Cancer Epidemiology 61 (2019) 129–132

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Defining a mutational signature for endometrial cancer screening and early detection

Laura Costas[a,*,1], Luis Palomero[b,1], Yolanda Benavente[a,c], Magdalena Guardiola[a], Jon Frias-Gomez[a], Miquel Ángel Pavón[a,c], Maite Climent[d], José Manuel Martinez[d], Marc Barahona[d], Mónica Salinas[e], Marta Pineda[e], Ilaria Bianchi[f], Jaume Reventós[g], Gabriel Capellà[e,h], Mireia Diaz[a,h], August Vidal[h,i], Josep Maria Piulats[h,j], Jordi Ponce[d], Joan Brunet[e,h,k], Francesc Xavier Bosch[a,h], Xavier Matias-Guiu[h,i], Laia Alemany[a,c], Silvia de Sanjosé[l,1], Álvaro Aytés[b,*,1], on behalf of the Screenwide Team

[a] Cancer Epidemiology Research Programme, IDIBELL, Catalan Institute of Oncology, Hospitalet de Llobregat, Barcelona, Spain
[b] Program Against Cancer Therapeutic Resistance (ProCURE), IDIBELL, Catalan Institute of Oncology, Hospitalet de Llobregat, Barcelona, Spain
[c] Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública, CIBERESP), Spain
[d] Department of Gynecology and Obstetrics, Hospital Universitari de Bellvitge, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain
[e] Hereditary Cancer Program, IDIBELL, Catalan Institute of Oncology, Hospitalet de Llobregat, Barcelona, Spain
[f] ASSIR Delta, Direcció d'Atenció Primària Costa de Ponent, SAP Delta del Llobregat, Gerència Territorial Metropolitana Sud / ICS, Barcelona, Spain
[g] Departament de Ciències Bàsiques, Universitat Internacional de Catalunya, Barcelona, Spain
[h] Centro de Investigación Biomédica en Red en Cáncer (CIBERONC), Madrid, Spain
[i] Department of Pathology, Hospital Universitari de Bellvitge, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain
[j] Department of Medical Oncology, IDIBELL, Catalan Institute of Cancer, Hospitalet de Llobregat, Barcelona, Spain
[k] Medical Sciences Department, School of Medicine, University of Girona, Girona, Spain
[l] PATH, Seattle, United States

ARTICLE INFO

Keywords:
Endometrial cancer
Early detection
Screening
Genomics
Biomarkers
Algorithm

ABSTRACT

Introduction: The current availability of genomic information represents an opportunity to develop new strategies for early detection of cancer. New molecular tests for endometrial cancer may improve performance and failure rates of histological aspirate-based diagnosis, and provide promising perspectives for a potential screening scenario. However, the selection of relevant biomarkers to develop efficient strategies can be a challenge.

Materials and methods: We developed an algorithm to identify the largest number of patients with endometrial cancer using the minimum number of somatic mutations based on The Cancer Genome Atlas (TCGA) dataset.

Results: The algorithm provided the number of subjects with mutations (sensitivity) for a given number of biomarkers included in the signature. For instance, by evaluating the 50 most representative point mutations, up to 81.9% of endometrial cancers can be identified in the TCGA dataset. At gene level, a 92.9% sensitivity can be obtained by interrogating five genes.

Discussion: We developed a computational method to aid in the selection of relevant genomic biomarkers in endometrial cancer that can be adapted to other cancer types or diseases.

