

RESEARCH

Open Access



Prediction of occult tumor progression via platelet RNAs in a mouse melanoma model: a potential new platform for early detection of cancer

Yue Yin^{1,2*} , Ruilan Jiang³, Mingwang Shen⁴, Zhaofang Li², Ni Yan², Junqiao Feng^{2,5}, Hong Jiang², Jiaxin Lv², Lijuan Shi², Lina Wang², Xi Liu², Kaiyun Zhang^{6,7} and Di Chen⁸

Abstract

Background: Cancer screening provides the opportunity to detect cancer early, ideally before symptom onset and metastasis, and offers an increased opportunity for a better prognosis. The ideal biomarkers for cancer screening should discriminate individuals who have not developed invasive cancer yet but are destined to do so from healthy subjects. However, most cancers lack effective screening recommendations. Therefore, further studies on novel screening strategies are urgently required.

Methods: We used a simple suboptimal inoculation melanoma mouse model to obtain 'pre-diagnostic samples' of mice with macroscopic melanomas. High-throughput sequencing and bioinformatic analysis were employed to identify differentially expressed RNAs in platelet signatures of mice injected with a suboptimal number of melanoma cells (eDEGs) compared with mice with macroscopic melanomas and negative controls. Moreover, 36 genes selected from the eDEGs via bioinformatics analysis were verified in a mouse validation cohort via quantitative real-time PCR. LASSO regression was utilized to generate the prediction models with gene expression signatures as the best predictors for occult tumor progression in mice.

Results: These RNAs identified from eDEGs of mice injected with a suboptimal number of cancer cells were strongly enriched in pathways related to immune response and regulation. The prediction models generated by 36 gene qPCR verification data showed great diagnostic efficacy and predictive value in our murine validation cohort, and could discriminate mice with occult tumors from control group (area under curve (AUC) of 0.935 (training data) and 0.912 (testing data)) (gene signature including *Cd19*, *Cdkn1a*, *S100a9*, *Tap1*, and *Tnfrsf1b*) and also from macroscopic tumor group (AUC of 0.920 (training data) and 0.936 (testing data)) (gene signature including *Ccr7*, *Cd4*, *Kmt2d*, and *Ly6e*).

Conclusions: Our proof-of-concept study provides evidence for potential clinical relevance of blood platelets as a platform for liquid biopsy-based early detection of cancer.

Keywords: Platelet RNA, Early detection, Occult tumor, Liquid biopsy, Mouse model

Background

Elegant studies have been designed to identify cancer biomarkers or signatures of indicating how patients would respond to therapies [1, 2]. These previous studies aim to diagnose malignancy or to examine the prognosis

*Correspondence: lycwr@hotmail.com

¹ Department of Histology and Embryology, School of Basic Medicine, Guilin Medical University, Guilin, Guangxi, China
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

such as future metastasis or long-term survival. However, the ideal biomarkers for the early detection of cancer should discriminate individuals with occult cancer that is destined to progress from healthy subjects.

Traditional cancer screening methods have demonstrated low accuracy and efficacy, while novel cancer markers such as circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA), which offer new genomic approaches through liquid biopsies, still have limited efficiency [3–6]. Indeed there are previous longitudinal studies using pre-diagnostic serums to screen for novel biomarkers of early cancer detection. These studies utilized pre-diagnostic serum to detect tumor-specific antigens or auto-antibodies for cancer risk prediction with limited sensitivity [7–10]. Novel cancer markers such as ctDNA have also promised to be a sensitive and specific test for cancer screening [11, 12]. However, ctDNA testing has several limitations for a screening platform compared with platelet RNA testing such as its limited abundance in blood, difficulty in extraction, dependency on known hotspots and its high expense. Therefore, further studies searching for new blood-based biomarkers for early cancer detection are now urgently required. Blood platelets, which are traditionally known for their function in hemostasis and thrombosis, have emerged as important participants in tumor pathogenesis [13, 14]. Recent studies have indicated significant platelet involvement in cancer growth and metastasis [15, 16]. It has been reported that tumor-educated platelets (TEPs) may have potential for cancer companion diagnostics [17–20]. However, whether platelets could serve as a platform for cancer risk assessment or early disease diagnostics still merits further investigation.

Methods

Mice

C57BL/6 mice were bred in the Laboratory Animal Center, Health Science Center, Xi'an Jiaotong University. All mice were female and aged between 6 and 8 weeks at the beginning of all experiments. Animal experiments were approved by the Animal Ethics Committee of Xi'an Jiaotong University. All animal experiments were set with a maximum endpoint at a tumor volume of 1000 mm³.

B16F10 cell line

B16F10 cells negative for mycoplasma contamination were cultured and passaged in RPMI-1640 medium containing 10% fetal calf serum (FCS) (Gemini, USA) at 37 °C/5% CO₂. For animal inoculations, cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Hyclone, USA) supplemented with 10% FCS and 5 µg/ml plasmocin prophylactic reagent (InvivoGen, Cat.: ant-mpp, USA) at 37 °C/5% CO₂ to induce better melanin

production. B16F10 cell line was generously provided by Dr Hui Zhang from Institute of Human Virology, Sun Yat-sen University and originally purchased from the ATCC.

B16F10 melanoma inoculation

For melanoma inoculation, B16F10 melanoma cells were harvested by washing with phosphate buffer saline (PBS), then incubating cells at 37 °C for 1–2 min with 1 × Trypsin/EDTA (Gibco, USA) solution and washing with Hanks' balanced saline solution (HBSS) twice. For B16F10 inoculation, the right flanks of mice were shaved with a mini-razor and cells (2 × 10³, 5 × 10³, 1 × 10⁴ and 1 × 10⁵ cells per mouse for each group) were suspended in 100 µl HBSS and then injected under the right flank subcutaneously using a 30G needle. Tumor formation was monitored by inspecting via a magnifying scope and measured with a caliper periodically (tumor volume was estimated using this formula: volume = length × width × height × 0.5) [21].

Blood sample collection

Blood samples were collected from retro-orbital sinuses of C57BL/6 mice 14 days after inoculation. For terminal or nonterminal blood collection, mice were fully anesthetized with isoflurane and blood samples were collected by puncturing the retro-orbital sinuses of mice using microhematocrit capillary tubes. Blood was collected into a tube containing the anti-coagulant EDTA. After nonterminal blood collection (less than 1% of body weight, approximately 150–200 µl), the tube was withdrawn and a slight pressure was put on the eye with a sterile cotton swab to ensure hemostasis. After terminal blood collection, mice were euthanized by cervical dislocation.

Isolation of platelet and PBMC RNA

Blood platelets isolation was performed as described previously [20, 22]. Briefly, anti-coagulated blood was centrifuged at 180 × g at room temperature for 10 min, yielding platelet-rich plasma. Platelets were isolated from the platelet-rich plasma by centrifugation at room temperature for 10 min at 1250 × g. The platelet pellet was lysed in TRIzol Reagent (Invitrogen, ThermoFisher Scientific, USA) and frozen at – 80 °C for future use.

The bottom layer of the centrifuged blood sample from the first step of platelet isolation was further used for peripheral blood mononuclear cell (PBMC) isolation using mouse PBMC isolation kit (TBD science, Tianjin) following the manufacturer's instructions. Briefly, the aforementioned bottom layer was mixed with the same volume of diluting solution provided by the manufacturer and the mixture was carefully layered on the PBMC isolation reagent of the same volume in a sterile centrifuge

tube. The tube was then centrifuged at $950 \times g$ for 30 min at room temperature. PBMC layer was transferred into a new tube from the interphase with a transfer pipette and washed twice by mixing with 10 ml washing solution (also provided by the manufacturer) followed by centrifuging at $250 \times g$ for 10 min. The final PBMC pellet was lysed in TRIzol and kept in -80°C for further use.

Next generation sequencing

Next generation sequencing was performed in Novogene (Tianjin, China). Briefly, platelet or PBMC RNA samples were assessed for quantity, purity and integrity using NanoPhotometer[®] spectrophotometer (IMPLEN, CA, USA) and RNA Nano 6000 Assay Kit of Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Samples were pooled to satisfy the quantity criteria of RNA sequencing and a minimum amount of 20 ng RNA per pooled sample was used as input material for the RNA sample preparations. For platelet samples, the numbers of mice sacrificed for 5 pooled samples in three groups were as follows: 3, 4, 5, 3 and 5 mice for O group (optimal inoculation group, mice inoculated with 1×10^5 B16F10 cells); 10, 10, 9, 11 and 10 mice for S group (suboptimal inoculation group, mice inoculated with 2×10^3 B16F10 cells); 10, 5, 5, 10 and 10 mice for C group (negative control group, mice injected with HBSS). For PBMC samples, the numbers of mice sacrificed for 5 pooled samples in three groups were as follows: 3, 4, 5, 3 and 5 mice for O group; 5, 4, 5, 5 and 5 mice for S group; 5 mice for each sample in C group.

Sequencing libraries were generated using NEBNext[®] UltraTM RNA Library Prep Kit for Illumina[®] (NEB, USA) following the manufacturer's recommendations with index codes added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation of mRNA was carried out followed by cDNA synthesis. Sequencing libraries were created by converting RNA to cDNA via reverse transcription and adding specialized adapters to both ends. Next, library fragments were purified and then PCR was performed with Phusion High-Fidelity DNA polymerase. PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Finally, qualified libraries were sequenced on an Illumina Novaseq platform.

Sequencing data analysis

Novogene provided sequence alignment, data mapping and differential expression analysis. Briefly, raw data of fastq format were processed to obtain clean reads with high quality by removing reads containing adapter, reads containing poly-N and reads of low quality from raw

data. Clean reads were aligned to mus musculus reference genome using Hisat2 v2.0.5 alignment program. The R package featureCounts v1.5.0-p3 was applied to count the reads numbers mapped to each gene [23]. And then FPKM (fragments per kilobase of transcript sequence per millions base pairs sequenced) of each gene was calculated based on the length of the gene and read count mapped to this gene.

