappalli SP, Saunders MA, Van Loan CF, Alter O. A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. *PLoS One*. 2011; 6: e28072. doi: [10.1371/journal.pone.0028072](https://doi.org/10.1371/journal.pone.0028072) PMID: [22216090](https://pubmed.ncbi.nlm.nih.gov/22216090/)
5. Golub GH, Van Loan CF. *Matrix Computations*. 4th ed. Baltimore, MD: Johns Hopkins University Press; 2012.
6. Horn RA, Johnson CR. *Matrix Analysis*. 2nd ed. Cambridge, UK: Cambridge University Press; 2012.
7. Van Loan CF. Generalizing the singular value decomposition. *SIAM J Numer Anal*. 1976; 13: 76–83. doi: [10.1137/0713009](https://doi.org/10.1137/0713009)
8. Paige CC, Saunders MA. Towards a generalized singular value decomposition. *SIAM J Numer Anal*. 1981; 18: 398–405. doi: [10.1137/0718026](https://doi.org/10.1137/0718026)
9. Van Loan CF. Computing the CS and the generalized singular value decompositions. *Numer Math*. 1985; 46: 479–491. doi: [10.1007/BF01389653](https://doi.org/10.1007/BF01389653)
10. Bai Z, Demmel JW. Computing the generalized singular value decomposition. *SIAM J Sci Comput*. 1993; 14: 1464–1486. doi: [10.1137/0914085](https://doi.org/10.1137/0914085)

11. Friedland S. A new approach to generalized singular value decomposition. *SIAM J Matrix Anal Appl.* 2005; 27: 434–444. doi: [10.1137/S0895479804439791](https://doi.org/10.1137/S0895479804439791)
12. Alter O, Brown PO, Botstein D. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proc Natl Acad Sci USA.* 2003; 100: 3351–3356. doi: [10.1073/pnas.0530258100](https://doi.org/10.1073/pnas.0530258100) PMID: [12631705](https://pubmed.ncbi.nlm.nih.gov/12631705/)
13. Lee CH, Alpert BO, Sankaranarayanan P, Alter O. GSVD comparison of patient-matched normal and tumor aCGH profiles reveals global copy-number alterations predicting glioblastoma multiforme survival. *PLoS One.* 2012; 7: e30098. doi: [10.1371/journal.pone.0030098](https://doi.org/10.1371/journal.pone.0030098) PMID: [22291905](https://pubmed.ncbi.nlm.nih.gov/22291905/)
14. Wiltshire RN, Rasheed BK, Friedman HS, Friedman AH, Bigner SH. Comparative genetic patterns of glioblastoma multiforme: potential diagnostic tool for tumor classification. *Neuro Oncol.* 2000; 2: 164–173. doi: [10.1093/neuonc/2.3.164](https://doi.org/10.1093/neuonc/2.3.164) PMID: [11302337](https://pubmed.ncbi.nlm.nih.gov/11302337/)
15. Misra A, Pellarin M, Nigro J, Smirnov I, Moore D, Lamborn KR, et al. Array comparative genomic hybridization identifies genetic subgroups in grade 4 human astrocytoma. *Clin Cancer Res.* 2005; 11: 2907–2918. doi: [10.1158/1078-0432.CCR-04-0708](https://doi.org/10.1158/1078-0432.CCR-04-0708) PMID: [15837741](https://pubmed.ncbi.nlm.nih.gov/15837741/)
16. Curran WJ Jr, Scott CB, Horton J, Nelson JS, Weinstein AS, Fischbach AJ, et al. Recursive partitioning analysis of prognostic factors in three Radiation Therapy Oncology Group malignant glioma trials. *J Natl Cancer Inst.* 1993; 85: 704–710. doi: [10.1093/jnci/85.9.704](https://doi.org/10.1093/jnci/85.9.704) PMID: [8478956](https://pubmed.ncbi.nlm.nih.gov/8478956/)
17. Gorlia T, van den Bent MJ, Hegi ME, Mirimanoff RO, Weller M, Cairncross JG, et al. Nomograms for predicting survival of patients with newly diagnosed glioblastoma: prognostic factor analysis of EORTC and NCIC trial 26981–22981/CE.3. *Lancet Oncol.* 2008; 9: 29–38. doi: [10.1016/S1470-2045\(07\)70384-4](https://doi.org/10.1016/S1470-2045(07)70384-4) PMID: [18082451](https://pubmed.ncbi.nlm.nih.gov/18082451/)
