Translational Oncology 13 (2020) 100837

Contents lists available at ScienceDirect

Translational Oncology

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

Original research

Liquid biopsy by NGS: Differential presence of exons (DPE) is related to metastatic potential in a colon-cancer model in the rat

Dolores C. García-Olmo [a,b,1], Ramón Peiró-Pastor [c,1], María G. Picazo [b], Susana Olmedillas-López [d], Mariano García-Arranz [d], Begoña Aguado [c], Damián García-Olmo [e,*]

[a] Centre de Recerca Experimental Biomédica Aplicada (CREBA), IRBLleida, Lleida, Spain
[b] Experimental Research Unit, University Hospital of Albacete, Spain
[c] Genomics and NGS Service, Centro de Biología Molecular Severo Ochoa (CBMSO), CSIC-UAM, Madrid, Spain
[d] New Therapies Laboratory, Health Research Institute-Fundación Jiménez Díaz University Hospital (IIS-FJD), Madrid, Spain
[e] Department of Surgery, Fundación Jiménez Díaz University Hospital, Madrid, Spain

ARTICLE INFO

Article history:
Received 9 May 2020
Received in revised form 9 July 2020
Accepted 16 July 2020
Available online xxxx

ABSTRACT

Differential presence of exons (DPE) is a method of interpretation of exome sequencing, which has been proposed to design a predictive algorithm with clinical value in patients with colorectal cancer (CRC). The goal of the present study was to examine the reproducibility in a rat model of metastatic colon cancer. DHD/K12-TRb cells were injected in syngenic immunocompetent BD-IX rats. Cells were from two stocks with low and normal metastatic potential, and injected into two separate groups of rats. Five to ten weeks after injection, blood samples were taken prior euthanasia and whole exome sequencing performed. Through DPE analysis, we identified a set of exons whose differential presence in plasma allowed us to compare both groups of tumor-bearing animals. A verification test was performed to confirm that the algorithm was able to classify extracted samples into their corresponding groups of origin. The highest mean probability was 0.8954. In conclusion, the DPE analysis in tumor-bearing animals was able to discriminate between different disease status, which fully supports previous results in CRC patients.

Introduction

The use of liquid biopsy in the management of cancer disease has been one of the most encouraging expectative in the last two decades. It is very probable that analysis on liquid biopsy will not replace tumor biopsies, however, they can provide additional insight into important clinically relevant information, and there is a real need for clinically relevant and reproducible targets. Notwithstanding tumor cells have been the focus of a high number of studies on liquid biopsy, the genetic profiling of cell-free DNA (cfDNA) in the blood is the main target of liquid biopsy [1].

Levels of cfDNA in plasma have been repeatedly examined in cancer patients [2] and tumor-bearing animals [3] and, in spite of heterogeneous results among studies, most of them have shown a link between cancer progression and releasing of cfDNA into the blood. Discrepancies among results from different studies are mainly regarding the correlation of such DNA with tumor stage, location, size or response to therapy. The kind of nucleic acid and the specific sequence used as the detection target, as well as the methods applied, are probably the main reasons for such disparity.

In the last decade, the use of next-generation sequencing (NGS) and sophisticated bioinformatics analysis, have opened new possibilities to increase the potential utility of cfDNA as clinical tools for diagnosis, prognosis and management of cancer patients, pointing to a precise and personalized medicine. Whole genome sequencing (WGS) or exome sequencing (WES) of tumor cells and plasma cfDNA using NGS is, theoretically, an unbiased approach that might provide extensive and comprehensive genomic information about cancer, through the detection of mutations (SNPs) and some structural alterations, such as indels or copy number aberrations [4].

To date, the most explored approaches using NGS have been the detection of cancer-specific mutations, and the discovering of them without prior knowledge of the tumor mutation landscape [5]. However, detection of cancer-specific mutations by NGS has not yet achieved the high sensitivity obtained by qPCR and ddPCR [3,6] which, in addition, have a substantially lower cost. The scope for error in NGS is problematic in rare mutation detection, due to the relatively high sequencing error rate or PCR errors introduced during library construction [1]. It is important to keep in mind that the sensitivity is highly relevant in the context of liquid biopsies, in which

* Corresponding author at: Department of Surgery, New Therapies Laboratory, Foundation Health Research Institute-Fundación Jiménez Díaz University Hospital (FIIS-FJD), Avda. Reyes Católicos, 2, 28040 Madrid, Spain.
E-mail address: damian.garcia@uam.es. (D. García-Olmo).
[1] Dolores C. García-Olmo and Ramón Peiró-Pastor have equally contributed to this work.

http://dx.doi.org/10.1016/j.tranon.2020.100837

there is a tiny fraction of mutated nucleic acids mixed with a majority of non-mutated molecules.

Numerous factors can impact the accuracy, sensitivity and specificity of cfDNA sequencing by NGS: the sample preparation, the target enrichment library construction, the sequencing platform, and the bioinformatics tools. In fact, a side-by-side comparison study showed how different may be the results of NGS of cfDNA obtained from the same patients in different laboratories, as reviewed by Posey et al. [7]. In summary, despite the potential utilities of NGS in cfDNA, its inherent limitations should not be overlooked.

Currently, many groups are designing approaches to circumvent such limitations, and get the maximum value of NGS to detect cancer-specific mutations, as reviewed by Kastrisiou et al. [8]. However, it is probably necessary to open the mind to other type of approaches for getting the maximum value of the broad information resulting from NGS techniques.

Our group has launched a method of interpretation of exome sequencing, based on a general overview of cfDNA in plasma, which allows to perform it at a shallower depth and, more interestingly, allowed us to design a predictive algorithm with clinical value in patients with colorectal cancer [9]. This method was based on the discovery that patients with and without metastasis have a different content of exons in their plasmas, which led to a new concept named as “differential presence of exons” (DPE) [9].

The goal of the present study was to examine the reproducibility and the utility of DPE approach in an animal model of metastatic colorectal cancer, with syngenic immunocompetent rats. By this way to validate the DPE proposal, we could avoid the heterogeneity and the possible bias related to any study of a human group.

Materials and methods

Cell culture

DHD/K12-PROb cells were used (rat colon cancer cells; also called DHD/K12-TRb cells; European Collection of Cell Cultures - ECACC no. 90062901, referred to as DHD cells herein) as malignant cells, which were purchased to the ECACC (Salisbury, UK). Cells were cultured in monolayer in a mixture of DMEM and Ham's F10 (1:1, v/v; Gibco-BRL, Life Technologies Ltd., Paisley, Scotland), supplemented with 10% fetal bovine serum (Gibco-BRL) and gentamycin (0.005%; Gibco-BRL). Cells were passaged after dispersion in 0.125% trypsin in EDTA.