Differential expression analysis was performed using DESeq2 R package 1.16.1 [24]. The resulting *P* values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rates. Genes with *P* values < 0.05 were assigned as differentially expressed for pairwise comparisons. Differentially expressed genes for subsequent screening of distinct markers of early-early cancer (eDEGs) must meet these requirements: $\log_2(\text{S vs. C Fold Change}) > 0$, $P < 0.05$, $\log_2(\text{S vs. O Fold Change}) > 0$, $P < 0.05$; or $\log_2(\text{S vs. C Fold Change}) < 0$, $P < 0.05$, $\log_2(\text{S vs. O Fold Change}) < 0$, $P < 0.05$. O group, optimal inoculation group, mice inoculated with 1×10^5 B16F10 cells; S group, suboptimal inoculation group, mice inoculated with 2×10^3 B16F10 cells; C group, negative control group, mice injected with HBSS.

Bioinformatics analyses and data visualizations

Data visualizations were performed using R (version 3.6.3) [25]. Heatmaps and clusterings were generated using pheatmap package [26]. Dot plots and bubble plots were generated using ggplot2 and corplot packages [27, 28]. Pathway enrichment analyses of differentially expressed genes in suboptimal inoculation group (eDEGs) were performed using clusterProfiler package with reference from KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways with *P* values adjusted by Benjamini and Hochberg method [29, 30]. The protein-protein interaction (PPI) network of eDEGs was retrieved from STRING database [31] and reconstructed using Cytoscape [32]. Each node's degree of connectivity in the network was calculated. Molecular COMplex DETection (MCODE) [33] was used to find gene clusters based on topology locating densely connected regions.

Quantitative real-time PCR

Blood platelet RNA was isolated as described above. Platelet RNA was then converted to complementary cDNA using PrimeScript RT Master Mix (Takara) according to the manufacturer's instructions. Quantitative real-time PCR (qPCR) was performed with TB Green Premix Ex Taq (Takara, Beijing) using LightCycler 96 System (Roche Life Science, USA) with parameters adjusted according to the PCR cyclers and the enzyme's manuals. The reaction process was as follows: preincubation at 95°C for 30 s; 40 cycles of 5 s at 95°C and 30 s at

55 °C for annealing and extension; melting at 95 °C for 5 s, 60 °C for 60 s and 95 °C for 1 s; cooling for 30 s at 50 °C. Cycle-threshold (Ct) values were determined for each gene and normalized to the housekeeping genes *Actb*, *Gapdh*, *Gusb*, *Gnas* and *Oaz1*, which were validated previously as reliable reference genes for platelet RNA qPCR [34, 35].

Statistic analyses

All statistical analyses were performed in PASW Statistics 18 (SPSS, USA) or Prism 8 (Graphpad, USA) and were two-sided. All experiments were performed with replicates as indicated, either with representative data shown, or with pooled data shown. Figures with pooled data from multiple experiments included all experiments performed. All data reflected multiple independent experiments with at least 3 mice per experiment, in which similar results were obtained.

Kaplan–Meier curves were generated to illustrate the relationship between percentage of tumor-free mice and time after inoculation. Mantel-Cox tests were used to test statistic significance.

Joinpoint software version 4.8.0.1 (National Cancer Institute, USA) was used to analyze tumor growth data for multi-phase regression in order to determine when the tumor went from occult state into fast progressing phase [36]. A maximum of 1 joinpoint was allowed based on the number of data points and previous studies on the role of angiogenesis in tumor dormancy [37]. The statistic significance of the change in tumor growth trend over time was tested using a Monte Carlo Permutation method embedded in the Joinpoint software [36]. Blood samples from mice whose tumor became visible more than 5 days after blood collection (19 days post-inoculation) were categorized as early-early group (E group). Alternatively, samples from mice with a micro-tumor (tumor volume < 1 mm³) on the day of blood collection (14 days post-inoculation) that did not progress for at least 20 days since inoculation (joinpoint ≥ 20, Permutation test *P* value < 0.05) were also classified as early-early group (E group). Blood samples from mice with macroscopic and palpable tumors (tumor volume > 30 mm³) were categorized as melanoma group (M group) and blood samples from mice injected with HBSS were in negative control group (C group).

Statistic analyses of qPCR results were performed via SPSS 18.0. Kruskal–Wallis non-parametric test was executed and adjusted *P* value below 0.05 was assigned as significant. Samples with more than 10 genes with invalid Ct values (no signal within 40 cycles of PCR due to low RNA quantity) out of 36 markers were excluded. Then genes with invalid Ct values in at least 5 samples in both dependent variable groups (probably due to low expression levels

of certain chosen markers in platelets) were not included for subsequent variable selection via LASSO binary regression analysis.

The Least Absolute Shrinkage and Selection Operator (LASSO) model is a shrinkage method for regression with high-dimensional predictors, which can preserve valuable variables from a large and potentially multicollinear set of variables, and avoid overfitting. This method is suited for analyzing gene expression data where multicollinearity of selected genes in related biological pathways may occur. We performed LASSO binary logistic regression using glmnet R package [38]. Data were randomly divided into training set (70%) and testing set (30%). We utilized ten-fold cross-validation to select the penalty term λ with the alpha of 1. The binomial deviance was set as measures of the predictive performance of the fitted models. The built-in function in glmnet package produced the λ that minimized the binomial deviance. The coefficients of selected variables were obtained through the penalizing process. The seed was set to 10 for data replication. The prediction score formulas for the discrimination of early-early tumor (occult tumor, E group) from negative control group (C group) or from macroscopic melanoma group (M group) were established as follows: Score = Intercept + \sum Coefficient \times (Ct_{Variable} - Ct_{Ref}).

Ridge regression and elastic net regression were also performed as well as LASSO regression for comparison of models. For ridge regression, data were also randomly divided into training set (70%) and testing set (30%). We utilized ten-fold cross-validation to select the penalty term λ with the alpha of 0. The binomial deviance was set as measures of the predictive performance of the fitted models. The built-in function in glmnet package produced the λ that minimized the binomial deviance. The coefficients of variables were obtained through the penalizing process. For elastic net regression, we used caret and tidyverse packages to determine the optimal alpha and lambda combination [39, 40]. The seed was set to 10 for data replication. The prediction score formulas were as described above. Root mean square error (RMSE) were calculated and compared between 3 regression models.

Receiver operating characteristic (ROC) curves were constructed and area under curve (AUC) was calculated using SPSS 18.0. Probability statistics for ROC were calculated according to the prediction score formulas generated from LASSO regression analyses described above: Probability = $e^{\text{Score}} / (1 + e^{\text{Score}})$.

Results

Inoculation of suboptimal numbers of tumor cells can induce delayed melanoma formation in mice

To investigate these issues, we established a melanoma mouse model by inoculating C57BL/6 mice with a

suboptimal number of tumor cells, which were named an “early-early” mouse model since the tumors in our model were occult or microscopic before they rapidly grew and became macroscopic. C57BL/6 mice were subcutaneously injected with several concentrations of B16F10 cells. In contrast with the rapid tumor development in mice from the group injected with a optimal number of 1×10^5 cells per mouse established by previous studies [41, 42], inoculation of mice with lower numbers of cells postponed the onset of tumor formation, with variable growth kinetics, as evidenced by the dispersion of growth curves (Fig. 1a). In groups injected with 1×10^5 cells and 1×10^4 cells per mouse, all mice developed tumors which became visible in 2 weeks and 3 weeks post-inoculation respectively (Fig. 1a, b). However, injection with suboptimal numbers of cells (5×10^3 cells or 2×10^3 cells per mouse) induced tumors in some subjects that remained small and did not progress for as long as 6 weeks after inoculation (Fig. 1a, b). Around 76% of mice (100 out of 131) injected with 5×10^3 cells per mouse developed tumors that became visible at 2–6 weeks after inoculation, while only 13% of mice (8 out of 60) injected with 2×10^3 cells per mouse formed visible tumors within 6 weeks post-inoculation (Fig. 1b, c). Moreover, around 24% of mice from the group injected with 5×10^3 cells per mouse and 87% of mice injected with 2×10^3 cells did not develop melanomas within 6 weeks after inoculation and remained tumor-free for a prolonged period of 15 weeks post-inoculation (Fig. 1b, c).

Since some tumors developed late (Fig. 1a), these “late-developer” mice harbored occult or microscopic melanomas after inoculation for weeks before they progressed into macroscopic tumors afterwards. We proposed that the “pre-diagnostic” blood samples from mice inoculated with a suboptimal number of tumor cells could be used for screening novel “early-early” cancer biomarkers.

Our data showed that samples from mice inoculated with higher number of cells could not be used for RNA sequencing for “early-early” model, since tumors were already in late stage (Fig. 1). Since the group inoculated with 2×10^3 B16F10 cells could yield a very low proportion of mice with visible tumors by 14 days for blood sample collection (Fig. 1), we could euthanize most of the mice for RNA sequencing. So we chose 2×10^3 B16F10 cells for inoculation of “early-early” group for maximum usage of mice.

