RESEARCH

A flexible penalized integrated analysis of mRNA and miRNA expression levels as biomarkers for endometrial cancer classification

Yang Xu[1], Donghai Dai[2] and Kai Wang[3*]

*Correspondence:
kai-wang@uiowa.edu
[3]Department of Biostatistics, College of Public Health, University of Iowa, 145 N. Riverside Drive, 52242 Iowa City, USA
Full list of author information is available at the end of the article

Abstract

Background: The expression levels of mRNAs and miRNAs in cancer tumors contain valuable information for cancer classification. The availability of The Cancer Genome Atlas (TCGA) data provides a unique opportunity for extracting such information. An integrative method that incorporates information from both mRNA and microRNA and relates them to clinical features is desirable.

Results: We propose a flexible LASSO method for identifying mRNA and miRNA signatures for important clinical features in endometrial tumors. These features are pathological stage, histological type, and grade.

Conclusions: The proposed method has identified mRNAs and miRNAs that are known to be risk factors in endometrial cancer. It also identified some novel ones which warrant further investigation. This approach is applicable to studies that involve more types of features than two.

Keywords: LASSO; endometrial cancer; stage; grade; type; microRNA; mRNA; cancer classification

Background

Endometrial cancer is the most common gynecological malignancy in US. There are 49,560 new cases and 8,190 deaths in 2013 alone [1]. While the prognosis for the early diseases is largely favorable, the outcome for the advanced and recurrent cases are poor. Development of effective biomarkers for stratification of patients for the best treatment regimens is thus critical to reduce mortality. However, current clinical classification and risk stratification are largely based on the histopathological features. There are no molecular markers in clinical use that can effectively predict patient outcomes or guide clinical management.

One promising molecular feature is tumor gene expression analysis. Differential gene expression has been reported in many types of cancer [2–6] including endometrial cancer [7]. However, most gene expression analyses were performed on arrays containing a limited set of genes. A promising emerging high throughput technique is the next generation RNA-sequencing.

Another important molecular feature is microRNA (miRNA) expression. MiRNA are short RNA molecules that do not encode any protein but regulate gene expression at the post-transcriptional level. MiRNAs regulate the expression of their target mRNAs through direct base-pairing [8]. More than 1000 miRNAs have been identified in humans. It has been predicted that more than 60% of mRNAs may be

regulated by miRNAs and that a single miRNA may target as many as 200 mRNAs [9–11]. MiRNAs are involved in the control of major cellular processes, such as cell proliferation, metabolism, developmental timing, stem cell division, differentiation, apoptosis and oncogenesis [12–17].

MiRNA expression profiling has been used for classification of different types of cancer and for predicting clinical outcomes. Some important works have been done in endometrial cancer as well [18–22]. A signature of 20 miRNAs is found to be uniquely expressed in endometrial cancer when compared with benign disease [23]. A signature of 17 miRNAs has been found to distinguish between endometrioid and serous adenocarcinoma, two most common types of endometrial cancer [23]. Differential expression patterns have been identified in endometrial and ovarian cancers, as well as in endometrioid and serous types of carcinoma, representing their organ of origin [24]. While gene expression and miRNA expression are well-studied separately as the molecular features of oncogenic processes, their integrated role has not been well-elucidated. Integration of these two molecular profiling as one signature could be used to develop novel biomarkers and could have important clinical implication.

The Cancer Genome Atlas (TCGA) project is a large-scale collaborative effort aimed at understanding the cancer genome in a comprehensive manner (`http://cancergenome.nih.gov`). Using the latest genome analysis technologies including next-generation sequencing, it has generated tremendous amount of genomic data on a number of cancer tissues collected with the most stringent criteria. Findings from this valuable resource has contributed significantly to our knowledge of various types of human cancer [25], including endometrial cancer [26–29]. We take advantage of this wealth of data for an integrated analysis of miRNAs and mRNAs in relation to endometrial tumor classifications including stage, grade, and type.

The LASSO [30] is a promising statistical method designed to disentangle "large data" like the one we are facing. A naive approach to integrated analysis with both mRNA and miRNA expression might be to use them simultaneously as predictors. However, since miRNA and mRNA expression are from different "domains" and differ in number by folds, the effect of one type of predictors can be dominated by the other type of predictors (see Discussion). To avoid this problem, we introduce an extra parameter such that mRNA and miRNA expressions are regulated separately. This modification to the regular LASSO allows us to study the association of mRNA and miRNA to a clinical feature *simultaneously*. The co-expression of certain mRNAs and miRNAs can serve as a signature for endometrial cancer classification. These molecular-level biomarkers should be complementary to the traditional histopathological features.

This report is organized as follows. In the next section, we describe our proposed method in detail where the foci are on on how the clinical outcomes, which are distributed either binomial or multinomial, are modeled and the cross-validation procedure for selecting the values for the two tuning parameters. After that, the data used for the analysis is described. The result section reports our findings. In the discussion section we provide a summary of our work and provide directions of future generalizations.