18. Cox DR. Regression models and life-tables. *J Roy Statist Soc B.* 1972; 34: 187–220.
19. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Amer Statist Assn.* 1958; 53: 457–481. doi: [10.1080/01621459.1958.10501452](https://doi.org/10.1080/01621459.1958.10501452)
20. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002; 12: 996–1006. doi: [10.1101/gr.229102](https://doi.org/10.1101/gr.229102) PMID: [12045153](https://pubmed.ncbi.nlm.nih.gov/12045153/)
21. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* 2004; 5: 557–572. doi: [10.1093/biostatistics/kxh008](https://doi.org/10.1093/biostatistics/kxh008) PMID: [15475419](https://pubmed.ncbi.nlm.nih.gov/15475419/)
22. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov.* 2002; 1: 727–730. doi: [10.1038/nrd892](https://doi.org/10.1038/nrd892) PMID: [12209152](https://pubmed.ncbi.nlm.nih.gov/12209152/)
23. Silljé HH, Takahashi K, Tanaka K, Van Houwe G, Nigg EA. Mammalian homologues of the plant Toulsted gene code for cell-cycle-regulated kinases with maximal activities linked to ongoing DNA replication. *EMBO J.* 1999; 18: 5691–5702. doi: [10.1093/emboj/18.20.5691](https://doi.org/10.1093/emboj/18.20.5691) PMID: [10523312](https://pubmed.ncbi.nlm.nih.gov/10523312/)
24. Pellegrini M, Cheng JC, Voutila J, Judelson D, Taylor J, Nelson SF, et al. Expression profile of CREB knockdown in myeloid leukemia cells. *BMC Cancer.* 2008; 8: 264. doi: [10.1186/1471-2407-8-264](https://doi.org/10.1186/1471-2407-8-264) PMID: [18801183](https://pubmed.ncbi.nlm.nih.gov/18801183/)
25. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. *SIAM J Matrix Anal Appl.* 2000; 21: 1253–1278. doi: [10.1137/S0895479896305696](https://doi.org/10.1137/S0895479896305696)
26. Omberg L, Golub GH, Alter O. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proc Natl Acad Sci USA.* 2007; 104: 18371–18376. doi: [10.1073/pnas.0709146104](https://doi.org/10.1073/pnas.0709146104) PMID: [18003902](https://pubmed.ncbi.nlm.nih.gov/18003902/)
27. Omberg L, Meyerson JR, Kobayashi K, Drury LS, Diffley JFX, Alter O. Global effects of DNA replication and DNA replication origin activity on eukaryotic gene expression. *Mol Syst Biol.* 2009; 5: 312. doi: [10.1038/msb.2009.70](https://doi.org/10.1038/msb.2009.70) PMID: [19888207](https://pubmed.ncbi.nlm.nih.gov/19888207/)
28. Kolda TG, Bader BW. Tensor decompositions and applications. *SIAM Rev.* 2009; 51: 455–500. doi: [10.1137/07070111X](https://doi.org/10.1137/07070111X)
29. Vandewalle J, De Lathauwer L, Comon P. The generalized higher order singular value decomposition and the oriented signal-to-signal ratios of pairs of signal tensors and their use in signal processing. In: *Proc ECCTD'03—European Conf on Circuit Theory and Design*; 2003. pp. I-389–I-392.