Platelet mRNA profiles of mice inoculated with a suboptimal number of tumor cells are distinct from those of both healthy and tumor-bearing mice

To screen for novel “early-early” cancer biomarkers in our melanoma model, we collected blood samples from optimal inoculation group (1×10^5 cells per mouse, O group), suboptimal inoculation group (2×10^3 cells per mouse, S group), and negative control group (C group) on day 14 post-injection, when optimal inoculation group all developed palpable tumors and suboptimal inoculation group did not form visible tumors yet, indicating the tumors

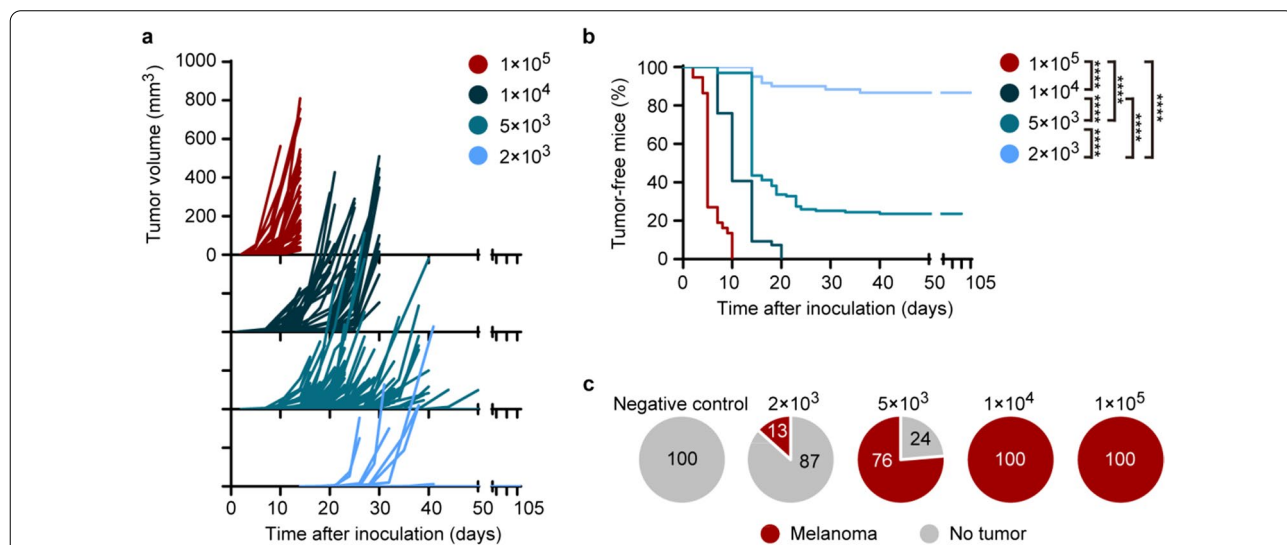


Fig. 1 Tumor growth kinetics in C57BL/6 mice inoculated with different numbers of B16F10 cells. **a** Tumor growth curves after inoculation with optimal and suboptimal numbers of B16F10 cells. **b, c** Proportion of tumor-free and tumor-bearing mice following inoculation with B16F10 cells. Data pooled from $n=8$ (1×10^5 cells per mouse, red), $n=12$ (1×10^4 cells per mouse, dark blue), $n=29$ (5×10^3 cells per mouse, blue) and $n=14$ (2×10^3 cells per mouse, light blue) biologically independent experiments with $n=37$ mice (1×10^5 cells per mouse), $n=54$ mice (1×10^4 cells per mouse), $n=131$ mice (5×10^3 cells per mouse) and $n=60$ mice (2×10^3 cells per mouse). **** $P < 10^{-10}$, log-rank Mantel-Cox test (**b**)

might still be occult (Fig. 2a). Mice were autopsied after blood collection and examined under a magnifying scope as well as pathological examination to confirm there were no visible tumors in S group (Additional file 1: Fig. S1a).

Recently, it has been reported that tumor-educated platelets (TEPs) may have potential for cancer companion diagnostics [17–20]. Therefore, we isolated blood platelets for further study as well as peripheral blood mononuclear cells (PBMCs) for comparison. Both platelet RNA and PBMC RNA were isolated and evaluated for quantity and quality. Total platelet RNA samples of 20 mice in O group, 50 mice in S group and 40 mice in C group were pooled into 5 samples per group in order to meet the quantity criteria of RNA sequencing. The disparity in platelet RNA quantity of mice from different groups was probably due to higher platelet production via thrombocytosis in tumor-bearing mice [43]. Total PBMC RNA samples of 20 mice in O group, 24 in S group and 25 in C group were also pooled into 5 samples per group to guarantee sufficient RNA quantity. Pooled platelet and PBMC RNA samples were then

processed for RNA sequencing. Platelet RNA sequencing yielded a mean read count of around 53 million clean reads per sample, while PBMC RNA sequencing yielded about 44 million clean reads per sample. After genome mapping of RNA reads, we identified among the platelet RNAs known platelet-abundant genes, such as beta-2 microglobulin (*B2m*), ferritin heavy polypeptide 1 (*Fth1*), platelet factor 4 (*Pf4*), pro-platelet basic protein (*Pppb*) and thymosin, beta 4, X chromosome (*Tmsb4x*) (Additional file 1: Fig. S1b), which yielded much higher read counts than average level. The obtained platelet RNA profiles correlated with PBMC RNA profiles, but the correlation between platelet and PBMC RNA profiles was much less prominent than that between samples within the platelet or PBMC group (Additional file 1: Fig. S1c). Moreover, mRNAs such as *B2m*, *Tmsb4x*, *Pppb*, regulator of G-protein signaling 18 (*Rgs18*), cathepsin B (*Ctsb*), calreticulin (*Calr*), eukaryotic translation elongation factor 2 (*Eef2*) and insulin-like growth factor binding protein 4 (*Igfbp4*), which were previously reported differentially expressed

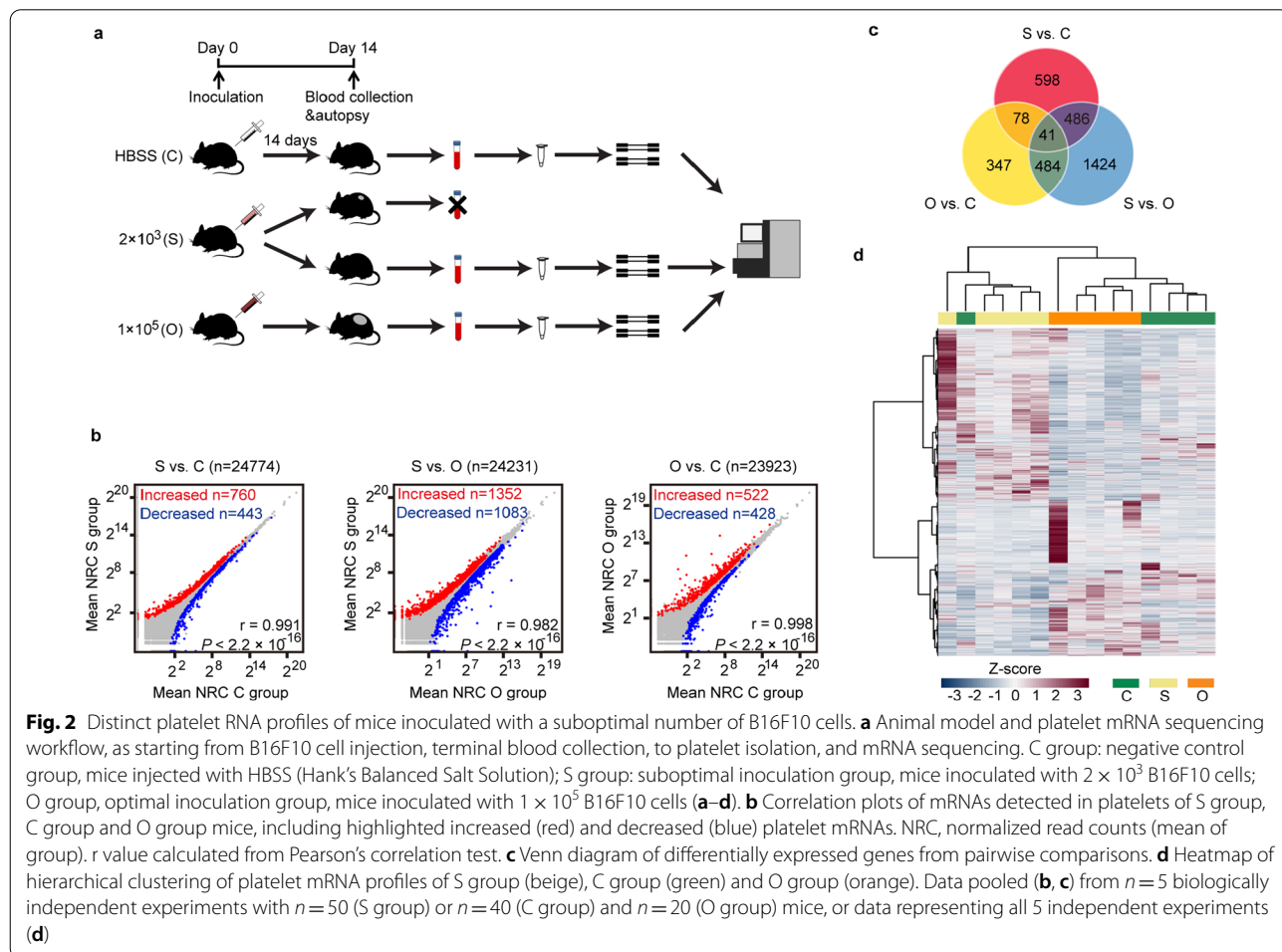


Fig. 2 Distinct platelet RNA profiles of mice inoculated with a suboptimal number of B16F10 cells. **a** Animal model and platelet mRNA sequencing workflow, as starting from B16F10 cell injection, terminal blood collection, to platelet isolation, and mRNA sequencing. C group: negative control group, mice injected with HBSS (Hank’s Balanced Salt Solution); S group: suboptimal inoculation group, mice inoculated with 2×10^3 B16F10 cells; O group, optimal inoculation group, mice inoculated with 1×10^5 B16F10 cells (**a–d**). **b** Correlation plots of mRNAs detected in platelets of S group, C group and O group mice, including highlighted increased (red) and decreased (blue) platelet mRNAs. NRC, normalized read counts (mean of group). r value calculated from Pearson’s correlation test. **c** Venn diagram of differentially expressed genes from pairwise comparisons. **d** Heatmap of hierarchical clustering of platelet mRNA profiles of S group (beige), C group (green) and O group (orange). Data pooled (**b, c**) from $n = 5$ biologically independent experiments with $n = 50$ (S group) or $n = 40$ (C group) and $n = 20$ (O group) mice, or data representing all 5 independent experiments (**d**)

genes between platelet and PBMC profiles [44], were also differentially expressed in our sequencing data (Additional file 1: Fig. S1d).