We used two cell populations: the first (named herein as “Old DHD”) was continuously cultured in a steadily manner (two passes per week) for more than 5 years. In prior own assays, we detected that when those old DHD cells were injected in rats, metastases were slower developed than the usual and previously standardized way [3]. The last cell population (named as “New DHD”) was bought and subjected to no more than 5 passes before being used for this study.

Animals

Both male and female BD-IX rats were used, from 6 to 7 weeks of age. They were taken from a colony established at the authors' animal facility from founders purchased from a commercial breeder (Charles River, Barcelona, Spain). The entire study was approved by the Ethics Committee for Animal Research of Castilla-La Mancha University (Spain; code number: 2013-0705). All animals were housed and handled according to the approved protocols, under supervision by the staff of the animal facility of the University General Hospital of Albacete. The studies were conducted in accordance with European and Spanish laws (Directive 2010/63/UE and Real Decreto 53/2013, respectively). As recommended by the Federation of European Laboratory Animal Science Associations (FELASA), mice in the animal facility were tested periodically to ensure that the colony remained free of pathogens. From birth to the end of the experiments, all rats had unlimited access to water and standard rat chow (Teklad Lab Animal Diets; Harlan Laboratories, Barcelona, Spain).

Implantation of tumors and design of experiments

Tumors were generated in the thoracic region by unilateral subcutaneous injection of DHD cells into the right side of the chest. One day before implantation, cells were trypsinized and seeded in new flasks at a medium confluence (30–40%), so that at the next day, cells were in the exponential-growth phase at implantation. Immediately before cell injection in animals, cells were trypsinized, washed, and resuspended in phosphate-buffered saline (PBS). Then, 0.25 ml of this suspension, containing 1×10^6 cells, was injected per rat.

All DHD cells were manipulated and injected in animals in a similar way, and two animal groups were established, according to the kind of injected cells: Group O (n = 7), injected with “Old-DHD cells” and euthanized when an endpoint criterium was reached; and Group N (n = 7), injected with “New-DHD cells” (Fig. 1). This last group was divided into two subgroups, according to the moment of the euthanasia: Subgroup N1 (n = 4): euthanized when an endpoint criterium was reached; Subgroup N2 (n = 3): euthanized earlier, at 5–6 weeks after cells inoculation. Each group was composed of male and female animals in similar proportions.

Animals were inspected daily by the veterinarian of the animal facility. The body weight and the growth of subcutaneous tumors were monitored and recorded weekly. The greatest diameter of each tumor was measured with electronic calipers. Our previous experience with this model demonstrated that higher-frequency analysis is unnecessary [3]. The protocol for handling of tumor-bearing animals was designed according to recommendations for ethical cancer research [10].

Axillary lymph nodes and lungs were inspected visually for the presence or absence of metastases, which was recorded.

Animals were euthanized when reached at least one of the following endpoint criteria: (i) greatest diameter of primary tumor higher than 4 cm; (ii) ulceration of primary tumor; or (iii) weight loss higher than 10%.

Collection of blood samples

At the moment of the euthanasia, rats were anesthetized with an intraperitoneal injection of a mixture of ketamine (75 mg/kg) and xylazin (10 mg/kg). Then blood (approximately 4.0 to 4.5 ml per rat) was withdrawn by cardiac puncture and a lethal dose of sodium thiopental was administered intracardially immediately afterwards. All blood samples were collected in tubes with EDTA and subjected to centrifugation (1800 $\times g$, 10 min, 4 °C). Plasma was collected and subjected to a second centrifugation at 3000 $\times g$ for 10 min. The plasma samples were aliquoted and stored at −80 °C prior to extraction of DNA.

Extraction of cfDNA from plasma samples

DNA was extracted from 1-ml samples of plasma with a commercial kit (QIAamp Circulating Nucleic Acid Kit; QIAGEN GmbH, Hilden, Germany) according to the instructions from the manufacturer. DNA samples were then stored at −20 °C.

Library preparation, exome-capture, and sequencing

Concentration, quantity, and integrity of cfDNA were estimated prior to use. The size distribution of cfDNA fragments was determined using a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Library preparation and specific exome capture were performed using a custom EZ-Developer capture system prepared according to the coordinates of the complete rat exome (Roche NimbleGen, Basel, Switzerland). Libraries were hybrid-captured using biotinylated probes. Adapter DNA sequences were placed on both ends, and exomes were sequenced on the Illumina NextSeq500 platform (Illumina, Inc., San Diego, CA, USA) with 75 bp paired-end reads. Library preps and sequencing were performed at the Genomic Facility of the Scientific Park of Madrid, Spain.

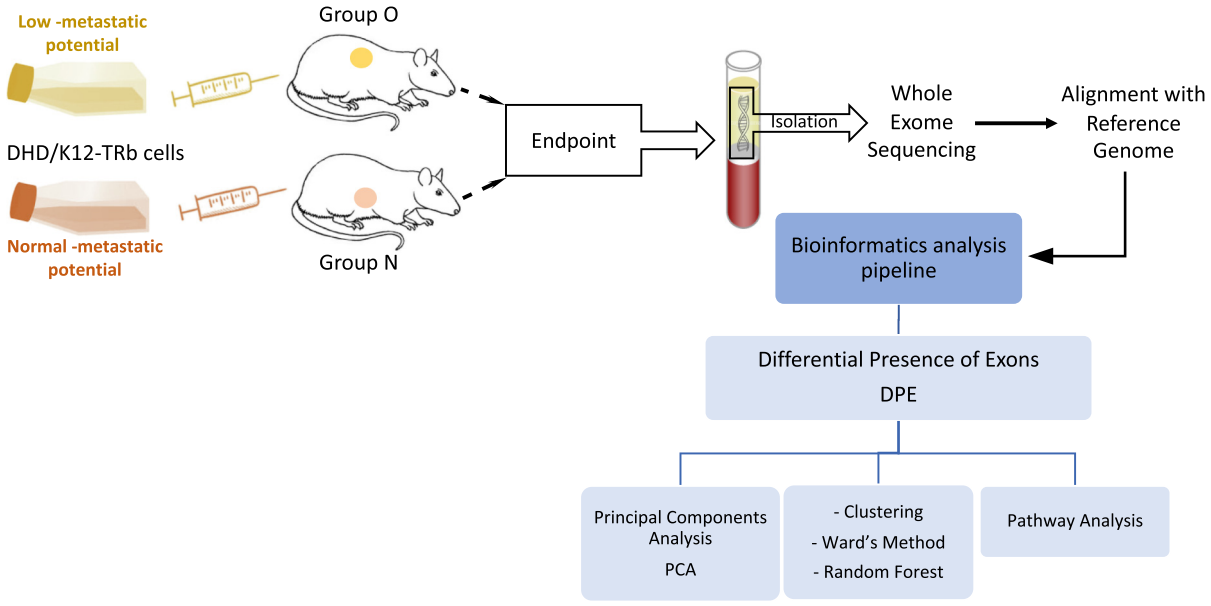

Fig. 1. Schematic representation of the experimental design. Two sublines of DHD cells were injected in BD-IX rats, which were classified into two groups: group O, injected with the named "old cells", with low-metastatic potential; and group N, injected with the named "new cells", with normal-metastatic potential. When animals met the endpoint criteria, blood samples were taken and plasma cfDNA samples were subjected to whole exome sequencing. Then, bioinformatics analyses were performed.

