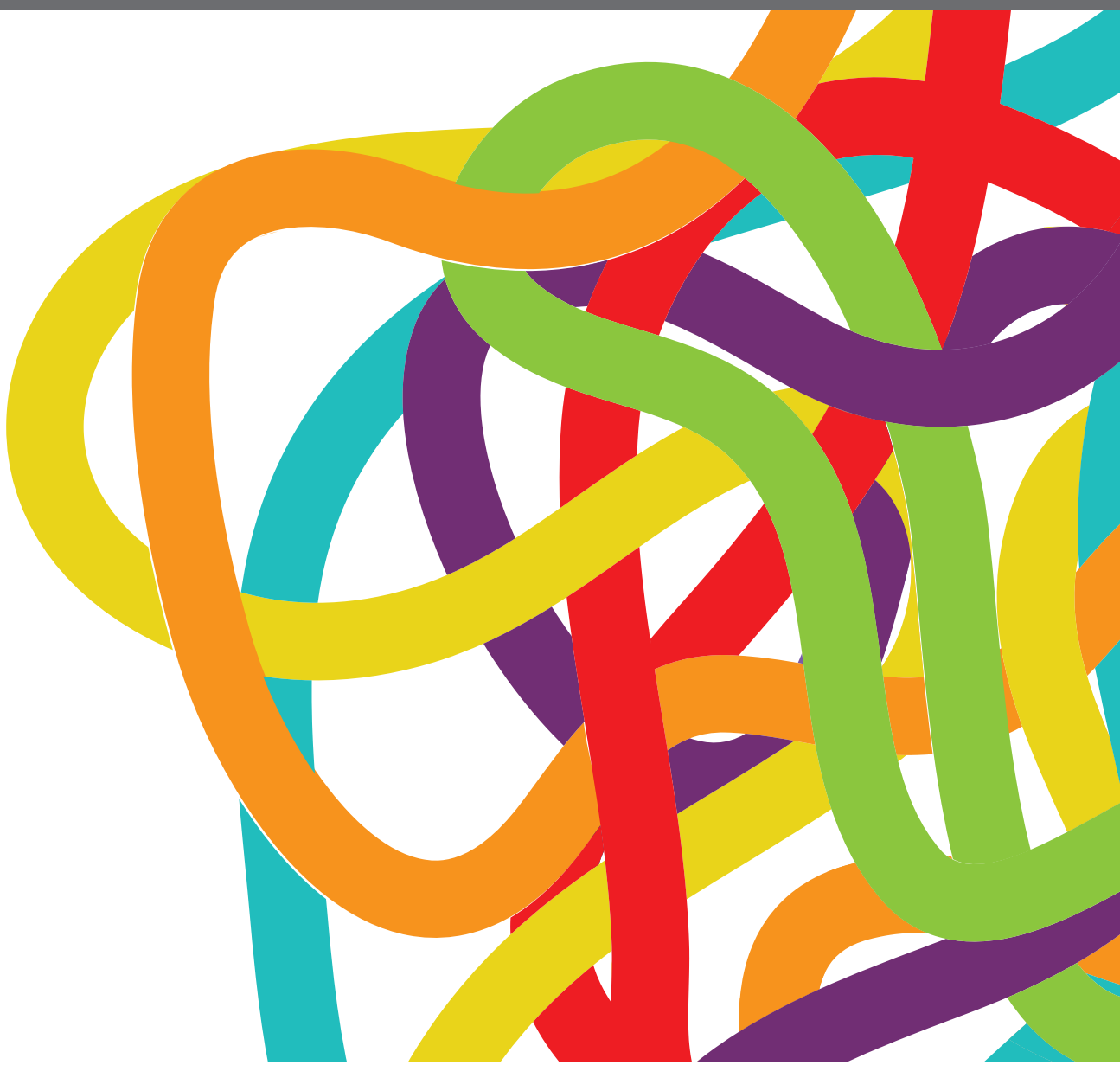
NEXT GENERATION SEQUENCING BASED DIAGNOSTIC APPROACHES IN CLINICAL ONCOLOGY

EDITED BY: Anton A. Buzdin, Ira Ida Skvortsova, Xinmin Li and Ye Wang
PUBLISHED IN: Frontiers in Oncology and Frontiers in Genetics



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-575-4
DOI 10.3389/978-2-88966-575-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

NEXT GENERATION SEQUENCING BASED DIAGNOSTIC APPROACHES IN CLINICAL ONCOLOGY

Topic Editors:
Anton A. Buzdin, I.M. Sechenov First Moscow State Medical University, Russia
Ira Ida Skvortsova, Innsbruck Medical University, Austria
Xinmin Li, University of California, Los Angeles, United States
Ye Wang, Qingdao University Medical College, China

Dr. Anton Buzdin (AB) is employed by Omicsway Corp. (USA). AB received grants from Amazon and Microsoft Azure to support cloud computations.

Dr. Xinmin Li is director of JCCC Shared Genomics Resource, the University of California, Los Angeles, CA

Dr. Ye Wang is Director of Gene testing Department (Core Lab) of Qingdao Central Hospital, the Second Affiliated Hospital of Qingdao University

Citation: Buzdin, A. A., Skvortsova, I. I., Li, X., Wang, Y., eds. (2021). Next Generation Sequencing Based Diagnostic Approaches in Clinical Oncology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-575-4

Table of Contents

frontiers in Oncology

EDITORIAL
published: 28 January 2021
doi: 10.3389/fonc.2020.635555

Editorial: Next Generation Sequencing Based Diagnostic Approaches in Clinical Oncology

Anton Buzdin[1,2,3,4*], *Ira Ida Skvortsova*[5,6,7], *Xinmin Li*[8] *and Ye Wang*[9]

[1] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia, [2] World-Class Research Center "Digital Biodesign and Personalized Healthcare", Sechenov First Moscow State Medical University, Moscow, Russia, [3] Translational Genome Bioinformatics Laboratory, Moscow Institute of Physics and Technology (National Research University), Moscow, Russia, [4] Research Department, OmicsWay Corp., Walnut, CA, United States, [5] Therapeutic Radiology and Oncology, Medical University of Innsbruck, Innsbruck, Austria, [6] Group for Experimental and Translational Radiooncology, Tyrolean Cancer Research Institute, Innsbruck, Austria, [7] PathoBiology Group, European Organization for Research and Treatment of Cancer (EORTC), Brussels, Belgium, [8] Department of Pathology & Laboratory Medicine, University of California Los Angeles (UCLA) Technology Center for Genomics & Bioinformatics, Los Angeles, CA, United States, [9] Clinical Laboratory, Qingdao Central Hospital, The Second Affiliated Hospital of Medical College of Qingdao University, Qingdao, China

Keywords: cancer diagnostics, genomics, transcriptomics, bioinformatics, biomarkers, theranostics, next generation sequencing, RNA sequencing

OPEN ACCESS

Edited and reviewed by:
Claudio Sette,
Catholic University of the Sacred Heart, Rome, Italy

***Correspondence:**
Anton Buzdin
buzdin@oncobox.com

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *30 November 2020*
Accepted: *14 December 2020*
Published: *28 January 2021*

Citation:
Buzdin A, Skvortsova II, Li X and Wang Y (2021) Editorial: Next Generation Sequencing Based Diagnostic Approaches in Clinical Oncology. Front. Oncol. 10:635555. doi: 10.3389/fonc.2020.635555

Editorial on the Research Topic

Next Generation Sequencing Based Diagnostic Approaches in Clinical Oncology

Next Generation Sequencing (NGS) technologies transformed cancer genetics by providing unprecedented access to big genomic and transcriptomic data (1, 2). Curation of cancer genetic profiles has led to many successful applications crosslinking mutations or polymorphisms with individual tumor response to therapies. Obviously valuable, these approaches however still cannot generate clinically actionable information for most of the cancer patients. In turn, high throughput RNA (transcriptome) analysis can be considered a rising star that may complement mutational screenings (1, 3). A combination of both approaches may be synergistic for many tasks in molecular diagnostics in oncology (4–6). Cancer mutation and expression biomarkers can not only help to set a diagnosis but also to identify appropriate personalized molecular-based treatment (7–9). Furthermore, novel bioinformatic approaches enable squeezing more and more clinically meaningful data from large genetic datasets (10).

Wang et al. reviewed current progress in using RNA sequencing (RNAseq) methods in cancer research, prognosis, and molecular diagnostics. The techniques of RNAseq have progressed rapidly from bulk RNAseq, laser-captured micro-dissected (LCM) RNAseq, and single-cell (SS) RNAseq to digital spatial RNA profiling, spatial transcriptomics, and direct *in situ* sequencing. These different technologies have their unique strengths, weaknesses, and suitable applications in the field of clinical oncology. For example, bulk RNAseq is a cost-effective mature technology that can be used for confident whole-transcriptome biomarker recovery and targeted capture of gene fusions. LCM-RNAseq interrogates tissue heterogeneity by dissecting cell type specific populations, whereas SS-RNAseq can characterize different cell types inhabiting and surrounding tumors. Zhigalova et al.

applied RNAseq to characterize T-cell receptor (TCR) and immunoglobulin repertoire in an HKP1 $Kras^{G12D}p53^{-/-}$ syngeneic mouse model of lung cancer after anti-PD-1 treatment, and observed decreased TCR diversity in response to therapy. However, repertoire diversity was then restored in progressing disease but remained decreased in good responders to therapy in both CD4+ and CD8+ subsets, thus suggesting their potential usefulness as prognostic biomarkers.

Borger et al. used RNAseq and bioinformatic analysis with Oncofinder (11) for characterization of molecular processes accompanying ALPPS, a novel two-staged hepatectomy that dramatically accelerates liver regeneration and enables extensive liver tumor resection. ALPPS specific signature included activation of cell survival branch of IGF1R signaling pathway, proliferation branch of ILK Pathway, and the IL-10 Pathway, whereas the transcriptional branch of the Interferon pathway was downregulated ($p < 0.05$). The PAK- and ILK-associated pathways were also activated at an earlier time point, reflecting acceleration of liver regeneration ($p < 0.001$). Molecular therapies influencing those pathways could potentially improve the performance of ALLPS in the future.

Poddubskaya et al. used bulk transcriptomics to identify effective experimental drug combination for an advanced recurrent *ALK*-positive lung cancer patient using Oncobox algorithm (12) based on ranking of tumor-upregulated drug target genes. A 48 y.o. male patient received personalized off-label combination of crizotinib + bevacizumab + docetaxel that led to additional stabilization for 22 months. The patient survival after developing resistance to ALK inhibitor was longer for 16 months than previously reported average survival for such cases, thus evidencing effectiveness of RNA-guided prescription of cancer therapies in terms of survival and quality of life.

By experimentally *in vitro* validating meta-analysis of RNAseq data for esophageal squamous cell carcinoma (ESCC), Tang et al. identified *LINC01614* non-coding (NC) RNA that was a poor prognosis biomarker, being upregulated in ESCCs. The knockdown of *LINC01614* expression significantly inhibited the migration of ESCC cells by restricting EMT. He et al. found that another NC RNA, circular RNA *hsa_circ_0007843*, in colon cancer SW480 cells can serve as a molecular sponge by interacting with the microRNA miR-518c-5p that interacts with matrix metallopeptidase 2 (MMP2). Overexpression of *hsa_circ_0007843* promoted tumor cell growth, invasion, and migration, whereas its downregulation had opposite effects. The possible explanation for those effects is removing the miR-518c-5p inhibitory block on MMP2 translation.

Finally, Borisov and Buzdin concentrate on the phenomenon that machine learning (ML) methods that had obviously revolutionized many fields are still poorly applicable to molecular oncology. Since plausible explanation is the small size of related datasets, the authors consider using dynamic data trimming (13) to filter for more relevant and informative feature sets to apply ML for finding robust transcriptomic biomarkers using available clinically annotated datasets, e.g. (14).

Conversely, Matrone et al. reported application of cancer DNA sequencing to identify the primary origin of metastases in the patient's thyroid gland in case of concomitant lung adenocarcinoma. The patient was diagnosed with both poorly differentiated thyroid cancer for the huge involvement of the neck and concomitant lung adenocarcinoma. Clinical features, imaging evaluation and available tumor markers couldn't support a well-defined diagnosis, and the histologic features of the thyroid and lung biopsies confirmed the figure of two different tumors. However, the NGS analysis showed a G12C mutation in *KRAS* gene in both tissues, which is highly prevalent in lung but not thyroid cancers. Therefore, the lung origin of the disease was deduced, and the patient was addressed to the appropriate therapeutic strategy.

Yuan et al. explored *ERBB2* exon 20 insertion that is a refractory oncogenic driver, by panel NGS of 59 or 1,021 genes, in 112 lung cancer patients, including 18 patients receiving afatinib treatment. There were 66% of patients having *TP53* co-mutation, and *FOXA1* was the most prevalent co-amplified gene (in 5.5% of the cases). Patients with co-occurring *TP53* mutation showed approximately twice shorter overall survival (OS): median OS of 14.5 *versus* 30.3 months, $p = 0.04$. *ERBB2* exon 20 insertion also related to shorter progression-free survival (PFS): median PFS of 1.2 *versus* 4.3 months, $p < 0.05$.

In a subgroup of 44 cases with *ALK* gene fusions out of total 1349 lung cancer patients, Liu et al. identified *ALK* fusion partners using a customized NGS panel. The most common partner was well-known gene *EML4*, but also a new *ALK* fusion partner *HMBOX1* was identified. The copy number alterations were found in ~30% of the cases, and the most commonly amplified genes were *MDM2* and *TERT*.

Furthermore, Belardinilli et al. used target gene panel sequencing to characterize 639 formalin-fixed paraffin-embedded (FFPE) metastatic colorectal cancer (mCRC) specimens and identified blocks of significantly cooccurring mutations. This led to a novel stratification of mCRC patients in eight groups characterized by specific mutational association patterns, which was validated on a literature dataset of The Cancer Genome Atlas (TCGA) project.

By performing whole-genome sequencing, Gao et al. investigated genetic alterations in spinal schwannoma and identified mutations in *ATM, CHD4, FAT1, KMT2D, MED12, NF2*, and *SUFU* genes, and homozygous deletion was observed in *NF1, NF2*, and *CDKN2C*. Hippo signaling pathway was most significantly affected by the mutations identified. Shahid et al. used 54-gene panel for target DNA sequencing of 26 acute myeloid leukemia (AML) samples by using ultra-deep NGS with ~5,000-coverage. Novel somatic mutations were identified, including those repeated in several genetically unrelated cancers, *i.e. STAG2* L526F and *BCORL1* A400V.

Of note, Nikitin et al. showed that Lynch Syndrome-related germline mutations (LS-mutations) in DNA mismatch repair pathway genes are statistically significantly associated with breast cancer (BC). To this end, they performed Targeted NGS of genes *MLH1, MSH2, MSH6, EPCAM*, and *PMS2* in a cohort of 492 healthy donors, 711 patients with hereditary BC, and 60 patients with sporadic BC. ~10% patients with hereditary BC had at least

one germline mutation, and ~5% had predicted pathogenic mutations in these genes.

Gao et al. explored in detail the influence of sample preparation method on the results of target gene panel NGS. A 22-gene panel with 103 hotspots was used to detect mutations in paired FFPE and fresh frozen tissue specimens from 118 patients with colorectal cancer. They found that 99% of the patients one or more detectable variants, with 226 variants in FFPE and 221 in fresh tissue. Of the totally 129 individual variants identified, 96 variants were common for both FFPE and fresh biosamples, 27 were specific for FFPE, and 6 for fresh tissue specimens. These findings suggest that when there is a choice, fresh frozen specimens should be most probably considered as the preferred type of biosamples because of a number (27/129) of apparently artifact variants identified specific for the FFPE tissue samples.

Another technological aspect of NGS was investigated by Petrackova et al. who used statistical tests to assess the sequencing coverage thresholds for a robust variant calling, and published a coverage calculator. Using the sequencing error only, the recommended minimum depth of coverage was 1,650 along with a threshold of no less than 30 variant reads for a targeted clinical NGS mutation analysis. This points to an important problem of data quality standardization in clinical NGS. However, in practice this threshold can be only applied to target panel gene sequencing methods, whereas it is so far unrealistic for the whole-genome or even exome sequencing approaches.

Finally, epigenetic aspects of using NGS in oncology were considered by Jovčevska who reviewed current progress in sequencing and machine learning applications for the analysis of DNA methylation in glioblastoma.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

AB contribution was supported by the Russian Science Foundation grant 18-15-00061. YW was sponsored by the National Natural Science Foundation of China (No. 81670822), and Qingdao Key Research Project (No. 19-6-1-3-nsh).

REFERENCES

1. Buzdin A, Sorokin M, Garazha A, Glusker A, Aleshin A, Poddubskaya E, et al. RNA sequencing for research and diagnostics in clinical oncology. *Semin Cancer Biol* (2020) 60:311–23. doi: 10.1016/j.semcancer.2019.07.010
2. Buzdin A, Sorokin M, Garazha A, Sekacheva M, Kim E, Zhukov N, et al. Molecular pathway activation – New type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin Cancer Biol* (2018) 53:110–24. doi: 10.1016/j.semcancer.2018.06.003
3. Rodon J, Soria JC, Berger R, Miller WH, Rubin E, Kugel A, et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat Med* (2019) 25:751–8. doi: 10.1038/s41591-019-0424-4
4. Negro G, Aschenbrenner B, Brezar SK, Cemazar M, Coer A, Gasljevic G, et al. Molecular heterogeneity in breast carcinoma cells with increased invasive capacities. *Radiol Oncol* (2020) 54:103–18. doi: 10.2478/raon-2020-0007
5. Skvortsova I. Special Issue "Enigmatic tumor properties associated with metastatic spread" seminars in cancer biology, volume XX. *Semin Cancer Biol* (2020) 60:iii–iv. doi: 10.1016/j.semcancer.2019.10.021
6. Skvortsova II, Span PN. Editorial: Advances in Biological Understanding of Tumor Radiation Resistance. *Front Oncol* (2020) 10:754. doi: 10.3389/fonc.2020.00754
7. Suntsova M, Gaifullin N, Allina D, Reshetun A, Li X, Mendeleeva L, et al. Atlas of RNA sequencing profiles for normal human tissues. *Sci Data* (2019) 6:36. doi: 10.1038/s41597-019-0043-4
8. Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R. Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer Treat Rev* (2020) 86(102019):1–11. doi: 10.1016/j.ctrv.2020.102019
9. Moisseev A, Albert E, Lubarsky D, Schroeder D, Clark J. Transcriptomic and genomic testing to guide individualized treatment in chemoresistant gastric cancer case. *Biomedicines* (2020) 8(3):67. doi: 10.3390/biomedicines8030067
10. Buzdin A, Sorokin M, Poddubskaya E, Borisov N. High-Throughput Mutation Data Now Complement Transcriptomic Profiling: Advances in Molecular Pathway Activation Analysis Approach in Cancer Biology. *Cancer Inform* (2019) 18:1176935119838844. doi: 10.1177/1176935119838844.
11. Buzdin AA, Zhavoronkov AA, Korzinkin MB, Venkova LS, Zenin AA, Smirnov PY, et al. Oncofinder, a new method for the analysis of intracellular signaling pathway activation using transcriptomic data. *Front Genet* (2014) 5:55:55. doi: 10.3389/fgene.2014.00055
12. Tkachev V, Sorokin M, Garazha A, Borisov N, Buzdin A. Oncobox Method for Scoring Efficiencies of Anticancer Drugs Based on Gene Expression Data. In: *Methods in molecular biology (Clifton, N.J.)*. (2020) p. 235–55. doi: 10.1007/978-1-0716-0138-9_17
13. Tkachev V, Sorokin M, Borisov C, Garazha A, Buzdin A, Borisov N. Flexible data trimming improves performance of global machine learning methods in omics- based personalized oncology. *Int J Mol Sci* (2020) 21(3):713. doi: 10.3390/ijms21030713
14. Borisov N, Sorokin M, Tkachev V, Garazha A, Buzdin A. Cancer gene expression profiles associated with clinical outcomes to chemotherapy treatments. *BMC Med Genomics* (2020) 13:111. doi: 10.1186/s12920-020-00759-0

Conflict of Interest: Author AB was employed by the company OmicsWay Corp.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

OPINION
published: 17 July 2019
doi: 10.3389/fonc.2019.00658

New Paradigm of Machine Learning (ML) in Personalized Oncology: Data Trimming for Squeezing More Biomarkers From Clinical Datasets

Nicolas Borisov[1]* *and Anton Buzdin*[1,2,3]

[1] Department of Personalized Medicine, I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia, [2] Department of Genomics and Postgenomic Technologies, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, [3] Department of Bioinformatics and Molecular Networks, OmicsWay Corporation, Walnut, CA, United States

Keywords: bioinformatics, personalized medicine, oncology, P vs. N problem, machine learning

OPEN ACCESS

Edited by:
Steven M. Lipkin, Weill Cornell Medicine, Cornell University, United States

Reviewed by:
Shicheng Guo, Marshfield Clinic Research Institute, United States
Parvin Mehdipour, Tehran University of Medical Sciences, Iran
Michele Milella, University of Verona, Italy

***Correspondence:**
Nicolas Borisov
borisov@oncobox.com

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *22 May 2019*
Accepted: *05 July 2019*
Published: *17 July 2019*

Citation:
Borisov N and Buzdin A (2019) New Paradigm of Machine Learning (ML) in Personalized Oncology: Data Trimming for Squeezing More Biomarkers From Clinical Datasets. Front. Oncol. 9:658. doi: 10.3389/fonc.2019.00658

INTRODUCTION

Personalized medicine has a huge potential of transforming healthcare standards when selection of therapies according to standard guidelines often fails, which can be the case in oncology (1, 2), endocrinology (3, 4), neurology (3), treatment of infectious diseases (5, 6) and hemostatic disorders (7, 8). Nowadays, personalized approach can be based on a solid fundament of big biomedical data obtained for an individual patient, analyzed vs. comparable datasets for other individual cases with known clinical outcome. This can help, for example, developing new criteria for predicting response of a cancer patient to a certain treatment.

The analysis of Big Data in oncology can benefit significantly from being empowered by *machine learning* (ML) techniques (9–13) tailored for solving this *"P* vs. *N"* problem. ML is usually defined as the study of algorithmically-built mathematical models that have been fitted for the portion of data called the *training dataset*, to make predictions for the similarly-obtained and similarly structured data called the *test* or *validation dataset*. Major principles of ML have been formulated more than half a century ago and transformed methodology in many areas such as engineering, physics, banking, defense, agriculture, and meteorology (11, 14). Efficiencies of ML-based predictor/classifier models are described by specific quality metrics such as sensitivity (Sn), specificity (Sp), area under ROC curve (AUC), accuracy rate (ACC), Matthews correlation coefficient (MCC), or by *p*-values from statistical tests distinguishing one class from another (15).

However, it was only in the beginning of XXI century when such ML on Big Data became possible in biomedicine, still not having a groundbreaking effect (16). This delay is most probably due to relatively recent emergence of experimental methods generating big biomedical data connected with the sufficiently developed IT infrastructure. Among those game-changing experimental methods the major role was played by next-generation sequencing (NGS) and novel mass-spectrometry approaches which enabled performing whole genome-, transcriptome-, proteome-, and metabolome analyses relatively fast and cheap (17–19), see **Figure 1A**. This allowed to feed ML methods with big biomedical data thus generating beneficial outputs, also in the field of clinical medicine. For example, over 150 scientific papers have been indexed in the PubMed repository during last 24 months mentioning *machine learning* and *drug sensitivity*[1].

Here we will focus on applying ML for personalized medicine, primarily oncology, dealing with attempts to generate as much as possible treatment response biomarkers from mediocre datasets. From the point of view of classical ML approaches, most if not all of the available clinical genetic

[1]This is the result of a PubMed query: https://www.ncbi.nlm.nih.gov/pubmed/?term=machine$+$learning$+$drug$+$sensitivity

datasets are insufficient for solving the task of differentiating, e.g., treatment responders from non-responders (9, 20). Numbers of features measured by NGS (e.g., mutations or gene expression values) are far greater than numbers of individual patients with traced clinical outcomes involved in each respective dataset. To generate statistically significant predictions, this requires extensive reduction of a pool of features to be considered, to make their number not exceeding the number of individuals analyzed (16). To increase the number of individuals, the datasets can be merged using cross-dataset harmonization. Different methods can be used to harmonize data obtained using the same (21, 22) or different experimental platforms (23, 24), or even using multiple platforms (25) (**Figure 1**).

ML INPUT DATA AND WORKFLOW

For ML applications dealing with prediction of patient's individual response(s) on drugs and different treatment

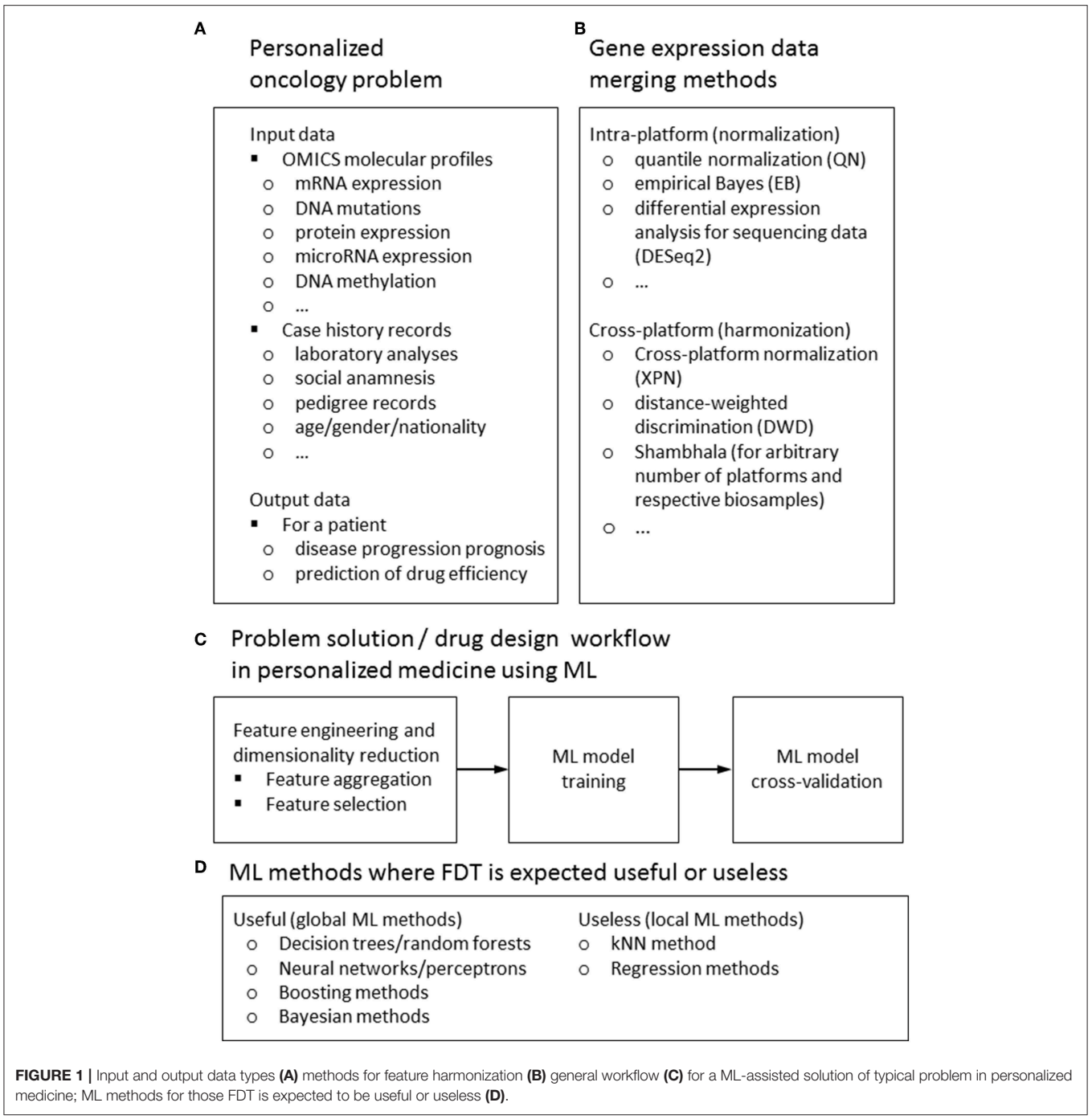


FIGURE 1 | Input and output data types **(A)** methods for feature harmonization **(B)** general workflow **(C)** for a ML-assisted solution of typical problem in personalized medicine; ML methods for those FDT is expected to be useful or useless **(D)**.

regimens, two types of data are most frequently used (**Figure 1A**):

1. Various multi-omics data, i.e., mRNA, microRNA, and protein expression levels, mutations in genomic DNA and epigenetic profiles (primarily DNA methylation) (26, 27). These data may be compared with the analogous types of data obtained on cell cultures in relation to sensitivity to therapeutics/treatment regimens, such as the data taken from the Broad Institute (28) and CancerRxGene (29) projects. These examples include, respectively, either changes of gene expression profiles influenced by the addition of drugs to cell culturing media, or gene expression and polymorphism/mutation profiles for many cell lines linked with their measured sensitivities to cancer drugs (30). These datasets are regarded plausible models for training ML drug sensitivity classifiers because they have thousands of individual "cases"—pairs cell culture/drug, each profiled in several replicates.
2. Alternatively, other types of data can be used including gender, age, results of clinical laboratory tests, functional diagnostics data (ECG, EEG etc.), risk factors, social anamnesis, and other electronic health records.

A typical workflow of ML drug sensitivity assay includes the following steps (**Figure 1C**):

1. *Data reduction, feature selection, and building on the training dataset*. Usually, in the collected raw data, the number of features (*NF*) exceeds the number of cases (*NC*), so that to provide a robust ML model, one must reduce the data to make the number of selected features (*NS*) lower than *NC* or at least comparable to it. This goal can be achieved in several ways. The raw data may be aggregated, e.g., in molecular pathways (2); or co-expressed/co-mutated clusters (31). Sometimes, the co-expression- and pathway topology-based analysis may be combined (32).
   Alternatively, they can be filtered according to specific functional of statistical traits (e.g., only the genes coding for tyrosine kinases are left; or genes with the highest abilities to discriminate responders from non-responders in training datasets) (33). The statistical methods for feature selection may include Pearson chi-squared test (34) or correlation test (34, 35). Other options are variance thresholding (VT), genetic algorithms (36), univariate feature selection (UFE), recursive feature elimination (RFE), principal component analysis (PCA) (35), CUR matrix (37) decomposition (27) and covariate regression (38).
2. *Applying ML algorithm*. The following methods may be used: support vector machines, SVM (2, 27, 39), *k* nearest neighbors, kNN (39), decision trees, DT (34, 39) or random forest, RF (39, 40). Alternatively, one can use artificial neural network, ANV (39), elastic net (41), back propagation networks (42), naïve Bayesian (27), logistic (27, 39), penalized (43), and lasso (43) regression models. In some cases, the hybrid global-local approaches, like combination of decision trees, random forests/SVM with kNN are used (2, 33, 39, 44, 45).
3. *Cross-validation and performance quality check*. The data obtained with the training dataset are then validated using independent validation dataset. For the cross-validation of machine learning methods, 5- or 10-fold cross validations are most commonly used. For datasets with smaller number of preceding cases (*NC*) the leave-one-out (LOO) scheme is preferable (2, 33, 43).

SHIFTING THE PARADIGM

The demonstrated performance of ML classifiers was high for problems like age recognition based on biochemical markers (41), but significantly lower for predictions of drug response in cancer patients (27, 46), with the exception of few reports based on very small patient cohorts (43).

A new paradigm recently emerged of considering flexible rather than fixed sets of features that are fitted individually for every comparison of a biosample with the pool of controls/training datasets (33). This can be done by means of *data trimming*[2]—sample-specific removal of features. The irrelevant features in a sample that don't have significant number of neighboring hits in the training dataset are removed from further analyses. In a pilot application for the SVM method of ML and high throughput gene expression data, this enabled to dramatically increase number and quality of biomarkers predicting responses to chemotherapy treatments for 10/10 cohorts of 46–235 cancer patients (33). Among them, in 3/10 cases basic ML applications were impossible to generate biomarkers of a sufficient quality.

The application of *flexible data trimming* (FDT) procedure prevents ML classifier from extrapolation by excluding non-informative features. Contrary to other complex data transfer techniques, this approach is heuristic, based on a common geometrical sense. For each point of a validation dataset, it takes into account only the proximal points of the training dataset. Thus, for every point of a validation dataset, the training dataset is adjusted to form a floating window. That was why we called (33) our FDT method *FLOating Window Projective Separator* (*FloWPS*).

DISCUSSION

Certainly, FloWPS is not the only possible method of data reduction for ML in oncology. In the pilot study, a simple PCA-based alternative was tried, which was less successful (33).

One of the major limitations of FloWPS is that it can be time-consuming at the level of optimization of data trimming parameters. The required computational time for such optimization grows cubically with the number of preceding cases in the training dataset. For example, for a 31 Gb RAM and 8·4.20 GHz CPUs computer running the Python FloWPS code for a dataset of 46 samples (33) takes ∼20 s, whereas for a bigger dataset of 235 samples (33) it requires already few hours.

[2]Data trimming is the process of removing or excluding extreme values, or outliers, from a dataset (47).

SVM is one of the most popular methods of ML nowadays (9, 48). However, using data trimming procedure has dramatically improved its performance for the task of classification cancer drug responders and non-responders. This means that it can be highly beneficial for the other ML methods as well. The FDT method simultaneously combines the advantages of both global (like SVM) and local (like kNN) methods of ML, and successfully acts even when purely local and global approaches fail. Due to its hybrid (global + local) nature, we expect that FloWPS may be also effective for other global ML methods such as decision trees/random forests, neural networks/multi-layer perceptrons, decision trees/random forests and boosting or Bayesian methods for ML, but may be useless for purely local approaches such as kNN or regression models (**Figure 1D**).

In its first published application, the data trimming could operate with high throughput gene expression or mutation profiles (33). However, it can be used for any type of Big Data in biomedicine, but not only. In this opinion paper, we speculate that this new concept has a potential to broadly introduce the use of ML in personalized oncology and, possibly, significantly expand its presence in many other fields.

AVAILABILITY OF CODE

The R package flowpspkg.tar.gz for FloWPS method and README manual are available at GitLab through the link: https://gitlab.com/borisov_oncobox/flowpspkg.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This publication was supported by the Russian Science Foundation grant no. 18-15-00061.

ACKNOWLEDGMENTS

The authors wish to acknowledge Victor Tkachev (Omicsway Corp.) for insightful discussion on perspectives of ML for biomedicine and Constantin Borisov (National Research University—Higher School of Economics, Moscow, Russia) for preparing the code for R package flowpspkg.tar.gz.

REFERENCES

1. Artemov A, Aliper A, Korzinkin M, Lezhnina K, Jellen L, Zhukov N, et al. A method for predicting target drug efficiency in cancer based on the analysis of signaling pathway activation. *Oncotarget.* (2015) 6:29347–56. doi: 10.18632/oncotarget.5119
2. Borisov N, Tkachev V, Suntsova M, Kovalchuk O, Zhavoronkov A, Muchnik I, et al. A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency. *Cell Cycle.* (2018) 17:486–91. doi: 10.1080/15384101.2017.1417706
3. Hokama M, Oka S, Leon J, Ninomiya T, Honda H, Sasaki K, et al. Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb Cortex.* (2014) 24:2476–88. doi: 10.1093/cercor/bht101
4. Kaizer EC, Glaser CL, Chaussabel D, Banchereau J, Pascual V, White PC. Gene expression in peripheral blood mononuclear cells from children with diabetes. *J Clin Endocrinol Metab.* (2007) 92:3705–11. doi: 10.1210/jc.2007-0979
5. De Maio F, Squeglia F, Goletti D, Delogu G. The mycobacterial HBHA protein: a promising biomarker for tuberculosis. *Curr Med Chem.* (2019) 26:2051–60. doi: 10.2174/0929867325666181029165805
6. Sanchez-Schmitz G, Stevens CR, Bettencourt IA, Flynn PJ, Schmitz-Abe K, Metser G, et al. Microphysiologic human tissue constructs reproduce autologous age-specific BCG and HBV primary immunization *in vitro*. *Front Immunol.* (2018) 9:2634. doi: 10.3389/fimmu.2018.02634
7. Lewis DA, Stashenko GJ, Akay OM, Price LI, Owzar K, Ginsburg GS, et al. Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thromb Res.* (2011) 128:536–40. doi: 10.1016/j.thromres.2011.06.003
8. Lewis DA, Suchindran S, Beckman MG, Hooper WC, Grant AM, Heit JA, et al. Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism. *Thromb Res.* (2015) 135:659–65. doi: 10.1016/j.thromres.2015.02.003
9. Bartlett P, Shawe-Taylor J. Generalization performance of support vector machines and other pattern classifiers. In: Bernhard S, Christopher JCB, Alexander JS, editors. *Advances in Kernel Methods: Support Vector Learning.* Cambridge, MA: MIT Press (1999). p. 43–54.
10. Minsky ML, Papert SA. *Perceptrons - Expanded Edition: An Introduction to Computational Geometry*. Boston, MA: MIT press Boston (1987).
11. Sammut C, Webb GI, eds. *Encyclopedia of Machine Learning*. New York, NY; London: Springer (2010).
12. Tikhonov AN, Arsenin VI. *Solutions of Ill-posed Problems*. Washington, DC; New York, NY; Winston; distributed solely by Halsted Press (1977).
13. Toloşi L, Lengauer T. Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics.* (2011) 27:1986–94. doi: 10.1093/bioinformatics/btr300
14. Rozonoer L, Mirkin B, Muchnik I, eds. Braverman readings in machine learning. In: *Key Ideas from Inception to Current State: International Conference Commemorating the 40th Anniversary of Emmanuil Braverman's Decease, Boston, MA Invited Talks*. Cham: Springer International Publishing (2018).
15. Chicco D. Ten quick tips for machine learning in computational biology. *BioData Min.* (2017) 10:35. doi: 10.1186/s13040-017-0155-3
16. Robin X, Turck N, Hainard A, Lisacek F, Sanchez J-C, Müller M. Bioinformatics for protein biomarker panel classification: what is needed to bring biomarker panels into *in vitro* diagnostics? *Expert Rev Proteomics.* (2009) 6:675–89. doi: 10.1586/epr.09.83
17. Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther.* (2012) 22:271–4. doi: 10.1089/nat.2012.0367
18. Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Ann Rev Biochem.* (2011) 80:273–99. doi: 10.1146/annurev-biochem-061308-093216
19. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. *Genomics.* (2009) 93:105–11. doi: 10.1016/j.ygeno.2008.10.003
20. Bishop CM. *Pattern Recognition and Machine Learning.* Corrected at 8th printing 2009. New York, NY: Springer (2009).
21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
22. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* (2003) 19:185–93. doi: 10.1093/bioinformatics/19.2.185

23. Huang H, Lu X, Liu Y, Haaland P, Marron JS. R/DWD: distance-weighted discrimination for classification, visualization and batch adjustment. *Bioinformatics.* (2012) 28:1182–3. doi: 10.1093/bioinformatics/bts096
24. Shabalin AA, Tjelmeland H, Fan C, Perou CM, Nobel AB. Merging two gene-expression studies via cross-platform normalization. *Bioinformatics.* (2008) 24:1154–60. doi: 10.1093/bioinformatics/btn083
25. Borisov N, Shabalina I, Tkachev V, Sorokin M, Garazha A, Pulin A, et al. Shambhala: a platform-agnostic data harmonizer for gene expression data. *BMC Bioinformatics.* (2019) 20:66. doi: 10.1186/s12859-019-2641-8
26. Buzdin A, Sorokin M, Garazha A, Sekacheva M, Kim E, Zhukov N, et al. Molecular pathway activation - new type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin Cancer Biol.* (2018) 53:110–24. doi: 10.1016/j.semcancer.2018.06.003
27. Turki T, Wei Z. Learning approaches to improve prediction of drug sensitivity in breast cancer patients. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL: IEEE (2016). p. 3314–20.
28. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* (2012) 483:603–7. doi: 10.1038/nature11003
29. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* (2013) 41:D955–61. doi: 10.1093/nar/gks1111
30. Venkova L, Aliper A, Suntsova M, Kholodenko R, Shepelin D, Borisov N, et al. Combinatorial high-throughput experimental and bioinformatic approach identifies molecular pathways linked with the sensitivity to anticancer target drugs. *Oncotarget.* (2015) 6:27227–38. doi: 10.18632/oncotarget.4507
31. Tarca A, Draghici S, Bhatti G, Romero R. Down-weighting overlapping genes improves gene set analysis. *BMC Bioinformatics.* (2012) 13:136. doi: 10.1186/1471-2105-13-136
32. Ozerov IV, Lezhnina KV, Izumchenko E, Artemov AV, Medintsev S, Vanhaelen Q, et al. *In silico* pathway activation network decomposition analysis (iPANDA) as a method for biomarker development. *Nat Commun.* (2016) 7:13427. doi: 10.1038/ncomms13427
33. Tkachev V, Sorokin M, Mescheryakov A, Simonov A, Garazha A, Buzdin A, et al. FLOating-window projective separator (FloWPS): a data trimming tool for support vector machines (SVM) to improve robustness of the classifier. *Front Genetics.* (2019) 9:717. doi: 10.3389/fgene.2018.00717
34. Cho H-J, Lee S, Ji YG, Lee DH. Association of specific gene mutations derived from machine learning with survival in lung adenocarcinoma. *PLoS ONE.* (2018) 13:e0207204. doi: 10.1371/journal.pone.0207204
35. Wang Z, Yang H, Wu Z, Wang T, Li W, Tang Y, et al. *In silico* prediction of blood-brain barrier permeability of compounds by machine learning and resampling methods. *ChemMedChem.* (2018) 13:2189–201. doi: 10.1002/cmdc.201800533
36. Soufan O, Kleftogiannis D, Kalnis P, Bajic VB. DWFS: a wrapper feature selection tool based on a parallel genetic algorithm. *PLoS ONE.* (2015) 10:e0117988. doi: 10.1371/journal.pone.0117988
37. Mahoney MW, Drineas P. CUR matrix decompositions for improved data analysis. *Proc Natl Acad Sci USA.* (2009) 106:697–702. doi: 10.1073/pnas.0803205106
38. Menden MP, Casale FP, Stephan J, Bignell GR, Iorio F, McDermott U, et al. The germline genetic component of drug sensitivity in cancer cell lines. *Nat Commu.* (2018) 9:3385. doi: 10.1038/s41467-018-05811-3
39. Yosipof A, Guedes RC, García-Sosa AT. Data mining and machine learning models for predicting drug likeness and their disease or organ category. *Front Chem.* (2018) 6:162. doi: 10.3389/fchem.2018.00162
40. Mamoshina P, Kochetov K, Putin E, Cortese F, Aliper A, Lee W-S, et al. Population specific biomarkers of human aging: a big data study using South Korean, Canadian and Eastern European patient populations. *J Gerontol Series A Biol. Sci. Med. Sci.* (2018) 73:1482–90. doi: 10.1093/gerona/gly005
41. Mamoshina P, Volosnikova M, Ozerov IV, Putin E, Skibina E, Cortese F, et al. Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. *Front. Genet.* (2018) 9:242. doi: 10.3389/fgene.2018.00242
42. Zhang L, Zhang H, Ai H, Hu H, Li S, Zhao J, et al. Applications of machine learning methods in drug toxicity prediction. *Curr Top Med Chem.* (2018) 18:987–97. doi: 10.2174/1568026618666180727152557
43. Kim YR, Kim D, Kim SY. Prediction of acquired taxane resistance using a personalized pathway-based machine learning method. *Cancer Res Treat.* (2018) 51:672–84. doi: 10.4143/crt.2018.137
44. Borisov N, Tkachev V, Muchnik I, Buzdin A. Individual drug treatment prediction in oncology based on machine learning using cell culture gene expression data. In: *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics.* Newark, NJ: ACM Press (2017). pp. 1–6.
45. Borisov N, Tkachev V, Buzdin A, Muchnik I. Prediction of drug efficiency by transferring gene expression data from cell lines to cancer patients. In: Rozonoer L. Mirkin B, Muchnik I, editors *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*, Cham: Springer International Publishing (2018). pp. 201–12.
46. Mulligan G, Mitsiades C, Bryant B, Zhan F, Chng WJ, Roels S, et al. Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood.* (2007) 109:3177–88. doi: 10.1182/blood-2006-09-044974
47. Data Trimming. *The SAGE Encyclopedia of Communication Research Methods* Thousand Oaks, CA: SAGE Publications, Inc.
48. Mitchell TM. *Machine Learning. International edn*. New York, NY: McGraw-Hill (1997).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

PERSPECTIVE
published: 04 September 2019
doi: 10.3389/fonc.2019.00851

Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics

*Anna Petrackova[1], Michal Vasinek[2], Lenka Sedlarikova[1], Tereza Dyskova[1], Petra Schneiderova[1], Tomas Novosad[2], Tomas Papajik[3] and Eva Kriegova[1]**

[1] Department of Immunology, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc, Czechia, [2] Department of Computer Science, Faculty of Electrical Engineering and Computer Science, Technical University of Ostrava, Ostrava, Czechia, [3] Department of Hemato-Oncology, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc, Czechia

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Robert Charles Day,
University of Otago, New Zealand
Ole Winther,
Technical University of Denmark, Denmark

***Correspondence:**
Eva Kriegova
eva.kriegova@email.cz

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: 07 June 2019
Accepted: 19 August 2019
Published: 04 September 2019

Citation:
Petrackova A, Vasinek M, Sedlarikova L, Dyskova T, Schneiderova P, Novosad T, Papajik T and Kriegova E (2019) Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. Front. Oncol. 9:851. doi: 10.3389/fonc.2019.00851

The insufficient standardization of diagnostic next-generation sequencing (NGS) still limits its implementation in clinical practice, with the correct detection of mutations at low variant allele frequencies (VAF) facing particular challenges. We address here the standardization of sequencing coverage depth in order to minimize the probability of false positive and false negative results, the latter being underestimated in clinical NGS. There is currently no consensus on the minimum coverage depth, and so each laboratory has to set its own parameters. To assist laboratories with the determination of the minimum coverage parameters, we provide here a user-friendly coverage calculator. Using the sequencing error only, we recommend a minimum depth of coverage of 1,650 together with a threshold of at least 30 mutated reads for a targeted NGS mutation analysis of ≥3% VAF, based on the binomial probability distribution. Moreover, our calculator also allows adding assay-specific errors occurring during DNA processing and library preparation, thus calculating with an overall error of a specific NGS assay. The estimation of correct coverage depth is recommended as a starting point when assessing thresholds of NGS assay. Our study also points to the need for guidance regarding the minimum technical requirements, which based on our experience should include the limit of detection (LOD), overall NGS assay error, input, source and quality of DNA, coverage depth, number of variant supporting reads, and total number of target reads covering variant region. Further studies are needed to define the minimum technical requirements and its reporting in diagnostic NGS.

Keywords: next-generation sequencing, variant allele frequency (VAF), coverage depth calculator, sequencing error, small subclones, *TP53* gene

INTRODUCTION

Next-generation sequencing (NGS) has rapidly expanded into the clinical setting in haemato-oncology and oncology, as it may bring great benefits for diagnosis, selection of treatment, and/or prognostication for many patients (1). Recently, several articles about the validation of deep targeted NGS in clinical oncology were published (2, 3), including a comprehensive recommendation by the Association for Molecular Pathology and the College of American Pathologists (1). However, the lack of standardization of targeted NGS methods still limits their implementation in clinical practice (4).

One challenge in particular is the correct detection of mutations present at low variant allele frequencies (VAF) and standardization of sequencing coverage depth (1, 5, 6). This is especially important for mutations that have clinical impacts at subclonal frequencies (1) such as the case of *TP53* gene mutations (*TP53*mut) in chronic lymphocytic leukemia (CLL) (7, 8). *TP53* aberrations (*TP53*mut and/or chromosome 17p deletion) are among the strongest prognostic and predictive markers guiding treatment decisions in CLL (9). Nowadays, the European Research Initiative on Chronic Lymphocytic Leukemia (ERIC) recommends detecting *TP53*mut with a limit of detection (LOD) of at least 10% VAF (10), and a growing body of evidence exists dedicated to the clinical impact of small *TP53* mutated subclones in CLL (7, 8).

Sanger sequencing and deep targeted NGS are currently the techniques most used for *TP53*mut analysis (10) as well as for analysis of other genes with clinical impacts at low allele frequencies. Although Sanger sequencing provides a relatively accessible sequencing approach, it lacks the sensitivity needed to detect subclones due to its detection limit of 10–20% of mutated alleles (10). NGS-based analysis has thus gained prominence in diagnostic laboratories for the detection of somatic variants and various technical developments of error correction strategies, both computational and experimental, are being developed for the accurate identification of low-level genetic variations (11). We therefore address the importance of the correct determination of sequencing depth in diagnostic NGS in order to obtain a confident and reproducible detection, not only of low VAF variants. Finally, we performed a dilution experiment to confirm our theoretical calculations, and we close by discussing our experience with diagnostic detection of *TP53*mut in CLL patients and further perspectives about NGS standardization in cancer diagnostics.

NGS SEQUENCING DEPTH AND ERROR RATE

NGS sequencing depth directly affects the reproducibility of variant detection: the higher the number of aligned sequence reads, the higher the confidence to the base call at a particular position, regardless of whether the base call is the same as the reference base or is mutated (1). In other words, individual sequencing error reads are statistically irrelevant when they are outnumbered by correct reads. Thus, the desired coverage depth should be determined based on the intended LOD, the tolerance for false positive or false negative results, and the error rate of sequencing (1, 11).

Using a binomial distribution, the probability of false positive and false negative results for a given error rate as well as the intended LOD can be calculated, and the threshold for a variant calling for a given depth can be estimated (1). For example, given a sequencing error rate of 1%, a mutant allele burden of 10%, and a depth of coverage 250 reads, the probability of detecting 9 or fewer mutated reads is, according to the binomial distribution, 0.01%. Hence, the probability of detecting 10 or more mutated reads is 99.99% (100–0.01%), and the threshold for a variant calling can be defined. In other words, a coverage depth of 250 with a threshold of at least 10 mutated reads will have a 99.99% probability that 10% of the mutant allele load will not be missed by the variant calling (although it can be detected in a different proportion). In this way, the risk of a false negative result is greatly minimized. On the other hand, the probability of false positives heavily depends on the sequencing error rate (as the accuracy of all analytical measurements depends on the signal-to-noise ratio) (1, 11). In our example, the probability of a false positive result is 0.025%; however, the rate of false positives is not negligible when decreasing the LOD to the value close to the error rate. Conventional intrinsic NGS error rates range between 0.1 and 1% (Phred quality score of 20–30) (1, 11) depending on the sequencing platform, the GC content of the target regions (12), and the fragment length, as shown in Illumina paired-end sequencing (13). Therefore, the detection of variants at VAFs $<2\%$ is affected by a high risk of a false positive result, regardless of the coverage depth. It is also important to mention that the sequencing error rate applies only for errors produced by sequencing itself and does not include other errors introduced during DNA processing and library preparation, particularly during amplification steps, which further increase error rates (1, 11).

MINIMUM SEQUENCING COVERAGE IN CLINICAL SETTINGS

There is currently no consensus on the minimum required coverage in a clinical setting using deep targeted resequencing by NGS, and so each laboratory has to set its own parameters in order to meet sufficient quality (1, 5). To date, only a few studies have recommended the minimum coverage criteria for deep targeted NGS in clinical oncology: 500 depth of coverage and a LOD of 5% (2), 300–500 depth of coverage without defying the LOD (3), 250 depth and a LOD of 5% with threshold adjustment to 1,000 depth of coverage is required in cases of heterogeneous variants in low tumor cellularity samples (1), and 100 depth with at least 10 variant reads and a LOD of 10% (10). According to the binominal data distribution, a coverage depth of 250 should indeed be sufficient to detect 5% VAF with a threshold of variant supporting reads ≥ 5 (**Figure 1**). On the other hand, NGS analysis with a coverage depth of 100 along with a requirement of at least 10 variant supporting reads as recommended by the ERIC consortium (10) would

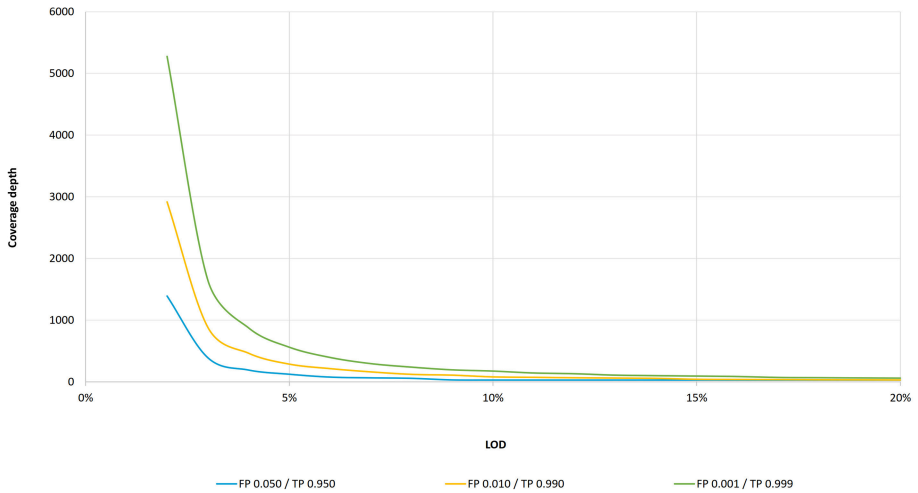


FIGURE 1 | LOD as a function of coverage depth according to the binomial distribution. Coverage depth needed to maintain an intended LOD (within 3–20% VAF range) for three cumulative probability settings: for false positive probability of 0.001 and true positive of 0.999, a LOD of 20% is achieved at 61 coverage depth, a LOD of 10% at 175, a LOD of 5% at 562, and a LOD of 3% at 1,650. For the false positive probability of 0.010 and true positive of 0.990, a LOD of 20% is achieved at 31, a LOD of 10% at 81, a LOD of 5% at 288, and a LOD of 3% at 886 coverage depth, respectively. For the false positive probability of 0.050 and true positive of 0.950, a LOD of 20% is achieved at 30, a LOD of 10% at 30, a LOD of 5% at 124, and a LOD of 3% at 392 coverage depth, respectively.

result in a false negative of 45% for samples with a LOD of 10%. To confirm these theoretic calculations, we performed two independent dilution experiments to estimate the performance of *TP53* NGS analysis to detect 10% VAF at a depth of coverage of 100 reads. Indeed, we detected 30% of false negatives (5 positive samples of 7 true-positive samples and 9 positive samples of 13 true-positive samples) in two independent sequencing runs. Unfortunately, the false negative rate is often underestimated in targeted resequencing. Also, a recent study investigating inter-laboratory results of somatic variant detection with VAFs between 15 and 50% in 111 laboratories with reported LODs of 5–15% (6) shows that major errors in diagnostic NGS may arise from false negative results, even in samples with high mutation loads (6). Of three concurrent false positive results, all variants were correctly detected but mischaracterised (6). Since laboratories have not been asked to report coverage depth for other regions than the identified variants (6), we may only assume that low coverage or high variant calling thresholds contributed to the false negative results. These results further highlight the need for standardized coverage depth parameters in diagnostic NGS, taking into account sequencing errors as well as assay-specific errors.

FREQUENCY OF *TP53* SUBCLONAL MUTATIONS IN CLL DETECTED THROUGH DIAGNOSTIC NGS

In order to evaluate the occurrence of low VAF in real-world settings, we reviewed our cohort of CLL patients examined for *TP53*mut in our diagnostic laboratory. The *TP53*mut were assessed as reported previously (14, 15). Briefly, *TP53* (exons 2–10 including 2 bp intronic overlap, 5′ and 3′UTR) was analyzed using 100 ng gDNA per reaction. Amplicon-based libraries were sequenced as paired-end on MiSeq (2x151, Illumina) with minimum target read depths of 5,000x. The LOD of *TP53*mut was set up to 1%, and the variants in the range 1–3% were confirmed by replication. Written informed consent was obtained from all the patients who were enrolled in accordance with the Helsinki Declaration, and the study was approved by the local ethical committee.

Of the diagnostic cohort of 859 CLL patients (April 2016–April 2019), 25% (215/859) were positive for *TP53*mut, and of those, 52.6% (113/215) carried variants with VAF at 10% or lower. In line with our observations, a recent study (8) reported the presence of 63 and 84% low burden (Sanger negative) *TP53*muts in CLL patients at the time of diagnosis and at the time of treatment, respectively, and confirmed the negative impact on the overall survival of *TP53*muts above 1% VAF at the time of treatment (8).

CALCULATOR FOR DIAGNOSTIC NGS SETTINGS FOR DETECTION OF SUBCLONAL MUTATIONS

To assist laboratories with the determination of the minimum proper coverage parameters, we are providing a simple, user-friendly theoretical calculator (software) based on the binomial distribution (**Figure 2**), described in the **Supplementary File**. A web (or desktop) application and stand-alone source codes in R are accessible on Github: https://github.com/mvasinek/olgen-coverage-limit. Using this calculator, the correct parameters of sequencing depth and the corresponding minimum number of variant reads for a given sequencing error rate and intended LOD can easily be determined. Moreover, users can also take into account other errors by simply adding assay-specific errors to the sequencing error rate and using this overall error as an input

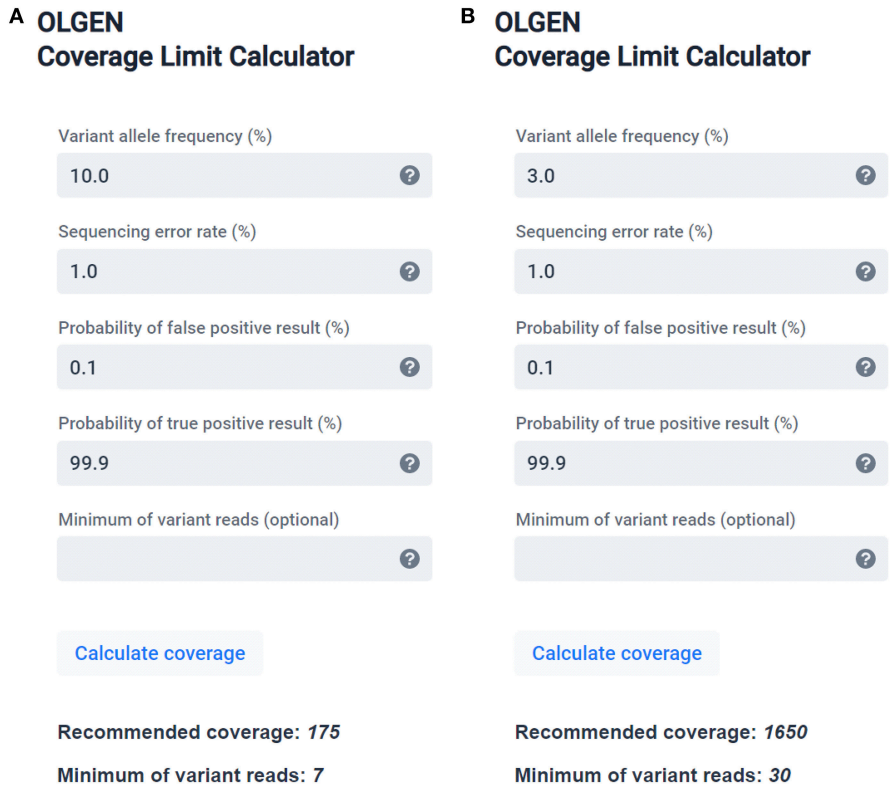


FIGURE 2 | OLGEN Coverage Limit calculator—a simple theoretical calculator suitable for determining the correct sequencing depth and corresponding minimum number of variant reads according to the binomial distribution for a given sequencing error rate and intended LOD recommended for diagnostic NGS. Examples of calculated sequencing depths and the corresponding minimum number of variant reads recommended for variants with **(A)** 10% VAF and 99.9% probability of detection and **(B)** 3% VAF and 99.9% probability of detection.

to the calculator. For example, in our case of *TP53* mutational analysis we calculated with the overall error of ∼1.16%, thus we set up our minimum coverage depth requirements to 2,000 with threshold of minimum 40 reads for 3% VAF.

DISCUSSION

Although diagnostic NGS has gained prominence in clinical settings for the assessment of somatic mutations in cancer, insufficient standardization of sequencing parameters still limits its implementation in clinical practice (1), mainly for variants present at low allele frequencies (4). We, therefore, addressed the technical question of correctly determining the sequencing depth in diagnostic NGS in order to obtain confident and reproducible detections of low VAF variants. In particular, we performed theoretical calculations to determine the optimum depth of coverage for the desired probability of detection of variants at low allele frequencies, taking into account the sequencing error rate. Moreover, we confirmed these theoretical calculations by conducting dilution experiments. Based on these observations, we recommend a depth of coverage of 1,650 or higher (together with the respective threshold of at least 30 mutated reads) to call ≥3% variants to achieve a 99.9% probability of variant detection, using the conventional NGS sequencing error only. Variants in the 1–3% VAF range can only be called if the obtained sequence data is of high quality (average Q30 > 90%) and/or when the variants are confirmed by replication or the orthogonal method (1, 11, 16). We are also providing a simple, user-friendly theoretical calculator (software) to assist laboratories with resolving the correct sequencing depth and the corresponding minimum number of variant reads while taking into account the sequencing error rate. Our simple calculator may help to minimize the false positive and false negative results in diagnostic NGS.

Nevertheless, correct sequencing depth is also influenced by assay-specific factors (1). Errors can occur at many stages during DNA processing and library preparation. The most common are amplification errors introduced during NGS library preparation (1, 12, 17). Other common sources of errors have to do with library complexity (the number of independent DNA molecules analyzed), DNA quality, and target region complexity etc. All potential assay-specific errors should be addressed through test design, method validation, and quality control.

Currently, emerging error correction strategies, both computational and experimental, are being developed in order to mitigate the high error rates in diagnostic NGS (11). So far, among the most promising error correction methods are UMI (Unique Molecular Identifiers), which correct for

PCR errors (18), and signal-to-noise correction approaches (11). These advances attempt to reduce the LOD, thereby increasing sequencing accuracy needed for future opportunities in NGS diagnosis.

In order to improve the standardization in diagnostic NGS, the estimation of correct coverage depth is a recommended starting point when assessing thresholds surrounding a particular NGS assay. Nevertheless, there is still lack of published guidance regarding the minimum technical requirements and its reporting in NGS, particularly important in detection of clonal and subclonal mutations in cancer diagnostics. This is mainly due to the broad range of library preparation approaches, and numerous variables playing a role in each specific NGS assay, that are difficult to standardize, together with inter-laboratory variability. Therefore, the definition of minimum technical requirements and its reporting in NGS is highly desirable. Based on our experience in diagnostic NGS in haemato-oncology, we suggest to report at least following technical parameters: LOD, overall error of NGS assay (or at least sequencing error rate), the amount of DNA input, source, and quality of DNA, minimum coverage depth and the percentage of targeted bases sequenced at this minimum depth, total number of target reads covering variant region and number of reads supporting the variant. Special emphasis should be given to NGS standardization of the formalin-fixed paraffin-embedded (FFPE) samples (19, 20).

Taken together, our study highlights the importance of correct sequencing depth and the minimum number of reads required for reliable and reproducible detection of variants with low VAF in diagnostic NGS. The calculation of correct sequencing depth for a given error rate using our user-friendly theoretical calculator (software) may help to minimize the false positive and false negative results in diagnostic NGS, in situations related to subclonal mutations among others. The rigorous testing and standardized minimum requirements for diagnostic NGS is particularly desirable to ensure correct results in clinical settings.

DATA AVAILABILITY

The datasets generated for this study are available on reasonable request to the corresponding author.

AUTHOR CONTRIBUTIONS

AP and EK designed the study, interpreted the results, and wrote the manuscript. AP, LS, TD, and PS performed NGS analysis. TP collected the patient samples and clinical data. MV performed bioinformatics analysis and wrote the calculator code. TN prepared web application. All authors read and approved the final version of manuscript.

FUNDING

Grant support: MZ CR VES16-32339A, in part by the MH CZ—DRO (FNOl, 00098892).

ACKNOWLEDGMENTS

We apologize to the many authors whose articles could not be cited because of reference limits.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00851/full#supplementary-material

REFERENCES

1. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. (2017) 19:341–65. doi: 10.1016/j.jmoldx.2017.01.011
2. D'Haene N, Le Mercier M, De Neve N, Blanchard O, Delaunoy M, El Housni H, et al. Clinical validation of targeted next generation sequencing for Colon and Lung cancers. *PLoS ONE*. (2015) 10:e0138245. doi: 10.1371/journal.pone.0138245
3. Deans ZC, Costa JL, Cree I, Dequeker E, Edsjo A, Henderson S, et al. Integration of next-generation sequencing in clinical diagnostic molecular pathology laboratories for analysis of solid tumours; an expert opinion on behalf of IQN path ASBL. *Virchows Arch*. (2017) 470:5–20. doi: 10.1007/s00428-016-2025-7
4. Ivanov M, Laktionov K, Breder V, Chernenko P, Novikova E, Telysheva E, et al. Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: intrinsic obstacles and possible solutions. *J Transl Med*. (2017) 15:22. doi: 10.1186/s12967-017-1125-8
5. Bacher U, Shumilov E, Flach J, Porret N, Joncourt R, Wiedemann G, et al. Challenges in the introduction of next-generation sequencing (NGS) for diagnostics of myeloid malignancies into clinical routine use. *Blood Cancer J*. (2018) 8:113. doi: 10.1038/s41408-018-0148-6
6. Merker JD, Devereaux K, Iafrate AJ, Kamel-Reid S, Kim AS, Moncur JT, et al. Proficiency testing of standardized samples shows very high interlaboratory agreement for clinical next-generation sequencing-based oncology assays *Arch Pathol Lab Med*. (2019) 143:463–71. doi: 10.5858/arpa.2018-0336-CP
7. Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Fama R, et al. Clinical impact of small *TP53* mutated subclones in chronic lymphocytic leukemia. *Blood*. (2014) 123:2139–47. doi: 10.1182/blood-2013-11-539726
8. Brieghel C, Kinalis S, Yde CW, Schmidt AY, Jonson L, Andersen MA, et al. Deep targeted sequencing of *TP53* in chronic lymphocytic leukemia: clinical impact at diagnosis and at time of treatment. *Haematologica*. (2019) 104:789–96. doi: 10.3324/haematol.2018.195818
9. Campo E, Cymbalista F, Ghia P, Jager U, Pospisilova S, Rosenquist R, et al. *TP53* aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics. *Haematologica*. (2018) 103:1956–68. doi: 10.3324/haematol.2018.187583
10. Malcikova J, Tausch E, Rossi D, Sutton LA, Soussi T, Zenz T, et al. ERIC recommendations for *TP53* mutation analysis in chronic lymphocytic leukemia-update on methodological approaches and results interpretation. *Leukemia*. (2018) 32:1070–80. doi: 10.1038/s41375-017-0007-7
11. Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet*. (2018) 19:269–85. doi: 10.1038/nrg.2017.117
12. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent,

Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. (2012) 13:341. doi: 10.1186/1471-2164-13-341
13. Tan G, Opitz L, Schlapbach R, Rehrauer H. Long fragments achieve lower base quality in Illumina paired-end sequencing. *Sci Rep*. (2019) 9:2856. doi: 10.1038/s41598-019-39076-7
14. Obr A, Prochazka V, Jirkuvova A, Urbankova H, Kriegova E, Schneiderova P, et al. *TP53* mutation and complex karyotype portends a dismal prognosis in patients with mantle cell lymphoma. *Clin Lymphoma Myeloma Leuk*. (2018) 18:762–8. doi: 10.1016/j.clml.2018.07.282
15. Turcsanyi P, Kriegova E, Kudelka M, Radvansky M, Kruzova L, Urbanova R, et al. Improving risk-stratification of patients with chronic lymphocytic leukemia using multivariate patient similarity networks. *Leuk Res*. (2019) 79:60–8. doi: 10.1016/j.leukres.2019.02.005
16. Shin HT, Choi YL, Yun JW, Kim N, Kim SY, Jeon HJ, et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. *Nat Commun*. (2017) 8:1377. doi: 10.1038/s41467-017-01470-y
17. Ma X, Shao Y, Tian L, Flasch DA, Mulder HL, Edmonson MN, et al. Analysis of error profiles in deep next-generation sequencing data. *Genome Biol*. (2019) 20:50. doi: 10.1186/s13059-019-1659-6
18. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. (2017) 27:491–9. doi: 10.1101/gr.209601.116
19. McDonough SJ, Bhagwate A, Sun Z, Wang C, Zschunke M, Gorman JA, et al. Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. *PLoS ONE*. (2019) 14:e0211400. doi: 10.1371/journal.pone.0211400
20. Ascierto PA, Bifulco C, Palmieri G, Peters S, Sidiropoulos N. Pre-analytic variables and tissue stewardship for reliable next-generation sequencing (NGS) clinical analysis. *J Mol Diagn*. (2019)21:756–67. doi: 10.1016/j.jmoldx.2019.05.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 September 2019
doi: 10.3389/fgene.2019.00884

MUDENG Expression Profiling in Cohorts and Brain Tumor Biospecimens to Evaluate Its Role in Cancer

*Juhyun Shin[1,2], Jun-Ha Choi[2], Seunghwa Jung[2], Somi Jeong[2], Jeongheon Oh[2], Do-Young Yoon[3], Man Hee Rhee[4], Jaehong Ahn[5], Se-Hyuk Kim[6] and Jae-Wook Oh[1,2]**

[1] Animal Resources Research Center, Konkuk University, Seoul, South Korea, [2] Department of Stem Cell and Regenerative Biotechnology, Konkuk University, Seoul, South Korea, [3] Department of Bioscience and Biotechnology, Konkuk Institute of Technology, Konkuk University, Seoul, South Korea, [4] Department of Veterinary Medicine, College of Veterinary Medicine, Kyungpook National University, Daegu, South Korea, [5] Department of Ophthalmology, Ajou University School of Medicine, Suwon, South Korea, [6] Department of Neurosurgery, Ajou University School of Medicine, Suwon, South Korea

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Fang Wang,
University of Texas MD Anderson Cancer Center, United States
Youxin Ji,
Qingdao Central Hospital, China

***Correspondence:**
Jae-Wook Oh
ohjw@konkuk.ac.kr

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Genetics

Received: *29 May 2019*
Accepted: *21 August 2019*
Published: *19 September 2019*

Citation:
Shin J, Choi J-H, Jung S, Jeong S, Oh J, Yoon D-Y, Rhee MH, Ahn J, Kim S-H and Oh J-W (2019) MUDENG Expression Profiling in Cohorts and Brain Tumor Biospecimens to Evaluate Its Role in Cancer. Front. Genet. 10:884. doi: 10.3389/fgene.2019.00884

Mu-2-related death-inducing gene (MUDENG, *MuD*) has been reported to be involved in the tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-associated apoptotic pathway of glioblastoma multiforme (GBM) cells; however, its expression level, interactors, and role in tumors are yet to be discovered. To investigate whether *MuD* expression correlates with cancer progression, we analyzed The Cancer Genome Atlas (TCGA) database using UALCAN and Gene Expression Profiling Interactive Analysis (GEPIA). Differential expression of *MuD* was detected in 6 and 10 cancer types, respectively. Validation performed using data from the Gene Expression Omnibus database showed that MuD expression is downregulated in KIRC tumor and correlate with higher chance of survival. Upregulation of *MuD* expression in GBM tumors was detected through GEPIA and high MuD expression correlated with higher survival in proneural GBM, whereas the opposite was observed in classical GBM subtype. GBM biospecimens analysis shows that MuD protein level was upregulated in three of six specimens, whereas mRNA level remained relatively unaltered. Therefore, *MuD* may exert differential effects according to subtypes, and/or be subjected to post-translational regulation in GBM. Correlation analysis between GBM cohort database and experiments using GBM cell lines revealed its positive effect on regulation of protein phosphatase 2 regulatory subunit B'Epsilon (*PPP2R5E*) and son of sevenless homolog 2 (*SOS2*). STRING database analysis indicated that the components of adaptor protein complexes putatively interacted with MuD but showed no correlation in terms of survival of patients with different GBM subtypes. In summary, we analyzed the expression of *MuD* in publicly available cancer patient data sets, GBM cell lines, and biospecimens to demonstrate its potential role as a biomarker for cancer prognosis and identified its candidate interacting molecules.