Since our samples were pooled from about 10 mice into 1 sequencing sample, the S group inevitably included mice that may never develop tumors. This strategy would still allow us to identify the differentially expressed genes between S group, C group and O group, only with lower fold-change results, which is why we set the $\log_2(\text{fold-change})$ to 0 as threshold for differentially expressed gene screening.

A total of 760 out of 24,774 mRNAs were increased and 443 out of 24,774 mRNAs were decreased in platelet samples of S group as compared to samples of C group, while presenting a strong correlation between these platelet profiles ($r=0.991$, Pearson's correlation) (Fig. 2b left). A total of 1352 out of 24,231 mRNAs were increased and 1,083 mRNAs were decreased in S group compared with O group ($r=0.982$, Pearson's correlation) (Fig. 2b middle). Out of 23,923 mRNAs, 522 were increased and 428 were decreased in O group compared to C group ($r=0.998$, Pearson's correlation) (Fig. 2b right). For PBMC samples, a total of 239 out of 31,625 mRNAs were increased and 251 out of 31,625 mRNAs were decreased in PBMC samples of S group as compared to samples of C group, also presenting a strong correlation between these PBMC mRNA profiles ($r=0.996$, Pearson's correlation) (Additional file 1: Fig. S2a left). A total of 732 out of 31,263 mRNAs were increased and 1,477 mRNAs were decreased in S group compared with O group ($r=0.833$, Pearson's correlation) (Additional file 1: Fig. S2a middle). Out of 26,630 mRNAs, 1,676 were increased and 868 were decreased in O group compared to C group ($r=0.803$, Pearson's correlation) (Additional file 1: Fig. S2a right). We detected in platelets 3,458 and in PBMCs 3,694 differentially expressed protein coding and non-coding RNAs by multiple pairwise comparisons, which were used for subsequent investigations (Fig. 2c, Additional file 1: Fig. S2b). Hierarchical clustering based on differentially detected platelet mRNAs distinguished 3 sample groups with minor overlap, while clustering based on PBMC mRNAs could not quite discriminate S group from C group (Fig. 2d, Additional file 1: Fig. S2c).

Blood platelets provide novel biomarkers to predict occult tumor progression in mice

To screen for distinct markers of occult melanoma, we searched for genes differentially expressed in S group compared with both C group and O group (eDEGs, “e” as in “early-early mouse model”) (Fig. 3a, details in “Methods”). Compared with 524 eDEGs (436 were protein-coding genes) from platelet data, there were only 149 genes (only 63 were protein-coding genes) in PBMC data that fulfilled our criteria (Additional file 2: Tables S9, Additional file 3: Table S10). KEGG pathway analysis revealed that these differentially expressed mRNAs from platelets of suboptimal inoculation group (eDEGs) were strongly enriched for biological processes related with immune response or regulation, such as “cytokine receptor interaction” and “cell adhesion molecules” (Fig. 3b Additional file 1: Table S1), whereas the differentially expressed genes in PBMC mRNAs were only enriched for two biological processes with low gene counts in each pathway (Additional file 1: Fig. S2d, Additional file 1: Table S2). Therefore, platelet RNA profiles might provide a platform for screening novel biomarkers of occult tumor. Furthermore, platelets were the optimum biosource for screening new biomarkers for early cancer detection since PBMC RNA profiles failed to yield enough eDEGs or significantly enriched pathways. To better understand the interplay among the identified eDEGs in platelets, we obtained the protein–protein interaction (PPI) network using the online STRING tool [31]. The complicated network was made up of 25 modules, including 351 nodes and 1,148 edges and the top three significant modules were selected for further analysis (Fig. 3c–e). The first module contained 26 nodes and 270 edges, including toll-like receptor 9 (*Tlr9*), intercellular adhesion molecule 1 (*Icam1*), chemokine (C–C motif) receptor 7 (*Ccr7*), interferon gamma (*Ifng*), etc. (Fig. 3c). In the second and third module, there were only 10 nodes found and the degree values of genes (edges) were much lower than those in the first module (Fig. 3d, e). Combining literature search with our bioinformatics analyses, we finally selected 36 genes from our platelet data for subsequent experimental validation (Fig. 3f).

We established a validation cohort using our “early-early” melanoma model to test the diagnostic efficacy and predictive value of the aforementioned 36 biomarkers.

(See figure on next page.)

Fig. 3 Bioinformatics analyses of differentially expressed genes from suboptimal inoculation group. **a** Schematics of screening strategy for differentially expressed genes for screening early cancer biomarkers (eDEGs). Eligible genes differentially expressed between S (suboptimal inoculation group, mice inoculated with 2×10^3 cells) and C group (negative control group, mice injected with HBSS), and also between S and O group (optimal inoculation group, mice inoculated with 1×10^5 cells), shown in red columns (eligible eDEGs criteria see details in “Methods”). **b** Top GO terms of pathway enrichment analysis of eligible eDEGs with reference from KEGG pathways. Adjusted P value < 0.05 , Benjamini and Hochberg method. **c–e** Top 3 PPI networks modules of eligible eDEGs. Color of a node in the PPI network: \log_2 (Fold change, FC) value of normalized read counts of genes from S group compared with C group; Size of a node: number of interacting proteins with the designated protein (**c–e**). **f** Panel of 36 genes screened from mRNA sequencing data besides 5 reference genes

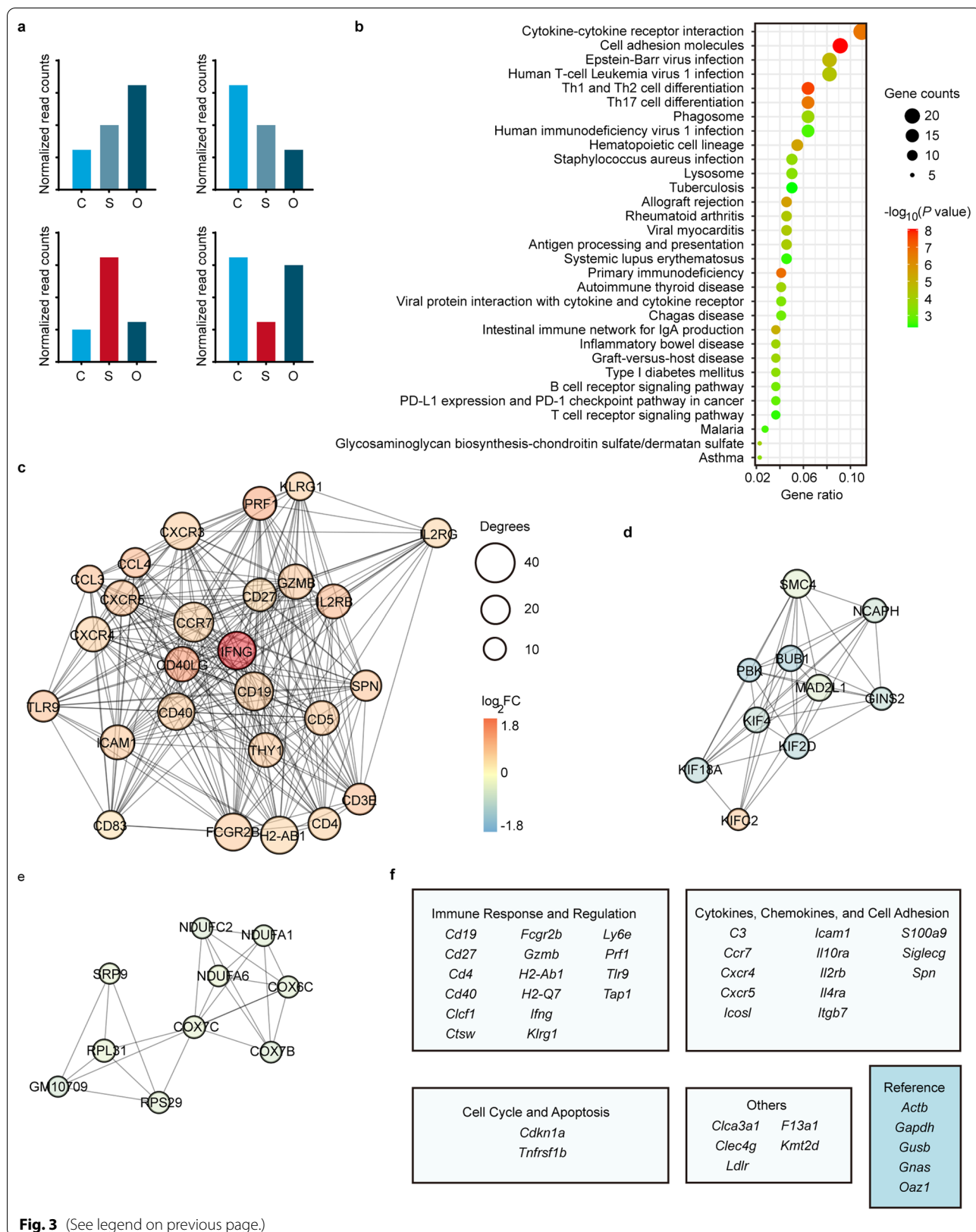


Fig. 3 (See legend on previous page.)