Methods

Modelling mRNAs and miRNAs in Lasso regression by two penalty parameters

Suppose that there are p mRNA expressions and q miRNA expressions. Let n be the number of patient samples. The three clinical phenotypes we consider include stage, histological type, and neoplasm histologic grade. All these outcomes are categorical: there are four levels (I, II, III and IV) for clinical stage, two levels (endometrioid endometrial adenocarcinoma and serous endometrial adenocarcinoma) for histological type, and three levels (1, 2 and 3) for neoplasm histologic grade. Therefore we apply generalized linear models when constructing the loss function.

For histological type, the systematic component is

$$\beta_0 + \sum_{j=1}^{p+q} \beta_j x_j,$$

where x_j is the expression level of a mRNA ($j \leq p$) or a miRNA ($j > p$) and β_j is its regression coefficient. Let i be the index of subjects. We consider the following penalized logistic regression:

$$\min_{\boldsymbol{\beta}} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \log L(\boldsymbol{\beta}; \{y_i\}, \{x_{ij}\}) + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=p+1}^{p+q} |\beta_j| \right\}. \tag{1}$$

Here $L(\boldsymbol{\beta}; \{y_i\}, \{x_i\})$ denotes the usual logistic regression likelihood with $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p+q})^t$, a column vector of regression coefficients. The penalty consists of two parts, one on mRNAs and the other on miRNAs, with separate tuning parameters λ_1 and λ_2, respectively. This makes it different from standard LASSO penalty which involves only one tuning parameter by requiring $\lambda_1 = \lambda_2$. We are considering jointly the effect of mRNA and miRNA on clinical phenotypes. Having two tuning parameters allow their effects to be regularized separately.

To fit model (1), we make use of R package `glmnet` [31] by taking advantage of its option `penalty.factor` in combination with the `lambda` option for the `glmnet` function. We have written an R code that invokes the `glmnet` function with specified values for λ_1 and λ_2. More importantly, this R code implements a cross-validation procedure for selecting values for λ_1 and λ_2 (see next subsection).

For clinical stage and neoplasm histologic grade that are multinomial, we use the multi-logit model implemented in `glmnet`. Specifically, this implementation specifies

$$\Pr(y = l|\mathbf{x}) = \frac{\exp\{\beta_{0l} + \mathbf{x}^t \boldsymbol{\beta}_l\}}{\sum_{k=1}^{K} \exp\{\beta_{0k} + \mathbf{x}^t \boldsymbol{\beta}_k\}}$$

where $\mathbf{x}$ is the vector of predictors and K is the number of categories for response y. Unlike the traditional multi-logit model such as baseline-category logit model, this `glmnet` implementation is over-parameterized. However it constitutes no difficulty for regularized regression [31].

Let $L(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K; \{y_i\}, \{x_i\})$ denotes the multi-logit regression likelihood with $\boldsymbol{\beta}_k = (\beta_{1k}, \ldots, \beta_{p+q,k}$. We consider the following multi-logit LASSO regression

problem:

$$\min_{\boldsymbol{\beta}} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \log L(\boldsymbol{\beta}; \{y_i\}, \{x_{ij}\}) + \lambda_1 \sum_{j=1}^{p}\sum_{k=1}^{K} |\beta_{jk}| + \lambda_2 \sum_{j=p+1}^{p+q}\sum_{k=1}^{K} |\beta_{jk}| \right\}. \quad (2)$$

Similar to problem (1), the minimization in (2) is solved by making use of the R package `glmnet`.

Cross-validation

The values for the tuning parameters λ_1 and λ_2 are determined by 5-fold cross-validation as follows. Data are randomly partitioned into 5 subsets. Because the response variables are categorical, this partitioning is carried out within each category of a response so that the sample sizes of each response category in each subset are (roughly) proportional to their values in the whole data [32]. For example, there are three groups for neoplasm histological grade: 1, 2 and 3. Random partitioning was carried in each subgroup so that all training and validation sets would have sample size ratio as close as possible to the whole sample.

Specifically, the procedure of cross-validation goes as follows:

1. Randomly partition the data in the way just described.
2. Each subset is then used in turn as the testing data while the remaining 4 subsets as the training data.
3. For each combination (λ_1, λ_2) where λ_1 ranges from 0.01 to 0.99 with step 0.01 and λ_2/λ_1 ranges from 0.1 to 0.9 with step 0.1, a set of mRNAs and miRNAs are obtained from the training data. These selected predictors are applied to the testing data by using the `predict` function with option `type="class"` in the `glmnet` R package. This option classifies each observation into the most likely response category. We then calculated the misclassification error rate (MER), which is defined as

$$\text{MER} = \frac{\text{number of misclassification}}{\text{number of validation sample size}}.$$

The λ_1 and λ_2 pair that gives the lowest MER is used for the final analysis. The search range for λ_1 and that for λ_2/λ_1 are determined based based on our preliminary analyses with wider range but cruder step. The range of λ_2/λ_1 is also hinted by results from regular LASSO (i.e. $\lambda_2/\lambda_1 = 1$) where selected features are predominantly mRNAs. For $\lambda_2/\lambda_1 > 1$, the situation is exacerbated.