1. Introduction

The rapid advances in medical sciences, including genomics and bioinformatics, promises a new era of biomedical research and precision medicine [1,2]. The term "*precision public health*" has consequently been coined to describe disease prevention by applying novel technologies and developing targeted personalized public health programs. New genomic and epigenomic-based diagnostics have shown the capacity for early detection of gynecologic cancers in minimally invasive sampling methods [3,4]. However, novel approaches that improve

* Corresponding authors at: Catalan Institute of Oncology, IDIBELL, Cancer Epidemiology Research Program, Epidemiología 2ª Planta, Granvia de L'Hospitalet 199-203, 08908 L'Hospitalet de Llobregat, Spain.
E-mail addresses: lcostas@iconcologia.net (L. Costas), aaytes@idibell.cat (Á. Aytés).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.canep.2019.06.003
Received 9 November 2018; Received in revised form 29 March 2019; Accepted 11 June 2019
Available online 22 June 2019
1877-7821/

patient outcomes are generally associated with increased costs, complicating its applicability at the general population level in cost-effective programs. Targeted sequencing allows the interrogation of a large number of genetic variants but it has a relatively high cost [5]. An early detection technology focusing on few selected variants could potentially be more efficient, although the selection of relevant biomarkers is a challenge. We developed an algorithm to identify the minimum number of genetic variants that capturing the largest number of endometrial cancer patients to support the development of new potential preventive strategies.

2. Methods

We adapted a greedy algorithm that provided the number of subjects with mutations (sensitivity) for a given number of variants included in the signature. The algorithm was implemented and executed over the complete set of somatic mutations included at The Cancer Genome Atlas (TCGA) for endometrial cancer (TCGA-UCEC) [6]. Type I cancers (endometrioid histology) are the most common group, while type II tumours (non-endometrioid) represent 10–15% of endometrial cancers. In the present analysis, we included genomic data from 548 endometrial cancer cases, including 411 endometrioid, 115 serous and 22 mixed histology tumors.

2.1. Definition of the initial set of somatic mutations included in the pipeline

The somatic mutation data of all endometrial cancers at the TCGA (version DR-7) identified with any of the variant callers (MuSE, MuTect2, SomaticSniper or VarScan2) were downloaded from Genomic Data Commons data portal. Starting from a pool of 545,989 non silent mutations (missense, nonsense, insertions and deletions), 63,472, corresponding to 501 genes, were removed using Kohane et al. [7] criteria to exclude those not related to cancer development. After exclusions, 481,930 mutations in 18,403 genes from 530 different samples with identified mutations were included in the second step. The set of mutations per patient was validated against the summary of mutations detected in the corresponding TCGA OncoGrid for endometrial cancer, which is a publicly available graph at the TCGA website that summarizes the mutational landscape for most frequently mutated genes per disease.

2.2. Iterative execution of this greedy algorithm

The algorithm was run at the point mutation level and at gene level following these steps. Considering a set of *S* samples, the most frequent *p* mutation is chosen in each iteration –if more than one mutation is detected one of them is chosen randomly–. The set of samples harboring this mutation is designated as *s,* and the corresponding percentage of samples harboring the mutation is the score for the *p* mutation. Then, a new *S*' subset is considered by removing the *s* samples previously detected. In a new iteration this same algorithm is applied to this new *S*' subset, and a new score is computed, which reflects the total number of samples that are identified by adding that new *p* mutation to the panel. This iterative algorithm is executed until all samples are detected (see Fig. 1A). The percentage of identified samples is also described by histologic subtype (endometrioid, serous and mixed). To improve the internal validity of the signature, we applied bootstrapping techniques: 400 iterations were performed on randomized subsets of 40 samples. Each iteration generated a partial signature in which each mutation (*p*) has a score.

2.3. Consolidation of candidate signature

This last step of the algorithm consists of consolidating the partial signatures and eliminating those mutations that have shown poor results considering the averages of detections in the different iterations of the bootstrap. For this, a new score of all mutations is generated based on the average of the scores of the partial signatures. Given that the algorithm selects the more frequent mutations and the bootstrapping further eliminates those mutations that are not observed consistently across the different variant callers, the algorithm ultimately selects those mutations with higher variant allele fractions. Copies of the code of the algorithm are available upon request from the corresponding authors.