30. Ayhan A, Kurman RJ, Yemelyanova A, Vang R, Logani S, Seidman JD, et al. Defining the cut point between low-grade and high-grade ovarian serous carcinomas: a clinicopathologic and molecular genetic analysis. *Am J Surg Pathol.* 2009; 33: 1220–1224. doi: [10.1097/PAS.0b013e3181a24354](https://doi.org/10.1097/PAS.0b013e3181a24354) PMID: [19461510](https://pubmed.ncbi.nlm.nih.gov/19461510/)
31. Prisco MG, Zannoni GF, De Stefano I, Vellone VG, Tortorella L, Fagotti A, et al. Prognostic role of metastasis tumor antigen 1 in patients with ovarian cancer: a clinical study. *Hum Pathol.* 2012; 43: 282–288. doi: [10.1016/j.humpath.2011.05.002](https://doi.org/10.1016/j.humpath.2011.05.002) PMID: [21835429](https://pubmed.ncbi.nlm.nih.gov/21835429/)
32. Harries M, Gore M. Chemotherapy for epithelial ovarian cancer—treatment at first diagnosis. *Lancet Oncol.* 2002; 3: 529–536. doi: [10.1016/S1470-2045\(02\)00846-X](https://doi.org/10.1016/S1470-2045(02)00846-X) PMID: [12217790](https://pubmed.ncbi.nlm.nih.gov/12217790/)

33. Pujade-Lauraine E, Hilpert F, Weber B, Reuss A, Poveda A, Kristensen G, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol*. 2014; 32: 1302–1308. doi: [10.1200/JCO.2013.51.4489](https://doi.org/10.1200/JCO.2013.51.4489) PMID: [24637997](https://pubmed.ncbi.nlm.nih.gov/24637997/)
34. Engler DA, Gupta S, Growdon WB, Drapkin RI, Nitta M, Sergent PA, et al. Genome wide DNA copy number analysis of serous type ovarian carcinomas identifies genetic markers predictive of clinical outcome. *PLoS One*. 2012; 7: e30996. doi: [10.1371/journal.pone.0030996](https://doi.org/10.1371/journal.pone.0030996) PMID: [22355333](https://pubmed.ncbi.nlm.nih.gov/22355333/)
35. Ikeda T, Zhang J, Chano T, Mabuchi A, Fukuda A, Kawaguchi H, et al. Identification and characterization of the human long form of Sox5 (*L-SOX5*) gene. *Gene*. 2002; 298: 59–68. doi: [10.1016/S0378-1119\(02\)00927-7](https://doi.org/10.1016/S0378-1119(02)00927-7) PMID: [12406576](https://pubmed.ncbi.nlm.nih.gov/12406576/)
36. Bourdon V, Naef F, Rao PH, Reuter V, Mok SC, Bosl GJ, et al. Genomic and expression analysis of the 12p11–p12 amplicon using EST arrays identifies two novel amplified and overexpressed genes. *Cancer Res*. 2002; 62: 6218–6223. PMID: [12414650](https://pubmed.ncbi.nlm.nih.gov/12414650/)
37. Lee LA, Lee E, Anderson MA, Vardy L, Tahinci E, Ali SM, et al. Drosophila genome-scale screen for PAN GU kinase substrates identifies Mat89Bb as a cell cycle regulator. *Dev Cell*. 2005; 8: 435–442. doi: [10.1016/j.devcel.2004.12.008](https://doi.org/10.1016/j.devcel.2004.12.008) PMID: [15737938](https://pubmed.ncbi.nlm.nih.gov/15737938/)
38. Blanco P, Sargent CA, Boucher CA, Howell G, Ross M, Affara NA. A novel poly(A)-binding protein gene (*PABPC5*) maps to an X-specific subinterval in the Xq21.3/Yp11.2 homology block of the human sex chromosomes. *Genomics*. 2001; 74: 1–11. doi: [10.1006/geno.2001.6530](https://doi.org/10.1006/geno.2001.6530) PMID: [11374897](https://pubmed.ncbi.nlm.nih.gov/11374897/)