Keywords: Mu-2-related death-inducing gene, the cancer genome atlas, gene expression omnibus, patient survival, glioblastoma multiforme, cancer cell apoptosis

INTRODUCTION

Glioblastoma multiforme (GBM) is the most common and malignant form of primary brain tumor (Huse and Holland, 2010; Siegel et al., 2017). Despite recent advances in surgical and other therapeutic techniques, the median survival of patients with GBM is as low as 12 to 15 months (Jung et al., 2014; Ostrom et al., 2014). Aside from the conventional therapies, the selective induction of apoptosis in target cancer cells with pro-apoptotic cytokines, such as tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) (Merino et al., 2007) seems promising, as this strategy exhibited low toxicity to non-cancerous cells, including brain cells, in clinical trials (Stuckey and Shah, 2013). However, the use of TRAIL is controversial because it is thought to induce apoptosis not only in cultured normal human hepatocytes but also in normal brain tissues (Jo et al., 2000). Therefore, the applicability of TRAIL for the treatment of brain cancer by combinatorial drug treatment strategies should be carefully monitored to improve its therapeutic efficacy (Stuckey and Shah, 2013).

One of the hallmarks and causes of GBM complexity is cellular heterogeneity, which poses a challenge for disease diagnosis and treatment (Friedmann-Morvinski, 2014; Inda et al., 2014). The molecular profiling of The Cancer Genome Atlas (TCGA) divides GBM into four distinctive subtypes, namely, classical, neural, proneural, and mesenchymal (Verhaak et al., 2010). Both classical and mesenchymal subtypes are aggressive in nature. Whereas the classical subtype is characterized by overexpression of epidermal growth factor receptor (EGFR), the mesenchymal subtype shows decreased neurofibromin 1 (NF1) expression and high transforming growth factor-β (TGF-β) and nuclear factor-κB (NF-κB) activities. The neural subtype is controversial because it is thought to originate from the substantial contamination of GBM samples with healthy brain tissue. Tumorigenesis of the proneural subtype starts from the frontal cortex of the cerebrum and often displays amplification of platelet-derived growth factor receptor α (PDGFRα) and mutations of isocitrate dehydrogenase 1/2 (IDH1/2) and tumor protein 53 (TP53) (Verhaak et al., 2010). Patients with proneural subtype exhibit the best prognosis but may have the worst disease outcomes in the absence of *IDH1* mutations (Phillips et al., 2006; Verhaak et al., 2010; Akan et al., 2012). Although several efforts have been directed to identify the critical driver pathways and therapeutic targets specific for each GBM subtype, very little progress has been made in this direction. A recent report revealed increased sensitivity of patients with proneural GBM to cyclin-dependent kinase 4/6 (CDK4/6) inhibitor treatment (Li et al., 2017) and significantly faster recurrence after bevacizumab treatment in patients with classic GBM (Hovinga et al., 2019), indicating the importance of careful evaluation of the subtypes before treatment.

The mu-2-related death-inducing gene (MUDENG, *MuD*), also called as the adaptor-related protein complex 5 subunit Mu 1 (AP5M1), was identified as a putative component of the fifth adaptor protein (AP) complex involved in endosomal transport (Hirst et al., 2011). *MuD* was reported to be involved in the apoptotic pathway in HeLa (Lee et al., 2008), Jurkat (Lee et al., 2008; Shin et al., 2013), and B-JAB (Lee et al., 2008) cell lines. Subsequent studies demonstrated the cleavage of MuD by active caspase-3 during TRAIL-induced apoptotic signaling (Shin et al., 2013), and the subsequent activation of the anti-apoptotic function of MuD near the BH3-interacting domain death agonist (BID) and B-cell lymphoma 2 (Bcl2) junction (Choi et al., 2016). These studies suggest a possible role for MuD in cancer cells apoptotic signaling.

In the present study, we used UALCAN and GEPIA, two web-based tools that allow in-depth analyses of RNA-sequencing data from TCGA database to assess *MuD* expression in cancer cohorts. In addition, we used the microarray data from the Gene Expression Omnibus (GEO) database to validate the selected results. We conducted an integrated analysis using 12 human brain tumor samples and GBM cancer cell lines. Furthermore, we identified the differential expression of *MuD* in tumors as well as the correlation between *MuD* expression and survival in cancer types, including specific GBM subtypes. We also identified the candidate interacting genes that were validated in GBM cell lines.

MATERIALS AND METHODS

Data Sources

The TCGA database curated by the National Institute of Health (NIH) comprises 2.5 petabytes of data on cohorts from 33 different tumor types, including genomic profiles from microarrays and next-generation sequencing (NGS) (Tomczak et al., 2015). MET500 is a database of NGS data from 500 patients with cancers of 30 primary sites (Robinson et al., 2017). Genotype-tissue expression project (GTEx) is a database of NGS and includes the microarray data collected from nearly 1,000 individuals (Consortium, 2013). As test sets, we used data sets available from the GEO database (Clough and Barrett, 2016). E-GEOD-53757 (von Roemeling et al., 2014) and E-GEOD-22541 (Wuttig et al., 2012) was used for KIRC validation (Wuttig et al., 2012), E-GEOD-70951 (Quigley et al., 2017) and E-GEOD-10886 (Parker et al., 2009) for BRCA validation, E-GEOD-68465 (Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma et al., 2008) for LUAD validation and E-GEOD-23400 (Su et al., 2011) for ESCA.

Statistical Analysis

Analysis of the TCGA data was carried out with UALCAN (RRID: SCR_015827) (Chandrashekar et al., 2017), and GEPIA (Tang et al., 2017). The differential expression of *MuD* and patient survival were analyzed with PanCan analysis and expression on box plot, respectively. Cox proportional hazard analysis was performed with GBM-BioDP provided at the Glioblastoma Bio Discovery Portal (https://gbm-biodp.nci.nih.gov/) (Celiku et al., 2014). Patients were divided based on the diagnosed GBM subtype and further stratified into four quartiles as per *MuD* expression level. For each group, a Cox proportional hazard model was used to plot the survival of the patients from the first quartile versus those from the fourth quartile using age and O-6-methylguanine-DNA methyltransferase (*MGMT*) methylation

status as covariates. The microarray data from E-GEOD-53757 and E-GEOD-070951 processed with MAS.5 and limma (RRID : SCR_010943) in R, respectively, were visualized as heat maps using ClustVis (https://biit.cs.ut.ee/clustvis/) (Metsalu and Vilo, 2015). Heat maps were row-centered and unit variance scaling was applied for rows. Principal components were calculated using the NIPALS PCA method included in pcaMethods R package, and heatmaps were plotted using heatmap R package (version 0.7.7). Differential expression and survival plots were plotted using survminer R package (version 0.4.4) after processing with limma package and z-score (value-mean normal value/normal SD) calculated by R. Student's *t*-test was used to analyze differences between groups in the real-time quantitative polymerase chain reaction (RT-qPCR) and immunoblot data.

Sample Collection From Human Brain Tumors

This study was approved by the Konkuk University Institutional Review Board (IRB; 7001355-124 201512-E-041), and all patients signed IRB-approved consent forms. The biospecimens used in the present study were provided by the Ajou Human Bio-Resource Bank (Suwon, Korea), a member of the National Biobank of Korea, supported by the Ministry of Health and Welfare. All samples derived from the National Biobank of Korea were obtained with informed consent under institutional review board-approved protocols. We obtained 12 tissues from the following patients (**Table 1**): 10 patients diagnosed with glioblastoma grade IV, and healthy tissues of six of these patients; one patient diagnosed with oligodendroglioma grade II; and one patient diagnosed with ependymoma grade II. Samples were stored below −80°C until nucleic acid and protein extraction.

Cell Lines and Cell Culture

The U251-MG cell line (NCI-DTP Cat U-251, RRID: CVCL 0021) was obtained from Dr Benveniste EN (University of Alabama at Birmingham, Birmingham, AL, USA). The U251-MG *MuD* knock-out (KO) line β18 was generated using clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 plasmid SpCas9-2A-puro (PX459) V 2.0 provided by Feng Zhang (RRID: Addgene_6288) and single guide RNA 5′-ACACTAATTAGTGGCGGACG-3′ designed with CRISPR DESIGN (http://crispr.mit.edu/). U251-MG cells stably expressing (SE) GFP alone (C1) and GFP-MuD (C1MuD) were generated by transfection using Lipofectamine 2000 and

TABLE 1 | Differential *MuD* expression and survival correlation revealed by UALCAN and GEPIA.

	Project name	UALCAN	GEPIA	GR(U)	GRG)	S (U)	S (G)
BRCA	Breast invasive carcinoma	N (n = 114) T (n = 1094)	N (n = 291) T (n = 1085)	NS	NS	C (HL)	No
CHOL	Cholengiocarcinoma	N (n = 9) T (n = 36)	N (n = 9) T (n = 36)	Up	NS	No	No
COAD	Colon adenocarcinoma	N (n = 41) T (n = 286)	N (n = 349) T (n = 275)	Down	NS	No	No
DLBC	Lymphoid neoplasm diffuse large B-cell lymphoma	–	N (n = 337) T (n = 47)	–	Up	No	No
ESCA	Esophageal carcinoma	N (n = 11) T (n = 184)	N (n = 286) T (n = 182)	Up	Up	No	No
GBM	Glioblastoma multiforme	N (n = 5) T (n = 156)	N (n = 207) T (n = 163)	–	Up	No	No
KICH	Kidney chromophobe	N (n = 25) T (n = 67)	N (n = 53) T (n = 66)	NS	Up	No	No
KIRC	Kidney renal clear cell carcinoma	N (n = 72) T (n = 533)	N (n = 100) T (n = 523)	Down	Down	.C (HH)	C (HH)
KIRP	Kidney renal papillary cell carcinoma	N (n = 32) T (n = 290)	N (n = 60) T (n = 286)	Down	NS	No	No
LGG	Lower-grade glioma	N (n = 248) T (n = 265)	N (n = 207) T (n = 518)	NS	Up	No	No
LUAD	Lung adenocarcinoma	N (n = 29) T (n = 519)	N (n = 347) T (n = 483)	NS	NS	C (HL)	No
OV	Ovarian serous cystadenocarcinoma	–	N (n = 88) T (n = 426)	–	Up	No	No
PAAD	Pancreatic Adenocarcinoma	N (n = 4) T (n = 178)	N (n = 171) T (n = 179)	NS	Up	No	No
READ	Rectum adenocarcinoma	N (n = 11) T (n = 166)	N (n = 92) T (n = 318)	Down	NS	No	No
STAD	Stomach adenocarcinoma	N (n = 34) T (n = 415)	N (n = 34) T (n = 415)	NS	Up	No	No
THYM	Thymoma	N (n = 2) T (n = 120)	N (n = 339) T (n = 118)	NS	Up	No	No

Abbreviations of cancer types based on TCGA are given in the left column with project name. UALCAN and GEPIA sample sizes for normal (N) and tumor (T) tissues are indicated for comparison as available or excluded if normal tissue data were unavailable. Gene regulation (GR) upon detection was listed as Up or Down as relevant and according to UALCAN (U) and GEPIA (G). Survival (S) correlation with expression is listed as correlating (C). High expression = Low survival chance (HL) and High expression = High survival chance (HH).

subsequently selected with G418 sulfate (200 μg/ml; Invitrogen, USA). The cells were maintained in minimum essential media (MEM; Gibco, USA) supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 μg/ml streptomycin (Welgene, Daegu, Korea). All cell lines were cultivated at 37°C in a humid 5% CO_2 chamber and subcultured every 3 days after they reached 80% to 90% confluency. The cells were not subcultured beyond 20 passages.

RT-qPCR Analysis

RNA and proteins were extracted using Nucleospin® RNA/Protein (Macherey-Nagel, BMS, Korea) according to the manufacturer's instructions. cDNA was synthesized using AccuPower® RT/PCR PreMix (Bioneer, Korea). qPCR was performed with SYBR qPCR Mix (CellSafe, Yongin, Korea) on a CFX96 Real-Time System (Bio-Rad, BMS, Korea). Data were analyzed with the Pfaffl method (Pfaffl, 2001) using glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene as reference.

Immunoblot Analysis

Sample lysates were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis, and the separated bands were transferred onto polyvinylidene difluoride membranes. The membranes were blocked with 5% non-fat milk and incubated with MuD monoclonal antibody (Wagley et al., 2013) at 4°C overnight. The blots were subsequently incubated with a horseradish peroxidase (HRP)-labeled anti-human IgG at room temperature (15–20°C) for 2 h. The immunoreactive bands were detected with an enhanced chemiluminescence substrate (Dogen, Seoul, Korea), and band intensities were measured with ImageJ. All primers and antibodies used are listed in **Supplementary Table 1**.

RESULTS

Differential Expression of *MuD* in 14 Cancers Types

Our previous study suggested the involvement of *MuD* gene in the apoptotic pathway of the GBM cell line U251-MG induced by TRAIL (Choi et al., 2016). To investigate the role of *MuD* in cancer, we analyzed *MuD* expression data from the TCGA database (Cancer Genome Atlas Research et al., 2013), which included 35 different cancer types with normalized RNA expression for 33,096 cases as of December 2018. The analysis was performed with UALCAN, a web-based tool that facilitates in-depth analysis of the TCGA, and MET500 transcriptome databases (Chandrashekar et al., 2017) and GEPIA, which use the TCGA and GTEx projects databases to compare gene expression between tumor and normal tissues (Tang et al., 2017). Differential regulation of *MuD* gene expression was detected in six cancer types with UALCAN and 10 cancer types using GEPIA for a total of 14 cancer types. In most cases, the sample size was larger in the GEPIA database than in UALCAN. *MuD* expression in tumors was upregulated as compared with that in normal tissues in 9 of the 10 cancer types identified by GEPIA versus only two of the eight cancer types identified with UALCAN. *MuD* expression was downregulated only in kidney renal clear cell carcinoma (KIRC) tumor tissues, as per GEPIA analysis, but three additional cancer types were identified with UALCAN. Both of these tools detected *MuD* upregulation in the tumor tissues from patients with esophageal carcinoma (ESCA) and downregulation in patients with KIRC (**Table 1**). To validate if tumor purity of the TCGA tumors might affect the outcome, consensus measurement of purity (CPE) as previously described (Aran et al., 2015) was used to select GBM and KIRC-TCGA tumor >0.9 and 0.7 based on tumor purity distribution of the samples. Results consent with UALCAN results, suggesting that divergent result from UALCAN and GEPIA is not due to TCGA tumor quality (**Supplementary Figure 1**).

Validation Using Test Data Sets

As mentioned above, both tools revealed downregulation of *MuD* in KIRC tumor tissues but failed to detect any significant dysregulation in breast invasive carcinoma. To validate these findings, we selected two test sets from EMBL-EBI ArrayExpress database. Microarray data of patients with renal clear cell carcinoma (E-GEOD-53757) and breast adenocarcinoma (E-GEOD-70951) as control were analyzed. Except for two tumor samples, all tissues from patients with renal cancer showed significantly downregulated *MuD* expression levels relative to the matched normal tissues (**Figure 1A**). In contrast, *MuD* expression in the control data set was differently regulated in tumor tissue as compared with normal tissue (**Figure 1B**), indicative of the absence of any correlation between *MuD* expression and tumor identity. According to both portal, ESCA was up-regulated in tumor. This was validated using E-GEOD-23400 (**Supplementary Figure 2**).

MuD Expression Patterns Correlated With Survival in Three Different Cancer Types Based on TCGA Database

We investigated the correlation between *MuD* expression and patient survival in selected TCGA cohorts. Kaplan-Meier survival curves were generated using UALCAN for three cancer types based on the information in the TCGA database (**Figure 2**). UALCAN use statistical analysis that divided patients into two groups, comparing the higher quartile to the rest based on *MuD* expression. Among patients with invasive breast cancer (BRCA), those with high *MuD* expression (n = 271) revealed significantly lower survival ($p < 0.005$) than the controls (n = 810). In LUAD, a similar pattern was observed in the cohort characterized with high MuD expression (n = 128) relative to the corresponding controls (n = 374) ($p < 0.05$). In KIRC, survival was significantly lower in the cohort with high MuD expression (n = 134) ($p < 0.0001$) than in controls (n = 397) (**Figures 2A**, **B**, and **C** left plot). However, when equal number of samples were used to analyze survival chance, both BRCA and LUAD-TGCA lost their significance ($p > 0.1$) (**Figures 2A**, **B**, and **C**, middle plot). Validation using E-GEOD data shows that BRCA and LUAD

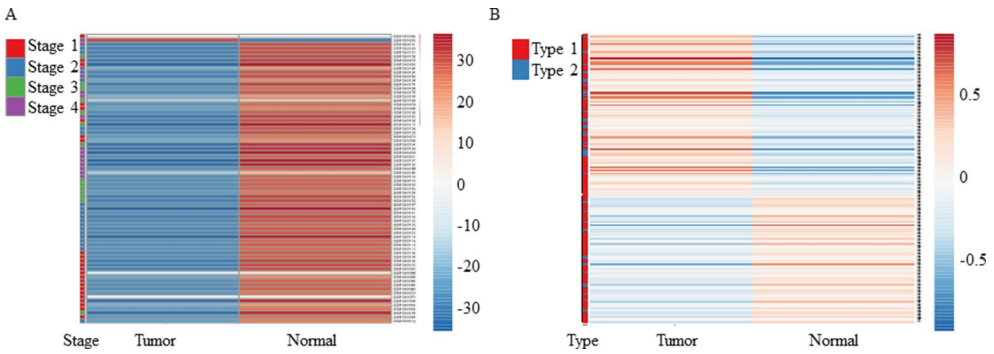


FIGURE 1 | Heatmap showing the expression of *MuD* from microarray data. *MuD* expression in kidney renal clear cell carcinoma **(A)** and adjacent kidney tissue as well as in breast cancer **(B)** and adjacent breast tissue. Clustering was performed with ClustVis using NIPALS PCA method. Red–white–blue scale was used to depict the normalized expression level. Red, blue, green, and violet color bars were used in clear cell renal carcinoma to represent classified stages, whereas red and blue bars were used in breast cancer to represent the diagnosis type.

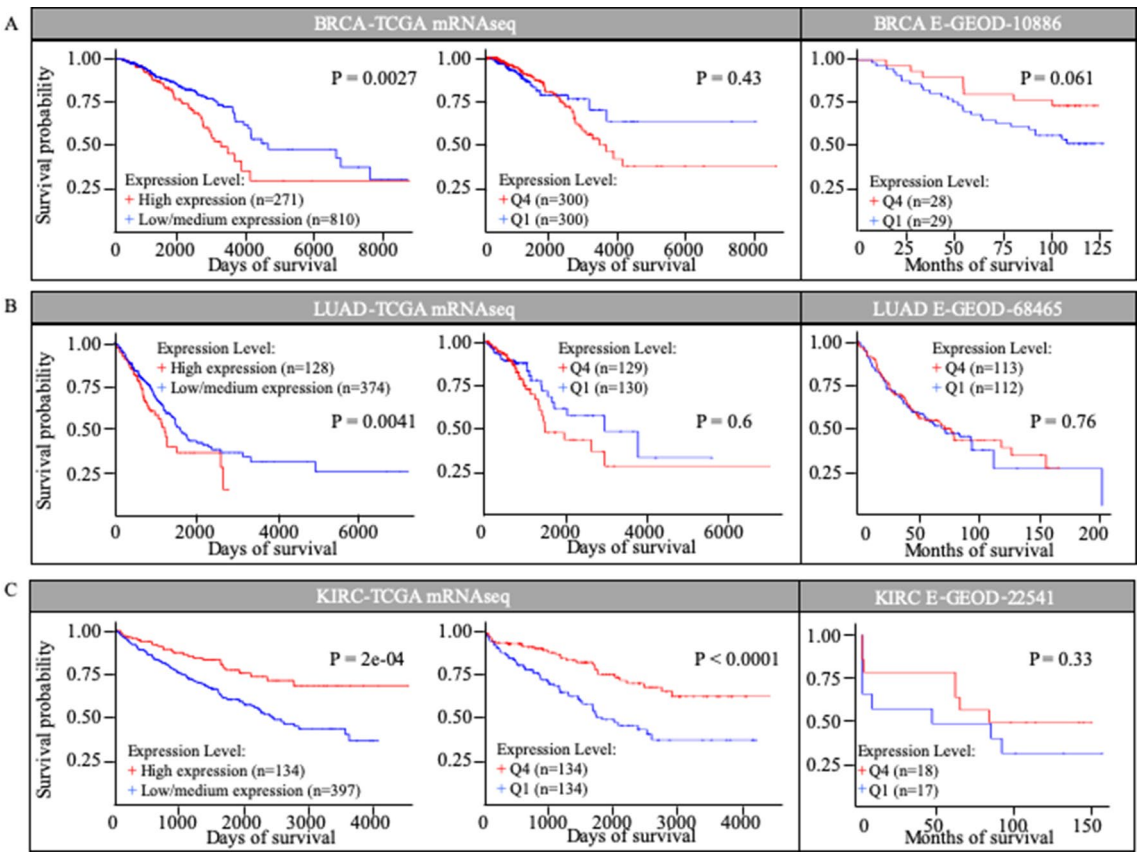


FIGURE 2 | Kaplan-Meier survival curves in TCGA cohorts and E-GEOD data sets. Survival curves based on MuD expression was plotted for breast invasive carcinoma (BRIA) **(A)**, lung adenocarcinoma (LUAD) **(B)**, and kidney renal clear cell carcinoma (KIRC) **(C)**.

outcomes from TCGA database are divergent from testing data set. Although KIRC data from E-GEOD-22514 lack vital status needed to analyze survival, this can be interfered from data of months free of tumor and total follow-up months as previously tested(Chang et al., 2018). Although *p* value was high due to the small numbers of samples, higher MuD expression correlated with higher survival, as it was for TCGA data (**Figure 2C**). Therefore, validation shows that although UALCAN is more sensitive in detecting potential correlation between expression and survival, GEPIA give a more robust outcome, possibly due to the fact that UALCAN analyze survival with unbalanced numbers of samples.

MuD Expression Pattern Correlated With High Survival in Proneural GBM Subtypes and Low Survival in Classical GBM Subtypes

Differential gene regulation in GBM tissues was detected with GEPIA but not UALCAN (**Table 1** and **Supplementary Figure 3**). As patients with GBM can be divided into subtypes with distinct molecular characteristics (Verhaak et al., 2010), we examined the survival of patients with different subtypes based on *MuD* expression. We analyzed 422 GBM samples available from TCGA. Patients were divided based on GBM subtype and further stratified into four quartiles based on *MuD* expression levels. For each subgroup, the Cox proportional hazard model was used to plot the survival of patients in the first quartile versus those in the fourth quartile, with age and *MGMT* methylation status as covariates (**Figure 3**). Interestingly, patients with proneural GBM from the fourth quartile showed significantly higher survival ($p < 0.005$), with a log-rank $p < 0.005$ and a hazard ratio (HR) value less than 1 (HR = 0.182). The age HR was slightly higher than 1 (HR = 1.05) and a high significance was observed ($p < 0.005$), suggesting that age may have a minor negative impact on the survival of patients with proneural GBM. The opposite results were observed in patients with classical subtype GBM, wherein the expression HR was higher than 1 (HR = 2.531) and a moderate significance was reported ($p < 0.1$). Interestingly, MGMT methylation was significantly more beneficial (HR = 4.67, $p < 0.05$) than age (HR = 1.005) in this group.

MuD Expression Correlated With That of *EXOC5*, *PPP2R2E*, and *SOS2* and *MuD* Overexpression Upregulated *PPP2R5E* and *SOS2*

We investigated the tumor-related genes in GBM tissues that showed correlation with *MuD* expression. Based on UALCAN results, we identified exocyst complex component 5 (*EXOC5*), protein phosphatase 2 regulatory subunit B'Epsilon (*PPP2R5E*), and son of sevenless homolog 2 (*SOS2*) to exhibit high correlations with *MuD* expression in GBM tumors from TCGA patients (Pearson's correlation coefficient > 0.79) (**Figure 4A**). Patient survival data based on *EXOC5*, *PPP2R5E*, and *SOS2* expression levels were available in the TCGA-GBM database, and the analysis with Cox proportional hazard model revealed that the high expression levels of these genes were associated with high survival in patients with proneural GBM at a log-rank *P* value cutoff of 0.05 (*PP2R5E* and *SOS2*) or close to 0.05 (*EXOC5*)

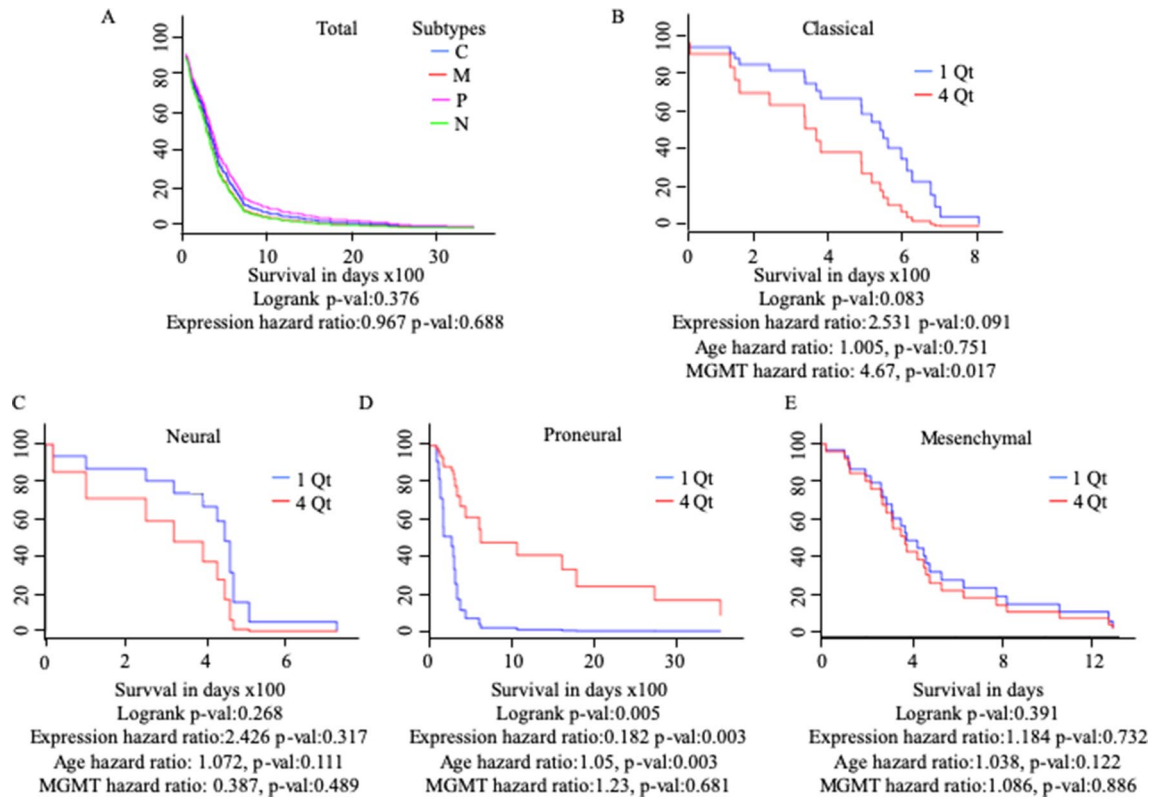


FIGURE 3 | Cox proportional analysis of 422 patients with GBM from the TCGA database divided into subtypes based on *MuD* expression level, as analyzed with GBM Bio Discovery Portal for total GBM **(A)**, Classical subtype **(B)**, Neural subtype **(C)**, Proneural subtype **(D)** and Mesenchymal subtype **(E)**. Patients were ranked into four quartiles based on *MuD* expression level. The survival rate of the patients from the first quartile with the lowest ranked *MuD* expression was compared to that of the patients from the fourth quartile. Age and *MGMT* promoter methylation status were used as covariates. C, classical; M, mesenchymal; P, proneural; N, neural; QT, quartile.

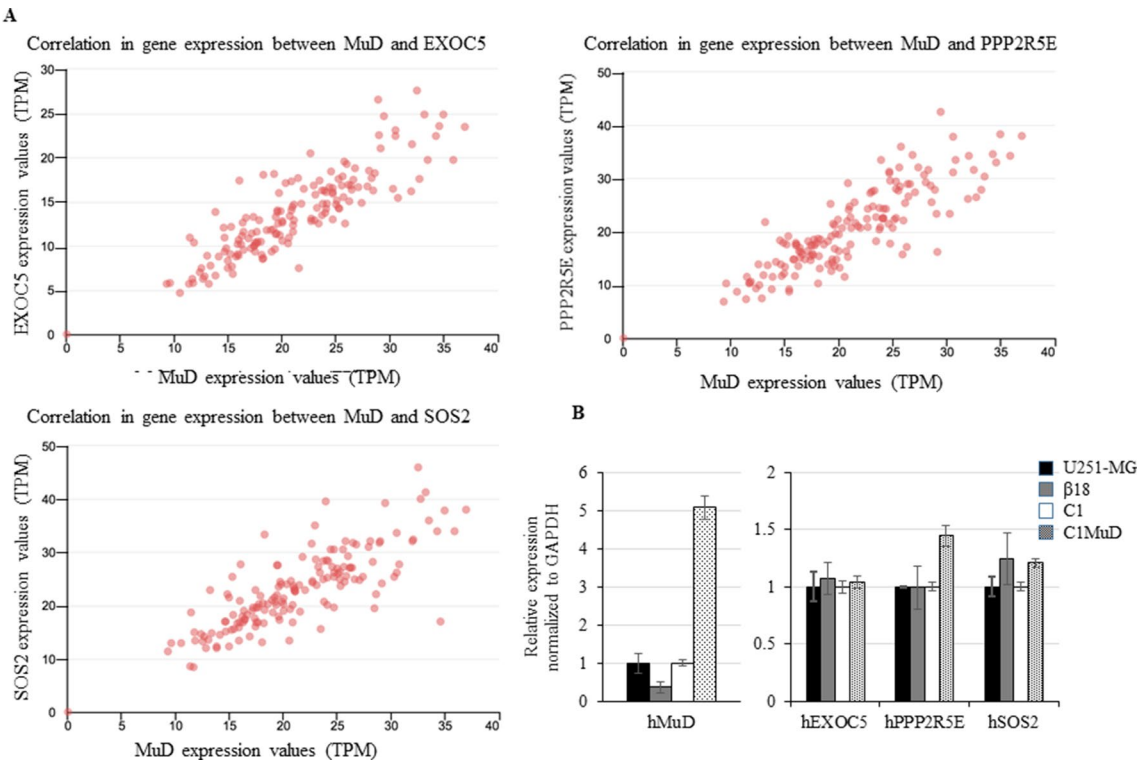


FIGURE 4 | Correlation between *EXOC5*, *PPP2R5E*, and *SOS2* expression levels and *MuD* levels. Expression in GBM tissues from TCGA database **(A)** and expression of *MuD*, *EXOC5*, *PPP2R5E*, and *SOS2* in the GBM *MuD* KO line β18 and *MuD* overexpression line C1MuD as compared with that in their respective controls **(B)**.

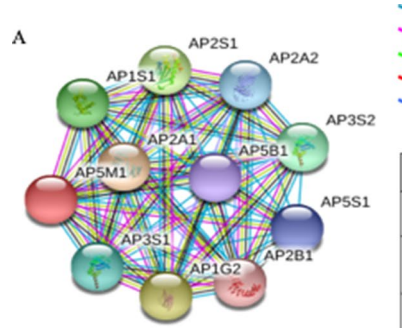
(**Supplementary Figure 4**). Correlation in BRCA, KIRC, LUAD, ESCA, and CHOL were also displayed as a table for the upmost linked 10 genes (**Supplementary Table 2**). EXOC5 shows high correlation with MuD in all subtypes except KIRC. To further investigate the correlation between *MuD* and these genes, we used a GBM cell line, U251-MG, a CRISPR-Cas9-generated *MuD* KO line β18, a plasmid transfection line containing pEGFP-C1 (C1), and a line SE *MuD* following MuD-GFP-C1 transfection (C1MuD). Although *MuD* KO failed to affect the expression of *EXOC5*, *PPO2R5E*, and *SOS2*, *MuD* stable expression increased *PPP2R5E* and *SOS2* expression levels to some extent (**Figure 4B**). The expression of these genes was also investigated in the biospecimens mentioned below, but the Pearson's correlation coefficient was insignificant, probably owing to the small sample number (**Supplementary Figure 5**).

Correlation Analysis Suggested the Possible Interactions Between *MuD* and Other Proteins That Affected Prognosis in Patients With GBM

To examine the network of proteins that potentially interact with MuD, we used STRING (Szklarczyk et al., 2017) (**Figure 5**). The identified interactors were found to be other components of the AP complexes. Survival data were unavailable for the identified proteins, AP5B1, AP5S1, AP5Z1 (other putative components of the fifth AP complex) and AP4B1 and AP4S1 (components of the fourth AP complex). The components of the fourth AP complex showed opposite patterns in terms of survival and MuD expression, as the first quartile patients with AP4M1 had higher chances of survival. Another component of the fourth complex, AP4E1 showed a similar pattern with MuD in proneural subclass but revealed a different pattern in the mesenchymal subclass, wherein survival chance was highly correlated with the fourth quartile patients of AP4E1. Components of the first AP complex (AP1G2 and AP1S1) as well as the third (AP3S2) and fourth (AP4S1) complexes showed no correlation between expression and survival in patients with proneural GBM (**Supplementary Figure 6**). This discrepancy among components of similar complexes suggests the possibility of additional roles of MuD aside from its involvement with the component of AP5. BioGrid (Chatr-Aryamontri et al., 2017) shows that besides other AP complex subunits, DDB1 and CUL4-associated factor 4 (DCAF4), spatacsin vesicle trafficking-associated (SPG11), general transcription factor IIIC subunit 3 (GTF3C3), and paired box protein PAX-6 (PAX6) are candidate protein interactors for MuD in human cells. Further investigation is warranted to validate the interactions of these components with MuD.

Expression Analysis Revealed Putative Post-Translational Regulation of *MuD*

To investigate MuD expression patterns in human brain tumor tissues, we isolated RNA and protein from six tumor



Known interaction: from curated databases
Known interaction: experimentally determined
Predicted interaction: gene neighborhood
Predicted interaction: gene fusions
Predicted interaction: gene co-occurence

Protein	Full name		Size	Score
AP5B1	AP-5 complex subunit beta-1	Armadillo-like helical domain containing.	878 aa	0.989
AP5S1	AP-5 complex subunit sigma-1;	Required for efficient homologous recombination DNA double-strand break repair.	200 aa	0.989
AP4M1	AP-4 complex subunit mu-1	Control the vesicular transport of proteins in different trafficking pathways.	453 aa	0.983
AP5Z1	AP-5 complex subunit zeta-1	Armadillo-like helical domain containing.	807 aa	0.982
AP4B1	AP-4 complex subunit beta-1	May be involved in the targeting of proteins from the trans-Golgi network (TGN) to the endosomal-lysosomal system.	739 aa	0.944
AP1G2	AP-1 complex subunit gamma-like 2	May function in protein sorting in late endosomes or multivesucular bodies (MVBs). Involved in MVB-assisted maturation of hepatitis B virus (HBV)	785 aa	0.943
AP4E1	AP-4 complex subunit epsilon-1	May be involved in the targeting of proteins from the TGN to the endosomal-lysosomal system.	1137 aa	0.943
AP1S1	AP-1 complex subunit sigma-1A	Plays a role in protein sorting in the late TGN and/or endosomes.	158 aa	0.93
AP4S1	AP-4 complex subunit sigma-1	May be involved in the targeting of proteins from the TGN to the endosomal-lysosomal system.	159 aa	0.93
AP3S2	AP-3 complex subunit sigma-2	Required to target cargos into vesicles assembled at cell bodies for delivery into neurites and nerve terminals.	193 aa	0.929

FIGURE 5 | Analysis of candidate proteins interacting with MuD. STRING analysis showing interactions between MuD (AP5M1) and other adaptor protein complex subunits **(A)**. BioGRID analysis revealing the additional putative MuD interacting proteins **(B)**.

biospecimens (four GBMs, one oligodendroglioma, and one ependymoma) and matched normal tissues (**Table 2**). Of these, four pairs exhibited MuD upregulation in the tumors as compared with that in the matched normal tissues, with high significance in three samples (**Figure 6A**, lane 3, 4, and 6). Two additional GBM samples (**Figure 4B**, lanes 7 and 9) and oligodendroglioma and ependymoma samples (**Figure 6B**, lanes 11, and 12) showed higher MuD expression than normal tissues. These findings imply that MuD expression may be upregulated in brain tumors, at least in GBM and other glioma tissues. Interestingly, neither MuD mRNA and protein levels nor subtypes showed any significant correlation, as MuD was only upregulated in patient 5 (**Figure 7A**). All but one (NEFL) subtype marker showed upregulation in patient 5 (**Figure 7B**) as compared with that in the other patients (**Figure 7C**), suggestive of the possible failure of gene downregulation in the tumor from this patient. As MuD protein expression was upregulated in at least four patients, there is a possibility of putative post-translational control of MuD in

TABLE 2 | Clinical characteristics of patients with brain tumor in this study. GB, glioblastoma; ODG, oligodendroglioma; ED, ependymoma.

Patients	Tissue	Tissue bcode	Diagnosis	WHO grade	Age	Gender	Tissue	Comment
1	ANC-13-0005	25502083	GB	4	81–85	Male	Pair	High necrosis rate
2	ANC-13-0028	25502722	GB	4	81–85	Female	Pair	–
3	ANC-13-0018	25680967	GB	4	61–65	Female	Pair	–
4	ANC-13-0019	25371685	GB	4	61–65	Male	Pair	–
5	ANC-13-0027	25333994	GB	4	71–75	Female	Pair	–
6	ANC-13-0052	25072638	GB	4	56–60	Female	Pair	–
7	ANC-13-0014	25502189	GB	4	71–75	Male	Tumor	–
8	ANC-13-0022	25502458	GB	4	76–80	Female	Tumor	–
9	ANC-14-0032	25756981	GB	4	41–45	Male	Tumor	–
10	ANC-15-0009	25227244	GB	4	56–60	Male	Tumor	–
11	ANC-14-0041	25537678	ODG	2	36–40	Female	Tumor	Low tumor cells percentage
12	ANC-15-0061	25022741	ED	2	51–55	Female	Tumor	–

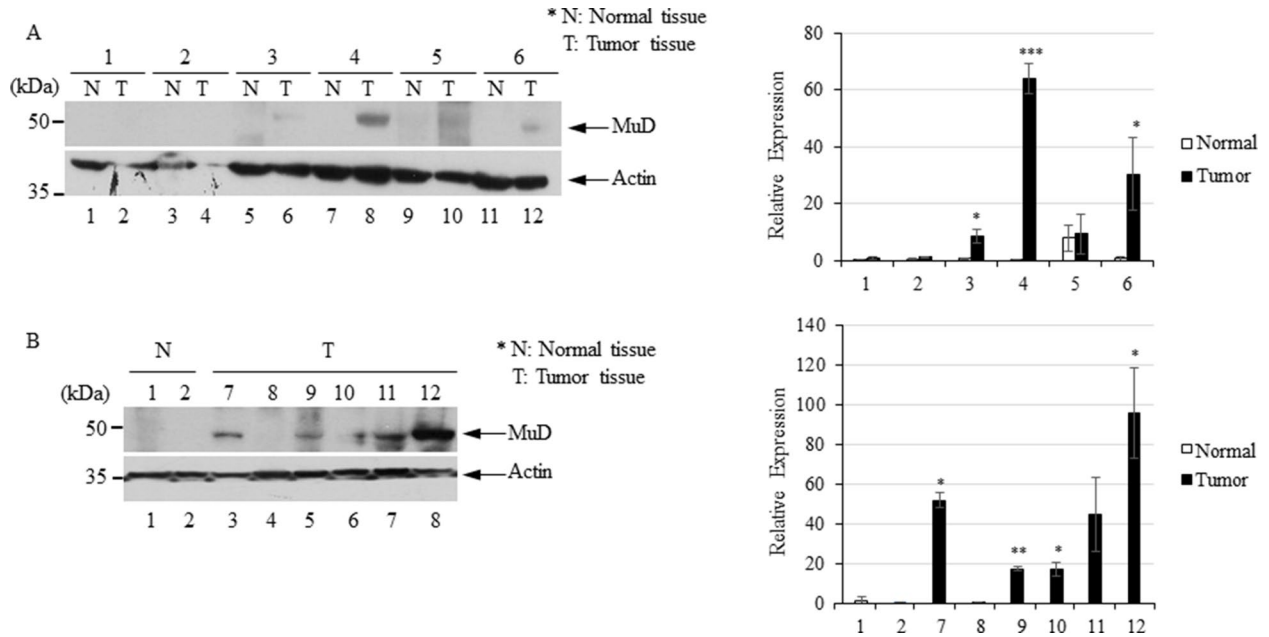


FIGURE 6 | Immunoblot analysis of MuD expression in six pairs of GBM **(A)** and GBM and astrocytoma biospecimens **(B)**. N, normal tissue; T, tumor tissue $^{*}p < 0.01$, $^{**}p < 0.005$, $^{***}p < 0.001$.

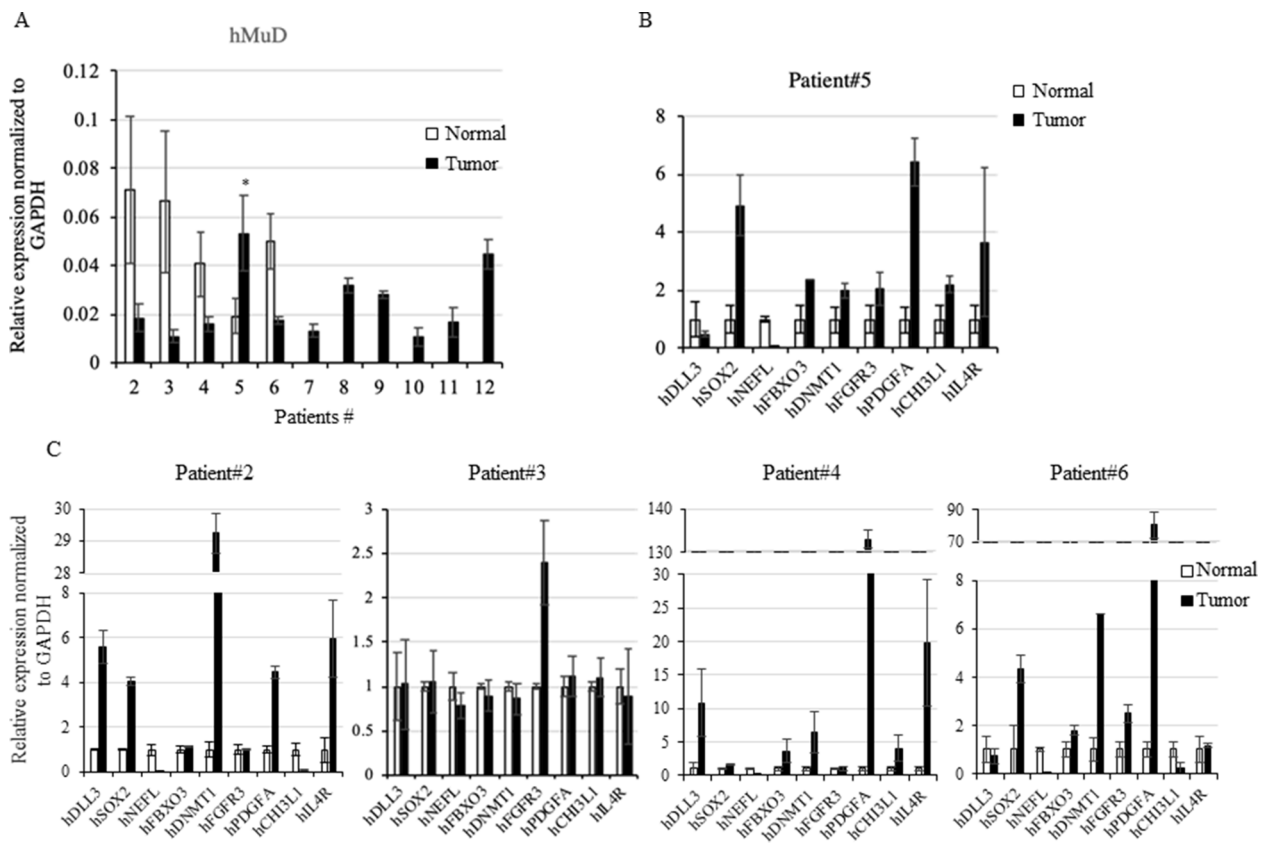


FIGURE 7 | Rt-qPCR analysis of *MuD* and subtypes marker genes expression. Expression of *MuD* mRNA analyzed with RT-qPCR in patient tissues across GBM and astrocytoma brain tumor tissues **(A)**. Expression of subtype markers in tissues from patient 5 **(B)** and other patients expression of subtype markers **(C)**. mRNA expression was normalized to *GAPDH* levels ($^{*}p < 0.05$).

brain tumors without ruling out the chances of contamination from neighboring tissues.

DISCUSSION

MuD was identified in a screening approach aimed to reveal any novel genes involved in Fas-mediated apoptosis (Kawasaki and Taira, 2002). *MuD* plays a specific role in several cancer cells (Lee et al., 2008; Choi et al., 2016), and is thought to exert its functions through apoptotic signaling, considering its cleavage by active caspase-3 upon TRAIL stimulation (Shin et al., 2013). However, the detailed roles of *MuD* in tumors remain to be elucidated. Herein, our findings based on the careful analyses of both metadata and data from patients with brain tumor suggest that any alterations in *MuD* expression could be associated with tumor progression and survival in selected cancer types.

We used two web-based portals to evaluate the potential role of *MuD* in cancer. Although the key data source for both UALCAN and GEPIA data is TCGA, only two cancer types were detected in both portals, including KRIC and ESCA. This discrepancy between the two portals may be related to the different data processing method or owing to the use of additional data from MET500 and GTEx projects, respectively. Although more cancer types were found to exhibit *MuD* upregulation in tumors using GTEx, UALCAN was more prone to detect cancer types and survival correlation with *MuD* downregulation. This result may be partially attributed to the fact that GTEx projects collected more samples, including healthy tissue data, for most cancer types, whereas both MET500 and GTEx projects had no data on patient survival. As a consequence, drastic changes in p and log-rank values were reported. Nevertheless, we identified two cancer types, wherein *MuD* may exhibit a role in tumor development and serve as a potential biomarker. Our results also demonstrate that although large-scale data analysis may be useful in finding novel oncogenes and new biomarkers, the data should be carefully validated.

In a previous study, we reported the anti-apoptotic function of *MuD* in GBM cell lines (Choi et al., 2016) and investigated the correlation between *MuD* expression and cancer progression. We failed to observe any correlation between *MuD* expression and overall survival in patients with GBM. However, GEPIA analysis suggested the upregulation of *MuD* in GBM tumors, and MuD protein levels were markedly upregulated in human brain tumor tissues, including 10 GBMs. *MuD* mRNA level from the same tissues showed no correlation except one at $p < 0.05$. Although this observation may be related to the small number of biospecimens investigated, additional regulation of *MuD* may occur at the post-translational level. This hypothesis is consistent with our previous finding that MuD was downregulated following TRAIL stimulation without any alteration in the *MuD* mRNA level (Choi et al., 2016).

Proneural GBM differs from other GBM subgroups with respect to gene expression patterns and responses to drug treatment (Chen and Xu, 2016). The proneural GBM cohort showed significantly improved prognosis as compared with patients with other subtypes (Verhaak et al., 2010) but failed to respond to immunotherapy as efficiently as the mesenchymal GBM cohort, presumably owing to TGF-R2 deficiency (Beier et al., 2012). Our findings showed that higher MuD expression levels were associated with prolonged survival in patients with proneural GBM; however, this correlation could not be extended to all patients with GBM. Although we failed to notice any correlation between GBM subtypes and MuD expression in our biospecimens, probably owing to the small sample size and the markers investigated, *MuD* expression might exert differential effects based on GBM subtypes and *MuD* may serve as a potential target gene specifically for the treatment of proneural subtype.

Analyses of MuD protein level, localization, and interactions with other putative proteins suggest its importance as a component of the putative AP5 complex (Hirst et al., 2011). Clathrin AP complexes play crucial roles in protein sorting in diverse post-Golgi pathways and are involved in endocytosis (McMahon and Boucrot, 2011). In particular, the AP1 complex is involved in trafficking between the trans-Golgi network (TGN) and endosomes (Hirst et al., 2012), AP2 is associated with endocytosis (McMahon and Boucrot, 2011), and AP3 mediates trafficking between the TGN/endosome and the vacuole/lysozyme system (Dell'Angelica, 2009). AP4 was thought to play a role in vacuolar sorting in plant cells (Fuji et al., 2016) and interact with Tepsin (Frazier et al., 2016). AP complexes are involved in several diseases, including X-linked mental retardation (Tarpey et al., 2006), Alzheimer's disease (Burgos et al., 2010), and Hermansky–Pudlak syndrome (Dell'Angelica et al., 1999). Whereas these complexes showed no correlation with MuD, the two putative partner proteins of MuD, AP4M1, and AP4E1, are components of the AP4 complex and showed correlation with cancer prognosis. A recent report showed that AP4 promotes oncogenic phenotype and drug resistance in breast cancer through the regulation of a novel oncogene, lysosomal-associated protein transmembrane-4 beta (*LAPTM4B*) (Wang et al., 2018), and induces prostate cancer proliferation though L-plastin regulation (Chen et al., 2017). These studies suggest that AP complexes may play a role in cancer cell proliferation.

Aside from its role as a component of the AP5 complex, MuD is involved in cancer pathogenesis (Merino et al., 2007; Johnstone et al., 2008; Cullen and Martin, 2015). MuD is implicated in TRAIL-induced apoptosis signaling (Lee et al., 2008; Shin et al., 2013; Choi et al., 2016). Studies have shown BID and Bcl2 as molecules acting upstream and downstream of MuD, respectively (Choi et al., 2016), suggesting that MuD may perform a novel role in the cancer apoptotic pathway. Our analysis identified several candidate genes, including *EXOC5*, *PPP2R5E*, and *SOS2*, and showed that *PPP2R5E* and *SOS2* expression levels correlated with MuD level in tumor cells to some extent. We failed to detect any correlation between *MuD* and *EXOC5* expression; however, *EXOC5* is adjacent to *MuD* (*EXOC5* 5′-UTR starts at Chr14: 57,268,899 and *MuD* 5′-UTR starts at Chr14: 57,268,888). As *MuD* KO or stable expression showed no effect on *EXOC5* expression, there is a possibility

that the correlation between *EXOC5* and *MuD* expression may be related to the positional effect. A previous study using a KO mouse model showed that the deletion of *EXOC5* led to apoptosis and disorganization of hair cell stereocilia bundles (Lee et al., 2018). *PPPR2RE* is known as a tumor suppressor gene, and its downregulation induces growth inhibition and apoptosis in gastric cancer cells (Liu et al., 2014). *SOS2* encodes a Ras-specific guanine nucleotide exchange factor (Esteban et al., 2000), and its downregulation decreases the level of Ras and activation of MAP kinase kinase1/2 (MEKK1/2), ultimately inhibiting TNFα-induced apoptosis (Kurada et al., 2009). Not only *PPP2R5E* and *SOS2* expression levels slightly correlated with that of *MuD* in GBM cell lines but also high expression of these genes was associated with longer survival among patients with proneural GBM at a moderate level (**Supplementary Figure 7**). Further study to validate the involvement of these genes in tumor generation linked to *MuD* is in progress.

Based on the database analyses, we propose that *MuD* expression may be upregulated in ESCA and downregulated in KIRC. Further studies should carefully validate these results to evaluate *MuD* as a biomarker with a putative prognostic role. In addition, *MuD* may play a role in the survival of patients with proneural GBM and could be linked to candidate gene regulation. Taken together, our study suggests a novel role for *MuD* in cancer.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The biospecimens utilized in the present study were provided by the Ajou Human Bio-Resource Bank (Suwon, Korea), a member of the National Biobank of Korea, which is supported by the Ministry of Health and Welfare. All samples derived from the National Biobank of Korea were obtained with informed consent under institutional review board-approved protocols. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

J-WO and JS conceived, designed and wrote the manuscript. Computational analysis was performed by JS. J-HC and JS performed biospecimen analysis. Samples were diagnosed and contributed by JA and S-HK. All authors mentioned above and SJ, JO, D-YY, MR critically reviewed and approved the final version of the manuscript.

FUNDING

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-S201801S00057 (JWO); NRF-2016R1D1A1B03935382 (JWO); NRF-S201806S00067 (JS)).

ACKNOWLEDGMENTS

The biospecimens and data used for this study were provided by the Biobank of Ajou University Hospital, a member of Korea Biobank Network. We sincerely acknowledge the public database: TCGA, GEO, GTEx and MET500 and Ajou Human Bio-Resource Bank for biospecimens.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00884/full#supplementary-material

REFERENCES

Akan, P., Alexeyenko, A., Costea, P. I., Hedberg, L., Solnestam, B. W., Lundin, S., et al. (2012). Comprehensive analysis of the genome transcriptome and proteome landscapes of three tumor cell lines. *Genome Med.* 4 (11), 86. doi: 10.1186/gm387

Aran, D., Sirota, M., and Butte, A. J. (2015). Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* 6, 8971. doi: 10.1038/ncomms9971

Beier, C. P., Kumar, P., Meyer, K., Leukel, P., Bruttel, V., Aschenbrenner, I., et al. (2012). The cancer stem cell subtype determines immune infiltration of glioblastoma. *Stem Cells Dev.* 21 (15), 2753–2761. doi: 10.1089/scd.2011.0660

Burgos, P. V., Mardones, G. A., Rojas, A. L., daSilva, L. L., Prabhu, Y., Hurley, J. H., et al. (2010). Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. *Dev. Cell* 18 (3), 425–436. doi: 10.1016/j.devcel.2010.01.015

Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45 (10), 1113–1120. doi: 10.1038/ng.2764

Celiku, O., Johnson, S., Zhao, S., Camphausen, K., and Shankavaram, U. (2014). Visualizing molecular profiles of glioblastoma with GBM-BioDP. *PLoS One* 9 (7), e101239. doi: 10.1371/journal.pone.0101239

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19 (8), 649–658. doi: 10.1016/j.neo.2017.05.002

Chang, P., Bing, Z., Tian, J., Zhang, J., Li, X., Ge, L., et al. (2018). Comprehensive assessment gene signatures for clear cell renal cell carcinoma prognosis. *Med. (Baltimore)* 97 (44), e12679. doi: 10.1097/MD.0000000000012679

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N. K., et al. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* 45 (D1), D369–D379. doi: 10.1093/nar/gkw1102

Chen, C., Cai, Q., He, W., Lam, T. B., Lin, J., Zhao, Y., et al. (2017). AP4 modulated by the PI3K/AKT pathway promotes prostate cancer proliferation and metastasis of prostate cancer *via* upregulating l-plastin. *Cell Death Dis.* 8 (10), e3060. doi: 10.1038/cddis.2017.437

Chen, Y., and Xu, R. (2016). Drug repurposing for glioblastoma based on molecular subtypes. *J. Biomed. Inform.* 64, 131–138. doi: 10.1016/j.jbi.2016.09.019

Choi, J. H., Lim, J. B., Wickramanayake, D. D., Wagley, Y., Kim, J., Lee, H. C., et al. (2016). Characterization of MUDENG, a novel anti-apoptotic protein. *Oncogenesis* 5, e221. doi: 10.1038/oncsis.2016.30

Clough, E., and Barrett, T. (2016). The gene expression omnibus database. *Methods Mol. Biol.* 1418, 93–110. doi: 10.1007/978-1-4939-3578-9_5

Consortium, G. T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi: 10.1038/ng.2653

Cullen, S. P., and Martin, S. J. (2015). Fas and TRAIL 'death receptors' as initiators of inflammation: implications for cancer. *Semin. Cell Dev. Biol.* 39, 26–34. doi: 10.1016/j.semcdb.2015.01.012

Dell'Angelica, E. C. (2009). AP-3-dependent trafficking and disease: the first decade. *Curr. Opin. Cell Biol.* 21 (4), 552–559. doi: 10.1016/j.ceb.2009.04.014

Dell'Angelica, E. C., Shotelersuk, V., Aguilar, R. C., Gahl, W. A., and Bonifacino, J. S. (1999). Altered trafficking of lysosomal proteins in Hermansky–Pudlak syndrome due to mutations in the beta 3A subunit of the AP-3 adaptor. *Mol. Cell* 3 (1), 11–21. doi: 10.1016/S1097-2765(00)80170-7

Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma , Shedden, K., Taylor, J. M., Enkemann, S. A., Tsao, M. S., Yeatman, T. J., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* 14 (8), 822–827. doi: 10.1038/nm.1790

Esteban, L. M., Fernandez-Medarde, A., Lopez, E., Yienger, K., Guerrero, C., Ward, J. M., et al. (2000). Ras-guanine nucleotide exchange factor sos2 is dispensable for mouse growth and development. *Mol. Cell Biol.* 20 (17), 6410–6413. doi: 10.1128/MCB.20.17.6410-6413.2000

Frazier, M. N., Davies, A. K., Voehler, M., Kendall, A. K., Borner, G. H., Chazin, W. J., et al. (2016). Molecular basis for the interaction between AP4 beta4 and its accessory protein, Tepsin. *Traffic* 17 (4), 400–415. doi: 10.1111/tra.12375

Friedmann-Morvinski, D. (2014). Glioblastoma heterogeneity and cancer cell plasticity. *Crit. Rev. Oncog.* 19 (5), 327–336. doi: 10.1615/CritRevOncog.2014011777

Fuji, K., Shirakawa, M., Shimono, Y., Kunieda, T., Fukao, Y., Koumoto, Y., et al. (2016). The adaptor complex AP-4 regulates vacuolar protein sorting at the trans-Golgi Network by interacting with vacuolar sorting receptor1. *Plant Physiol.* 170 (1), 211–219. doi: 10.1104/pp.15.00869

Hirst, J., Barlow, L. D., Francisco, G. C., Sahlender, D. A., Seaman, M. N., Dacks, J. B., et al. (2011). The fifth adaptor protein complex. *PLoS Biol.* 9 (10), e1001170. doi: 10.1371/journal.pbio.1001170

Hirst, J., Borner, G. H., Antrobus, R., Peden, A. A., Hodson, N. A., Sahlender, D. A., et al. (2012). Distinct and overlapping roles for AP-1 and GGAs revealed by the "knocksideways" system. *Curr. Biol.* 22 (18), 1711–1716. doi: 10.1016/j.cub.2012.07.012

Hovinga, K. E., McCrea, H. J., Brennan, C., Huse, J., Zheng, J., Esquenazi, Y., et al. (2019). EGFR amplification and classical subtype are associated with a poor response to bevacizumab in recurrent glioblastoma. *J. Neurooncol.* 142 (2), 337–345. doi: 10.1007/s11060-019-03102-5

Huse, J. T., and Holland, E. C. (2010). Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. *Nat. Rev. Cancer* 10 (5), 319–331. doi: 10.1038/nrc2818

Inda, M. M., Bonavia, R., and Seoane, J. (2014). Glioblastoma multiforme: a look inside its heterogeneous nature. *Cancers (Basel)* 6 (1), 226–239. doi: 10.3390/cancers6010226

Jo, M., Kim, T. H., Seol, D. W., Esplen, J. E., Dorko, K., Billiar, T. R., et al. (2000). Apoptosis induced in normal human hepatocytes by tumor necrosis factor-related apoptosis-inducing ligand. *Nat. Med.* 6 (5), 564–567. doi: 10.1038/75045

Johnstone, R. W., Frew, A. J., and Smyth, M. J. (2008). The TRAIL apoptotic pathway in cancer onset, progression and therapy. *Nat. Rev. Cancer* 8 (10), 782–798. doi: 10.1038/nrc2465

Jung, K. W., Won, Y. J., Kong, H. J., Oh, C. M., Lee, D. H., and Lee, J. S. (2014). Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2011. *Cancer Res. Treat.* 46 (2), 109–123. doi: 10.4143/crt.2014.46.2.109

Kawasaki, H., and Taira, K. (2002). A functional gene discovery in the Fas-mediated pathway to apoptosis by analysis of transiently expressed randomized hybrid-ribozyme libraries. *Nucleic Acids Res.* 30 (16), 3609–3614. doi: 10.1093/nar/gkf476

Kurada, B. R., Li, L. C., Mulherkar, N., Subramanian, M., Prasad, K. V., and Prabhakar, B. S. (2009). MADD, a splice variant of IG20, is indispensable for MAPK activation and protection against apoptosis upon tumor necrosis factor-alpha treatment. *J. Biol. Chem.* 284 (20), 13533–13541. doi: 10.1074/jbc.M808554200

Lee, B., Baek, J. I., Min, H., Bae, S. H., Moon, K., Kim, M. A., et al. (2018). Exocyst complex member EXOC5 is required for survival of hair cells and spiral ganglion neurons and maintenance of hearing. *Mol. Neurobiol.* 55 (8), 6518–6532. doi: 10.1007/s12035-017-0857-z

Lee, M. R., Shin, J. N., Moon, A. R., Park, S. Y., Hong, G., Lee, M. J., et al. (2008). A novel protein, MUDENG, induces cell death in cytotoxic T cells. *Biochem. Biophys. Res. Commun.* 370 (3), 504–508. doi: 10.1016/j.bbrc.2008.03.139

Li, M., Xiao, A., Floyd, D., Olmez, I., Lee, J., Godlewski, J., et al. (2017). CDK4/6 inhibition is more active against the glioblastoma proneural subtype. *Oncotarget* 8 (33), 55319–55331. doi: 10.18632/oncotarget.19429

Liu, X., Liu, Q., Fan, Y., Wang, S., Liu, X., Zhu, L., et al. (2014). Downregulation of PPP2R5E expression by miR-23a suppresses apoptosis to facilitate the growth of gastric cancer cells. *FEBS Lett.* 588 (17), 3160–3169. doi: 10.1016/j.febslet.2014.05.068

McMahon, H. T., and Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* 12 (8), 517–533. doi: 10.1038/nrm3151

Merino, D., Lalaoui, N., Morizot, A., Solary, E., and Micheau, O. (2007). TRAIL in cancer therapy: present and future challenges. *Expert. Opin. Ther. Targets* 11 (10), 1299–1314. doi: 10.1517/14728222.11.10.1299

Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43 (W1), W566–W570. doi: 10.1093/nar/gkv468

Ostrom, Q. T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., et al. (2014). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro. Oncol.* 16 Suppl 4, iv1–iv63. doi: 10.1093/neuonc/nou223

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27 (8), 1160–1167. doi: 10.1200/JCO.2008.18.1370

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29 (9), e45. doi: 10.1093/nar/29.9.e45

Phillips, H. S., Kharbanda, S., Chen, R., Forrest, W. F., Soriano, R. H., Wu, T. D., et al. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9 (3), 157–173. doi: 10.1016/j.ccr.2006.02.019

Quigley, D. A., Tahiri, A., Luders, T., Riis, M. H., Balmain, A., Borresen-Dale, A. L., et al. (2017). Age, estrogen, and immune response in breast adenocarcinoma and adjacent normal tissue. *Oncoimmunology* 6 (11), e1356142. doi: 10.1080/2162402X.2017.1356142

Robinson, D. R., Wu, Y. M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., et al. (2017). Integrative clinical genomics of metastatic cancer. *Nature* 548 (7667), 297–303. doi: 10.1038/nature23306

Shin, J. N., Han, J. H., Kim, J. Y., Moon, A. R., Kim, J. E., Chang, J. H., et al. (2013). MUDENG is cleaved by caspase-3 during TRAIL-induced cell death. *Biochem. Biophys. Res. Commun.* 435 (2), 234–238. doi: 10.1016/j.bbrc.2013.04.075

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics, 2017. *CA Cancer J. Clin.* 67 (1), 7–30. doi: 10.3322/caac.21387

Stuckey, D. W., and Shah, K. (2013). TRAIL on trial: preclinical advances in cancer therapy. *Trends Mol. Med.* 19 (11), 685–694. doi: 10.1016/j.molmed.2013.08.007

Su, H., Hu, N., Yang, H. H., Wang, C., Takikita, M., Wang, Q. H., et al. (2011). Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin. Cancer Res.* 17 (9), 2955–2966. doi: 10.1158/1078-0432.CCR-10-2724

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi: 10.1093/nar/gkx247

Tarpey, P. S., Stevens, C., Teague, J., Edkins, S., O'Meara, S., Avis, T., et al. (2006). Mutations in the gene encoding the Sigma 2 subunit of the adaptor protein 1

complex, AP1S2, cause X-linked mental retardation. *Am. J. Hum. Genet.* 79 (6), 1119–1124. doi: 10.1086/510137

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Pozn.)* 19 (1A), A68–A77. doi: 10.5114/wo.2014.47136

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17 (1), 98–110. doi: 10.1016/j.ccr.2009.12.020

von Roemeling, C. A., Radisky, D. C., Marlow, L. A., Cooper, S. J., Grebe, S. K., Anastasiadis, P. Z., et al. (2014). Neuronal pentraxin 2 supports clear cell renal cell carcinoma by activating the AMPA-selective glutamate receptor-4. *Cancer Res.* 74 (17), 4796–4810. doi: 10.1158/0008-5472.CAN-14-0210

Wagley, Y., Choi, J. H., Wickramanayake, D. D., Choi, G. Y., Kim, C. K., Kim, T. H., et al. (2013). A monoclonal antibody against human MUDENG protein. *Monoclon. Antib. Immunodiagn. Immunother.* 32 (4), 277–282. doi: 10.1089/mab.2013.0015

Wang, L., Meng, Y., Xu, J. J., and Zhang, Q. Y. (2018). The transcription factor AP4 promotes oncogenic phenotypes and cisplatin resistance by regulating LAPTM4B expression. *Mol. Cancer Res.* 16 (5), 857–868. doi: 10.1158/1541-7786.MCR-17-0519

Wuttig, D., Zastrow, S., Fussel, S., Toma, M. I., Meinhardt, M., Kalman, K., et al. (2012). CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed by genome-wide expression analyses of primary tumors and metastases. *Int. J. Cancer* 131 (5), E693–E704. doi: 10.1002/ijc.27419

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

CASE REPORT
published: 15 October 2019
doi: 10.3389/fonc.2019.01026

Transcriptomics-Guided Personalized Prescription of Targeted Therapeutics for Metastatic ALK-Positive Lung Cancer Case Following Recurrence on ALK Inhibitors

Elena Poddubskaya[1,2], Alexey Bondarenko[1], Alexander Boroda[2], Evgenia Zotova[2], Alex Glusker[2], Svetlana Sletina[1], Luidmila Makovskaia[3], Philipp Kopylov[2,4], Marina Sekacheva[2], Alexey Moisseev[2] and Madina Baranova[1,4]*

[1] Clinical Center Vitamed, Moscow, Russia, [2] I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia, [3] Faculty of Fundamental Medicine, Lomonosov Moscow State University, Moscow, Russia, [4] FSBEI FPE Russian Medical Academy of Continuing Professional Education MOH, Moscow, Russia

OPEN ACCESS

Edited by:
Ira Ida Skvortsova, Innsbruck Medical University, Austria

Reviewed by:
Giulio Rossi, Azienda Unità Sanitaria Locale (AUSL) della Romagna, Italy
Xiaofei Song, University of Texas MD Anderson Cancer Center, United States

***Correspondence:**
Alexey Moisseev
moisseevalexey@hotmail.com

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *02 August 2019*
Accepted: *23 September 2019*
Published: *15 October 2019*

Citation:
Poddubskaya E, Bondarenko A, Boroda A, Zotova E, Glusker A, Sletina S, Makovskaia L, Kopylov P, Sekacheva M, Moisseev A and Baranova M (2019) Transcriptomics-Guided Personalized Prescription of Targeted Therapeutics for Metastatic ALK-Positive Lung Cancer Case Following Recurrence on ALK Inhibitors. Front. Oncol. 9:1026. doi: 10.3389/fonc.2019.01026

Non-small cell lung carcinoma (NSCLC) is the major cause of cancer-associated mortality. Identification of rearrangements in anaplastic lymphoma kinase (*ALK*) gene is an effective instrument for more effective targeted therapy of NSCLC using ALK inhibitors dramatically raising progression-free survival in the *ALK*-mutated group of patients. However, the tumors frequently develop resistance to ALK inhibitors. We describe here a case of 48 y.o. male patient with *ALK*-positive NSCLC who was clinically managed for 6.5 years from the diagnosis. The tumor was surgically resected, but 8 months later multiple brain metastases were discovered. The patient started receiving platinum-based chemotherapy and then was enrolled in a clinical trial of second-generation ALK inhibitor ceritinib, which resulted in a 21 months stabilization. Following disease relapse, the patient was successfully managed for 33 months with different lines of chemo- and local ablative therapies. Chemotherapy regimens, including off-label combination of crizotinib + bevacizumab + docetaxel, were selected using the cancer transcriptome data-guided bioinformatical decision support system Oncobox. These therapies led to additional stabilization for 22 months. Survival of our patient after developing resistance to ALK inhibitor was longer for 16 months than previously reported average survival for such cases. This case shows that transcriptomic-guided sequential personalized prescription of targeted therapies can be effective in terms of survival and quality of life in *ALK*-mutated NSCLC.