3. Results

The algorithm provided the sensitivity in this particular dataset (percentage of patients with mutations) given a specified number of variants included in a signature. Signatures at point mutation level and gene level, overall and by endometrial cancer subtype, are summarized in Table 1. While a signature of 5 point mutations only identified 35.4% of all endometrial cancers in the TCGA dataset, up to 81.9% could be identified with a signature of 50 point mutations, including 84.4% endometrioid cancers, 75.7% serous carcinomas, and 68.2% of mixed histology. Sensitivity data with larger signatures are available at Supplemental Table 1.

At the gene level, 57.3% of endometrial cancers carried a mutation in exonic regions of the tumor suppressor *PTEN* (signature A). As expected, *PTEN* identified most of the endometrioid endometrial cancers (72.3%), while by adding *TP53* (signature B) most of non-endometrioid cancers were identified (77.3% serous and 86.1% mixed). Signature E included 5 genes (*PTEN*, *TP53*, *PIK3CA, ARID1A* and *CTNNB1*), and identified 92.9% of the endometrial cancers in the dataset. Trends in sensitivity by increasing the number of mutations are shown in Fig. 1B (signatures at point mutations level) and Fig. 1C (at gene level). Similarly, trends in sensitivity stratified by endometrial cancer subtype, and corresponding oncogrids summarizing frequencies of each mutation can be found in Supplemental Fig. 1.

4. Discussion

We adapted a computational algorithm to define the minimum possible number of somatic genetic variants that allow identifying the largest number of endometrial cancer patients. In particular, we used an algorithm that provides the estimated sensitivity at a given number of somatic mutations, selecting those that are able to detect the highest number of endometrial cancer cases based on the TCGA dataset. In the present study, by evaluating the top 50 most representative point mutations, up to 81.9% sensitivity can be achieved, while 92.9% sensitivity could be obtained by interrogating the exonic regions of five genes. The landscape of mutations observed in the signatures of the present study was consistent with that observed in the published literature using the TCGA [6] and other genomic datasets [8–11]. In brief, *PTEN* mutations are common in endometrioid adenocarcinomas and loss of *PTEN* expression is suggested to be one of the earliest events in endometrial carcinogenesis [6]. Mutations in *PIK3CA* and *KRAS* are also frequent in endometrioid cancers, while mutations in *TP53* are usually present in non-endometrioid subtypes [6,12].

Some recent studies have shown feasibility in discriminating between endometrial cancer and benign conditions in Pap brush samples and other minimally invasive sampling methods using genomics [3,13] and epigenomics [4,14–17]. These molecular tests may help refining current diagnostic algorithms among women with postmenopausal bleeding by improving performance and failure rate of aspirate-based diagnosis. Importantly, they are better tolerated by women and therefore could be more appropriate in screening settings among asymptomatic populations. Given their high underlying risk, women with Lynch syndrome or obese women would probably benefit the most from new screening approaches. Unfortunately, the costs of these technologies are a shortcoming. Cost-effectiveness analyses can help evaluating the best preventive approaches and implementing different screening and early