39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000; 25: 25–29. doi: [10.1038/75556](https://doi.org/10.1038/75556) PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
40. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009; 10: 48. doi: [10.1186/1471-2105-10-48](https://doi.org/10.1186/1471-2105-10-48) PMID: [19192299](https://pubmed.ncbi.nlm.nih.gov/19192299/)
41. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144: 646–674. doi: [10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013) PMID: [21376230](https://pubmed.ncbi.nlm.nih.gov/21376230/)
42. Karnoub AE, Weinberg RA. Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol*. 2008; 9: 517–531. doi: [10.1038/nrm2438](https://doi.org/10.1038/nrm2438) PMID: [18568040](https://pubmed.ncbi.nlm.nih.gov/18568040/)
43. Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. *Nature*. 1999; 400: 464–468. doi: [10.1038/22780](https://doi.org/10.1038/22780) PMID: [10440377](https://pubmed.ncbi.nlm.nih.gov/10440377/)
44. Sibley CR, Seow Y, Saayman S, Dijkstra KK, El Andaloussi, Weinberg MS, et al. The biogenesis and characterization of mammalian microRNAs of mirtron origin. *Nucleic Acids Res*. 2012; 40: 438–448. doi: [10.1093/nar/gkr722](https://doi.org/10.1093/nar/gkr722) PMID: [21914725](https://pubmed.ncbi.nlm.nih.gov/21914725/)
45. Waldman T, Kinzler KW, Vogelstein B. p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res*. 1995; 55: 5187–5190. PMID: [7585571](https://pubmed.ncbi.nlm.nih.gov/7585571/)
46. Bulavin DV, Saito S, Hollander MC, Sakaguchi K, Anderson CW, Appella E, et al. Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *EMBO J*. 1999; 18: 6845–6854. doi: [10.1093/emboj/18.23.6845](https://doi.org/10.1093/emboj/18.23.6845) PMID: [10581258](https://pubmed.ncbi.nlm.nih.gov/10581258/)
47. Anglesio MS, Arnold JM, George J, Tinker AV, Tothill R, Waddell N, et al. Mutation of ERBB2 provides a novel alternative mechanism for the ubiquitous activation of RAS-MAPK in ovarian serous low malignant potential tumors. *Mol Cancer Res*. 2008; 6: 1678–1690. doi: [10.1158/1541-7786.MCR-08-0193](https://doi.org/10.1158/1541-7786.MCR-08-0193) PMID: [19010816](https://pubmed.ncbi.nlm.nih.gov/19010816/)
48. Klein HL. The consequences of Rad51 overexpression for normal and tumor cells. *DNA Repair*. 2008; 7: 686–693. doi: [10.1016/j.dnarep.2007.12.008](https://doi.org/10.1016/j.dnarep.2007.12.008) PMID: [18243065](https://pubmed.ncbi.nlm.nih.gov/18243065/)
49. Diaz F, Bourguignon LY. Selective down-regulation of IP<sub>3</sub> receptor subtypes by caspases and calpain during TNF $\alpha$ -induced apoptosis of human T-lymphoma cells. *Cell Calcium*. 2000; 27: 315–328. doi: [10.1054/ceca.2000.0126](https://doi.org/10.1054/ceca.2000.0126) PMID: [11013462](https://pubmed.ncbi.nlm.nih.gov/11013462/)
50. Iorio MV, Visone R, Di Leva G, Donati V, Petrocca F, Casalini P, et al. MicroRNA signatures in human ovarian cancer. *Cancer Res*. 2007; 67: 8699–8707. doi: [10.1158/0008-5472.CAN-07-1936](https://doi.org/10.1158/0008-5472.CAN-07-1936) PMID: [17875710](https://pubmed.ncbi.nlm.nih.gov/17875710/)