Data

Permission was obtained to access The Cancer Genome Atlas (TCGA) database (`http://cancergenome.nih.gov/cancersselected/endometrial`). The mRNA and miRNA expression data were downloaded from Uterine Corpus Endometrial Carcinoma (UCEC) of TCGA database. There are 433 patients who have both mRNA and miRNA expression data. Two histologic types (endometrioid and serous), four stages (i, II, III, and IV), and three grades (1, 2, and 3) were used in the analysis of expression level of mRNAs and miRNAs. Sample size of each clinical

data is summarized in table 1. As a quality control measure, mRNAs and miRNAs whose non-detection (i.e., 0 expression level) rate not less than 90% are removed. Finally, there are 19604 mRNAs and 677 miRNAs for this study. Their expression level are expressed as the number of reads per kilobase per million (RPKM) defined by:

$$\text{RPKM} = \frac{\text{number of reads on a gene}}{(\text{length of gene in kb}) \times (\text{total number of reads in millions})}.$$

Results and Discussion

The complete list of the selected mRNAs and miRNAs by the two-λ LASSO is reported in the additional file. Table 2 provides a summary. Venn diagrams of these identified mRNAs and miRNAs are presented in figures 1, 2, and 3. These Venn diagrams are prepared by using the program VennPlex [33].

Some of the identified miRNAs has been reported before. Hsa-let-7g increases the odds for grade 2 tumor, but decreases the odds for grade 3 tumor. Let-7i increase the odds for grade 3 tumor. Interestingly, increased expression of let-7b and 7g both decreases the odds for endometrioid type and increase the odds for serous type of endometrial cancer, suggesting that let-7 family member could be potentially a quantitative marker to distinguish tumor grade and histological types in endometrial cancer. Let-7 family has been shown to be involved in lung cancers and is associated with poor prognosis [34]. Let-7 genes are located at regions deleted in many human cancers. Let-7 is reported to inhibit the Ras family of oncogenes and c-Myc gene function [35, 36]. MiR-221 has been implicated in many cancers including stomach, colon, pancreas, liver, bladder, thyroid cancer, glioblastoma [37], and our own study of endometrioid endometrial cancer (data not published). MiR-9-1 and 9-2 have been among the signature for serous tumor. MiR-9 is often downregulated in colorectal cancer and medulloblastoma [38]. Epigenetic change of miR-9-1 is associated with poor prognosis in colorectal cancer with increased lymph node metastasis. Additionally, miR-29c and miR-22 is among the signature for high grade tumor (grade 3). MiR-29c is shown to be over-expressed in many cancer [39] and miR-22 is associated with chemoresistance in ovarian cancer [40]. MiR132 also increases the odds for high grade tumor and is reported to be associated with gemcitabline resistance [41]. In addition, 4 miRNAs (hsa-mir-185 and has-mir-429 for grade 3, hsa-mir-205 and has-mir-221 for type) are among 18 miRNAs reported previously [42].

Overlaps between the identified mRNAs and Molecular Signatures Database (MSigDB, v4.0) [43] were computed. Several mRNA sets are involved in come oncogene pathways that may contribute to cancer pathology. For stage I, XPO5, STK38, RTN3, CIRBP, TCF25, and SORC53 have similar promoter site containing a motif of nuclear respirator factor 1, which is believed to regulate endogenous estrogen receptor β, a well-studied receptor in human breast cancer [44]. For grade 2, HSPG2 and KRT13 genes were found to be up-regulated upon active Stat3, a signal transducer and activator of transcription 3 frequently detected in wide range of human cancer, including endometrial cancer [45]. For grade 3, RAE1, TRAF5, CD1C, ADAM28, PDLIM1 and CHST2 genes were up-regulated in B cells, which may indicate enhancement of cancer cell immunization activity [42]. For histological

type, CDR2L and EPHB2 were classified as cancer gene neighborhood of PRKACA, a protein kinase, whose aberrant function could cause various human cancer [46]. None of the identified mRNAs were found to be targets of the identified miRNAs.

Conclusions

We have proposed a two-parameter LASSO method for an integrated analysis that relates miRNA and mRNA expression levels to important clinical outcomes of endometrial cancer. The flexibility of this method allows us to select miRNAs and mRNAs simultaneously in a way the effect of miRNAs is not dominated by mRNAs as the latter is far more numerous. Indeed, if no distinction between mRNA and miRNA is made and a regular LASSO that has single λ is used, the majority of the selected features are mRNA (table 3).

For the endometrial cancer data we used, this two-λ LASSO method has identified some interesting results. Some of them have been implicated in previous studies. Some are related to known genes by gene set enrichment analysis. These results warrant further investigation.

In summary, the proposed two-λ LASSO method provides a flexible way for identifying mRNAs and miRNAs signatures for endometrial cancer. We are generalizing this approach to include more types of genetic features such as SNP genotypes, copy number variations, and methylation. We are also applying this method to other types of cancer.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
DD and KW participated in study design. YX performed statistical analyses. All authors aided in the manuscript preparation.

Acknowledgements
We thank Dr. Patrick Breheny for bring the `penalty.factor` option of the `glmnet` function in the `glmnet` package to our attention.

Author details
[1]Interdisciplinary Graduate Program in Informatics, University of Iowa, 3087 Main Library, 52242 Iowa City, USA. [2]Department of Obstetrics and Gynecology, Carver College of Medicine, University of Iowa, 200 Hawkins Drive, 52242 Iowa City, USA. [3]Department of Biostatistics, College of Public Health, University of Iowa, 145 N. Riverside Drive, 52242 Iowa City, USA.