Mice inoculated in our previous experiments were divided into 3 groups according to their tumor development status on the day of blood collection (day 14 post-inoculation) (Fig. 4a). Mice with occult tumors on day 14 post-inoculation, which were validated by subsequent

tumor progression at least 5 days after blood collection, were categorized as early-early tumor group (E group) (Additional file 1: Fig. S3, Additional file 1: Table S3, details in “Methods”). Mice with macroscopic tumors were classified as melanoma group (M group) and mice

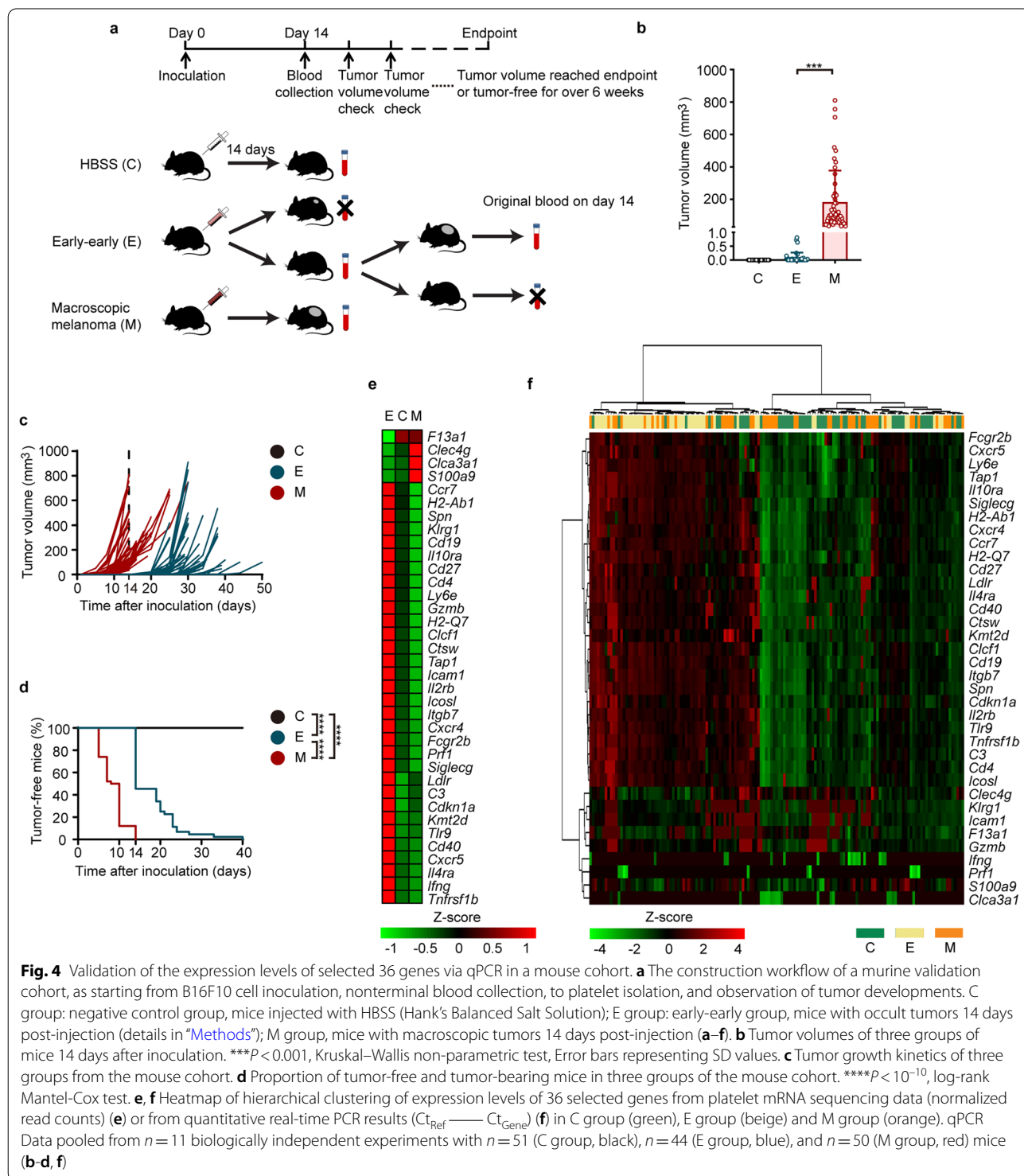


Fig. 4 Validation of the expression levels of selected 36 genes via qPCR in a mouse cohort. **a** The construction workflow of a murine validation cohort, as starting from B16F10 cell inoculation, nonterminal blood collection, to platelet isolation, and observation of tumor developments. C group: negative control group, mice injected with HBSS (Hank’s Balanced Salt Solution); E group: early-early group, mice with occult tumors 14 days post-injection (details in “Methods”); M group, mice with macroscopic tumors 14 days post-injection (**a–f**). **b** Tumor volumes of three groups of mice 14 days after inoculation. ***P < 0.001, Kruskal–Wallis non-parametric test, Error bars representing SD values. **c** Tumor growth kinetics of three groups from the mouse cohort. **d** Proportion of tumor-free and tumor-bearing mice in three groups of the mouse cohort. ****P < 10⁻¹⁰, log-rank Mantel–Cox test. **e, f** Heatmap of hierarchical clustering of expression levels of 36 selected genes from platelet mRNA sequencing data (normalized read counts) (**e**) or from quantitative real-time PCR results (Ct_{Ref} — Ct_{Gene}) (**f**) in C group (green), E group (beige) and M group (orange). qPCR Data pooled from n = 11 biologically independent experiments with n = 51 (C group, black), n = 44 (E group, blue), and n = 50 (M group, red) mice (**b, d, f**)

injected with HBSS were in negative control group (C group) (Fig. 4a). Tumor volumes measured on blood collection day showed that E group mice had no visible tumor or only small tumors (volume < 1 mm³) that did not progress for more than 2 weeks post-inoculation, verified by multi-phase regression analysis via Joinpoint program (Fig. 4b, Additional file 1: Fig. S3, Additional file 1: Table S3) [36]. Tumor growth kinetics demonstrated that E group mice eventually developed macroscopic melanomas afterwards, much later than M group (Fig. 4c, d).

Quantitative real-time PCR (qPCR) experiments were performed to validate the selected 36 genes from eDEGs in our mouse validation cohort (Additional file 1: Table S4). The normalized expression levels of the 36 genes were mostly in accordance with our previous sequencing data, except chloride channel accessory 3A1 (*Clca3a1*), coagulation factor XIII, A1 subunit (*F13a1*), *Ifng*, perforin 1 (pore forming protein) (*Prf1*) and S100 calcium binding protein A9 (*S100a9*), which yielded non-significant results (Additional file 1: Fig. S4, Additional file 1: Table S5). Hierarchical clustering of Δ Ct values could discriminate E group from C group and M group, which was consistent with our sequencing data (Fig. 4e, f). Although most selected genes could be validated in qPCR experiments, the quantities of several platelet RNA samples were too low for 36-gene-panel qPCR experiments. Therefore, 10 samples were excluded for subsequent analysis. Moreover, some markers such as *Clca3a1*, *Ifng* or *Prf1*, yielded invalid Ct value in over 20% of samples from each group, probably due to low abundance in platelets (Additional file 1: Table S5). Therefore, 7 genes including *Clca3a1*, *F13a1*, granzyme B (*Gzmb*), *Icam1*, *Ifng*, killer cell lectin-like receptor subfamily G, member 1 (*Klrg1*), and *Prf1*, were not used for subsequent regression analysis of E and C group. However, *Gzmb* was included in regression analysis of E and M group since it yielded valid Ct results in more than 90% of samples in each group (Additional file 1: Table S5). Ultimately there were 29 genes and 30 genes included as independent variables in subsequent regression analyses of E vs. C group and E vs. M group.

LASSO binomial logistic regression was applied to generate the prediction model with a multi-gene expression signature as the best predictor for occult tumor progression in mice. Cross-validation was carried out in 10 folds to prevent overfitting (internal training sets and internal validation sets constructed randomly) (Additional file 1: Fig. S5a, b). We also compared LASSO with ridge and elastic net regression, which yielded similar gene signatures with LASSO (Additional file 1: Fig. S6, Tables S6–S8). Finally the optimal gene signature consisting of CD19 antigen (*Cd19*), cyclin-dependent kinase inhibitor 1A (P21) (*Cdkn1a*), *S100a9*, transporter 1, ATP-binding

cassette, sub-family B (MDR/TAP) (*Tap1*), tumor necrosis factor receptor superfamily, member 1b (*Tnfrsf1b*) for E vs. C group, and *Ccr7*, CD4 antigen (*Cd4*), lysine (K)-specific methyltransferase 2D (*Kmt2d*), lymphocyte antigen 6 complex, locus E (*Ly6e*) for E vs. M group, as well as the corresponding coefficients were identified by the regularization process of LASSO regression (Additional file 1: Fig. S5c). Predictive scores for tumor progression were calculated from qPCR data using the training set of 63 mice for E vs. C group and 60 mice for E vs. M group. The scores were then tested in the validation set of 27 and 25 mice for E vs. C and for E vs. M group respectively. The biomarker score formula for E vs. C group could discriminate E group from C group with an area under curve (AUC) of 0.935 (training data) and 0.912 (testing data) (Fig. 5a). Moreover, the score formula for E vs. M group could also distinguish E group from M group with an AUC of 0.920 (training data) and 0.936 (testing data) (Fig. 5b).