Data analysis

Fourteen samples were sequenced (7 from each group of tumor-bearing animals), obtaining a total of around 543 M of 2 × 76 nt reads. The read depth varied in a range from 59× to 104×, with an average read depth of 82× per sample. Reads quality analyses were performed using the FastQC software (Brabaham Bioinformatics) [11]. Reads were aligned against the *Rattus norvegicus* genome (6.0.88) using Bowtie aligner [12] allowing up to 3 mismatches (parameters: -v 3 -k 1 –best).

Detection of differentially present exons (DPE)

Sequencing data were processed as in a standard RNA-seq analysis, but trying to highlight DPE rather than differentially expressed genes. This analysis was done with 'EdgeR' R package [13]. Reads counts by exon were calculated with HTSeq-count [14]. Edger's background statistical methods were based on generalized linear models (glm), which test for DPE using either likelihood ratio tests (LRT) [13] or quasi-likelihood F-tests (QLF) [15]. The DPE for a threshold p-value of $p \leq 0.005$ was selected, and MA plots for such DPE were drawn.

DPE clustering and principal components analysis (PCA)

Normalized DPE counts (counts per million, CPM) were calculated for every exon for each sample using EdgeR. These counts were used to cluster the samples using Ward's method [16], to verify that all samples were behaving properly. Additionally, principal components analysis (PCA) was performed with an in-house R script.

Random Forest (RF) classification

Random Forest (RF) classification was implemented with an in-house R script using the 'randomForest' package [17]. Two samples from both O and N groups were randomly selected and extracted, using the five remaining samples of each group (10 samples in total) as a 'training set' to generate a predictive algorithm. One hundred classifications were performed by iteration of this process, calculating the mean value of the obtained probabilities. The accuracy of the resulting model was tested by checking its ability to correctly classify randomly extracted samples into their corresponding groups of origin.

Pathway analysis

Data were analyzed with Ingenuity Pathway Analysis (IPA®, Qiagen, http://www.qiagenbioinformatics.com; [18]), a software application for the analysis and interpretation of data derived from omics experiments. The list of over-represented genes for each group were imported into IPA and mapped to the IPA Knowledge Database. We performed the Core Analysis for predicting pathways and molecular functions affected, based on gene expression. The significance of the association between the dataset and the specific pathways was determined by Right Tailed Fisher's Exact Test ($p < 0.05$). A Wilcoxon–Mann–Whitney test was performed for paired samples to analyze whether the percentage of genes from both groups were differentially distributed in the main function categories.

Statistical analysis

For variables related to the clinical course of tumors in the animals, comparisons of means were made by Student's *t*-test for normal distributions and by the Kruskal-Wallis tests for non-normal distributions. A p-value of less than 0.05 was accepted as evidence of statistical significance. Qualitative variables were compared using the $\chi 2$ test and Fisher's exact test when any frequency was less than 5. Two-tailed p-values of less than 0.05 were considered evidence of statistical significance. Such statistical analyses were performed with StatPlus, version 2009 (AnalystSoft) and OpenEpi (Open Source Epidemiologic Statistics for Public Health, Version 2.3.1, www.OpenEpi.com).

Right Tailed Fisher's Exact Test was used for IPA with a cutoff p-value of $p < 0.05$. A Wilcoxon–Mann–Whitney test for paired samples was also performed for gene distribution in IPA main function categories. LRT and QLF tests were performed for DPE with a p-value of $p \leq 0.005$.

Data availability

Whole-exome sequencing data that supports the findings of this study has been deposited in the European Nucleotide Archive (ENA) with the accession number PRJEB37894.

The writing of this article based on an animal study has followed the ARRIVE Guidelines [19].

Results

Body weight and growth of tumors

In the two groups, the mean of body weight was increasing along the experiment, except in the last two weeks tested in the Group N (Fig. 2A). The initial body weight ranged from 112 g to 154 g, and the final weight (at euthanasia), from 205 to 395 g. There was no statistically significant difference between the two groups. As expected, the mean body weight of male animals (n = 7) was higher than that of female ones (n = 7; Fig. 2B), with statistically significant differences, at all weeks except for the moment of cells injection.

Tumors were detectable in rats from the first week after inoculation of DHD cells and grew at an apparently uniform rate for the duration of the experiment (Fig. 3A). The means of the tumor diameters were similar among groups at all weeks. However, there was statistically significant difference between male and female rats at all weeks, except for the first one (Fig. 3B).

Macroscopic observations

Lung macrometastases were not detected in animals in group O and in subgroup N2. However, such metastases were macroscopically visible in all rats in subgroup N1 There was no difference between male and female animals.

cfDNA isolation from plasma and NGS

Circulating cfDNA was successfully extracted from all plasma samples, obtaining disparate concentrations of DNA ranging from 0.02 to 97.35 ng/μL. There was not statistically significant difference between groups.

BioAnalyzer plots revealed a cfDNA size > 1 kb. All samples were subjected to a mechanical fragmentation process using Bioruptor® Sonication System (Diagenode, Liège, Belgium) to obtain fragments with a length of around 145 bp. Adapter DNA sequences were placed yielding a total length of 170 bp.

The total number of reads per sample ranged from 33 to 45 million. Quality analyses of reads using the FastQC software (Phred + 33 quality score) revealed that more than 85% of the bases had a >Q30 quality. The Supplementary Table 1 shows the total number of reads, depth of exome coverage, GC content, and percentage of aligned and unaligned reads of each sample.