Keywords: NSCLC, ALK, transcriptomics, personalized oncology, gene expression

BACKGROUND

Lung cancer is the most common type of cancer and the main factor of cancer-related mortality. According to the reports of World Health Organization and International Agency for Research on Cancer, in 2018 there were ∼2.1 million new registered cases of lung cancer and ∼1.8 million associated deaths (1). Non-small-cell lung carcinoma (NSCLC) is diagnosed in about 80–85% of all lung cancer cases. NSCLC response to standard chemotherapy (typically including treatment with platinum agents) is relatively poor with the median survival time of <1 year after diagnosis (2).

However, the identification of proper molecular biomarkers in NSCLC allowed to increase patient survival by specifically selecting targeted therapeutics. One of these biomarkers is rearrangement within the anaplastic lymphoma kinase gene *ALK* which is present in about 4–7% of all NSCLC cases (3, 4). Worldwide, ∼40,000 new NSCLC cases with mutated *ALK* are detected annually. The current standards of care for patients with advanced *ALK*-mutated NSCLC include therapy with targeted ALK-specific therapeutic crizotinib and other selective ALK inhibitors (5). The median overall survival from the onset of treatment with crizotinib can reach 31 months (6). Although response rate on selective ALK inhibitors is high, there is a major problem of acquiring resistance to these therapeutics within 1–2 years from the onset of therapy (5).

In this report, we describe a case of *ALK*-positive NSCLC with brain metastases. The patient was under observation for 6.5 years and was treated by resection surgery, stereotactic radiosurgery, combination chemotherapies, and by several lines of targeted therapies. After the acquisition of resistance to crizotinib, two other targeted therapeutics were individually selected using a bioinformatic decision support system Oncobox based on the analysis of gene expression and activation of molecular pathways in the patient's tumor biosample (7, 8).

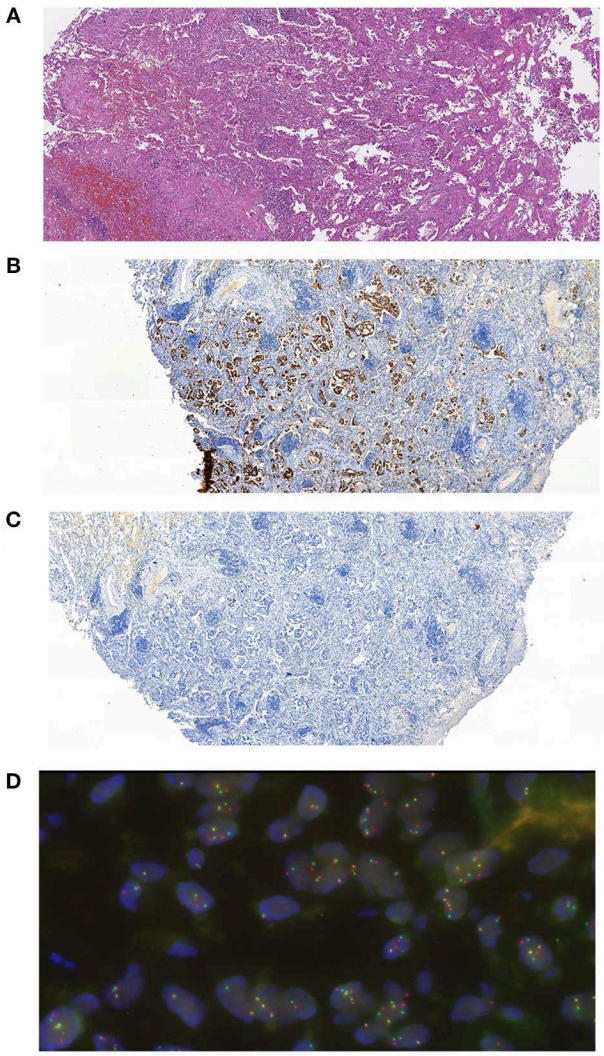


FIGURE 1 | Histological evaluation of the patent's tumor. **(A)** Hematoxylin and eosin staining microphotograph. **(B)** Immunohistochemical staining for TTF-1 (SPT24). **(C)** Immunohistochemical staining for p40 (DeltaNp63). **(D)** FISH analysis for ALK-EML4 translocation.

CASE PRESENTATION

The patient was 48 y.o. male diagnosed in January 2012 with *ALK* mutation-positive NSCLC, stage IIA, T2b N0 M0. The tumor has demonstrated positive immunostaining for TTF-1 (SPT24) and negative for p40 (DeltaNp63). ALK translocation was detected using FISH (**Figure 1**). The patient had a 10 pack-year smoking history but stopped smoking 3 years before the diagnosis. The patient underwent resection surgery (lower lobe of the right lung) and received 4 cycles of vinorelbine + cisplatin (25 mg/m² IV on days 1, 8, 15, and 22 of a 28-day cycle with IV cisplatin 100 mg/m² on day 1) as adjuvant therapy from February to May 2012.

Four months later (September 2012) the patient's condition worsened and multiple brain metastases were discovered (maximum size – 2.9 × 3.5 cm). In October-November 2012 the patient underwent whole brain radiation therapy (linear accelerator, a dose of 40 Gy in 2 Gy fractions) that resulted in a short-term stabilization with subsequent deterioration of the patient's condition.

In April 2013, following confirmation of *EML4-ALK* translocation, the patient was enrolled in the clinical trial NCT01283516 and was prescribed with a second-generation ALK inhibitor ceritinib (750 mg PO daily). Ceritinib therapy resulted in a reduction of brain metastases and the patient's performance status improved significantly. Five months later (September 2013) the patient was able to return to his professional occupation. In February 2015, after 21 progression-free months we observed an increase in the size of brain metastases and the patient was excluded from the NCT01283516 protocol according to exclusion criterion of neurologically unstable metastases.

In March-June 2015, the patient received four cycles of pemetrexed + cisplatin therapy (500 mg/m² IV on day 1 of each 21-day cycle), which resulted in a reduction of several lesions

TABLE 1 | Brain lesions progression.

Lesion\MRI date	2015.02.17	2015.04.13	2015.07.20	2015.09.10	2016.08.04	2016.01.24	2016.11.25
Therapy	End of ceritinib	Pemetrexed + cisplatin	Topotecan, crizotinib	Crizotinib	Crizotinib, SRS + dexamethasone	Crizotinib	Crizotinib + bevacizumab
Lesion 1, right temporoparietal region	65 × 28 × 30 mm	65 × 23 × 25 mm	52 × 10 × 18 mm	61 × 29 × 29 mm	61 × 35 × 32 mm	61 × 34 × 30 mm	51 × 35 × 30 mm
Lesion 2, right occipital lobe.	8 × 9 mm	8 × 7 mm	6 × 7 mm	10 × 14 mm	10 × 14 mm	20 × 18 mm	20 × 17 mm
Lesion 3, right occipital lobe.	19 × 14 mm	19 × 14 mm	15 × 10 mm	25 × 17 mm	18 × 17 mm	14 × 12 mm	13 × 12 mm
Lesion 4, left parietal lobe, parasagittal.				5 × 5 mm	5 × 5 mm	5 × 5 mm	5 × 6 mm
Lesion 5, head of left caudate nucleus, paraventricular.	6 × 4 mm	6 × 5 mm	3 × 1 mm	6 × 6 mm	3 × 3 mm	Couldn't be visualized.	3 × 3 mm
Lesion 6, left temporal pole.				7 × 7 mm	16 × 11 mm	16 × 1.1 mm	16 × 11 mm
Lesion 7, left hippocampus.					4 × 4 mm	3 × 3 mm	3 × 3 mm
Lesion 8, left hippocampus.					4 × 4 mm	3 × 3 mm	3 × 3 mm
Sum of maximal diameters	99 mm	98 mm	77 mm	118 mm	125 mm	122 mm	115 mm
RECIST		Stabilization	Stabilization	Progression	Progression	Stabilization	Stabilization

Changes in size of eight brain lesions from February 2015 till November 2016. Red color indicates increase of lesions in comparison to previous evaluation, green color indicates decrease.

(MRI 2015.04.13, **Table 1**). After that four cycles of topotecan (2.3 mg/m^2 PO days 1–5 of 21-day cycle) were prescribed followed by targeted therapy with first-generation anti-ALK drug crizotinib (250 mg PO twice a day). In July 2015, MRI evaluation revealed reduction of several metastases (**Table 1**, **Figure 2**).

After 3 progression-free months, in September 2015, the patient's condition worsened (headaches, unstable walking). MRI examination showed an increase in size for all lesions previously identified (**Table 1**). We performed stereotactic radiosurgery for lesions in the right temporoparietal region (CyberKnife, a dose of 30 Gy in 6 Gy fractions) and other lesions (CyberKnife, a dose of 6 Gy). Dexamethasone (4 mg daily) was prescribed as adjuvant therapy.

To identify further possible options of chemo- and targeted therapy, we profiled gene expression in the patient's tumor biopsy using microarrays. Bioinformatical platform Oncobox was used to select potentially effective targeted drugs (9, 10). Following Oncobox report, bevacizumab was added to the treatment regimen since May 2016 (550 mg 7.5 mg/kg IV once in 3 weeks). This resulted in a decrease of several brain lesions (**Table 1**, MRI 2016.08.04, 2016.10.24, and 2016.11.25; **Figure 2**, 2016.11.25). In November 2016, the patient underwent another stereotactic radiosurgery for a lesion in the left parietal lobe (Novalis, a dose of 21.9 Gy).

In March 2017, after 10 progression-free months two lung metastases were discovered. Based on the Oncobox report, we added docetaxel to the treatment regimen (80 mg/m^2 IV day 1 of 21-day cycle; 3 cycles). Due to significant adverse effects observed for the patient, the docetaxel dose was reduced to 60 mg/m^2 for the next 3 cycles. In May 2017, CT examination showed reduction of lung lesions.

Following severe pneumonia in March 2018, the patient's condition significantly worsened. Examination in April indicated growth of all previously discovered lesions and the appearance of new metastases. Immunohistochemical testing revealed that ∼98% of tumor cells were PD-L1-positive. The treatment scheme was changed in May 2018 and the patient received two infusions of monoclonal anti-PD-1 antibody pembrolizumab (200 mg IV day 1 of 21-day cycle). However, the patient's condition further deteriorated soon after immunotherapy administration. The patient died of brain edema in July 2018. The history of all treatments is summarized on **Table 2**.

MATERIALS AND METHODS

The patient provided informed written consent for gene expression analysis of his cancer biosample and for presentation of relevant clinical and molecular data in this paper. The tumor tissue sample used for gene expression analysis was obtained during lung resection in February 2012 and stored in the form of formalin-fixed paraffin-embedded (FFPE) tissue block at the room temperature. For RNA extraction, we used five 250-μm-thick consecutive sections of FFPE block with tumor cell content >80%. Gene expression profiling was performed using microarray platform CustomArray (USA) according to the Manufacturer's protocol, except for the addition of the dNTP mix containing biotinylated dUTP to the amplification reaction (final proportion of dTTP/biotin-dUTP was 5-to-1).

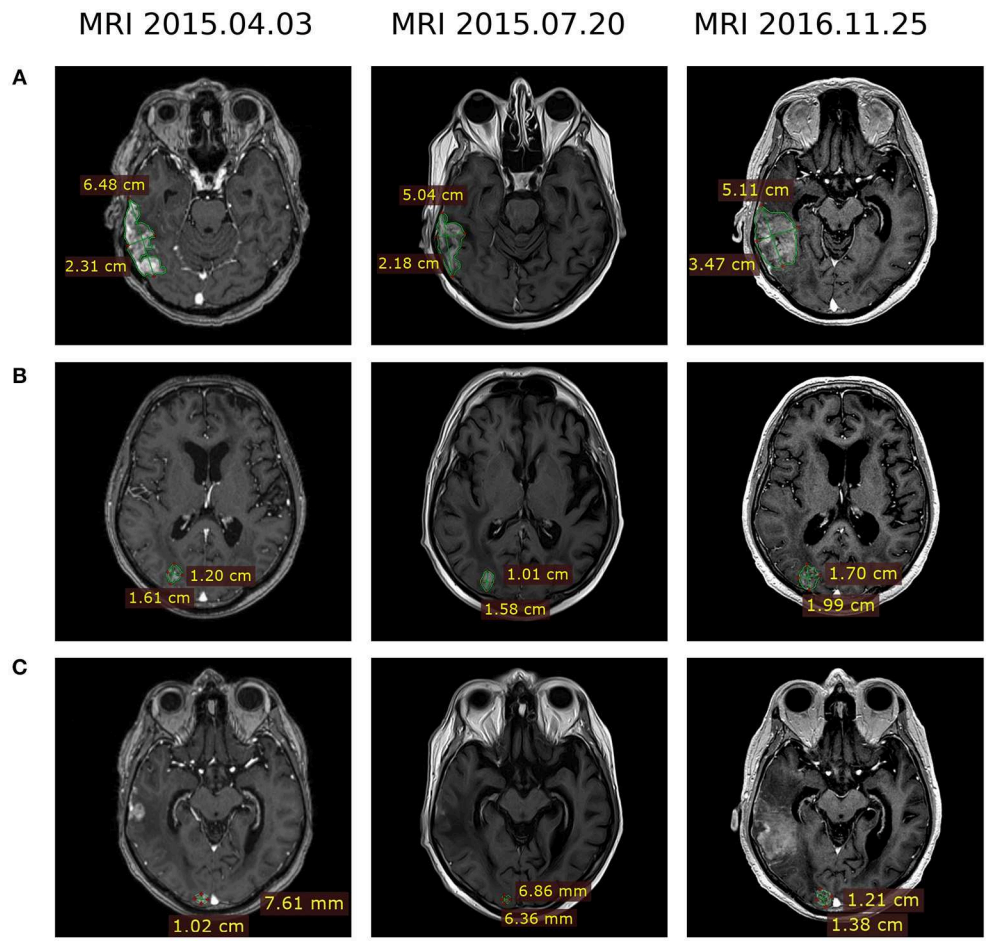


FIGURE 2 | MRI evaluation of brain metastases dynamics at therapy lines 3 and 4 after development of the ceritinib resistance. Treatment regimens were pemetrexed + cisplatin (2015.04.03), topotecan followed by crizotinib (2015.07.20) and crizotinib + bevacizumab (2016.11.25). **(A)** Lesion 1, right temporoparietal region. **(B)** Lesion 2, right occipital lobe. **(C)** Lesion 3, right occipital lobe.

The expression profile of 3,682 human genes was measured and deposited in Gene Expression Omnibus repository with ID GSE133605. For gene expression normalization, four healthy lung tissue gene expression profiles from unrelated donors (GEO: GSM862609-GSM862612) were used as the reference (9). The signaling pathway activation analysis and prioritizing of targeted therapeutics were made using Oncobox bioinformatical platform, as previously described in Sanders et al. (10) and Poddubskaya et al. (11).

DISCUSSION

We report here the case of *ALK*-mutated NSCLC treated with six lines of therapy including several molecular-targeted drugs. Gene expression analysis complemented genetic *ALK* testing and was useful for selecting further treatment options.

The first line therapy was resection surgery and vinorelbine + cisplatin, which is the standard treatment for stage II NSCLC (12). The second line was monotherapy with ceritinib—the second-generation anti-ALK targeted drug currently recommended as the first-line therapy for *ALK*-mutated NSCLC (12) that was in the clinical trials back in 2013.

The patient developed resistance after 21 progression-free months. Acquired resistance is well documented for crizotinib, a first-generation anti-ALK target drug (13). Usually, ceritinib or other second-generation anti-ALK targeted therapeutics are used to overcome crizotinib resistance (14), and ceritinib resistance can potentially be reversed by using afatinib, another second-generation inhibitor of ALK (15). But back in 2015, second-generation ALK-inhibitors were at the different stages of clinical trials and the following treatment strategy was accepted.

First, pemetrexed + cisplatin polychemotherapy started because this regimen combination was reported to be effective and well-tolerated in NSCLC patients with brain metastases (16). Then the patient was treated with topotecan, a Topoisomerase I inhibitor that previously showed encouraging results in the NSCLC treatment (17). Finally, the patient was prescribed with crizotinib. To date discontinuation on ALK inhibitor therapy

TABLE 2 | Outline of the patient treatment strategy.

Date	Drug	Oncobox predicted drug efficiency score	Response	Progression free survival (months)
LINE 1				
01–09.2012	Lung resection; vinorelbine + cisplatin	−7.65 (vinorelbine); NA (cisplatin)	No progression	7
10–11.2012	Whole brain radiation therapy		Progression	
LINE 2				
04.2013–02.2015	Ceritinib	0.6 (ceritinib)	Partial response	21
LINE 3				
03–06.2015	Pemetrexed + cisplatin	−0.5 (pemetrexed); NA (cisplatin)	Stabilization	3
06.2015–09.2015	Topotecan (4 cycles), crizotinib	0.55 (topotecan); 6.15 (crizotinib)	Stabilization	3
09.2015–05.2016	Stereotactic radiosurgery to multiple lesions; dexamethasone; crizotinib is continued	NA (dexamethasone); 6.15 (crizotinib)	Stabilization	8
LINE 4				
05.2016–03.2017	Crizotinib + bevacizumab Stereotactic radiosurgery to a single lesion	6.15 (crizotinib); 9.45 (bevacizumab)	Stabilization	10
LINE 5				
03.2017–03.2018	Crizotinib + bevacizumab + docetaxel	6.15 (crizotinib); 9.45 (bevacizumab); 0.45 (docetaxel)	Stabilization until severe pneumonia	12
LINE 6				
20.05.2018, 11.06.2018	Pembrolizumab	NA (pembrolizumab)	Not evaluable	No Data

is not recommended because of possible disease flare (18, 19). Stereotactic radiosurgery was performed because local ablative therapy with continued administration of ALK inhibitors was effective according to the previous reports (20).

It was recently demonstrated that genomic and transcriptomic profiling are potentially useful for improving therapy recommendations and patient outcomes (21). Identification of further therapeutic schemes including combinations crizotinib + bevacizumab and crizotinib + bevacizumab + docetaxel was based on the bioinformatic analysis of tumor gene expression profile. First, bevacizumab was added to the treatment scheme based on its positive simulated Drug Efficiency Score (**Supplementary Table 1**), which was calculated based on gene expression and activation level of molecular pathways in the patient's tumor using the Oncobox platform. It revealed that Ras signaling pathway was upregulated in the patient's tumor (**Figure 3**). The Ras pathway contains targets of both crizotinib and bevacizumab, so potentially the achieved clinical benefit of their combination may be linked with dual inhibition of this pathway.

Previously, activation of Raf-MEK-ERK signaling axis of Ras pathway was shown to be crucial for the *ALK* mutation-positive tumor cell survival and dual ALK-MEK inhibition was proposed as a new approach to battle tumor drug resistance (22). However, in the current tumor case the Raf-MEK-ERK axis was downregulated (**Figure 3**) and based on these data the dual ALK-MEK inhibition therapy would not be recommended.

Bevacizumab and other anti-vascular endothelial growth factor monoclonal antibodies were approved for the treatment of NSCLC (23). Recently, clinical investigation of crizotinib + bevacizumab combined therapy for advanced NSCLC reported a median progression-free survival of 13 months (24). In agreement with these results, in the case of our patient crizotinib + bevacizumab treatment resulted in 10 progression-free months.

When the patient progressed on crizotinib + bevacizumab therapy, docetaxel was added to the treatment regimen based on its positive simulated Drug Efficiency Score (**Supplementary Table 1**) and because of its different mechanism of action compared to the other therapeutics used. Docetaxel binds to microtubules, thereby interfering with cell proliferation and promoting cancer cell death. Docetaxel has been also approved for NSCLC (25) and bevacizumab + docetaxel polychemotherapy had a mean progression-free survival of 6 months for NSCLC in a published clinical investigation (26). However, to our knowledge, there are no previous reports on molecular-guided therapy with triple combination crizotinib + bevacizumab + docetaxel that resulted in 12 progression-free months in our case.

The next planned line of therapy was treatment with anti-PD-1 immunotherapeutic pembrolizumab since most of the patient's cancer cells were PD-1-positive. Unfortunately, severe pneumonia most likely accelerated further progression of the disease, and efficacy of the anti-PD-1 therapy couldn't be assessed due to the swift discontinuation of this treatment plan.

Overall, the patient lived for 78 months (6.5 years) after the diagnosis and 70 months after the discovery of brain metastases. The patient studies of ceritinib resistance development are only

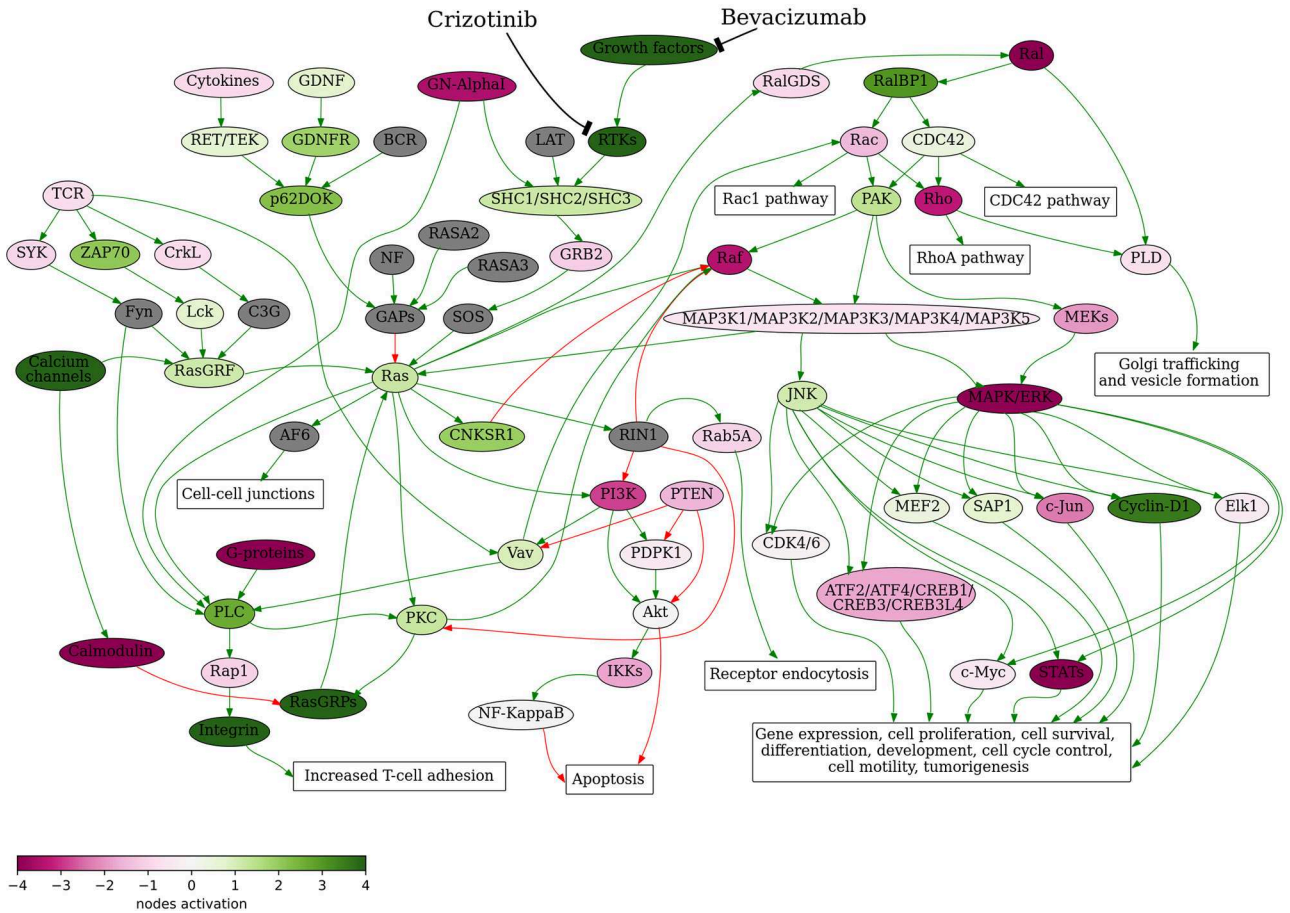


FIGURE 3 | Ras signaling pathway shown as an interacting network. This pathway was hyperactivated in the patient's tumor tissue. Green arrows indicate activation, red arrows–inhibition. Transcript nodes are shown in ovals, interface molecular pathways and cellular effects–in rectangles. The color depth of transcript nodes reflects the extent of node activation (logarithms of the case-to-normal (CNR) expression rate for each node, in which "normal" is a geometric average between expression levels in normal tissue samples). Molecular targets of crizotinib and bevacizumab are indicated by black arrows. Visualization was implemented using Oncobox software.

represented by several published clinical cases (27–29) and cannot be used to directly evaluate the effectiveness of our approach. However, there are far more literature data available for crizotinib. For male ALK mutation-positive patients treated with one or more lines of ALK inhibitors the median overall survival after stage IV diagnosis was found to be 48 months (30), while in the case of our patient the overall survival was 70 months. The patient's survival since the start of therapy with crizotinib (line 3) was 36 months which exceeds previously reported median values of 31 (6) and 16.6 months (31). Moreover, the median overall survival after progression on crizotinib was reported to be 25 months on next-generation ALK-inhibitors and only 6.4 months on the other therapies (31). However, our patient lived for 41 months after developing resistance to ceritinib and for 33 months after developing resistance to crizotinib, which is higher than both above estimates.

Therefore, this case suggests that the drug efficiency scoring based on gene expression profiling of the patient's tumor biopsy biomaterial could potentially complement the standard mutation analysis for the management of advanced cancer patients with NSCLC. In turn, the Oncobox platform can be potentially helpful for selecting effective treatment regimens also to the other types of solid tumors as previously shown for metastatic cholangiocarcinoma and ovarian cancer (9, 11).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE133605.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB) at Clinical Center Vitamed, Moscow, Russia. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

EP, MB, SS, ABor, AG, and MS collected and interpreted patient data. EP, MB, and ABon were involved in clinical management. EZ and PK performed molecular analyses. LM and AM analyzed and interpreted the MRI data. AM, ABor, EZ, MS, PK, and EP wrote the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01026/full#supplementary-material

Supplementary Table 1 | Oncobox balanced efficiency scores and pathway activation levels.

REFERENCES

1. IARC – International Agency for Research on Cancer. *Cancer Today*. Available online at: http://gco.iarc.fr/today/online-analysis-pie (accessed June 17, 2019).
2. Millett RL, Elkon JM, Tabbara IA. Directed therapies in anaplastic lymphoma kinase-rearranged non-small cell lung cancer. *Anticancer Res*. (2018) 38:4969–75. doi: 10.21873/anticanres.12815
3. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*. (2007) 448:561–6. doi: 10.1038/nature05945
4. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*. (2007) 131:1190–203. doi: 10.1016/j.cell.2007.11.025
5. Alice T. Shaw JAE. ALK in lung cancer: past, present, and future. *J Clin Oncol*. (2013) 31:1105–11. doi: 10.1200/JCO.2012.44.5353
6. Tsimafeyeu I, Moiseenko F, Orlov S, Filippova E, Belonogov A, Nebesnykh A, et al. Overall survival of patients with ALK-positive metastatic non–small-cell lung cancer in the Russian federation: nationwide cohort study. *J Glob Oncol*. (2019) 5:1–7. doi: 10.1200/JGO.19.00024
7. Sorokin M, Kholodenko R, Suntsova M, Malakhova G, Garazha A, Kholodenko I, et al. Oncobox bioinformatical platform for selecting potentially effective combinations of target cancer drugs using high-throughput gene expression data. *Cancers*. (2018) 10:365. doi: 10.3390/cancers10100365
8. Buzdin A, Sorokin M, Garazha A, Sekacheva M, Kim E, Zhukov N, et al. Molecular pathway activation - new type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin Cancer Biol*. (2018) 53:110–24. doi: 10.1016/j.semcancer.2018.06.003
9. Poddubskaya EV, Baranova MP, Allina DO, Smirnov PY, Albert EA, Kirilchev AP, et al. Personalized prescription of tyrosine kinase inhibitors in unresectable metastatic cholangiocarcinoma. *Exp Hematol Oncol*. (2018) 7:21. doi: 10.1186/s40164-018-0113-x
10. Sanders YY, Ambalavanan N, Halloran B, Zhang X, Liu H, Crossman DK, et al. Altered DNA methylation profile in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. (2012) 186:525–35. doi: 10.1164/rccm.201201-0077OC
11. Poddubskaya EV, Baranova MP, Allina DO, Sekacheva MI, Makovskaia LA, Kamashev DE, et al. Personalized prescription of imatinib in recurrent granulosa cell tumor of the ovary: case report. *Cold Spring Harb Mol Case Stud*. (2019) 5:a003434. doi: 10.1101/mcs.a003434
12. National Comprehensive Cancer Network. *Non-Small Cell Lung Cancer (Version 5.2019)*. (2019). Available online at: https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf
13. Casaluce F, Sgambato A, Sacco PC, Palazzolo G, Maione P, Rossi A, et al. Resistance to crizotinib in advanced non-small cell lung cancer (NSCLC) with ALK rearrangement: mechanisms, treatment strategies and new targeted therapies. *Curr Clin Pharmacol*. (2016) 11:77–87. doi: 10.2174/1574884711666160502124134
14. Muller IB, De Langen AJ, Honeywell RJ, Giovannetti E, Peters GJ. Overcoming crizotinib resistance in ALK-rearranged NSCLC with the second-generation ALK-inhibitor ceritinib. *Expert Rev Anticancer Ther*. (2016) 16:147–57. doi: 10.1586/14737140.2016.1131612
15. Chen H, Zhang Q, Zhang Y, Jia B, Zhang B, Wang C. Afatinib reverses ceritinib resistance (CR) in ALK/ROS1-positive non-small-cell lung cancer cell (NSCLC) via suppression of NRG1 pathway. *Onco Targets Ther*. (2018) 11:8201–9. doi: 10.2147/OTT.S173008
16. Barlesi F, Gervais R, Lena H, Hureaux J, Berard H, Paillotin D, et al. Pemetrexed and cisplatin as first-line chemotherapy for advanced non-small-cell lung cancer (NSCLC) with asymptomatic inoperable brain metastases: a multicenter phase II trial (GFPC 07-01). *Ann Oncol*. (2011) 22:2466–70. doi: 10.1093/annonc/mdr003
17. Vennepureddy A, Atallah J-P, Terjanian T. Role of topotecan in non-small cell lung cancer: a review of literature. *World J Oncol*. (2015) 6:429–36. doi: 10.14740/wjon950e
18. Kuriyama Y, Kim YH, Nagai H, Ozasa H, Sakamori Y, Mishima M. Disease flare after discontinuation of crizotinib in anaplastic lymphoma kinase-positive lung cancer. *Case Rep Oncol*. (2013) 6:430–3. doi: 10.1159/000354756
19. Pop O, Pirvu A, Toffart A-C, Moro-Sibilot D. Disease flare after treatment discontinuation in a patient with EML4-ALK lung cancer and acquired resistance to crizotinib. *J Thorac Oncol*. (2012) 7:e1–2. doi: 10.1097/JTO.0b013e318257fc1d
20. Weickhardt AJ, Scheier B, Burke JM, Gan G, Lu X, Bunn PA Jr, et al. Local ablative therapy of oligoprogressive disease prolongs disease control by tyrosine kinase inhibitors in oncogene addicted non-small cell lung cancer. *J Thorac Oncol*. (2012) 7:1807–14. doi: 10.1097/JTO.0b013e3182745948
21. Rodon J, Soria J-C, Berger R, Miller WH, Rubin E, Kugel A, et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat Med*. (2019) 25:751–8. doi: 10.1038/s41591-019-0424-4
22. Gorjan Hrustanovic TGB. RAS-MAPK in ALK targeted therapy resistance. *Cell Cycle*. (2015) 14:3661–2. doi: 10.1080/15384101.2015.1096103
23. Sandler AB, Johnson DH, Herbst RS. Anti-vascular endothelial growth factor monoclonals in non-small cell lung cancer. *Clin Cancer Res*. (2004) 10:4258s−62s. doi: 10.1158/1078-0432.CCR-040023
24. Yang B, Cui Z, Meng X, Huang Z, Hu Y. Crizotinib with bevacizumab as first-line therapy in patients with advanced non-small-cell lung cancer harboring EML4-ALK fusion variant mutation: a prospective exploratory study. *J Clin Oncol*. (2018) 36:e21186. doi: 10.1200/JCO.2018.36.15_suppl.e21186
25. He X, Wang J, Li Y. Efficacy and safety of docetaxel for advanced non-small-cell lung cancer: a meta-analysis of Phase III randomized controlled trials. *Onco Targets Ther*. (2015) 8:2023–31. doi: 10.2147/OTT.S85648
26. Kurishima K, Watanabe H, Ishikawa H, Satoh H, Hizawa N. A retrospective study of docetaxel and bevacizumab as a second- or later-line chemotherapy for non-small cell lung cancer. *Mol Clin Oncol*. (2017) 7:131–4. doi: 10.3892/mco.2017.1282
27. Tchekmedyian N, Ali SM, Miller VA, Haura EB. Acquired ALK L1152R mutation confers resistance to ceritinib and predicts response to alectinib. *J Thorac Oncol*. (2016) 11:e87–8. doi: 10.1016/j.jtho.2016.03.018
28. Toyokawa G, Inamasu E, Shimamatsu S, Yoshida T, Nosaki K, Hirai F, et al. Identification of a novel ALK G1123S mutation in a patient with ALK-rearranged non–small-cell lung cancer exhibiting resistance to ceritinib. *J Thorac Oncol*. (2015) 10:e55–7. doi: 10.1097/JTO.0000000000000509
29. Mehlman C, Chaabane N, Lacave R, Kerrou K, Ruppert AM, Cadranel J, et al. Ceritinib ALK T1151R resistance mutation in lung cancer

with initial response to brigatinib. *J Thorac Oncol*. (2019) 14:e95–6. doi: 10.1016/j.jtho.2018.12.036

30. Pacheco JM, Gao D, Smith D, Purcell T, Hancock M, Bunn P, et al. Natural history and factors associated with overall survival in stage IV ALK-rearranged non-small cell lung cancer. *J Thorac Oncol*. (2019) 14:691–700. doi: 10.1016/j.jtho.2018.12.014
31. Duruisseaux M, Besse B, Cadranel J, Pérol M, Mennecier B, Bigay-Game L, et al. Overall survival with crizotinib and next-generation ALK inhibitors in ALK-positive non-small-cell lung cancer (IFCT-1302 CLINALK): a French nationwide cohort retrospective study. *Oncotarget*. (2017) 8:21903–17. doi: 10.18632/oncotarget.15746

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 19 November 2019
doi: 10.3389/fonc.2019.01206

Exploration of the Transcriptional Landscape of ALPPS Reveals the Pathways of Accelerated Liver Regeneration

Pieter Borger[1], Marcel Schneider[1], Lukas Frick[1], Magda Langiewicz[1], Maksim Sorokin[2,3,4], Anton Buzdin[2,3,4,5], Ekaterina Kachaylo[1], Rolf Graf[1], Bostjan Humar[1] and Pierre-Alain Clavien[1]*

[1] Laboratory of the Swiss Hepato-Pancreato-Biliary (HPB) and Transplantation Center, Department of Surgery, University Hospital Zürich, Zurich, Switzerland, [2] OmicsWay Corp., Walnut, CA, United States, [3] I.M. Sechenov First Moscow State Medical University, Moscow, Russia, [4] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, [5] Oncobox Ltd., Moscow, Russia

OPEN ACCESS

Edited by:
Ernesto Augusto Bueno Da Fonseca Lima, University of Texas at Austin, United States

Reviewed by:
Xiaofei Song, University of Texas MD Anderson Cancer Center, United States
Marcello Donati, University of Catania, Italy

***Correspondence:**
Pieter Borger
peterborger@hotmail.com

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *10 June 2019*
Accepted: *23 October 2019*
Published: *19 November 2019*

Citation:
Borger P, Schneider M, Frick L, Langiewicz M, Sorokin M, Buzdin A, Kachaylo E, Graf R, Humar B and Clavien P-A (2019) Exploration of the Transcriptional Landscape of ALPPS Reveals the Pathways of Accelerated Liver Regeneration. Front. Oncol. 9:1206. doi: 10.3389/fonc.2019.01206

Background and Aims: ALPPS (associating liver partition and portal vein ligation for staged hepatectomy), a novel 2-staged hepatectomy, dramatically accelerates liver regeneration and thus enables extensive liver tumor resection. The signaling networks underlying the ALPPS-induced accelerated regeneration process are largely unknown.

Methods: We performed transcriptome profiling (TP) of liver tissue obtained from a mouse model of ALPPS, standard hepatectomy (68% model), and additional control surgeries (sham, PVL and Tx). We also performed TP using human liver biopsies ($n = 5$) taken from the occluded lobe and the future liver remnant (FLR) during the first step of ALPPS surgery (4–5 h apart). We used Oncofinder computational tools, which covers 378 ISPs, for unsupervised, unbiased quantification of ISP activity.

Results: Gene expression cluster analysis revealed an ALPPS specific signature: the IGF1R Signaling Pathway (Cell survival), the ILK Pathway (Induced cell proliferation), and the IL-10 Pathway (Stability determination) were significantly enriched, whereas the activity of the Interferon Pathway (Transcription) was reduced ($p < 0.05$). Further, the PAK- and ILK-associated ISPs were activated at an earlier time point, reflecting significant acceleration of liver regeneration ($p < 0.001$). These pathways, which were also recovered in human liver biopsies, control cell growth and proliferation, inflammatory response, and hypoxia-related processes.

Conclusions: ALPPS is not a straightforward addition of portal vein ligation (PVL) plus transection—it is more. The early stages of normal and accelerated liver regeneration are clearly discernible by a significantly increased and earlier activation of a small number of signaling pathways. Compounds mimicking these responses may help to improve the ALPPS method and further reduce the hospitalization time of the patient.

Keywords: two-staged hepatectomy, ALPPS, transcriptome profiling, signaling pathways, Oncofinder

INTRODUCTION

Liver regeneration is controlled by a cooperating, redundant system of biological networks performing an assortment of tasks, which together result in a coordinated response to replenish lost liver tissue. The regenerating capacity of the liver inspired liver surgeons to develop 2-staged strategies to perform extended liver resections to clean the organ from multiple tumors. In the conventional 2-staged strategies, the complete restoration of a functional liver may take up to 8–10 weeks (1, 2). Recently, a novel two-staged strategy has been introduced: ALPPS (for: *A*ssociating *L*iver *P*artition and *P*ortal vein ligation for *S*taged hepatectomy) (3). In the first stage of the ALPPS procedure, the portal vein is ligated, followed by removal of any tumor in the future liver and an *in situ* split of the parenchyma between the healthy and diseased liver (partition or transection) is performed. In the second step, the isolated, deportalized liver is removed, leaving behind a tumor-free, hypertrophic liver remnant (4–6). The main advantage of the ALPPS strategy is that the accelerated regeneration of the liver remnant, which reaches a body-sustaining size within 7–10 days, enables the prompt elimination of the major tumor load (4, 7).

The biological processes underlying liver regeneration processes are complex. They involve (thyroid) hormones (8), cytokines [IL-6 (9) and TNF (10)], growth factor responses [HGF, TGF-β, epidermal growth factor (EGF)] (11, 12), glucose- and bile acid metabolism (13, 14), and platelet-derived factors, such as serotonin (15). Further studies in knockout mouse models revealed several key switches in signal-transduction systems, which either delay or accelerate the regenerating process. NF-κB (16), nuclear receptors [FXR (14) and CAR (17)] have been described as accelerators, whereas p21 (18), Socs3 (19), and Tob1 (20) act as repressors.

Recently, we reported that that Indian hedgehog (*Ihh*), a secreted ligand important for fetal development, is a crucial mediator of the regenerative acceleration triggered by ALPPS surgery (21). Despite our increasing knowledge of the interwoven biological signaling networks underlying normal and accelerated liver regeneration, comprehensive whole genome analyses are required for full recognition of the underlying pathways. The newly developed bioinformatics tool OncoFinder enables quantitative measurement of intracellular signaling pathway (ISP) activation based on whole genome expression data (22, 23). The advantage of OncoFinder over alternative tools, such as Metacore and Ingenuity Pathway Analysis (IPA), is that it quantifies the pathway activation strength (PAS) (22–24). PAS values represent the cumulative value of perturbations in a signaling pathway and serve as reliable indicators of pathological changes to the intracellular signaling machinery. The PAS value itself serves as a robust new biomarker that can distinguish between the pathway activation profiles in different tissues (23, 25).

Abbreviations: ALPPS, Associating Liver Partition and Portal Vein Ligation Surgery; LLLx, left lateral lobe resection; PAS, pathway activation strength; PVL, portal vein ligation; Tx, parenchymal transection; TSH, two-stage hepatectomy; FLR, functional liver remnant.

The aim of our present study was to assess comprehensively the ISPs underlying normal and accelerated liver regeneration in two established mouse models (26, 27), and in liver tissue obtained from humans before and after the first step of ALPPS surgery. The data designate that our murine ALPPS model—despite differences—reflects ALPPS-induced accelerated liver regeneration in humans.

METHODS

Surgery

ALPPS surgery in C57BL/6 mice ($n = 3$ for all procedures) was performed as described earlier (26). In brief, a 90% PVL was performed, leaving a 10% functional remnant consisting of the left and a part of the right middle lobe. Then, a partial 80% transection was done through the middle lobe along the demarcation line of the occluded/non-occluded parenchyma. The left lateral lobe (LLL, 25% of liver volume) was also resected to simulate the cleaning of the liver from smaller tumors as often carried out in human ALPPS (21, 26). ALPPS surgery is associated with some initial injury (serum ALT at around 5,000 U/I 1 day post-operation), which however declines over time toward zero at day 7 post-operation. Serum HMGB1, released by necrotic cells, is not elevated at any time post-operation, indicating the absence of significant necrosis as confirmed by histology on day 2. Very similar findings are observed for PVL surgery, indicating that ALPPS does not augment injury (26). Following 68% hepatectomy, injury is negligible, with ALT <100 U/I 1 day post-operation (27). Serum bilirubin is not elevated following ALPPS, PVL, or 68% hepatectomy (26, 27). Liver weight gain is clearly evident already at 4 h post ALPPS, steeply rising to reach a plateau at 24 h (with step 2—resection of ligated parts—usually performed at day 2 in mice). PVL also induces—to a lesser extent—early liver weight gain; however, a low plateau is reached already at 8 h post PVL (21). After 68% hepatectomy, liver starts to gain weight more slowly, with its strongest gains toward 48 h. This time point coincides with the hepatocellular mitotic peak, which follows cell cycle entry around 16–20 h post 68% hepatectomy (27). In contrast, hepatocytes enter the cell cycle already at 4 h post ALPPS, with a first mitotic peak at 8 h. After PVL, cell cycle entry also occurs early, but only at levels similar to transection (which does not induce regeneration), and low numbers of mitoses are observed from 12 h onwards (21). Therefore, ALPPS surgery accelerates mouse liver regeneration both in time and magnitude relative to other liver surgeries.

Standard hepatectomy (partial 68% hepatectomy) in mice was performed as described by Lehmann (27). In short, a midline incision was performed, and the liver was freed from ligaments. The pedicle of the left lobe was ligated (silk, 6/0) and resected. After cholecystectomy (Prolene, 8/0; Ethicon, Neuchatel, Switzerland), the middle lobe was ligated in 2 steps (silk 6/0) and resected. All animal experiments conformed to the Swiss Federal Animal Regulations and were approved by the Veterinary Office of Zurich. Animals aged 10–12 weeks were kept on a 12-h day/night cycle with free access to food and water. C57BL/6 mice were obtained from Envigo (Horst, The Netherlands). All animals were part of the same shipment, same

age and gender, randomized, and part of the same project to ensure similar conditions.

Tissue Specimen

Mice liver tissue was collected at different time points after surgery (as indicated throughout the manuscript). Human liver tissue specimens (biopsies) were obtained from the Department of Visceral and Transplant Surgery, University Hospital Zurich with the approval of the local ethics committees (Nr: 2015-0547, Cantonal Ethics Committee, Zurich) and written consent of all patients (for characteristics see **Supplementary Data S-XI**). After laparotomy and initial inspection of the abdomen, liver punch biopsies were taken of the future liver remnant (FLR) immediately before starting ALPPS step 1 [biopsy 1]. Then, partial ALPPS was performed as described earlier (7). Biopsy 2 was taken of the FLR immediately before closure of the laparotomy [end of ALPPS step 1]. All biopsies were snap frozen in liquid nitrogen and stored at −80°C. RNA isolation and sequencing was performed as described below.

RNA Isolation and Sequencing

RNA isolation was performed using the TRIzol Method as described by the manufacturers (Life Technologies, Switzerland). An equimolar RNA pool was prepared from liver tissue and/or biopsies using DNA Column Clean-up (Qiagen, Basel, CH). Then, 1 μg of total RNA was used for library preparation according to the Illumina TruSeq stranded mRNA sample preparation protocol (Illumina). The resulting mRNA library was sequenced on an Illumina Hiseq 2,500 sequencer (Functional Genomics Center Zürich, Zürich, CH). Sequenced reads were aligned to the mouse and human (hg19) reference genome with TopHat (version 2.0.10) 27 by using the -G (GTF file of Ensembl release 75) option. Furthermore, the aligned reads were used to quantify mRNA expression by using HTSeq-count (version 0.6.1)28 with hg19 GTF (Ensembl release 75). All data have been deposited to the European Nucleotide Archive (ENA) under the accession code PRJEB15593.

Differential Expression Analysis

Raw reads were filtered by quality >30 score through FASTX toolkit and then trimmed at $5'$ and $3'$ in order to remove index and adapter. Only the remaining reads were used for alignment with the human genome assembly (GRCh37), where we employed TopHat v2.0.14 (28).

Source Datasets

The signaling pathways knowledge base developed by SABiosciences (https://www.slideshare.net/elsavonlicy/pathway-mapreferenceguide) was used to determine structures of intracellular pathways for OncoFinder (22, 23).

Functional Annotation of Gene Expression Data

For the functional annotation of the primary gene expression data, we applied our original algorithm termed OncoFinder (22, 23). It enables calculation of the Pathway Activation Strength (PAS), a value that serves as a qualitative measure of pathway activation. Briefly, the algorithm utilizes the following formula to evaluate pathway activation:

$$PAS_p = \sum_n ARR_{np} \cdot BTIF_n \cdot \lg(CNR_n) \quad (1)$$

Here the *case-to-normal ratio*, *CNRn*, is the ratio of expression levels for a given gene (*n)* in the sample to the mean value for the control group. The Boolean flag of *BTIF* (*beyond tolerance interval flag*) equals to zero when the *CNR* value has passed simultaneously the two criteria that demark the significantly perturbed expression level from essentially normal. First, the expression level for the sample lies within the tolerance interval, where $p > 0.05$. Second, the value of *CNR* differs from 1 considerably, *CNR* 0.66 or *CNR* 1.5. The discrete value of *ARR* (*activator / repressor role*) reflects the functional role of a protein *n* in the pathway (22). The pathway-specific PAS values calculated by Oncofinder are more reliable than single gene analysis and improves the robustness of experimental transcriptomics data (23).

Statistical Tests

The PAS values for each normal sample were obtained using the whole set of these normal samples as a reference. Distribution of PAS values was estimated, assuming its Gaussian behavior. Then, for each pathway of each sample, the probability that its PAS value comes from this estimated distribution was calculated. Additionally, *p*-values for each pathway of the entire group of samples were calculated using Wilcoxon rank-sum test. Principal component analyses were performed using the MADE4 package (29). Hierarchical clustering heatmaps with Pearson distance and average linkage were generated using heatmap.2 function from "gplots" package (30). Pearson tau correlation matrices were calculated in R 3.1.1 using a function of standard library "cor" with the default settings. Correlation diagrams were built using a function "corrplot" from the package "corrplot" sorted with respect to hierarchical clustering. Similarities between the pathways according to the content of similar genes were calculated using the Jaccard coefficient. The Jaccard coefficient measures similarity between finite sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets. Venn diagrams were constructed using Venny 2.1 [http://bioinfogp.cnb.csic.es/tools/venny/].

miRNA Target Prediction

Focusing on differentially expressed pre-miRNAs present in our datasets, we predicted their putative mRNA targets considering only experimentally validated miRNA-mRNA interactions using the Ingenuity Pathway Analysis (IPA) suite (Qiagen, Redwood City, Calif; https://www.slideshare.net/elsavonlicy/pathway-mapreferenceguide). We used miRTarBase for predicting targets of miRNAs and assessed the influence of miRNAs on ISPs using our method MiRImpact (31). Among all miRNA-targeted mRNAs, only genes having at least 10 reads (read count ≥10) were considered true targets for differentially expressed miRNAs in BSM cells.

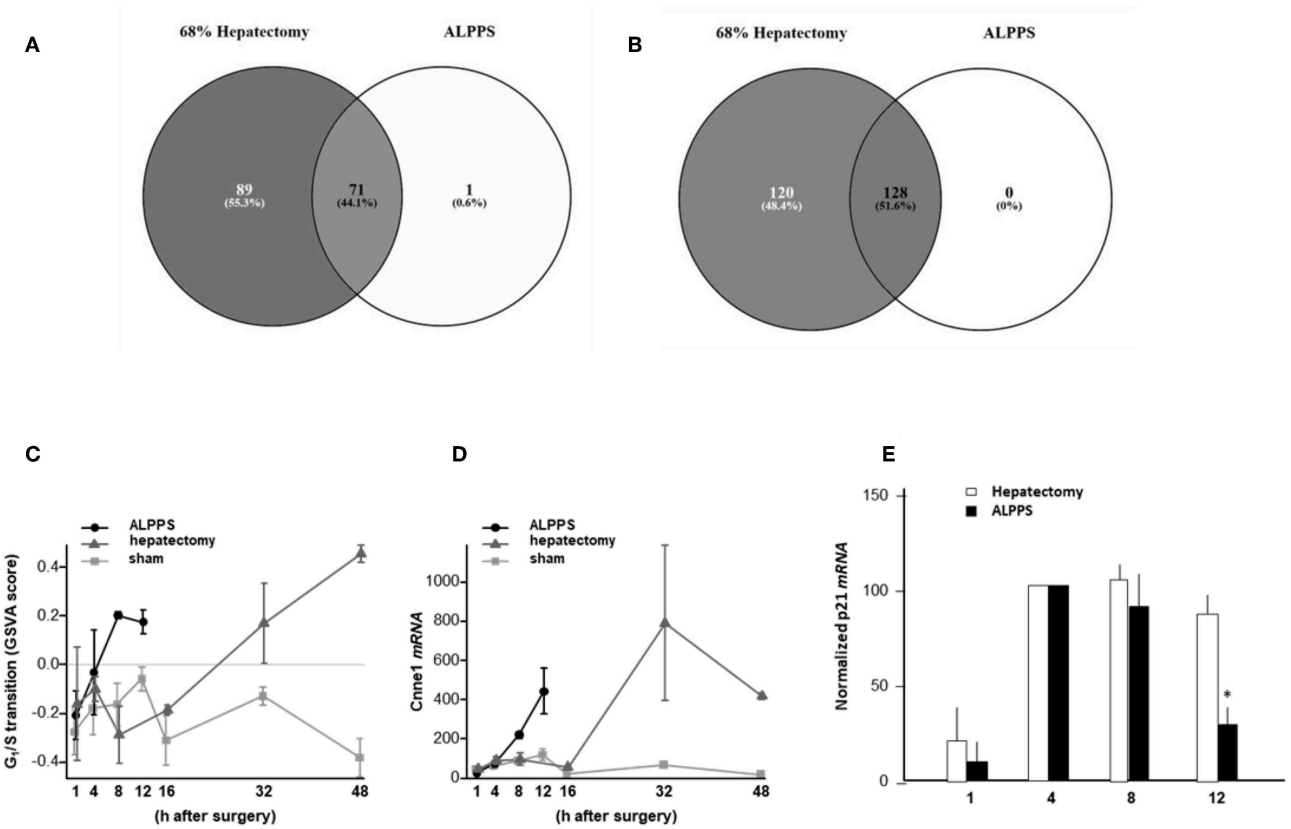


FIGURE 1 | Venn diagrams showing the total of unique and common ISPs after 68% hepatectomy and ALPPS surgery. **(A)** Unique and common ISPs activated/repressed 4 h after surgery. **(B)** Common ISPs activated/repressed 8 h after surgery (For details see: **Table 2**). **(C)** Per-sample enrichment scores for genes involved in the G1/S transition phase of the mitotic cell cycle (the GO_BP_MM_G1_S_TRANSITION_OF_MITOTIC_CELL_CYCLE signature from the GSKB database), calculated using gene set variation analysis (GSVA). **(D)** Cyclin E1 mRNA (Ccne1) mRNA expression. **(E)** Expression of p21 mRNA after ALPPS and 68% hepatectomy (as indicated); $^*p < 0.005$.

Gene Set Variation Analysis of Mitotic Cell Cycle Signatures

To compare the kinetics of the expression of mitotic cell cycle genes between ALPPS and standard hepatectomy, gene signatures for different phases of the mitotic cell cycle were selected from the MSigDB and GSKB online databases. For access to the GSKB database, the gskb R package (version 1.10.0) was used. Individual-sample gene set enrichment scores were calculated using Gene Set Variation Analysis with the GSVA R package, version 1.26.0. All gene set enrichment analyses were performed within the R statistical programming environment, version 3.4.3.

RESULTS

ALPPS Surgery Induced an Earlier Cell Cycle Entry

To elucidate the molecular pathways responsible for accelerated liver regeneration, hepatic RNA isolated at several time points after ALPPS, PVL, transection, LLLx, and sham operation was deep-sequenced. Pathway activation strength (PAS) profiles were established using the normalized gene expression levels of liver-expressed genes with the OncoFinder algorithm. First, we analyzed the activation status of 378 ISPs. Using the PAS values, we built hierarchical clustering heat maps with Euclidean distance and average linkage for all groups (For heatmaps see: **Supplementary Data S-I**). We identified the common and unique ISPs for ALPPS and normal liver regeneration as induced by standard hepatectomy. As shown in **Figure 1A**, 160 and 72 ISPs were significantly affected 4 h after 68% hepatectomy and ALPPS, respectively. While 89 ISPs were unique for standard hepatectomy, only the Interferon main pathway was specific for the ALPPS procedure (PAS:−0.479740209). Likewise, 120 ISPs were shared between the two procedures, with no unique ISPs identified for ALPPS 8 h after surgery (**Figure 1B**; **Supplementary Data S-II**). These data indicate that the two procedures predominantly differ quantitatively, not qualitatively. The earlier activation of the *Cell Cycle-pathway (metaphase-anaphase)* 8 and 12 h after ALPPS surgery (**Supplementary Data S-III**) signifies the initiation of chromosomal replication and segregation during cell divisions. It associated with a much earlier cell cycle entrance relative to standard hepatectomy (**Figure 1C**; **Supplementary Data S-IV**),

TABLE 1 | Comparing the intracellular signaling pathways (ISPs) of standard (68%) hepatectomy and ALPPS 4 h after surgery.

Intracellular signaling pathway	PAS (68% Hx)	PAS (ALPPS)
Androgen receptor pathway (gonadotropin regulation)	0.080962033	0.170591759
Androgen receptor pathway (histone modification)	0.080962033	0.170591759
Androgen receptor pathway (prostate differentiation and development)	0.080962033	0.170591759
Androgen receptor pathway (sexual differentiation and sexual maturation at puberty)	0.080962033	0.170591759
ATM main pathway	0.069585724	0.164331506
ATM pathway (G2_M checkpoint arrest)	0.192334187	0.758265977
BRCA 1 main pathway	−0.01618013	−0.173732307
EGFR main pathway	0.062796782	0.101805817
ErbB family main pathway	0.05751811	0.172579526
GPCR pathway (gene expression)	0.068563364	0.123092979
HGF main pathway	0.062203501	0.112692194
HGF pathway (cell cycle progression)	0.337396472	0. 546430817
Hypoxia pathway EMT 1	0.084001749	0. 691549538
Hypoxia pathway EMT 2	0.084001749	0. 691549538
Hypoxia pathway EMT 3	0.084001749	0. 691549538
Hypoxia pathway EMT 4	0.084001749	0. 691549538
ILK Main Pathway	0.074326641	0.17275502
ILK Pathway (Apoptosis)	0.091947966	0. 169660241
ILK Pathway (Cell adhesion, cell motility, opsonization)	0.095965334	0. 218655438
ILK pathway (cell cycle proliferation)	0.08622972	0.180086957
ILK pathway (cell migration, retraction)	0.094442075	0.206157845
ILK pathway (cell motility)	0.080839955	0.19368415
ILK pathway (cytoskeletal reorganization)	0.115000716	0.249836663
ILK pathway (G2-phase arrest)	0.08622972	0.180086957
ILK pathway (induced cell proliferation)	0.197724934	0.186502548
ILK pathway (regulation of intermediate filaments)	0.106090655	0.23979623
ILK pathway (regulation of junction assembly of desmosomes)	0.095197611	0.216906194
ILK pathway (wound healing)	0.095197611	0.224489964
IL-10 pathway (stability determination)	0.053478054	1. 848219244
IL-2 main pathway	0.017026437	0. 113993181
Integrin signaling main pathway	0.067439974	0.145964253
JNK pathway (apoptosis, inflammation, tumorigenesis, cell migration)	0.079482132	0.207123801
JNK pathway (insulin signaling)	−0.07241052	−0.426061537
MAPK signaling main pathway	0.0519986	0.108187021
MAPK signaling pathway (cell survival, inflammation, apoptosis, osmoregulation)	0.212081896	0.229612073
MAPK signaling pathway (gene expression)	0.091194477	0.149905425
mTOR pathway (actin organization)	0.059587334	0.128667337
p53 signaling (negative) main pathway	0.084470457	0.14626007
PAK main pathway	0.034655373	0.117954615
SMAD (negative) main pathway	0.139663741	0.315152085
SMAD (positive) main pathway	0.139663741	0.315152085
TGF beta pathway (SnON degradation)	0.166857414	0.576950547
TGF beta pathway (tumorigenesis)	0.238367734	0. 512365322
TGF beta pathway (tumor suppression)	0.238367734	0. 512365322
TNF (positive) pathway (gene expression, cell survival)	0.157317821	0.326862624
VEGF pathway (actin reorganization)	0.114882119	0.070794654

Only ISPs with PAS values ≥0.1 are shown. The most significantly affected ISPs (PAS values ≥ 0.5) are highlighted.

an earlier increased expression of cyclin E1 (**Figure 1D**) and a decline of the cell cycle inhibitor p21 mRNA (**Figure 1E**). To identify pathways that are most significantly different between ALPPS and 68% hepatectomy, we performed a more stringent analysis only including those pathways having PAS values ≥0.1. We subsequently obtained 47 and 69 affected pathways after 4 and 8 h, respectively. **Table 1** displays the PAS values of these ISPs, demonstrating that the ALPPS procedure affects the same ISPs as standard hepatectomy, but to a significantly higher extent. Of particular interest are the highly significant increased PAS values of the *ATM Pathway (G2_M Checkpoint Arrest)*, the *HGF Pathway (Cell cycle progression)*, the *EMT-associated hypoxia pathways*, and the *IL-10 Pathway (Stability determination)*. In addition, the three branches of the *TGFβ Pathway* and 10 branches of the *ILK Pathway* were significantly increased 4 h after ALPPS surgery (**Table 1**). The same set of ISPs were retrieved comparing ALPPS-specific pathways to those significantly affected 32 and 48 h after 68% hepatectomy (**Supplementary Data S-III**).

Dissecting ALPPS-Induced Pathway Activation Profiles

To further elucidate the molecular pathways responsible for accelerated liver regeneration, hepatic RNA isolated during the first 12 h after ALPPS, PVL, transection, LLLx, and sham operation was analyzed. Unsupervised hierarchical clustering separated the samples into two early groups (≤4 h post OP) and two late groups (≥8 h post OP). The dendrogram presentation revealed that ALPPS samples isolated 4 h after surgery grouped together with all late samples of PVL and ALPPS (≥8 h post OP), indicating an accelerated biological response after the ALPPS procedure (**Supplementary data S-XII**). Principal Component Analyses confirmed that the major dissimilarities between surgical procedures occurred 4 h past surgery (data not shown). In accordance with the expression data, the most distinctive differences between the groups were observed 4 h after the surgical procedures were performed. Principal Component Analyses also confirmed that the major dissimilarities between

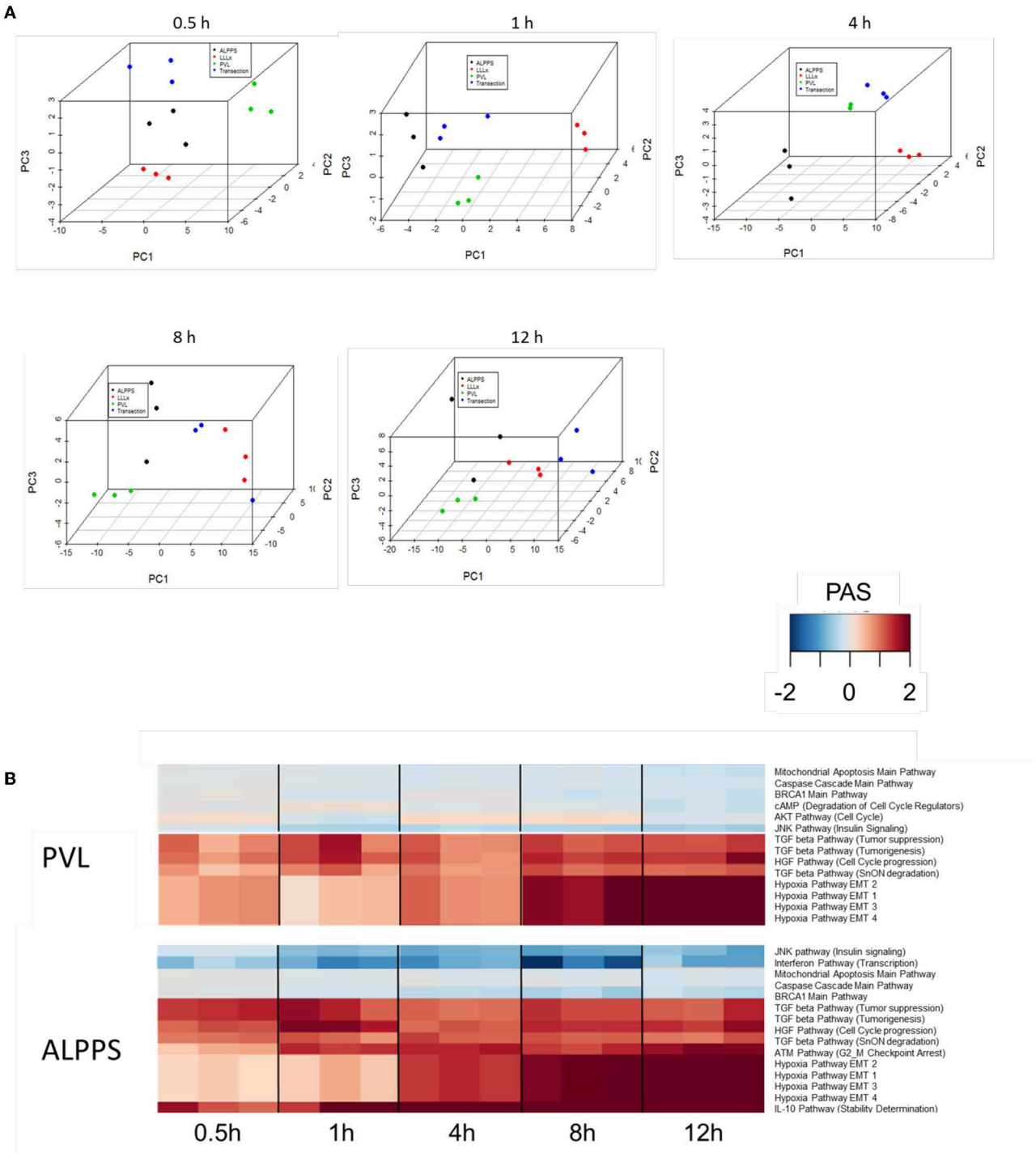


FIGURE 2 | (A) Results of the principal component analysis 0.5, 1, 4, 8, and 12 h after surgery. Green dots, left lateral liver lobectomy (LLLx); Blue dots, Transection; Red dots, portal vein ligation (PVL); Black dots, ALPPS. **(B)** Kinetics of the most significant PVL- and ALPPS-induced ISPs over time.

surgical procedures occurred 4 h after surgery (**Figure 2A**). The observation that the samples of the same surgical procedures hardly segregate (colored dots) underlines the excellent reproducibility of the different surgical procedures. The PAS data showed that ALPPS surgery significantly affected 72 ISPs, whereas PVL and transection affected 88 and 46 ISPs, respectively. The kinetics of the most significant PVL- and ALPPS-induced ISPs are presented in **Figure 2B**. Next, we

constructed Venn diagrams to find common and distinctive ISPs for ALPPS, PVL, and transection. As shown in **Table 2**, all surgical procedures involve unique ISPs. After transection, three unique ISPs are affected: the *Glucocorticoid Receptor Pathway (Gene expression),* the *Growth Hormone Pathway (Gene expression),* and the *IL-2 Pathway (Actin reorganization).* The ALPPS procedure was marked by three unique upregulated pathways [*IGF1R Signaling Pathway (Cell survival),* the *ILK Pathway (Induced cell proliferation),* the *IL-10 Pathway (Stability determination),* and the downregulation of the *Interferon Pathway (Transcription)*], whereas the PVL procedure was **b**y twenty unique ISPs (For all ISPs see: **Supplementary Data S-V**; For over-time heatmaps see: **Supplementary Data S-VI**). Remarkably, the ISPs affected by ALPPS are not just the sum of the ISPs affected by PVL plus transection, indicating that the ALPPS procedure is synergistic rather than additive. To define time-dependent ALPPS-specific ISP signatures indicated by the shifted transcriptional landscape, we plotted the PAS values as a function of time. Four hours after surgery, ten ISPs presented with a significantly increased PAS value (**Figure 3A**), five ISPs were significantly decreased (**Figure 3B**), whereas 19 ISPs demonstrated unique profiles (**Figure 3C**).