Discussion

Although platelets have been suggested as a valuable platform for cancer diagnostics [20, 45, 46], studies have yet to address their potential as a cancer screening platform. We used mouse melanoma cell line B6F10 to inoculate immune-competent wild-type C57BL/6 mice, for unlike studies with immune-compromised mice, mice with intact immune system could better simulate the interaction between cancer and the immune system such as “cancer immunoediting” in tumor immune microenvironment. By using a mouse model that is simple, affordable and efficient, we identified differentially expressed RNAs in platelet signatures of mice injected

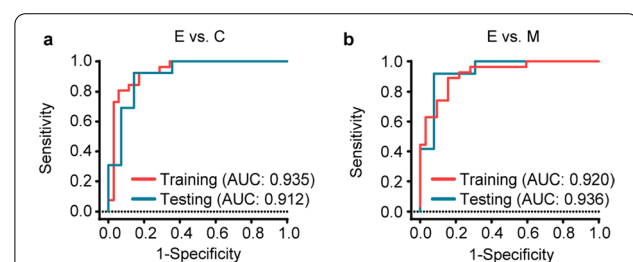


Fig. 5 ROC curves of the diagnostic performances of gene predictors for occult tumor progression in mice. ROC curves for the diagnostic performances of the prediction score formulas generated from LASSO regression in the mouse cohort. ROC curves for the discrimination of early-early tumor (occult tumor, E group) from negative control group (C group) (a, E vs. C) or from macroscopic melanoma group (M group) (b, E vs. M). Probability statistics calculated according to the prediction score formulas generated from LASSO regression analyses: Probability = $e^{\text{Score}} / (1 + e^{\text{Score}})$. 95% CI of AUC: training data 0.872–0.999, testing data 0.799–1.000 (a); training data 0.852–0.988, testing data 0.837–1.000 (b)

with a suboptimal number of tumor cells, compared with mice with large melanomas and negative controls. These genes presented strong positive correlations with RNAs implicated in immune response and regulation. This possibly reflects the interactions between tumor cells and the immune system in the early stage of tumorigenesis. Moreover, the lack of enriched biological pathways from PBMC samples suggests platelets are the optimum bio-source for early detection of cancer.

Finally, our study selects the optimal gene-expression-signature for the prediction of cancer risk via quantitative real-time PCR via LASSO regression. CD19 is an essential receptor for B cell antigen receptor (BCR) signal transduction. Co-ligation of CD19 could enhance mitogen-activated protein kinase activity and cell proliferation, and could also negatively regulate BCR signaling. Anti-CD19 chimeric antigen receptor T cells are currently used in transformational therapy for aggressive B-cell lymphomas [47, 48]. CDKN1A (P21) is a member of cyclin-dependent kinase inhibitors and it has been regarded as a tumor suppressor by regulating the cell cycle and maintaining genomic stability. The downregulation of CDKN1A is linked to poor prognosis in multiple cancers [49]. However, its overexpression is also found in a variety of human cancers [49]. Moreover, the upregulation of CDKN1A and its frequently cytoplasmic relocation correlate positively with poor prognosis in gastric cancer. The role of CDKN1A on tumorigenesis depends on the cellular context, its subcellular localization and posttranslational modifications. The application of CDKN1A as a prognostic marker and a therapeutic target in cancer still require further investigation [49]. S100A9 (calgranulin B, Calprotectin) is a Ca(2+) binding protein involved in inflammatory processes. S100A9 was elevated in inflammation and various human cancers [50]. Recent studies have shown the presence of S100A9 and inflammatory factors in the tumor micro-environment. The function of S100A9 depends on its concentration and location. S100A9 at high extracellular concentrations could induce the apoptosis pathway in cancer cells, while at lower levels S100A9 seem to promote proliferation of tumor cells [50]. However, at high intracellular concentrations, S100A9 induces a reduction in cancer cell invasion capacity by regulating the epithelial–mesenchymal transition–the mesenchymal–epithelial transition (EMT–MET) signaling cascades [50]. The molecular mechanism of pro- and anti-tumor properties of S100A9 is still unknown. TAP1 is associated with antigen processing of major histocompatibility complex class I peptides for recognition by tumor-specific cytotoxic T lymphocytes. TAP1 overexpression might be an indicator of aggressive breast cancer and was also significantly associated with poor prognosis in colorectal cancer

[51–53]. Moreover, decreased TAP1 protein expression was significantly associated with low infiltration of lymphocytes and macrophages [51, 52]. Bioinformatic study with large datasets demonstrated a correlation between the TAP1 gene and tumor progression and a significant negative correlation for TAP1 gene expression and the survival rate in different cancer types [53]. TNFRSF1B belongs to TNF receptor (TNFR) superfamilies and plays an important role in protective immunity, inflammatory and tumor immunology. TNFRSF1B can induce downstream signaling pathways such as NF- κ B and PI3K/Akt activation when interacted with its ligand TNF α [54]. TNF α -mediated co-stimulation supports TCR/CD28-mediated T cell activation and survival [54]. Furthermore, ligation of TNFRSF1B inhibits regulatory T cell differentiation by suppressing Smad3-dependent Foxp3 transcription [54]. CCR7 when ligated with its ligands, could induce the homing of T cells to a lymph node. Therefore, the increased expression of CCR7 has an anti-cancer effect via cytotoxic TIL in tumors [55]. However, CCR7 could also enhance proliferation and stemness of cancer cells. The mechanisms of the tumor-promoting effect of CCR7 include the induction of tumor angiogenesis by activating NF- κ B and increasing VEGF expression, epithelial–mesenchymal transition of cancer cells and migration of cancer cells to metastasis sites [55]. Higher expression of CCR7 is also associated with worse prognosis in diffuse large B-cell lymphoma [55]. CD4 is a coreceptor with the T-cell receptor on the T lymphocyte and CD4(+) T cell help signals are relayed to CD8(+) T cells by specific dendritic cells to optimize cytotoxic T lymphocyte (CTL) response [56]. Deficient CD4(+) T cell help can reduce the response of CTLs and maximizing CD4(+) T cell help can improve outcomes in cancer immunotherapy [56]. KMT2D belongs to the lysine methyltransferase (KMT2) family of histone modifying proteins, which play essential roles in regulating developmental pathways. The KMT2A-D proteins are important for RNA Polymerase II-dependent transcription and KMT2 mutations have been linked to multiple cancers [57]. Recent studies have also provided evidence for KMT2 protein participation in epigenetic gene regulation and in carcinogenesis [57]. LY6E belongs to the LY6 gene family, which represent novel biomarkers for poor cancer prognosis and play essential roles in cancer progression and immune escape [58]. LY6E expressions are increased in bladder cancer, gastric cancer, etc. [58]. The LY6E gene has also been associated with more aggressive stem like cells in hepatocellular carcinoma, pancreatic carcinoma, etc. [58]. Recent data suggest that increased expression of LY6E is associated with poor overall survival of renal papillary cell carcinoma and pancreatic ductal adenocarcinoma [58]. These genes in our final gene signature are

mostly protein coding genes involved in cancer immunology. It is possible that platelet mRNA could be used in cancer risk prediction in other types of cancers.

Indeed there are previous longitudinal studies using pre-diagnostic serums to screen for novel biomarkers of early cancer detection. However, these studies utilized pre-diagnostic serum to detect tumor-specific antigens or auto-antibodies for cancer risk prediction with limited sensitivity [7–10]. Novel cancer markers such as circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA) offer new genomic approaches to screen for cancer through liquid biopsies. However, recent studies indicate CTCs assay cannot differentiate between patients with early-stage malignancy and people with no cancer and it has limited specificity as a screening tool [3, 4]. On the other hand, ctDNA has promised to be a sensitive and specific test for cancer screening [11, 12]. Still, ctDNA testing has several limitations for a screening platform compared with platelet RNA testing. First, the quantity of ctDNA is very limited even in cancer patients, not to mention in patients with early-stage cancer, while blood platelets are quite abundant. So the volume of blood needed in platelet testing is about 0.1 ml while ctDNA testing requires at least 10 ml. Second, the isolation and conversion process may cause damages to ctDNA, while platelet isolation procedure is simple and sample is stable and easy for storage. Third, ctDNA extraction requires an expensive kit while platelet isolation needs no expensive consumables. The subsequent sequencing analysis of ctDNA is also more expensive than platelet sequencing in our study. Furthermore, our study used LASSO regression to select the optimum gene-expression-signature for the prediction of cancer risk via quantitative real-time PCR. Hence our strategy with the prediction models including 4 or 5 biomarkers as variables is much more cost-effective than ctDNA testing for hundreds of hotspots. Fourth, ctDNA analysis could only detect frequently mutated genes in common cancers. The evolutionary and heterogeneity nature of cancer demands a large amount of possible mutations to be screened to achieve a consistent biomarker. Platelet biomarkers, on the other hand, are genes correlated with immune response and regulation according to our findings. Hence, platelet RNA testing may not be affected by cancer type or heterogeneity. Fifth, our platelet RNA prediction model could discriminate early-stage cancer from both healthy control and macroscopic tumor group, while biomarkers or screening models from previous studies often cannot distinguish samples from different stages of cancer. Thus platelet RNA testing may easily determine the best window for possible intervention. Last but not the least, platelet RNA testing described in our proof-of-concept study takes hours via qPCR while ctDNA testing

takes days or weeks via next-generation sequencing and require skilled biology and bioinformatics technicians. Hence platelet testing is much less time-consuming and requires less training of technicians. This demonstrates the potential of platelets as a non-invasive screening platform for the detection of occult cancer.

The sensitivity and specificity of our model could further improve by including more samples or increasing RNA quantities to avoid invalid qPCR results from low-abundant genes, or by employing machine learning of large sequencing data for validation. Furthermore, the eDEGs are mostly immune-related, not tumor-specific. Hence it is possible platelets-based liquid biopsy could enable simultaneous early detection of cancers from multiple organs of origin. Since it has been shown that platelet profiles are influenced by tumor type [20], it is feasible to add tumor type markers into the gene-panel to determine the origin of cancer. It would also be interesting to investigate platelet profiles in immunoediting animal models to further understand the role of platelets in cancer-immune interactions.