DPE analysis

We identified a set of exons whose differential presence in plasma allowed us to compare group O vs group N1 and group N2. There was no difference between male and female animals. DPE was analyzed with

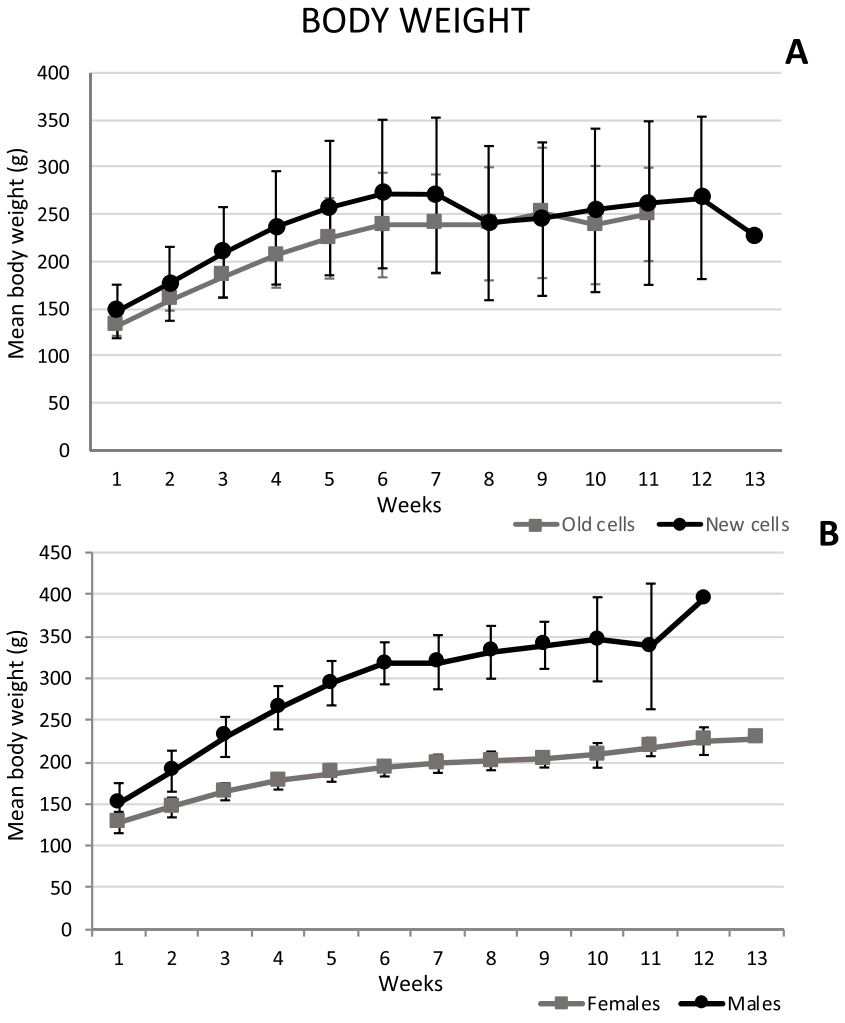

Fig. 2. Rats growth during the study. Values shown are mean body weight (±SD). A. Separate curves for animals injected with "Old" and "New" DHD cells (with low and normal metastatic potential, respectively). B. Separate curves for female and male animals.

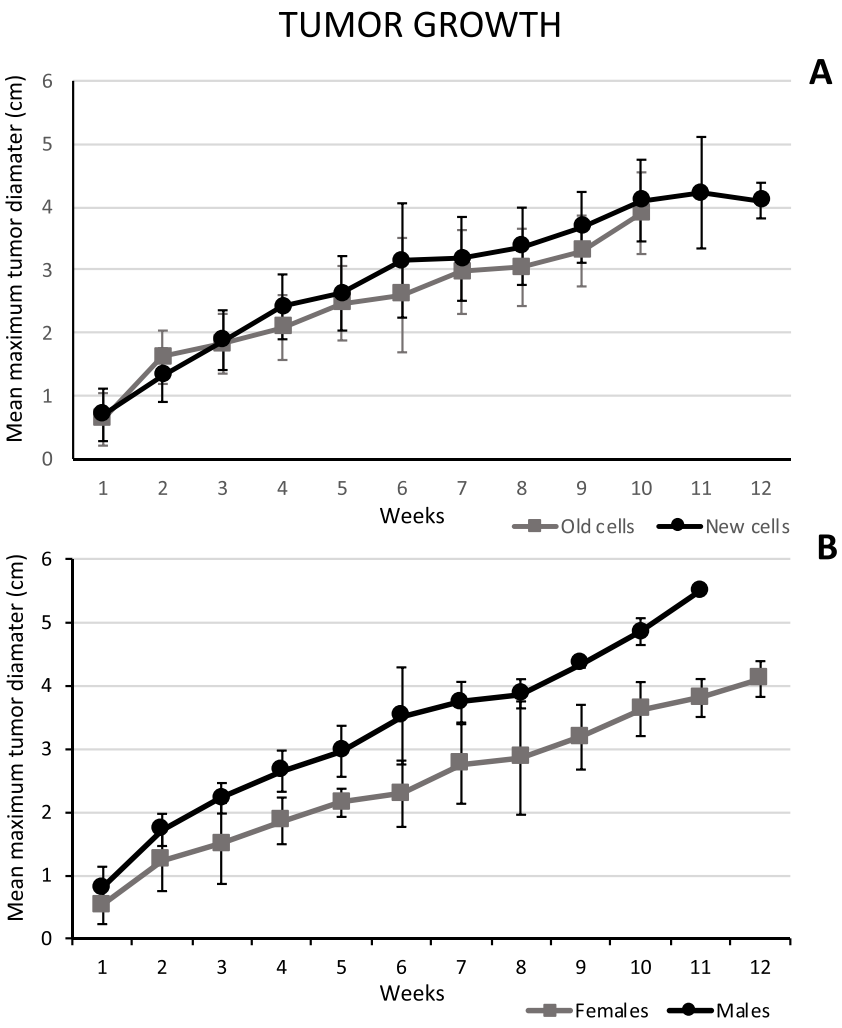

Fig. 3. Tumor growth in rats after subcutaneous injection of DHD cells. Values shown are mean maximum diameters of tumors (±SD). A. Separate curves for animals injected with "Old" and "New" DHD cells (with low and normal metastatic potential, respectively). B. Separate curves for female and male animals.

EdgeR, using either QLF or LRT methods, with a threshold of p-value ≤0.005 for each comparison. For O vs N1 comparison, a total of 102 and 244 exons were obtained respectively, yielding 258 exons overall, including unique and common exons. For O vs N2 comparison, a total of 331 and 783 exons were obtained, yielding 793 unique and common exons. A total amount of 1028 differentially present exons were found, considering both comparisons and methods.

MA plots for both comparisons and methods are showed in Fig. 4, in which the exons identified as DPEs in each comparison are in red. Those statistically significant exons were located on the margins of the point cloud, as expected.

Then, we examined in which genes were located those exons. Considering that some exons belong to the same gene, finally we identified a total of 937 genes, of which 447 were over-represented in group N1, and 481 in group O (for O vs N1 comparison), and 413 were over-represented in group N2, and 521 in group O (for O vs N2 comparison; Supplementary Tables 2 and 3).

DPE clustering

Clustering of normalized DPE was performed by Ward's method, which gives a distances tree shown in Fig. 5. Samples were properly clustered, maintaining the separation between the O, N1 and N2 samples. Next, PCAs were performed and Fig. 6 shows a bidimensional plot with the first two principal components. As can be seen, groups O, N1 and N2 are clearly separated and clustered properly.