References

1. American Cancer Society: Cancer Facts N Figures 2013. American Cancer Society, Atlanta, GA (2013)
2. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V.: Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature **406**, 536–540 (2000)
3. Akavia, U.D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H.C., Pochanard, P., Mozes, E., Garraway, L.A., Pe'er, D.: An integrated approach to uncover drivers of cancer. Cell **143**(6), 1005–10017 (2010)
4. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**, 531–537 (1999)
5. van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., Friend, S.H.: Gene expression profiling predicts clinical outcome of breast cancer. Nature **415**(6871), 530–536 (2002)
6. Bettuzzi, S., Scaltriti, M., Caporali, A., Brausi, M., D'Arca, D., Astancolle, S., Davalli, A. P. an Corti: Successful prediction of prostate cancer recurrence by gene profiling in combination with clinical data: a 5-year follow-up study. Cancer Res **63**(13), 3469–3472 (2003)

7. Risinger, J.I., Maxwell, G.L., Chandramouli, G.V., Jazaeri, A., Aprelikova, O., Patterson, T., Berchuck, A., Barrett, J.C.: Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer. Cancer Res **63**, 6–11 (2003)
8. Ventura, A., Jacks, T.: MicroRNAs and cancer: short RNAs go a long way. Cell **136**, 586–591 (2009)
9. Friedman, R.C., Farh, K.K., Burge, C.B., Bartel, D.P.: Most mammalian mRNAs are conserved targets of microRNAs. Genome Res **19**, 92–105 (2009)
10. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J.: miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res **34**, 140–144 (2006)
11. Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., Johnson, J.M.: Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature **433**, 769–773 (2005)
12. Cheng, A.M., Byrom, M.W., Shelton, J., Ford, L.P.: Antisense inhibition of human miRNAs and indications for an involvement of mirna in cell growth and apoptosis. Nucleic Acids Res **33**, 1290–1297 (2005)
13. Xu, P., Guo, M., Hay, B.A.: MicroRNAs and the regulation of cell death. Trends Genet **20**, 617–624 (2004)
14. Guo, H., Ingolia, N.T., Weissman, J.S., Bartel, D.P.: Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature **466**, 835–840 (2010)
15. Krutzfeldt, J., Stoffel, M.: MicroRNAs: a new class of regulatory genes affecting metabolism. Cell Metab **4**, 9–12 (2006)
16. Nimmo, R.A., Slack, F..J.: An elegant miRror: microRNAs in stem cells, developmental timing and cancer. Chromosoma **118**, 405–418 (2009)
17. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., Croce, C.M.: Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A **101**, 2999–3004 (2004)
18. Boren, T., Xiong, Y., Hakam, A., Wenham, R., Apte, S., Wei, Z., Kamath, S., Chen, D.T., Dressman, H., Lancaster, J.M.: MicroRNAs and their target messenger RNAs associated with endometrial carcinogenesis. Gynecol Oncol **110**, 206–215 (2008)
19. Chung, T.K., Cheung, T.H., Huen, N.Y., Wong, K.W., Lo, K.W., Yim, S.F., Siu, N.S., Wong, Y.M., Tsang, P.T., Pang, M.W., Yu, M.Y., To, K.F., Mok, S.C., Wang, V.W., Li, C., Cheung, A.Y., Doran, G., Birrer, M.J., Smith, D.I., Wong, Y.F.: Dysregulated microRNAs and their predicted targets associated with endometrioid endometrial adenocarcinoma in hong kong women. Int J Cancer **124**, 1358–1365 (2009)
20. Wu, W., Lin, Z., Zhuang, Z., Liang, X.: Expression profile of mammalian microRNAs in endometrioid adenocarcinoma. Eur J Cancer Prev **18**, 50–55 (2009)
21. Hiroki, E., Akahira, J., Suzuki, F., Nagase, S., Ito, K., Suzuki, T., Sasano, H., Yaegashi, N.: Changes in microrna expression levels correlate with clinicopathological features and prognoses in endometrial serous adenocarcinomas. Cancer Sci **101**, 241–249 (2010)
22. Ratner, E.S., Tuck, D., Richter, C., Nallur, S., Patel, R.M., Schultz, V., Hui, P., Schwartz, P.E., Rutherford, T.J., Weidhaas, J.B.: MicroRNA signatures differentiate uterine cancer tumor subtypes. Gynecol Oncol **118**, 251–257 (2010)
23. Devor, E.J., Hovey, A.M., Goodheart, M.J., Ramachandran, S., Leslie, K.K.: microRNA expression profiling of endometrial endometrioid adenocarcinomas and serous adenocarcinomas reveals profiles containing shared, unique and differentiating groups of microRNAs. Oncol Rep **26**, 995–1002 (2011)
24. Zorn, K.K., Bonome, T., Gangi, L., Chandramouli, G.V., Awtrey, C.S., Gardner, G.J., Barrett, J.C., Boyd, J., Birrer, M.J.: Gene expression profiles of serous, endometrioid, and clear cell subtypes of ovarian and endometrial cancer. Clin Cancer Res **11**, 6422–6430 (2005)
25. Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., Leiserson, M.D., Miller, C.A., Welch, J.S., Walter, M.J., Wendl, M.C., Ley, T.J., Wilson, R.K., Raphael, B.J., Ding, L.: Mutational landscape and significance across 12 major cancer types. Nature **502**, 333–339 (2013)
26. Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., Yau, C., Laird, P.W., Ding, L., Zhang, W., Mills, G.B., Kucherlapati, R., Mardis, E.R., Levine, D.A.: Integrated genomic characterization of endometrial carcinoma. Nature **497**, 67–73 (2013)
27. Kuhn, E., Wu, R.C., Guan, B., Wu, G., Zhang, J., Wang, Y., Song, L., Yuan, X., Wei, L., Roden, R.B., Kuo, K.T., Nakayama, K., Clarke, B., Shaw, P., Olvera, N., Kurman, R.J., Levine, D.A., Wang, T.L., Shih, I.: Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses. J. Natl. Cancer Inst **104**, 1503–1513 (2012)
28. Le, G.M., O'Hara, A.J., Rudd, M.L., Urick, M.E., Hansen, N.F., O'Neil, N.J., Price, J.C., Zhang, S., England, B.M., Godwin, A.K., Sgroi, D.C., Hieter, P., Mullikin, J.C., Merino, M.J., Bell, D.W.: Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. Nat Genet **44**, 1310–1315 (2012)
29. Liang, H., Cheung, L.W., Li, J., Ju, Z., Yu, S., Stemke-Hale, K., Dogruluk, T., Lu, Y., Liu, X., Gu, C., Guo, W., Scherer, S.E., Carter, H., Westin, S.N., Dyer, M.D., Verhaak, R.G., Zhang, F., Karchin, R., Liu, C.G., Lu, K.H., Broaddus, R.R., Scott, K.L., Hennessy, B.T., Mills, G.B.: Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. Genome Res **22**, 2120–2129 (2012)
30. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B **58**(1), 267–288 (1996)
31. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software **33**(1), 1–22 (2010)
32. Hess, K.R., Anderson, K., Symmans, W.F., Valero, N. V. Ibrahim, Mejia, J.A., Booser, D., Theriault, R.L., Buzdar, A.U., Dempsey, P.J., Rouzier, R., Sneige, N., Ross, J.S., Vidaurre, T., Gómez, H.L., Hortobagyi, G.N., Pusztai, L.: Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol **24**(26), 4236–4244 (2006)
33. Cai, H., Chen, H., Yi, T., Daimon, C.M., Boyle, J.P., Peers, C., Maudsley, S., Martin, B.: VennPlex–A novel venn diagram program for comparing and visualizing datasets with differentially regulated datapoints. PLoS