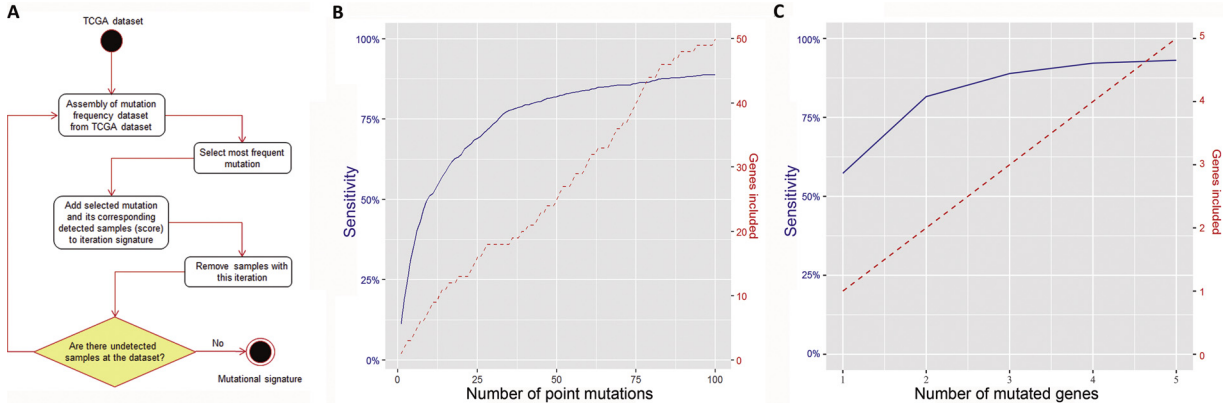

Fig. 1. Bioinformatics approach for signature generation. (A) Activity diagram of the greedy algorithm used for mining the TCGA dataset and generation of the signature. (B) Sensitivity as new mutations (point mutations) are incorporated by the algorithm. (C) Sensitivity as new mutations (full genes) are incorporated by the algorithm.

Table 1
Minimum set of somatic mutations to identify endometrial cancer in the TCGA dataset.

Minimum set of mutations (signatures)	Genes involved (gene symbol)	Sensitivity			
		Overall	Endometrioid histology	Serous histology	Mixed histology
Point mutations					
Signature 1 = 5 point mutations	*KRAS,PTEN,PIK3CA,PPP2R1A*	35.4%	39.4%	23.5%	22.7%
Signature 2 = 10 point mutations	Signature 1 + *RPL22,SETD1B,CTNNB1,RNF43*	51.1%	59.1%	24.3%	40.9%
Signature 3 = 15 point mutations	Signature 2 + *JAK1,FGFR2,BCOR*	58.6%	67.9%	27.8%	45.5%
Signature 4 = 20 point mutations	Signature 3 + *TP53,DOCK3*	64.1%	71.5%	40.0%	50.0%
Signature 5 = 25 point mutations	Signature 4 + *SPOP,ARID1A,POLE*	69.0%	74.7%	52.2%	50.0%
Signature 6 = 30 point mutations	Signature 5 + *FBXW7,SOX17*	73.4%	77.4%	61.7%	59.1%
Signature 7 = 35 point mutations	Same genes than Signature 6	77.7%	80.3%	70.4%	68.2%
Signature 8 = 40 point mutations	Signature 7 + *ACVR2A,KCNG4*	79.4%	81.8%	73.0%	68.2%
Signature 9 = 45 point mutations	Signature 8 + *CTCF,ZBTB20*	80.5%	82.7%	74.8%	68.2%
Signature 10 = 50 point mutations	Signature 9 + *CHD4,SOS1, SMAD7*	81.9%	84.4%	75.7%	68.2%
Genes					
Signature A = 1 mutated gene	*PTEN*	57.3%	72.3%	40.9%	7.0%
Signature B = 2 mutated genes	Signature A + *TP53*	81.6%	80.5%	77.3%	86.1%
Signature C = 3 mutated genes	Signature B + *PIK3CA*	88.9%	88.6%	90.9%	89.6%
Signature D = 4 mutated genes	Signature C + *ARID1A*	92.2%	92.9%	90.9%	89.6%
Signature E = 5 mutated genes	Signature D + *CTNNB1*	92.9%	93.9%	90.9%	89.6%

For simplicity, signatures are shown in incremental groups of five point mutations. We arbitrarily show signatures up to fifty point mutations and five genes (trends in sensitivity by increasing number of mutations can be observed in Fig. 1). Different point mutations in the same gene can occur across signatures, although genes are only represented once (for instance, mutations in *PTEN* are found in signature 1, but also in signature 4).

detection strategies. We adapted the algorithm in order to evaluate whether assessing few selected mutations could still result in an adequate diagnostic performance, and therefore to support the development of new potential technologies for endometrial cancer prevention. Ideally, targeting less than ten variants could result in a relatively cheap early detection technology. However, our results indicate that a maximum sensitivity of 51% would be obtained by interrogating this reduced number of point mutations. This sensitivity is insufficient in an early detection cancer setting, especially considering that these biomarkers are expected to be in lower concentrations in the minimally invasive sampling methods that could be used in preventive strategies [5]. Targeting 50 point mutations substantially increased the estimated sensitivity, although at an expected higher cost and possibly requiring a technology able to sequence larger DNA regions, similarly to targeted exome sequencing, or with a high multiplexing capacity.