These results encouraged us to develop a predictive algorithm to classify samples. To achieve this goal, a RF classification was obtained after 100 iterations (Fig. 7), extracting two randomly selected samples from each group (O and N), and generating a predictive model, with the 10 remaining samples (five per group) as a training set. A verification test was performed to confirm that the algorithm was able to classify extracted samples into their corresponding groups of origin, by calculating the average probabilities of belonging to one group or another (Supplementary Table 4). The extracted samples were correctly identified, with a mean value of $p = 0.7268$ with a standard deviation of $S = 0.1167$, and the highest mean probability was $p_{\_max} = 0.8954$.

We also made an alternative approach, extracting three samples from both O and N groups and using the four remaining samples of each group (8 samples in total) as a 'training set'. Such approach had not significantly improved the prediction; specifically, we obtained a value of $p = 0.7111$ with a standard deviation of $S = 0.1277$, and the highest mean probability was $p_{\_max} = 0.9006$ (Supplementary Fig. 1).

Pathway analysis

We performed a pathway analysis using IPA to gain further insight into the functional annotation of the 937 over-represented genes. Such analysis that showed that over-represented genes in groups O and N2 were related to three major biological functions: 'mitochondrial dysfunction', 'oxidative phosphorylation', and 'sirtuin signaling pathway', while genes that were over-represented in group N1 were related to major biological functions

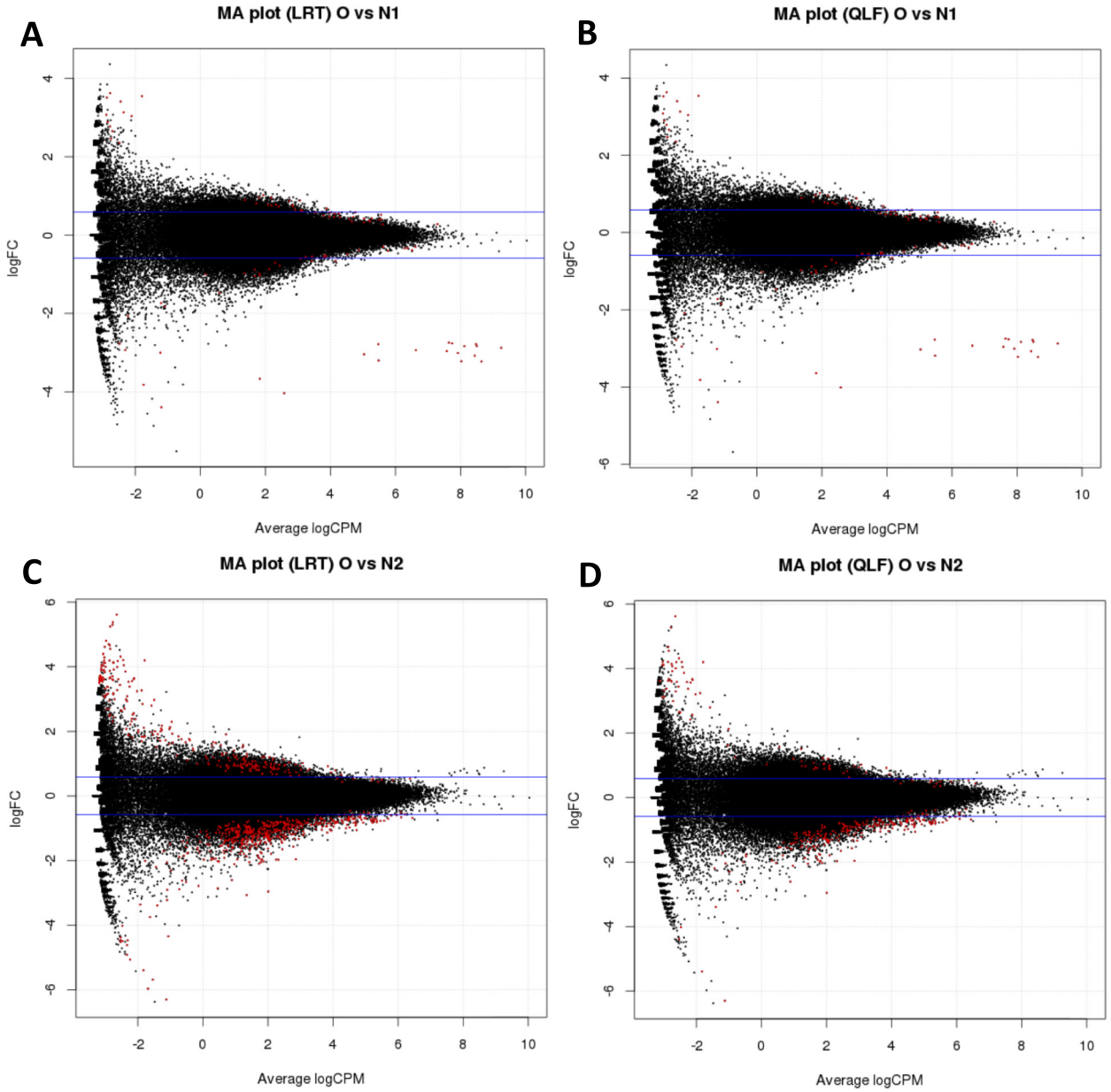

Fig. 4. MA plots for selected differentially present exons combining two different methods: Likelihood Ratio Tests (LRT) and Quasi-Likelihood F-tests (QLF) and two different comparatives (group O vs group N1, and group O vs group N2). The log ratio of fold-change (FC) is plotted on the y-axis, and the average of the normalized counts (counts per million; CPM) is plotted on the x-axis. Differentially-present exons are highlighted in red (DPEs; $p \leq 0.005$). A. LRT method for the comparison of group O vs group N1 (244 DPEs). B. QLF method for the comparison of group O vs group N1 (102 DPEs). C. LRT method for the comparison of group O vs group N2 (783 DPEs). D. QLF method for the comparison of group O vs group N2 (331 DPEs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that include 'molecular mechanisms of cancer' (Supplementary Tables 5 to 8, and Supplementary Figs. 2 to 5). Regarding networks, for O vs N1 comparison, over-represented genes in group N1 were associated to 12 networks whereas those in group O were related to other 25 different networks, and for O vs N2 comparison, over-represented genes in group N2 were associated to 13 networks whereas those in group O were related to 25 networks (Supplementary Tables 9 to 12).

Discussion

The use of NGS-based techniques in blood has opened a wide field for exploring new precise - even personalized- biomarkers, with potential clinical utility in cancer. In fact, in the last decade, there have been many efforts to explore this sophisticated tool, and to generate accurate and feasible techniques for the management of the cancer disease [1,8]. However, we should not overlook that information resulting from NGS might be valuable, not only for clinical use, but also to deepen in the high-complex mechanism of the cancer progression.