Identification of ALPPS-Specific microRNAs

Our dataset contained 73 differentially expressed long non-coding (lnc)RNA transcripts, of which 60 were identified as pre-miRNAs (**Supplementary Data S-VII**). Assuming pre-miRNA sequences as the precursors of mature miRNAs, we further analyzed their expression patterns and effects on ISPs. After removal of possible miRNA-targeted mRNA transcripts not expressed in the liver, using miRTarBase software we identified 28 known miRNAs, which together target 2,003 experimentally validated mRNAs (**Supplementary Data S-VIII**). How the distribution of miRNA expression patterns changes over time after ALPPS, PVL, and transection is presented in **Figure 4**. We identified six miRNAs, which may serve prominent roles in the biological processes involved in accelerated liver regeneration. *Mmu-miR 466i-3p* and *mmu-miR 466i-5p* were exclusively expressed 1 h after ALPPS surgery, whereas *mmu-miR 675-3p* and *mmu-miR 675-5p* were exclusively observed 12 h after ALPPS surgery. In addition, *mmu-miR 3470a* and *mmu-miR-3470b* were detected 4 h after ALPPS surgery, whereas they first appeared in the PVL samples 8 h post-surgery.

ISPs Analysis in Human ALPPS

Finally, we assessed to what extent the observations with our mouse model of ALPPS reflect the biology of accelerated liver regeneration in humans. Considering the difficulty of sampling humans' hepatic samples, we isolated RNA from liver biopsies taken just before initiating and concluding the first stage of the ALPPS procedure (4–5 h apart; $n = 5$). Again, we built ISPs with normalized RNAseq expression data using the OncoFinder software and plotted the PAS values of the 35 most affected ISPs (**Figure 5A**). The heatmap clearly demarcated two groups with similar ISP activation patterns representing the biopsies taken from the occluded lobe and the FLR, respectively. Principal Component Analyses confirmed the major dissimilarities between the two groups (**Figure 5B**). The five most activated ISPs in the FLR were identified as the *STAT3 Pathway* [*G1_to_S_Cell_Cycle_Progression* and the *STAT3 Pathway (Anti-Apoptosis)*], the *JAK-STAT Pathway (Gene_Expression_via_MYC)*, the *IL-10 Pathway (Stability determination)*, the *Estrogen Pathway (Vasodilatation)*, and the *Akt-Signaling Pathway (Regulation of Na^+ transport)*. The five most repressed pathways were identified as the *Glucocorticoid Receptor Signaling Pathway (Cell_Cycle_Arrest)*, the *Glucocorticoid Receptor Signaling Pathway (Histone_Deacetylation)*, the *EGF Pathway (Rab5_Regulation_Pathway)*, the *IGF1R Signaling Pathway (Glucose_Uptake)*, and the *BRCA1 Pathway (Base_Excision_Repair)*. A comprehensive list of all affected pathways, including the PAS values, can be found in the **Supplementary Data (S-IX)**. Finally, we compared the ALPPS-specific pathways identified in our mouse model to those

TABLE 2 | Activated and repressed ISPs unique for Transection, ALPPS, and PVL 4 h after surgery.

ISPs exclusively in ALPPS	PAS
IGF1R signaling pathway (cell survival)	0.09677963
ILK pathway (induced cell proliferation)	0.18650255
IL-10 pathway (stability determination)	1.84821924
Interferon pathway (transcription)	−0.47974021
ISPs exclusively in PVL	
AKT pathway (cell cycle)	0.09431465
Androgen receptor pathway (cell survival and cell growth)	0.1243162
ATM pathway (cell survival)	0.05295231
cAMP main pathway	0.05587309
cAMP pathway (degradation of cell cycle regulators)	−0.04355851
Erythropoeitin main pathway	0.04643793
Hedgehog pathway (repression of Hh, BMP)	0.08050423
HGF Pathway (Anoikis)	0.20617418
HIF1-Alpha main pathway	0.06045947
HIF1Alpha pathway (gene expression)	0.14085828
HIF1Alpha pathway (NOS pathway)	0.16252878
HIF1Alpha pathway (Pyruvate)	0.14085828
HIF1Alpha pathway (VEGF pathway)	0.13205464
IL-10 main pathway	0.27795708
Integrin signaling pathway (cytoskeleton contraction integrin modulation cell invasion and migration)	0.12105071
Interferon main pathway	0.03386061
IP3 main pathway	0.03666389
MAPK family pathway (gene Expression)	0.03057693
RAS main pathway	0.04229486
TGF beta pathway (post-transcriptional G1 arrest)	0.13784324
ISPs exclusively in transection	
Glucocorticoid receptor pathway (gene expression)	−0.01557421
Growth hormone pathway (gene expression)	0.02746121
IL-2 pathway (actin reorganization)	0.4806433

Positive PAS, activated ISPs; Negative PAS values, repressed ISPs. ISPs with PAS. The most significantly affected ISPs (up and down) are highlighted.

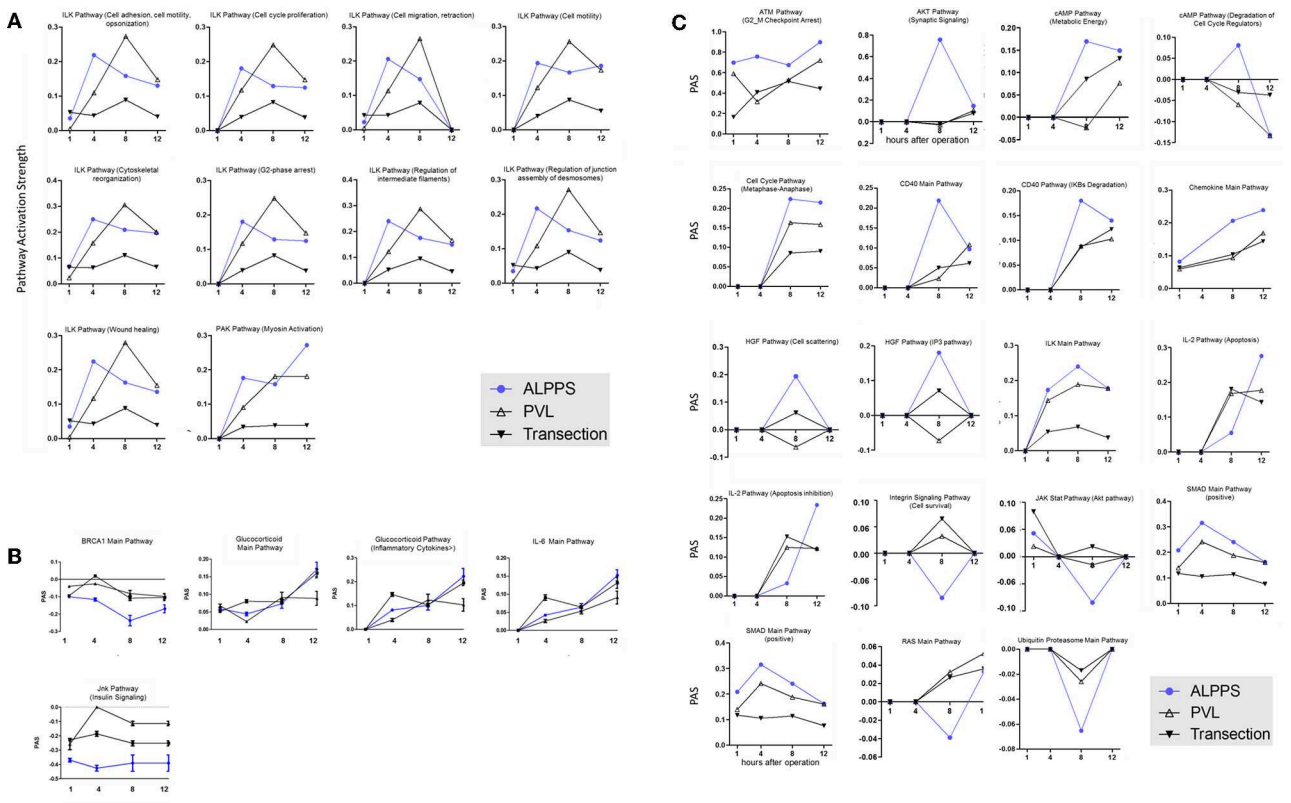


FIGURE 3 | PAS values for common ISPs as a function of time after Transection-, PVL-, and ALPPS-surgery. **(A)** Relative to PVL and Transection, 10 ISPs presented with a significantly increased PAS value 4 h after ALPPS surgery. **(B)** Significantly decreased ISPs 4 h after ALPPS surgery (relative to either PVL or transection). **(C)** ISPs with a unique profile after ALPPS surgery.

obtained with human liver samples. As demonstrated in **Figure 5C**, two of the four ALPPS-specific ISPs of mice are also upregulated in humans, immediately after ALPPS surgery step 1 [*IGF1R Signaling Pathway (Cell survival)* and the *IL-10 Pathway (Stability determination)*]. In contrast, the *ILK Pathway (Induced cell proliferation)* and the *IFN main pathway* were not significantly affected. In the human samples, we also retrieved the ISPs that were highly and significantly enriched in ALPPS when compared to standard hepatectomy, including the *ATM Pathway (G2_M Checkpoint Arrest)*, the *HGF Pathway (Cell cycle progression)*, the *EMT-associated hypoxia pathways*, and the ensuing activation of the *HIF1Alpha pathway via Jun_CREB3* (**Figure 5C**).

DISCUSSION

The novel two-staged hepatectomy ALPPS, which combines PVL plus a parenchymal liver transection, has gained increasing interest among liver surgeons. Its importance lies in the observation that ALPPS induces accelerated liver hypertrophy, so that hepatectomy can be completed within a very short time frame (7–10 days post stage 1). Previously, we reported that the *Ihh* gene is one of the 50 most upregulated genes 4 h after ALPPS surgery in mice and present in serum of patients shortly after the ALPPS procedure (21). We here observed that *Hedgehog signaling* is not unique to ALPPS surgery, since it is also activated during normal liver regeneration after 68% hepatectomy—a finding in accordance with earlier studies (32). Importantly, however, hedgehog signaling after 68% hepatectomy was induced primarily through *shh*, while *ihh* is the dominant morphogen after ALPPS (21).

Our current data demonstrate that ALPPS surgery is unique in that it induced an earlier activation of the cell cycle, signified by increased expression of cyclin E1 and a decline of the cell cycle inhibitor p21. In search of an ALPPS-specific signature, we observed four ISPs that demarcated the ALPPS procedure from PVL and transection. After ALPPS surgery, the *IGF1R Signaling Pathway (Cell survival)*, the *ILK Pathway (Induced cell proliferation)*, and the *IL-10 Pathway (Stability determination)* were significantly enriched, whereas the activity of the *Interferon Pathway (Transcription)* was reduced. The *IGF1R signaling pathway* for cell survival has been associated with early wound healing and liver cell proliferation via the IGF1R/IRS1/ERK axis and activation of the cyclins A1 and D1. Disrupting hepatic IGF1R signaling has been shown to significantly impair hepatocyte proliferation in a mouse model of liver regeneration, and IGFBP-1 null mutants show abnormal

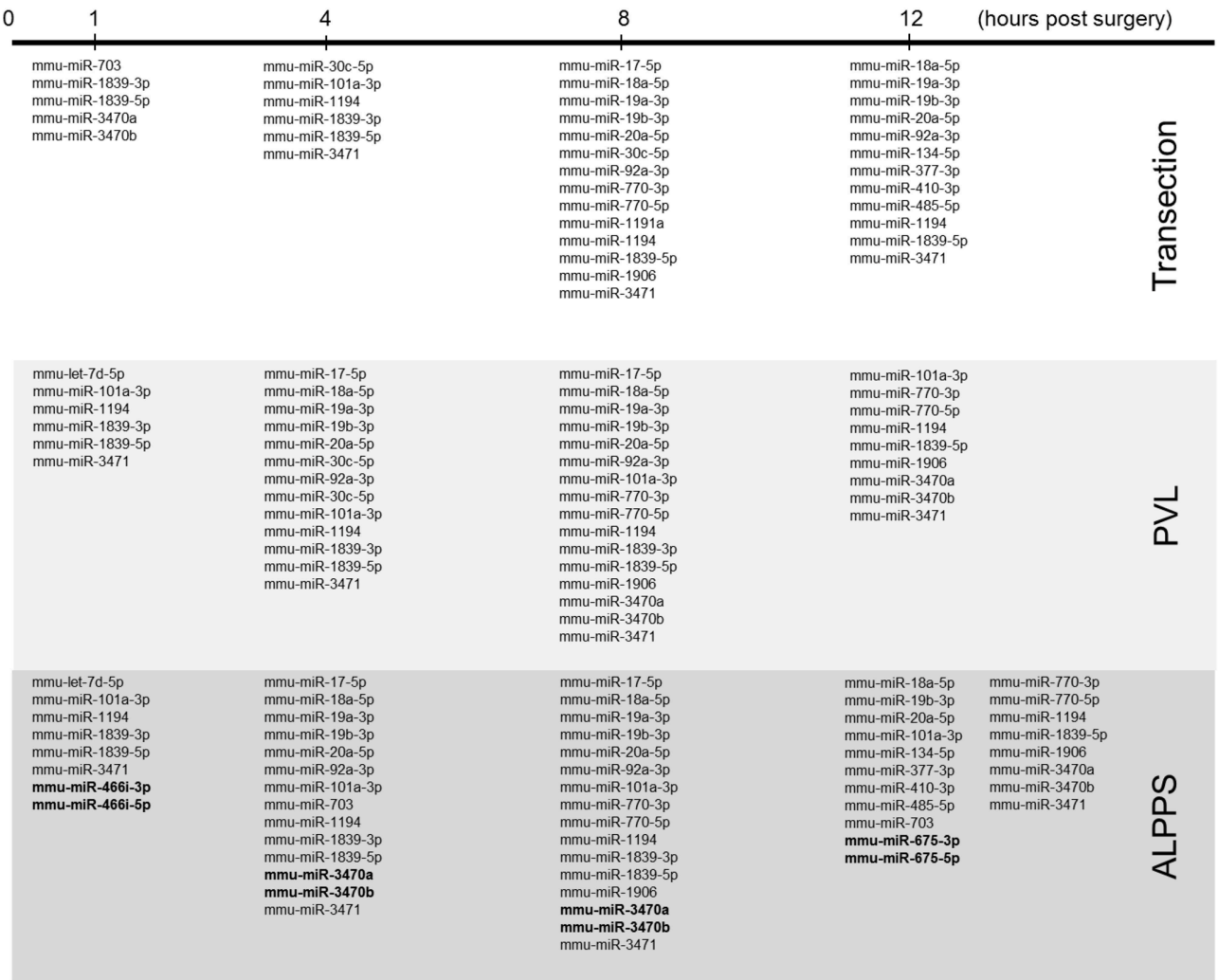


FIGURE 4 | miRNA expression patterns as a function of time after surgery (Transection, PVL and ALPPS). miRNAs specific for ALPPS, as well as miRNAs expressed at an earlier time point, are highlighted in bold.

liver regeneration (33). Increased IGF1R signaling may thus be reflecting an accelerated regeneration process. The integrin-linked kinase (ILK) is a protein involved in transmitting extracellular matrix signals and has been associated with the termination of liver regeneration. Mice lacking hepatic ILK expression cannot appropriately complete the liver regeneration process, and DNA synthesis in such mice is prolonged (34). Our data now indicate that a specific branch of the ILK pathway is also crucial to the early phase of ALPPS-induced accelerated liver regeneration, possibly contributing to the early matrix production. ALPPS surgery is also unique in that it decreased the *IFN-Main Pathway* and concomitantly induced a branch of the *IL10 Pathway*. Earlier, it was reported that IFN-γ deficiency in mice enhanced liver regeneration responses (35). In hepatocytes, IFN-γ activates Stat-mediated signaling, leading to the activation of p53, which together form transcriptionally active protein complexes to induce the cell cycle inhibitor p21 (36). The reduced IFN signaling may thus denote accelerated cell cycle progression due to diminished p21 activity. Further, the dampening of IFN signaling implies a role for IFN-γ producing cells, such as T lymphocytes, NK and NKT cells. In response to surgery-inflicted injury, these cells may receive signals not to enter the liver or to silence their activity. The activation of NK and NKT cells clearly impeded liver regeneration, which also involved IFN-γ mediated STAT1 signaling (35). The increased activity of the *IL10 pathway* is indeed suggestive of such scenarios, since it functions to diminish inflammatory responses, hence preparing an environment suitable for augmented regeneration as observed after ALPPS surgery.

In addition to these ALPPS-specific signaling pathways, we observed 10 ISPs activated at an earlier time point, paralleling the hepatic mRNA levels peaking at 4 h post ALPPS surgery. With nine distinctive branches, the ILK pathways predominated these shifted ISPs. The integrin-linked kinase

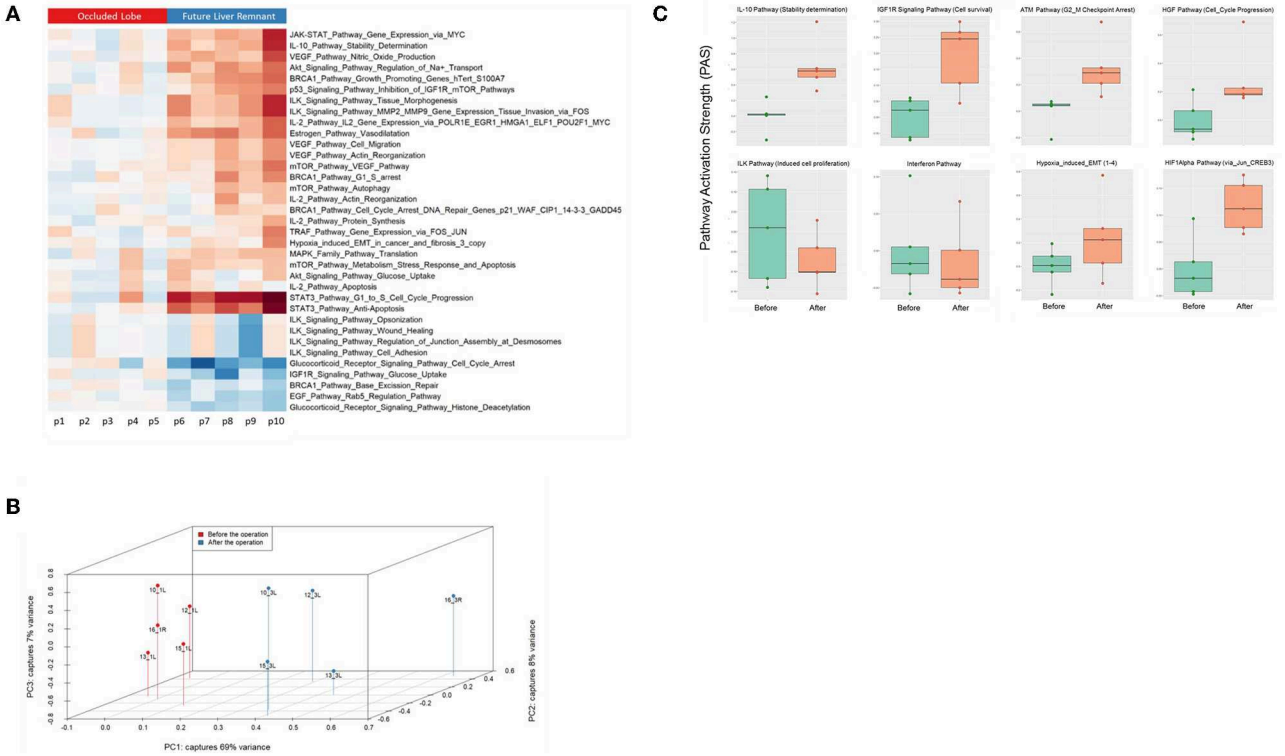


FIGURE 5 | (A) Hierarchical clustering heatmap based on the analysis of 374 intracellular signaling pathway activation profiles in human liver tissue (occluded lobe compared to future liver remnant). Red, positive PAS scores; blue, negative PAS scores for a given pathway in a given sample. **(B)** Results of the principal component analysis before and after the first step of ALPPS surgery of the human liver samples. **(C)** PAS values of ALPPS-specific pathways identified in our murine ALPPS model compared to those obtained with human liver samples.

(ILK) is a PI3-kinase-dependent, serine/threonine protein kinase that interacts with the cytoplasmic domains of both β1 and β3 integrins, possesses kinase activity, and controls an assortment of biological activities (37). The wide-ranging functionality of ILK pathways in numerous processes related to cell proliferation and tissue remodeling (see: **Figure 3A**), as well as their early and increased PAS values after ALPPS surgery, picture ILK as a plausible key regulator of accelerated liver regeneration. The earlier activation of one branch of the *PAK pathway (myosin activation)*, which is also a characteristic feature of ALPPS, may reflect the above findings, since ILK and PAK1 cross-regulate each other and together regulate cytoskeletal dynamics. A function for PAKs has also been established in cell cycle progression in cancers, in which PAKs are overexpressed or hyper-activated (38).

Comparing standard hepatectomy with ALPPS surgery, we demonstrated that 4 h after surgery the two procedures predominantly differ quantitatively, not qualitatively. The only ALPPS-specific signature was—as described above—the suppression of the *IFN Main Pathway*, which clearly demarcated ALPPS from standard hepatectomy. The majority of the affected ISPs demonstrated a low but detectable activity after normal liver regeneration. After ALPPS surgery, however, the activities of these ISPs were dramatically increased. Of particular interest are the four branches of the hypoxia-induced pathway, which are 8.2-fold higher in ALPPS than in standard hepatectomy. Hypoxia signaling has been implicated in regulating the transition that is necessary to produce the extracellular matrix but also to initiate the regenerative capacity of EMT-like liver cells (39). Although it has been proposed that *in situ* liver partition contributes to hypoxia (40), our data explicitly demonstrate that ligation of the portal vein suffices to induce the hypoxia pathways (**Supplementary Date S-I**). Hypoxia of the FLR is thus an immediate early event after PVL, probably due to an excess of oxygen-poor blood from the portal vein, and a major trigger for the accelerated ALPPS-induced regenerative response.

The importance of hypoxia as a driver of liver regeneration is increasingly being recognized. Although HIFs have traditionally been in the focus, HIF1A has been reported not to react toward hypoxia in the liver. Interestingly, HIF1A's cellular location was associated with peroxisomes rather than the nucleus upon exposure of hepatocytes to hypoxia (41). Conceivably, early changes due to the altered portal flow may provide triggers (such as activation of Kupffer cells, or the re-organization of matrix) for regeneration but so far are unproven. When the parenchyma expands, however, hypoxia develops in analogy to a growing tumor. After 70%HX, hypoxia leads to HIF2A (but not HIF1A) activation, which promotes hepatocyte mitosis and induces VEGF production for the later angiogenic phase (42). Thus, hypoxia imprints timely order on regeneration, with the

start of the angiogenic phase coupled to the successful completion of hepatocyte division. Recently, unspecific activation of HIFs by ethyl-3,4-dihydroxybenzoate was shown to accelerate PVL regeneration to ALPPS levels but had no discernible effect on liver metastases (40, 43). Induction of the HIF pathway may thus be a first clinical approach to improving liver regeneration after 68% hepatectomy or PVL.

Further, three branches of the *TGFβ-pathway*, the *IL10-Pathway* (stability determination), and the *ATM-pathway* (G2-M_Checkpoint_Arrest) are significantly enriched after the ALPPS procedure. At the molecular level, one of the key mediators of regenerative responses is the secreted cytokine transforming growth factor-β (TGFβ) (44). The same branch of the *IL-10-pathway (stability_determination)* is again recovered. Relative to normal hepatectomy, the activity of the ATM main pathway and its specific branch for G2-M checkpoint arrest were significantly increased 4 h post ALPPS surgery. The G2-M DNA damage checkpoint is an important cell cycle checkpoint ensuring that cells do not initiate mitosis before repairing damaged DNA after replication. Cells that have a defective G2-M checkpoint enter mitosis before repairing their DNA, leading to death after cell division.

Our comprehensive analysis demonstrated that the distinct surgical procedures underlying ALPPS surgery—PVL and transection—are discernible by a distinctive set of activated and repressed ISPs. It also revealed that ALPPS surgery is not merely an addition of the pathways induced by PVL plus Transection. Although the molecular mechanism for this synergism is currently unclear, it might be explained by the presence and involvement of several unique ALPPS-specific microRNAs. MicroRNAs (miRNAs) are short, single-stranded RNAs that modify gene expression at the post-transcriptional level and are heavily involved in the spatiotemporal control of gene expression during the entire process of liver regeneration. The number, nature and level of expressed miRNAs profoundly depends on the organism studied, the activation status and microenvironment of cells, as well as other unidentified factors. In rats and mice, several miRNAs have been linked to liver regeneration, including miR19a, miR21, and miR214 (45). Here, we recognized 28 precursor transcripts of 28 miRNA present after liver surgery (Transection, PVL, and ALPPS) and identified four that were uniquely and specifically enriched after ALPPS surgery. *Mmu-miR 466i-3p* and *mmu-miR 466i-5p* were uniquely expressed immediately after ALPPS surgery, indicating they may underlie the initiation of the accelerated regeneration process by deactivating the mRNAs of cell cycle inhibitors. Indeed, *miR 466i-3p* and *mmu-miR 466i-5p* intervene with many targets of ISPs that keep the cells committed and prevent them from preparing for cell cycle entry, including the *PAK Pathway (Actin_Organization)*, the *p38 pathway (Actin-Cytoskeleton_reorganization)*, the *MAPK signaling pathways*, and the *Jnk Pathway (Gene expression Apoptosis Inflammation_Tumorigenesis_Cell-Migration)* (see **Supplementary Data S-X**). Likewise, *mmu-miR 3470a* and *mmu-miR-3470b* were uniquely detected 4 h after the ALPPS procedure and accordingly coincided with the ALPPS-specific shift described earlier (21). *Mmu-miR 675-3p* and *mmu-miR 675-5p*, both exclusively observed 12 h after ALPPS surgery, exert a narrower range of biological activities, mainly silencing anti-proliferative pathways. Indeed, *miR-675* and its precursor long-non-coding RNA, H19, contribute to increased proliferation, and apoptosis inhibition (46). These ALPPS-specific miRNAs may cooperate so that the earlier described ALPPS-specific ISPs become prevalent—unhindered from competing inhibitory signals and/or ISPs. Also of interest is the observation that six of the 28 detected miRNAs target the PTEN gene (*mmu-miR-17-5p, mmu-miR-18a-5p, mmu-miR-19a-3p, mmu-miR-19b-3p, mmu-miR-20a-5p, and mmu-miR-410-3p*), thus enhancing PI3-Kinase/Akt signaling and fatty acid metabolism. The downregulation of PTEN may fuel liver growth after hepatectomy due to increased β-oxidation (47).

Comparing the human data to the murine data, we observed that they have several ISPs in common and that the direction of change is the same, indicating our ALPPS model reflects the major changes of accelerated regeneration in humans at the molecular level. That not all ALPPS-specific ISPs observed in mice are retrieved in human liver tissue may reflect distinct kinetics of the liver regeneration process: sufficient regeneration in our mice model is achieved 2 days post-surgery, while it takes 7–10 days in humans. Still, the data signify that our murine ALPPS model—despite differences—may be useful to gain insight into the molecular background underlying ALPPS-induced accelerated liver regeneration in humans. Considering the close match between ISPs in mice and humans, deeper exploration of our ALPPS model may disclose essential leads to the development of potential therapeutic strategies targeting specific ISPs, in particular those that counteract responses of the immune system.

We realize that our results require further biological validation, since RNAseq data do not give insights into protein expression changes or activities of proteins belonging to a specific pathway. Still, our study provides a comprehensive framework of the signaling pathways involved in normal and accelerated liver regeneration, which is now available for further exploration.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the ENA, PRJEB15593.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Cantonal Ethics Committee, Stampfenbachstrasse 121, 8090 Zürich Switzerland. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Veterinary Office Zürich, Zollstrasse 20 8090 Zürich, Switzerland.

AUTHOR CONTRIBUTIONS

PB: experimental design, fund raising, data analysis, and writing. MSc: experimental design, patients, and writing. LF, MSo, and

AB: data analysis. ML and EK: animals and experiments. RG and BH: fund raising, consultancy, and animal studies. P-AC: fund raising, study design, patients, and consultancy.

FUNDING

This study was funded by the Clinical Research Priority Program from the University of Zürich non-resectable liver tumors—from palliation to cure and the Swiss National Science Foundation (310030_169382).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01206/full#supplementary-material

REFERENCES

1. Jaeck D, Bachellier P, Nakano H, Oussoultzoglou E, Weber JC, Wolf P, et al. One or two-stage hepatectomy combined with portal vein embolization for initially non-resectable colorectal liver metastases. *Am J Surg*. (2003) 185:221–9. doi: 10.1016/S0002-9610(02)01373-9
2. Clavien PA, Petrowsky H, DeOliveira ML, Graf R. Strategies for safer liver surgery and partial liver transplantation. *N Engl J Med*. (2007) 356:1545–59. doi: 10.1056/NEJMra065156
3. Schnitzbauer AA, Lang SA, Goessmann H, Nadalin S, Baumgart J, Farkas SA, et al. Right portal vein ligation combined with *in situ* splitting induces rapid left lateral liver lobe hypertrophy enabling 2-staged extended right hepatic resection in small-for-size settings. *Ann Surg*. (2012) 255:405–14. doi: 10.1097/SLA.0b013e31824856f5
4. Schadde E, Ardiles V, Slankamenac K, Tschuor C, Sergeant G, Amacker N, et al. ALPPS offers a better chance of complete resection in patients with primarily unresectable liver tumors compared with conventional-staged hepatectomies: results of a multicenter analysis. *World J Surg*. (2014) 38:1510–9. doi: 10.1007/s00268-014-2513-3
5. Clavien PA, Lillemoe KD. Note from the editors on the ALPPS e-Letters-to-the-Editor. *Ann Surg*. (2012) 256:552. doi: 10.1097/SLA.0b013e318266fa1f
6. de Santibanes E, Clavien PA. Playing play-doh to prevent postoperative liver failure: the "ALPPS" approach. *Ann Surg*. (2012) 255:415–7. doi: 10.1097/SLA.0b013e318248577d
7. Petrowsky H, Györi G, DeOliveira M, Lesurtel M, Clavien PA. Is partial-ALPPS safer than ALPPS? A single-center experience. *Ann Surg*. (2015) 261:e90–2. doi: 10.1097/SLA.0000000000001087
8. Leffert HL, Koch KS, Moran T, Rubalcava B. Hormonal control of rat liver regeneration. *Gastroenterology*. (1979) 76:1470–82. doi: 10.1016/0016-5085(79)90418-9
9. Selzner N, Selzner M, Odermatt B, Tian Y, Van Rooijen N, Clavien PA. ICAM-1 triggers liver regeneration through leukocyte recruitment and Kupffer cell-dependent release of TNF-alpha/IL-6 in mice. *Gastroenterology*. (2003) 124:692–700. doi: 10.1053/gast.2003.50098
10. Feingold KR, Soued M, Grunfeld C. Tumor necrosis factor stimulates DNA synthesis in the liver of intact rats. *Biochem Biophys Res Commun*. (1988) 153:576–82. doi: 10.1016/S0006-291X(88)81134-3
11. Russell WE. Transforming growth factor beta (TGF-beta) inhibits hepatocyte DNA synthesis independently of EGF binding and EGF receptor autophosphorylation. *J Cell Physiol*. (1988) 135:253–61. doi: 10.1002/jcp.1041350212
12. Jia C. Advances in the regulation of liver regeneration. *Expert Rev Gastroenterol Hepatol*. (2011) 5:105–21. doi: 10.1586/egh.10.87
13. Bucher ML, Swaffield MN. Regulation of hepatic regeneration in rats by synergistic action of insulin and glucagon. *Proc Natl Acad Sci USA*. (1975) 72:1157–60. doi: 10.1073/pnas.72.3.1157
14. Huang W, Ma K, Zhang J, Qatanani M, Cuvillier J, Liu J, et al. Nuclear receptor-dependent bile acid signaling is required for normal liver regeneration. *Science*. (2006) 312:233–6. doi: 10.1126/science.1121435
15. Lesurtel M, Graf R, Aleil B, Walther DJ, Tian Y, Jochum W, et al. Platelet-derived serotonin mediates liver regeneration. *Science*. (2006) 312:104–7. doi: 10.1126/science.1123842
16. Cressman DE, Greenbaum LE, Haber BA, Taub R. Rapid activation of post-hepatectomy factor/nuclear factor kappa B in hepatocytes, a primary response in the regenerating liver. *J Biol Chem*. (1994) 269:30429–35.
17. Columbano A, Ledda-Columbano GM, Pibiri M, Piga R, Shinozuka H, De Luca V, et al. Increased expression of c-fos, c-jun and LRF-1 is not required for *in vivo* priming of hepatocytes by the mitogen TCPOBOP. *Oncogene*. (1997) 14:857–63. doi: 10.1038/sj.onc.1200891
18. Stepniak E, Ricci R, Eferl R, Sumara G, Sumara I, Rath M, et al. c-Jun/AP-1 controls liver regeneration by repressing p53/p21 and p38 MAPK activity. *Genes Dev*. (2006) 20:2306–14. doi: 10.1101/gad.390506
19. Riehle KJ, Campbell JS, McMahan RS, Johnson MM, Beyer RP, Bammler TK, et al. Regulation of liver regeneration and hepatocarcinogenesis by suppressor of cytokine signaling 3. *J Exp Med*. (2008) 205:91–103. doi: 10.1084/jem.20070820
20. Ho KJ, Do NL, Otu HH, Dib MJ, Ren X, Enjyoji K, et al. Tob1 is a constitutively expressed repressor of liver regeneration. *J Exp Med*. (2010) 207:1197–208. doi: 10.1084/jem.20092434
21. Langiewicz M, Schlegel A, Saponara E, Linecker M, Borger P, Graf R, et al. Hedgehog pathway mediates early acceleration of liver regeneration induced by a novel two-staged hepatectomy in mice. *J Hepatol*. (2017) 66:560–70. doi: 10.1016/j.jhep.2016.10.014
22. Buzdin AA, Zhavoronkov AA, Korzinkin MB, Venkova LS, Zenin AA, Smirnov PY, et al. Oncofinder, a new method for the analysis of intracellular signaling pathway activation using transcriptomic data. *Front Genet*. (2015) 5:55. doi: 10.3389/fgene.2014.00055
23. Borisov N, Suntsova M, Sorokin M, Garazha A, Kovalchuk O, Aliper A, et al. Data aggregation at the level of molecular pathways improves stability of experimental transcriptomic and proteomic data. *Cell Cycle*. (2017) 16:1810–23. doi: 10.1080/15384101.2017.1361068
24. Alexandrova E, Nassa G, Corleone G, Buzdin A, Aliper AM, Terekhanova N, et al. Large-scale profiling of signaling pathways reveals an asthma specific signature in bronchial smooth muscle cells. *Oncotarget*. (2016) 7:25150–61. doi: 10.18632/oncotarget.7209
25. Buzdin AA, Zhavoronkov AA, Korzinkin MB, Roumiantsev SA, Aliper AM, Venkova LS, et al. The OncoFinder algorithm for minimizing the errors introduced by the high-throughput methods of transcriptome analysis. *Front Mol Biosci*. (2014) 1:8. doi: 10.3389/fmolb.2014.00008
26. Schlegel A, Lesurtel M, Melloul E, Limani P, Tschuor C, Graf R, et al. ALPPS: from human to mice highlighting accelerated and novel mechanisms of liver regeneration. *Ann Surg*. (2014) 260:839–47. doi: 10.1097/SLA.0000000000000949
27. Lehmann K, Tschuor C, Rickenbacher A, Jang JH, Oberkofler CE, Tschopp O, et al. Liver failure after extended hepatectomy in mice is mediated by a p21-dependent barrier to liver regeneration. *Gastroenterology*. (2012) 143:1609–19. doi: 10.1053/j.gastro.2012.08.043
28. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. (2013) 14:R36. doi: 10.1186/gb-2013-14-4-r36
29. Culhane AC, Thioulouse J, Perriere G, Higgins DG. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics*. (2005) 21:2789–90. doi: 10.1093/bioinformatics/bti394
30. Scales M, Jager R, Migliorini G, Houlston RS, Henrion MY. VisPIG—a web tool for producing multi-region, multitrack, multi-scale plots of genetic data. *PLoS ONE*. (2014) 9:e107497. doi: 10.1371/journal.pone.0107497
31. Artcibasova AV, Korzinkin MB, Sorokin MI, Shegay PV, Zhavoronkov AA, Gaifullin N, et al. MiRImpact, a new bioinformatic method using complete microRNA expression profiles to assess their overall influence on the

activity of intracellular molecular pathways. *Cell Cycle*. (2016) 15:689–98. doi: 10.1080/15384101.2016.1147633
32. Ochoa B, Syn WK, Delgado I, Karaca GF, Jung Y, Wang J, et al. Hedgehog signaling is critical for normal liver regeneration after partial hepatectomy in mice. *Hepatology*. (2010) 51:1712–23. doi: 10.1002/hep.23525
33. Desbois-Mouthon C, Wendum D, Cadoret A, Rey C, Leneuve P, Blaise A, et al. Hepatocyte proliferation during liver regeneration is impaired in mice with liver-specific IGF-1R knockout. *FASEB J*. (2006) 20:773–5. doi: 10.1096/fj.05-4704fje
34. Apte U, Gkretsi V, Bowen WC, Mars WM, Luo JH, Donthamsetty S, et al. Enhanced liver regeneration following changes induced by hepatocyte-specific genetic ablation of integrin-linked kinase. *Hepatology*. (2009) 50:844–51. doi: 10.1002/hep.23059
35. Sun R, Gao B. Negative regulation of liver regeneration by innate immunity (natural killer cells/interferon-gamma). *Gastroenterology*. (2004) 127:1525–39. doi: 10.1053/j.gastro.2004.08.055
36. Horras CJ, Lamb CL, Mitchell KA. Regulation of hepatocyte fate by interferon-γ. *Cytokine Growth Factor Rev*. (2011) 22:35–43. doi: 10.1016/j.cytogfr.2011.01.001
37. Hannigan GE, McDonald PC, Walsh MP, Dedhar S. Integrin-linked kinase: not so 'pseudo' after all. *Oncogene*. (2011) 30:4375–85. doi: 10.1038/onc.2011.177
38. Molli PR, Li DQ, Murray BW, Rayala SK, Kumar R. PAK signaling in oncogenesis. *Oncogene*. (2009) 28:2545–55. doi: 10.1038/onc.2009.119
39. Xue Z, Wu X, Ming L. Hepatic regeneration and the epithelial to mesenchymal transition. *World J Gastroenterol*. (2013) 19:1380–6. doi: 10.3748/wjg.v19.i9.1380
40. Schadde E, Tsatsaris C, Swiderska-Syn M, Breitenstein S, Urner M, Schimmer R, et al. Hypoxia of the growing liver accelerates regeneration. *Surgery*. (2017) 161:666–79. doi: 10.1016/j.surg.2016.05.018
41. Khan Z, Michalopoulos GK, Stolz DB. Peroxisomal localization of hypoxia-inducible factors and hypoxia-inducible factor regulatory hydroxylases in primary rat hepatocytes exposed to hypoxia-reoxygenation. *Am J Pathol*. (2006) 169:1251–69. doi: 10.2353/ajpath.2006.060360
42. Kron P, Linecker M, Limani P, Schlegel A, Kambakamba P, Lehn JM, et al. Hypoxia-driven Hif2a coordinates mouse liver regeneration by coupling parenchymal growth to vascular expansion. *Hepatology*. (2016) 64:2198–209. doi: 10.1002/hep.28809
43. Harnoss JM, Platzer LK, Burhenne J, Radhakrishnan P, Cai J, Strowitzki MJ, et al. Prolyl hydroxylase inhibition enhances liver regeneration without induction of tumor growth. *Ann Surg*. (2017) 265:782–91. doi: 10.1097/SLA.0000000000001696
44. Karkampouna S, Ten Dijke P, Dooley S, Julio MK. TGFβ signaling in liver regeneration. *Curr Pharm Des*. (2012) 18:4103–13. doi: 10.2174/138161212802430521
45. Yi PS, Zhang M, Xu MQ. Role of microRNA in liver regeneration. *Hepatobiliary Pancreat Dis Int*. (2016) 15:141–6. doi: 10.1016/S1499-3872(15)60036-4
46. Liu G, Xiang T, Wu QF, Wang WX. Long non-coding RNA H19-Derived miR-675 enhances proliferation and invasion via RUNX1 in gastric cancer cells. *Oncol Res*. (2016) 23:99–107. doi: 10.3727/096504015X14496932933575
47. Kachaylo E, Tschuor C, Calo N, Borgeaud N, Ungethüm U, Limani P, et al. PTEN down-regulation promotes β-oxidation to fuel hypertrophic liver growth after hepatectomy in mice. *Hepatology*. (2017) 66:908–21. doi: 10.1002/hep.29226

Conflict of Interest: AB was employed by company OmicsWay Corp., Walnut, CA, United States and Oncobox ltd., Moscow, Russia. MSo was employed by company OmicsWay Corp., Walnut, CA, United States.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

CASE REPORT
published: 17 December 2019
doi: 10.3389/fonc.2019.01390

The Molecular Signature More Than the Site of Localization Defines the Origin of the Malignancy

*Antonio Matrone[1], Liborio Torregrossa[2], Elisa Sensi[2], Daniele Cappellani[1], Walter Baronti[1], Raffaele Ciampi[1], Eleonora Molinaro[1], Clara Ugolini[2], Aleksandr Aghababyan[3], Luigi De Napoli[3], Francesco Latrofa[1], Gabriele Materazzi[3], Fulvio Basolo[2], Paolo Vitti[1] and Rossella Elisei[1]**

[1] Unit of Endocrinology, Department of Clinical and Experimental Medicine, Pisa University Hospital, Pisa, Italy, [2] Anatomic Pathology Section, Department of Surgical, Medical, Molecular Pathology and Critical Area, Pisa University Hospital, Pisa, Italy, [3] Unit of Endocrine Surgery, Department of Surgical, Medical, Molecular Pathology and Critical Area, Pisa University Hospital, Pisa, Italy

The diagnosis of the primary origin of metastases to the thyroid gland is not easy, in particular in case of concomitant lung adenocarcinoma which shares several immunophenotypical features. Although rare, these tumors should be completely characterized in order to set up specific therapies. This is the case of a 64-years-old woman referred to our institution for a very advanced neoplastic disease diagnosed both as poorly differentiated/anaplastic thyroid cancer (PDTC/ATC) for the huge involvement of the neck and concomitant lung adenocarcinoma (LA). Neither the clinical features and the imaging evaluation nor the tumor markers allowed a well-defined diagnosis. Moreover, the histologic features of the thyroid and lung biopsies confirmed the synchronous occurrence of two different tumors. The molecular analysis showed a c.34G>T (p.G12C) mutation in the codon 12 of K-RAS gene, in both tissues. Since, this mutation is highly prevalent in LA and virtually absent in PDTC/ATC the lung origin of the malignancy was assumed, and the patient was addressed to the correct therapeutic strategy.

Keywords: cancer, thyroid cancer, lung cancer, mutation—genetics, molecular oncology

OPEN ACCESS

Edited by:
Ira Ida Skvortsova, Innsbruck Medical University, Austria

Reviewed by:
Enke Baldini, Sapienza University of Rome, Italy
Ricardo Celestino, University of Porto, Portugal

****Correspondence:***
Rossella Elisei
rossella.elisei@med.unipi.it

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *28 May 2019*
Accepted: *25 November 2019*
Published: *17 December 2019*

Citation:
Matrone A, Torregrossa L, Sensi E, Cappellani D, Baronti W, Ciampi R, Molinaro E, Ugolini C, Aghababyan A, De Napoli L, Latrofa F, Materazzi G, Basolo F, Vitti P and Elisei R (2019) The Molecular Signature More Than the Site of Localization Defines the Origin of the Malignancy. Front. Oncol. 9:1390. doi: 10.3389/fonc.2019.01390

BACKGROUND

Metastases to thyroid gland (TGM) from other primary tumors are rare entities and frequently reported in autoptic series (1, 2). The incidence of TGM widely varies in different series and has been related both to the site of the primary tumors and the ethnicity of the patients. TGM are more frequently caused by breast, lung, and kidney cancer (3–5). TGM has been reported also in case of gastrointestinal tract cancer, mainly in the Asian population (6, 7). The clinical presentation could be characterized by a rapid growth of a thyroid nodule with marked symptoms of dysphagia and/or dysphonia. In other cases, TGM show a clinical silent course without specific symptoms and were occasionally discovered after imaging procedures (neck US, CT scan, ^{18fdg}PET-CT, etc.) during the assessment of the primary tumor (2, 4). Concomitant metastases in other organs are hardly ever present, and thyroid function is usually not compromised. Differential diagnosis with primary thyroid cancer may be done by fine needle aspiration cytology followed by immunocitochemistry evaluation. However, it is not easy to distinguish thyroid cancer from a thyroid metastasis particularly if derived from lung adenocarcinoma. Indeed, thyroid cancer and

lung adenocarcinoma could have several common histopathologic features such as pale nuclei, finely granular chromatin, occasional nuclear grooves, and intra-cytoplasmic inclusions (2, 8, 9) that can make difficult a well-defined distinction between the two cancer entities even after the biopsy. Thus, in these cases, other techniques are required to perform a differential diagnosis that is crucial for choosing the best therapeutic strategy. We report a case of lung TGM for which we needed to turn to a next generation sequencing method to analyze specific molecular alterations to distinguish a primary thyroid cancer from a TGM.

CASE PRESENTATION

A 64-years-old woman was referred to our institution for a second opinion about a very advanced neoplastic disease involving both the thyroid and the lung. The initial diagnosis was a poorly differentiated/anaplastic thyroid cancer (PDTC/ATC) with lung and bone metastases. The tumor mass was considered not surgically resectable because of the infiltration of the thyroid and cricoid cartilages, as well as of the carotid artery and the right jugular vein. An emergency tracheotomy with position of the 6 mm Shiley cannula because of the severe dyspnea, was performed. After the exclusion of contraindications and the placement of peripherally central catheter (PICC) in superior vena cava, a chemotherapy with Taxol (80 mg/m^2) and Carboplatinum (AUC = 2) once weekly for 4 weeks, every 6 weeks, was performed and completed.

When the patient was referred to our Department, we discovered that she was a strong and old smoker (about 20 cigarettes daily for about 40 years). The physical examination showed hypotonic and hypotrophic muscles, and the presence of the enlargement of the neck with a large stiff mass; some palpable lymph nodes were evident in latero-cervical compartment, bilaterally. No other abnormalities were noted at physical examination except for the presence of tracheotomy. The patient was in medical therapy with atenolol for sporadic arrhythmia, steroids as anti-edema and omeprazole for gastric pain prevention.

To assess the neoplastic status of the patient, we performed a total body CT scan with i.v. contrast medium that showed the presence of multiple metastatic lymph nodes (max diam 17 mm in the right latero-cervical region) and a large thyroid gland with increased dimension in particular in the right lobe (**Figure 1A**). Moreover, the CT scan showed a large mass (9.0 × 4.5 cm) in the mediastinum, linked to the thyroid mass, involving the left upper lung lobe and occluding the left pulmonary artery branch (**Figure 1B**). Multiple metastatic lymph nodes were evident in paratracheal and paraesophageal regions. Two sclerotic lesions of L11 and D6 were also discovered.

Surgery was not indicated for the huge infiltration of the nearby vital structures. Radiotherapy counseling excluded the possibility to perform an external beam radiotherapy (EBRT) for the extension of the disease. Orthopedic counseling did not indicate any treatment for L11 and D6 lesions because there was no evidence of vertebral fractures. The only therapeutic option could be a systemic therapy, the most possible specific for the type of malignancy.

Serum tumor markers were all elevated and did not provide any specific information. In particular, Ca 125 was 35.9 U/ml (<35), Ca 15.3 was 169.3 U/ml (<25), Ca 19.9 was 59.3 U/ml (<39), CEA was 8.3 ng/ml (<5.2), CYFRA 21.1 was 8.8 ng/ml (<3.3).

Bronchoscopy showed the paralysis of left vocal cord with hypomobility of the right, arytenoids edema with reduced glottic space; the exploration of the bronchial tube showed the infiltration of the left upper lobe bronchus (LULB) with massive stenosis. Then, we performed a LULB brushing and a biopsy. To confirm the histological diagnosis, we decided to perform a Tru-Cut biopsy with 3 biopsies samples (2 left and 1 right) on the neck mass.

The thyroid Tru-Cut biopsy showed a poorly differentiated thyroid carcinoma (PDTC) with several anaplastic areas (**Figure 1C**). The immunohistochemistry for the thyroid specific proteins was TTF-1 focally positive, Tg, and PAX-8 negative (**Figures 1E–G**). The bronchus brushing showed the presence of an adenocarcinoma with widespread mucinous aspects, confirmed by biopsy (**Figure 1D**). The immunohistochemistry was CK7 positive, TTF-1, Napsin A, Tg, and PAX-8 negative (**Figures 1H–J**).

We than hypothesized that the patient could have two different types of cancer, PDTC/ATC of the thyroid and a mucinous adenocarcinoma of the lung. To solve this question, we decided to perform the molecular analysis of both the thyroid and bronchus biopsies.

The analysis was performed with a next generation sequencing method based on DNA extraction and MALDI-TOF mass spectrometry from paraffin-embedded tissues using CE-IVD validated kits "Myriapod Lung Status" on the "MassARRAY system" (Sequenom). Data were analyzed using the software "MassARRAY Analyzer 4" and "iGenetics Myriapod" (Diatech Pharmacogenetics). The analyzed genes were selected from a series of genes whose alterations are different in the two types of the tumors (**Table 1**) (10, 11).

Both, thyroid (**Figures 1K,L**) and lung (**Figures 1M,N**) samples showed the substitution c.34G>T (p.G12C) in the codon 12 of K-RAS gene, while all the other mutations analyzed were negative. The patient was referred to the oncology unit who agreed with our hypothesis that the thyroid mass could be a metastasis of the lung adenocarcinoma. The patient was submitted to other 4 cycles of chemotherapy with the same scheme previously performed, but she died 3 months later.

DISCUSSION

There are some oncologic cases whose diagnoses are not easy because of the severe degree of de-differentiation of the tumor cells. This differential diagnosis is particularly difficult when a thyroid malignant lesion and a lung malignant lesion are simultaneously present in the same patient. Immunohistochemistry for some specific thyroid genes and corresponding proteins can be helpful, but we have to consider

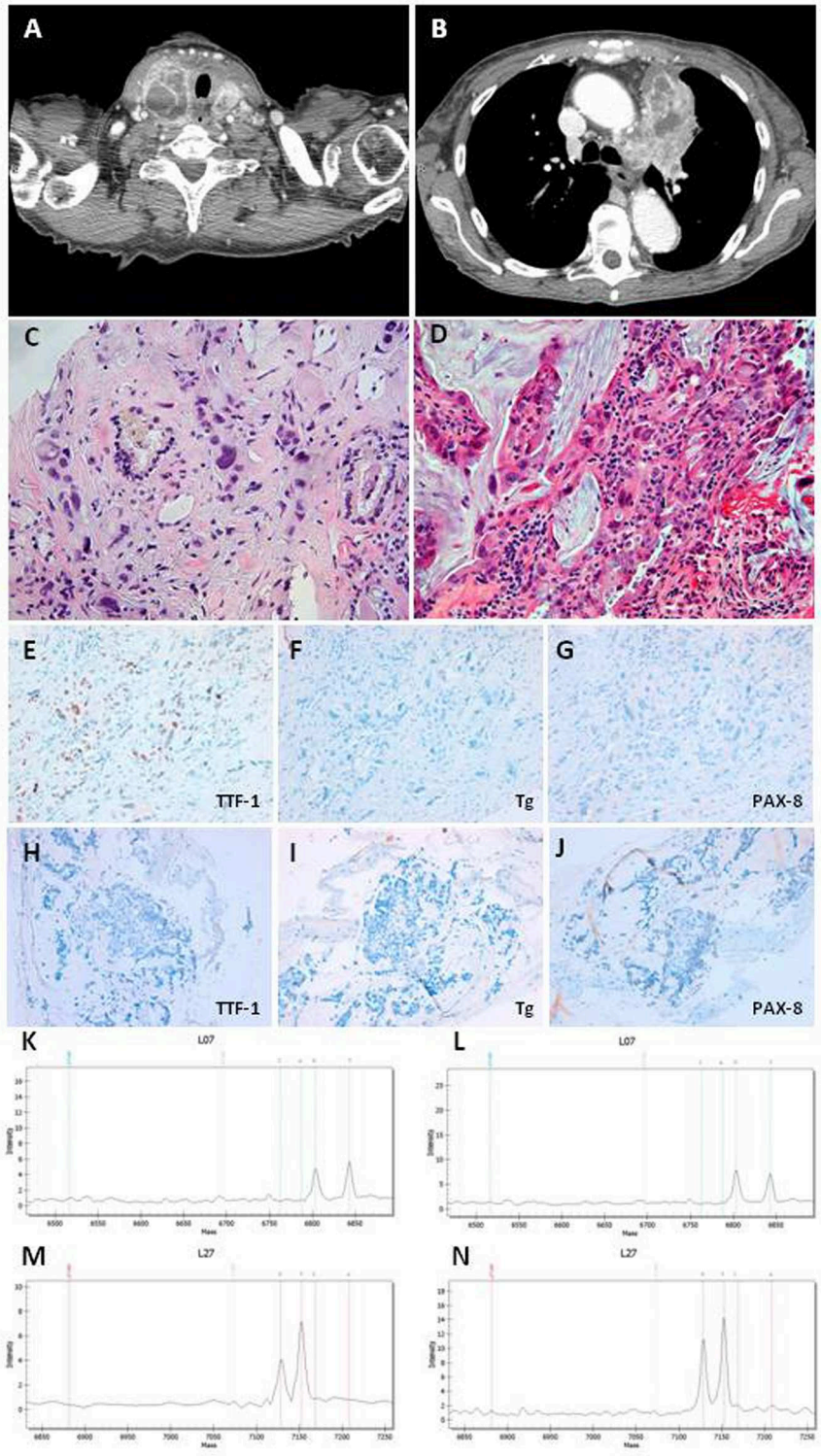


FIGURE 1 | (A) CT scan section of thyroid neoplasia. **(B)** CT scan section of lung neoplasia. **(C)** Undifferentiated carcinoma with marked nuclear atypia intermixed to thyroid follicles in thyroid Tru-cut. Magnifications: x200. **(D)** Adenocarcinoma with mucinous differentiation in bronchial biopsy. Magnifications: x200. **(E–G)** Immunohistochemistry in thyroid Tru-cut showing focal immunoreactivity for TTF-1 and absence of immunoreactivity for Thyroglobulin and PAX-8 markers. Original magnifications: x200. **(H–J)** Immunohistochemistry in bronchial biopsy showing absence of immunoreactivity for TTF-1, Thyroglobulin, and PAX-8 markers. Original magnifications: x200. **(K,L)** Mass spectrometry of thyroid neoplasia with substitution c.34G>T (p.G12C) in the codon 12 of kRAS gene. **(M,N)** Mass spectrometry of lung neoplasia with substitution c.34G>T (p.G12C) in the codon 12 of kRAS gene.

TABLE 1 | Genetic mutations of LA, PDTC, and ATC.

	LA	PDTC	ATC
EGFR*	14%	/	/
ALK*	/	4%	/
RET*	/	6%	/
MET	7%	/	/
nRAS*	/	21%	18%
hRAS	/	5%	6%
kRAS*	33%	2%	/
BRAF*	10%	33%	45%
PIK3CA*	7%	2%	18%
pTEN	/	4%	15%
EIF1AX	/	11%	9%
TERT*	/	40%	73%
NF1	11%	/	9%
TSH-R	/	2%	6%
STK11	17%	1%	6%
PAx8/PPAR gamma	/	4%	/
TP53	46%	8%	73%
ATM	/	7%	9%
RB1	4%	1%	9%
PI3K/AKT	/	11%	39%
SWI/SNF	/	6%	36%
HMTs	/	7%	24%
MMR	/	2%	12%
KEAP1	17%	/	/

*LA, Lung Adenocarcinoma; PDTC, Poorly Differentiated Thyroid Carcinoma; ATC, Anaplastic Thyroid Carcinoma; *Genes analyzed in our samples after DNA extraction with MALDI-TOF mass spectrometry [EGFR (Exon 18, mutation and deletion of codon 709 and 719; Exon 19, mutation and deletion of codon 744–759; Exon 20, mutation and insertion of codon 767–775 and 790; Exon 21, mutation of codon 833, 835, 848, 854, 858, and 861); kRAS (Mutation codon 12, 13, 61); nRAS (Mutation in codon 12 and 61); BRAF (Mutation in codon 466, 469, 594, 597, 600); PIK3CA (Mutation in codon 542, 545, 1,043, 1,047); ALK (Mutation in codon 1,156, 1,196, 1,269); ERBB2 (Mutation in exon 20); DDR2 (Mutation in codon 239, 638, 768); MAP2K1 (Mutation in codon 56, 57, 67); RET (Mutation in codon 918); TERT (Mutation C228T)].*

that thyroid transcription factor-1 (TTF-1) can be expressed by both thyroid cancer and lung adenocarcinoma (12) while thyroglobulin (Tg) is specific of thyroid cancer but, in PDTC, and even more in ATC cases, its expression is almost completely lost. Similarly, some serum tumor marker, such as CYFRA 21-1, could be elevated in the majority of lung carcinoma (13) but, although to a lesser extent, also in some cases of PDTC/ATC (14). Based on these considerations, neither the highly elevated values of serum tumor markers (e.g., CYFRA 21-1, CEA, Ca 125, Ca 15.3, or Ca 19.9) nor the positivity at immunohistochemistry of some thyroid specific genes (e.g., TTF-1 and PAX-8) were able to distinguish the two histologies.

Both thyroid cancer and lung adenocarcinoma can have mutations of genes involved in the MAP kinase pathways and some of them are almost exclusive of either lung adenocarcinoma or PDTC/ATC (**Table 1**) (10, 11, 15). It is indeed known that K-RAS is frequently mutated in lung adenocarcinoma and very rare or completely absent in PDTC and ATC, respectively (10, 11, 16). Moreover, the most common type of K-RAS mutations in thyroid cancer are Q61K, Q61R, and G12V (10, 17). It is worth noting that in several recent papers, which analyzed more than one thousand molecular profiles of PDTC/ATC by next generation sequencing, no K-RAS G12C mutation was found (10, 16, 18, 19). At variance, K-RAS p.G12C mutation is one of the most common type of K-RAS mutation in LA (11).

Currently, the presence of a K-RAS mutation identifies a subgroup of lung adenocarcinoma with a very poor prognosis (20), as confirmed by the very short survival of our patient.

Otherwise, in a prospective way, the importance of distinguishing the two tumors could be crucial to perform the right treatment. Indeed, the association of dabrafenib and trametinib demonstrated a robust clinical activity for the treatment of ATC with BRAF V600E mutation and was approved in USA from FDA (21). Also in case of LA, molecular targeted therapy are available and, the results from the phase III SELECT-1 trial including farnesyl transferase inhibition and synthetic lethality partners such as STK 33 showed promising biological activity against LA, in particular in p.G12C K-RAS mutated patients (22).

Moreover, in our case, not only the positivity of K-RAS mutation in both specimens, strongly supported the hypothesis that the thyroid mass could be a metastasis of the lung adenocarcinoma, but also its position in the lung lobe. The localization in the upper lobe is peculiar of the LA (23, 24), while PDTC/ATC metastases are usually multiple and located at the basis of the lung.

CONCLUSIONS

This case represents a typical, although rare, case in which the diagnosis of the malignancy was based on the molecular signature more than the localization of the tumor mass and its immunohistochemical and biochemical features. The case is of interest since strongly demonstrates that nowadays the histology, especially in more complicated and not well-differentiated cases that can require specific therapies, needs to be associated with a molecular pathology analysis.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The study was approved by Ethical Committee (CEAVNO—Comitato Etico Area Vasta Nord-Ovest). A written informed consent was obtained from the patient for the publication of this case report and any potentially-identifying information/images.

AUTHOR CONTRIBUTIONS

AM, PV, and RE contributed conception and design of the study. AM, DC, WB, FL, and EM collected clinical data and followed the patient during the clinical course. AA, LD, and GM performed

the biopsies. LT, CU, and FB evaluated the histological specimens. ES performed the molecular analysis. RC contributed to the interpretation of the molecular data. AM and RE wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Ministero dell' Istruzione, dell' Università'e della Ricerca (MIUR, investigator grant 2015 project code PRIN 2015HPMLFY) for the continuing research support.

REFERENCES

1. Shimaoka K, Sokal JE, Pickren JW. Metastatic neoplasms in the thyroid gland. Pathological and clinical findings. *Cancer.* (1962) 15:557–65. doi: 10.1002/1097-0142(196205/06)15:3<557::AID-CNCR2820150315>3.0.CO;2-H
2. Papi G, Fadda G, Corsello SM, Corrado S, Rossi ED, Radighieri E, et al. Metastases to the thyroid gland: prevalence, clinicopathological aspects and prognosis: a 10-year experience. *Clin Endocrinol.* (2007) 66:565–71. doi: 10.1111/j.1365-2265.2007.02773.x
3. Brady LW, O'Neill EA, Farber SH. Unusual sites of metastases. *Semin Oncol.* (1977) 4:59–64.
4. Wood K, Vini L, Harmer C. Metastases to the thyroid gland: the Royal Marsden experience. *Eur J Surg Oncol.* (2004) 30:583–8. doi: 10.1016/j.ejso.2004.03.012
5. Nixon IJ, Coca-Pelaz A, Kaleva AI, Triantafyllou A, Angelos P, Owen RP, et al. Metastasis to the thyroid gland: a critical review. *Ann Surg Oncol.* (2017) 24:1533–9. doi: 10.1245/s10434-016-5683-4
6. Kim TY, Kim WB, Gong G, Hong SJ, Shong YK. Metastasis to the thyroid diagnosed by fine-needle aspiration biopsy. *Clin Endocrinol.* (2005) 62:236–41. doi: 10.1111/j.1365-2265.2005.02206.x
7. Lam KY, Lo CY. Metastatic tumors of the thyroid gland: a study of 79 cases in Chinese patients. *Arch Pathol Lab Med.* (1998) 122:37–41.
8. Hegerova L, Griebeler ML, Reynolds JP, Henry MR, Gharib H. Metastasis to the thyroid gland: report of a large series from the Mayo Clinic. *Am J Clin Oncol.* (2015) 38:338–42. doi: 10.1097/COC.0b013e31829d1d09
9. Zhu YZ, Li WP, Wang ZY, Yang HF, He QL, Zhu HG, et al. Primary pulmonary adenocarcinoma mimicking papillary thyroid carcinoma. *J Cardiothorac Surg.* (2013) 8:131. doi: 10.1186/1749-8090-8-131
10. Landa I, Ibrahimpasic T, Boucai L, Sinha R, Knauf JA, Shah RH, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Investig.* (2016) 126:1052–66. doi: 10.1172/JCI85271
11. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* (2014) 511:543–50. doi: 10.1038/nature13385
12. Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. *Am J Surg Pathol.* (2002) 26:767–73. doi: 10.1097/00000478-200206000-00010
13. Pujol JL, Grenier J, Daures JP, Daver A, Pujol H, Michel FB. Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. *Cancer Res.* (1993) 53:61–6.
14. Giovanella L, Treglia G, Verburg FA, Salvatori M, Ceriani L. Serum cytokeratin 19 fragments: a dedifferentiation marker in advanced thyroid cancer. *Eur J Endocrinol.* (2012) 167:793–7. doi: 10.1530/EJE-12-0660
15. Bellevicine C, Vigliar E, Malapelle U, Carelli E, Fiorelli A, Vicidomini G, et al. Lung adenocarcinoma and its thyroid metastasis characterized on fine-needle aspirates by cytomorphology, immunocytochemistry, and next-generation sequencing. *Diagn Cytopathol.* (2015) 43:585–9. doi: 10.1002/dc.23264
16. Romei C, Tacito A, Molinaro E, Piaggi P, Cappagli V, Pieruzzi L, et al. Clinical, pathological and genetic features of anaplastic and poorly differentiated thyroid cancer: a single institute experience. *Oncol Lett.* (2018) 15:9174–82. doi: 10.3892/ol.2018.8470
17. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell.* (2014) 159:676–90. doi: 10.1016/j.cell.2014.09.050.
18. Chen H, Luthra R, Routbort MJ, Patel KP, Cabanillas ME, Broaddus RR, et al. Molecular profile of advanced thyroid carcinomas by next-generation sequencing: characterizing tumors beyond diagnosis for targeted therapy. *Mol Cancer Therap.* (2018) 17:1575–84. doi: 10.1158/1535-7163.MCT-17-0871
19. Pozdeyev N, Gay LM, Sokol ES, Hartmaier R, Deaver KE, Davis S, et al. Genetic analysis of 779 advanced differentiated and anaplastic thyroid cancers. *Clin Cancer Res.* (2018) 24:3059–68. doi: 10.1158/1078-0432.CCR-18-0373
20. Wood K, Hensing T, Malik R, Salgia R. Prognostic and predictive value in KRAS in non-small-cell lung cancer: a review. *JAMA Oncol.* (2016) 2:805–12. doi: 10.1001/jamaoncol.2016.0405
21. Subbiah V, Kreitman RJ, Wainberg ZA, Cho JY, Schellens JHM, Soria JC, et al. Dabrafenib and trametinib treatment in patients with locally advanced or metastatic BRAF V600-mutant anaplastic thyroid cancer. *J Clin Oncol.* (2018) 36:7–13. doi: 10.1200/JCO.2017.73.6785
22. Lindsay CR, Jamal-Hanjani M, Forster M, Blackhall F. KRAS: reasons for optimism in lung cancer. *Eur J Cancer.* (2018) 99:20–7. doi: 10.1016/j.ejca.2018.05.001
23. Del Ciello A, Franchi P, Contegiacomo A, Cicchetti G, Bonomo L, Larici AR. Missed lung cancer: when, where, and why? *Diagn Interv Radiol.* (2017) 23:118–26. doi: 10.5152/dir.2016.16187
24. Byers TE, Vena JE, Rzepka TF. Predilection of lung cancer for the upper lobes: an epidemiologic inquiry. *J Natl Cancer Inst.* (1984) 72:1271–5.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 February 2020
doi: 10.3389/fgene.2020.00009

Hsa_Circ_0007843 Acts as a mIR-518c-5p Sponge to Regulate the Migration and Invasion of Colon Cancer SW480 Cells

Jin Hua He[1], Ze Ping Han[1†], Jin Gen Luo[2†], Jian Wei Jiang[3†], Jia Bin Zhou[1], Wei Ming Chen[1], Yu Bing Lv[1], Meng Ling He[1], Lei Zheng[4], Yu Guang Li[1] and Ji Dong Zuo[5*]*

[1] Department of Laboratory Medicine, Central Hospital of Panyu District, Guangzhou, China, [2] Digesting Center, Central Hospital of Panyu District, Guangzhou, China, [3] Department of Biochemistry, Medical College, Jinan University, Guangzhou, China, [4] Department of Laboratory Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, China, [5] Department of Gastrointestinal Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Xiaofeng Mu,
Qingdao Central Hospital, China
Hongxing Liu,
Lu Daopei Hospital, China

***Correspondence:**
Ji Dong Zuo
zuojd@126.com
Yu Guang Li
lyg_py@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Genetics

Received: *04 June 2019*
Accepted: *06 January 2020*
Published: *25 February 2020*

Citation:
He JH, Han ZP, Luo JG, Jiang JW, Zhou JB, Chen WM, Lv YB, He ML, Zheng L, Li YG and Zuo JD (2020) Hsa_Circ_0007843 Acts as a mIR-518c-5p Sponge to Regulate the Migration and Invasion of Colon Cancer SW480 Cells. Front. Genet. 11:9. doi: 10.3389/fgene.2020.00009

Circular RNA (circRNA), a type of RNA that is widely expressed in mammalian cells, is considered to be essential in tumorigenesis. CircRNA can regulate target gene expression by interacting with the corresponding microRNA (miRNA). Our preliminary results showed that the expression levels of 1,817 circRNAs were significantly different in colon cancer tissue compared with paracancerous tissue, of which 1,236 were upregulated and 581 were downregulated. By using RT-PCR, we confirmed that the expression of hsa_circ_0007843, hsa_circ_0010575, hsa_circ_0007331, and hsa_circ_0001615 was significantly higher in colon cancer tissue than in normal colonic tissue; however, the expression levels of hsa_circ_0014879 and hsa_circRNA_401801 were not significantly different between normal and neoplastic colonic tissue. Among the circRNAs that were confirmed to be upregulated in colon cancer tissue, hsa_circ_0007843 was also found to be highly expressed in colon cancer SW480 cells. Overexpression of hsa_circ_0007843 promoted the invasion and migration of SW480 cells, whereas its downregulation suppressed their invasion and migration. Overexpression of hsa_circ_0007843 promoted tumor growth, whereas its downregulation inhibited tumor growth. We found that hsa_circ_0007843 interacted with miR-518c-5p and suppressed its expression, and miR-518c-5p interacted with matrix metallopeptidase 2 (MMP2) and promoted its expression and translation. Taken together, this study demonstrated that hsa_circ_0007843 acted as an miRNA sponge to regulate MMP2 expression by removing the inhibitory effect of miR-518c-5p on MMP2 gene translation, which further affected the invasive capability of SW480 cells.