Conclusions

Combined, our study provides evidence for potential clinical relevance of blood platelets as a platform for liquid biopsy-based early detection of cancer.

Abbreviations

eDEGs: Differentially expressed genes in suboptimal inoculation group compared with optimal inoculation group and control group; qPCR: Quantitative real-time PCR; ROC curve: Receiver operating characteristic (ROC) curve; C group: Negative control group, mice injected with HBSS (Hank's Balanced Salt Solution); S group: Suboptimal inoculation group, mice inoculated with 2×10^3 B16F10 cells; O group: Optimal inoculation group, mice inoculated with 1×10^5 B16F10 cells; E group: Early-early group, mice with occult tumors 14 days post-injection; M group: Mice with macroscopic tumors 14 days post-injection.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-022-03268-z>.

Additional file 1: Fig. S1. Platelet RNA profiles of mice inoculated with B16F10 cells are consistent with previous studies. a, Images of mice after terminal blood collection. Representative images of mice in O group (optimal inoculation group, mice inoculated with 1×10^5 B16F10 cells) (left) and S group (suboptimal inoculation group, mice inoculated with 2×10^3 B16F10 cells) with HE-stained histological images of tissue from inoculation site (right). b, Platelet mRNA sequencing data of known platelet-abundant genes with a dashed line showing the average read count of our data. c, Pearson's correlation (color bar) matrix of our mRNA sequencing data of platelets and PBMCs (columns and rows). S: platelet suboptimal inoculation group; O: platelet optimal inoculation group; C: platelet negative control group. PS: PBMC suboptimal inoculation group; PO: PBMC optimal inoculation group; PC: PBMC negative control group. d, Heatmap of previously reported differentially expressed genes between platelets and PBMCs from our sequencing data. **Fig. S2.** PBMC RNA profiles of mice inoculated with an optimal or suboptimal number of B16F10

cells. a, Correlation plots of mRNAs detected in PBMCs of suboptimal inoculation group (S group, mice inoculated with 2×10^3 B16F10 cells), negative control group (C group, mice injected with HBSS) and optimal inoculation group (O group, mice inoculated with 1×10^5 B16F10 cells) mice, including highlighted increased (red) and decreased (blue) PBMC mRNAs. NRC, normalized read counts (mean of group). r value calculated from Pearson's correlation test. b, Venn diagram of differentially expressed genes from pairwise comparisons. c, Heatmap of hierarchical clustering of PBMC mRNA profiles of S group (beige), C group (green) and O group (orange). Data pooled from (a, b) $n = 5$ biologically independent experiments with $n = 24$ (S group) or $n = 25$ (C group) and $n = 20$ (O group) mice or data representing all 5 independent experiments (c). d, Top GO terms of pathway enrichment analysis of eligible eDEGs (strategy same as Fig. 3a, details in "Methods") in PBMCs with reference from KEGG pathways. Adjusted P value < 0.05 , Benjamini and Hochberg method. **Fig. S3.** An example of the selected models from Joinpoint multi-phase regression analyses. Tumor growth data from one mouse in early-early tumor group (E group). **Fig. S4.** qPCR verifications of the expression levels of selected 36 genes in the mouse cohort. Violin plots of normalized gene expression levels ($\Delta\Delta Ct(\text{Ref-Gene})$) of 36 selected genes in three groups of the mouse cohort. C: negative control; E: early-early tumor; M: macroscopic melanoma. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, Kruskal-Wallis test. Data without significance tags representing non-significant for analyses between three groups (h, n, u, ad, ae) (detailed statistics including n values and P values see Table S5). **Fig. S5.** LASSO regression model construction and variable selection for predicting occult tumor progression in mouse cohort. a, b, Ten-fold cross-validation for the selection of the penalty term λ with the binomial deviance as measures of the predictive performance of the fitted models. The dependent variable groups: E vs. C (a) or E vs. M (b). c, Coefficients derived from LASSO regression. E, early-early tumor (occult tumor that progressed into macroscopic tumor later); C, negative control group; M, macroscopic melanoma group. Numbers of samples included in LASSO regression: E, $n = 40$; C, $n = 50$; M, $n = 45$. Numbers of variables included in LASSO regression: E vs. C, $n = 29$; E vs. M, $n = 30$. The prediction score formulas for the discrimination of E group from C group (ScoreEC) or from M group (ScoreEM) established as follows: $\text{ScoreEC} = 6.337 - 0.156 \times (\text{CtCd19} - \text{CtRef}) - 0.345 \times (\text{CtCdkn1a} - \text{CtRef}) + 0.187 \times (\text{CtS100a9} - \text{CtRef}) - 1.582 \times (\text{CtTap1} - \text{CtRef}) + 0.615 \times (\text{CtTnfrsf1b} - \text{CtRef})$; $\text{ScoreEM} = 4.664 - 0.149 \times (\text{CtCcr7} - \text{CtRef}) - 0.900 \times (\text{CtCd4} - \text{CtRef}) + 0.326 \times (\text{CtKmt2d} - \text{CtRef}) - 0.313 \times (\text{CtLy6e} - \text{CtRef})$. **Fig. S6.** Ridge and elastic net regression model construction and variable selection for predicting occult tumor progression in mouse cohort. a, b, Ten-fold cross-validation for the selection of the penalty term λ with the binomial deviance as measures of the predictive performance of the fitted models for ridge regression (coefficients derived from LASSO regression see Table S6). The dependent variable groups: E vs. C (a) or E vs. M (b). ROC curves for the diagnostic performances of the prediction score formulas generated from ridge (c, d) and elastic net (f, g) regression in the mouse cohort. ROC curves for the discrimination of early-early tumor (occult tumor, E group) from negative control group (C group) (c, f, E vs. C) or from macroscopic melanoma group (M group) (d, g, E vs. M). The prediction score formulas for the discrimination of E group from C group (ScoreEC) or from M group (ScoreEM) established as follows: $\text{Score} = \text{Intercept} + \sum \text{Coefficient} \times (\text{CtVariable} - \text{CtRef})$. Probability statistics calculated according to the prediction score formulas generated from regression analyses: Probability = $e^{\text{Score}} / (1 + e^{\text{Score}})$. 95% CI of AUC: training data 0.879-0.997, testing data 0.806-1.000 (c); training data 0.891-1.000, testing data 0.822-1.000 (d); training data 0.861-0.994, testing data 0.807-1.000 (f); training data 0.887-0.995, testing data 0.901-1.000 (g). e, Elastic net regression's optimal alpha and lambda combination. E, early-early tumor (occult tumor that progressed into macroscopic tumor later); C, negative control group; M, macroscopic melanoma group. Numbers of samples included in regression: E, $n = 40$; C, $n = 50$; M, $n = 45$. Numbers of variables included in regression: E vs. C, $n = 29$; E vs. M, $n = 30$. **Table S1.** Enriched KEGG pathways of differentially expressed mRNAs in platelets of suboptimal inoculation group. **Table S2.** Enriched KEGG pathways of differentially expressed mRNAs in PBMCs of suboptimal inoculation group. **Table S3.** Joinpoint multi-phase regression statistics for all mice in early-early group of the mouse cohort.

Table S4. Primers for qPCR experiments. **Table S5.** Statistic analyses of the normalized expression levels of selected 36 genes in the mouse cohort.

Table S6. Coefficients derived from ridge regression. **Table S7.** Coefficients derived from elastic net regression. **Table S8.** Root mean square error (RMSE) analyses of ridge, lasso and elastic net regression.

Additional file 2: Table S9. Differentially expressed mRNAs in platelets of suboptimal inoculation group compared with optimal inoculation group and control group.

Additional file 3: Table S10. Differentially expressed mRNAs in PBMCs of suboptimal inoculation group compared with optimal inoculation group and control group.

Acknowledgements

A special thank you goes to Dr. Le Ma, Dr. Liang Bai, Dr. Bei Han, Dr. Jing Han, Dr. Jing Lin, Dr. Weiqing Wang, Yixue Li, Qiao Peng, Jianing Wu and Ruilei Cheng for their help in this study. The pre-print version of this article is present on <https://www.biorxiv.org/content/10.1101/2021.02.09.430530v1>. This article is not published nor is under publication elsewhere.

Authors' contributions

Y.Y. and J.R. designed the study and performed the experiments. L.Z., Y.N., F.J., J.H., L.J., S.L., W.L., L.X., Z.K., and C.D. participated in animal maintenance, animal blood sample collection and/or tumor cell inoculation. Y.Y. and S.M. performed data analyses and interpretation. Y.Y. wrote the manuscript. All authors read and approved the final manuscript.

Funding

Financial support was provided by Xinjiang Guan Run Logistics Co., Ltd (Funding No.: 20181224), J.R., the National Natural Science Foundation of China (Funding No.: 81602698), and the Science and Technology of Shaanxi Province (2019JQ-591).

Availability of data and materials

The datasets generated and analyzed during the current study are available in the GEO database. The accession number for the raw sequencing data reported in this study is GEO: GSE160650.

Declarations

Ethics approval and consent to participate

Animal experiments were approved by the Animal Ethics Committee of Xi'an Jiaotong University.

Consent for publication

Not applicable.

Competing interests

Y.Y., J.R. and S.M. are inventors of a patent (No. 202011118562.8) for the animal model used in this study and the application for this animal model in cancer research. J.R. partially funded this study.

Author details

¹Department of Histology and Embryology, School of Basic Medicine, Guilin Medical University, Guilin, Guangxi, China. ²School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an, China. ³Xi'an, China. ⁴China-Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an, Shaanxi 710061, People's Republic of China. ⁵National and Local Joint Engineering Research Center of Biodiagnosis and Biotherapy, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710004, Shaanxi, China. ⁶Xi'an People's Hospital (Xi'an Fourth Hospital), Xi'an, China. ⁷The First Affiliated Hospital of Medical School of Xi'an Jiaotong University, Xi'an, China. ⁸Institute of Basic Medical Sciences, Xi'an Medical University, Xi'an 710021, China.