Using NGS-based techniques in tumors, Gerlinger et al. [20] got strong evidences of the intratumor heterogeneity (ITH), suggesting that such phenomenon might lead to underestimation of the tumor genomics landscape and might present major challenges to personalized-medicine and biomarker development. With respect to colorectal cancer (CRC), the idea of clonality - lengthily assumed after the establishment of the multi-step carcinogenesis model [21] - has recently been revised by a series of studies demonstrating IHT in advanced cancer, as well as in precancerous lesions [22,23]. Specifically, it has been suggested that ITH is an inherent characteristic of colorectal tumors that arises early and continuously increases during growth, and is not significantly constrained by clonal selection [22]. Such heterogeneity might affect the possibility to reliably identify

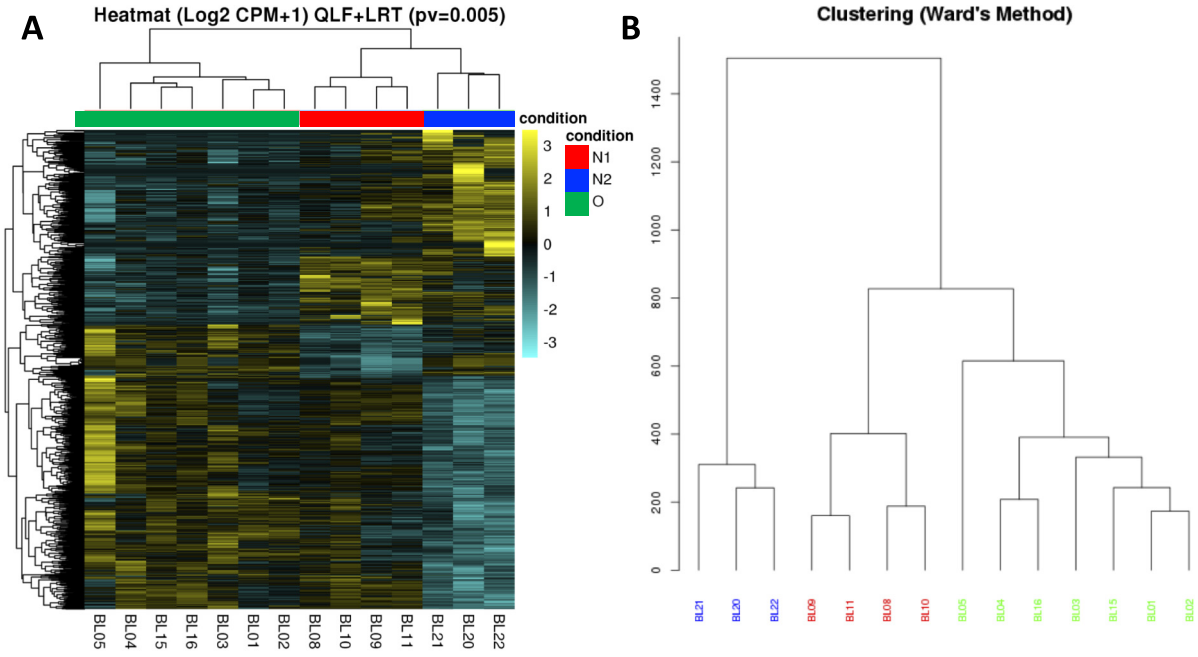

Fig. 5. Clustering of samples using normalized values of differentially present exons (DPEs). A. Heatmap representation of the presence levels of DPEs. Presence levels are shown on a log2(CPM + 1) scale to avoid possible null values, covering a range from 3 (yellow, highest level) to −3 (cyan, lowest level). Color bars at the top of the plot indicate experimental group of the samples. B. Clustering of samples by Ward's Method. Animals were individually named as BL-number and, in this figure, are marked by different colors according to the group in which were included: O: green, N1: red; and N2: blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

somatic mutations in cfDNA, as suggested in previous studies [24]. However, since plasma receives DNA from the various heterogeneous tumor clones in the body, sequencing of plasma DNA from a different point of view than mutational status, might be a readily available and noninvasive method for studying and monitoring tumoral heterogeneity and progression.

For this reason, the technologically-advanced approaches focused on the overview of cfDNA circulating in plasma might result in a deeper understanding of cfDNA biology -which is essential for the practical application of liquid biopsy-, and even in useful and reproducible tools for clinical management. In this sense, through WES techniques, we have previously shown that a series of CRC patients with and without metastasis have a different content of exons (DPE) in their plasmas [9]. Such DPE profiles led to a DPE algorithm that was capable of providing a predictive model, by which, metastatic and non-metastatic patients were correctly clustered and clearly separated [9].

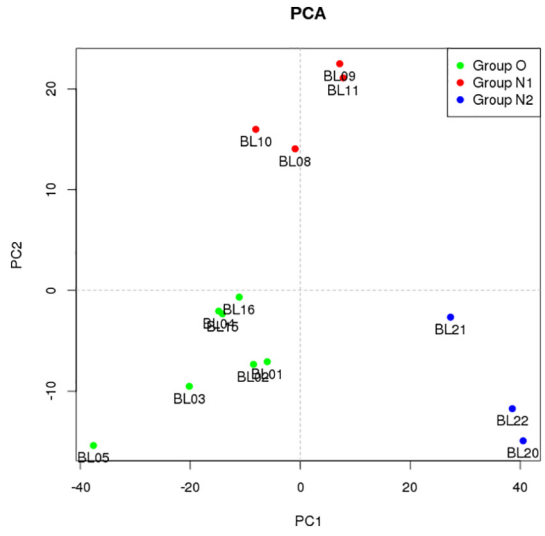

Fig. 6. Bidimensional principal components analysis (PCA) plot. Group O, N1 and N2 samples are marked by different colors: O: green, N1: red; and N2: blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the current study, we have fully corroborated the predictive value of the DPE algorithm in a rat model of CRC, in a very-similar way. We have used a well-known rat model of metastatic CRC, which have been repeatedly used by our group [3,25–28]. We have reported that lung metastases are macroscopically found between the 8th and the 11th week after injection of tumor cells, regardless the site of injection [27,28]. In the current study, the model had a special feature, because we used a cell subline which had spontaneously hampered the development of metastases when was injected in rats. We took advantage of this change to compare rats with the usual metastatic tumors (grouped herein in group N) and rats with the hampered disease (group O). Animals in group N that were euthanized when reached ethical endpoint criteria (subgroup N1), had lung macro metastases, whereas such metastases were not found in any animal in the group O.