Keywords: hsa_circ_0007843, matrix metallopeptidase 2, miR-518c-5p, colon cancer, invasion, migration

Abbreviations: ceRNA, competing endogenous RNA; circRNA, circular RNA; miRNA, microRNA; MMP2, metalloproteinase 2; NC, negative control; NCI, negative control inhibitor; RT-PCR, reverse transcription polymerase chain reaction; UTR, untranslated region.

INTRODUCTION

Colon cancer, a common gastrointestinal tumor, remains a serious threat to human health with the 3rd highest morbidity and mortality rate worldwide (Siegel et al., 2017). The development of colon cancer is a multistep process in which the abnormal expression of genes plays an important role (Qi and Ding, 2018). Although colon cancer has been studied widely, its pathogenesis is not fully understood. Hence, in-depth studies are required for the further elucidation of the molecular mechanism involved in the carcinogenesis of human colon cancer, which will provide new insights for the diagnosis, prognosis, and treatment of colon cancer.

Circular RNA (circRNA), a newly discovered non-coding RNA, has become a research hotspot in relevant fields. Unlike a linear RNA molecule that contains a 5′ cap and a 3′ polyadenylation tail, circRNA forms a covalently closed continuous loop and therefore does not have 5′ and 3′ ends (Memczak et al., 2013; Wang et al., 2018). CircRNAs are largely expressed in the eukaryotic transcriptome and participate in the regulation of gene expression (Guttman and Rinn, 2012; Holdt et al., 2018) and can therefore be considered essential in tumorigenesis (Arnaiz et al., 2018).

CircRNAs can regulate target gene expression by interacting with corresponding microRNAs (miRNAs) (Zhang et al., 2015). CircRNAs can function as miRNA "sponges" and regulate gene transcription (Hansen et al., 2013; Sang et al., 2018). For example, circRNA-ITCH increases ITCH gene expression by acting as an miRNA sponge by interacting with miR-7, miR-17, and miR-214. Overexpression of ITCH promotes the ubiquitination and degradation of phosphorylated Dv12 protein, thereby inhibiting the Wnt signaling pathway and suppressing the progression of cancer (Li et al., 2015). circRNA-000284 affects proliferation and invasion of cervical cancer cell *via* sponging miR-506 (Ma et al., 2018).

In the present study, we retrospectively collected colon cancer tissue samples from newly diagnosed patients. By using microarray analysis, we examined the circRNA expression profile of colon cancer tissue. CircRNAs whose expression levels were confirmed to be altered in colon cancer tissue were subjected to further *in vitro* analysis. By studying the biological function of the identified circRNAs and the molecular mechanism involved, we hope to provide invaluable information for the diagnosis, prognosis, and treatment of colon cancer.

MATERIALS AND METHODS

Sample Collection

Clinical tissue specimens from patients newly diagnosed with colon cancer were obtained from the archives of the Central Hospital of Panyu District with informed consent from the patients and with the approval of the institutional Ethics Committee (Guangdong, China). In this study, we collected tumor tissue and paired normal adjacent tissue from 30 patients with colon cancer. All of the samples were collected at the General Surgery Department of Panyu Central Hospital of Guangzhou from 2013 to 2018. The patients ranged in age from 45 to 82 years, with an average age of 65.66 ± 12.14 years (20 men, average age 61.8 ± 10.39 years; 10 women, average age 65.6 ± 13.69 years). None of the patients received radiotherapy or chemotherapy before surgical resection of the pathological tissue. All of the tissue samples were immediately snap-frozen in liquid nitrogen after surgical excision and stored at −80°C until total RNA was extracted for further experimentation. Basic information of the 30 selected subjects selected in this study is shown in **Table 1**.

Cell Culture

293T, FIHC, HCT-116, HT-29, SW480, LOVO, and SW620 cell lines were purchased from Yu Jia Bio Technology Co. (Guangzhou, China). The cell lines were cultured in Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham (DMEM/F12) containing 10% fetal bovine serum and were maintained in a humidified incubator at 37°C with 5% CO_2. The medium was changed every other day, and the cells were grown to 80–90% confluence and digested with 0.25% trypsin and sub-cultured. Cells in logarithmic growth with 95% viability were subjected to further experimentation.

CircRNA Microarray Analysis

Total RNA was extracted from the colon cancer tissue samples and paracancerous tissues (separated by a margin of 5 cm) with the TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instruction.

The tissue samples were sent to KangChen Bio-tech Inc. (Shanghai, China) for circRNA microarray analysis using a human 8×15K circRNA array (Arraystar Inc., Rockville, MD), which contains 9114 circRNA probes. Each circRNA was identified by using a specific probe that targets the specific splice junction of circRNA. Sample labeling and array hybridization were performed according to the manufacturer's protocol (Arraystar, Inc.). The R software package (R version 3.1.2) was used to normalize the raw data and subsequent data processing. Two groups of profile differences (tumor samples vs. paracancerous tissue samples) and the absolute fold change for each circRNA were computed.

Quantitative Real-Time RT-PCR

M-MLV reverse transcriptase was used for the reverse transcription of mRNA to cDNA, which was later used as the

TABLE 1 | The bascial data analysis of the 30 patients.

Indicator	
Male (n)	20
Female (n)	10
Mean age of male (years)	61.8 ± 10.39
Mean age of female (years)	65.6 ± 13.69
TNM stage (I/II/III/IV)	6/1/2/18/3
Tumor volume (<5 cm) (n)	12
Tumor volume (>5 cm) (n)	18

PCR template. Each experiment was performed in triplicate. The expression levels of U6 and β-actin were used as an internal control for mRNAs. The primers used in quantitative real-time PCR analysis are shown in **Table 2**.

Cell Transfection and Grouping

The design and synthesis of small interfering RNAs (siRNAs) for hsa_circ_0007843 and matrix metallopeptidase 2 (MMP2), the construction of a lentiviral vector overexpressing miR-518c-5p, and the construction of a lentiviral vector expressing si-hsa_circ_0007843 were conducted by Sangon Biotech Co., Ltd. (Shanghai, China). A lentiviral vector overexpressing hsa_circ_0007843 was produced by Jisai Biotech Co., Ltd. (Guangzhou, China). An inhibitor for miR-518c-5p was purchased from RiboBio Co., Ltd. (Guangzhou, China). Transfection was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions (**Table 3**). The transfected cells were divided into six groups: control group (no infection), hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843), si_hsa_circ_0007843 group (infected with lentiviral vector expressing si_hsa_circ_0007843), miR-518c-5p group (infected with lentiviral vector overexpressing miR-518c-5p), si-MMP2 group (transfected with si- MMP2), and NC group (infected with lentiviral vector expressing negative control sequence).

Colony Formation Assay

The cells were divided into four groups: control group (no infection), hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843), si_hsa_circ_0007843 group (infected with lentiviral vector expressing si_hsa_circ_0007843), and NC group (infected with lentiviral vector expressing negative control sequence). SW480 cells in logarithmic growth were plated at a density of 100 cells/well in a 6-well plate and incubated at 37°C with 5% CO_2 for 1 week. The supernatant in each well was discarded and the cells were washed twice with 1× phosphate-buffered saline (PBS). The cells were fixed for 15 min in 4% paraformaldehyde and stained for 15 min with crystal violet in methanol. Upon discarding the staining solution, the plates were allowed to air-dry, and the colonies were observed under a microscope.

Scratch-Wound Assay

The cells were divided into four groups: control group (no infection), hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843), si_hsa_circ_0007843 group (infected with lentiviral vector expressing si_hsa_circ_0007843), and NC group (infected with lentiviral vector expressing negative control sequence). Cells in logarithmic growth were plated and allowed to grow until 100% confluence was reached. The cell monolayer was scraped in a straight line with a pipette tip to create a gap. Debris was removed by washing the cells 3 times with 1× PBS. Culture medium was added and the cells were allowed to grow for 24 and 48 h. Microscopic images were taken at 0, 24, and 48 h. For each image, the distance between the two edges of the scratch was measured using ImageJ software.

Transwell Assay

The cells were divided into four groups: control group (no infection), hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843), si_hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843 siRNA), and NC group (infected with lentiviral vector expressing negative control sequence). Matrigel (Beijing Xia Si Biotechnology Co., Ltd., Beijing, China) was diluted in pre-chilled serum-free medium at a volume ratio of 1:3, and 40 μL pre-chilled serum-free medium was added into the pre-chilled transwell chamber. The Matrigel was solidified by incubation at 37°C for 2 h. The excess liquid was removed from the chamber and 100 and 600 μL serum-free medium was added to the upper and lower chamber, respectively. The plate containing the transwell chambers was incubated overnight at 37°C. On the day after cell transfection, 1.0×10^5 cells were re-suspended in 100 μL serum-free DMEM-F12. The cells were added to the upper transwell chamber, and 600 μL complete medium was added to the lower chamber at the same time. After incubation at 37°C with 5% CO_2 for 24 and 48 h, the surface cells and Matrigel in the upper chamber were removed using a cotton swab and the cells were observed under an inverted microscope. The cells at the lower surface of the upper chamber were stained with crystal violet and washed with 33% acetic acid. The absorbance of the solution was measured at 570 nm.

TABLE 2 | The primer for gene.

Gene	The sequence for primer	Length (bp)
β-actin(H)	F:5′-GTGGCCGAGGACTTTGATTG-3′; R:5′;-CCTGTAACAACGCATCTCATATT-3′;	73
hsa_circ_0007843	F:5′;-TCCGAAGATGGCTGAATGTG-3′; R:5′;-TCCCAATCAGGCCGCTCT-3′;	151
hsa_circ_0010575	F:5′;-CTGCCATCCAGGTGTGAA-3′; R:5′;-AGTCGTGGACGAGGAAGC-3′;	137
hsa_circ_401801	F:5′;-CAGTTTGCTGTTCATGGAGAC-3′; R:5′;-GGTGGGGACTGGTGCTAT-3′;	121
hsa_circ_0001615	F:5′;-TGATCGAACTGGCAGACG-3′; R:5′;-CTCCAGGAACACTTTGAGGA-3′;	128
hsa_circ_0007331	F:5′;-GAATGGGATTCGAGACCTG-3′; R:5′;-TTCTTCCAAAGCTGCCTGT-3′;	122
hsa_circ_0014879	F:5′;-TCTCCCTGTACGTTCTTATCTGC-3′; R:5′;-CTGCTCCCTTTGCTGGACATC-3′;	197
β-actin(H)	F:5′;-GTGGCCGAGGACTTTGATTG-3′; R:5′;-CCTGTAACAACGCATCTCATATT-3′;	73
miR-518c-5p	F:5′;-ATGGTTCGTGGGTCTCTGGAGGGAAGCACTTTC-3′; R:5′;-GTGCAGGGTCCGAGGT-3′;	89
MMP2	F:5′;-GATGCCGCCTTTAACTGG-3′; R:5′;-TCAGCAGCCTAGCCAGTCG-3′;	278
RT(miRNAs)	5′;-CTCAACTGGTGTCGTGGAGTCGGCAATTCAGTTGAGTTCCCAT-3′;	

TABLE 3 | The siRNA sequence.

Gene	siRNA (5′- 3′)
hsa_circ_0007843	sense: GGCAGCATACAGGAAGATGAA antisense: UUCAUCUUCCUGUAUGCUGCC
Negative control sequence	sense: UUCUCCGAACGUGUCACGUUUC antisense: GAAACGUGACACGUUCGGAGAA
MMP2	Sense: ACUUUUCUCCUCUUUUUUCCU antisense: GAAAAAAGAGGAGAAAAGUGG

In Vivo Treatment

The cells were divided into four groups: control group (no infection), hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843), si_hsa_circ_0007843 group (infected with lentiviral vector expressing si-hsa_circ_0007843), and NC group (infected with lentiviral vector expressing negative control sequence). A total of 20 athymic BALB/c nude mice (weight, 18–20 g) were purchased from Guangdong Medical Laboratory Animal Center (Foshan, China; animal production license, NO: 440035458020). Five cell lines (SW480, SW480-NC, hsa_circ_0007843-SW480, and si-circ_0007843-SW480) were digested with 0.25% trypsin, washed with PBS, counted by trypan blue staining, and adjusted to a concentration of 1.0×10^6 cells/mL, and 0.1-mL aliquots were used each time. After mixing with Matrigel (Beijing Xia Si Biotechnology Co., Ltd.), the cells were injected subcutaneously between the abdominal ribs of specific pathogen free-grade male nude mice aged up to 4 weeks. The tumor weight of the mice was observed.

Luciferase Reporter Assay

The cells were divided into the following groups: control group (psiCHECK-2-cirRNA/psiCHECK-2-mRNA, psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut), inhibitor group (inhibitor for miRNA + psiCHECK-2-cirRNA/psiCHECK-2-mRNA, psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut), NC group (negative control sequence + psiCHECK-2-cirRNA/psiCHECK-2-mRNA or psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut), NCI group (inhibitor for negative control + psiCHECK-2-cirRNA/psiCHECK-2-mRNA or psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut sequence), and hsa_miR-518c-5p group (hsa_miR-518c-5p mimics + psiCHECK-2-cirRNA/psiCHECK-2-mRNA or psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut). Genomic DNA was extracted from SW480 cells and used as the template for PCR, and *Xho*I and *Not*I restriction sites were introduced. The PCR amplification product was double-digested with the respective enzymes and cloned into the psiCHECK-2 vector to generate psiCHECK-2-cirRNA/psiCHECK-2-miRNA. A point mutation was introduced into the ligation product to generate the psiCHECK-2-cirRNA-mut/psiCHECK-2-miRNA-mut vector. 293T cells were transfected with miRNA mimics, negative control vector, miRNA inhibitor, psiCHECK-2-cirRNA/psiCHECK-2-miRNA vector, or psiCHECK-2-cirRNA/psiCHECK-2-miRNA-Mut vector (He et al., 2019).

A dual-luciferase assay (Promega, Madison, WI) was then performed according to the manufacturer's instruction.

Western Blot Analysis

The cells were divided into four groups: control group (no infection), si_hsa_circ_0007843 group (infected with lentiviral vector expressing si-hsa_circ_0007843), NC group (infected with lentiviral vector expressing negative control sequence), and miR-518c-5p group (infected with lentiviral vector overexpressing miR-518c-5p). When the cells reached 80% confluence, the supernatant was discarded and the cells were washed with pre-chilled 1× PBS. Then, 320 μL cell lysis buffer (RIPA with 3.2 μL PMSF) was added to the cells in order to extract cellular protein. After incubation for 30 min on ice, the cells were scraped into a 1.5 mL centrifuge tube and centrifuged for 15 min at 12,000 rpm, 4°C. Protein quantification was performed using a spectrophotometer (NanoDrop ND-1000; Thermo Fisher Scientific, Waltham, MA). Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and electrotransferred to polyvinylidene fluoride membranes. The membranes were incubated with a primary antibody against MMP2 at 4°C overnight, washed extensively with 0.1% Tween-20 in PBS, and incubated with a secondary antibody conjugated to horseradish peroxidase (1:1,000) at room temperature for 3 h. Immunolabeling was visualized using the ECL system.

Statistical Analysis

Data are expressed as the mean ± standard deviation and analyzed using SPSS20.0 statistical software. Statistical comparisons were performed using one-way analysis of variance.

RESULTS

Hsa_circ_0007843 Is Highly Expressed in Colon Cancer Tissue and SW480 Cells

Total RNA was isolated from 4 colon cancer tissue samples and 4 paracancerous tissue samples that were separated by a margin of 5 cm. By using Arraystar circRNA microarray analysis, we identified 1817 circRNAs with clearly different expression profiles in colon cancer tissue, of which 1,236 were upregulated and 581 were downregulated (**Figures 1A–C**). Six circRNAs with the most significant change in expression were selected for subsequent study. Their corresponding chromosomal locations were analyzed using bioinformatics software. As shown in **Table 4**, most of the circRNAs corresponded to protein-coding exons. The potential miRNA targets of these circRNAs were also predicted (**Figure 1D**).

By using RT-PCR, we further verified the expression profiles of the 6 circRNAs in 26 paired colon cancer and paracancerous tissue samples The results indicated that the expression levels of hsa_circ_0007843, hsa_circ_0010575, hsa_circ_0007331, and hsa_circ_0001615 were higher in colon cancer tissue compared with normal colonic tissue ($P < 0.05$); however, the expression levels of hsa_circ_0014879 and hsa_circRNA_401801 were not different between normal and neoplastic colonic tissues

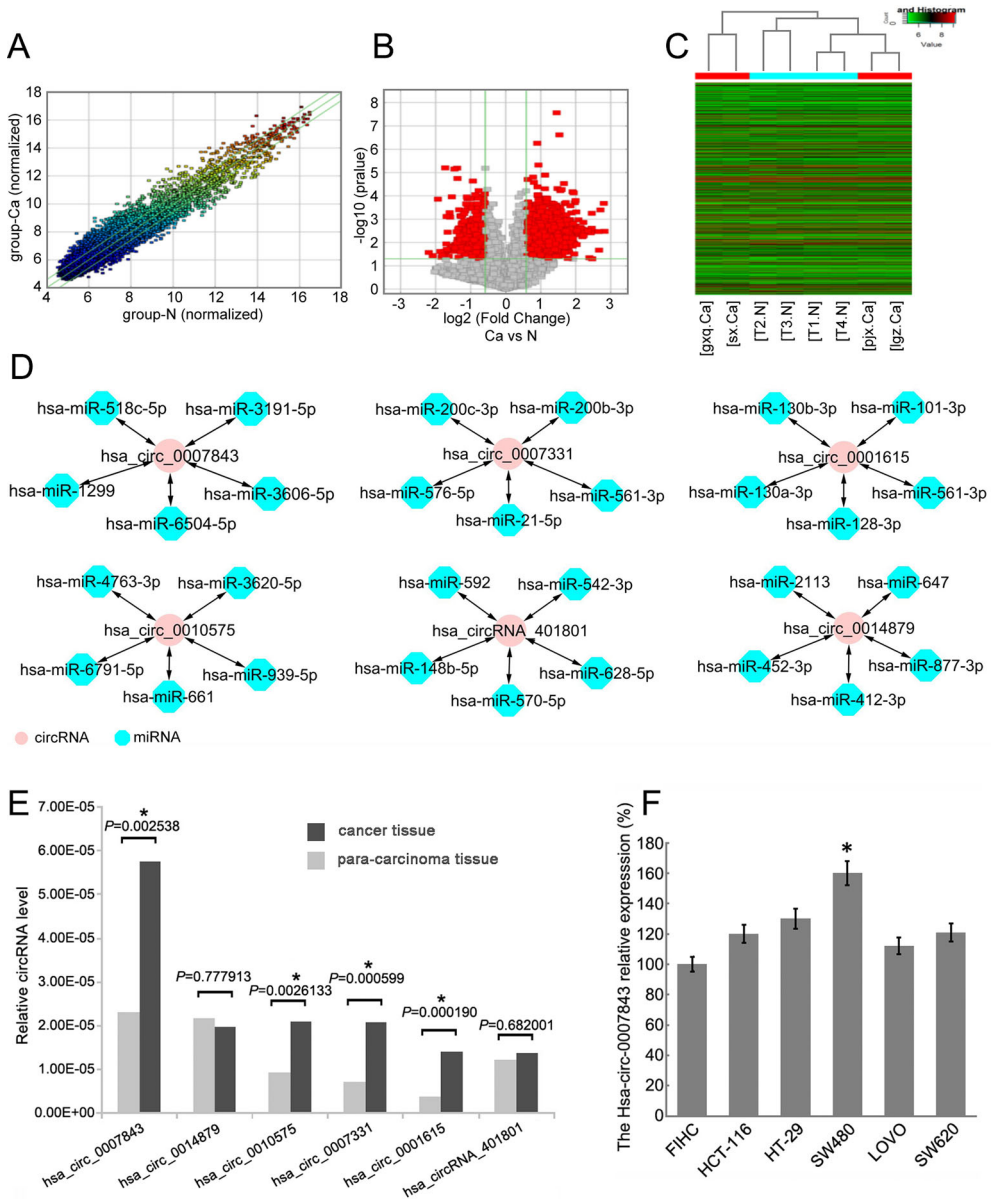


FIGURE 1 | Hsa_circ_0007843 expression is upregulated in colon cancer tissue and cell lines. **(A)** Scatter-plot showing circRNA expression variation between colon cancer tissue samples and paracancerous tissue samples. The values of the X and Y axes represent the normalized signal values. The circRNAs above the top green line and below the bottom green line indicates a greater than 2.0-fold change of the circRNAs between the two compared samples. **(B)** Volcano plots showing the differential expression of circRNAs. **(C)** Hierarchical clustering heat map showing the different circRNA expression profiles of the 8 samples. **(D)** miRNA targets of the 6 circRNAs predicted by bioinformatics software. **(E)** Relative expression of the 6 circRNAs in colon cancer and paracancerous tissues. *$P < 0.05$. **(F)** Relative expression of hsa_circ_0007843 in colon cancer cell lines by RT-PCR. Each bar represents the mean of 3 independent experiments. *$P < 0.05$.

($P > 0.05$). Among the circRNAs that were confirmed to be upregulated in colon cancer tissue, hsa_circ_0007843 expression was significantly increased (**Figure 1E**). *In vitro* analysis showed that hsa_circ_0007843 was highly expressed in colon cancer SW480 cells compared to FIHC, HCT-116, HT-29, LOVO, and SW620 cells (**Figure 1F**). Hence, hsa_circ_0007843 has the potential to be utilized as a biomarker for the early diagnosis of colon cancer.

Hsa_circ_0007843 Affects Colon Cancer SW480 Cell Colony Formation, Invasion, and Migration

In order to clarify the effect of hsa_circ_0007843 on SW480 cells, we infected the cells with hsa_circ_0007843 siRNA and a lentiviral vector overexpressing hsa_circ_0007843. At 1 week after infection, we observed that overexpression of hsa_circ_0007843 significantly promoted SW480 cell colony

TABLE 4 | Differential expression of circRNA in colon cancer and paracancerous tissues was screened by CircRNA microarray.

Regulation	circRNA	Chrom	Strand	txStart	txEnd	circRNA_type	best_transcript	GeneSymbol	*P*-value
Up	hsa_circ_0007843	chr11	–	128993340	129034322	Exonic	NM_001142685	ARHGAP32	0.000399138
Up	hsa_circ_0007331	chr3	–	195101737	195112876	Exonic	NM_012287	ACAP2	0.003367649
Up	hsa_circ_0001615	chr6	–	79752559	79770535	Exonic	NM_017934	PHIP	0.000555769
Up	hsa_circ_0010575	chr1	–	22157474	22207995	Exonic	NM_005529	HSPG2	0.002443045
Up	hsa_circ_0014879	chr1	–	160206924	160231148	Exonic	NM_015726	DCAF8	0.003808792
Up	hsa_circRNA_401801	chr10	–	14643337	14643566	Intronic	ENST00000181796	FAM107B	0.04496253

formation, whereas suppressing hsa_circ_0007843 expression inhibited SW480 cell colony formation (**Figures 2A, B**). Scratch-wound and transwell assays were performed to examine the effect of hsa_circ_0007843 on SW480 cell migration and invasion. Our results showed that the migration ability of SW480 cells was significantly enhanced in the hsa_circ_0007843 group compared with the si_hsa_circ_0007843, control, and NC groups (**Figures 2C, D**). In addition, the transwell assay demonstrated that the number of invading cells was obviously increased in the hsa_circ_0007843 group compared with the si_hsa_circ_0007843 group (**Figures 2E, F**). Therefore, these results suggested that overexpressing hsa_circ_0007843 promoted colon cancer SW480 cell colony formation, invasion, and migration, while suppressing its expression inhibited SW480 cell colony formation, invasion, and migration.

Hsa_circ_0007843 Affects Tumor Growth

SW480 tumor xenografts were established in athymic nude mice to evaluate the effects of hsa_circ_0007843 on cancer cell growth *in vivo*. Compared with the untreated animals, the application of si-hsa_circ_0007843 and hsa_circ_0007843 significantly affected tumor mass, whereas no effect was observed in the negative control group (**Figures 2G, H**). No body weight loss or diarrhea was detected and all animals (treated and untreated) survived. These results showed that reducing hsa_circ_0007843 expression effectively inhibited colon cancer growth *in vivo*, while overexpression of hsa_circ_0007843 effectively promoted colon cancer growth *in vivo*.

Hsa_circ_0007843 Interacts with miR-518c-5p, and miR-518c-5p Interacts with MMP2

RT-PCR was conducted to determine the relative expression levels of miR-518c-5p and MMP2 in the colon cancer cell lines. The results indicated that MMP2 expression was upregulated in SW480 cells, whereas miR-518c-5p expression was downregulated in these cells (**Figure 3**). To further confirm the interaction between hsa_circ_0007843, miR-518c-5p, and MMP2, hsa_circ_0007843 wild-type (hsa_circ_0007843-WT), hsa_circ_0007843 mutant-type (hsa_circ_0007843-Mut), MMP2-3′ untranslated region (UTR) wild-type (MMP2-3′ UTR-WT), and MMP2-3′UTR mutant type (MMP2-3′UTR-Mut) expression vectors were constructed and co-transfected with miR-518c-5p mimics into 293T cells, after which luciferase activity was examined. Our results indicated that luciferase activity was significantly decreased in 293T cells co-transfected with hsa_circ_0007843-WT, MMP2-3′UTR-WT, and miR-518c-5p mimics, suggesting that hsa_circ_0007843 might interact with miR-518c-5p, and miR-518c-5p might interact with MMP2. Furthermore, luciferase activity was not affected in 293T cells co-transfected with hsa_circ_0007843-Mut, MMP2-3′UTR-Mut, and miR-518c-5p mimics, suggesting that no interaction existed (**Figure 4**).

Hsa_circ_0007843 Suppresses miR-518c-5p Expression, and miR-518c-5p Promotes MMP2 Expression

In order to elucidate the regulatory mechanism involving hsa_circ_0007843, miR-518c-5p, and MMP2, we downregulated hsa_circ_0007843 expression in SW480 cells and found that miR-518c-5p expression was upregulated, while MMP2 expression was downregulated (**Figures 5A, B, G**). At 48 h after infection of SW480 cells with a lentiviral vector overexpressing miR-518c-5p, the expression levels of hsa_circ_0007843 and MMP2 were decreased (**Figures 5C, D, H**). At 48 h after SW480 cells were infected with MMP2 siRNA, the expression level of hsa_circ_0007843 was decreased; however, miR-518c-5p expression was increased (**Figures 5E, F**). Because hsa_circ_0007843 could interact with miR-518c-5p, and because miR-518c-5p could interact with MMP2, we concluded that hsa_circ_0007843 interacted with miR-518c-5p to suppress its expression, while miR-518c-5p interacted with MMP2 to promote its expression.

DISCUSSION

In the present study, by using circRNA microarray and RT-PCR analyses, we identified four circRNAs (hsa_circ_0007843, hsa_circ_0010575, hsa_circ_0007331, and hsa_circ_0001615) that were significantly upregulated in colon cancer tissue. Among these four circRNAs, the expression of hsa_circ_0007843 was the most significantly increased and thereby selected for further study. Hsa_circ_0007843 is encoded by the ARHGAP32 (Rho GTPase activating protein 32arhgap32) gene, which is located at chr11:128993340-129034322 (http://www.circbase.org/) and has a role in the development of breast cancer (Ina et al., 2012). Our results showed that hsa_circ_0007843 overexpression could promote SW480 cell invasion and migration, whereas the downregulation of hsa_circ_0007843 expression could suppress SW480 cell invasion and migration. These results suggest that hsa_circ_0007843 plays a role in colon cancer development

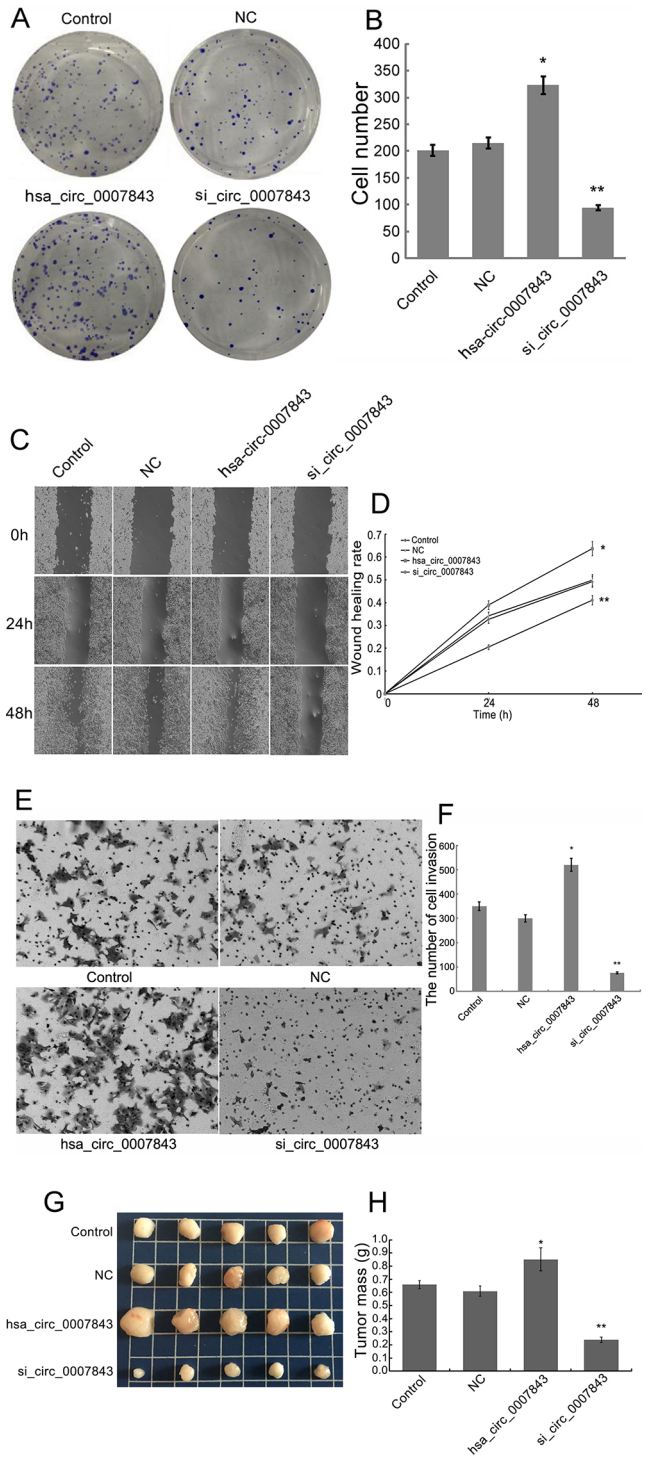


FIGURE 2 | Hsa_circ_0007843 exerts oncogenic effects on colon cancer SW480 cells. **(A)** Images from the colony formation assay with different groups of SW480 cells (×40). **(B)** Bar figures representing the number of colonies formed in each group; each bar represents the mean of 3 independent experiments. **(C)** Images from the scratch-wound assay with different groups of SW480 cells (×40). **(D)** Wound healing rates for the different groups of SW480 cells; each curve represents the mean of 3 independent experiments. **(E)** Images from the transwell assay with different groups of SW480 cells (×40). **(F)** Bar figures representing the number of invaded SW480 cells in each group; each bar represents the mean of 3 independent experiments. **(G)** Representative images of tumor-bearing xenograft mice. **(H)** Tumor weight of the nude mice in each group; each bar represents the mean of 3 independent experiments. $^{*}P < 0.05$, $^{**}P < 0.01$ compared with the control and NC groups. Control (no infection), hsa_circ_0007843 (infected with lentiviral vector expressing hsa_circ_0007843), si_hsa_circ_0007843 (infected with lentiviral vector expressing hsa_circ_0007843 siRNA), and NC (infected with lentiviral vector expressing negative control sequence).

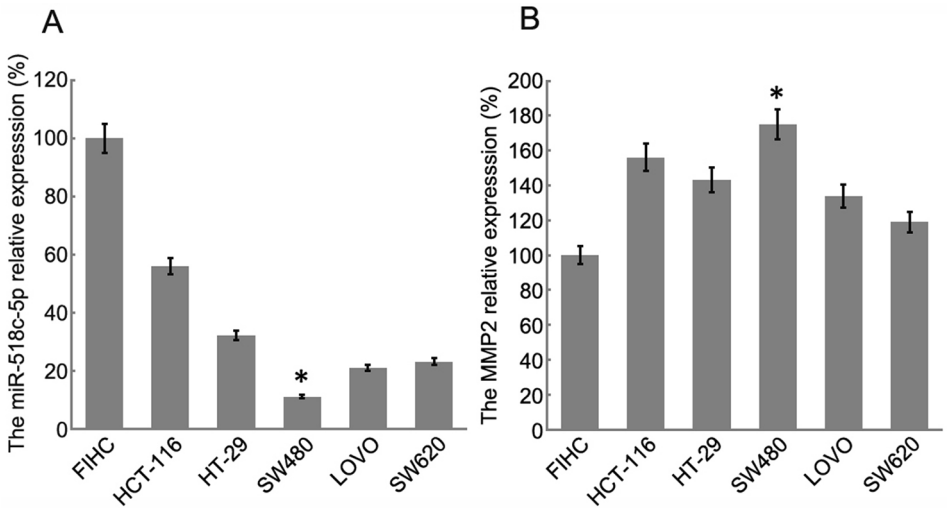


FIGURE 3 | Relative expression of miR-518c-5p and MMP2 in colon cancer cell lines by RT-PCR. **(A)** Relative expression of miR-518c-5p; each bar represents the mean of 3 independent experiments. *$P < 0.05$. **(B)** Relative expression of MMP2; each bar represents the mean of 3 independent experiments. *$P < 0.01$.

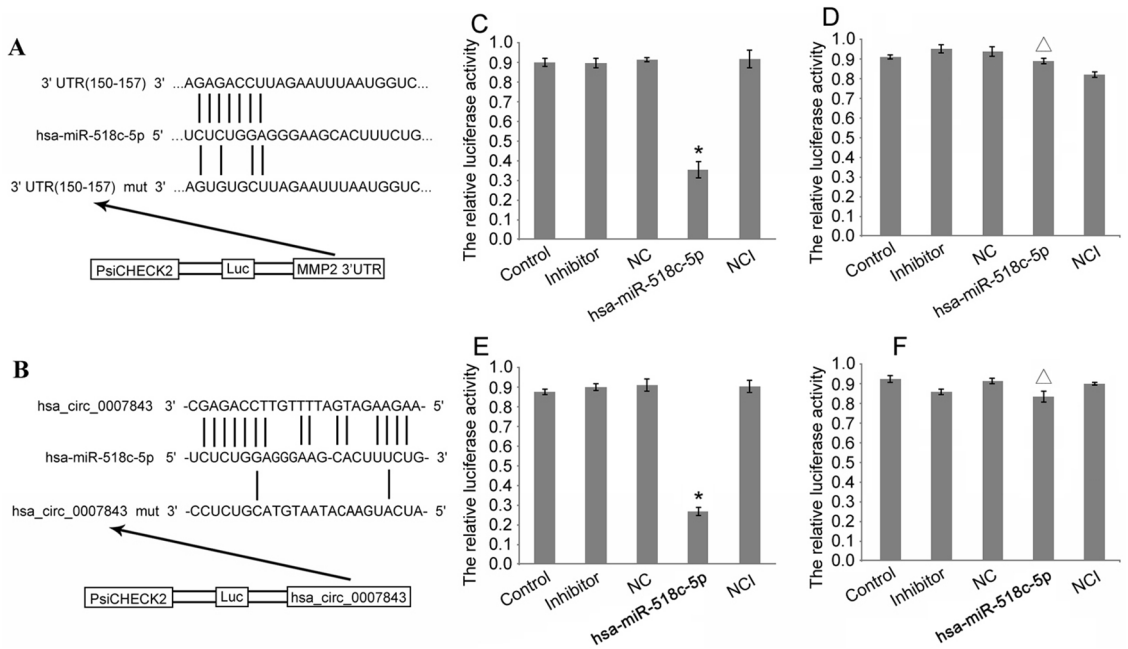


FIGURE 4 | Construction of double fluorescent reporter gene vector to verify the interactions between hsa_circ_0007843, miR-518c-5p, and MMP2. **(A)** Target sequences for miR-518-5p in the 3′-UTR region of wild-type MMP2. **(B)** Target sequences for miR-518-5p on hsa_circ_0007843. **(C)** Comparison of luciferase activity in cells of different groups that were co-transfected with MMP2-3′UTR-WT; each bar represents the mean of 3 independent experiments. *$P < 0.05$. **(D)** Luciferase activity in cells of different groups that were co-transfected with MMP2-3′UTR-Mut; each bar represents the mean of 3 independent experiments. $^{\Delta}P > 0.05$. **(E)** Comparison of luciferase activity in cells of different groups that were co-transfected with hsa_circ_0007843-WT; each bar represents the mean of 3 independent experiments. *$P < 0.05$. **(F)** Luciferase activity in cells of different groups that were co-transfected with hsa_circ_0007843-Mut; each bar represents the mean of 3 independent experiments. $^{\Delta}P > 0.05$. Control group (psiCHECK-2-cirRNA/psiCHECK-2-mRNA, psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut), inhibitor (inhibitor for miRNA + psiCHECK-2-cirRNA/psiCHECK-2-mRNA, psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut), NC group (negative control sequence + psiCHECK-2-cirRNA/psiCHECK-2-mRNA or psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut), NCI group (inhibitor for negative control + psiCHECK-2-cirRNA/psiCHECK-2-mRNA or psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut sequence), hsa_miR-518c-5p group (hsa_miR-518c-5p mimics + psiCHECK-2-cirRNA/psiCHECK-2-mRNA or psiCHECK-2-cirRNA-mut/psiCHECK-2-mRNA-mut).

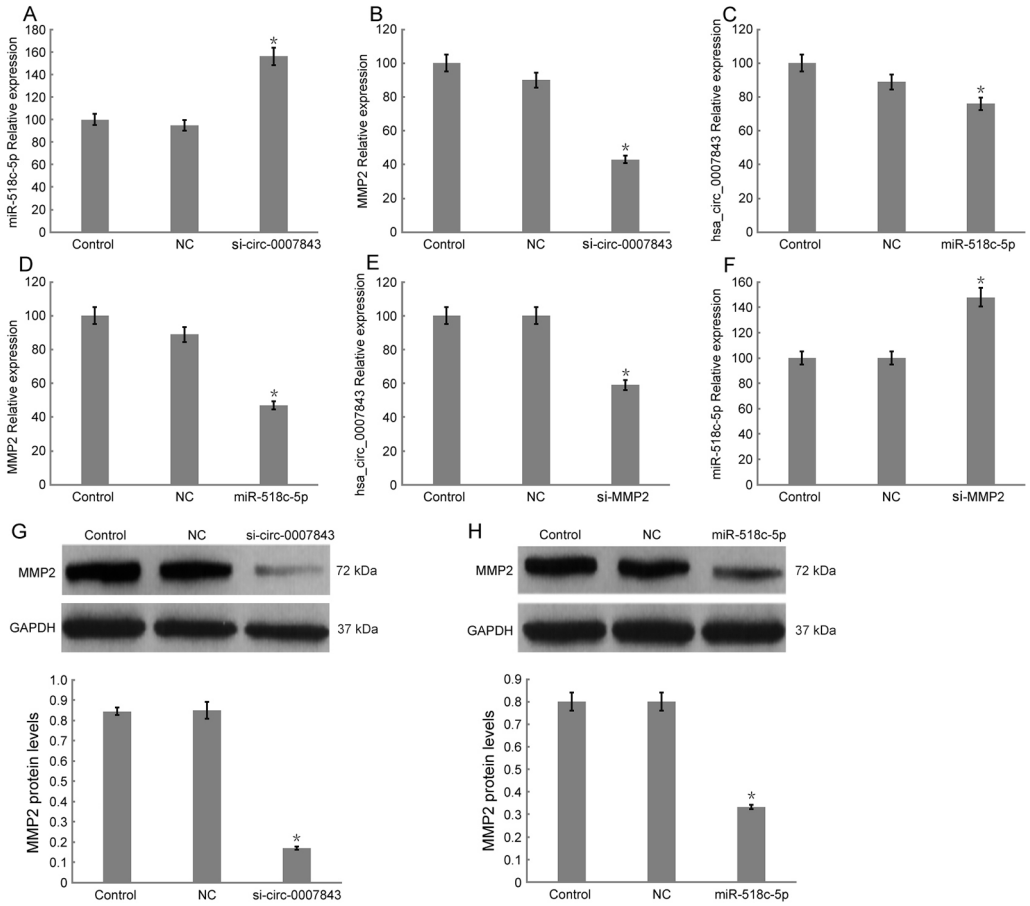


FIGURE 5 | Relative expression levels of hsa_circ_0007843, miR-518c-5p, and MMP2 in SW480 cells by RT-PCR and western blotting. **(A)** Relative expression of miR-518c-5p upon downregulation of hsa_circ_0007843 expression; each bar represents the mean of 3 independent experiments. *$P < 0.05$. **(B)** Relative expression of MMP2 upon downregulation of hsa_circ_0007843 expression; each bar represents the mean of 3 independent experiments. *$P < 0.01$. **(C)** Relative expression of hsa_circ_0007843 upon overexpression of miR-518c-5p; each bar represents the mean of 3 independent experiments. *$P < 0.05$. **(D)** Relative expression of MMP2 upon overexpression of miR-518c-5p; each bar represents the mean of 3 independent experiments. *$P < 0.01$. **(E)** Relative expression of hsa_circ_0007843 upon downregulation of MMP2 expression; each bar represents the mean of 3 independent experiments. *$P < 0.01$. **(F)** Relative expression of miR-518c-5p upon downregulation of MMP2 expression; each bar represents the mean of 3 independent experiments. *$P < 0.01$. **(G)** Corresponding expression of MMP2 protein upon downregulation of hsa_circ_0007843 expression. **(H)** Corresponding expression of MMP2 protein upon overexpression of miR-518c-5p. Control group (no infection), si_hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843 siRNA), hsa_circ_0007843 group (infected with lentiviral vector expressing hsa_circ_0007843), NC group (infected with lentiviral vector expressing negative control sequence), and miR-518c-5p group (infected with lentiviral vector overexpressing miR-518c-5p).

and could be considered as a potential biomarker for the early diagnosis of colon cancer. A previous study demonstrated that the expression of circRNAs is highly stable and shows tissue and developmental stage specificity (Li et al., 2015). Moreover, circRNAs are considered to be potential biomarkers for cancer diagnosis.

Li et al. reported that hsa_circ_002059 expression was significantly decreased in gastric cancer tissue, and its preoperative and post-operative levels in plasma were significantly different, which could be affected by distant metastasis, tumor-node-metastasis staging, sex, and age (Li et al., 2015). A total of 698 circRNAs were shown to be dysregulated in laryngeal carcinoma, with hsa_circ_100855 having the most significant increase in expression (Xuan et al., 2016). Hsa_circ_0001649 expression was found to be downregulated in hepatocellular carcinoma, and was considered to be a potential biomarker for the diagnosis and treatment of hepatocellular carcinoma (Qin et al., 2016). Wang et al. demonstrated that the expression of hsa_circ_001988 was decreased in colon cancer, which might be associated with the proliferation and eosinophilic invasion of colon cancer cells (Wang et al., 2015). Thus, hsa_circ_0007843 could play important roles in the pathogenesis and development of colon cancer.

An increasing number of studies have shown that some circRNAs have specific binding sites for miRNAs, and

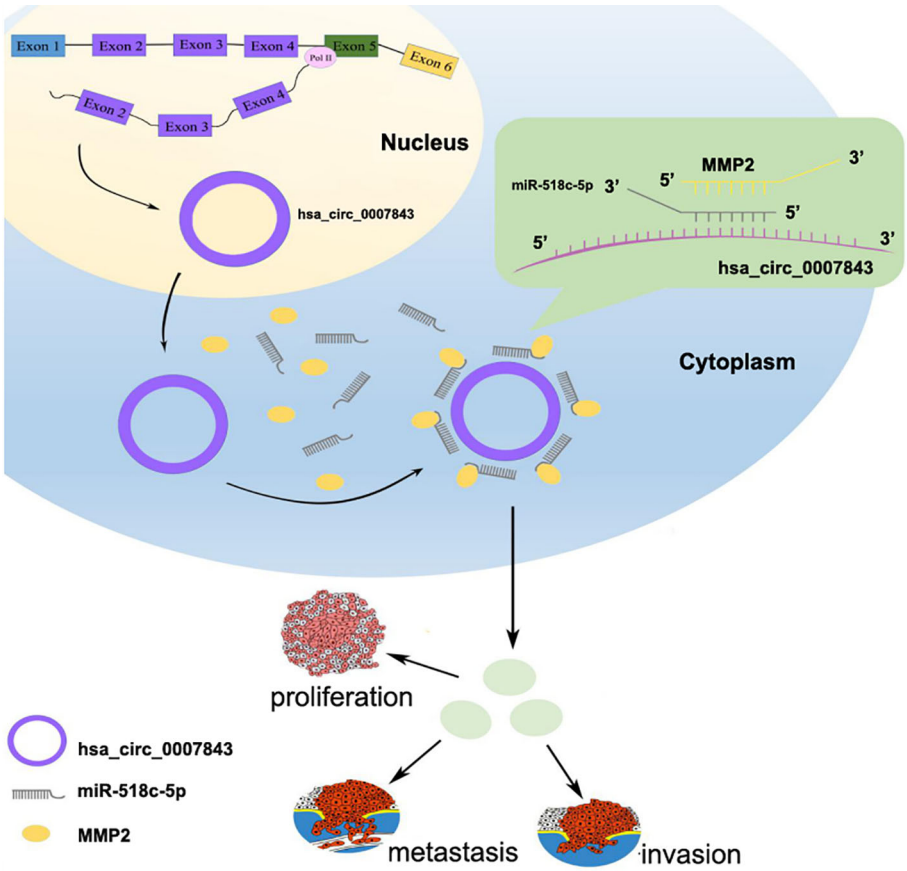


FIGURE 6 | Hsa_circ_0007843/miR-518c-5p/MMP2 regulation axis effects on invasion and migration of SW480 cells.

miRNAs can suppress the degradation of the corresponding mRNAs by competitively binding to endogenous circRNAs (Hansen et al., 2013). In the present study, we demonstrated that the expression levels of hsa_circ_0007843 and MMP2 were upregulated in SW480 cells, whereas miR-518c-5p expression was downregulated in these cells. These results suggested that hsa_circ_0007843 acted as an miRNA sponge to regulate MMP2 expression by removing the inhibitory effect of miR-518c-5p on MMP2 gene translation. Bioinformatics analysis and a luciferase reporter assay revealed that the 3′ UTRs of hsa_circ_0007843 and MMP2 share identical miR-518c-5p response elements and could therefore bind competitively to miR-518c-5p. Using RT-PCR and western blotting analyses, we demonstrated that hsa_circ_0007843 suppressed miR-518c-5p expression, while miR-518c-5p promoted MMP2 expression. MMP2 is considered to be fundamental for the degradation of the extracellular matrix. Its expression level was found to be clearly upregulated in colon cancer, which might contribute to colon cancer cell invasion and migration (Schwegmann et al., 2016; Hsu et al., 2017). Hsa_circ_000984 acts as a competing endogenous RNA (ceRNA) by competitively binding to miR-106b and effectively upregulating CDK6 expression, thereby inducing a series of malignant phenotypes of tumor cells (Xu et al., 2017). In addition, hsa_circ_0055625 increases the growth of colon cancer cells by sponging miR-106b-5p (Zhang et al., 2019). CircCCDC66 expression is elevated in polyps and colon cancer and is associated with poor prognosis; circCCDC66 controls multiple pathological processes, including cell proliferation, migration, invasion, and anchorage-independent growth (Hsiao et al., 2017).

In conclusion, our results suggest that hsa_circ_0007843, along with oncogenic characteristics, might be a potential biomarker for the diagnosis and treatment of colon cancer. Hsa_circ_0007843 overexpression promoted SW480 cell invasion and migration, whereas its downregulation suppressed SW480 cell invasion and migration. Hsa_circ_0007843 suppressed miR-518c-5p expression, while miR-518c-5p promoted MMP2 expression. Therefore, hsa_circ_0007843 acted as an miRNA sponge to regulate MMP2 expression by removing the inhibitory effect of miR-518c-5p on MMP2 gene translation, which further affected the invasion and migration of SW480 cells (**Figure 6**). Our findings suggested that the hsa_circ_0007843/miR-518c-5p/MMP2 regulation axis might play a critical role in the progression and development of colon cancer. The ceRNA network and pathway involved

might be novel clinical markers and therapeutic targets for colon cancer patients.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by archives of the Central Hospital of Panyu District. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by JiNan University.

AUTHOR CONTRIBUTIONS

JH, ZH, YBL, MH and JBZ performed the experiments. JJ, JGL and LZ analyzed the data. JH and ZH wrote the manuscript. YGL and JDZ designed the study and revised the manuscript. WC and MH provided the reagents. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the Medical and Health Science and Technology Project of Panyu District, Guangzhou (No. 2017-Z04-18;2018-Z04-59;2019-Z04-02), Science and Technology Planning Project of Guangdong Province (No. 2017ZC0372), Guangzhou Health and Family Planning Commission Program (No. 20181A011118; 20192A011027; No. 20191A011119), Project of Guangdong Administration of Traditional Chinese Medicine (No. 20192073), Natural Science Foundation of Guangdong Province (No. 2018A0303130191), and Guangzhou Science and Technology Plan Project (No. 201904010044).

ACKNOWLEDGMENTS

We are grateful to Lei Zheng for providing technical guidance.

REFERENCES

Arnaiz, E., Sole, C., Manterola, L., Iparraguirre, L., Otaegui, D., and Lawrie, C. H. (2018). CircRNAs and cancer: biomarkers and master regulators. *Semin. Cancer Biol.* 58, 90–99. doi: 10.1016/j.semcancer.2018.12.002

Guttman, M., and Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339–346. doi: 10.1038/nature10887

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microrRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993

He, J. H., Han, Z. P., Yu g, L., and Zheng, L. (2019). CDX2/mir-145-5p/SENP1 pathways affect LNCaP cells invasion and migration. *Front. Oncol.* 9, 477. doi: 10.3389/fonc.2019.00477

Holdt, L. M., Kohlmaier, A., and Teupser, D. (2018). Molecular roles and function of circular RNAs in eukaryotic cells. *Cell Mol. Life Sci.* 75, 1071–1098. doi: 10.1007/s00018-017-2688-5

Hsiao, K. Y., Lin, Y. C., Gupta, S. K., Chang, N., Yen, L., Sun, H. S., et al. (2017). Noncoding effects of circular RNA CCDC66 promote colon cancer growth and metastasis[J]. *Cancer Res.* 77, 2339. doi: 10.1158/0008-5472.CAN-16-1883

Hsu, H. H., Kuo, W. W., Day, C. H., Shibu, M. A., Li, S. Y., Chang, S. H., et al. (2017). Taiwanin E inhibits cell migration in human LoVo colon cancer cells by suppressing MMP-2/9 expression *via* p38 MAPK pathway. *Environ. Toxicol.* 32, 2021–2031. doi: 10.1002/tox.22379

Ina, Schulte, EM, Batty, JC, Pole, KA, Blood, Mo, S., et al. (2012). Structural analysis of the genome of breast cancer cell line ZR-75-30 identifies twelve expressed fusion genes. *BMC Genomics* 13, 719. doi: 10.1186/1471-2164-13-719

Li, F., Zhang, L., Li, W., Deng, J., Zheng, J., An, M., et al. (2015). Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/β-catenin pathway. *Oncotarget* 6, 6001–6013. doi: 10.18632/oncotarget.3469

Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., et al. (2015). Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* 25, 981–984. doi: 10.1038/cr.2015.82

Li, P., Chen, S., Chen, H., Mo, X., Li, T., Shao, Y., et al. (2015). Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clin. Chim. Acta* 444, 132–136. doi: 10.1016/j.cca.2015.02.018

Ma, H. B., Yao, Y. N., Yu, J. J., et al. (2018). Extensive profiling of circular RNAs and the potential regulatory role of circRNA-000284 in cell proliferation and invasion of cervical cancer *via* sponging miR-506. *Am. J. Transl. Res.* 10 (2), 592–604.

Memczak, S. 1., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928

Qi, L., and Ding, Y. (2018). Screening of differentiation-specific molecular biomarkers for colon cancer. *Cell Physiol. Biochem.* 46, 2543–2550. doi: 10.1159/000489660

Qin, M., Liu, G., Huo, X., Tao, X., Sun, X., Ge, Z., et al. (2016). Hsa_circ_0001649: A circular RNA and potential novel biomarker for hepatocellular carcinoma. *Cancer Biomarkers* 16, 161–169. doi: 10.3233/CBM-150552

Sang, M., Meng, L., Liu, S., Ding, P., Chang, S., Ju, Y., et al. (2018). Circular RNA ciRS-7 maintains metastatic phenotypes as a ceRNA of miR-1299 to target MMPs. *Mol. Cancer Res.* 16, 1665–1675. doi: 10.1158/1541-7786.MCR-18-0284

Schwegmann, K., Bettenworth, D., Hermann, S., Faust, A., Poremba, C., Foell, D., et al. (2016). Detection of early murine colorectal cancer by MMP-2/-9-guided fluorescence endoscopy. *Inflammatory Bowel Dis.* 22, 82–91. doi: 10.1097/MIB.0000000000000605

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics, 2017. *CA Cancer J. Clin.* 67, 7–30. doi: 10.3322/caac.21387

Wang, X. 1., Zhang, Y. 2., Huang, L. 1., Zhang, J., Pan, F., Li, B., et al. (2015). Decreased expression of hsa_circ_001988 in colorectal cancer and its clinical significances. *Int. J. Clin. Exp. Pathol.* 8, 16020–16025.

Wang, Y., Lu, T., Wang, Q., Liu, J., and Jiao, W. (2018). Circular RNAs: Crucial regulators in the human body (Review). *Oncol. Rep.* 40, 3119–3135. doi: 10.3892/or.2018.6733

Xu, X. W., Zheng, B. A., Hu, Z. M., Qian, Z. Y., and Huang, C. J. (2017). Circular RNA hsa_circ_000984 promotes colon cancer growth and metastasis by

sponging miR-106b[J]. *Oncotarget* 8, 91674–91683. doi: 10.18632/oncotarget.21748

Xuan, L., Qu, L., Zhou, H., Wang, P., Yu, H., Wu, T., et al. (2016). Circular RNA: a novel biomarker for progressive laryngeal cancer. *Am. J. Transl. Res.* 8, 932–939.

Zhang, C., Wu, H., Wang, Y., Zhao, Y., Fang, X., Chen, C., et al. (2015). Expression patterns of circularRNAs from primary kinase transcripts in the mammary glands of lactating rats. *J. Breast Cancer* 18, 235–241. doi: 10.4048/jbc.2015.18.3.235

Zhang, J., Liu, H., Zhao, P., Zhou, H., and Mao, T. (2018). Has_circ_0055625 from circRNA profile increases colon cancer cell growth by sponging miR-106b-5p [J]. *J. Cell. Biochem.* 120, 3027–3037. doi: 10.1002/jcb.27355

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 13 March 2020
doi: 10.3389/fonc.2020.00310

Comparison of Fresh Frozen Tissue With Formalin-Fixed Paraffin-Embedded Tissue for Mutation Analysis Using a Multi-Gene Panel in Patients With Colorectal Cancer

Xian Hua Gao[1†], Juan Li[2†], Hai Feng Gong[1†], Guan Yu Yu[1], Peng Liu[1], Li Qiang Hao[1], Lian Jie Liu[1], Chen Guang Bai[3*] and Wei Zhang[1*]*

[1] Department of Colorectal Surgery, Changhai Hospital, Shanghai, China, [2] Department of Nephrology, Changhai Hospital, Shanghai, China, [3] Department of Pathology, Changhai Hospital, Shanghai, China

Background: Next generation sequencing (NGS)-based multi-gene panel tests have been performed to predict the treatment response and prognosis in patients with colorectal cancer (CRC). Whether the multi-gene mutation results of formalin-fixed paraffin-embedded (FFPE) tissues are identical to those of fresh frozen tissues remains unknown.

Methods: A 22-gene panel with 103 hotspots was used to detect mutations in paired fresh frozen tissue and FFPE tissue from 118 patients with CRC.

Results: In our study, 117 patients (99.2%) had one or more variants, with 226 variants in FFPE tissue and 221 in fresh frozen tissue. Of the 129 variants identified in this study, 96 variants were present in both FFPE and fresh frozen tissues; 27 variants were found in FFPE tissues only; 6 variants were found only in fresh frozen tissues. The mutation results demonstrated >94.0% concordance in all variants, with Kappa coefficient >0.500 in 64.3% (83/129) of variants. At the gene level, concordance ranged from 73.8 to 100.0%, with Kappa coefficient >0.500 in 81.3% (13/16) of genes.

Conclusions: The results of mutation analysis performed with a multi-gene panel and FFPE and fresh frozen tissue were highly concordant in patients with CRC, at both the variant and gene levels. There were, however, some important differences in mutation results between the two tissue types. Therefore, fresh frozen tissue should not routinely be replaced with FFPE tissue for mutation analysis with a multi-gene panel. Rather, FFPE tissue is a reasonable alternative for fresh frozen tissue when the latter is unavailable.

Keywords: colorectal cancer, multi-gene panel, gene mutation, fresh frozen tissue, formalin-fixed paraffin-embedded tissue

OPEN ACCESS

Edited by:
Xinmin Li,
University of California, Los Angeles, United States

Reviewed by:
Pasquale Pisapia,
University of Naples Federico II, Italy
Manfred Dietel,
Charité Medical University of Berlin, Berlin, Germany

***Correspondence:**
Lian Jie Liu
lianjieliu@yeah.net
Chen Guang Bai
bcg709@126.com
Wei Zhang
weizhang2000cn@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *03 June 2019*
Accepted: *21 February 2020*
Published: *13 March 2020*

Citation:
Gao XH, Li J, Gong HF, Yu GY, Liu P, Hao LQ, Liu LJ, Bai CG and Zhang W (2020) Comparison of Fresh Frozen Tissue With Formalin-Fixed Paraffin-Embedded Tissue for Mutation Analysis Using a Multi-Gene Panel in Patients With Colorectal Cancer. Front. Oncol. 10:310. doi: 10.3389/fonc.2020.00310

INTRODUCTION

The survival of patients with colorectal cancer (CRC) has improved greatly in recent decades, because of advancements in surgical technique, chemoradiotherapy, and targeted therapy (1, 2). However, CRC remains the third most common cancer and the fourth cause of cancer-related death worldwide (3). Molecular biomarkers have been reported to play a vital role in the early diagnosis of CRC, and individualized treatment for metastatic CRC (4, 5). In the past decade, somatic gene mutations of *KRAS* (KRAS proto-oncogene, GTPase), *NRAS* (NRAS proto-oncogene, GTPase), and *BRAF* (B-Raf proto-oncogene, serine/threonine kinase) have been used to predict the outcomes of *EGFR* (epidermal growth factor receptor)—targeted therapy for metastatic CRC (4). Research studies of targeted therapy and individualized medicine have identified additional genes associated with the development and treatment of CRC (6). The classification of CRC based on multiple gene detection may help to explain inter-individual differences in treatment response and long-term outcomes. For example, the *UGT1A* (UDP glucuronosyltransferase family 1 member A complex locus) polymorphism has been reported to predict drug toxicity (delayed severe diarrhea) in patients with CRC who receive irinotecan-based chemotherapy (7, 8). Somatic mutations of *MLH1* (mutL homolog 1), *MSH2* (mutS homolog 2), *MSH6* (mutS homolog 6), and *PMS2* (PMS1 homolog 2, mismatch repair system component) in CRC tissue could be used to screen for Lynch syndrome (9, 10). Moreover, the detection of multiple gene mutations could offer more options for new targeted treatment efforts in drug-resistant patients (11). Hence, the need for combined detection of multiple gene mutations has acquired increasing urgency.

Next generation sequencing (NGS) is an efficient and rapid tool for the detection of single-nucleotide mutations and small-fragment insertion/deletions. This approach has become the standard technique for multiple gene detection (12). Compared with the single-gene mutation detection, multiple gene detection with NGS is a timely and cost-effective technique that requires a small amount of DNA (13, 14). In our previous study (15), we established a 22-gene panel, which included 103 amplicons targeting the variants found to be most common in CRC. Those 22 genes included *KRAS, BRAF, TP53* (tumor protein p53), *EGFR, CTNNB1* (catenin beta 1), *DDR2* (discoidin domain receptor tyrosine kinase 2), *ERBB2* (erb-b2 receptor tyrosine kinase 2), *ERBB4* (erb-b2 receptor tyrosine kinase 4), *FBXW7* (F-box and WD repeat domain containing 7), *FGFR1* (fibroblast growth factor receptor 1), *FGFR2* (fibroblast growth factor receptor 2), *FGFR3* (fibroblast growth factor receptor 3), *AKT1* (AKT serine/threonine kinase 1), *ALK* (ALK receptor tyrosine kinase), *MAP2K1* (mitogen-activated protein kinase kinase 1), *MET* (MET proto-oncogene, receptor tyrosine kinase), *NOTCH1* (notch receptor 1), *NRAS, PIK3CA* (phosphatidylinositol-4,5- bisphosphate 3-kinase catalytic subunit alpha), *PTEN* (phosphatase and tensin homolog), *SMAD4* (SMAD family member 4), *and STK11* (serine/threonine kinase 11). Use of this panel requires only 10 ng of DNA and a single-tube multiplex polymerase chain reaction (PCR).

Fresh frozen tissue is the preferred sample to detect gene mutation due to its superiority in preserving DNA. However, fresh frozen tissue is often not available in clinical practice, as the associated protocol requires that resected tissue be snap-frozen in liquid nitrogen 30–60 min after surgical resection. Moreover, the cost of preserving fresh frozen tissue is relatively high, as it requires the maintenance of a constant ultralow temperature. Compared with fresh frozen tissue, formalin-fixed paraffin-embedded (FFPE) tissue has several advantages: (1) preservation of the cellular and architectural morphology; (2) the possibility of storage at room temperature for several years; (3) easy availability, as FFPE blocks are routinely prepared in the pathology departments of most centers. Therefore, FFPE tissue has become the sample type used most commonly for molecular testing (11). However, the fixation and archiving process in FFPE often leads to the cross-linking, degradation, and fragmentation of DNA molecules. These alterations inevitably affect the quality and quantity of DNA extracted from FFPE tissue, which in turn would adversely impact the accuracy with which gene mutations are detected (16–19). Nevertheless, the detection of *EGFR* and *KRAS* mutations in DNA extracted from FFPE tissue has proven to be highly accurate (20). The detection of gene mutations using NGS may also be performed with DNA from FFPE samples (13, 14, 16, 21).

In our previous study, we showed the utility of this 22-gene panel when used with NGS to detect gene mutations in FFPE tissue from 207 patients with CRC (15). However, whether the gene mutation results of this multi-gene panel are concordant between FFPE and fresh frozen tissue remains unknown. In this study, gene mutations were detected in paired FFPE and fresh frozen primary tumor tissue from patients with CRC using this 22-gene panel. The results obtained were compared between tissue types at the variant and gene levels.

MATERIALS AND METHODS

Inclusion Criteria

Patients who satisfied all of the following criteria were included: (1) pathologically proven primary CRC adenocarcinoma; (2) history of radical surgery for a primary tumor at Changhai Hospital (Shanghai, China) during the period from March 2015 to November 2016; (3) availability of FFPE and fresh frozen primary tumor tissue samples.

Exclusion Criteria

Patients were excluded if they met any one of the following criteria: (1) history of preoperative radiotherapy; (2) insufficient FFPE or fresh frozen tumor tissue for the extraction of DNA; (3) history of local tumor excision; (4) recurrent CRC; (5) personal history of other tumors; (6) hereditary CRC.

Patients

All patients with CRC satisfying the above-mentioned criteria were enrolled in the study. All the relevant clinicopathological information was prospectively maintained in an electronic CRC database. The study was approved by the ethical committee at Changhai Hospital. All included patients provided their

written informed consent. All patients were followed up after surgery every 3 months for the first 2 years, every 6 months for the next 3 years, and every year from that point onward.

Processing of Tumor Tissue

All tumor tissue samples were obtained from surgically resected primary CRC tumor specimens. The resected tissue was divided into two parts. One part was frozen in liquid nitrogen <30 min after surgery and stored at −80°C until the time of DNA extraction; the other part was fixed in 4% formalin for 24–72 h and embedded in paraffin, and then stored at room temperature. The blocks containing FFPE tissues were divided to obtain ten consecutive sections with thickness of 10 μm. Hematoxylin and eosin staining was performed on the first section. Two pathologists examined each stained section individually and estimated the percentage of neoplastic cells. Sections with ≥ 40% neoplastic cells were considered as eligible. The remaining nine sections were pooled into a 1.5-mL tube.

DNA Isolation

The QIAamp DNA Mini Kit(Qiagen) and GeneRead DNA FFPE kit (Qiagen) were used to extract DNA from fresh frozen and FFPE tissues, respectively, according to the manufacturer's protocols. The detailed methods used for DNA extraction from FFPE tissue samples were reported previously by our group (15). Briefly, FFPE tissue samples were dewaxed with deparaffinization solution, then incubated with lysis buffer for 1 h. After incubation at 90°C to remove cross-links, Uracil-N-Glycosilase was added for the specific removal of deaminated cytosine residues from the DNA. For fresh frozen tissues, 25 mg of tumor tissues were used for DNA extraction. After being homogenized with the Bioprep-24 homogenizer, the tissues were incubated in lysis buffer at 56°C until the tissues were completely lysed. RNase A was used to digest the RNA molecules from the genomic DNA. For both FFPE and fresh frozen tissue, DNA was eluted and quantified with a Qubit 3.0 fluorometer and dsDNA HS assay kit (Life Technologies).

DNA Amplification and Sequencing

For sequencing, 20 ng DNA extracted from each FFPE or fresh frozen sample was used for library construction. In brief, gene-specific PCR was used to amplify 103 regions in the first round, followed by purification via size selection. The details of the first round of PCR were showed in **Table 1**. Subsequently, the second round of PCR ("indexing PCR") was conducted (**Table 2**). The 22-gene panel was purchased from Pillar Biosciences, USA (ONCO/Reveal Lung & Colon Cancer Panel). The details of the 22-gene panel were same as described previously (14, 15). This process involves the addition of Illumina index adaptors to purify the products for sample tracking and sequencing. Lastly, the libraries were eluted in 22 μL nuclease-free water. The final libraries were quantified using a Qubit 3.0 fluorometer and dsDNA HS assay kit (Life Technologies), as per the manufacturer's protocol. The MiSeq was used for sequencing libraries according to the manufacturer's protocol. Each gene library was normalized to 4 nM and combined at equal volume (4 μL). The mixed library was then denatured with 0.2 N NaOH and diluted up to the concentration of 15 pM for sequencing with MiSeq Reagent Kit v2 at 300 cycles (or 20 pM for v3). Sample quality of FFPE tissues was determined by the amount of amplifiable DNA using qPCR and sequencing libraries were evaluated using Bioanalyzer.

TABLE 1 | The procedure details of the first round of PCR ("gene-specific PCR").

Temperature	Time	Cycles
95°C	15 min	1
95°C	1 min	5
60°C	6 min	
95°C	30 s	18
72°C	3 min	
8°C	Hold	1

TABLE 2 | The procedure details of the second round of PCR ("indexing PCR").

Temperature	Time	Cycles
95°C	2 min	1
95°C	30 s	5
66°C	30 s	
72°C	60 s	
72°C	5 min	1
8°C	Hold	1

Variant Analysis

Subsequently, they were annotated with the 1000 genomes (https://www.internationalgenome.org/home), which is one of the most frequently used databases in genetic research. Variants were selected for known single nucleotide polymorphisms (SNPs) and synonymous mutations. All non-coding, silent, synonymous, unknown and common germline variants were filtered out, as well as all variants present in 1,000 G data (22). Moreover, all variants at a locus with coverage of <200, or variants with a variant frequency <0.05 were excluded. The remaining mutations were assessed using the Catalog of Somatic Mutations in Cancer (COSMIC) database (14, 23). The SIFT (http://sift.jcvi.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) online databases were used to analyze the clinical significance of these variants.