Received: 18 September 2021 Accepted: 19 January 2022
Published online: 05 February 2022

References

- Luke JJ, Flaherty KT, Ribas A, Long GV. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol*. 2017;14:463–82.
- Gaebe K, Li AY, Das S. Clinical biomarkers for early identification of patients with intracranial metastatic disease. *Cancers (Basel)*. 2021. <https://doi.org/10.3390/cancers13235973>.
- Schiffman JD, Fisher PG, Gibbs P. Early detection of cancer: past, present, and future. *Am Soc Clin Oncol Educ Book*. 2015. https://doi.org/10.14694/EdBook_AM.2015.35.57.
- Franken B, de Groot MR, Mastboom WJ, Vermes I, van der Palen J, Tibbe AG, Terstappen LW. Circulating tumor cells, disease recurrence and survival in newly diagnosed breast cancer. *Breast Cancer Res*. 2012;14:R133.
- Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Lubner B, Alani RM, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6:224ra224.
- Newman AM, Bratman SV, To J, Wynne JF, Eclow NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20:548–54.
- Guida F, Sun N, Bantiss LE, Muller DC, Li P, Taguchi A, Dhillon D, Kundnani DL, Patel NJ, Yan Q, et al. Assessment of lung cancer risk on the basis of a biomarker panel of circulating proteins. *JAMA Oncol*. 2018;4: e182078.
- Anderson GL, McIntosh M, Wu L, Barnett M, Goodman G, Thorpe JD, Bergan L, Thornquist MD, Scholler N, Kim N, et al. Assessing lead time of selected ovarian cancer biomarkers: a nested case-control study. *J Natl Cancer Inst*. 2010;102:26–38.
- Mao J, Ladd J, Gad E, Rastetter L, Johnson MM, Marzbani E, Childs JS, Lu H, Dang Y, Broussard E, et al. Mining the pre-diagnostic antibody repertoire of TgMMTV-neu mice to identify autoantibodies useful for the early detection of human breast cancer. *J Transl Med*. 2014;12:121.
- Udgata S, Takenaka N, Bamlet WR, Oberg AL, Yee SS, Carpenter EL, Herman D, Kim J, Petersen GM, Zaret KS. THBS2/CA19–9 detecting pancreatic ductal adenocarcinoma at diagnosis underperforms in pre-diagnostic detection: Implications for biomarker advancement. *Cancer Prev Res (Phila)*. 2020. <https://doi.org/10.1158/1940-6207.CAPR-20-0403>.
- Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, Barda N, Zuzarte PC, Heisler L, Sundaravadanam Y, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. 2018;559:400–4.
- Chen X, Gole J, Gore A, He Q, Lu M, Min J, Yuan Z, Yang X, Jiang Y, Zhang T, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat Commun*. 2020;11:3475.
- George JN. Platelets. *Lancet*. 2000;355:1531–9.
- Leslie M. Cell biology. Beyond clotting: the powers of platelets. *Science*. 2010;328:562–4.
- Schlesinger M. Role of platelets and platelet receptors in cancer metastasis. *J Hematol Oncol*. 2018;11:125.
- Haemmerle M, Stone RL, Menter DG, Afshar-Kharghan V, Sood AK. The platelet lifeline to cancer: challenges and opportunities. *Cancer Cell*. 2018;33:965–83.
- Calverley DC, Phang TL, Choudhury QG, Gao B, Oton AB, Weyant MJ, Geraci MW. Significant downregulation of platelet gene expression in metastatic lung cancer. *Clin Transl Sci*. 2010;3:227–32.
- McAllister SS, Weinberg RA. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat Cell Biol*. 2014;16:717–27.
- Nilsson RJ, Balaj L, Hulleman E, van Rijn S, Pegtel DM, Walraven M, Widmark A, Gerritsen WR, Verheul HM, Vandertop WP, et al. Blood platelets contain tumor-derived RNA biomarkers. *Blood*. 2011;118:3680–3.
- Best MG, Sol N, Kooi I, Tannous J, Westerman BA, Rustenburg F, Schellen P, Verschueren H, Post E, Koster J, et al. RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell*. 2015;28:666–76.
- Tomayko MM, Reynolds CP. Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother Pharmacol*. 1989;24:148–54.
- Im JH, Muschel RJ. Protocol for murine/mouse platelets isolation and their reintroduction in vivo. *Bio-Protoc*. 2017;7: e2132.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Team RC: R: a Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
- Kolde R. pheatmap: Pretty Heatmaps. 2019.
- Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
- Wei T, Simko V. R package "corrplot": visualization of a correlation matrix. 2017.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
- Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015;31:608–9.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607–13.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504.
- Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform*. 2003;4:2.
- Alhasan AA, Izuogu OG, Al-Balool HH, Steyn JS, Evans A, Colzani M, Ghevaert C, Mountford JC, Marenah L, Elliott DJ, et al. Circular RNA enrichment in platelets is a signature of transcriptome degradation. *Blood*. 2016;127:e1–11.
- Zsori KS, Muszbek L, Csiki Z, Shemirani AH. Validation of reference genes for the determination of platelet transcript level in healthy individuals and in patients with the history of myocardial infarction. *Int J Mol Sci*. 2013;14:3456–66.
- Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*. 2000;19:335–51.
- Naumov GN, Akslen LA, Folkman J. Role of angiogenesis in human tumor dormancy: animal models of the angiogenic switch. *Cell Cycle*. 2006;5:1779–87.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
- Kuhn M. *caret: Classification and regression training*. Astrophysics Source Code Library; 2015.
- Wickham H, Mara A, Bryan J, Chang W, McGowan L, François R, Grol-emund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the tidyverse. *J Open Sour Softw*. 2019;4:1686.
- Park SL, Buzzai A, Rautela J, Hor JL, Hochheiser K, Efferm M, McBain N, Wagner T, Edwards J, McConville R, et al. Tissue-resident memory CD8(+) T cells promote melanoma-immune equilibrium in skin. *Nature*. 2019;565:366–71.
- Denapoli PMA, Zanetti BF, Dos Santos AA, de Moraes JZ, Han SW. Preventive DNA vaccination against CEA-expressing tumors with anti-idiotypic scFv6.C4 DNA in CEA-expressing transgenic mice. *Cancer Immunol Immunother*. 2017;66:333–42.
- Stone RL, Nick AM, McNeish IA, Balkwill F, Han HD, Bottsford-Miller J, Rupaimoole R, Armaiz-Pena GN, Pecot CV, Coward J, et al. Paraneoplastic thrombocytosis in ovarian cancer. *N Engl J Med*. 2012;366:610–8.
- Kissopoulou A, Jonasson J, Lindahl TL, Osman A. Next generation sequencing analysis of human platelet PolyA+ mRNAs and rRNA-depleted total RNA. *PLoS ONE*. 2013;8: e81809.
- Best MG, Wesseling P, Wurdinger T. Tumor-educated platelets as a noninvasive biomarker source for cancer detection and progression monitoring. *Cancer Res*. 2018;78:3407–12.

46. Best MG, Sol N, In 't Veld S, Vancura A, Muller M, Niemeijer AN, Fejes AV, Tjon Kon Fat LA, Huis In 't Veld AE, Leurs C, et al. Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell*. 2017;32:238–252239.
47. Abramson JS. Anti-CD19 CAR T-cell therapy for B-cell non-Hodgkin lymphoma. *Transfus Med Rev*. 2020;34:29–33.
48. Li X, Ding Y, Zi M, Sun L, Zhang W, Chen S, Xu Y. CD19, from bench to bedside. *Immunol Lett*. 2017;183:86–95.
49. Kreis NN, Louwen F, Yuan J. The multifaceted p21 (Cip1/Waf1/CDKN1A) in cell differentiation, migration and cancer therapy. *Cancers (Basel)*. 2019. <https://doi.org/10.3390/cancers11091220>.
50. Shabani F, Farasat A, Mahdavi M, Gheibi N. Calprotectin (S100A8/S100A9): a key protein between inflammation and cancer. *Inflamm Res*. 2018;67:801–12.
51. Henle AM, Nassar A, Puglisi-Knutson D, Youssef B, Knutson KL. Downregulation of TAP1 and TAP2 in early stage breast cancer. *PLoS ONE*. 2017;12:e0187323.
52. Ling A, Löfgren-Burström A, Larsson P, Li X, Wikberg ML, Öberg Å, Stenling R, Edin S, Palmqvist R. TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. *Oncoimmunology*. 2017;6:e1356143.
53. Tabassum A, Samdani MN, Dhali TC, Alam R, Ahammad F, Samad A, Karpinski TM. Transporter associated with antigen processing 1 (TAP1) expression and prognostic analysis in breast, lung, liver, and ovarian cancer. *J Mol Med (Berl)*. 2021;99:1293–309.
54. So T, Ishii N. The TNF-TNFR family of co-signal molecules. *Adv Exp Med Biol*. 2019;1189:53–84.
55. Korbecki J, Grochans S, Gutowska I, Barczak K, Baranowska-Bosiacka I. CC chemokines in a tumor: a review of pro-cancer and anti-cancer properties of receptors CCR5, CCR6, CCR7, CCR8, CCR9, and CCR10 ligands. *Int J Mol Sci*. 2020. <https://doi.org/10.3390/ijms21207619>.
56. Borst J, Ahrends T, Bąbala N, Melief CJM, Kastenmüller W. CD4(+) T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol*. 2018;18:635–47.
57. Fagan RJ, Dingwall AK. COMPASS ascending: emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. *Cancer Lett*. 2019;458:56–65.
58. Upadhyay G. Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells. *Front Immunol*. 2019;10:819.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