Similar approaches have been used to select cancer biomarkers driving breast cancer heterogeneity [18], or to analyze microarray datasets with information on transcriptome [19,20], among others. We did not provide a formal statistics to evaluate the threshold in which sensitivity significantly stops increasing with incrementing number of mutations, since our aim was to describe the expected sensitivity at different sets of mutations. The maximum number of mutations to be tested will actually be defined by the cost and capacity of the resultant technology. Given that this is an *in silico* approach based on data from tumour samples, the selection of somatic variants could be further refined considering the biological pathways of the disease and using data from the evaluation of the biospecimen of interest. Given that endometrial hyperplasias are recognized precursors of endometrial cancer, applying the algorithm using a series of these cases will be also relevant. Importantly, the TCGA dataset may not be representative of all endometrial cases, although sequencing data from consecutive cases is not available for replication of the signatures. For comparison at the gene level, McConechy et al observed that 90.1% of 393 endometrial cancer cases showed somatic mutations in nine genes (*ARID1A, PTEN, PIK3CA, KRAS, CTNNB1, PPP2R1A, BRAF, TP53,* and *PPP2R5C*), which is slightly lower to our results [11]. In summary, we provide a simple method to select a reduced number of somatic genetic variants from a large genomic dataset that can be used in other cancer types or scenarios. This may be of help assisting researchers selecting specific variants for further study or to develop novel technologies in personalized early detection strategies.

Authors statement

All authors contributed to the planning of the study. LC wrote the manuscript and constructed Table 1 and Supplemental Table 1. LP performed the bioinformatic analysis, constructed Figs. 1 and S1. YB critically reviewed the bioinformatic code. All authors critically reviewed the manuscript and approved of the final draft, tables and figures.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgements

We thank Miquel Ángel Pujana for his valuable contributions. This work was partially supported by the public grants from Spanish Ministry of Economy and Competitiveness - Carlos III Institute of Health (PIE16/00049, PI17/01179, PI19/01835), Catalan Government (AGAUR-2017SGR1085, PERIS-SLT006/17/76), and European Regional Development Fund-ERDF.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.canep.2019.06.003.

References

[1] M.J. Khoury, Precision Public Health and Precision Medicine: Two Peas in a Pod | | Blogs | CDC, (2015) (Accessed 3 October 2018), https://blogs.cdc.gov/genomics/2015/03/02/precision-public/.

[2] M.J. Khoury, M.F. Iademarco, W.T. Riley, Precision public health for the era of precision medicine, Am. J. Prev. Med. 50 (2016) 398–401, https://doi.org/10.1016/j.amepre.2015.08.031.

[3] Y. Wang, L. Li, C. Douville, J.D. Cohen, T.-T. Yen, I. Kinde, K. Sundfelt, S.K. Kjær, R.H. Hruban, I.-M. Shih, T.-L. Wang, R.J. Kurman, S. Springer, J. Ptak, M. Popoli, J. Schaefer, N. Silliman, L. Dobbyn, E.J. Tanner, A. Angarita, M. Lycke, K. Jochumsen, B. Afsari, L. Danilova, D.A. Levine, K. Jardon, X. Zeng, J. Arseneau, L. Fu, L.A. Diaz, R. Karchin, C. Tomasetti, K.W. Kinzler, B. Vogelstein, A.N. Fader, L. Gilbert, N. Papadopoulos, Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers, Sci. Transl. Med. 10 (2018) eaap8793, https://doi.org/10.1126/scitranslmed.aap8793.