One **8**(1), 53368 (2013)
34. Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T., Calin, G.A., Liu, C.G., Croce, C.M., Harris, C.C.: Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell **9**, 189–198 (2006)
35. Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., Slack, F.J.: RAS is regulated by the let-7 microRNA family. Cell **120**, 635–647 (2005)
36. Sampson, V.B., Rong, N.H., Han, J., Yang, Q., Aris, V., Soteropoulos, P., Petrelli, N.J., Dunn, S.P., Krueger, L.J.: MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res **67**, 9762–9770 (2007)
37. Negrini, M., Nicoloso, M.S., Calin, G.A.: MicroRNAs and cancer–new paradigms in molecular oncology. Curr Opin Cell Biol **21**, 470–479 (2009)
38. Cho, W.C.: Micrornas: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. Int J Biochem Cell Biol **42**, 1273–1281 (2010)
39. Navon, R., Wang, H., Steinfeld, I., Tsalenko, A., Ben-Dor, A., Yakhini, Z.: Novel rank-based statistical methods reveal microRNAs with differential expression in multiple cancer types. PLoS. ONE **4**, 8003 (2009)
40. Yang, N., Kaur, S., Volinia, S., Greshock, J., Lassus, H., Hasegawa, K., Liang, S., Leminen, A., Deng, S., Smith, L., Johnstone, C.N., Chen, X.M., Liu, C.G., Huang, Q., Katsaros, D., Calin, G.A., Weber, B.L., Butzow, R., Croce, C.M., Coukos, G., Zhang, L.: MicroRNA microarray identifies let-7i as a novel biomarker and therapeutic target in human epithelial ovarian cancer. Cancer Res **68**, 10307–10314 (2008)
41. Boren, T., Xiong, Y., Hakam, A., Wenham, R., Apte, S., Chan, G., Kamath, S.G., Chen, D.T., Dressman, H., Lancaster, J.M.: MicroRNAs and their target messenger RNAs associated with ovarian cancer response to chemotherapy. Gynecol Oncol **113**, 249–255 (2009)
42. Banno, K., Yanokura, M., Kisu, I., Yamagami, W., Susumu, N., Aoki, D.: MicroRNAs in endometrial cancer. Int J Clin Oncol **18**(2), 186–192 (2013)
43. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, B.L. S. Ebert, Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A **102**(43), 15545–15550 (2005)
44. Ivanova, M.M., Luken, K.H., Zimmer, A.S., Lenzo, F.L., Smith, R.J., Arteel, W., M., Kollenberg, T.J., Mattingly, K.A., Klinge, C.M.: Tamoxifen increases nuclear respiratory factor 1 transcription by activating estrogen receptor beta and AP-1 recruitment to adjacent promoter binding sites. FASEB J **25**(4), 402–416 (2011)
45. Chen, C.L., Hsieh, F.C., Lieblein, J.C., Brown, J., Chan, C., Wallace, J.A., Cheng, G., Hall, B.M., Lin, J.: Stat3 activation in human endometrial and cervical cancers. Br J Cancer **96**(4), 591–599 (2007)
46. Almeida, M.Q., Stratakis, C.A.: How does cAMP/protein kinase A signaling lead to tumors in the adrenal cortex and other tissues? Mol Cell Endocrinol **336**(1-2), 162–168 (2011)