By the DPE algorithm, those two groups were purely distinguished, in a similar way that previously reported in CRC patients [9], although the absolute values of the analyses were different, as expected in a different species. In the DPE analysis of plasma of CRC human patients, we found a lesser number of differentially-present exons than in rats (379 exons vs 1028, respectively). However, in the two species the algorithm was able to classify extracted samples into their corresponding groups of origin, by calculating the average probabilities of belonging to one of the groups. In tumor-bearing rats, the highest mean probability to correctly classify a sample was higher than in CRC patients (0.8954 vs 0.68).

In addition, we established a subgroup in group N (subgroup N2) in which rats were early euthanized when lung metastases were not macroscopically detected. Clustering of normalized DPE by Ward's method gave

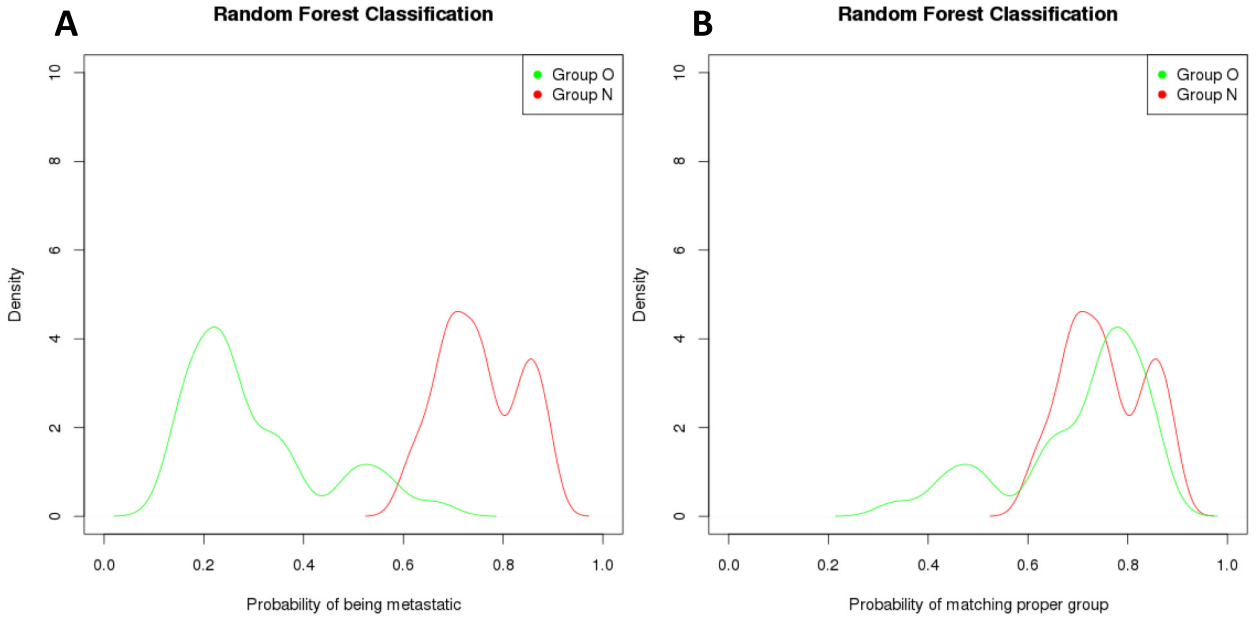

Fig. 7. Kernel Density Estimation (KDE) of the distribution of probabilities obtained from 100 iterations of a Random Forest (RF) classification. In each iteration, two randomly selected samples from each group (O and N, joining N1 and N2 groups), were extracted and classified with a predictive RF model using the 10 remaining samples (five per group) as a training set. A. KDE distribution of the probabilities of being metastatic. B. KDE distribution of the probabilities of being correctly assigned to the corresponding group of origin. Group O is marked in green and Group N in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a distances tree in which samples were properly clustered, maintaining the separation between the O, N2 and N1 samples. Probably, if N2 rats were euthanized later, with metastases, samples would have been clustered in the N1 area. These results suggest that the DPE profile is dynamic and change with the cancer progression in a similar way among animals.

The functional analysis of the over-represented exons in each group did not lead to any conclusion about the reason of such abundance and its connection of cancer progression, if any. Nonetheless, the fact is that a differential detection of exons was able to distinguish between tumor-bearing animals with different tumor status, which lead to hypothesize that changes in molecular behavior of tumors are more extensive and intensive than believed, and might be monitored by a DPE analysis.

In addition, the fact that certain exons were more represented than others, and that such profiles were repeated among animals in the same group, support the idea of an active and intelligent release of nucleic acids by cells. Historically, the most-persistent theory about the origin of cfDNA has been the cell death [29], however, if this phenomenon were the unique or even the predominant, the expected results would have been a uniform releasing of exons or in a random appearance, without predominance of none.

The active and intelligent releasing of cfDNA links with the theory that such phenomenon is not a waste phenomenon, but a step of a putative pathway for tumor progression [30]. Horizontal transfer of such cfDNA between cells has been proposed as a pivotal mechanism in the development of metastasis both in vitro and in vivo, a phenomenon called genometastasis [30,31], which has been proved in mixed in vivo and in vitro models [32,33]. Thus, under this view, the cfDNA found in plasma of animals and patients might have been, to a greater or lesser extent, actively secreted by both tumor and non-tumor cells, and may contribute to metastasis.

Summarizing, the DPE analysis in tumor-bearing animals was able to discriminate between different disease status, which fully supports previous results in CRC patients. DPE analysis is affordable from WES results with a minimal bioinformatic structure. Hence, our findings encourage to analyze under this perspective any NGS results obtained in series of cancer patients, to gain complementary approaches of the status of the illness. In addition, the results found here support the theory of an active and directed releasing of cfDNA.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2020.100837.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The next-generation sequencing (NGS) data analysis has been performed by the Genomics and NGS Core Facility at the Centro de Biología Molecular Severo Ochoa (CBMSO, CSIC-UAM) which is part of the CEI UAM + CSIC, Madrid, Spain (http://www.cbm.uam.es/genomica).

Credit author statement

Dolores C. García-Olmo: Conceptualization, Methodology, Investigation, Resources, Data curation, Writing - original draft, Visualization, Supervision, Project administration, Funding acquisition.

Ramón Peiró-Pastor: Software, Validation, Formal analysis, Resources, Data curation, Writing - Original draft, Visualization.

María G. Picazo: Investigation, Resources, Data curation, Writing – review & editing.

Susana Olmedillas: Resources, Writing – review & editing, Project administration.

Mariano García-Arranz: Resources, Writing – review & editing, Project administration, Funding acquisition.

Begoña Aguado: Resources, Writing – review & editing, Project administration.

Damián García-Olmo: Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Financial support

This study was funded by two grants from "Instituto de Salud Carlos III", Spain (FIS; refs. PS09/01815 and PI13/01924).