[4] J.N. Bakkum-Gamez, N. Wentzensen, M.J. Maurer, K.M. Hawthorne, J.S. Voss, T.N. Kroneman, A.O. Famuyide, A.C. Clayton, K.C. Halling, S.E. Kerr, W.A. Cliby, S.C. Dowdy, B.R. Kipp, A. Mariani, A.L. Oberg, K.C. Podratz, V. Shridhar, M.E. Sherman, Detection of endometrial cancer via molecular analysis of DNA collected with vaginal tampons, Gynecol. Oncol. 137 (2015) 14–22, https://doi.org/10.1016/j.ygyno.2015.01.552.

[5] J.C.M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J.D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: towards implementation of circulating tumour DNA, Nat. Rev. Cancer 17 (2017) 223–238, https://doi.org/10.1038/nrc.2017.7.

[6] Cancer Genome Atlas Research Network, C. Kandoth, N. Schultz, A.D. Cherniack, R. Akbani, Y. Liu, H. Shen, A.G. Robertson, I. Pashtan, R. Shen, C.C. Benz, C. Yau, P.W. Laird, L. Ding, W. Zhang, G.B. Mills, R. Kucherlapati, E.R. Mardis, D.A. Levine, Integrated genomic characterization of endometrial carcinoma, Nature 497 (2013) 67–73, https://doi.org/10.1038/nature12113.

[7] I.S. Kohane, M. Hsing, S.W. Kong, Taxonomizing, sizing, and overcoming the incidentalome, Genet. Med. 14 (2012) 399–404, https://doi.org/10.1038/gim.2011.68.

[8] H.B. Salvesen, S.L. Carter, M. Mannelqvist, A. Dutt, G. Getz, I.M. Stefansson, M.B. Raeder, M.L. Sos, I.B. Engelsen, J. Trovik, E. Wik, H. Greulich, T.H. Bø, I. Jonassen, R.K. Thomas, T. Zander, L.A. Garraway, A.M. Oyan, W.R. Sellers, K.H. Kalland, M. Meyerson, L.A. Akslen, R. Beroukhim, Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 4834–4839, https://doi.org/10.1073/pnas.0806514106.

[9] A. Berg, E.A. Hoivik, S. Mjøs, F. Holst, H.M.J. Werner, I.L. Tangen, A. Taylor-Weiner, W.J. Gibson, K. Kusonmano, E. Wik, J. Trovik, M.K. Halle, A.M. Øyan, K.-H. Kalland, A.D. Cherniack, R. Beroukhim, I. Stefansson, G.B. Mills, C. Krakstad, H.B. Salvesen, Molecular profiling of endometrial carcinoma precursor, primary and metastatic lesions suggests different targets for treatment in obese compared to non-obese patients, Oncotarget 6 (2015) 1327–1339, https://doi.org/10.18632/oncotarget.2675.

[10] W.J. Gibson, E.A. Hoivik, M.K. Halle, A. Taylor-Weiner, A.D. Cherniack, A. Berg, F. Holst, T.I. Zack, H.M.J. Werner, K.M. Staby, M. Rosenberg, I.M. Stefansson, K. Kusonmano, A. Chevalier, K.K. Mauland, J. Trovik, C. Krakstad, M. Giannakis, E. Hodis, K. Woie, L. Bjorge, O.K. Vintermyr, J.A. Wala, M.S. Lawrence, G. Getz, S.L. Carter, R. Beroukhim, H.B. Salvesen, The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis, Nat. Genet. 48 (2016) 848–855, https://doi.org/10.1038/ng.3602.