Statistical Analysis

All data were analyzed using the Statistical Package for Social Sciences (SPSS 22.0, Chicago, IL). Categorical parameters were recorded as frequency and percentage; continuous parameters were described as mean and standard deviation, or median and interquartile range (IQR), as appropriate. Concordance rate and Kappa coefficient were used to compare FFPE and fresh frozen tissue in terms of mutation results, at both the

variant and gene levels. $P < 0.05$ (two-sided) was considered as statistically significant.

RESULTS

Clinicopathological Features of Patients Included in the Study

Of the 118 patients with CRC included in this study, 72 (61%) were men. The median age (IQR) was 62 (53–69) years. Most of the patients had TNM stage II ($n = 36$) or III ($n = 47$) disease. Eleven (9.3%) patients received preoperative chemotherapy (**Table 3**). The storage period of FFPE tissues ranged from 3 to 24 months, with a median period of 10 months.

TABLE 3 | Clinicopathological characteristics of patients included in the study.

Parameters		*N*(%)
Gender	Male	72 (61.0%)
	Female	46 (39.0%)
Gross type	Protruding	24 (20.3%)
	Ulcerative	91 (77.1%)
	Infiltrative	3 (2.5%)
Tumor position	Right sided colon cancer	26 (22.0%)
	Left sided colon cancer	40 (33.9%)
	Rectal cancer	52 (44.1%)
Differentiation	Well	0 (0.0%)
	Moderate	95 (80.5%)
	Poor	23 (19.5%)
T	1	2 (1.7%)
	2	17 (14.4%)
	3	81 (68.6%)
	4	18 (15.3%)
N	0	57 (48.3%)
	1	36 (30.5%)
	2	25 (21.2%)
M	0	95 (80.5%)
	1	23 (19.5%)
TNM	I	12 (10.2%)
	II	36 (30.5%)
	III	47 (39.8%)
	IV	23 (19.5%)
Tumor deposit	No	97 (82.2%)
	Yes	21 (17.8%)
Perineural invasion	No	106 (89.8%)
	Yes	12 (10.2%)
Vascular invasion	No	101 (85.6%)
	Yes	17 (14.4%)
Preoperative chemotherapy	No	107 (90.7%)
	Yes	11 (9.3%)
Age (years, IQR)		62 (53–69)
Diameter (cm, IQR)		4.1 (3.0–6.0)
CEA (ng/mL, IQR)		4.7 (2.0–15.4)
CA19-9 (U/mL, IQR)		5.4 (13.0–32.9)

IQR, interquartile range.

Comparisons of Gene Mutation Characteristics Between FFPE and Fresh Frozen Tissue

The median and interquartile of total coverage across all genes were 3739 (2148–5866) reads for fresh frozen tissues, which were significantly higher than those of FFPE tissues [2,814 (1,784–3,936) reads] ($P < 0.001$). In our study, 117 patients (99.2%) had one or more variants, with 226 variants in FFPE tissues and 221 in fresh frozen tissues. All variants identified in the FFPE and fresh frozen tissues are showed in **Supplemental Table 1**.

FFPE tissue analysis revealed at least one variant in 112 patients (94.9%), yielding a total of 226 variants. Among 112 patients, 44, 39, 19, 7, 2, and 1 had one, two, three, four, six and seven variants, respectively. The genes mutated most frequently were *TP53* (54.2%), *KRAS* (47.5%), *PIK3CA* (21.2%), and *FBXW7* (15.3%). No mutations were identified in *ALK, FGFR1, FGFR3, MET, NOTCH1, or STK11* (**Figure 1**).

The analysis of fresh frozen tissue revealed the presence of at least one variant in 105 patients (89.0%), yielding a total of 221 variants. Among the 105 patients, 39, 36, 18, 5, 6, and 1 patients had one, two, three, four, five and six variants, respectively. The list of most frequently mutated genes was similar to that developed for FFPE tissues: *TP53* (72.9%), *KRAS* (45.8%), *PIK3CA* (22.9%), and *FBXW7* (12.7%). No mutations were identified in *ALK, FGFR1, FGFR3, MET, NOTCH1, or STK11* (**Figure 1**).

A comparison between FFPE and fresh frozen tissue in terms of variant characteristics and their clinical significance was shown in **Table 4**. The impact of all variants were evaluated using the Variant Impact Predictor Database (VIPdb). No obvious difference in variant impact was found between the two samples (**Table 4**).

Comparison of Gene Mutations Between FFPE and Fresh Frozen Tissue at the Variant Level

A total of 129 variants were identified in this study. Among these variants, 96 were present in both FFPE and fresh frozen tissue; 27 variants were present in FFPE tissues alone; 6 variants were present in fresh frozen tissue alone. Of the 27 variants that existed in FFPE only, 59.3% (16/27) were C>T/G>A or A>G/T>C transitions, 14.8%(4/27) were G>T/T>G transversions, 7.4%(2/27) were A>C transversions, 7.4% (2/27) were A>T/T>A transversions, 3.7%(1/27) was C>G transversion and 7.4% (2/27) were deletion. Among the 6 variants that existed in fresh frozen tissue only, 83.3% (5/6) were C>T/G>A or A>G/T>C transitions, and 16.7% (1/6) was A>C transversion. Comparison of the gene mutations at the variant level revealed that concordance rates were 100.0% for 38.0% (49/129) variants, and concordance rates were higher than 94.0% for all variants (**Supplemental Table 2**). Kappa coefficients were higher than 0.500 in 64.3% (83/129) of variants (**Supplemental Table 2**).

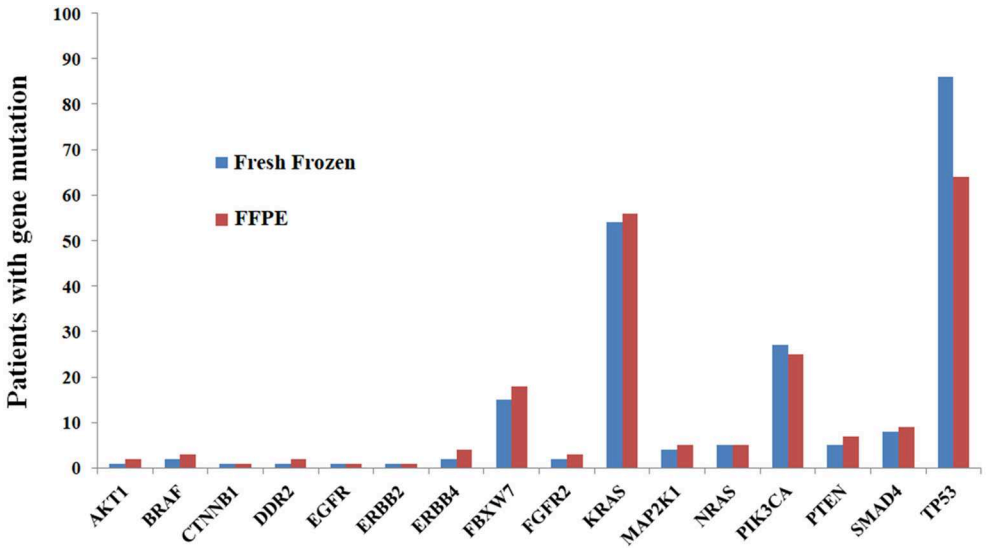


FIGURE 1 | Comparisons of mutation results between FFPE and fresh frozen tissue at the gene level, showing high concordance.

Comparison of Gene Mutations Between FFPE and Fresh Frozen Tissue at the Gene Level

The concordance rates of gene mutations ranged from 73.8 to 100.0% for all genes (**Table 5**, **Figure 1**). Kappa coefficients were >0.500 in 81.3% (13/16) of genes (**Table 5**).

DISCUSSION

In the present study, we found that 117 patients (99.2%) had one or more variants, with 226 variants in FFPE tissues and 221 in fresh frozen tissues. The concordance rates of gene mutation results in FFPE vs. fresh frozen tissue were higher than 94% for all variants, with Kappa coefficient >0.500 in 64.3% (83/129) variants. At the gene level, concordance ranged from 73.8 to 100.0%, with Kappa coefficient >0.500 in 81.3% (13/16) of genes. Our results indicated that the mutation results for FFPE and fresh frozen tissue were highly concordant at the variant and gene levels, but there were still some important differences in mutation results between the two tissue types.

The total coverage of fresh frozen tissues was significantly higher than that of FFPE tissues. We hypothesized that this could be due to the smaller DNA fragment size in FFPE tissues as some were below the detection limit. In our study, 226 variants were identified in FFPE tissues and 221 variants were identified in fresh frozen tissues. The more variants detected in FFPE samples might be caused by DNA damage during the formalin fixation process (e.g., fragmentation, degradation, crosslinking). Of the 27 variants identified only in FFPE tissues, 16 (59.3%) were C>T/G>A or A>G/T>C transitions. These 16 transitions may be artifacts secondary to postmortem deamination of cytosine or adenine to uracil or hypoxanthine residues (20). We took special care to decrease the rate of false-negative results by including only sections with >40% tumor cells, because adjacent normal cells usually have no mutations (20). In a study by Gallegos et al., the authors compared FFPE samples with fresh frozen tissue samples from 47 lung cancer patients in terms of EGFR and KRAS mutations. The authors showed that the success rate of PCR amplification was only 50% in FFPE tissues, with a false-positive rate up to 50% (19). The high false-positive rate may be related to tissue type and fixation method (18). The fixation and archiving processes required for FFPE often lead to the degradation and fragmentation of DNA. In our study, all PCR primers were designed to make sure that the PCR products were <200 bp, which may have contributed to the high concordance observed.

A few small studies previously compared paired FFPE and fresh frozen tissue from cancer patients in terms of microRNA expression, gene expression, and DNA methylation (24–27). De et al. compared the results of whole-exome sequencing in ten matched FFPE and fresh frozen tissue samples from patients with melanoma (28), and found that the average concordance rate was 43.2% over a total of 1,299 variants for the chosen 27 genes (28). The low concordance rate may be related to the length of PCR product. Spencer et al. compared the variants of 27 cancer-related genes between 16 pairs of fresh frozen and FFPE tissues from patients with lung carcinoma (29), and found that the concordance rate was up to 96.8% in the single-nucleotide variants. Solassol et al. compared the mutation status of *KRAS* in 33 patients with metastatic CRC, using paired fresh frozen and FFPE tumor tissues. The findings obtained revealed a concordance rate of only 81.9% (20). Betge et al. studied gene mutations in paired FFPE and fresh frozen tissue samples from hepatic metastases from 10 patients with CRC. The results revealed a high concordance between samples, with

TABLE 4 | Comparison of fresh frozen and FFPE tissues in terms of variant characteristics.

	FFPE tissue	Fresh frozen tissue
Total variants	226	221
Consequence of variants		
Missense variant	195	192
Stop gained	19	17
Frameshift variant	5	5
Splice variant	5	4
Inframe deletion	2	3
Inframe insertion	0	0
Impact of variants		
Moderate	197	195
High	29	26
Variant type		
SNV*	216	211
Deletion	8	7
Insertion	2	3
SIFT analysis		
Deleterious	159	170
Tolerated	36	22
Unknown	31	29
PolyPhen-2 analysis		
Benign	56	51
Probably damaging	114	120
Possibly damaging	25	21
Unknown	31	29
Clinical significances		
Likely benign	3	0
Likely pathogenic	8	17
Pathogenic	63	87
Uncertain significance	10	17
Unknown	142	100

**SNV, single-nucleotide variant (single nucleotide replacement); single nucleotide deletions and insertions were classified separately.*

TABLE 5 | Concordance of mutation results between fresh frozen and FFPE tissues, at the gene level.

Mutation Genes	a*	b*	c*	d*	Concordance#	Kappa	*P*
AKT1	1	0	1	116	99.2%	0.663	<0.001
BRAF	1	0	2	116	99.2%	0.796	<0.001
CTNNB1	0	0	1	118	100.0%	1.000	<0.001
DDR2	1	0	1	116	99.2%	0.663	<0.001
EGFR	1	1	0	114	98.3%	−0.009	0.925
ERBB2	0	0	1	105	100.0%	1.000	<0.001
ERBB4	2	1	2	113	97.5%	0.559	<0.001
FBXW7	5	3	13	64	90.6%	0.706	<0.001
FGFR2	1	0	2	113	99.1%	0.796	<0.001
KRAS	16	8	46	92	85.2%	0.679	<0.001
MAP2K1	1	0	4	99	99.0%	0.884	<0.001
NRAS	3	3	2	106	94.7%	0.372	<0.001
PIK3CA	10	12	18	90	83.1%	0.512	<0.001
PTEN	5	2	6	57	90.0%	0.575	<0.001
SMAD4	5	2	6	111	94.4%	0.602	<0.001
TP53	10	35	54	73	73.8%	0.481	<0.001

**a, Mutation identified in fresh frozen tissue, but not in FFPE tissue; b, Mutation identified in FFPE tissue, but not in fresh frozen tissue; c, Mutation identified in FFPE and fresh frozen tissue; d, Mutation was not identified in FFPE or fresh frozen tissue; #Concordance = (c+d)/(a+b+c+d).*

21 identical variants and 2 different variants (22). All of the above-mentioned studies used NGS to detect mutation, but the testing method were different. The various concordance rates may be related to the various factors such as primer, length of PCR product, tumor type, and testing methods. These results proved that remarkable differences existed between results of FFPE tissues and fresh frozen tissues, which were consistent with our finding. To the best of our knowledge, this is the first study of its kind to systematically compare the rate of mutation in paired fresh frozen and FFPE CRC tissue samples, at the gene and variant levels. Furthermore, our study had a relatively large sample size (118 patients), compared with previous relevant studies.

Although fresh frozen tissue is the gold standard for molecular analyses, its use in clinical practice is impractical because of its high cost and technical difficulty (20). Based on the results of this study, we suggest that archived tissues from pathology departments may be used for mutation detection with the 22-gene panel. The feasibility of using FFPE tissues for mutation detection will facilitate future studies of CRC. However, because the accuracy of mutation detection in FFPE tissues is influenced by multiple factors, it is important to standardize the procedure in order to minimize variability. Standardized protocols should be elaborated for sample preparation, storage room requirements, library preparation, evaluating the quality of extracted DNA, and the exclusion of poor-quality samples (22).

This study had some limitations. First, it was a retrospective study of stored FFPE tissues. In clinical practice, the fixation and embedding of specimens (which includes tissue thickness, fixative volume, and fixation time of 24–72 h) was not strictly controlled. Variation in the above factors may have affected the quality of preserved DNA, leading to inaccurate results (20). Second, the processes of fixation and embedding may result in deamination, leading to artifactual mutations or false-negative results. These factors may have had variable impacts at different tumor sites (20). In this study, we only took 10 consecutive sections of FFPE tumor tissues. We were therefore unable to rule out the possibility of intra-tumor heterogeneity in terms of DNA deamination. Third, although it is the largest study of its kind, larger prospective studies are required to validate our results. The number of variants will increase with sample size, leading to a more accurate evaluation of concordance, which is especially important for rare variants. Fourth, our study included FFPE specimens that had been stored in the pathology department for <2 years. The degradation and fragmentation of DNA in FFPE tissues may increase with time. Therefore, our observations cannot be extrapolated to all FFPE specimens.

CONCLUSION

The gene mutation results of a 22-gene panel showed high concordance between paired FFPE and fresh frozen tissue samples, at both the variant and gene levels. This indicates that FFPE tissues stored for <2 years may be used as an alternative to fresh frozen tissue for detecting gene mutations in patients with CRC. However, any alteration in the preparation or detection process may affect the accuracy of results. Factors that may be affected include fixation time, duration of storage of FFPE specimens, DNA sample quality, and tools used to analyze the variants. This should be taken into consideration when interpreting the findings presented above. Furthermore, the mutation results still showed some differences between tissues. Therefore, fresh frozen tissue should not routinely be replaced with FFPE for mutation analysis with a multi-gene panel; instead, FFPE is a reasonable alternative for fresh frozen tissue when the latter is unavailable. In addition, if the clinical response of *EGFR*-targeted therapy were not consistent with the mutation results based on FFPE tissue, gene mutation test might be performed again with fresh frozen tissue.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article can be found in the supplementary files, and are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethical committee at Changhai Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LL, CB, and WZ: conceptualization. XG, JL, and HG: data curation and writing—original draft. XG and JL: formal analysis. XG: funding acquisition. HG, GY, and PL: investigation. LH: methodology. CB: resources. LL and WZ: supervision. GY: validation. GY, PL, LH, LL, CB, and WZ: writing—review and editing.

FUNDING

This work was sponsored by Shanghai Pujiang Program (#2019PJD052) and the National Natural Science Foundation of China (#81572332 and #81572358).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00310/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD,Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69:7–34. doi: 10.3322/caac.21551
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*. (2019) doi: 10.3322/caac.21565
3. Stewart B, Wild, CP. (Eds.). *World Cancer Report 2014*. Lyon: International Agency for Research on Cancer (IARC) (2014).
4. Marmol I, Sanchez-de-Diego C, Pradilla Dieste A, Cerrada E,Rodriguez Yoldi MJ. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int J Mol Sci*. (2017) 18:197–236. doi: 10.3390/ijms18010197
5. Garde Noguera J, Jantus-Lewintre E, Gil-Raga M, Evgenyeva E, Macia Escalante S, Llombart-Cussac A, et al. Role of RAS mutation status as a prognostic factor for patients with advanced colorectal cancer treated with first-line chemotherapy based on fluoropyrimidines and oxaliplatin, with or without bevavizumab: a retrospective analysis. *Mol Clin Oncol*. (2017) 6:403–8. doi: 10.3892/mco.2017.1149
6. Goel G. Molecular characterization and biomarker identification in colorectal cancer: toward realization of the precision medicine dream. *Cancer Manag Res*. (2018) 10:5895–908. doi: 10.2147/CMAR.S162967
7. Liu D, Li J, Gao J, Li Y, Yang R,Shen L. Examination of multiple UGT1A and DPYD polymorphisms has limited ability to predict the toxicity and efficacy of metastatic colorectal cancer treated with irinotecan-based chemotherapy: a retrospective analysis. *BMC Cancer*. (2017) 17:437. doi: 10.1186/s12885-017-3406-2
8. Liu XH, Lu J, Duan W, Dai ZM, Wang M, Lin S, et al. Predictive value of UGT1A1*28 polymorphism in irinotecan-based chemotherapy. *J Cancer*. (2017) 8:691–703. doi: 10.7150/jca.17210
9. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, Goossens M, Ouchene H, Hendriks-Cornelissen SJ, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. *Gastroenterology*. (2014) 146:643–6e8. doi: 10.1053/j.gastro.2013.12.002
10. Ziada-Bouchaar H, Sifi K, Filali T, Hammada T, Satta D,Abadi N. First description of mutational analysis of MLH1, MSH2 and MSH6 in algerian families with suspected lynch syndrome. *Fam Cancer*. (2017) 16:57–66. doi: 10.1007/s10689-016-9917-1
11. Normanno N, Rachiglio AM, Roma C, Fenizia F, Esposito C, Pasquale R, et al. Molecular diagnostics and personalized medicine in oncology: challenges and opportunities. *J Cell Biochem*. (2013) 114:514–24. doi: 10.1002/jcb.24401
12. Rey JM, Ducros V, Pujol P, Wang Q, Buisine MP, Aissaoui H, et al. Improving mutation screening in patients with colorectal cancer predisposition using next-generation sequencing. *J Mol Diagn*. (2017) 19:589–601. doi: 10.1016/j.jmoldx.2017.04.005
13. Tops BB, Normanno N, Kurth H, Amato E, Mafficini A, Rieber N, et al. Development of a semi-conductor sequencing-based panel for genotyping of colon and lung cancer by the onconetwork consortium. *BMC Cancer*. (2015) 15:26. doi: 10.1186/s12885-015-1015-5
14. Dijkstra JR, Tops BB, Nagtegaal ID, van Krieken JH,Ligtenberg MJ. The homogeneous mutation status of a 22 gene panel justifies the use of serial sections of colorectal cancer tissue for external quality assessment. *Virchows Arch*. (2015) 467:273–8. doi: 10.1007/s00428-015-1789-5
15. Gao XH, Yu GY, Hong YG, Lian W, Chouhan H, Xu Y, et al. Clinical significance of multiple gene detection with a 22-gene panel in formalin-fixed paraffin-embedded specimens of 207 colorectal cancer patients. *Int J Clin Oncol*. (2019) 24:141–52. doi: 10.1007/s10147-018-1377-1
16. Esteve-Codina A, Arpi O, Martinez-Garcia M, Pineda E, Mallo M, Gut M, et al. A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS ONE*. (2017) 12:e0170632. doi: 10.1371/journal.pone.0170632
17. Suciu BA, Pap Z, Denes L, Brinzaniuc K, Copotoiu C,Pavai Z. Allele-specific PCR method for identification of EGFR mutations in non-small cell lung cancer: formalin-fixed paraffin-embedded tissue versus fresh tissue. *Rom J Morphol Embryol*. (2016) 57:495–500.

18. Lehmann U,Kreipe H. Real-time PCR analysis of DNA and RNA extracted from formalin-fixed and paraffin-embedded biopsies. *Methods*. (2001) 25:409–18. doi: 10.1006/meth.2001.1263
19. Gallegos Ruiz MI, Floor K, Rijmen F, Grunberg K, Rodriguez JA,Giaccone G. EGFR and K-ras mutation analysis in non-small cell lung cancer: comparison of paraffin embedded versus frozen specimens. *Cell Oncol*. (2007) 29:257–64. doi: 10.1155/2007/568205
20. Solassol J, Ramos J, Crapez E, Saifi M, Mange A, Vianes E, et al. KRAS mutation detection in paired frozen and Formalin-Fixed Paraffin-Embedded (FFPE) colorectal cancer tissues. *Int J Mol Sci*. (2011) 12:3191–204. doi: 10.3390/ijms12053191
21. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, et al. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. *J Mol Diagn*. (2013) 15:234–47. doi: 10.1016/j.jmoldx.2012.11.006
22. Betge J, Kerr G, Miersch T, Leible S, Erdmann G, Galata CL, et al. Amplicon sequencing of colorectal cancer: variant calling in frozen and formalin-fixed samples. *PLoS ONE*. (2015) 10:e0127146. doi: 10.1371/journal.pone.0127146
23. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*. (2004) 91:355–8. doi: 10.1038/sj.bjc.6601894
24. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA*. (2007) 13:1668–74. doi: 10.1261/rna.642907
25. Kalmar A, Wichmann B, Galamb O, Spisak S, Toth K, Leiszter K, et al. Gene expression analysis of normal and colorectal cancer tissue samples from fresh frozen and matched formalin-fixed, paraffin-embedded (FFPE) specimens after manual and automated RNA isolation. *Methods*. (2013) 59:S16–9. doi: 10.1016/j.ymeth.2012.09.011
26. Kalmar A, Peterfia B, Wichmann B, Patai AV, Bartak BK, Nagy ZB, et al. Comparison of automated and manual DNA isolation methods for DNA methylation analysis of biopsy, fresh frozen, and formalin-fixed, paraffin-embedded colorectal cancer samples. *J Lab Autom*. (2015) 20:642–51. doi: 10.1177/2211068214565903
27. Moran B, Das S, Smeets D, Peutman G, Klinger R, Fender B, et al. Assessment of concordance between fresh-frozen and formalin-fixed paraffin embedded tumor DNA methylation using a targeted sequencing approach. *Oncotarget*. (2017) 8:48126–37. doi: 10.18632/oncotarget.18296
28. De Paoli-Iseppi R, Johansson PA, Menzies AM, Dias KR, Pupo GM, Kakavand H, et al. Comparison of whole-exome sequencing of matched fresh and formalin fixed paraffin embedded melanoma tumours: implications for clinical decision making. *Pathology*. (2016) 48:261–6. doi: 10.1016/j.pathol.2016.01.001
29. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD,Duncavage EJ. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *J Mol Diagn*. (2013) 15:623–33. doi: 10.1016/j.jmoldx.2013.05.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

REVIEW
published: 09 April 2020
doi: 10.3389/fonc.2020.00447

Changing Technologies of RNA Sequencing and Their Applications in Clinical Oncology

Ye Wang[1], Michael Mashock[2], Zhuang Tong[3], Xiaofeng Mu[1,4], Hong Chen[5], Xin Zhou[5], Hong Zhang[2], Gexin Zhao[2], Bin Liu[3] and Xinmin Li[2]**

[1] Clinical Laboratory, Qingdao Central Hospital, The Second Affiliated Hospital of Medical College of Qingdao University, Qingdao, China, [2] Department of Pathology & Laboratory Medicine, UCLA Technology Center for Genomics & Bioinformatics, Los Angeles, CA, United States, [3] Cancer Hospital of China Medical University, Liaoning Cancer Hospital and Institute, Shenyang, China, [4] Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China, [5] Qiqihaer First Hospital, Qiqihar, China

RNA sequencing (RNAseq) is one of the most commonly used techniques in life sciences, and has been widely used in cancer research, drug development, and cancer diagnosis and prognosis. Driven by various biological and technical questions, the techniques of RNAseq have progressed rapidly from bulk RNAseq, laser-captured micro-dissected RNAseq, and single-cell RNAseq to digital spatial RNA profiling, spatial transcriptomics, and direct *in situ* sequencing. These different technologies have their unique strengths, weaknesses, and suitable applications in the field of clinical oncology. To guide cancer researchers to select the most appropriate RNAseq technique for their biological questions, we will discuss each of these technologies, technical features, and clinical applications in cancer. We will help cancer researchers to understand the key differences of these RNAseq technologies and their optimal applications.

Keywords: RNA sequencing, bulk RNAseq, LCM-RNAseq, single-cell RNAseq, digital spatial profiling, spatial transcriptomics, fourth-generation RNAseq, next generation sequencing

OPEN ACCESS

Edited by:
Ingrid A. Hedenfalk, Lund University, Sweden

Reviewed by:
Livia Garzia, McGill University, Canada
Parvin Mehdipour, Tehran University of Medical Sciences, Iran

****Correspondence:***
Bin Liu
B13889108173@163.com
Xinmin Li
xinminli@mednet.ucla.edu

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *12 November 2019*
Accepted: *13 March 2020*
Published: *09 April 2020*

Citation:
Wang Y, Mashock M, Tong Z, Mu X, Chen H, Zhou X, Zhang H, Zhao G, Liu B and Li X (2020) Changing Technologies of RNA Sequencing and Their Applications in Clinical Oncology. Front. Oncol. 10:447. doi: 10.3389/fonc.2020.00447

BULK RNAseq

Since Bulk RNAseq was developed over a decade ago (1), it has become a popular genomic tool in the life science field and is shaping nearly every aspect of our understanding of genomic functions (2). Bulk RNAseq is used in >60% of all next-generation sequencing projects, including whole genome sequencing (WGS), whole exome sequencing (WES), MethySeq, chromatin immunoprecipitation sequencing (ChIP-seq), and ATAC-seq. It is the most widely used genomic technique for studying the transcriptional landscape and altered molecular pathways in human cancers. RNAseq consists of four key steps: total RNA extraction, library construction, sequencing, and data analysis. The biological question and RNA quality will dictate the type of library employed, the selection of kits, sequencing type, and sequencing depth (**Figure 1**).

Based on the biological questions of the researcher, there are two types of bulk RNAseq that may be employed. The first type is simple RNAseq analysis aimed at identifying differentially expressed genes or markers (signatures), in order to understand molecular mechanisms implicated in various biological processes or to guide for diagnosis and treatment. Single-read sequencing (1 × 50 or 1 × 75) is appropriate for these types of RNAseq experiments, and 20–30 million reads/sample is usually a sufficient read depth. The majority of the libraries for these purposes are prepared using the poly-A RNA selection approach. The second type of bulk RNAseq is transcriptome

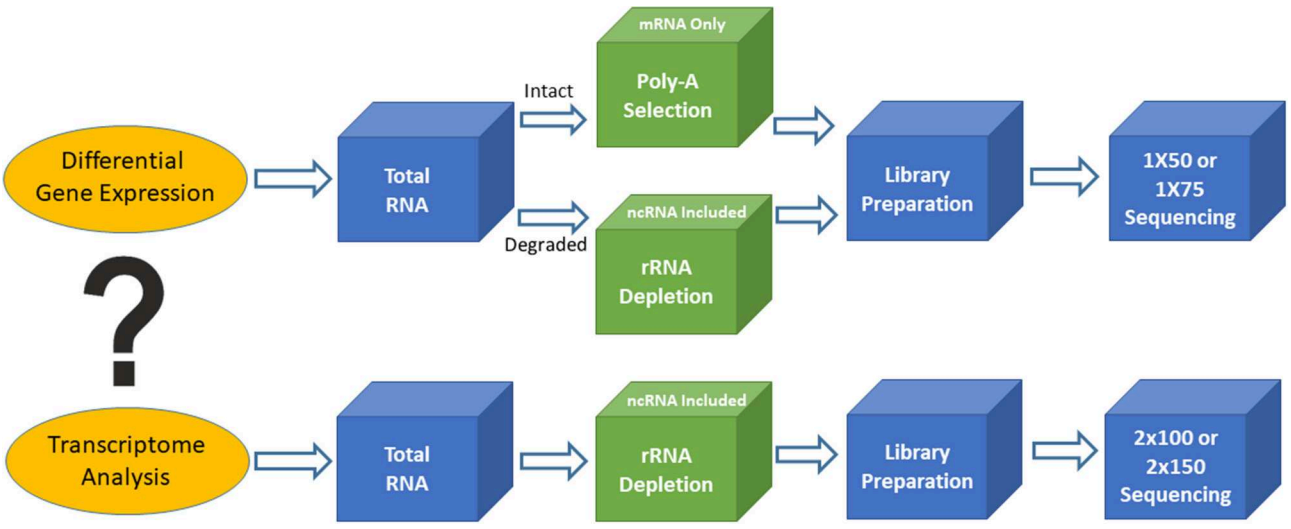


FIGURE 1 | Outline of two different types of RNA sequencing. Top is for differential gene expression and bottom is for transcriptome analysis.

sequencing, which not only achieves the goals of simple RNAseq, but also extends our knowledge of alternative splicing, point mutations, novel genes and transcripts, long non-coding RNAs, and fusion transcripts. Transcriptome analysis requires paired-end sequencing (2 × 100 or 2 × 150) at 40–50 million reads/sample from each direction, and the libraries are usually prepared using the rRNA depletion approach. The ENCODE guidelines (https://www.encodeproject.org/) provide various technical details for bulk RNAseq methodology and should be used for standards to assist in designing clinical RNAseq experiments with suggestions on sequencing depth, read length, replicates, and so on.

Bulk RNAseq is a cost effective and efficient tool for both cancer research and clinical applications (3–5). Today, clinical RNAseq is mainly used in novel gene fusion discoveries, panel-based accompanying gene fusion diagnosis, whole transcriptome-based biomarker (signature) discovery, and guidance for therapeutic treatment. A good example for using RNAseq data to detect novel, and clinically relevant, gene fusions involved in cancer used large-scale transcriptome analysis (6). The study employed an in-house developed bioinformatics pipeline to detect kinase gene fusions using nearly 7,000 cancer samples from The Cancer Genome Atlas. The study had immediate clinical implications as it led to the discovery of numerous novel and recurrent kinase gene fusions, for many of which approved or exploratory drugs now exist. Several other studies have provided additional evidence to support the discovery of novel gene fusions that have benefited directly from existing kinase inhibitors or new therapeutic opportunities (7–9). For example, whole transcriptome sequencing discovered novel FGFR gene fusions that subsequently led to the development of clinical trials of the tyrosine kinase inhibitors ponatinib and BGJ398, for the treatment of cancer patients with FGFR fusions (10).

The RNA-based gene fusion detection panel was one of the first RNAseq applications successfully translated into routine clinical practice (11). The Foundation One Heme is such a clinically validated panel, an integrated DNA/RNA profiling platform using targeted next-generation sequencing. This panel includes 265 genes frequently involved in gene fusions in various cancers, including FLT3, NPM1, CEBPA, BCRABL1, KIT, IDH2, IDH1, JAK2, MPL, PML-RARA, and MLL. The gene fusions detected by RNA sequencing can be validated by targeted DNA sequencing included in the Heme panel. The test can be used by physicians to identify targeted therapy options, detect alterations for prognosis, and sub-classify sarcoma diagnoses. Another clinically used popular gene fusion panel for a companion diagnosis is the Lung NGS Fusion Profile offered by NEO Genomics. This RNA-based next-generation sequencing panel detects translocations and fusions of six genes (ALK, NTRK1, NTRK2, NTRK3, RET, and ROS1) with known and novel fusion partners. Point mutations in select exons of these six genes are also frequently detected. In non-small cell lung carcinoma (NSCLC), the gene fusions of ALK, NTRK, RET, and ROS1 are detected with the approximate frequencies of 4–6, 1, 1–2, and 1–2%, respectively. Patients harboring such gene fusions may respond to several specific kinase inhibitors.

Another key clinical application of RNAseq is the discovery of biomarkers using whole transcriptome analysis. These biomarker signatures are used for cancer diagnosis, prognosis, and prediction. The clinical utility of gene expression signatures, developed by use of microarray, QRT-PCR, and other classic methods, have been well-established and used widely in routine clinical practice, including MammaPrint, OncotypeDX, and Prosigna for breast cancer, GeneFx for lung cancer, Prolaris for prostate cancer, and ColoPrint for colon cancer. The above commonly used, clinically validated signature panels can be potentially translated into RNAseq signature panels. In fact, translatability has been demonstrated by comparing

gene expression signatures in breast cancer between Affymetrix microarray and Illumina RNA-sequencing technology (12). In addition, systematic evaluation of RNAseq-based and microarray-based technology demonstrated that RNAseq is better in characterizing the transcriptome of cancer, and similar in clinical endpoint prediction, when compared with arrays. Zhang et al. (13) and Tom Lesluyes et al. (14) also used RNAseq technology with formalin-fixed paraffin-embedded (FFPE) tissue, a clinically more accessible sample type. This helped validate a prognostic signature of metastatic soft tissue sarcomas (CINSARC), developed using microarray with frozen tumor tissue, further demonstrating that CINSARC is a platform and material independent prognostic signature for metastatic sarcomas. Recently, some other RNAseq-based signatures have been developed and validated, such as the diagnostic signature for thyroid cancer (15), prognostic signatures for both Neuroblastoma (16), and Lung Adenocarcinoma (17), and predictive signatures for metastatic melanoma (18, 19).

Whole transcriptome RNAseq can also be used for guiding therapeutic treatment. Its feasibility and clinical utility in cancer were established in an early study of integrative clinical sequencing (whole exome and transcriptome analyses), which involved youths with relapsed or refractory cancer. This study identified potentially actionable findings in 46% of patients, some of those consequently changed treatment and genetic counseling (20). Later, Ronbinson et al. (21) demonstrated the broad utility of transcriptomic data in characterizing metastatic tumors and cancer treatment.

The RNAseq–based analyses have many advantages over DNA-based or other classic methods for clinical applications, including precise detail about base pairs, the ability to detect splicing variants, allele-specific expression, novel gene fusion, non-coding RNA, and novel RNAs. It is anticipated that RNAseq data will provide a more complete view of cancer-related genetic alterations and there is ample evidence to support such a view. For example, RNAseq identified an alternative breast cancer 1 (*BRCA1*) transcript in a subset of patients with breast cancer that was missed by conventional genomic analysis (22); Cabanski et al. (23) discovered a receptor tyrosine kinase (ROS1) gene fusion involved in a novel fusion partner with TMEM106B that was overlooked by standard FISH or PCR approaches; RNAseq analysis found that the germline allele-specific expression (ASE) of the transforming growth factor-beta (TGF-beta) type I receptor (TGFBR1) is associated with an increased risk of colon cancer (24); and RNAseq-based studies also show Long non-coding RNAs and miRNAs have prognostic potential in lung squamous cell carcinoma (25) and adenocarcinoma (26).

RNAseq can be applied to all tumor sample types including tumor cell lines, fresh or frozen tumor tissues, FFPE tumor tissues, and even liquid biopsy samples. The most accessible clinical tissue is FFPE tissue, which normally produces a limited amount of degraded RNA. This reality poses a challenge to produce high quality RNAseq data. Although RNAseq can reveal a more complete picture of genomic alterations in cancer, it is less commonly used compared to DNA sequencing in the clinical environment as RNA is less stable. With further technological advances, such as the development of new RNA preservative reagents, extraction methods, and RNA capture/hybridization protocols, the major hurdle of RNA stability will be overcome and the great potential of RNAseq in precision oncology will be fully realized.

It is important to keep in mind that bulk RNAseq reveals only an average gene expression profile from the studied tissue. Most tumors contain heterogeneous cell populations, including malignant cells, immune cells, fibroblasts, and vascular cells. This heterogeneity exists not only in the same tumor types from multiple patients, but also within various tumors from individual patients. The resulting average gene expression profile from tumor tissue can potentially weaken the true signals from a specific cell type that may drive tumorigenesis or resistance to treatments. For example, bulk RNAseq may have low detection sensitivity for biomarker discovery when the markers are only present in a specific cell type. This weakness can be addressed by alternative RNAseq technologies discussed in the following sections. The sensitivity issue for the companion diagnosis of gene fusion panels can be alleviated by increasing sequencing depth.

LASER CAPTURE MICRO-DISSECTED RNAseq

Many approaches have been developed in attempts to overcome the weaknesses of bulk RNAseq. One of the simplest approaches is laser capture micro-dissected RNAseq (LCM-RNAseq) (27). The key procedures of LCM-RNAseq consist of laser capture micro-dissection of cells of interest, followed by normal RNAseq, as illustrated in **Figure 2**. The majority of LCM-RNAseq employs FFPE materials, however the RNA extracted from FFPE materials are notoriously low quantity and quality and the LCM procedures further reduce the RNA integrity. As such, it is necessary to optimize LCM-RNAseq workflow by considering two critical factors: (1) optimizing the LCM procedures to minimize damage of RNA, which includes proper selection of the LCM instrument with IR laser, and (2) using a suitable RNAseq library construction kit that is optimized for the limited amount of degraded RNA (27), such as the SMARTer Stranded Total RNA-Seq Kit v2 (pico input mammalian; 250 pg−10 ng RNA input). This kit has a built-in CRIPR/CAS9-mediated rRNA depletion procedure without an independent rRNA depletion step. The kit is capable of depleting rRNA with only picograms of degraded RNA, which is not possible with library construction kits that employ a separate rRNA depletion step (normally requires >100 ng of RNA).

As the vast resource of clinical samples are FFPE tissues, further refining and improving existing LCM-RNAseq protocols could have far reaching impacts in both retrospective studies and current clinical testing of tumor samples. Recently, several new LCM-RNAseq methods were developed to address the constraints of the low input of degraded RNA derived from LCM FFPE tissues. By improving pre-amplification procedures, Singh et al. (28) claimed that sequencing data derived from as few as 10 LCM isolated single cells can reliably and sensitively measure

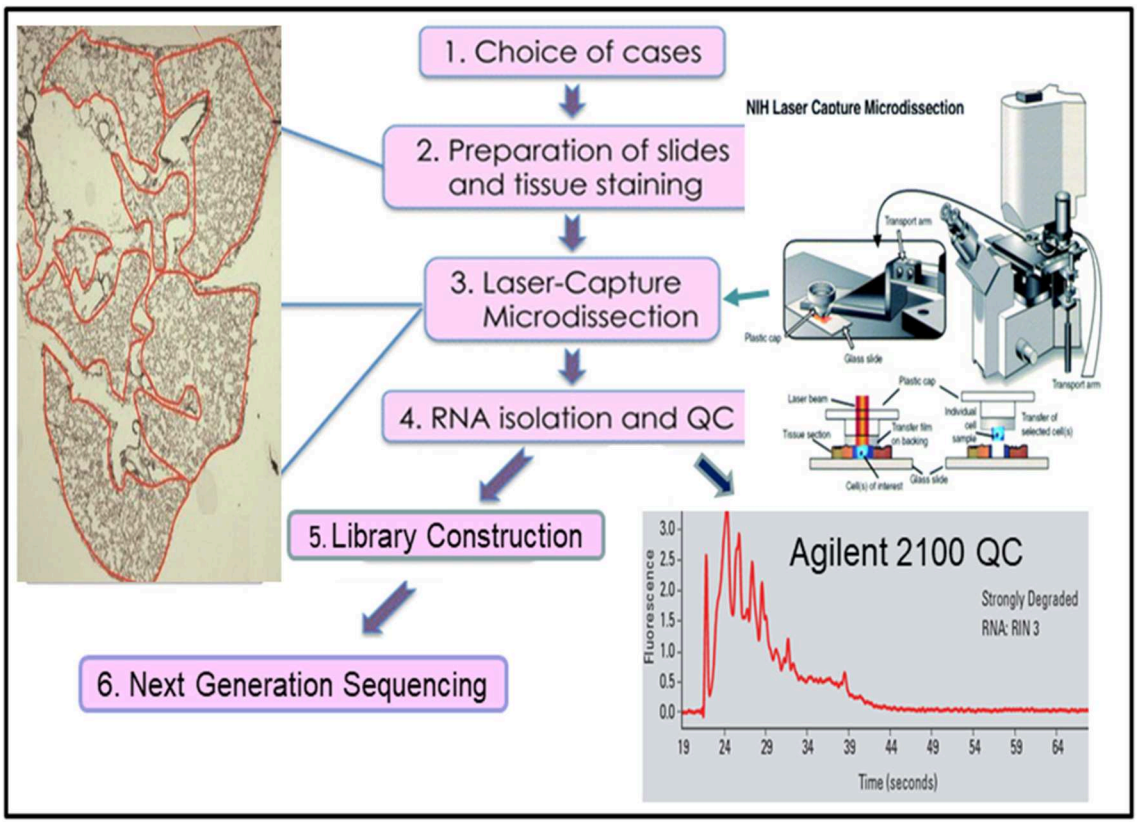


FIGURE 2 | The workflow of laser capture micro-dissected RNA sequencing.

cell-state heterogeneity in tumor tissues. A new LCM- Smart-3SEQ method was also developed that can quantify transcript abundance better with a low amount of degraded RNA from LCM cells (29). FFPEcap-seq is another method specifically designed for sequencing capped 5′ ends of RNA derived from FFPE samples. This 5′ capped RNAs-based method can also detect enhancer RNAs that arise from distal regulatory regions in addition to accurately capturing mRNA expression levels (30).

Due to technological limitations, the applications of LCM-RNAseq in cancer is relatively limited. The first application of LCM-RNAseq in cancer was performed in the normal-AIS-invasive adenocarcinoma progression model of lung cancer (31). The study established initial feasibility of LCM-RNASeq in a six patient sample population with an aim to identify biomarkers for lung cancer progression. Several studies have recently emerged, here we provide one example that highlights how LCM-RNAseq data can be used to deconvolve multiple cell type-specific gene expression profiles in cancer (32). In this study, the expression profiles of six specific tissue compartments of human glioblastoma (BGM) were analyzed using LCM-RNAseq techniques. These different compartments have interconnected complex networks and create a complex micro-environment that constantly gives signals to activate cell migration and promote cancer cell survival and proliferation (33). By isolating cell-specific gene expression signatures from different compartments, the authors found an overexpression of proangiogenic genes and pathways in pseudopalisading astrocytes cells. These overexpressed genes and pathways were known to promote cell survival and infiltrative growth, migration, and resistance to cancer-targeted therapies in GBM. Civita et al. (32) also observed a considerable up-regulation of growth factors signaling pathways in pseudopalisading cells compared to the tumor core. The data demonstrate that certain molecular events are region specific and different regions are molecularly interrelated. These findings provide potential targets for the development of new treatments and change current clinical management of BGM patients.

Although LCM-RNASeq can reveal cell population-specific gene expression profiles, it is associated with two practical issues. First, the procedure is time consuming and one can only work on a small number of cells at a time, thus data derived from 10 to 100 cells are generally less robust compared to the data obtained from $>1 \times 10^6$ cells of Bulk RNASeq. Secondly, the RNA yield is of low quantity and highly degraded, which requires more PCR cycles for amplification and thereby leading to poor quality RNAseq data with high PCR duplicates and possibly a biased gene expression profile. These issues need to be addressed by further technological improvement or alternative technologies. Interestingly, He et al. (11) develop a new algorithm (ADVOCATE) by using LCM-RNASeq derived

data from the malignant epithelium and stroma of pancreatic ductal adenocarcinoma (PDA). These LCM-RNASeq derived algorithms can predict the compartment–specific expression profiles from bulk RNASeq profiles of PDA. This pilot work provides a framework for potentially analyzing the cellular heterogeneity of cancer and expanding the utility of the large collection of bulk gene expression data.

SINGLE-CELL RNAseq

There are a number of different technologies available today for single-cell RNAseq (scRNAseq) (34). The Fluidigm C1 microfluidics system represents one of the early scRNAseq technologies. This system can process only 96 single cells in a single run over 1 day. The throughput is increased to 800 single cells by using improved Fluidigm IFC chips (35). Recently, microdroplet-based single-cell sequencing systems have become dominant players. Among those, the 10x Genomics Chromium® system is one of the most popular systems for high-quality scRNAseq. In contrast to LCM-RNAseq, the Chromium technology enables rapid analysis of the gene expression profiles of up to 10,000 individual cells in one experiment.

The Chromium scRNAseq workflow, similar to other microdroplet scRNAseq systems, is illustrated in **Figure 3**. The first critical step for scRNAseq is to isolate viable, individual cells from targeted tissues or cultured cells. Then, the Chromium microfluidics system is used to inexpensively generate hundreds of thousands of microdroplets, called GEMs, which are aqueous microdroplets surrounded by oil. Each GEM has a volume of ∼2 nl that includes all necessary reagents for reverse transcription (RT) and also contains a bead conjugated with a specific 80-base pair (bp) oligo sequence. This oligo sequence has several components, including adaptor sequences for next-generation sequencing (NGS) (Read 1), a cell-specific 10x barcode for identifying which cell the RNA comes from, the random molecular tags for identifying and quantifying unique mRNA transcripts [i.e., unique molecular identifiers (UMI)], and polyoligo-dT primers for mRNA binding. Following cell

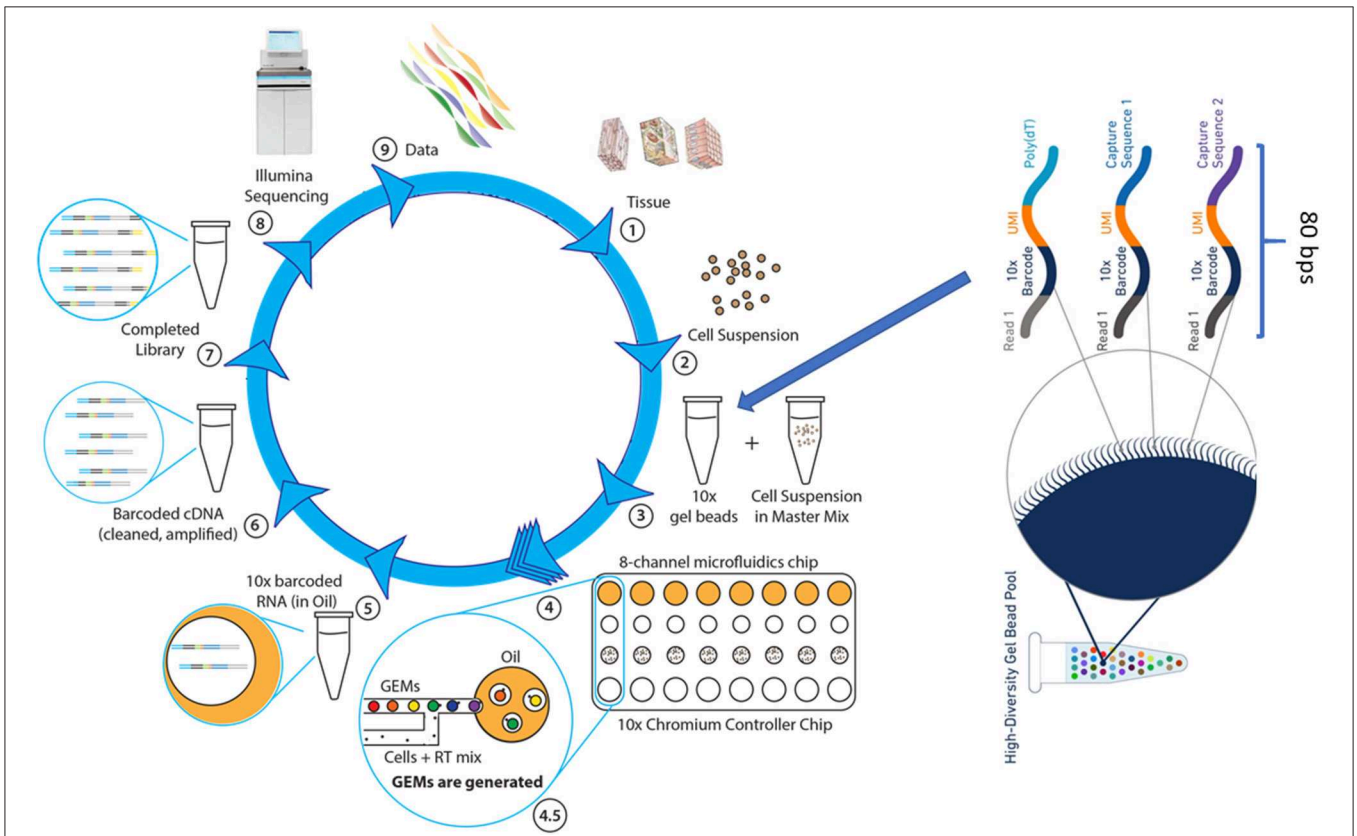


FIGURE 3 | The single cell RNASeq. (1) The dissociation of tissue cells and removal of dead cells and cell debris, (2) Viable cells are resuspended in the desired buffer at a correct concentration. (3) Cell suspension is combined with RT reagents and, along with gel beads and immersion oil, introduced into Chromium Controller chip. (4) Microfluidics chip generates single cell GEMs, a gel bead bound to a cell's RNA molecules. (4.5) Gel beads and cell suspension, in RT mix, are pushed into the immersion oil. (5) GEMs are transferred into PCR tubes and undergo RT-PCR to produce cDNA. (6) The cDNA, suspended in oil, is released from GEMs, removed from oil, and amplified via PCR. (7) Libraries are completed by fragmenting the cDNA to proper insert size, followed by end repair, A-tailing, and ligation of Illumina read 2 index, all occurring in a single PCR step. (8) Sample-specific index are added and the sequence-ready libraries are sequenced by using Illumina sequencer (NextSeq 500, HiSeq3000/4000 or NovaSeq6000). (9) The 10x single cell data analysis pipeline employs Cell Ranger to align reads and perform cluster and gene expression analysis, followed by Cell Loupe Browser to visualize and analyze the Cell Ranger data output.

lysis, the oligo-dT on the beads hybridizes to the poly-A tail of the released mRNA, and then RT reactions are carried out within the microdrops. At this step, the bead-specific oligo sequence is incorporated into the cDNA, which is used to align sequence reads back to a specific cell. GEMs are then broken, cDNAs are pooled together, amplified, and purified, followed by NGS library preparation using 10x Genomics protocols. The libraries can be sequenced by using Illumina sequencers at the following settings with Novaseq 100 cycle sequencing kit: 28 bps for read 1 (sequencing the cell-specific 10x barcode), 8 bps for i7 (sequencing the sample barcode), and 91 bps for read 2 (sequencing single-cell RNA). The 10x Genomics Chromium system and its associated library construction kits are commercially available. This system offers significant advantages over other microfluidics systems, such as Dropseq and Fluidigm C1, primarily in data quality and throughput. Therefore, it has become an important analytical tool for researchers in many disciplines, particularly in cancer research.

One key feature of tumors is their heterogeneity. The analysis of bulk RNAseq is complicated by significant infiltration of stroma and other type of cells in the tumor. Given the quantitative nature of gene expression data, it can be difficult to deconvolve the functionally relevant signals from average signals derived from bulk RNAseq. The scRNAseq technology offers a complementary and powerful tool to dissect intratumoral transcriptomic heterogeneity (36), important for therapeutic response. A good example of such is an early study on drug resistance in a model of drug tolerance with a metastatic breast cancer cell line. By analyzing untreated, stressed, and drug-tolerant cell groups, authors demonstrated that drug-tolerant cells contain specific RNA variants in genes involved in microtubule organization, stabilization, cell adhesion and cell surface signaling (37). This drug-tolerant-specific RNA variants were absent in untreated or stressed cells. The generation of specific RNA variants increases heterogeneity and ensure the survival of a minority population that efficiently converse stress-tolerant cells back to normal cells. Single cell analysis can also provide insightful clue for tumor treatment. Due to the intratumoral heterogeneity, a given targeted therapy often eliminates a specific subpopulation of tumor cells while leaving others unharmed. To overcome this challenge, therapeutic strategies that can target multiple tumor subpopulations are critical. By analyzing numerous drug target pathways in various cell populations in metastatic renal cell carcinoma, Kim et al. used scRNAseq technology to successfully develop an optimized combinatorial therapeutic strategy that showed significantly improved response *in vitro* and *in vivo* compared to monotherapies (38).

Another major application of single cell sequencing is to characterize known cell types, subtypes, and previously unknown cell types within and surrounding tumors, and to identify the gene signature for given cell types (39–41). These studies have facilitated dissection of complex pathways in heterogeneous tumor tissues and have provided guidance for cancer treatment. Here, we highlight how single cell sequencing technology was used to identify new cell types and biomarkers in T cell infiltration. The status of T cell infiltration and their characteristics are associated with different prognostic outcomes (42) and it is important to the development of immunotherapies and the prediction of their clinical responses in cancers. In 2017, Zheng et al. (40) performed a comprehensive analysis of infiltrating lymphocytes in liver cancer and reveals distinctive functional composition of T cells in hepatocellular carcinoma (HCC). The study identified 11 large subsets as well as unique subpopulations, such as $CD8^+FOXP3^+$ regulatory-like cells and clonal TCRs, at single-cell level. They also identified *LAYN*, an HCC-associated Treg marker gene, which is associated with tumor-infiltrating exhausted $CD8^+$ T cells and poor prognosis. The authors have made this comprehensive single T cell database publicly available for the wide research community (http://hcc.cancer-pku.cn).

Today, single-cell transcriptomic analysis has revolutionized our understanding of cancer biology, including tumor heterogeneity and their therapeutic implications. However, the major limitation of the technology is the level of detail that can be resolved from the captured mRNA data. Although the 10X genomics chromium system can capture up to 10,000 cells in a single experiment, it can only recover a few thousand unique transcripts from a single cell. By deeper sequencing, this problem can be alleviated to a certain degree, but is still far less than ideal for full transcriptome analysis. In conclusion, bulk RNAseq, LCM-RNAseq, and single-cell RNAseq all suffer from a common weakness—lost critical spatial information due to the micro-dissection or cell dissociation at the early stage of these protocols, which impacts the understanding of cell functionality and pathological changes (43). These limitations can be addressed by recently developed spatial profiling technologies as discussed below.

DIGITAL SPATIAL PROFILING

Each organ of a complex organism consists of diverse cell types that often interact in highly structured manners under distinct microenvironments. Such highly structured spatial heterogeneity enables the organism to function correctly and efficiently. To fully understand gene functions in a given cell type, one must study gene expression in the context of the location of the cells in the tissue (44). However, none of the technologies discussed above can provide this critical spatial information. Traditionally, immunohistochemistry and *in situ* hybridization have been used to reveal spatial gene expression in tissue sections, but the throughput of these procedures is limited to the analysis of only one or a few genes at a time.

Recently developed digital spatial profiling (DSP) technology has made it possible to resolve spatial gene expression with significantly improved throughput. DSP is based on the nCounter® barcoding technology from NanoString Technologies to enable spatially resolved, digital characterization of mRNA expression in a highly multiplexed assay (up to 1,000 RNA targets). The key technology of the assay relies upon RNA hybridization probes conjugated to photo cleavable oligonucleotide tags. After binding of probes to their targeted mRNA on the slide-mounted FFPE tissue sections, the slide is

imaged. Then, the oligonucleotide tags are released from the regions of interest on the tissue via UV exposure. Released tags are quantified in an nCounter® assay. The counts of a specific tag, representing a specific mRNA, are mapped back to tissue location (defined region of interest), yielding a spatially-resolved digital profile of mRNA abundance (**Figure 4**).

The GSP technology from NanoString Technologies, called GeoMx DSP, is commercially available and has already shown valuable applications in elucidating tumor microenvironments, immuno-oncology biomarker discovery, and optimizing immunotherapy by using targeted small gene panels. Ihle et al. (45) characterized tumor microenvironment of lytic and blastic bone metastases in prostate cancer patients using DSP technology, and found a distinct set of immune cell populations and signaling pathways specifically present in lytic or blastic types of prostate cancer. The immune cells in blastic lesions were enriched for pSTAT3 and JAK-STAT pathway related genes while pAKT activity and PI3K-AKT pathway related genes were more active in lytic-type lesions. The direct implication of this finding is that the targeted therapies for pAKT or pSTAT3 can potentially be considered. In addition, the immune checkpoints, such as PD-L1, were identified in blastic prostate cancer, which can now be considered as a new therapeutic target for blastic prostate patients with bone metastases.

Two landmark studies demonstrated that DSP is also a powerful tool for biomarker discoveries and optimizing therapeutic strategy (46, 47). In these studies, authors demonstrated that the combined ipilimumab and nivolumab therapies had high response rates with more lymphoid infiltration, whereas treatment with nivolumab monotherapy had modest responses with a more clonal and diverse T cell infiltration in responders, respectively, and that low RNA expression of the IFN-γ signature was associated with relapse after combinational therapies (ipilimumab + nivolumab), while none of the patients with a high or intermediate IFN-γ signature has relapsed in high-risk melanoma patients (47). Both studies identified promising biomarkers for further

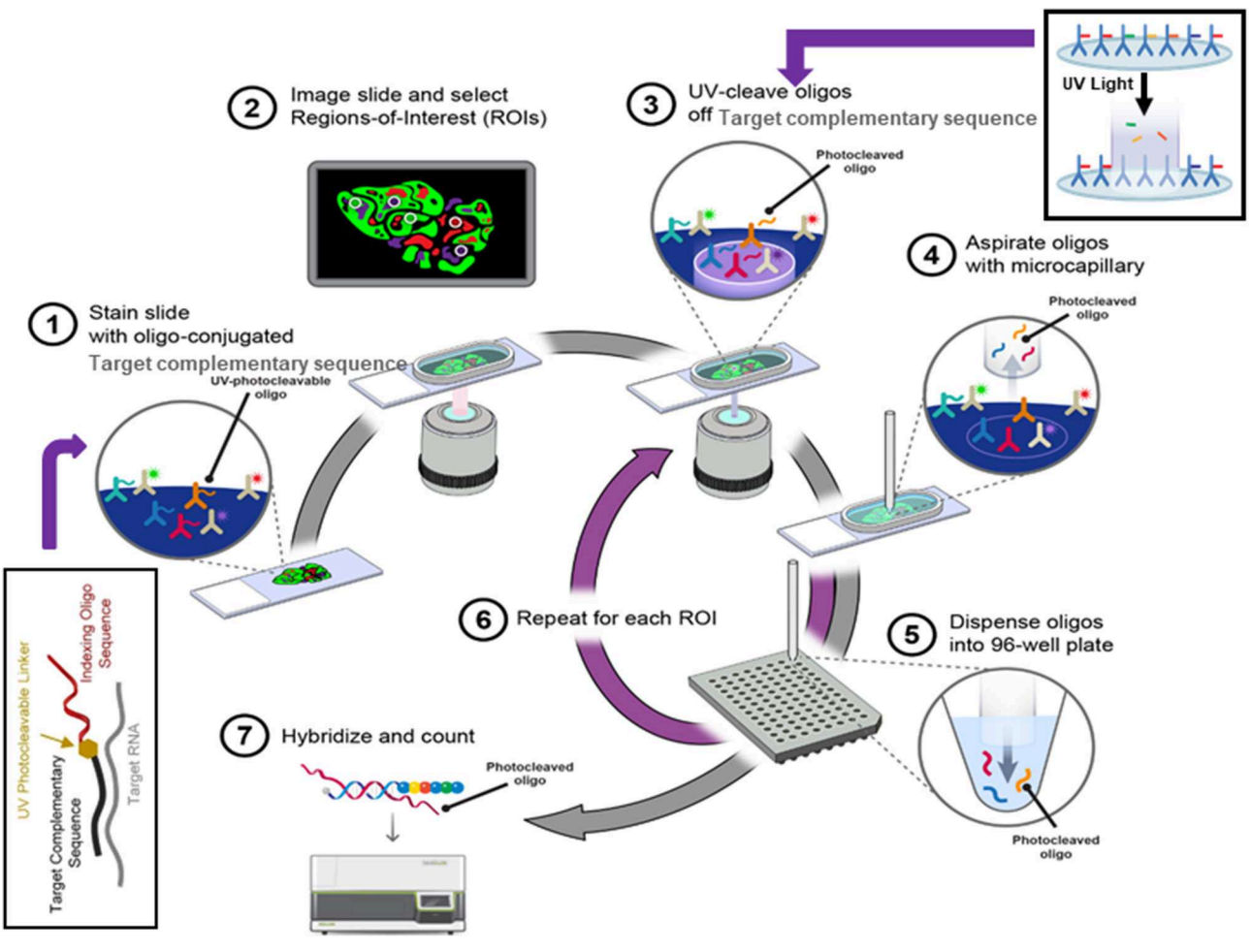


FIGURE 4 | The digital spatial profiling. (1) Apply high-plex oligo-labeled probes to FFPE slide. (2) Use visible wavelength low-plex imaging to establish tissue "geography." Select regions-of-interest (ROIs) for high-plex profiling. (3) UV-release oligo tags at selected ROIs. (4, 5) Collect and dispense released tags in microtiter plate. (6) Repeat the procedures for each ROI. (7) Index, hybridize, and count the tags per ROI and analyze the data with nSolver™ Advanced Analysis Software.

validation and offered possible solutions for optimizing immunotherapy strategy.

The DSP technology is becoming popular in the digital gene expression space. We expect that more applications of this technology in clinical oncology will emerge in the next few years. However, the relatively small number of mRNA targets that can be investigated simultaneously, the requirement of pre-knowledge of the gene, and inability to reveal sequencing information have limited its applications on a broader scale. Given that spatial gene expression is critical for understanding cell identity and function in tissue content, there is compelling reason to expect innovation to continue in the field.

SPATIAL TRANSCRIPTOMICS

A new spatial transcriptome technology has been in active development for several years. This technology overcomes the limitations of DSP technology by allowing scientists to study the whole transcriptome spatially (43). It can theoretically provide information similar to bulk transcriptome analysis along with spatial content.

Spatial transcriptomics adopted a strategy that integrates the features of microarray and the barcoding system of 10x Genomics. Briefly, a fresh-frozen tissue section is imaged and placed on a patterned, barcoded oligo-dT microarray slide. The capture probes on the slide include a T7 promoter for *in vitro* transcription (IVT), a partial Illumina handle for the sequencing, a spatial barcode for RNA localization, a UMI for removing amplification duplicates, and oligo-dT sequences for capturing mRNA. The tissue is then fixed and permeabilized to release RNA, which binds to adjacent oligo-dT sequences of capture probes. During the cDNA synthesis, the spatial barcodes indicating the location of each spot on the array are incorporated into the cDNA. One strand of double stranded cDNA, which contains the information of where the cDNA came from, is cleaved off from the array. The libraries are completed off the chip and then sequenced. The spatial barcode allows each read to be mapped to the correct spatial coordinates (**Figure 5**). Since spatial transcriptomics technology displays spatially-resolved whole transcriptome data on the original tissue section, scientists can choose all or any number of genes of interest to visualize and analyze.

The spatial transcriptomic analysis is generally applicable to fresh-frozen mammalian tissues and fresh plant tissues (48), and is potentially applicable to FFPE materials (44). The application of this powerful technology in the cancer arena is still limited and only tested in early technology access. One critical application is to investigate intratumor heterogeneity, which has posed a challenge to understanding tumor progression and treatment. Dr. Joakim Lundeberg's group has used this technology to explore the landscape of tumor heterogeneity in prostate cancer (49) and melanoma (50). By profiling 6,750 and 2,200 tissue regions in prostate and melanoma, respectively, they showed extraordinary gene expression heterogeneity between biopsies (distinct gene expression signature) and different regions within the biopsy (coexistence of several expression profiles) (50). The gene expression heterogeneity extends well-beyond cell type or tissue type. For example, the lymphoid area adjacent to the tumor region had a specific expression pattern (50), non-tumor tissue in close proximity to the tumor region displayed a gradient expression pattern and unique cancer

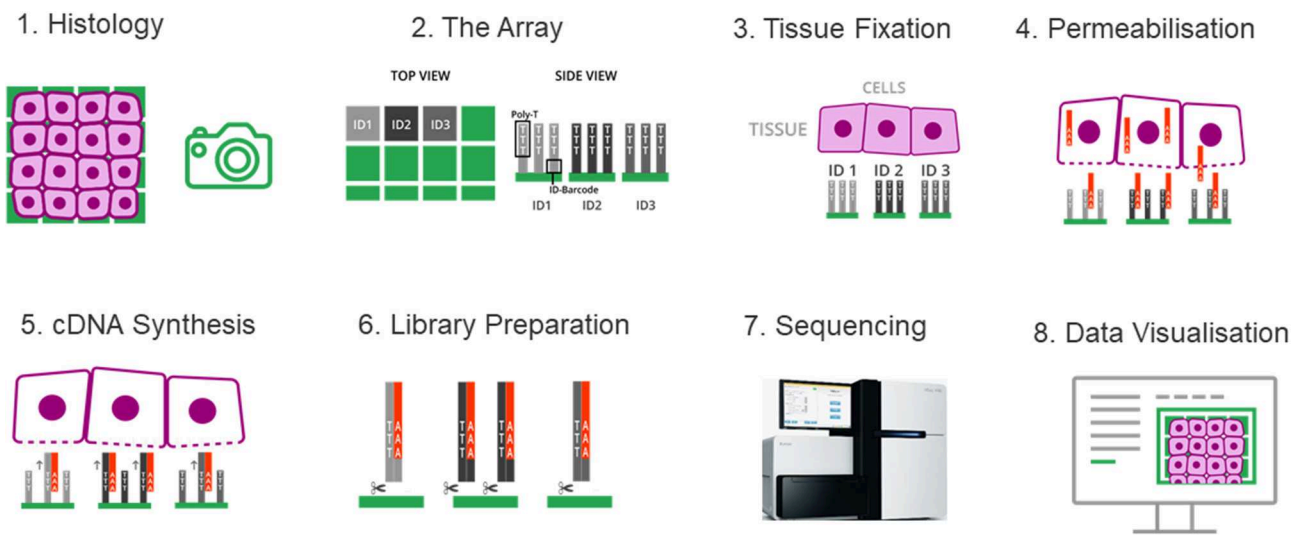


FIGURE 5 | The spatial transcriptomes. (1) A freshly frozen tissue section is prepared and attached onto the chip. (2) The chip contains an array of distinguishable capture probes. The Poly-T tails of these capture probes can bind the Poly-A tails of RNA molecules. (3) The tissue section is fixed and imaged, which makes it possible to overlay the cell tissue image and the gene expression data in a later step. (4) The tissue is permeabilized and RNA molecules can exit the cells through small holes created in the cell membrane, and bind to the adjacent capture probes on the chip. (5) cDNA synthesis is performed on the chip. (6) The cDNA-RNA-hybrids are cleaved off the chip, followed by library construction. (7) The libraries are sequenced. (8) Data are visualized to determine where genes are expressed and in what quantity.

expression profiles progress beyond pathologically defined tumor boundaries (49). All these findings suggest that the location-dependent gene expression is a reflection of cell-cell interactions in tumor microenvironment and certainly impact immune cells' function and therapy response. This is a key area that deserves further investigation in order to fully understand tumor metastasis.

Spatial transcriptomics technology has also been used in cancer diagnosis. Yoosuf et al. (51) used publicly available breast cancer spatial transcriptomics datasets, in combination with machine learning technique, to distinguish ductal carcinoma *in situ* (DCIS) from invasive ductal carcinoma (IDC). By identifying spatial transcriptomics signatures from known DCIS and IDC regions and training the machine learning method, they achieved a prediction accuracy of 95% for DCIS and 91% for IDC. This pilot study demonstrates the power of spatial transcriptomics in breast cancer diagnosis and subtype characterization.

The spatial transcriptomics technology does not require specialized equipment or pre-knowledge of gene sequences, and has a throughput higher than that of digital spatial profiling methods. The limitation of this technology is that the currently available product is not able to offer single-cell resolution as it is limited by the microarray spot size and spacing. However, in October, 2019, Stahl's group claimed that they have improved the spatial resolution of this technology by 1,400x so that it can now study spatial gene expression at the single cell level, opening the opportunity to detect tumor cells in the critical early stages (52). In today's clinical management of oncological patients, we need to quickly identify resistant clones during standard targeted therapies and discover robust, sensitive biomarkers to predict response to immunotherapy. Given the current resolution and sensitivity of spatial transcriptomics technology, this is an ideal technology to resolve these unmet needs.