References

[1] Y. Sato, R. Matoba, K. Kato, Recent advances in liquid biopsy in precision oncology research, Biol. Pharm. Bull. 42 (3) (2019) 337–342.
[2] V.A. Adalsteinsson, G. Ha, S.S. Freeman, A.D. Choudhury, D.G. Stover, H.A. Parsons, et al., Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors, Nat. Commun. 8 (1) (2017) 1324.
[3] D.C. García-Olmo, M.G. Picazo, I. Toboso, A.I. Asensio, D. García-Olmo, Quantitation of cell-free DNA and RNA in plasma during tumor progression in rats, Mol. Cancer 12 (2013) 8.
[4] K.C. Chan, P. Jiang, Y.W. Zheng, G.J. Liao, H. Sun, J. Wong, et al., Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing, Clin. Chem. 59 (1) (2013) 211–224.
[5] C.M. Stewart, D.W.Y. Tsui, Circulating cell-free DNA for non-invasive cancer management, Cancer Genet. 228–229 (2018) 169–179.
[6] S. Olmedillas Lopez, D.C. Garcia-Olmo, M. Garcia-Arranz, H. Guadalajara, C. Pastor, D. Garcia-Olmo, KRAS G12V mutation detection by droplet digital PCR in circulating cell-free DNA of colorectal cancer patients, International Journal of Molecular Sciences 17 (4) (2016).
[7] J.A.Y.C. Posey, Drivers of precision medicine: liquid biopsy and next-generation sequencing, Cancer Ther. Oncol. Int. J. 2 (2016) 555595.
[8] M. Kastrisiou, G. Zarkavelis, G. Pentheroudakis, A. Magklara, Clinical application of next-generation sequencing as a liquid biopsy technique in advanced colorectal cancer: a trick or a treat? Cancers (Basel) 11 (10) (2019).
[9] S. Olmedillas-López, D.C. García-Olmo, M. García-Arranz, R. Peiró-Pastor, B. Aguado, D. García-Olmo, Liquid biopsy by NGS: differential presence of exons (DPE) in cell-free DNA reveals different patterns in metastatic and nonmetastatic colorectal cancer, Cancer Med. 7 (5) (2018) 1706–1716.
[10] P. Workman, E.O. Aboagye, F. Balkwill, A. Balmain, G. Bruder, D.J. Chaplin, et al., Guidelines for the welfare and use of animals in cancer research, Br. J. Cancer 102 (11) (2010) 1555–1577.
[11] FastQC - a quality control tool for high throughput sequence data, Babraham Bioinformatics , Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Internet, cited June 20, 2017).
[12] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (3) (2009) R25.
[13] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics. 26 (1) (2010) 139–140.
[14] S. Anders, P.T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data, Bioinformatics. 31 (2) (2015) 166–169.
[15] S.P. Lund, D. Nettleton, D.J. McCarthy, G.K. Smyth, Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates, Stat Appl Genet Mol Biol. 11 (5) (2012).
[16] J. Ward, Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc. 58 (301) (1963) 9.
[17] A. Liaw, M. Wiener, Classification and regression by randomForest, R News 2 (3) (2002) 5.
[18] A. Krämer, J. Green, J. Pollard, S. Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis, Bioinformatics. 30 (4) (2014) 523–530.
[19] C. Kilkenny, W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman, Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research, PLoS Biol. 8 (6) (2010), e1000412.
[20] M. Gerlinger, A.J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, et al., Intratumor heterogeneity and branched evolution revealed by multiregion sequencing, N. Engl. J. Med. 366 (10) (2012) 883–892.
[21] E.R. Fearon, B. Vogelstein, A genetic model for colorectal tumorigenesis, Cell. 61 (5) (1990) 759–767.
[22] A. Sottoriva, H. Kang, Z. Ma, T.A. Graham, M.P. Salomon, J. Zhao, et al., A Big Bang model of human colorectal tumor growth, Nat. Genet. 47 (3) (2015) 209–216.
[23] T. Saito, A. Niida, R. Uchi, H. Hirata, H. Komatsu, S. Sakimura, et al., A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer, Nat. Commun. 9 (1) (2018) 2884.
[24] A.M. Rachiglio, R. Esposito Abate, A. Sacco, R. Pasquale, F. Fenizia, M. Lambiase, et al., Limits and potential of targeted sequencing analysis of liquid biopsy in patients with lung and colon carcinoma, Oncotarget. 7 (41) (2016) 66595–66605.
[25] D. Garcia-Olmo, J. Ontanon, D. Garcia-Olmo, M. Atienzar, M. Vallejo, Detection of genomically-tagged cancer cells in different tissues at different stages of tumor development: lack of correlation with the formation of metastasis, Cancer Lett. 140 (1–2) (1999) 11–20.
[26] D. Garcia-Olmo, L. Gutierrez-Gonzalez, R. Ruiz-Piqueras, M. Picazo, D. Garcia-Olmo, Detection of circulating tumor cells and of tumor DNA in plasma during tumor progression in rats, Cancer Lett. 217 (1) (2005) 115–123.
[27] D. Garcia-Olmo, M. Garcia-Rivas, D. Garcia-Olmo, M. Atienzar, Orthotopic implantation of colon carcinoma cells provides an experimental model in the rat that replicates the regional spreading pattern of human colorectal cancer, Cancer Lett. 132 (1–2) (1998) 127–133.
[28] D. Garcia-Olmo, M. Garcia-Rivas, D. Garcia-Olmo, J. Ontanon, The site of injection of tumor cells in rats does not influence the subsequent distribution of metastases, Oncol. Rep. 10 (4) (2003) 903–907.
[29] S. Jahr, H. Hentze, S. Englisch, D. Hardt, F.O. Fackelmayer, R.D. Hesch, et al., DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells, Cancer Res. 61 (4) (2001) 1659–1665.
[30] D.C. García-Olmo, D. García-Olmo, Biological role of cell-free nucleic acids in cancer: the theory of genometastasis, Crit. Rev. Oncog. 18 (1–2) (2013) 153–161.
[31] D. Garcia-Olmo, J. Ontanon, E. Martinez, Horizontal transfer of DNA and the "genometastasis hypothesis", Blood. 95 (2) (2000) 724–725.
[32] D.C. Garcia-Olmo, C. Dominguez, M. Garcia-Arranz, P. Anker, M. Stroun, J.M. Garcia-Verdugo, et al., Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells, Cancer Res. 70 (2) (2010) 560–567.
[33] D. Hamam, M. Abdouh, Z.H. Gao, V. Arena, M. Arena, G.O. Arena, Transfer of malignant trait to BRCA1 deficient human fibroblasts following exposure to serum of cancer patients, J. Exp. Clin. Cancer Res. 35 (2016) 80.