Figures

Figure 1 Selected mRNAs and miRNAs for stage by the two-λ LASSO method. Effect direction of a feature is based on the sign of its coefficient. Common elements in "Stage I" and "Stage III": L1CAM, MURC.

Figure 2 Selected mRNAs and miRNAs for grade by the two-λ LASSO method. Effect direction of a feature is based on the sign of its coefficient. Common elements in "Grade 1" and "Grade 2": hsa-mir-1303, hsa-mir-603, hsa-mir-3138. Common elements in "Grade 1" and "Grade 3": hsa-mir-628. Common elements in "Grade 2" and "Grade 3": hsa-mir-874, hsa-let-7g, hsa-mir-184.

Figure 3 A summary of mRNAs and miRNAs for the three clinical outcomes selected by the two-λ LASSO method. Effect direction of a feature is based on the sign of its coefficient. Common elements in "Stage" and "Grade": hsa-mir-3199-1. Common elements in "Grade" and "Type": hsa-let-7g, hsa-mir-34a, hsa-mir-660, hsa-mir-598. Common elements in "Stage" and "Type": L1CAM, SMPD4, PKIA.

Tables

Table 1 Sample size for each clinical feature for the 433 subjects

Stage	I	279
	II	34
	III	97
	IV	23
Grade	1	85
	2	101
	3	247
Histological Type	Endometrioid	337
	Serous	96

Table 2 A summary of the selected miRNAs and mRNAs by the two-λ LASSO method.

Clinical Feature		Effect	miRNA	mRNA	Total
Stage	I	$-$	1	23	24
		$+$	2	6	8
	II	$+$	1	3	4
	III	$+$	1	11	12
	IV	$+$	1	3	4
Grade	1	$-$	9	0	9
		$+$	7	16	23
	2	$-$	12	1	13
		$+$	19	6	25
	3	$-$	11	26	37
		$+$	20	13	33
Histological Type		$-$	20	3	23
		$+$	19	22	41

Table 3 A summary of the selected miRNAs and mRNAs by the one-λ LASSO method.

Clinical Feature		Effect	miRNA	mRNA	Total
Stage	I	$-$	0	14	14
		$+$	0	5	5
	III	$+$	1	7	8
Grade	1	$+$	0	16	16
	2	$-$	0	1	1
		$+$	1	4	5
	3	$-$	2	38	40
		$+$	3	17	20
Histological Type		$-$	0	5	5
		$+$	1	46	47

Additional Files

Additional file — Complete list of the identified mRNAs and miRNAs by the two-λ LASSO method
The file name is `Additional file.xlsx`. It is in Microsoft Excel format.