FOURTH-GENERATION RNAseq

The ultimate goal of RNAseq is a simple, robust, spatially-resolved transcriptomic analysis at a single-cell resolution. The recent developments of fourth-generation sequencing technologies, such as *in situ* sequencing (ISS) and fluorescent ISS (FISSEQ), have potential toward this final destiny (36, 52). The detailed technologies, promises, and consequences were reviewed by Ke et al. (53). The ISS method applied padlock probes combined with rolling circle amplification (RCA) to generate *in situ* amplified, targeted sequencing libraries that are subsequently sequenced via sequencing-by-ligation NGS chemistry (53). Through sequencing of a molecular barcode, consisting of four bases in the non-target hybridization part of the padlock probes, the ISS method can simultaneously sequence up to 256 unique transcripts. As this method uses target-specific padlock probes to create rolling circle amplification products, it is used only for sequencing known genes, such as gene panels. In contrast, the fluorescent *in situ* sequencing (FISSEQ) method uses random hexamers with a sequencing primer tag to initiate *in situ* RT. Different from cDNA in ISS, the resultant cDNAs are circularized using CircLigase. During RT, dUTP is introduced and the cDNAs are cross-linked to tissue with the reagent BS (PEG)9 to prevent diffusion of the cDNAs. After RCA, the products are sequenced by using the same sequencing by ligation techniques. By applying FISSEQ with a 30-base read length, Lee et al. obtained 156,762 reads covering 8,102 annotated genes in human primary fibroblasts (36).

Compared to ISS methods, FISSEQ generates random libraries and, in principle, allows an unbiased analysis of all cellular transcripts at a single-cell resolution. Practically, the number of transcripts detected in each cell is low (54), since the majority of sequenced molecules are rRNAs. In this regard, ISS technology uses targeted gene panels and thus the sensitivity of ISS is around two orders of magnitude higher than that of FISSEQ for any given gene (55).

Although ISS and FISSEQ technologies each have their own strengths in detection, these technologies are still in their very early developmental stages and many technical aspects need to be addressed before they can be applied in cancer research and clinical applications. The main bottlenecks are tissue preparation, optimized methods for improving efficiency, computational tools, and imaging scale. However, fourth-generation RNAseq provides a direct *in situ* sequencing approach. If technical

TABLE 1 | Key strengths, weaknesses, and current suitable applications of six RNASeq technologies in clinical oncology.

	Strengths	Weaknesses	Suitable applications
Bulk RNASeq	High throughput, cost effective, mature technology	Average gene expression profile, lack of spatial content	Whole transcriptome-based biomarker discovery, targeted RNAseq panel for gene fusion
LCM-RNAseq	Cell type specific gene expression profile	Time consuming, low quality data, lack of spatial content	Tumor heterogeneity by dissecting cell type specific population
Single cell RNASeq	>10,000 single cell gene expression profile	High cost, a limited number of unique transcripts, lack of spatial content	Tumor heterogeneity, cell type characterization, and discovery
Digital spatial profiling	Spatial information, applicable to FFPE materials	Limited to small number of genes (gene panel only), lack of sequencing information	Tumor microenvironments, immuno-oncology biomarker discovery and optimizing immunotherapy
Spatial transcriptomics	Whole transcriptome analysis with spatial and sequencing information	Long procedures, early stage of technology	Tumor heterogeneity, tumor microenvironments, optimizing immunotherapy
Fourth generation RNAseq	*In situ* sequencing with future potential	In-matured technology	Not demonstrated yet

obstacles can be addressed in the coming years, fourth-generation RNAseq can potentially become a straightforward method for high-throughput spatial transcriptomic analysis.

SUMMARY

Among six RNAseq technologies described above, each has its own strengths, weaknesses and suitable applications, as summarized in **Table 1**. We anticipate that bulk RNAseq will remain the primary choice for clinical oncology in the near future, the application of single cell sequencing will further expand when it becomes more cost-effective, and technologies with spatial content will be the final destiny in precision oncology.

AUTHOR CONTRIBUTIONS

YW and XL conceived and designed the project. MM and XZ wrote the part of bulk RNAseq. ZT and HC wrote the part of laser capture micro-dissected RNAseq. XM and HZ wrote the part of single cell RNAseq. GZ and BL wrote the part of digital spatial profiling. YW wrote the whole manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 81670822, 81370990, and 81800805), and Qingdao Key Research Project (Nos. 17-3-3-10-nsh and 19-6-1-3-nsh).

REFERENCES

1. Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res.* (2007) 17:69–73. doi: 10.1101/gr.5145806
2. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet.* (2019) 20:631–56. doi: 10.1038/s41576-019-0150-2
3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* (2009) 10:57–63. doi: 10.1038/nrg2484
4. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* (2011) 12:87–98. doi: 10.1038/nrg2934
5. Buzdin A, Sorokin M, Garazha A, Glusker A, Aleshin A, Poddubskaya E, et al. RNA sequencing for research and diagnostics in clinical oncology. *Semin Cancer Biol.* (2019) 60:311–23. doi: 10.1016/j.semcancer.2019.07.010
6. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. *Nat Commun.* (2014) 5:4846. doi: 10.1038/ncomms5846
7. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell.* (2012) 22:153–66. doi: 10.1016/j.ccr.2012.06.005
8. Roberts KG, Li Y, Payne-Turner D, Harvey RC, Yang YL, Pei D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med.* (2014) 371:1005–15. doi: 10.1056/NEJMoa1403088
9. Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer.* (2013) 13:772–87. doi: 10.1038/nrc3612
10. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin.* (2016) 66:75–88. doi: 10.3322/caac.21329
11. He J, Maurer HC, Holmstrom SR, Su T, Ahmed A, Hibshoosh H, et al. Transcriptional deconvolution reveals consistent functional subtypes of pancreatic cancer epithelium and stroma. *bioRxiv*. (2018). doi: 10.1101/288779
12. Fumagalli D, Blanchet-Cohen A, Brown D, Desmedt C, Gacquer D, Michiels S, et al. Transfer of clinically relevant gene expression signatures in breast cancer: from Affymetrix microarray to Illumina RNA-Sequencing technology. *BMC Genomics*. (2014) 15:1008. doi: 10.1186/1471-2164-15-1008
13. Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* (2015) 16:133. doi: 10.1186/s13059-015-0694-1
14. Lesluyes T, Perot G, Largeau MR, Brulard C, Lagarde P, Dapremont V, et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur J Cancer.* (2016) 57:104–11. doi: 10.1016/j.ejca.2015.12.027
15. Han LO, Li XY, Cao MM, Cao Y, Zhou LH. Development and validation of an individualized diagnostic signature in thyroid cancer. *Cancer Med.* (2018) 7:1135–40. doi: 10.1002/cam4.1397
16. Zhou JG, Liang B, Jin SH, Liao HL, Du GB, Cheng L, et al. Development and validation of an RNA-seq-based prognostic signature in neuroblastoma. *Front Oncol.* (2019) 9:1361. doi: 10.3389/fonc.2019.01361
17. Shukla S, Evans JR, Malik R, Feng FY, Dhanasekaran SM, Cao X, et al. Development of a RNA-seq based prognostic signature in lung adenocarcinoma. *J Natl Cancer Inst.* (2017) 109:1–9. doi: 10.1093/jnci/djw200
18. Chen PL, Roh W, Reuben A, Cooper ZA, Spencer CN, Prieto PA, et al. Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. *Cancer Discov.* (2016) 6:827–37. doi: 10.1158/2159-8290.CD-15-1545
19. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* (2015) 350:207–11. doi: 10.1126/science.aad0095
20. Mody RJ, Wu YM, Lonigro RJ, Cao X, Roychowdhury S, Vats P, et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. *JAMA.* (2015) 314:913–25. doi: 10.1001/jama.2015.10080
21. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. *Nature.* (2017) 548:297–303. doi: 10.1038/nature23306
22. Gambino G, Tancredi M, Falaschi E, Aretini P, Caligo MA. Characterization of three alternative transcripts of the BRCA1 gene in patients with breast cancer and a family history of breast and/or ovarian cancer who tested negative for pathogenic mutations. *Int J Mol Med.* (2015) 35:950–6. doi: 10.3892/ijmm.2015.2103
23. Cabanski CR, Magrini V, Griffith M, Griffith OL, McGrath S, Zhang J, et al. cDNA hybrid capture improves transcriptome analysis on low-input and archived samples. *J Mol Diagn.* (2014) 16:440–51. doi: 10.1016/j.jmoldx.2014.03.004
24. Valle L, Serena-Acedo T, Liyanarachchi S, Hampel H, Comeras I, Li Z, et al. Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. *Science.* (2008) 321:1361–5. doi: 10.1126/science.1159397
25. Tang RX, Chen WJ, He RQ, Zeng JH, Liang L, Li SK, et al. Identification of a RNA-seq based prognostic signature with five lncRNAs for lung squamous cell carcinoma. *Oncotarget.* (2017) 8:50761–73. doi: 10.18632/oncotarget.17098
26. Siriwardhana C, Khadka VS, Chen JJ, Deng Y. Development of a miRNA-seq based prognostic signature in lung adenocarcinoma. *BMC Cancer*. (2019) 19:34. doi: 10.1186/s12885-018-5206-8
27. Farris S, Wang Y, Ward JM, Dudek SM. Optimized method for robust transcriptome profiling of minute tissues using laser capture microdissection and low-input RNA-seq. *Front Mol Neurosci*. (2017) 10:185. doi: 10.3389/fnmol.2017.00185
28. Singh S, Wang L, Schaff DL, Sutcliffe MD, Koeppel AF, Kim J, et al. *In situ* 10-cell RNA sequencing in tissue and tumor biopsy samples. *Sci Rep*. (2019) 9:4836. doi: 10.1038/s41598-019-41235-9
29. Foley JW, Zhu C, Jolivet P, Zhu SX, Lu P, Meaney MJ et al. Gene expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ. *Genome Res.* (2019) 29:1816–25. doi: 10.1101/gr.234807.118

30. Vahrenkamp JM, Szczotka K, Dodson MK, Jarboe EA, Soisson AP, Gertz J. FFPEcap-seq: a method for sequencing capped RNAs in formalin-fixed paraffin-embedded samples. *Genome Res.* (2019) 29:1826–35. doi: 10.1101/gr.249656.119
31. Morton ML, Bai X, Merry CR, Linden PA, Khalil AM, Leidner RS, et al. Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens. *Lung Cancer*. (2014) 85:31–9. doi: 10.1016/j.lungcan.2014.03.020
32. Civita P, Franceschi S, Aretini P, Ortenzi V, Menicagli M, Lessi F, et al. Laser capture microdissection and RNA-seq analysis: high sensitivity approaches to explain histopathological heterogeneity in human glioblastoma FFPE archived tissues. *Front Oncol*. (2019) 9:482. doi: 10.3389/fonc.2019.00482
33. Wels J, Kaplan RN, Rafii S, Lyden D. Migratory neighbors and distant invaders: tumor-associated niche cells. *Genes Dev*. (2008) 22:559–74. doi: 10.1101/gad.1636908
34. Haque A, Engel J, Teichmann SA, Lonnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*. (2017) 9:75. doi: 10.1186/s13073-017-0467-4
35. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, Kathiriya IS, et al. Single-cell resolution of temporal gene expression during heart development. *Dev Cell.* (2016) 39:480–90. doi: 10.1016/j.devcel.2016.10.001
36. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, et al. Highly multiplexed subcellular RNA sequencing *in situ*. *Science.* (2014) 343:1360–3. doi: 10.1126/science.1250212
37. Lee MC, Lopez-Diaz FJ, Khan SY, Tariq MA, Dayn Y, Vaske CJ, et al. Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proc Natl Acad Sci USA*. (2014) 111:E4726–35. doi: 10.1073/pnas.1404656111
38. Kim KT, Lee HW, Lee HO, Song HJ, Jeong da E, Shin S, et al. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol*. (2016) 17:80. doi: 10.1186/s13059-016-0945-9
39. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. (2017) 356:eaah4573. doi: 10.1126/science.aah4573
40. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell*. (2017) 169:1342–56.e16. doi: 10.1016/j.cell.2017.05.035
41. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun*. (2017) 8:15081. doi: 10.1038/ncomms15081
42. Sharma P, Allison JP. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell.* (2015) 161:205–14. doi: 10.1016/j.cell.2015.03.030
43. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science.* (2016) 353:78–82. doi: 10.1126/science.aaf2403
44. Salmen F, Stahl PL, Mollbrink A, Navarro JF, Vickovic S, Frisen J, et al. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat Protoc.* (2018) 13:2501–34. doi: 10.1038/s41596-018-0045-2
45. Ihle CL, Provera MD, Straign DM, Smith EE, Edgerton SM, Van Bokhoven A, et al. Distinct tumor microenvironments of lytic and blastic bone metastases in prostate cancer patients. *J Immunother Cancer*. (2019) 7:293. doi: 10.1186/s40425-019-0753-3
46. Amaria RN, Reddy SM, Tawbi HA, Davies MA, Ross MI, Glitza IC, et al. Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nat Med.* (2018) 24:1649–54. doi: 10.1038/s41591-018-0197-1
47. Blank CU, Rozeman EA, Fanchi LF, Sikorska K, van de Wiel B, Kvistborg P, et al. Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. *Nat Med.* (2018) 24:1655–61. doi: 10.1038/s41591-018-0198-0
48. Giacomello S, Salmen F, Terebieniec BK, Vickovic S, Navarro JF, Alexeyenko A, et al. Spatially resolved transcriptome profiling in model plant species. *Nat Plants*. (2017) 3:17061. doi: 10.1038/nplants.2017.61
49. Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstrahle J, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun*. (2018) 9:2419. doi: 10.1038/s41467-018-04724-5
50. Thrane K, Eriksson H, Maaskola J, Hansson J, Lundeberg J. Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage III cutaneous malignant melanoma. *Cancer Res.* (2018) 78:5970–9. doi: 10.1158/0008-5472.CAN-18-0747
51. Yoosuf N, Navarro JF, Salmen F, Stahl PL, Daub CO. Identification and transfer of spatial transcriptomics signatures for cancer diagnosis. *Breast Cancer Res*. (2020) 22:6. doi: 10.1186/s13058-019-1242-9
52. Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods.* (2013) 10:64–7. doi: 10.1038/nmeth.2281
53. Ke R, Mignardi M, Hauling T, Nilsson M. Fourth generation of next-generation sequencing technologies: promise and consequences. *Hum Mutat.* (2016) 37:1363–7. doi: 10.1002/humu.23051
54. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science.* (2017) 358:64–9. doi: 10.1126/science.aan6827
55. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc.* (2015) 10:442–58. doi: 10.1038/nprot.2014.191

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 28 April 2020
doi: 10.3389/fonc.2020.00385

RNA-Seq-Based TCR Profiling Reveals Persistently Increased Intratumoral Clonality in Responders to Anti-PD-1 Therapy

Ekaterina A. Zhigalova[1†], Anna I. Izosimova[2†], Diana V. Yuzhakova[2†], Lilia N. Volchkova[2], Irina A. Shagina[3,4], Maria A. Turchaninova[2,3,4], Ekaterina O. Serebrovskaya[2,3,4], Elena V. Zagaynova[2], Dmitriy M. Chudakov[1,2,3,4] and George V. Sharonov[2,3,4]*

[1] Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia, [2] Laboratory of Genomics of Antitumor Adaptive Immunity, Privolzhsky Research Medical University, Nizhny Novgorod, Russia, [3] Genomics of Adaptive Immunity Department, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, [4] Molecular Technologies Department, Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia

OPEN ACCESS

Edited by:
Ira Ida Skvortsova, Innsbruck Medical University, Austria

Reviewed by:
Alvaro Teijeira, ETH Zürich, Switzerland
Liang Ding, St. Jude Children's Research Hospital, United States

***Correspondence:**
Dmitriy M. Chudakov
chudakovdm@mail.ru

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *21 November 2019*
Accepted: *04 March 2020*
Published: *28 April 2020*

Citation:
Zhigalova EA, Izosimova AI, Yuzhakova DV, Volchkova LN, Shagina IA, Turchaninova MA, Serebrovskaya EO, Zagaynova EV, Chudakov DM and Sharonov GV (2020) RNA-Seq-Based TCR Profiling Reveals Persistently Increased Intratumoral Clonality in Responders to Anti-PD-1 Therapy. Front. Oncol. 10:385. doi: 10.3389/fonc.2020.00385

Substantial effort is being invested in the search for peripheral or intratumoral T cell receptor (TCR) repertoire features that could predict the response to immunotherapy. Here we demonstrate the utility of MiXCR software for TCR and immunoglobulin repertoire extraction from RNA-Seq data obtained from sorted tumor-infiltrating T and B cells. We use this approach to extract TCR repertoires from RNA-Seq data obtained from sorted tumor-infiltrating $CD4^+$ and $CD8^+$ T cells in an HKP1 ($Kras^{G12D}p53^{-/-}$) syngeneic mouse model of lung cancer after anti-PD-1 treatment. For both subsets, we demonstrate decreased TCR diversity in response to therapy. At a later time point, repertoire diversity is restored in progressing disease but remains decreased in responders to therapy in both $CD4^+$ and $CD8^+$ subsets. These observations complement previous studies and suggest that stably increased intratumoral $CD4^+$ and $CD8^+$ T cell clonality after anti-PD-1/PD-L1 therapy could serve as a predictor of long-term response.

Keywords: tumor-infiltrating lymphocytes, TCR repertoire, RNA-Seq, anti-PD-1, T cell clonality, MiXCR

INTRODUCTION

Active tumor infiltration by $CD8^+$ and Th1 T cells has repeatedly been shown to correlate with improved clinical outcomes in a variety of cancers (1–4). At the same time, it remains a matter of debate which proportion of these infiltrating T cells is actually tumor-reactive and could participate in an antitumor response (5), and this proportion may differ between different cancer types and individual patients.

T cell receptor (TCR) repertoire analysis can reveal the clonal content of tumor-infiltrating T cells, the presence of large clonal expansions (6), and the presence of clusters of convergent TCR variants that potentially respond to the same antigen (7–10). However, the prognostic and predictive value of TCR repertoire profiling in cancer immunotherapy remains a matter of investigation.

In a seminal work by Tumeh et al. (6), it was shown that high intratumoral T cell clonality—indicating the presence of large clonal expansions—may be associated with clinical response to anti-PD-1 therapy in patients with advanced melanoma. Furthermore, responders demonstrated a tendency toward increased clonal expansion during therapy. Tamura et al. (11) likewise observed increased intratumoral T cell clonality in response to peptide vaccines and oxaliplatin-based chemotherapy in colorectal cancer patients who exhibited long periods of progression-free survival. A combination of neoadjuvant ipilimumab with high-dose IFNα2b in advanced melanoma showed higher efficiency for patients exhibiting increased T cell clonality in the primary tumor at 6–8 weeks following neoadjuvant therapy (12).

Several studies have also shown that the analysis of peripheral blood TCR repertoire clonality could assist in predicting therapeutic outcomes. In particular, response to anti-PD-1 therapy has been associated with the initial presence of clonal peripheral blood T cell expansions in metastatic melanoma (13), although the opposite was reported for PD-L1 blockade in urothelial cancer (14). In another study of metastatic urothelial cancer patients treated with anti-PD-L1, clinical response was associated with high intratumoral T cell clonality and induced peripheral blood expansion of major tumor-resident T cell clones (15).

Response to anti-CTLA-4 therapy has been linked with initially low peripheral blood TCR clonality in melanoma (13) and pancreatic ductal adenocarcinoma (16), with the latter study also observing increased presence of clonal expansions over the course of therapy (16). These results are in line with the logic of anti-CTLA-4 action via blocking regulatory T cell (T_{reg})-mediated suppression of antigen-presenting cells and interclonal competition between $CD4^+$ T cells (17–20). This allows multiple novel expansions to arise, thereby broadening the peripheral TCR repertoire (21). Although anti-CTLA-4 therapy has been associated with essential remodeling and diversification of peripheral TCR repertoires, it has also been reported that improved clinical outcomes may be associated with the persistence of initially high-frequency clones during therapy (22). Using the ALICE algorithm on the data described in Robert et al. (21) and Subudhi et al. (23), we have also recently shown that the number of TCR sequences actively involved in current immune response—as judged by the number of clusters of non-randomly met (non-public) homologous TCR variants—increases after anti-CTLA4 therapy (10), suggesting reactivation of immune response to diverse antigens.

Notably, an increase in intratumoral T cell clonality was also observed in response to targeted therapy with a BRAF inhibitor, and persistence of initially detected dominant T cell clones was associated with therapy response (24). In a B16 mouse melanoma model, expansion of $CD8^+$ T cells within the tumor—but not in the periphery—was associated with antitumor effects (25). In $FGFR2^{K660N}/p53^{mut}$ lung cancer mouse model, reduced TCR clonality was found in responders receiving anti-PD-1 therapy in combination with an FGFR inhibitor (26). Thus, the current data on the dependence of response to different immunotherapies on the clonal composition of T cell repertoires remain incomplete and somewhat contradictory.

A recent study on the HKP1 ($Kras^{G12D}p53^{-/-}$) immunocompetent, syngeneic mouse lung cancer model, which is histologically similar to human adenocarcinoma (27), used RNA-Seq analysis of fluorescence-activated cell sorting (FACS)-sorted tumor-infiltrating $CD4^+$ and $CD8^+$ T cells in order to reveal the intrinsic features of T cell behavior associated with early immune response to anti-PD-1 therapy (28). This work showed that response to anti-PD-1 treatment was correlated with T cell subset-specific alterations, although the clonality of T cells was not specifically analyzed. However, TCR transcripts are present in bulk RNA-Seq data and enriched in sorted T cell RNA-Seq data, and MiXCR software allows one to extract TCR CDR3 repertoires with near-maximal efficiency and accuracy (29–31).

Here, we show that MiXCR efficiently extracts TCR and immunoglobulin repertoires from RNA-Seq data obtained from sorted tumor-infiltrating T or B cells. We applied this approach to extract TCRα and β CDR3 repertoires from the tumor-infiltrating T cell RNA-Seq data reported by Mittal and colleagues (28), and compared clonality and diversity parameters in responders to anti-PD-1 therapy vs. progressors and untreated control mice. At 1 week after the start of therapy, we find that TCR diversity goes down in both $CD4^+$ and $CD8^+$ T cells, reflecting clonal expansion. At a later time point, about 2 weeks after the start of the therapy, diversity remains low for responders, but reverts back to high diversity in progressors, reflecting reduced clonal expansion. These data demonstrate that the primary response to anti-PD-1 immunotherapy, as expressed by clonal expansion of T cells, is insufficient to provide sustained response to therapy, and that stability of the intratumoral clonal T cell expansions acquired in the course of the treatment is associated with longer-term response.

RESULTS AND DISCUSSION

RNA-Seq based immune repertoire profiling may work well for sorted T or B cells, where the percentage of CDR3-covering reads is relatively high (30). This approach—preferably using relatively long (e.g., 100+100-nt) paired-end sequencing—makes it possible to combine transcriptomic profiling of tumor-infiltrating lymphocyte populations of interest with relatively deep profiling of immune repertoires.

To illustrate and validate this approach, we performed RNA-Seq analysis on sorted T and B cells from mouse B16F0 melanoma

TABLE 1 | CDR3 counts from sorted T and B cells from B16F0 melanoma.

Subset	Cells sorted	RNA-Seq reads	IGH/TCRβ CDR3 containing reads	IGH/TCRβ CDR3 clonotypes
$CD19^+$ B cells	6,000–25,000	5–10 million	10,000–50,000	1,000–4,000
$CD4^+$ T cells (excluding T_{reg})	8,000–35,000	5–10 million	400–2,000	300–700

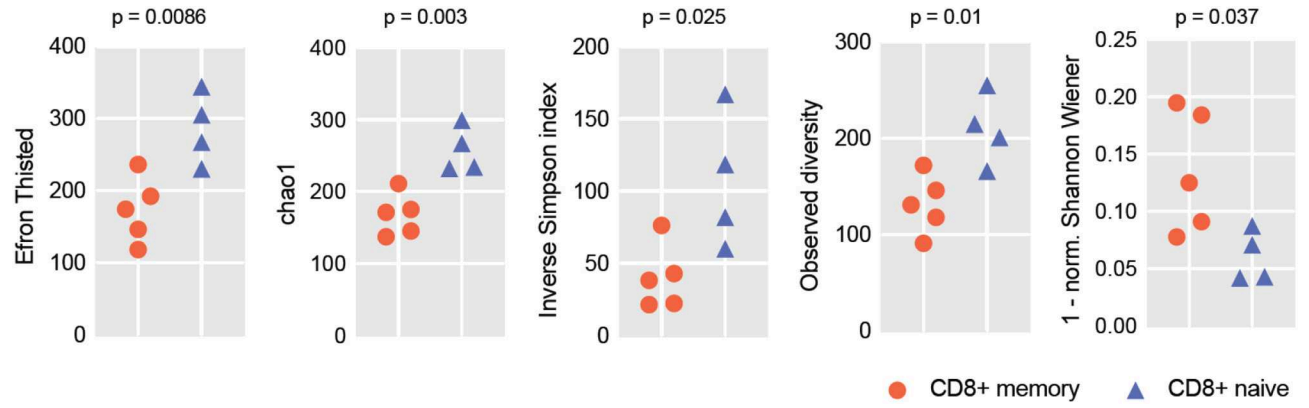


FIGURE 1 | RNA-Seq-based TCR repertoire diversity in naïve and memory CD8$^+$ T cells. Diversity metrics calculated for naïve and effector CD8$^+$ TCR α + β (combined) CDR3 repertoires are shown. Note that each sample is represented by 1,500 randomly chosen CDR3-covering sequencing reads for normalization. Observed diversity is a total number of unique clonotypes in a sample, so it takes into account all clonotypes. Chao1 depends mostly on the representation of singletons and doubletons—clonotypes represented by one and two reads, respectively. Both Chao1 and Efron-Thisted indices estimate relative total TCR diversity, similar to the estimation of species richness. These numbers should not be, of course, understood as true total diversity, but only as a lower bound calculated based on a given sample size. Inverse Simpson's Diversity Index takes into account both richness and evenness. [1-normalized Shannon Wiener] represents "clonality" metrics used in Tumeh et al. (6) and largely reflects the presence of large clonal expansions.

tumors. Tumors were excised and carefully cleaned from the outer fibrous capsule, which contains a large amount of immune cells, in order to focus on lymphocytes that have truly infiltrated the tumor tissue. CD4$^+$ T cells or CD19$^+$ B cells were sorted from the isolated tumor tissues, and all extracted RNA materials and all obtained cDNA were used to generate cDNA libraries using the Clontech Smart-Seq v4 Ultra Low Input RNA Kit. These were sequenced on an Illumina HiSeq (100+100-nt paired-end reads), and TCRβ or IGH CDR3 repertoires were extracted using the MiXCR software. **Table 1** shows typical counts of cells, obtained RNA-Seq reads, CDR3 reads, and CDR3 clonotypes extracted from RNA-Seq data for the sorted T and B cells.

To verify if the RNA-Seq-based repertoire analysis can be informative in terms of repertoire diversity and clonality, we extracted TCRα and TCRβ CDR3 repertoires from transcriptomes obtained for the human naïve and effector CD8+ T cells reported in Simoni et al. (32). For each sample, we extracted from 399 to 1,255 distinct TCRα and from 795 to 1,579 TCRβ CDR3 clonotypes. Due to the limited repertoire information extracted from some of the samples, we opted to analyze joint TCRα and β repertoires in order to get better averaged statistics for diversity metrics. For normalization (33), we down-sampled our clonesets by extracting an equal number of 1,500 randomly chosen CDR3-covering reads from each set of clones. Diversity and clonality metrics were analyzed using the VDJtools software (34); results are shown in **Figure 1**. This analysis clearly distinguished naïve and effector CD8+ subsets, as expected.

We next applied the same approach to the CD4$^+$ and CD8$^+$ sorted T cell RNA-Seq data obtained by Mittal and colleagues from their syngeneic mouse model of lung cancer (28). We extracted joint TCRα and β CDR3 repertoires from these data using the MiXCR RNA-Seq mode. These samples were divided into four groups: (1) controls treated with IgG2a; (2) early response to anti-PD-1 therapy, with therapy on days 7, 10, and 13 and tumors excised at day 14 post-implantation; (3) late response with regression, with therapy on days 7, 10, 13, and 16 and tumors excised at day 17 or 24 post-implantation; and (4) late response with progression, with the same treatment and excision regimen. Note that the early response group was not split into regression and progression due to the insufficient number of usable samples that would have sufficient depth of TCR repertoire analysis in terms of CDR3-covering reads.

Similar to the one described above, we analyzed joint TCRα and β repertoires, down-sampled by extracting an equal number of 500 randomly chosen CDR3-covering reads from each set of clones. Datasets with fewer than 500 reads were discarded.

Our comparison of the resulting diversity and clonality metrics yielded several findings (**Figure 2**, **Supplementary Figure 1**). First, the diversity of TCR repertoires in anti-PD-1-treated animals decreased at early time points compared to controls treated by IgG2a, both in the CD4$^+$ and CD8$^+$ subsets. Additionally, although the limited number of samples did not allow us to estimate statistically meaningful differences between responders and progressors at the early time point (day 14), there was a trend toward more reduced CD4$^+$ diversity among the responders. At the later time points (day 17 or 24), repertoire diversity was restored in both progressors and responders, and in both CD4$^+$ and CD8$^+$ subsets. However, the latter effect was much stronger for progressors, leading to statistically significant differences in repertoire diversity metrics between responders and progressors at the late time point for both T cell subsets.

These data complement the results obtained by Mittal and colleagues as well as the current knowledge on the association between TCR clonality and response to anti-PD-1 therapy. Sustained response is associated with stably increased clonality and decreased diversity in both CD4$^+$ and CD8$^+$ subsets over the course of therapy, while restoration of initial diversity seems to be associated with disease progression.

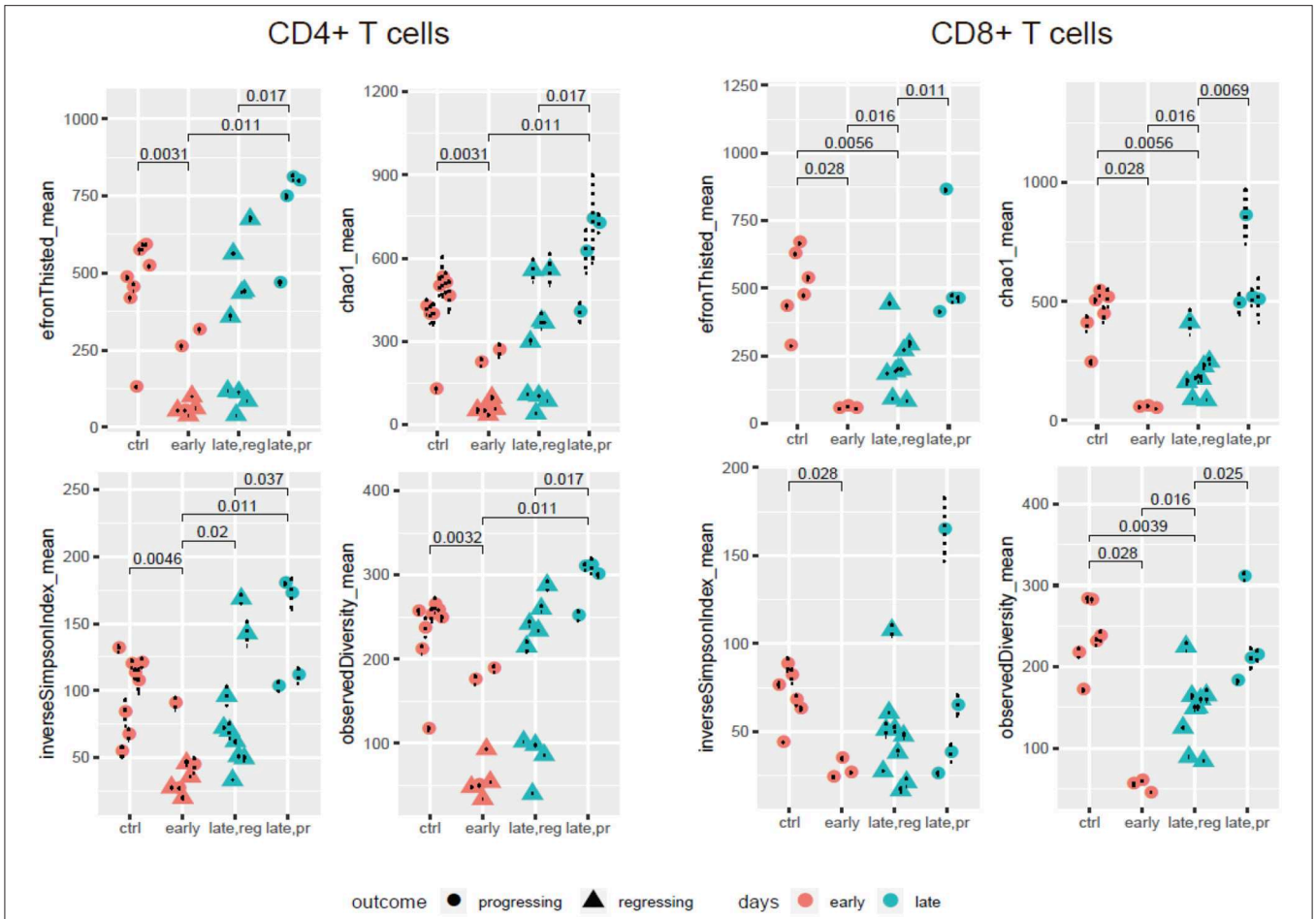


FIGURE 2 | TCR repertoire diversity in the course of anti-PD-1 therapy. Diversity metrics calculated for $CD4^+$ and $CD8^+$ TCR $\alpha + \beta$ (combined) CDR3 repertoires extracted from the four groups of samples: control; early response (day 14); late response, regressing (days 17 and 24); and late response, progressing (days 17 and 24). Each sample is represented by 500 randomly chosen CDR3-covering sequencing reads, for normalization.

One possible interpretation is that initial proliferation of previously suppressed $PD\text{-}1^+$ T cell clones specific to locally present immunogenic antigens—including both tumor and non-tumor (e.g., viral) antigens (32, 35)—in response to anti-PD1 therapy leads to temporarily increased clonality and decreased diversity. Notably, it has been shown that increased intratumoral TCR clonality in adenocarcinoma is uniquely correlated with $CD8^+PD\text{-}1^+$ T cell subsets, but not with bulk $CD8^+$ T cells (36). The same observation was noted in targeted analysis of $CD8^+PD1^+$ T cell subsets in NSCLC after anti-PD-1 treatment. Only intratumoral $CD8^+$ T lymphocyte populations with high PD-1 expression levels were characterized as having significantly increased clonality (37). Thus, the pool of oligoclonal $PD\text{-}1^+$ T cells could serve as a source for rapidly growing clonality in response to anti-PD-1 therapy.

On the other hand, some studies show that PD-1 blockade may unleash a novel tumor-specific TCR repertoire that was not previously observed in the tumor (38). Experiments tracking tumor-infiltrating TCR clones from patients with basal cell carcinoma and squamous cell carcinoma have demonstrated that clonal expansions of memory $CD8^+$ T cells with an exhausted phenotype referred to chronical activation before and after anti-PD-1 therapy are distinct from each other. The authors called this phenomenon "clonal replacement" and suggested that exhaustion of tumor-infiltrating T cells limits their renewal following checkpoint blockade. They further proposed that the T cell response relies on the expansion of novel tumor-specific T cell clones originating from non-tumor sites such as lymphoid organs or rare, unexpanded clones present within the tumor (38).

The association of a prolonged response to treatment with the stability of clonal expansions acquired during initial response, as we have observed in this work, makes intuitive sense and may be interpreted as evidence that a sustained immune response results in the stable presence and proliferation of tumor-specific $CD4^+$ and $CD8^+$ T cells.

Our results support the concept that monitoring intratumoral T cell clonality—for example, by measuring T cell clonality in excised tumor samples after neoadjuvant anti-PD-1/PD-L1 therapy (6, 11, 12, 39)—is a rational strategy for predicting long-term response.

Technically, we advocate for performing RNA-Seq on sorted T and B cell subsets as a means to simultaneously evaluate

both functional behavior and repertoire features in response to immunotherapy. MiXCR currently provides the most efficient computational solution for such analyses, either as an offline tool (https://mixcr.milaboratory.com) or as an Illumina BaseSpace cloud application (https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/milaboratory-mixcr-immune-repertoire-analyzer.html).

METHODS

Mouse Melanoma B16F0 Tumor Model

Experiments were carried out on C57BL/6-Foxp3^{eGFP} mice kindly provided by Prof. A. Rudensky (Memorial Sloan-Kettering Cancer Center) (40). Tumors were generated by subcutaneous (s.c.) injection of 10^5 B16F0 cancer cells in 300 μl PBS into the left flank of 5–7-month-old female mice. These tumor cells were initially grown in a DMEM medium supplemented with 10% fetal bovine serum (FBS), 0.06% L-glutamine, 50 units/ml penicillin, and 50 μg/ml streptomycin. Cells were cultured in an incubator at 37°C and 5% CO_2 and passaged two or three times per week. Before injection, cells were detached by trypsin, then counted and resuspended in PBS to a final concentration of 10^6 cells in 3 ml. Mice with tumor diameters ranging from 0.5 to 1.2 cm were sacrificed with isoflurane (Esteve, Italy), after which tumors were excised and prepared for further analysis. All animal experiments were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH Publications No. 8023, revised 1978). The experimental protocol was approved by the Ethical Committee of the Privolzhsky Research Medical University Academy, Russia (EC #6, granted April 17, 2019).

Mouse Melanoma Resection and Lymphocyte Isolation

Tumors were excised and cleaned from the outer tumor capsule. For lymphocyte isolation, excised tumor nodules or tumor parts were homogenized with a gentleMACS dissociator (Miltenyi Biotec, Germany) and incubated in 1–2 ml dissociation solution [RPMI medium supplemented with 417 μg/ml Liberase TL (Roche, Germany) and 10 μg/ml DNase I (Roche, Germany) for 30 min at 37°C in a shaker]. After dissociation, cell suspensions were passed through a 70-μm cell strainer and washed twice with 5 ml of an incubation buffer (PBS, pH 7.2 containing 0.5% bovine serum albumin and 2 mM EDTA).

Cell pellets were resuspended in 100 μl of an incubation buffer with the following antibodies (2 μl each): CD45-PerCP/Cy5.5 (Clone 30-F11, BioLegend), CD3-APC (Clone 145-2C11, BioLegend), CD4-V450 (Clone RM4-5, BD Biosciences), and CD19-PE/Cy7 (Clone 6D5, BioLegend); 400 μl of incubation buffer was added after 45–120 min staining at 4°C. $CD3^+CD4^+$ (excluding eGFP-positive T_{reg} cells) and $CD19^+$ subsets were sorted with a FACSAria III cell sorter (BD Biosciences) using the 85-μm nozzle directly into 200 μl of an RLT cell lysis buffer (Qiagen). After sorting, the samples were vortexed and then left at room temperature for 10 min to ensure cell lysis, and finally stored at −20°C.

Mouse Model of Non-small-cell Lung Cancer

Mittal and colleagues utilized an immunocompetent, syngeneic preclinical model of early-stage non-small-cell lung cancer (NSCLC), which has been shown to exhibit histological similarities to human adenocarcinoma (28). Mice were injected with 250 μg of a rat monoclonal blocking anti–PD-1 antibody or IgG2a intraperitoneally on days 7, 10, 13, and 16 after tumor implantation. Mice were sacrificed on day 14, day 17, or day 24 depending on the group. RNA was extracted from sorted tumor-infiltrating $CD4^+$ and $CD8^+$ T cells. RNA-Seq libraries were prepared with an Illumina TruSeq RNA Sample Preparation kit and sequenced on an Illumina HiSeq4000 with single-end 50-bp reads, 8 samples per lane. Samples were grouped based on the day of tumor excision (day 14—early tumors, days 17 and 24—late tumors) and tumor growth (progressing, regressing, or partially regressing). For reference, see Figure 7A in Markowitz et al. (28). There were 31 samples of intratumoral $CD8^+$ and $CD4^+$ T cells in total (**Supplementary Table 1**). Joint TCRα and β CDR3 repertoires were extracted from raw fastq files using MiXCR v2.1.11 (https://github.com/milaboratory/mixcr/releases/tag/v2.1.11).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The RNA sequencing data are available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/gds) under the accession number GSE114300.

ETHICS STATEMENT

All animal experiments were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH Publications No. 8023, revised 1978). The experimental protocol was approved by the Ethical Committee of the Privolzhsky Research Medical University Academy, Russia (EC #6, granted April 17, 2019).

AUTHOR CONTRIBUTIONS

EZh, MT, and DC worked on data analysis. AI, DY, LV, IS, and GS worked with mouse tumors and library preparation. ES, MT, EZa, DC, and GS wrote the manuscript. EZa and DC supervised the study.

FUNDING

The work was supported by grant of the Ministry of Education and Science of the Russian Federation (14.W03.31.0005).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00385/full#supplementary-material

REFERENCES

1. Pages F, Kirilovsky A, Mlecnik B, Asslaber M, Tosolini M, Bindea G, et al. *In situ* cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol*. (2009) 27:5944–51. doi: 10.1200/JCO.2008.19.6147
2. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol*. (2011) 29:1949–55. doi: 10.1200/JCO.2010.30.5037
3. Talebian Yazdi M, Van Riet S, Van Schadewijk A, Fiocco M, Van Hall T, Taube C, et al. The positive prognostic effect of stromal CD8+ tumor-infiltrating T cells is restrained by the expression of HLA-E in non-small cell lung carcinoma. *Oncotarget*. (2016) 7:3477–88. doi: 10.18632/oncotarget.6506
4. Ovarian Tumor Tissue Analysis C, Goode EL, Block MS, Kalli KR, Vierkant RA, Chen W, et al. Dose-response association of CD8+ tumor-infiltrating lymphocytes and survival time in high-grade serous ovarian cancer. *JAMA Oncol*. (2017) 3:e173290. doi: 10.1001/jamaoncol.2017.3290
5. Scheper W, Kelderman S, Fanchi LF, Linnemann C, Bendle G, De Rooij M, et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat Med*. (2019) 25:89–94. doi: 10.1038/s41591-018-0266-5
6. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. (2014) 515:568–71. doi: 10.1038/nature13954
7. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. (2017) 547:89–93. doi: 10.1038/nature22383
8. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature*. (2017) 547:94–8. doi: 10.1038/nature22976
9. Sidhom JW, Bessell CA, Havel JJ, Kosmides A, Chan TA, Schneck JP. ImmunoMap: a Bioinformatics Tool for T-cell Repertoire Analysis. *Cancer Immunol Res*. (2018) 6:151–62. doi: 10.1158/2326-6066.CIR-17-0114
10. Pogorelyy MV, Minervina AA, Shugay M, Chudakov DM, Lebedev YB, Mora T, et al. Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol*. (2019) 17:e3000314. doi: 10.1371/journal.pbio.3000314
11. Tamura K, Hazama S, Yamaguchi R, Imoto S, Takenouchi H, Inoue Y, et al. Characterization of the T cell repertoire by deep T cell receptor sequencing in tissues and blood from patients with advanced colorectal cancer. *Oncol Lett*. (2016) 11:3643–9. doi: 10.3892/ol.2016.4465
12. Tarhini A, Lin Y, Lin H, Rahman Z, Vallabhaneni P, Mendiratta P, et al. Neoadjuvant ipilimumab (3 mg/kg or 10 mg/kg) and high dose IFN-alpha2b in locally/regionally advanced melanoma: safety, efficacy and impact on T-cell repertoire. *J Immunother Cancer*. (2018) 6:112. doi: 10.1186/s40425-018-0428-5
13. Hogan SA, Courtier A, Cheng PF, Jaberg-Bentele NF, Goldinger SM, Manuel M, et al. Peripheral blood TCR repertoire profiling may facilitate patient stratification for immunotherapy against melanoma. *Cancer Immunol Res*. (2019) 7:77–85. doi: 10.1158/2326-6066.CIR-18-0136
14. Snyder A, Nathanson T, Funt SA, Ahuja A, Buros Novik J, Hellmann MD, et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. *PLoS Med*. (2017) 14:e1002309. doi: 10.1371/journal.pmed.1002309
15. Funt S, Charen AS, Yusko E, Vignali M, Benzeno S, Boyd ME, et al. Correlation of peripheral and intratumoral T-cell receptor (TCR) clonality with clinical outcomes in patients with metastatic urothelial cancer (mUC) treated with atezolizumab. *J Clin Oncol*. (2016) 34:3005. doi: 10.1200/JCO.2016.34.15_suppl.3005
16. Hopkins AC, Yarchoan M, Durham JN, Yusko EC, Rytlewski JA, Robins HS, et al. T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma. *JCI Insight*. (2018) 3:e122092. doi: 10.1172/jci.insight.122092
17. Kedl RM, Schaefer BC, Kappler JW, Marrack P. T cells down-modulate peptide-MHC complexes on APCs *in vivo*. *Nat Immunol*. (2002) 3:27–32. doi: 10.1038/ni742
18. Qureshi OS, Zheng Y, Nakamura K, Attridge K, Manzotti C, Schmidt EM. Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. *Science*. (2011) 332:600–3. doi: 10.1126/science.1202947
19. Akkaya B, Oya Y, Akkaya M, Al Souz J, Holstein AH, Kamenyeva O, et al. Regulatory T cells mediate specific suppression by depleting peptide-MHC class II from dendritic cells. *Nat Immunol*. (2019) 20:218. doi: 10.1038/s41590-018-0280-2
20. Ovcinnikovs V, Ross EM, Petersone L, Edner NM, Heuts F, Ntavli E, et al. CTLA-4-mediated transendocytosis of costimulatory molecules primarily targets migratory dendritic cells. *Sci Immunol*. (2019) 4:eaaw0902. doi: 10.1126/sciimmunol.aaw0902
21. Robert L, Tsoi J, Wang X, Emerson R, Homet B, Chodon T, et al. CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin Cancer Res*. (2014) 20:2424–32. doi: 10.1158/1078-0432.CCR-13-2648
22. Cha E, Klinger M, Hou Y, Cummings C, Ribas A, Faham M, et al. Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci Transl Med*. (2014) 6:238ra270. doi: 10.1126/scitranslmed.3008211
23. Subudhi SK, Aparicio A, Gao J, Zurita AJ, Araujo JC, Logothetis CJ, et al. Clonal expansion of CD8 T cells in the systemic circulation precedes development of ipilimumab-induced toxicities. *Proc Natl Acad Sci USA*. (2016) 113:11919–24. doi: 10.1073/pnas.1611421113
24. Cooper ZA, Frederick DT, Juneja VR, Sullivan RJ, Lawrence DP, Piris A. BRAF inhibition is associated with increased clonality in tumor-infiltrating lymphocytes. *Oncoimmunology*. (2013) 2:e26615. doi: 10.4161/onci.26615
25. Hosoi A, Takeda K, Nagaoka K, Iino T, Matsushita H, Ueha S, et al. Increased diversity with reduced "diversity evenness" of tumor infiltrating T-cells for the successful cancer immunotherapy. *Sci Rep*. (2018) 8:1058. doi: 10.1038/s41598-018-19548-y
26. Palakurthi S, Kuraguchi M, Zacharek SJ, Zudaire E, Huang W, Bonal DM, et al. The combined effect of FGFR inhibition and PD-1 blockade promotes tumor-intrinsic induction of antitumor immunity. *Cancer Immunol Res*. (2019) 7:1457–71. doi: 10.1158/2326-6066.CIR-18-0595
27. Choi H, Sheng J, Gao D, Li F, Durrans A, Ryu S, et al. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep*. (2015) 10:1187–201. doi: 10.1016/j.celrep.2015.01.040
28. Markowitz GJ, Havel LS, Crowley MJ, Ban Y, Lee SB, Thalappillil JS, et al. Immune reprogramming via PD-1 inhibition enhances early-stage lung cancer survival. *JCI Insight*. (2018) 3:e96836. doi: 10.1172/jci.insight.96836
29. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. (2015) 12:380–1. doi: 10.1038/nmeth.3364
30. Bolotin DA, Poslavsky S, Davydov AN, Frenkel FE, Fanchi L, Zolotareva OI, et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol*. (2017) 35:908–11. doi: 10.1038/nbt.3979
31. Bolotin DA, Poslavsky S, Davydov AN, Chudakov DM. Reply to "Evaluation of immune repertoire inference methods from RNA-seq data". *Nat Biotechnol*. (2018) 36:1035–6. doi: 10.1038/nbt.4296
32. Simoni Y, Becht E, Fehlings M, Loh CY, Koo SL, Teng KWW, et al. Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature*. (2018) 557:575–9. doi: 10.1038/s41586-018-0130-2
33. Izraelson M, Nakonechnaya TO, Moltedo B, Egorov ES, Kasatskaya SA, Putintseva EV, et al. Comparative analysis of murine T-cell receptor repertoires. *Immunology*. (2018) 153:133–44. doi: 10.1111/imm.12857
34. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput Biol*. (2015) 11:e1004503. doi: 10.1371/journal.pcbi.1004503
35. Kvistborg P, Shu CJ, Heemskerk B, Fankhauser M, Thrue CA, Toebes M, et al. TIL therapy broadens the tumor-reactive CD8(+) T cell compartment in melanoma patients. *Oncoimmunology*. (2012) 1:409–18. doi: 10.4161/onci.18851
36. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell*. (2017) 169:750–65 e717. doi: 10.1016/j.cell.2017.04.014

37. Thommen DS, Koelzer VH, Herzig P, Roller A, Trefny M, Dimeloe S. A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat Med*. (2018) 24:994–1004. doi: 10.1038/s41591-018-0057-z
38. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med*. (2019) 25:1251–9. doi: 10.1038/s41591-019-0522-3
39. Forde PM, Chaft JE, Smith KN, Anagnostou V, Cottrell TR, Hellmann MD, et al. Neoadjuvant PD-1 blockade in resectable lung cancer. *N Engl J Med*. (2018) 378:1976–86. doi: 10.1056/NEJMoa1716078
40. Fontenot JD, Rasmussen JP, Williams LM, Dooley JL, Farr AG, Rudensky AY. Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity*. (2005) 22:329–41. doi: 10.1016/j.immuni.2005.01.016

Conflict of Interest: DC is a co-founder of MiLaboratory LLC, which develops MiXCR software and has exclusive rights for its commercial distribution.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 May 2020
doi: 10.3389/fonc.2020.00560

Clinical Multigene Panel Sequencing Identifies Distinct Mutational Association Patterns in Metastatic Colorectal Cancer

*Francesca Belardinilli[1†], Carlo Capalbo[1†], Umberto Malapelle[2], Pasquale Pisapia[2], Domenico Raimondo[1], Edoardo Milanetti[3], Mahdavian Yasaman[1], Carlotta Liccardi[1], Paola Paci[4], Pasquale Sibilio[4], Francesco Pepe[2], Caterina Bonfiglio[5], Silvia Mezi[6], Valentina Magri[7], Anna Coppa[8], Arianna Nicolussi[8], Angela Gradilone[1], Marialaura Petroni[9], Stefano Di Giulio[1], Francesca Fabretti[1], Paola Infante[9], Sonia Coni[1], Gianluca Canettieri[1,10], Giancarlo Troncone[2] and Giuseppe Giannini[1,10]**

[1] Department of Molecular Medicine, University La Sapienza, Rome, Italy, [2] Department of Public Health, University Federico II, Naples, Italy, [3] Department of Physics, University La Sapienza, Rome, Italy, [4] Institute for System Analysis and Computer Science "Antonio Ruberti", National Research Council, Rome, Italy, [5] National Institute of Gastroenterology-Research Hospital, IRCCS "S. de Bellis", Bari, Italy, [6] Department of Radiological Oncological and Pathological Sciences, University La Sapienza, Rome, Italy, [7] Department of Surgery Pietro Valdoni, Faculty of Medicine and Dentistry, Sapienza University of Rome, Rome, Italy, [8] Department of Experimental Medicine, University La Sapienza, Rome, Italy, [9] Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Rome, Italy, [10] Pasteur Institute-Cenci Bolognetti Foundation, Rome, Italy

Extensive molecular characterization of human colorectal cancer (CRC) via Next Generation Sequencing (NGS) indicated that genetic or epigenetic dysregulation of a relevant, but limited, number of molecular pathways typically occurs in this tumor. The molecular picture of the disease is significantly complicated by the frequent occurrence of individually rare genetic aberrations, which expand tumor heterogeneity. Inter- and intratumor molecular heterogeneity is very likely responsible for the remarkable individual variability in the response to conventional and target-driven first-line therapies, in metastatic CRC (mCRC) patients, whose median overall survival remains unsatisfactory. Implementation of an extensive molecular characterization of mCRC in the clinical routine does not yet appear feasible on a large scale, while multigene panel sequencing of most commonly mutated oncogene/oncosuppressor hotspots is more easily achievable. Here, we report that clinical multigene panel sequencing performed for anti-EGFR therapy predictive purposes in 639 formalin-fixed paraffin-embedded (FFPE) mCRC specimens revealed previously unknown pairwise mutation associations and a high proportion of cases carrying actionable gene mutations. Most importantly, a simple principal component analysis directed the delineation of a new molecular stratification of mCRC patients in eight groups characterized by non-random, specific mutational association patterns (MAPs), aggregating samples with similar biology. These data were validated on a The Cancer Genome Atlas (TCGA) CRC dataset. The proposed stratification may provide great opportunities to direct more informed therapeutic decisions in the majority of mCRC cases.

Keywords: mCRC, NGS, molecular stratification, mutation, genes

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Valentina Silvestri,
Sapienza University of Rome, Italy
Alexandre Harle,
Institut de Cancérologie de Lorraine, France
Hongxing Liu,
Lu Daopei Hospital, China

***Correspondence:**
Giuseppe Giannini
giuseppe.giannini@uniroma1.it

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *02 October 2019*
Accepted: *27 March 2020*
Published: *07 May 2020*

Citation:
Belardinilli F, Capalbo C, Malapelle U, Pisapia P, Raimondo D, Milanetti E, Yasaman M, Liccardi C, Paci P, Sibilio P, Pepe F, Bonfiglio C, Mezi S, Magri V, Coppa A, Nicolussi A, Gradilone A, Petroni M, Di Giulio S, Fabretti F, Infante P, Coni S, Canettieri G, Troncone G and Giannini G (2020) Clinical Multigene Panel Sequencing Identifies Distinct Mutational Association Patterns in Metastatic Colorectal Cancer. Front. Oncol. 10:560. doi: 10.3389/fonc.2020.00560

INTRODUCTION

Colorectal carcinoma (CRC) is one of the most commonly diagnosed cancers worldwide (1, 2). A large proportion of patients develop distant metastasis, which contributes to the high mortality reported for this tumor. With the current standard approaches, the 5-year survival rate for metastatic CRC (mCRC) is about 13% (1–3). These oxaliplatin or irinotecan-based chemotherapy regimens allow a median overall survival (OS) of about 18–20 months (4, 5). Survival rates can be significantly improved by a "triplet" approach consisting of 5-FU, oxaliplatin, and irinotecan chemotherapy (6) and/or by the addition of targeted drugs, such as monoclonal antibodies directed against angiogenesis or EGFR pathway (7). Nonetheless, median OS for mCRC rarely exceeds 30–36 months (8–10). Unfortunately, individual responses to these therapeutic approaches may be dramatically different from patient to patient reflecting the broad inter- and intratumor molecular heterogeneity.

Historically, CRC represented the first model for multistep cancer evolution in which discrete and sequential genetic modifications in specific oncogenes and tumor-suppressor genes occur throughout cancer progression (11, 12). Next Generation Sequencing (NGS) provided significant advances in understanding the molecular basis of CRC (13–15) and indicated that genetic or epigenetic dysregulation of a relevant, but limited, number of molecular pathways typically occurs in human CRC (13, 15, 16). This molecular picture is complicated by the frequent occurrence of individually rare genetic aberrations, which further expand tumor heterogeneity (13–15).

Reflecting the different biology of CRCs, Guinney et al. recently proposed a molecular classification in four consensus molecular subtypes (CMS): CMS1-MSI immune, CMS2-canonical, CMS3-metabolic, and CMS4-mesenchymal (13). Although this might have implications for prognostication and therapy decisions, its immediate transfer to routine diagnostic/clinical settings is seriously challenging in terms of methodology, turnaround time, costs, and mindset. In fact, despite NGS and other omic approaches may disclose a huge amount of molecular details, still very few of them have yet acquired clinical relevance. In example, the use of anti-EGFR therapy is essentially dictated by the RAS (KRAS+NRAS) wild type status, in the clinical routine (17, 18), which however is largely insufficient for the positive selection of responsive patients (19, 20). Treatment with anti-VEGF antibodies is not driven by specific selection criteria due to lack of validated molecular biomarkers (21, 22). Other targeted approaches (i.e., BRAF or PI3K inhibitors used as single agents) failed due to resistance mechanisms (23). These evidences support the need for a paradigm shift in personalized medicine, as suggested by Dienstmann et al. (24): from a one-gene one-drug approach, to a multi-gene multi-drug perspective.

The use of multigene panel sequencing has been recently validated for clinical applications. In example, we introduced a 22 multigene panel sequencing, which includes the clinically relevant RAS and BRAF hotspots, as a routine for the predictive selection of mCRC patients to be subjected to anti-EGFR therapy (25–32). This implementation allowed us to accumulate a large dataset to ask the question of whether application of multigene panel sequencing to the standard diagnostics of mCRC could provide clinically useful information, with no extra-costs in terms of turnaround time and money.

On the basis of results obtained on 639 formalin-fixed and paraffin-embedded (FFPE) tumor samples, here we report that clinical genomic profiling with a multigene panel identifies distinct molecular association patterns (MAPs) and provides great opportunities to unveil co-occurrence of actionable gene mutations to direct more appropriate therapeutic decisions for the majority of mCRC patients.

PATIENTS AND METHODS

Specimen Collection

A total of 779 FFPE tumor samples from mCRC patients were collected from Policlinico Umberto I (Rome, Italy) and from the Department of Public Health, University Federico II, Naples, Italy. The large majority of samples (696/779) were from the primary site, while few (83/779) were from metastatic sites. All samples reached the molecular pathology labs with a medical prescription for determination of RAS/BRAF mutation status for predictive purposes. As such, only scattered clinical-pathological information was available for the two series. For this retrospective observational study all investigations were approved by the Ethics Committee of the University La Sapienza (Prot.: 88/18; RIF.CE:4903, 31-01-2018). All information regarding human material included in the study was managed using anonymous numerical codes, and all samples were handled in compliance with the principles outlined in the declaration of Helsinki. For samples collected at the Department of Public Health, University Federico II, we obtained written informed consent from all patients, in accordance with the general authorization to process personal data for scientific research purposes from "The Italian Data Protection Authority (http://www.garanteprivacy.it/web/guest/home/docweb/-/docwebdisplay/export/2485392).

DNA Extraction

Tissue samples with a content of tumor-vs.-non-tumor cells below 20% (evaluated at the observation of Hematoxylin and Eosin stained slides) were excluded from the analysis (33). The tumor area was macroscopically dissected to concentrate tumor tissue. Xylene was added once and ethanol was added twice to remove all paraffin from the tissue sample (34). DNA was extracted using QIAamp DNA FFPE Tissue kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. Eluted DNA was quantified with Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Van Allen Way, Carlasbad, CA 92008, USA) using Qubit™ dsDNA HS Assay Kit (Thermo-Fisher Scientific, Eugene, Oregon 96492, USA).

IT-PGM Sequencing and Variant Calling

IT-PGM sequencing was achieved as described (25, 27, 35). Approximately, 10 ng of DNA samples was required to construct barcoded and adaptor-ligated libraries using the Ion AmpliSeq Library kit 2.0 (Thermo Fisher Scientific, Van Allen Way,

Carlsbad, CA 92008 USA) and Ion Xpress Barcode Adapter 1-16 Kit (Thermo Fisher Scientific, Van Allen Way, Carlsbad, CA 92008 USA). The samples were analyzed using Ion AmpliSeq Colon and Lung Cancer Research Panel V2 (CLV2, Thermo Fisher Scientific, Guilford, CT 06437, USA) containing a single primer pool to amplify hotspots and targeted regions of 22 cancer genes frequently mutated in CRCs and NSCLCs (29). Templated spheres were prepared using 100 pM of each library using the Ion One Touch 2.0 machine (Thermo Fisher Scientific, Van Allen Way, Carlsbad, CA 92008 USA). Template-positive spheres were loaded into Ion chip 314 or Ion chip 316 and sequenced by IT-PGM machine (Thermo Fisher Scientific, Van Allen Way, Carlsbad, CA 92008 USA). Sequencing data were analyzed with the Ion Torrent Suite (Thermo Fisher Scientific, http://github.com/iontorrent/TS). Variants with a quality <30 were filtered out.

For the purpose of the study, we generated a mutational data set including only samples carrying mutations of established clinical relevance for KRAS (mutations at codon 12, 13, 59, 61, 117, and 146), BRAF (V600E) and PIK3CA (mutations in exon 10 and 21). For TP53, we included in the study samples carrying mutations with defined pathogenic significance according to ClinVar and/or well-established hotspot mutations. We excluded from the study 140 samples carrying variants of unknown clinical significance (VUS) in these genes. For all other genes, we listed all genetic alterations described as pathogenic, likely-pathogenic or predicted deleterious by *in silico* analysis, while benign polymorphisms were not considered.

When appropriate, PolyPhen-2 (**Poly**morphism **Phen**otyping v2; http://genetics.bwh.harvard.edu/pph2/), PROVEAN/SIFT (Sort Intolerant From Tolerant Subsitutions) http://provean.jcvi.org/protein_batch_submit.php?species=human) computational tools were used to predict the possible impact of the detected alterations on the structure and function of the protein (18, 19).

The reference sequence used are: KRAS NM_033360.3, TP53 NM_000546.5, PIK3CA NM_006218.3, BRAF NM_004333.4, NRAS NM_002524.4, FBXW7 NM_033632.3, SMAD4 NM_005359.5, PTEN NM_000314.6, MET NM_001127500.2, STK11 NM_000455.4, EGFR NM_005228.4, CTNNB1 NM_001904.3, AKT1 NM_001014431.1, ERBB2 NM_004448.3, ERBB4 NM_005235.2, FGFR1, NM_001174063.2, ALK NM_004304.4, MAP2K1 NM_002755.3, NOTCH1 NM_017617.4, DDR2 NM_001014796.3, FGFR3 NM_000142.4, FGFR2 NM_000141.4.

MSI Analysis

Determination of MSI status was investigated on 162 patients (72 of the 639 cases representing the main bulk of the study plus 90 additional cases collected at a later stage and analyzed separately). It was carried out by analysis of BAT25, BAT26, NR21, NR22, and NR24 mononucleotide repeats as previously described (36). Briefly, one PCR primer of each pair was labeled with 1 with either FAM, HEX, or NED fluorescent markers. PCR amplification was performed under the following conditions: denaturation at 94°C for 5 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. This was followed by an extension step at 72°C for 7 min. PCR products were run on ABI PRISM 3130xl Genetic Analyzer (16 capillary DNA sequencer, Applied Biosystem). Gene Mapper software 5 (version 5.0, Applied Biosystems, Van Allen Way, Carsvad, CA 92008, USA) was used to calculate the size of each fluorescent PCR product.

Statistical Analysis

The mutational data set was organized in a matrix composed by 20 columns and 639 rows where each row corresponds to a different sample and each column corresponds to one of 22 different genes whose mutational pattern was characterized. We performed a Principal Component Analysis (PCA) on this mutational dataset in order to classify mutational patterns based on their similarity. Each matrix element M_{ij} (where i is a generic sample and j is a generic gene) can assume the value 0 or 1 if the patient i has no mutation in the gene j or the mutation is present, respectively (37). Each principal component is a linear combination of optimally-weighted original variables, and so it is often possible to ascribe meaning to what the components represent. The statistical analysis was carried out with SPSS statistics or standard R software, version 2.13.1 (http://www.r-project.org).

Statistical analyses on gender, tumor type, tumor location, and MSI-H phenotype were performed on all cases for which appropriate information was available, using both the 639 and the 90 series.

The Pearson's Chi-square test and Fisher's exact test of association was used to determine the relationship between two categories which consist in coexistence of two mutations (pairwise association analysis). A $p < 0.05$ was considered statistically significant.

TCGA Network Data set

We downloaded gene somatic mutations for 625 patients from the TCGA data portal (https://portal.gdc.cancer.gov/) accessed December 2018 (38, 39). We cleared this dataset from samples carrying VUS, as we did for our dataset (see above). The resulting data set contained 412 patients with their mutational data of the 22 genes included in the CLV2 panel.

We employed the R package TCGAbiolinks (40) to retrieve patient's Microsatellite Instability (MSI) status from the legacy archive of GDC data portal (https://portal.gdc.cancer.gov/).

RESULTS

Mutation Profiling of mCRCs, Pairwise Associations, and Identification of Actionable Targets

Using a 22 gene panel NGS approach, we detected pathogenic mutations in at least one of the 22 targets in 523 out of 639 (81.8%) mCRC samples (**Table S1**). Mutation spectra and frequencies were in line with previous reports (14, 15, 31) (**Figure 1A**). Eleven genes displayed a mutation frequency >1.5% (mutation number >10), being *TP53* and *KRAS* the most frequently mutated genes (48.5 and 39.4%, respectively) (**Figure 1A**). Mutations occurred less frequently (<1.5%) in the other 11 genes (*CTNNB1*, *AKT1*, *ERBB2*, *ERBB4*, *FGFR1*, *ALK*,

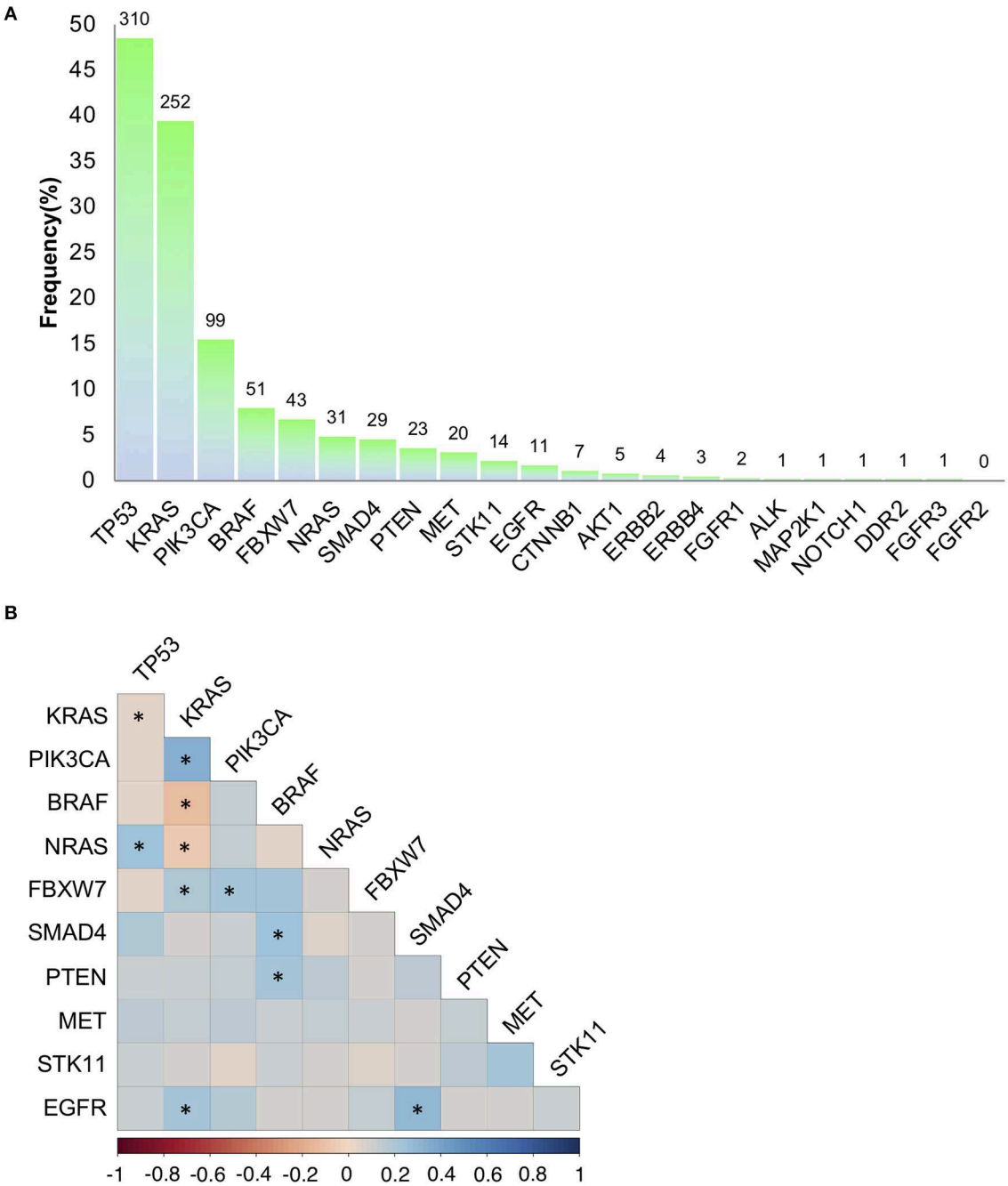


FIGURE 1 | Mutation frequencies and pairwise associations. **(A)** Mutation rates (and absolute numbers of the bars) in 639 metastatic colorectal cancers. **(B)** Correlation Plot describing pairwise association of the mutations occurring on the 11 genes with a mutation frequency >1.5% (mutation number >10). Statistical analysis is given **Table 1**. *$p < 0.05$.