[11] M.K. McConechy, J. Ding, M.C. Cheang, K. Wiegand, J. Senz, A. Tone, W. Yang, L. Prentice, K. Tse, T. Zeng, H. McDonald, A.P. Schmidt, D.G. Mutch, J.N. McAlpine, M. Hirst, S.P. Shah, C.-H. Lee, P.J. Goodfellow, C.B. Gilks, D.G. Huntsman, Use of mutation profiles to refine the classification of endometrial carcinomas, J. Pathol. 228 (2012) 20–30, https://doi.org/10.1002/path.4056.

[12] E. Kuhn, R.-C. Wu, B. Guan, G. Wu, J. Zhang, Y. Wang, L. Song, X. Yuan, L. Wei, R.B.S. Roden, K.-T. Kuo, K. Nakayama, B. Clarke, P. Shaw, N. Olvera, R.J. Kurman, D.A. Levine, T.-L. Wang, I.-M. Shih, Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses, J. Natl. Cancer Inst. 104 (2012) 1503–1513, https://doi.org/10.1093/jnci/djs345.

[13] I. Kinde, C. Bettegowda, Y. Wang, J. Wu, N. Agrawal, I.-M. Shih, R. Kurman, F. Dao, D.A. Levine, R. Giuntoli, R. Roden, J.R. Eshleman, J.P. Carvalho, S.K.N. Marie, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, L.A. Diaz, Evaluation of DNA from the papanicolaou test to detect ovarian and endometrial cancers, Sci. Transl. Med. 5 (2013) 167ra4, https://doi.org/10.1126/scitranslmed.3004952.

[14] R.-L. Huang, P.-H. Su, Y.-P. Liao, T.-I. Wu, Y.-T. Hsu, W.-Y. Lin, H.-C. Wang, Y.-C. Weng, Y.-C. Ou, T.H.-M. Huang, H.-C. Lai, Integrated epigenomics analysis reveals a DNA methylation panel for endometrial cancer detection using cervical scrapings, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 23 (2017) 263–272, https://doi.org/10.1158/1078-0432.CCR-16-0863.

[15] C.C. Chang, H.C. Wang, Y.P. Liao, Y.C. Chen, Y.C. Weng, M.H. Yu, H.C. Lai, The feasibility of detecting endometrial and ovarian cancer using DNA methylation biomarkers in cervical scrapings, J. Gynecol. Oncol. 29 (2018) e17, https://doi.org/10.3802/jgo.2018.29.e17.

[16] L.M.A. De Strooper, M. van Zummeren, R.D.M. Steenbergen, M.C.G. Bleeker, A.T. Hesselink, G.B.A. Wisman, P.J.F. Snijders, D.A.M. Heideman, C.J.L.M. Meijer, CADM1, MAL and miR124-2 methylation analysis in cervical scrapes to detect cervical and endometrial cancer, J. Clin. Pathol. 67 (2014) 1067–1071, https://doi.org/10.1136/jclinpath-2014-202616.

[17] G.-E. Kim, S.-S. Kweon, J.S. Lee, J.H. Lee, J.H. Nam, C. Choi, Quantitative assessment of DNA methylation for the detection of cervical and endometrial adenocarcinomas in liquid-based cytology specimens, Anal. Quant. Cytopathol. Histopathol. 34 (2012) 195–203.

[18] X. Dai, T. Hua, T. Hong, Integrated diagnostic network construction reveals a 4-gene panel and 5 cancer hallmarks driving breast cancer heterogeneity, Sci. Rep. 7 (2017) 6827, https://doi.org/10.1038/s41598-017-07189-6.

[19] V.V. Galatenko, A.E. Lebedev, I.N. Nechaev, M.Y. Shkurnikov, E.A. Tonevitskii, V.E. Podol'skii, On the construction of medical test systems using greedy algorithm and support vector machine, Bull. Exp. Biol. Med. 156 (2014) 706–709, https://doi.org/10.1007/s10517-014-2430-3.

[20] M. Song, S. Rajasekaran, A greedy algorithm for gene selection based on SVM and correlation, Int. J. Bioinf. Res. Appl. 6 (2010) 296–307, https://doi.org/10.1504/IJBRA.2010.034077.