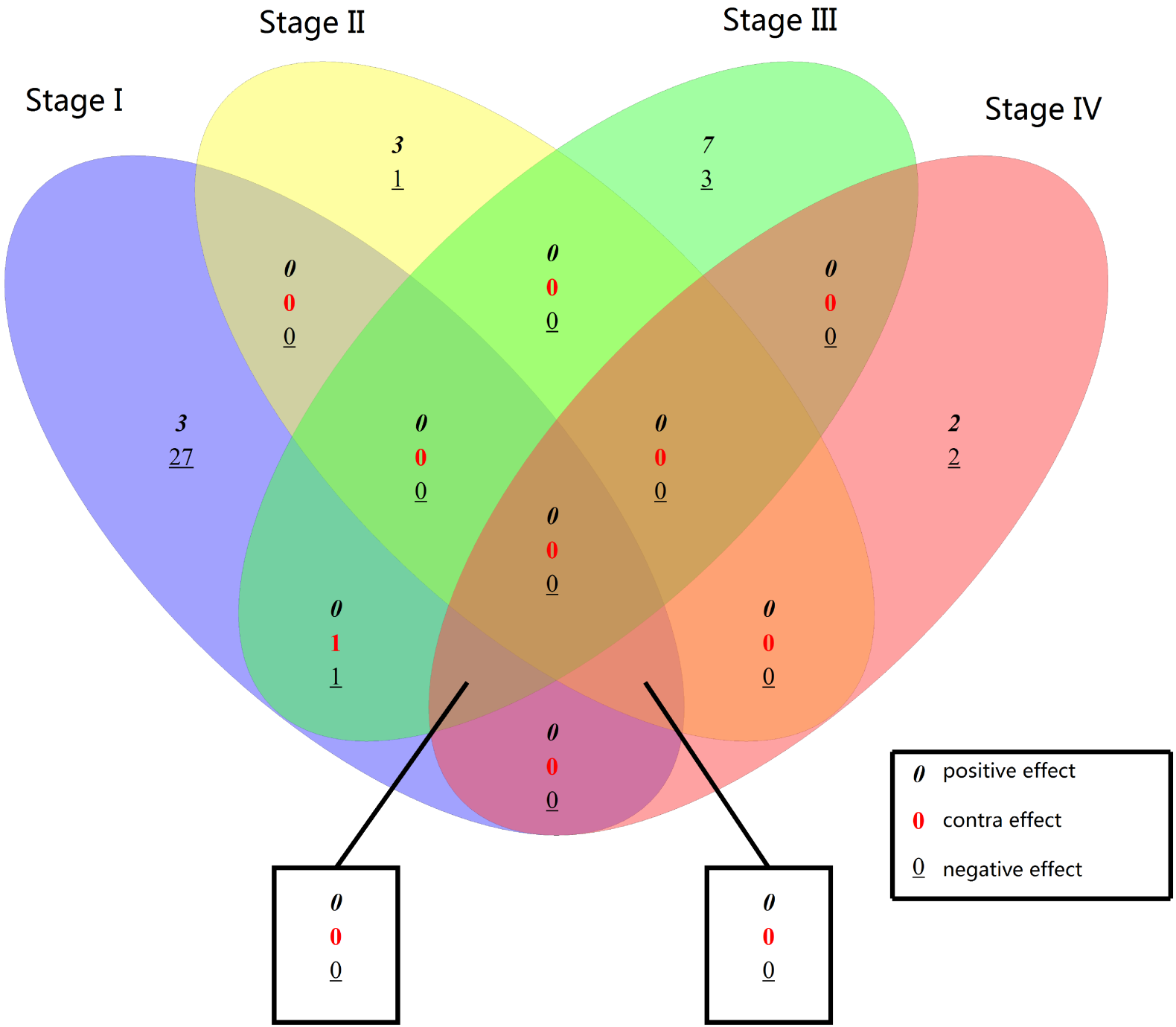
Stage I
Stage II
Stage III
Stage IV
3
27
3
1
7
3
2
2
0
0
0
0
1
1
positive effect
contra effect
negative effect

Figure 1

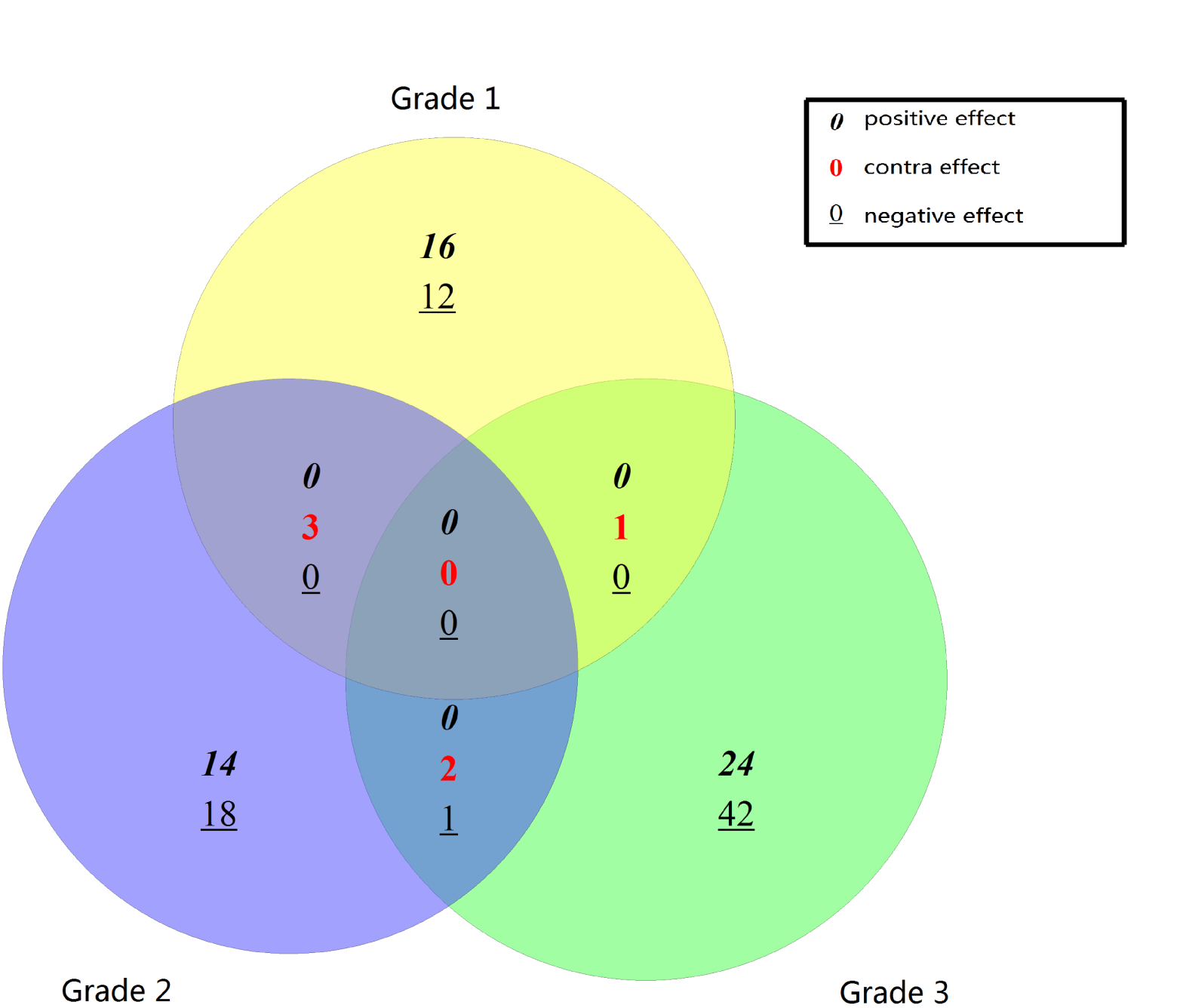
Grade 1
0 positive effect
0 contra effect
0 negative effect
16
12
0
3
0
0
0
0
0
1
0
0
0
2
1
14
18
24
42
Grade 2
Grade 3