51. Yang D, Sun Y, Hu L, Zheng H, Ji P, Pecot CV, et al. Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. *Cancer Cell*. 2013; 23: 186–199. doi: [10.1016/j.ccr.2012.12.020](https://doi.org/10.1016/j.ccr.2012.12.020) PMID: [23410973](https://pubmed.ncbi.nlm.nih.gov/23410973/)
52. Nagahara H, Vocero-Akbani AM, Snyder EL, Ho A, Latham DG, Lissy NA, et al. Transduction of full-length TAT fusion proteins into mammalian cells: TAT-p27<sup>Kip1</sup> induces cell migration. *Nat Med*. 1998; 4: 1449–1452. doi: [10.1038/4042](https://doi.org/10.1038/4042) PMID: [9846587](https://pubmed.ncbi.nlm.nih.gov/9846587/)

53. Kwon YH, Jovanovic A, Serfas MS, Tyner AL. The Cdk inhibitor p21 is required for necrosis, but it inhibits apoptosis following toxin-induced liver injury. *J Biol Chem*. 2003; 278: 30348–30355. doi: [10.1074/jbc.M300996200](https://doi.org/10.1074/jbc.M300996200) PMID: [12759355](https://pubmed.ncbi.nlm.nih.gov/12759355/)
54. Chu IM, Hengst L, Slingerland JM. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer*. 2008; 8: 253–267. doi: [10.1038/nrc2347](https://doi.org/10.1038/nrc2347) PMID: [18354415](https://pubmed.ncbi.nlm.nih.gov/18354415/)
55. Duncan TJ, Al-Attar A, Rolland P, Harper S, Spendlove I, Durrant LG. Cytoplasmic p27 expression is an independent prognostic factor in ovarian cancer. *Int J Gynecol Pathol*. 2010; 29: 8–18. doi: [10.1097/PGP.0b013e3181b64ec3](https://doi.org/10.1097/PGP.0b013e3181b64ec3) PMID: [19952944](https://pubmed.ncbi.nlm.nih.gov/19952944/)
56. Ahmed J, Meinel T, Dunkel M, Murgueitio MS, Adams R, Blasse C, et al. CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res*. 2011; 39: D960–D967. doi: [10.1093/nar/gkq910](https://doi.org/10.1093/nar/gkq910) PMID: [20952398](https://pubmed.ncbi.nlm.nih.gov/20952398/)
57. Romanova LY, Willers H, Blagosklonny MV, Powell SN. The interaction of p53 with replication protein A mediates suppression of homologous recombination. *Oncogene*. 2004; 23: 9025–9033. doi: [10.1038/sj.onc.1207982](https://doi.org/10.1038/sj.onc.1207982) PMID: [15489903](https://pubmed.ncbi.nlm.nih.gov/15489903/)
58. Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol*. 2010; 11: 196–207. doi: [10.1038/nrm2851](https://doi.org/10.1038/nrm2851) PMID: [20177395](https://pubmed.ncbi.nlm.nih.gov/20177395/)
59. Kumar GR, Shum L, Glaunsinger BA. Importin  $\alpha$ -mediated nuclear import of cytoplasmic poly(A) binding protein occurs as a direct consequence of cytoplasmic mRNA depletion. *Mol Cell Biol*. 2011; 31: 3113–3125. doi: [10.1128/MCB.05402-11](https://doi.org/10.1128/MCB.05402-11) PMID: [21646427](https://pubmed.ncbi.nlm.nih.gov/21646427/)
60. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*. 2009; 25: 2906–2912. doi: [10.1093/bioinformatics/btp543](https://doi.org/10.1093/bioinformatics/btp543) PMID: [19759197](https://pubmed.ncbi.nlm.nih.gov/19759197/)
61. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci USA*. 2013; 110: 4245–4250. doi: [10.1073/pnas.1208949110](https://doi.org/10.1073/pnas.1208949110) PMID: [23431203](https://pubmed.ncbi.nlm.nih.gov/23431203/)