Figure 2

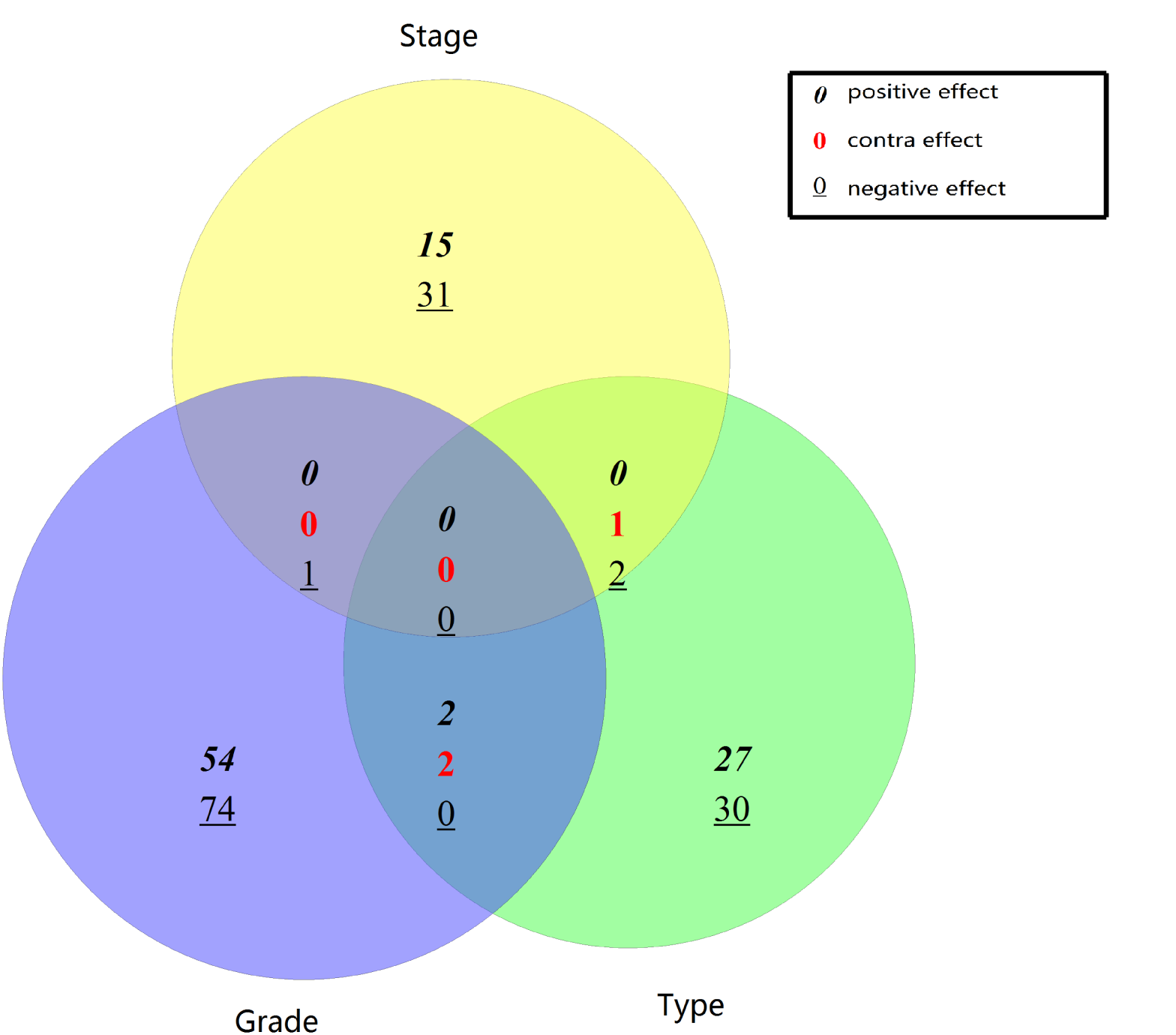
Stage
0 positive effect
0 contra effect
0 negative effect
15
31
0
0
1
0
0
0
0
1
2
2
2
0
54
74
27
30
Grade
Type

Figure 3

Additional files provided with this submission:

Additional file 1: Additional File.xlsx, 16K
http://www.biomedcentral.com/imedia/1310070062118824/supp1.xlsx