MAP2K1, *NOTCH1*, *DDR2*, *FGFR3,* and *FGFR2*), consistent with the "tail effect" associated with NGS profiling of tumor samples (15) (**Figure 1A**).

To investigate on mutation associations, we initially performed a pairwise association analysis for those genes with a mutation frequency >1.5%. In agreement with previous literature, *BRAF, KRAS*, *NRAS* mutations were mutually exclusive, while *PIK3CA* and *FBXW7* mutations frequently occurred in association with *KRAS* mutations. *BRAF* and *SMAD4* mutations were associated, while *TP53* and *KRAS* mutations were negatively associated (**Figure 1B**, **Table 1**). We also revealed previously unreported positive association between

TABLE 1 | Significant pairwise associations between most frequent gene mutations.

Gene[a]	Status	Wt (%)	Mut (%)	p^b	References
		KRAS (%)			
TP53	Mut	200 (51.7)	110 (43.7)	0.047	(44)
PIK3CA	Mut	38 (9.8)	61 (24.2)	<0.001	(43)
BRAF	Mut	50 (12.9)	1 (0.4)	<0.001	(43)
NRAS	Mut	30 (7.8)	1 (0.4)	<0.001	(43)
FBXW7	Mut	20 (5.2)	23 (9.1)	0.05	(31)
EGFR	Mut	3 (0.8)	8 (3.2)	0.03*	**New**
		PIK3CA (%)			
FBXW7	Mut	31 (5.7)	12 (12.1)	0.02	(45)
		BRAF (%)			
SMAD4	Mut	23 (3.9)	6 (11.8)	0.022*	(46)
PTEN	Mut	18 (3.1)	5 (9.8)	0.03*	**New**
		EGFR (%)			
SMAD4	Mut	26 (4.1)	3 (27.3)	0.011*	**New**
		TP53 (%)			
NRAS	Mut	9 (2.7)	22 (7.1)	0.01	**New**

[a] *The genes with an overall mutational rate higher than 1.5% (number of mutations >10) were considered for statistical analysis.* [b] *Chi-squared test.*
**Fisher exact test.*

TABLE 2 | Frequency of co-mutation in genes carrying actionable mutations.

	Status	No. of pts. (%)	No. of pts. (%) with additional mutations
KRAS	WT	387 (60.6)	270 (42.2)
	Mut	252 (39.4)	176 (27.5)
NRAS	WT	608 (95.2)	491 (76.8)
	Mut	31 (4.8)	26 (4.1)
BRAF	WT	588 (92.0)	471 (73.7)
	Mut	51 (8.0)	34 (5.3)
PIK3CA	WT	540 (84.5)	423 (66.2)
	Mut	99 (15.5)	89 (13.9)
EGFR	WT	629 (98.4)	512 (80.1)
	Mut	10 (1.6)	9 (1.4)
MET	WT	633 (99.1)	516 (80.7)
	Mut	6 (0.9)	6 (0.9)
PTEN	WT	618 (96.7)	501 (78.4)
	Mut	21 (3.3)	20 (3.1)
AKT1	WT	634 (99.2)	517 (80.9)
	Mut	5 (0.8)	4 (0.6)
ERBB2	WT	636 (99.5)	519 (81.2)
	Mut	3 (0.5)	3 (0.5)
ALK	WT	638 (99.8)	521 (81.5)
	Mut	1 (0.2)	1 (0.2)
MAP2K1	WT	638 (99.8)	521 (81.5)
	Mut	1 (0.2)	1 (0.2)

EGFR mutations and *KRAS* and *SMAD4* mutations, while *BRAF* mutations were significantly associated with *PTEN* mutations (**Figure 1B**, **Table 1**, **Table S2**). At variance from *KRAS*, *NRAS* mutations were significantly associated with *TP53* mutations.

Overall, 374/639 (58.5%) patients carried actionable gene mutations, as defined by Chakravarty et al. (41), and 153 patients carried druggable alterations. Importantly, the vast majority of patients positive or negative for specific actionable mutations frequently carried additional relevant genetic alterations (**Table 2**), which in principle could contribute to an individual variability in patients' responsiveness to standard and target-driven therapies. In example, only 27 (4.2% of the entire series) of the 99 patients carrying *PIK3CA* mutations were *RAS*/*BRAF* WT and only 9 of these (1.4% of the entire series) harbored exclusively *PIK3CA* mutations. On the same line, 17/639 (2.7%) patients carried only *BRAF* mutations, while 34 BRAF mutant samples also carried additional mutations.

Identification of Mutational Association Patterns (MAPs)

Although pairwise associations might provide interesting insights into the molecular nature of CRC and represents a step forward in considering the molecular complexity of cancer for prognostic and predictive purposes, we reasoned that a more comprehensive use of the entire mutational profile of each sample could help defining a novel and more precise classification of CRC.

Thus, we subjected our large dataset to a principal component analysis (PCA) with the aim to detect those genes which better classify the different samples based on their overall mutational profile. This approach clearly indicated that two genes (*TP53* and *KRAS*) could sharply cluster our samples into four different subsets: $TP53^{wt}/KRAS^{wt}$ samples, $TP53^{mut}/KRAS^{wt}$ samples, $TP53^{wt}/KRAS^{mut}$ samples, and $TP53^{mut}/KRAS^{mut}$ samples (**Figure 2**). While mutations in other genes could also aggregate our samples into distinct subsets (see for example *PIK3CA* and *BRAF*, **Figure 2**), they never reached the sharp effectiveness of *TP53* and *KRAS* mutations.

Thus, in accordance to PCA results, we stratified the 639 CRC cases into four different mutation association patterns (MAPs) based on *TP53* and *KRAS* mutation status (**Figure 3A**). Depending on the presence/absence of mutations in genes other than *TP53* and *KRAS*, each MAP could be further divided in two subsets leading to delineation of a total of eight different MAPs (**Figure 3A**).

This stratification promptly revealed that 18.2% of the samples carried no mutations in any of the 22 gene of the panel (MAP4.2), while 29.4% harbored only one mutation in either *KRAS* or *TP53* (MAP2.2 and MAP3.2, respectively). An additional 11.6% of the patients only carried *KRAS* and *TP53* mutations with no other alterations (MAP1.2), which indicates that a large fraction of the mCRC cohort is characterized by a very low mutation rate, as detectable by our gene panel sequencing.

The distribution of mutations in genes other than *TP53* and *KRAS* also occurred non-randomly among the MAPs (**Figures 3A,B**) clearly defining distinct molecular profiles. Indeed, the Pearson's Chi-square test and Fisher's exact test showed statistical significance for almost all the comparisons between the MAPs (**Table S3**). In details, the eight MAPs are characterized as follows.

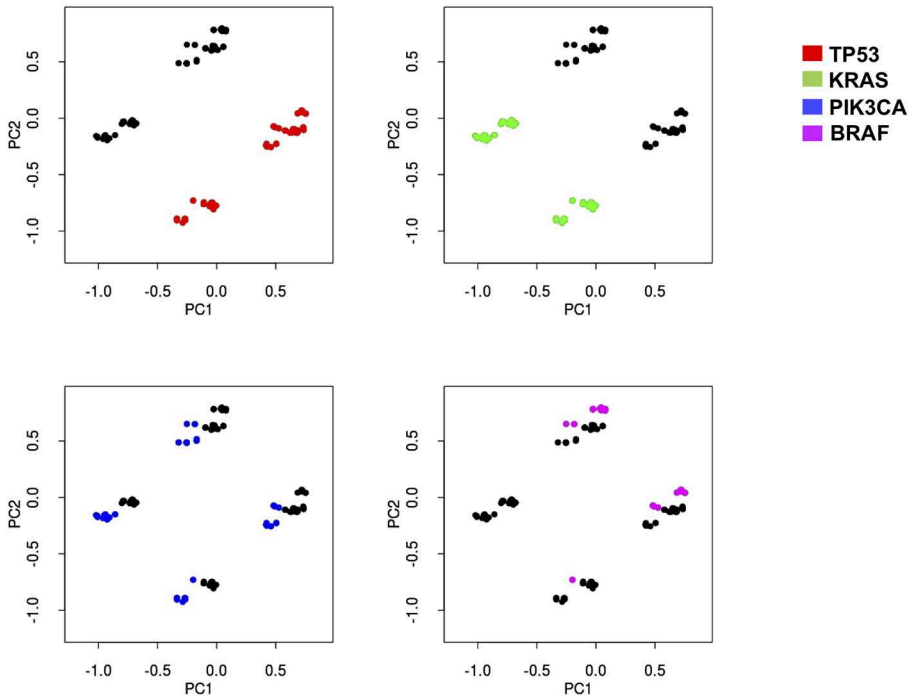


FIGURE 2 | Principal component analysis indicates that four different subsets of mCRC samples may be sharply identified based on KRAS and TP53 mutation status. Principal-component analysis of the sequencing results of 639 mCRCs indicates that the most represented genes in the first two principal components (PC) are able to better separate the data according to their variation. PC1 and PC2 contain 51% of variation in the data. KRAS, TP53, PIK3CA, and BRAF genes have been identified as the most important genes of PC1 and PC2. Each mutational profile has been projected in a two-dimensional space using the PC1 and PC2 to help appreciate sample separation. Each graph indicates how PCA analysis assembles patients (dots) in four distinct groups distinguishable in the two-dimensional space. Red dots, green dots, blue dots, and magenta dots represent samples with mutations in p53, KRAS, PIK3CA, or BRAF, respectively. While KRAS and TP53 mutations sharply map in the four distinct groups in the two dimensional-space, both PIK3CA and BRAF mutations are much less efficient in defining the identity of the four groups, thus indicating that the formers are more effective in creating sharp group separation.

MAP1.1. This MAP, characterized by the $TP53^{mut}/KRAS^{mut}$ status, showed a high rate of *PIK3CA* mutations (52.8%), rare (2.8%) $BRAF^{V600E}$ mutations and no *NRAS* alterations. We also found a relevant number of *FBXW7* mutations (27.8%), and some *PTEN* (11.1%) and *SMAD4* mutations (8.3%), most often mutually exclusive with *PIK3CA* mutations.

MAP1.2. This MAP was characterized by the $TP53^{mut}/KRAS^{mut}$ status, and no additional mutations.

MAP2.1. This MAP, characterized by $TP53^{wt}/KRAS^{mut}$ status, showed the highest frequency of *PIK3CA* mutations (62.7%). Intriguingly, 3 out of 5 E17K *AKT1* mutations occurred in *PIK3CA* WT samples in this MAP, concurring to the activation of the same pathway.

A fair amount of *FBXW7* mutations (19.4%) and a few *SMAD4* mutations (9.0%), but no *BRAF* and rare *NRAS* mutations (1.5%) occurred in MAP2.1. Coherent with the previously mentioned *KRAS* pairwise association, the rare *EGFR* mutations clustered in MAP1.1 and MAP2.1.

MAP2.2. This MAP was characterized by the $TP53^{wt}/KRAS^{mut}$ status, and no additional mutations.

MAP3.1. This MAP, characterized by $TP53^{mut}/KRAS^{wt}$ status, had a high frequency of *BRAF* (20.5%), combined with the highest frequency of *NRAS* (in a mutually exclusive way) and *SMAD4* mutations (25.0 and 18.2%, respectively). This group also showed *PIK3CA* mutations in 23.9% of the samples, at least partially non-overlapping with *BRAF*, *NRAS,* and *SMAD4* mutations, and the lowest frequency of *FBXW7* mutations (5.7%).

MAP3.2. This MAP was characterized by the $TP53^{mut}/KRAS^{wt}$ status, and no additional mutations.

MAP4.1. In this MAP, characterized by $TP53^{wt}/KRAS^{wt}$ status, we found the highest frequency of *BRAF* mutations (45.1%) and the lowest amount of *SMAD4* mutations (5.6%). It also showed mutations in *PIK3CA*, *NRAS,* and *FBXW7*, respectively, in 23.9, 11.3, and 21.1% of the samples.

MAP4.2. This MAP was characterized by absence of mutations.

Importantly, the analysis of microsatellite instability (MSI) on a test group of 162 samples revealed that 9 tumors were MSI-H. 6 out of 9 MSI-H samples clustered into MAP4.1, 2 in MAP3.1 and 1 in MAP4.2 (**Table 3**), which suggests that the proposed mutational stratification is able to aggregate samples with similar biology.

Mutation distribution of other genes did not vary significantly among MAPs and/or was too low to support major conclusions.

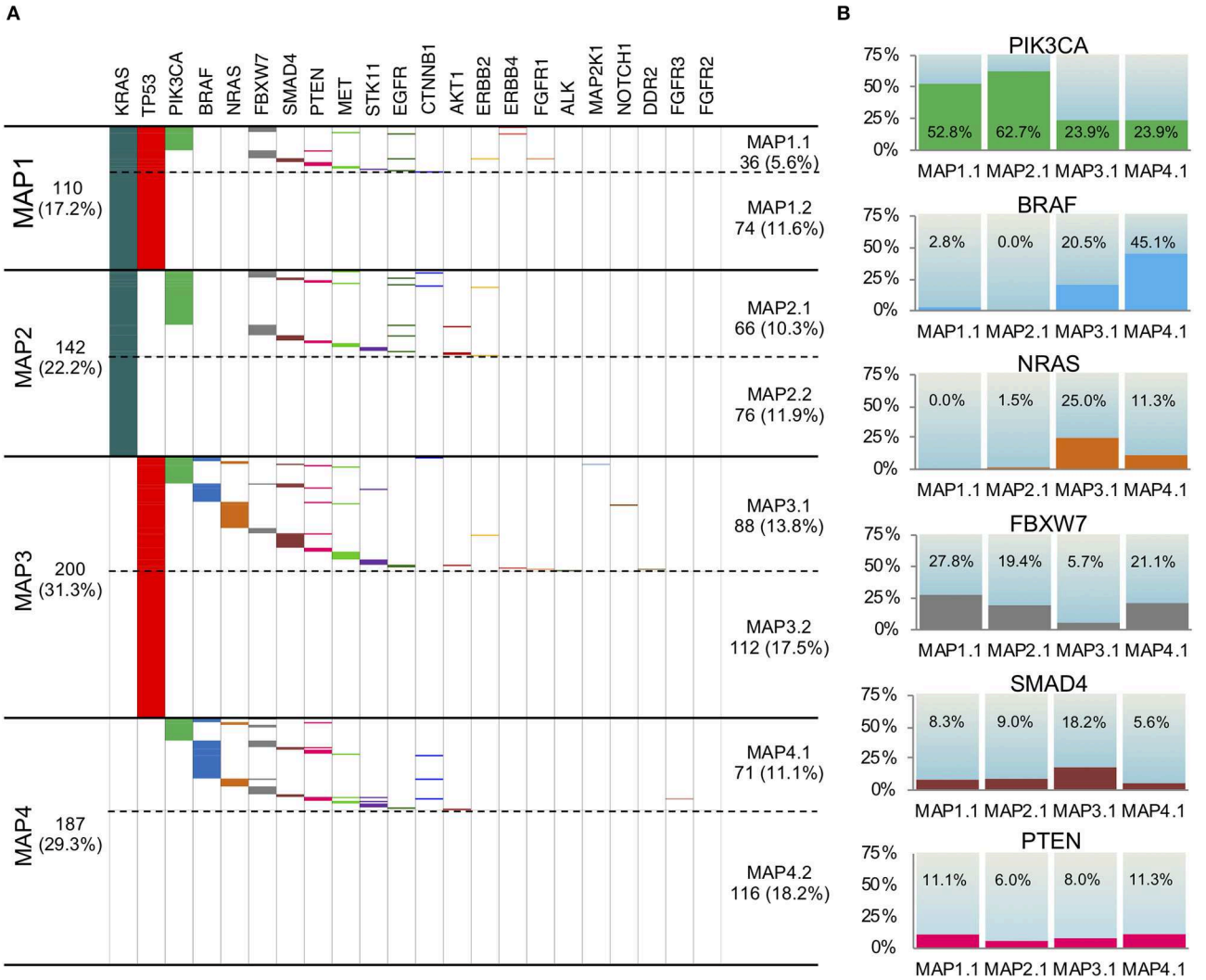


FIGURE 3 | Mutation Association Patterns of the 639 mCRC samples according to the 22 gene panel analysis. **(A)** The presence of a mutation is depicted with a specific color for each gene, while the absence is indicated in white. Four main patterns are obtained, depending on KRAS and TP53 status: MAP1, MAP2, MAP3, and MAP4. Depending on the presence/absence of mutations in genes other than TP53 and KRAS, each MAP could be further divided in two subMAPs (MAP1.1, MAP1.2, MAP 2.1, MAP 2.2, MAP 3.1, MAP 3.2, MAP 4.1, MAP 4.2). **(B)** The frequency of mutation in PIK3CA, BRAF, NRAS, FBXW7, SMAD4, and PTEN genes in the four subMAPs is reported. The Pearson's Chi-square test or Fisher's exact test were carried out and shown in **Table S3**.

Identification of MAPs on the TCGA Dataset

To validate MAPs in an external dataset, we accessed the TCGA public mutational data for CRC patients. After appropriate curation of the dataset in order to select all pathogenic mutations potentially identifiable by our multigene panel sequencing approach, we had 412 samples available for MAP stratification. Of interest, the TCGA dataset included all CRC stages, and only a minority of the cases were mCRC (**Figure S1A**), as already noted by others (14, 42).

The mutation frequencies on the 22 genes included in the CLV2 panel were largely similar between TCGA dataset and our mCRC cohort (**Figure S1B**). All different MAPs exist with rather similar rates, in TCGA dataset and our series, with MAP 4.1 and MAP4.2 representing sharp exceptions. Indeed, MAP 4.1 accounts for 11.1% of our series of mCRC samples, compared to 18.0% of the TCGA dataset, while MAP4.2 accounts for 18.2% in our series and 5.3% of the TCGA dataset. At variance from our cohort, TCGA dataset included 14% of MSI-H samples, which is consistent with its stage 1-to-stage 4 composition (14). The majority of these cases clustered in MAP4.1, possibly providing an explanation for the different MAP4 rates between the two datasets.

PIK3CA, BRAF, and NRAS mutation rates in the different MAPs display similar trends in our metastatic cohort and in the TCGA dataset (**Figures S2A,B**). We observed less consistency for the mutation rates of the less frequently mutated FBXW7, SMAD, and PTEN genes.

TABLE 3 | Associations between selected features and MAPs.

				MAPs								
		No.	**%**	**1.1**	**1.2**	**2.1**	**2.2**	**3.1**	**3.2**	**4.1**	**4.2**	***p***
Gender	M	346	60.1	29 (8.40%)	44 (12.7%)	35 (10.1%)	37 (10.7%)	44 (12.7%)	64 (18.5%)	34 (9.8%)	59 (17.1%)	0.301
	F	230	39.9	15 (6.5%)	37 (16.1%)	28 (12.2%)	26 (11.3%)	41 (17.8%)	33 (14.3%)	22 (9.6%)	28 (12.2%)	
Site	Rectum	89	15.5	9 (10.1%)	14 (15.7%)	17 (19.1%)	5 (5.6%)	9 (10.1%)	13 (14.6%)	6 (6.7%)	16 (18.0%)	0.058
	Colon	486	84.5	35 (7.2%)	67 (13.8%)	46 (9.5%)	58 (11.9%)	76 (15.6%)	84 (17.3)	50 (10.3%)	70 (14.4%)	
Side	Right	183	55.3	12 (6.6%)	31 (16.9%)	23 (12.6%)	18 (9.8%)	36 (19.7%)	15 (8.2%)	34 (18.6%)	14 (7.7%)	**<0.0001**
	Left	148	44.7	12 (8.10%)	17 (11.5%)	15 (10.1%)	19 (12.8%)	24 (16.2%)	36 (24.3%)	3 (2.0%)	22 (14.9%)	
MSI	absent	153	94.4	19 (12.4%)	31 (20.3%)	20 (13.1%)	14 (9.2%)	25 (16.3%)	24 (15.7%)	3 (2.0%)	17 (11.1%)	***<0.0001**
	present	9	5.6	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (22.2%)	0 (0.0%)	6 (66.7%)	1 (11.1%)	

**Fisher's exact test. Bold: statistically significant.*

Correlation Between Clinical-Pathological Features and MAPs

Next, we examined how MAP stratification correlated with gender and tumor site, the only variables we had available for a reasonable number of patients (**Table 3**). MAP stratification did not significantly correlate with gender. Concerning tumor site, differences in MAP distribution between colon and rectal localization were close to statistical significance, with a trend for MAP2.1 to be more represented and for MAP2.2 and MAP3.1 to be less represented in rectum compared to colon cancer (**Table 3**). Moreover, while MAP1 and MAP2 have similar frequencies among right-side and left-side CRC, MAP3.2 and MAP4.2 (accounting for by samples with no mutations or TP53 mutations only) were overrepresented in left-side CRC, and MAP4.1 was overrepresented in right-side CRC. Of relevance, the association of MAP 4.1 with right side remains significant even by omitting MSI-H cases (not shown).

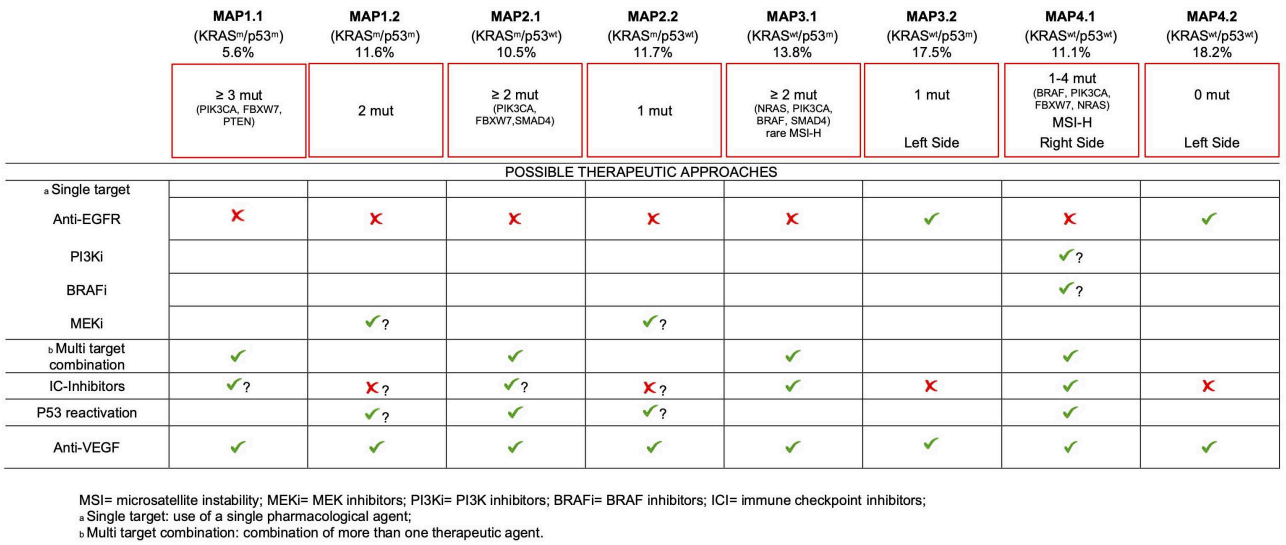
DISCUSSION

The response of mCRC to current therapeutic approaches is highly variable, reflecting the elevated heterogeneity of the disease (7). This, together with an increasing availability of targeted therapeutic approaches, stresses the need for more comprehensive molecular characterization of each tumor sample, in order to push forward the real achievement of personalized interventions. Despite it is clear that an extended molecular characterization of CRC patients may significantly impact on their clinical management (27–30), very little has entered the clinical routine, yet.

Here, we report that a clinical genomic profiling via multigene panel sequencing allowed identification of pairwise mutation associations and eight distinct MAPs, providing great opportunities to direct more informed therapeutic decisions, in the majority of mCRC cases.

Our data confirm previously reported pairwise gene mutation associations (31, 43–46) and unveil for the first time EGFR/KRAS, EGFR/SMAD4, BRAF/PTEN, and NRAS/TP53 positive associations. The biological or clinical meaning of these associations is difficult to trace, at the moment. In example, while mutations in the EGFR tyrosine-kinase domain are mutually exclusive with KRAS mutations and are positive predictive biomarkers for the efficacy of tyrosine kinase inhibitors in non-small cell lung cancer (NSCLC) (34, 47), these mutations are rare and scarcely relevant in predicting responses to antibody-based anti-EGFR therapy, in mCRC (48). 8/11 EGFR mutation occurring in our mCRC cohort coexist with KRAS mutations. On one side, this may imply that coexisting EGFR and KRAS activating mutation might provide additional advantages to tumor progression in mCRC. This is also supported by the frequent co-occurrence of KRAS and PIK3CA, leading to the constitutive activation of two different pathways downstream of the EGFR. Alternatively, EGFR and KRAS mutations could also exist in different subclonal mCRC populations in the same tumor, as a result of tumor heterogeneity, driving the activation of the same pathway. Addressing the details of tumor heterogeneity and clonality, by tumor multisampling and/or single cell sequencing, will be required to address these issues.

Clinical multigene panel sequencing may easily lead to the identification of actionable and targetable gene mutations (27–31). More importantly, it provides awareness that specific actionable/targetable mutations most frequently co-occur with additional relevant genetic alterations, which in principle could contribute to an individual variability in patients' responsiveness to standard and target-driven therapies. Overlooking this molecular complexity may account for treatment failures, when approved or innovative targeted approaches are used. The scant success of PI3K inhibitors in mCRC may be at least in part due to PIK3CA mutations co-occurring with RAS/BRAF mutations (more than 70% of the $PIK3CA^{mut}$ patients in our cohort) (49). It cannot be excluded that other gene mutations (occurring in an additional 18% of the $PIK3CA^{mut}$ patients in our cohort) may also provide primary resistance to PI3K inhibitors. Only 1.4% of our entire cohort carried exclusively *PIK3CA* mutations, possibly underscoring a subset of patients best suited for treatment with PI3K inhibitors. On the same line, 2.7% (17/639) of the patients carried only *BRAF* mutation, possibly representing the best subset for a target treatment with $BRAF^{V600E}$ inhibitors alone, or in combination with anti-EGFR (50, 51). Most patients carrying $BRAF^{V600E}$ also carried additional mutations, at least some of



MSI= microsatellite instability; MEKi= MEK inhibitors; PI3Ki= PI3K inhibitors; BRAFi= BRAF inhibitors; ICI= immune checkpoint inhibitors;
a Single target: use of a single pharmacological agent;
b Multi target combination: combination of more than one therapeutic agent.

FIGURE 4 | Molecular association pattern taxonomy and potential implications for therapies.

which might be expected to be involved in primary resistance to anti-BRAF therapies, providing contraindication to single target approaches. In line with this, a phase III 3-arm trial is currently exploring the effectiveness of a triplet therapy with the BRAF inhibitor plus MEK inhibitor associated with the anti EGFR antibody CETUXIMAB in *BRAF*^{V600E} mCRC, in the second or third-line setting (BEACON CRC NCT02928224) (23).

Most importantly, we believe that the greatest added value of clinical multigene panel sequencing may come from a more comprehensive use of the entire mutational profile of each patient to implement a more precise molecular stratification of mCRC. In this observational study, we developed a new stratification system into eight distinct MAPs characterized by non-random, specific mutational combinations. We validated these findings via TCGA data analysis, although few interesting differences emerged. In particular, the different rate of MSI-H cases and the different size of MAP4.1 may be due to the different stage composition between our cohort and TCGA dataset. Whether this is also relevant for the different size of MAP4.2 and the different distribution of some less frequently mutated genes remains to be determined.

Although we are aware that the lack of clinical data only allows for a speculative proposition, we believe that our comprehensive molecular stratification may provide the base for informed therapeutic decisions, for the majority of mCRC patients, as detailed below (**Figure 4**).

Firstly, about 50% of our cohort bears no mutations at all (MAP4.2) or just one mutation (MAP2.2 and MAP3.2). An additional 11.6% of patients is characterized by mutations limited to *KRAS* and *TP53* (MAP1.2). Even considering the almost ubiquitous activation of the WNT pathway due to mutations in APC, β-catenin or RNF43 genes (14), these data suggest that the majority of mCRC develop and progress with a low mutation load, which has significant implications for therapeutic interventions. Of interest, the little representation of MAP4.2 in the mixed-stages TCGA dataset compared to our mCRC cohort seems to suggest a higher tendency to cancer progression and a negative prognostic value to having no detectable pathogenic mutations, in addition to being less responsive to common therapies (27). This interesting hypothesis also needs to be confirmed in independent series.

MAP1.1 patients (as well as MAP1.2, MAP2.1, and MAP2.2 patients) are not eligible for anti-EGFR therapy, due to KRAS mutations. More in general, therapies directed against single targets are likely to fail in this group of patients due to primary resistance, as a consequence of having ≥3 gene mutations. Appropriate multitarget combinations should instead be considered for this group, taking advantage of the druggable mutations detected in each patient. Inhibitors of immune-checkpoints are effective in MSS patients (52). The presence of ≥3 pathogenic mutations/tumor in MAP1.1 may suggest a higher mutation rate (compared to MAP 2.2/4.2), raising the possibility to test the efficacy of checkpoint inhibitors, in this subset.

Due to the occurrence of multiple mutations, target driven drug combinations also need to be considered for MAP2.1. However, in this subset we noticed the highest frequency of *PIK3CA* mutations (42/66), 5/66 *PTEN* mutations and 3 out of the 5 *AKT1* mutations, mutually exclusive with *PIK3CA* mutations. Thus, the highest frequency of constitutive activation of the PI3K-AKT1 pathway occurs in this *TP53* WT subset. It has been shown that p53 may limit KRAS dependent transformation (53), suggesting that p53 inactivation may be required for KRAS-dependent cancer development. Nonetheless, KRAS and TP53 mutations are not positively associated in mCRC (31, 44), at variance with NRAS mutations. Of interest, PI3K-AKT axis impinges on MDM2, promoting an increased E3-ubiquitin ligase activity, ultimately leading to p53 functional inactivation via

increased degradation (54). Therefore, activation of the PI3K-AKT pathway provides a functional mean to inactivate p53, in KRAS mutant samples. Consistent with this, Singh et colleagues found mutually exclusive occurrence of *TP53* mutations and *PIK3CA* amplification in squamous cell carcinomas (55). This support the possibility that TP53 reactivation approaches, which are being tested elsewhere (56), could also find application in MAP2.1 mCRC (**Figure 4**).

Beside standard treatment including anti-VEGF, additional intervention is hard to prospect for MAP1.2 patients, since they lack targetable gene mutations, with the possible exception of MEK inhibitors. Mutant TP53 reactivation approaches are yet to come at the clinical level, but they will find potential application also in this mCRC subset. The role of immunotherapy in this subset and in MAP2.2 patients seems counterintuitive, due to the low mutation rate.

Targeting EGFR as a single strategy will probably be ineffective for most patients of MAP3.1 due to the frequent occurrence of PIK3CA, NRAS, BRAF, or SMAD4 mutations, all of which have been related to primary resistance to this approach (17, 43, 57, 58). Therefore, combination treatments should also be carefully planned in this subset. Importantly, few MSI-H patients fall in this group creating opportunities for immune system reactivation therapies.

In sharp contrast, MAP3.2 and MAP4.2 patients, largely coincident with the known "quadruple negative" mCRC subset (24, 59), are probably the most eligible to chemotherapy plus anti-EGFR therapies, since they lack known predictable resistance mechanisms. Of interest, these subsets are prevalent in the left colon, consistent with the observation that TP53 mutations and alternative mechanisms of activation of receptor tyrosine kinase pathways characterize tumors developing in the distal colon (14). These data also fit with the recently reported increased chance of response to anti-EGFR treatment in left colon mCRC (60, 61).

mCRCs of the MAP4.1 subset are predominantly localized to the right colon, where tumors appear to be less responsive to conventional therapies (60, 62). In this subset we detected the highest percentage of $BRAF^{V600E}$ mutant patients, suggesting multiple targeting of $BRAF^{V600E}$ and EGFR, perhaps also in combination with MEK inhibitors (50, 63). Anti-EGFR therapy alone should be possibly avoided, due to the frequent occurrence of primary resistance mutations in PIK3CA, NRAS, BRAF, or FBXW7 (17, 43, 57, 58). TP53 reactivation may also seem reasonable, in cases with NRAS and PIK3CA mutations, similar to MAP2.1 patients. Finally, MAP4.1 also includes the majority of MSI-H mCRC patients, which are most likely to benefit from immune checkpoint inhibitors (52).

Although we are aware that our clinical genomic profiling does not take into account copy number variations and genomic rearrangements that may lead to derangement of specific oncogenic/oncosuppressive pathways, these rarely occur in mCRC (14). It remains that the major limitation of our study is that we had no access to homogeneously collected clinical data, which clearly prevented us from reaching significant clinical conclusion. In example, we cannot infer whether any of the MAPs indicates a better response to anti-VEGF therapy, which is still orphan of biomarkers. Nonetheless, we believe that the simple and cost-effective molecular stratification of mCRC compatible with clinical settings described in this observational study will encourage us and others to design prospective studies to specifically address its effective value for more personalized therapeutic intervention of mCRC patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI SRA (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA614492/).

ETHICS STATEMENT

For this retrospective observational studies all investigations were approved by the Ethics Committee of the University La Sapienza (Prot.: 88/18; RIF.CE:4903, 31-01-2018). All information regarding human material included in the study was managed using anonymous numerical codes, and all samples were handled in compliance with the principles outlined in the declaration of Helsinki. For samples collected at the Department of Public Health, University Federico II, we obtained written informed consent from all patients, in accordance with the general authorization to process personal data for scientific research purposes from The Italian Data Protection Authority (http://www.garanteprivacy.it/web/guest/home/docweb/-/docwebdisplay/export/2485392).

AUTHOR CONTRIBUTIONS

FB performed NGS and statistical analyses, interpreted the results, and drafted the manuscript. CC recruited samples, collected clinical-pathologic data, and interpreted the results. UM, PPi, and FP recruited samples, performed NGS, and collected sequencing data. DR and EM performed statistical/bioinformatics analyses and interpreted the results. MY and CL performed NGS. PPa, PS, and CB performed statistical analyses. SM, VM, AC, AN, AG, MP, SD, FF, PI, SC, GC, and GT recruited samples and collected clinical-pathologic data. GG conceived, designed, coordinated the study, and drafted the manuscript. All authors reviewed, edited, and approved the manuscript for publication.

FUNDING

This work was supported by Italian Ministry of Education, Universities and Research-Dipartimenti di Eccellenza-L. 232/2016; AIRC grant (IG17734), Italian Ministry of University and Research, PRIN projects, Istituto Pasteur-Fondazione Cenci Bolognetti and Ricerca Scientifica di Ateneo La Sapienza, to GG. AIRC grant (IG17575) and Istituto Pasteur Fondazione Cenci Bolognetti to GC.

ACKNOWLEDGMENTS

FB was supported by Fondazione Umberto Veronesi; FF is a recipient of a fellowship of the Ph.D. Programme in Tecnologie Biomediche in Medicina Clinica, University La Sapienza.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00560/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69:7–34. doi: 10.3322/caac.21551
2. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin*. (2017) 67:177–93. doi: 10.3322/caac.21395
3. van Cutsem E, Cervantes A, Nordlinger B, Arnold D, ESMO Guidelines Working Group. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. (2014) 25(Suppl. 3):iii1–9. doi: 10.1093/annonc/mdu260
4. Colucci G, Gebbia V, Paoletti G, Giuliani F, Caruso M, Gebbia N, et al. Phase III randomized trial of FOLFIRI versus FOLFOX4 in the treatment of advanced colorectal cancer: a multicenter study of the Gruppo Oncologico Dell'Italia Meridionale. *J Clin Oncol*. (2005) 23:4866–75. doi: 10.1200/JCO.2005.07.113
5. Tournigand C, André T, Achille E, Lledo G, Flesh M, Mery-Mignard D, et al. FOLFIRI followed by FOLFOX6 or the reverse sequence in advanced colorectal cancer: a randomized GERCOR study. *J Clin Oncol*. (2004) 22:229–37. doi: 10.1200/JCO.2004.05.113
6. Golfinopoulos V, Salanti G, Pavlidis N, Ioannidis JPA. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: a meta-analysis. *Lancet Oncol*. (2007) 8:898–911. doi: 10.1016/S1470-2045(07)70281-4
7. Cremolini C, Schirripa M, Antoniotti C, Moretto R, Salvatore L, Masi G, et al. First-line chemotherapy for mCRC—a review and evidence-based algorithm. *Nat Rev Clin Oncol*. (2015) 12:607–19. doi: 10.1038/nrclinonc.2015.129
8. Bokemeyer C, Bondarenko I, Makhson A, Hartmann JT, Aparicio J, de Braud F, et al. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J Clin Oncol*. (2009) 27:663–71. doi: 10.1200/JCO.2008.20.8397
9. Saltz LB, Clarke S, Díaz-Rubio E, Scheithauer W, Figer A, Wong R, et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol*. (2008) 26:2013–9. doi: 10.1200/JCO.2007.14.9930
10. Van Cutsem E, Köhne C-H, Hitre E, Zaluski J, Chang Chien C-R, Makhson A, et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med*. (2009) 360:1408–17. doi: 10.1056/NEJMoa0805019
11. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. (1990) 61:759–67. doi: 10.1016/0092-8674(90)90186-I
12. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science*. (2013) 339:1546–58. doi: 10.1126/science.1235122
13. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. (2015) 21:1350–6. doi: 10.1038/nm.3967
14. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, et al. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell*. (2018) 33:125–36.e3. doi: 10.1016/j.ccell.2017.12.004
15. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. (2012) 487:330–7. doi: 10.1038/nature11252
16. Capece D, Verzella D, Tessitore A, Alesse E, Capalbo C, Zazzeroni F. Cancer secretome and inflammation: the bright and the dark sides of NF-κB. *Semin Cell Dev Biol*. (2018) 78:51–61. doi: 10.1016/j.semcdb.2017.08.004
17. Douillard J-Y, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med*. (2013) 369:1023–34. doi: 10.1056/NEJMoa1305275
18. van Cutsem E, Lenz H-J, Köhne C-H, Heinemann V, Tejpar S, Melezínek I, et al. Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J Clin Oncol*. (2015) 33:692–700. doi: 10.1200/JCO.2014.59.4812
19. Allegra CJ, Rumble RB, Hamilton SR, Mangu PB, Roach N, Hantel A, et al. Extended RAS gene mutation testing in metastatic colorectal carcinoma to predict response to anti-Epidermal growth factor receptor monoclonal antibody therapy: American Society of clinical oncology provisional clinical opinion update (2015). *J Clin Oncol*. (2016) 34:179–85. doi: 10.1200/JCO.2015.63.9674
20. Battaglin F, Puccini A, Ahcene Djaballah S, Lenz H-J. The impact of panitumumab treatment on survival and quality of life in patients with RAS wild-type metastatic colorectal cancer. *Cancer Manag Res*. (2019) 11:5911–24. doi: 10.2147/CMAR.S186042
21. Rodriguez-Pascual J, Cubillo A. Dynamic biomarkers of response to antiangiogenic therapies in colorectal cancer: a review. *Curr Pharmacogen Person Med*. (2017) 15:81–5. doi: 10.2174/1875692115666170815161754
22. Jubb AM, Harris AL. Biomarkers to predict the clinical efficacy of bevacizumab in cancer. *Lancet Oncol*. (2010) 11:1172–83. doi: 10.1016/S1470-2045(10)70232-1
23. Kopetz S, Desai J, Chan E, Hecht JR, O'Dwyer PJ, Maru D, et al. Phase II pilot study of Vemurafenib in patients with metastatic BRAF-mutated colorectal cancer. *J Clin Oncol*. (2015) 33:4032–8. doi: 10.1200/JCO.2015.63.2497
24. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer*. (2017) 17:79–92. doi: 10.1038/nrc.2016.126
25. Belardinilli F, Capalbo C, Buffone A, Petroni M, Colicchia V, Ferraro S, et al. Validation of the Ion Torrent PGM sequencing for the prospective routine molecular diagnostic of colorectal cancer. *Clin Biochem*. (2015) 48:908–10. doi: 10.1016/j.clinbiochem.2015.04.003
26. Malapelle U, Vigliar E, Sgariglia R, Bellevicine C, Colarossi L, Vitale D, et al. Ion torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *J Clin Pathol*. (2015) 68:64–8. doi: 10.1136/jclinpath-2014-202691
27. Capalbo C, Belardinilli F, Raimondo D, Milanetti E, Malapelle U, Pisapia P, et al. A simplified genomic profiling approach predicts outcome in metastatic colorectal cancer. *Cancers*. (2019) 11:147. doi: 10.3390/cancers11020147
28. Raimondi C, Nicolazzo C, Belardinilli F, Loreni F, Gradilone A, Mahdavian Y, et al. Transient disappearance of RAS mutant clones in plasma: a counterintuitive clinical use of EGFR inhibitors in RAS mutant metastatic colorectal cancer. *Cancers*. (2019) 11:42. doi: 10.3390/cancers11010042
29. Gao XH, Yu GY, Hong YG, Lian W, Chouhan H, Xu Y, et al. Clinical significance of multiple gene detection with a 22-gene panel in formalin-fixed paraffin-embedded specimens of 207 colorectal cancer patients. *Int J Clin Oncol*. (2019) 24:141–52. doi: 10.1007/s10147-018-1377-1
30. de Nicola F, Goeman F, Pallocca M, Sperati F, Pizzuti L, Melucci E, et al. Deep sequencing and pathway-focused analysis revealed multigene oncodriver signatures predicting survival outcomes in advanced colorectal cancer. *Oncogenesis*. (2018) 7:55. doi: 10.1038/s41389-018-0066-2
31. Malapelle U, Pisapia P, Sgariglia R, Vigliar E, Biglietto M, Carlomagno C, et al. Less frequently mutated genes in colorectal cancer: evidences from next-generation sequencing of 653 routine cases. *J Clin Pathol*. (2016) 69:767–71. doi: 10.1136/jclinpath-2015-203403
32. Del Vecchio F, Mastroiaco V, Di Marco A, Compagnoni C, Capece D, Zazzeroni F, et al. Next-generation sequencing: recent applications

to the analysis of colorectal cancer. *J Transl Med*. (2017) 15:246. doi: 10.1186/s12967-017-1353-y
33. Li W, Qiu T, Guo L, Ying J. Major challenges related to tumor biological characteristics in accurate mutation detection of colorectal cancer by next-generation sequencing. *Cancer Lett*. (2017) 410:92–9. doi: 10.1016/j.canlet.2017.09.014
34. Belardinilli F, Gradilone A, Gelibter A, Zani M, Occhipinti M, Ferraro S, et al. Coexistence of three EGFR mutations in an NSCLC patient: a brief report. *Int J Biol Markers*. (2018) 33:1724600818782200. doi: 10.1177/1724600818782200
35. Nicolussi A, Belardinilli F, Mahdavian Y, Colicchia V, D'Inzeo S, Petroni M, et al. Next-generation sequencing of BRCA1 and BRCA2 genes for rapid detection of germline mutations in hereditary breast/ovarian cancer. *PeerJ*. (2019) 7:e6661. doi: 10.7717/peerj.6661
36. Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, et al. Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology*. (2002) 123:1804–11. doi: 10.1053/gast.2002.37070
37. Jolliffe IT. *Principal Component Analysis*. 2nd ed. New York, NY: Springer-Verlag (2002). Available online at: https://www.springer.com/us/book/9780387954424 (accessed January 27, 2019).
38. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The cancer genome atlas pan-Cancer analysis project. *Nat Genet*. (2013) 45:1113–20. doi: 10.1038/ng.2764
39. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. (2008) 455:1061–8. doi: 10.1038/nature07385
40. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. (2016) 44:e71. doi: 10.1093/nar/gkv1507
41. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. *JCO Precis Oncol*. (2017) 2017. doi: 10.1200/PO.17.00011
42. Smeets D, Miller IS, O'Connor DP, Das S, Moran B, Boeckx B, et al. Copy number load predicts outcome of metastatic colorectal cancer patients receiving bevacizumab combination therapy. *Nat Commun*. (2018) 9:4112. doi: 10.1038/s41467-018-06567-6
43. de Roock W, Claes B, Bernasconi D, de Schutter J, Biesmans B, Fountzilas G, et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol*. (2010) 11:753–62.
44. Domingo E, Ramamoorthy R, Oukrif D, Rosmarin D, Presz M, Wang H, et al. Use of multivariate analysis to suggest a new molecular classification of colorectal cancer. *J Pathol*. (2013) 229:441–8. doi: 10.1002/path.4139
45. Korphaisarn K, Morris VK, Overman MJ, Fogelman DR, Kee BK, Raghav KPS, et al. FBXW7 missense mutation: a novel negative prognostic factor in metastatic colorectal adenocarcinoma. *Oncotarget*. (2017) 8:39268–79. doi: 10.18632/oncotarget.16848
46. Mehrvarz Sarshekeh A, Advani S, Overman MJ, Manyam G, Kee BK, Fogelman DR, et al. Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. *PLoS ONE*. (2017) 12:e0173345. doi: 10.1371/journal.pone.0173345
47. Chan BA, Hughes BGM. Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Transl Lung Cancer Res*. (2015) 4:36–54. doi: 10.3978/j.issn.2218-6751.2014.05.01
48. Heinemann V, Stintzing S, Kirchner T, Boeck S, Jung A. Clinical relevance of EGFR- and KRAS-status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR. *Cancer Treat Rev*. (2009) 35:262–71. doi: 10.1016/j.ctrv.2008.11.005
49. Alzahrani AS. PI3K/Akt/mTOR inhibitors in cancer: at the bench and bedside. *Semin Cancer Biol*. (2019) 59:125–32. doi: 10.1016/j.semcancer.2019.07.009
50. Capalbo C, Marchetti P, Coppa A, Calogero A, Anastasi E, Buffone A, et al. Vemurafenib and panitumumab combination tailored therapy in BRAF-mutated metastatic colorectal cancer: a case report. *Cancer Biol Ther*. (2014) 15:826–31. doi: 10.4161/cbt.28878
51. Ursem C, Atreya CE, van Loon K. Emerging treatment options for BRAF-mutant colorectal cancer. *Gastrointest Cancer*. (2018) 8:13–23. doi: 10.2147/GICTT.S125940
52. Cohen R, Pellat A, Boussion H, Svrcek M, Lopez-Trabada D, Trouilloud I, et al. Immunotherapy and metastatic colorectal cancers with microsatellite instability or mismatch repair deficiency. *Bull Cancer*. (2019) 106:137–42. doi: 10.1016/j.bulcan.2018.09.004
53. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*. (1997) 88:593–602. doi: 10.1016/S0092-8674(00)81902-9
54. Moll UM, Petrenko O. The MDM2-p53 interaction. *Mol Cancer Res*. (2003) 1:1001–8.
55. Singh B, Reddy PG, Goberdhan A, Walsh C, Dao S, Ngai I, et al. p53 regulates cell survival by inhibiting PIK3CA in squamous cell carcinomas. *Genes Dev*. (2002) 16:984–93. doi: 10.1101/gad.973602
56. Blandino G, Di Agostino S. New therapeutic strategies to treat human cancers expressing mutant p53 proteins. *J Exp Clin Cancer Res*. (2018) 37:30. doi: 10.1186/s13046-018-0705-7
57. Cremolini C, Pietrantonio F. How the lab is changing our view of colorectal cancer. *Tumori*. (2016) 102:541–7. doi: 10.5301/tj.5000551
58. Lupini L, Bassi C, Mlcochova J, Musa G, Russo M, Vychytilova-Faltejskova P, et al. Prediction of response to anti-EGFR antibody-based therapies by multigene sequencing in colorectal cancer patients. *BMC Cancer*. (2015) 15:808. doi: 10.1186/s12885-015-1752-5
59. Bardelli A, Siena S. Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer. *J Clin Oncol*. (2010) 28:1254–61. doi: 10.1200/JCO.2009.24.6116
60. Tejpar S, Stintzing S, Ciardiello F, Tabernero J, Van Cutsem E, Beier F, et al. Prognostic and predictiverRelevance of primary tumor location in patients with RAS wild-type metastatic colorectal cancer: retrospective analyses of the CRYSTAL and FIRE-3 trials. *JAMA Oncol*. (2017) 3:194–201. doi: 10.1001/jamaoncol.2016.3797
61. Lee GH, Malietzis G, Askari A, Bernardo D, Al-Hassi HO, Clark SK. Is right-sided colon cancer different to left-sided colorectal cancer? - a systematic review. *Eur J Surg Oncol*. (2015) 41:300–8. doi: 10.1016/j.ejso.2014.11.001
62. Holch JW, Ricard I, Stintzing S, Modest DP, Heinemann V. The relevance of primary tumour location in patients with metastatic colorectal cancer: a meta-analysis of first-line clinical trials. *Eur J Cancer*. (2017) 70:87–98. doi: 10.1016/j.ejca.2016.10.007
63. Corcoran RB, André T, Atreya CE, Schellens JHM, Yoshino T, Bendell JC, et al. Combined BRAF, EGFR, and MEK inhibition in patients with BRAFV600E-mutant colorectal cancer. *Cancer Discov*. (2018) 8:428–43. doi: 10.1158/2159-8290.CD-17-1226

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer VS declared a past co-authorship with the authors to the handling editor.

frontiers in Oncology

ORIGINAL RESEARCH
published: 08 May 2020
doi: 10.3389/fonc.2020.00726

The Genomic Characteristics of ALK Fusion Positive Tumors in Chinese NSCLC Patients

Shaokun Liu[1,2,3†], Tanxiao Huang[4†], Ming Liu[4], Wenlong He[1,2,3], YingShen Zhao[4], Lizhen Yang[1,2,3], Yingjiao Long[1,2,3], Dandan Zong[1,2,3], Huihui Zeng[1,2,3], Yuanyuan Liu[4], Wenting Liao[4], Jingxian Duan[4], Subo Gong[5] and Shifu Chen[4*]*

[1] Department of Pulmonary and Critical Care Medicine, The Second Xiangya Hospital, Central South University, Changsha, China, [2] Research Unit of Respiratory Disease, Central South University, Changsha, China, [3] Diagnosis and Treatment Center of Respiratory Disease, Central South University, Changsha, China, [4] HaploX Biotechnology Co., Ltd., Shenzhen, China, [5] Department of Geriatrics, The Second Xiangya Hospital, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Ira Ida Skvortsova, Innsbruck Medical University, Austria

Reviewed by:
Xiaofei Song, University of Texas MD Anderson Cancer Center, United States
Qi Zhao, Sun Yat-sen University Cancer Center (SYSUCC), China

***Correspondence:**
Subo Gong
gsb510@csu.edu.cn
Shifu Chen
chen@haplox.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *22 January 2020*
Accepted: *16 April 2020*
Published: *08 May 2020*

Citation:
Liu S, Huang T, Liu M, He W, Zhao Y, Yang L, Long Y, Zong D, Zeng H, Liu Y, Liao W, Duan J, Gong S and Chen S (2020) The Genomic Characteristics of ALK Fusion Positive Tumors in Chinese NSCLC Patients. Front. Oncol. 10:726. doi: 10.3389/fonc.2020.00726

Anaplastic lymphoma kinase (ALK) fusion events account for ∼3–7% genetic alterations in patients with non-small cell lung cancer (NSCLC). In this study, we identified the ALK fusion patterns and a novel ALK fusion partner in 44 ALK positive NSCLC patients using a customized HapOncoCDx panel, and identified ALK fusion partners. The most common partner is EML4, forming the variant 1 (v1, E13:A20, 18/44), variant 2 (v2, E20:A20, 5/44), and variant 3 (v3, E6:A20, 13/44). Moreover, we detected a new ALK fusion partner HMBOX1. At the mutation level, TP53 is the most frequently mutated gene (24%), followed by ALK (12%) and STED2 (12%). The median tumor mutation burden (TMB) of these samples is 2.29 mutations/Mb, ranging from 0.76 mut/Mb to 16.79 muts/Mb. We further elaborately portrayed the TP53 mutation sites on the peptide sequence of the encoded protein by lollipop. The mutational signature and copy number alterations (CNAs) of the samples were also analyzed. The CNA events were found in 13 (13/44) patients, and the most commonly amplified genes were MDM2 ($n = 4/13$) and TERT ($n = 4/13$). Together, these results may guide personalized clinical management of patients with ALK fusion in the era of precision medicine.

Keywords: NSCLC, ALK, NGS—next generation sequencing, copy number aberrations, genomic landscape

INTRODUCTION

Anaplastic lymphoma kinase (ALK) fusion events, which are the result of ALK rearrangements, account for ∼3–7% genetic alterations in non-small cell lung cancer (NSCLC) patients (1, 2). These oncogenic mutations could lead to the constitutive activation of the ALK tyrosine kinase domain, and further cause tumorigenesis (3). Hitherto, multiple ALK fusion partners have been identified, and the most normal one is echinoderm microtubule-associated protein-like 4 (EML4), which were observed in nearly 80% of all the ALK fusion cases (2). It is worth noting that more than a dozen of different EML4-ALK variants have been identified in NSCLC patients. The most common variants are variant 1 (v1, E13:A20), variant 2 (v2, E20:A20), and variant 3 (v3, E6:A20) (4).

Currently, ALK tyrosine kinase inhibitors (TKIs) are recommended for the treatment of NSCLC patients harboring ALK fusion (5). Prior to ALK TKIs treatment, the median overall survival (OS) of ALK fusion-positive metastatic NSCLC patients receiving chemotherapy was around 12 months (6). However, under the sequential treatment with ALK fusion TKIs, the OS of the patients were

extended to 5 years (7). The first approved is Crizotinib which is the first generation TKI. Compared to traditional chemotherapy, it improves the PFS and OS of ALK fusion NSCLC patients significantly (8). Nevertheless, nearly all of the patients would develop drug resistance within 2 years after the initial treatment. The drug resistance was possibly caused by a secondary mutation in the kinase domain of ALK, the activation of alternative pathways, ALK amplification, or epithelial-mesenchymal transition (9). To overcome the resistance, the second-generation ALK TKIs were developed including Ceritinib, Alectinib, and Brigatinib. They were approved for the treatment of metastatic NSCLC patients with ALK fusion and had progressed on or intolerant to Crizotinib. Notably, Ceritinib and Alectinib were approved for the first-line treatment of the ALK fusion positive metastatic NSCLC patients (10–16). In addition, as the third generation ALK inhibitor, Lorlatinib has also been approved for the treatment of metastatic NSCLC patients with ALK fusion, on condition that the disease has progressed on Crizotinib or at least one other ALK inhibitor such as Alectinib or Ceritinib for metastatic disease (17). It is worth noting that different ALK inhibitors have different potencies and spectrums against acquired resistance mutations (18).

In the era of precision medicine, the genomic profiles of the patients may guide the planning of treatment strategy. For the ALK fusion positive NSCLC patients, detailed genomic profiles can elucidate the fusion partner and the rearranged breakpoint. Moreover, the proposed resistant mutations are critical for clinical treatment guidelines. Furthermore, several studies reported that molecular profiling is also associated with the prognosis of patients. Christopoulos et al. reported that the concomitant TP53 mutation is a strong indicator of poor prognosis in ALK fusion positive NSCLC patients (19). Their study also reported that EML4-ALK fusion variant V3 was associated with a more aggressive phenotype and worse overall survival due to earlier failure of several therapeutic modalities. In addition, they found that V3 and TP53 double positive patients had a very high risk of death with a median OS of around 2 years.

With the development of next-generation sequencing technologies, it is becoming more affordable to obtain the genomic landscape of cancer patients. In this study, we aim to demonstrate the genomic landscape of ALK fusion-positive tumors in 44 Chinese NSCLC patients sequenced with our customized HapOncoCDx panel which involves hybridization capture and deep sequencing of all protein-coding exons of 464 cancer-associated genes and other selected introns of other oncogenes and tumor suppressor genes, and illustrate their genomic mutation patterns and characteristics, which potentially helps to develop treatment strategy.

MATERIALS AND METHODS

Patients and Samples

Forty-four patients were enrolled from 1349 NSCLC patients in this study. Tumor tissues were collected during surgery, and were formalin fixed, paraffin-embedded (FFPE) and archived. Peripheral blood (PBL) samples were also collected from each patient as control.

DNA Extraction

DNA samples were extracted from Formalin-fixed paraffin-embedded (FFPE) samples with QIAamp DNA FFPE tissue kit (Qiagen). Extraction of PBL DNA was conducted using the RelaxGene Blood DNA system (Tiangen Biotech Co., Ltd., Beijing, China) according to the manufacturer's protocol. All the DNA samples were quantified both using the Qubit 2.0 fluorometer and the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to the manufacturer's protocol.

Library Construction and Sequencing

One hundred nanogram of DNA from each sample was sheared by the dsDNA Fragmentase (New England BioLabs, Inc., Ipswich, MA, USA), and then performed size selection (150–250 bp) using Ampure XP beads (Beckman Coulter, Inc., Brea, CA, USA). Library construction was performed using the KAPA Library Preparation kit (Kapa Biosystems, Inc., Wilmington, MA, USA) according to the manufacturer's protocol. The concentration of the library were assessed using the e Qubit dsDNA HS Assay kit, and fragment length was determined on a 4200 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). Target enrichment was carried out using the Agilent SureSelect XT HS kit (Agilent Technologies) according to the manufacturer's Protocol. DNA sequencing was then performed on the Illumina Novaseq 6000 system at an average depth of 2000X.

Data Analysis and Variant Calling

Raw sequences were pre-processed by fastp version 0.18.0 (https://github.com/OpenGene/fastp) (20), and clean reads were aligned to the hg19 genome (GRch37) using Burrows-Wheeler Aligner maximal exact matches algorithm (21). The Gencore version 0.12.0 (https://github.com/OpenGene/gencore) (22) was used for removing duplicate reads. Pileup files with mapping quality $\geq$60 were generated using Samtools version 0.1.19 (http://www.htslib.org/) (23). Somatic variants were called using VarScan2 version 2.3.8 (http://varscan.sourceforge.net/) (24) [the minimum read depth 20; the variant allele frequency (VAF) threshold $\geq$0.01; somatic-$P \leq 0.01$; strand-filter $\geq$1; others, default parameters]. CNV kit with version 0.9.3 (https://github.com/etal/cnvkit) (25) was used for copy number variation detection, and GeneFuse version v0.6.1 (https://github.com/OpenGene/GeneFuse) (26) for structural variation detection. Maftools was used for visualizing somatic variant analysis (27).

RESULTS

Sample Collection and Patient Characteristics

Of the 1349 NSCLC cases, ALK rearrangements were detected in 44 cases (3.26 %). Those 44 Chinese patients with locally advanced or metastatic NSCLC were enrolled in this study, of which 20 (45.5%) were female. All patients carry ALK fusion events. Their mean age was 52.5 with ranging from 29 to 73. NGS was performed on 44 pairs of tumor and white blood cell samples. All the samples that passed the histology

quality control (HQC) yielded sufficient amounts of DNA for NGS.

Identification of ALK Rearrangements Using Targeted Sequencing

In order to identify ALK rearrangement from the DNA of patients' FFPE samples, we designed probes to cover the intron 18 and intron 19 of ALK, as well as introns of some well-known ALK fusion partners. We identified ALK rearrangements and corresponding breakpoints in the sequencing data of these patients. The statistical summary and breakpoints of the rearrangement events are listed in **Table 1** and shown in **Figures 1**, **2**, respectively. We found that 43 out of 44 patients had an EML4-ALK fusion, with variant 1 (v1, E13:A20), variant 2 (v2, E20:A20), and variant 3 (v3, E6:A20) detected in 18, 5, and 13 patients, respectively. We also identified one novel ALK fusion partner HMBOX1.

TABLE 1 | Fusion patterns of ALK.

Fusion type	Counts	Percent (%)
EML4-exon13-ALK-exon20	18	40.9
EML4-exon6-ALK-exon20	13	29.5
EML4-exon20-ALK-exon20	5	11.4
EML4-exon13-ALK-exon19	2	4.5
EML4-exon14-ALK-exon20	2	4.5
EML4-exon6-ALK-exon19	2	4.5
EML4-exon19-ALK-exon20	1	2.3
HMBOX1-exon4-ALK-exon20	1	2.3

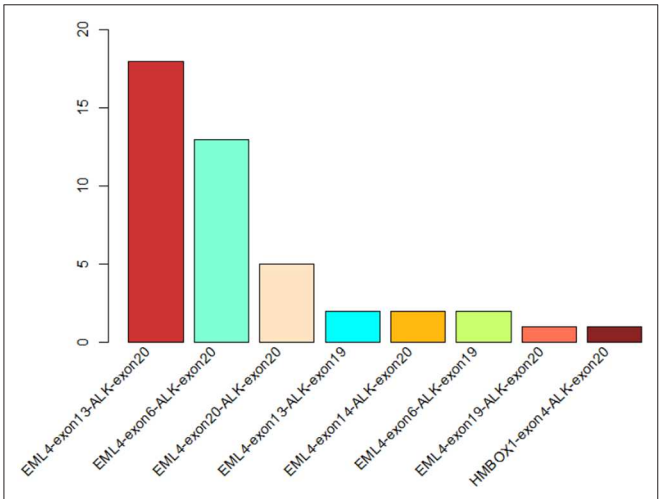


FIGURE 1 | The statistics of different ALK rearrangement forms. The number of each ALK fusion pattern identified in 44 NSCLC patients are shown in the barchart.

Mutational Profiles of ALK Fusion Positive NSCLC Patients

Genomic alterations were found in 34 ($n = 34/44$, 77.3%) samples with a total of 134 alterations identified including variants of non-synonymous mutations and splicing mutations. The detailed information is shown in **Figure 3A**. The mutation landscapes of ALK fusion positive NSCLC patients were highly heterogeneous. The median TMB was 2.29 mut/Mb with a range between 0.76 and 16.79 mut/Mb which is similar to the TMB of the TCGA NSCLC cohort.

We constructed a heatmap to demonstrate the somatic mutations occurred in the tumor tissues of the patients (**Figure 3A**). TP53 was most commonly altered ($n = 8/34$, 24%), followed by SETD2 ($n = 4$, 12%), ALK ($n = 4$, 12%), SYNE1 ($n = 3$, 9%), SMAD4 ($n = 3$, 9%), SLX4 ($n = 3$, 9%), NOTCH3 ($n = 3$, 9%), LRP1B ($n = 3$, 9%), EP300 ($n = 3$, 9%), and CTNNB1 ($n = 3$, 9%).

Other genomic alterations of low frequencies are AMER1 ($n = 2$, 6%), ARID1B ($n = 2$, 6%), CSF3R ($n = 2$, 6%), FAT1 ($n = 2$, 6%), FOXP2 ($n = 2$, 6%), KDM6A ($n = 2$, 6%), KMT2A ($n = 2$, 6%),LATS1 ($n = 2$, 6%), MAP2K4 ($n = 2$, 6%), NFEL2L2 ($n = 2$, 6%), NOTCH1($n = 2$, 6%), NTRK3 ($n = 2$, 6%), TERT ($n = 2$, 6%), and TGFBR2 ($n = 2$, 6%). Alterations in ABL1, ADH1B, ALDH2, APC, AR, ARID2, ATM, AURKA, BMPR1A, CACNA1C, CADM2, CAMTA1, CAPN2, CARD11, CDC73, CDK12, CREBBP, CSMD3, DNMT3A, EPHA3, ERBB4, ESR2, EWSR1, EXT1, EZH2, FGFR1, FLCN, FOXA1, FOXL2, GATA6, GPRIN2, HIF1A, HNF1B, JAK1, KDR, KMT2C, KMT2D, LATS2, MAP2K1, MDM4, MYCN, NF2, NSD1, NTRK1, PTEN, PTPRD, PZP, RARA, RIT1, RNF43, ROS1, SETBP1, SMARCA4, SMO, SOCS6, SOX2, SPEN, STAT3, STK11, SUZ12, TSHR, and U2AF1 were identified in one sample each ($n = 1$, 3%). We further compared our results with the MSK-IMPACT study (28), in which we extracted 45 ALK fusion positive cases that yielded 81 mutations. Overall, our results were highly consistent with the MSK-IMPACT findings, which showed that TP53 and ALK are the most frequently altered genes (**Figure 3B**).

We further studied their mutational signatures. We observed that C>T transition occurred most frequently, followed by C>A transition (**Figure 4**). This pattern is consistent with COSMIC signature 1 that had been found in most cancer samples.

Different driver gene mutations revealed inter-tumor heterogeneity. TP53 mutations in exon 5–8 were observed, and we further elaborately portrayed the TP53 mutation sites on the peptide sequence in a lollipop plot (**Figure 5**).

Copy Number Aberrations of ALK Fusion Positive NSCLC Patients

Somatic copy number alterations were found in 13 ($n = 13/44$, 29.5%) samples. A total of 22 alterations were identified, including gain and loss (**Figure 6**). MDM2 and TERT were most commonly amplified genes ($n = 4/13$, 31%), followed by CCND1, EPCAM, IKZF1, MET, MYCN, RICTOR ($n = 1$, 8%). Loss of copy number was most frequently observed in CD274, CDKN2A, JAK2 ($n = 2/13$, 15%), followed by FGFR1, FGFR3 ($n = 1$, 8%).

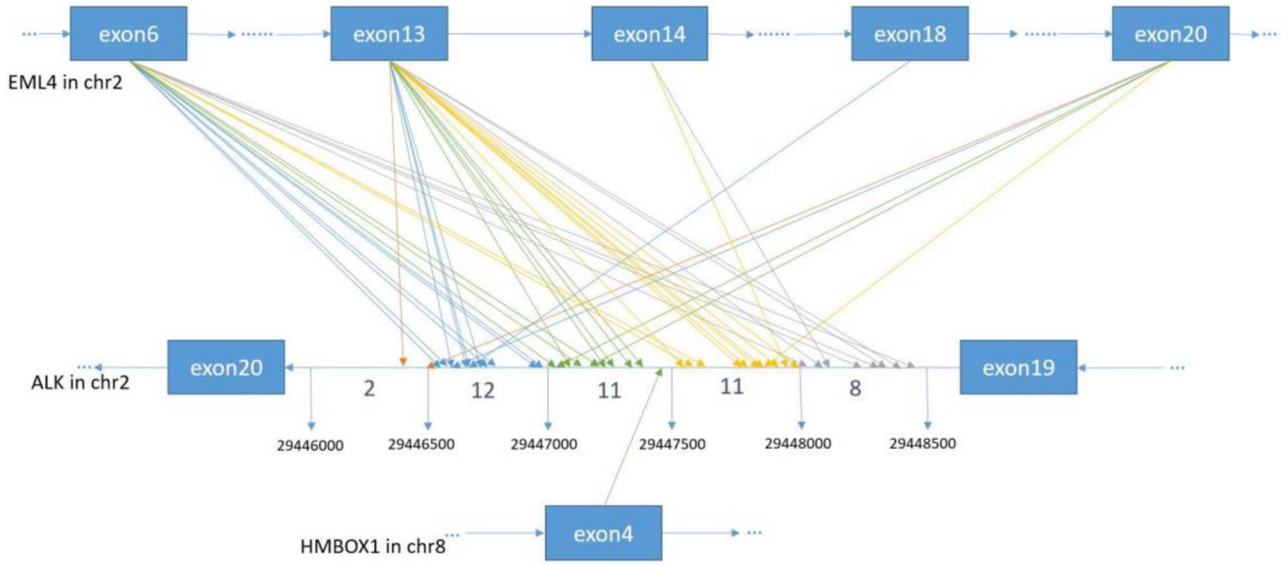


FIGURE 2 | The breakpoint distribution in ALK and the respective fusion partners. Each arrowed line represents one fusion event. The exact breakpoints of ALK in GRch37 are shown in the middle panel, while the fused exons of ALK fusion partners are shown in the top and bottom panels, respectively. The sequence of ALK is exhibited reversely (from right to left), while the sequences of EML4 and HMBOX1 are represented in the forward direction. The genomic region of ALK between 299446000 and 29448500 on Chromosome 2 is divided into regions every 500 bp. Breakpoint position in ALK locate between 29446000 and 29446500 with an orange arrow, between 29446501 and 29447000 with a blue arrow, between 29447001 and 29447500 with a green arrow, between 29447501 and 29448000 with an yellow arrow, between 29448001 and 29448500 with a gray arrow.

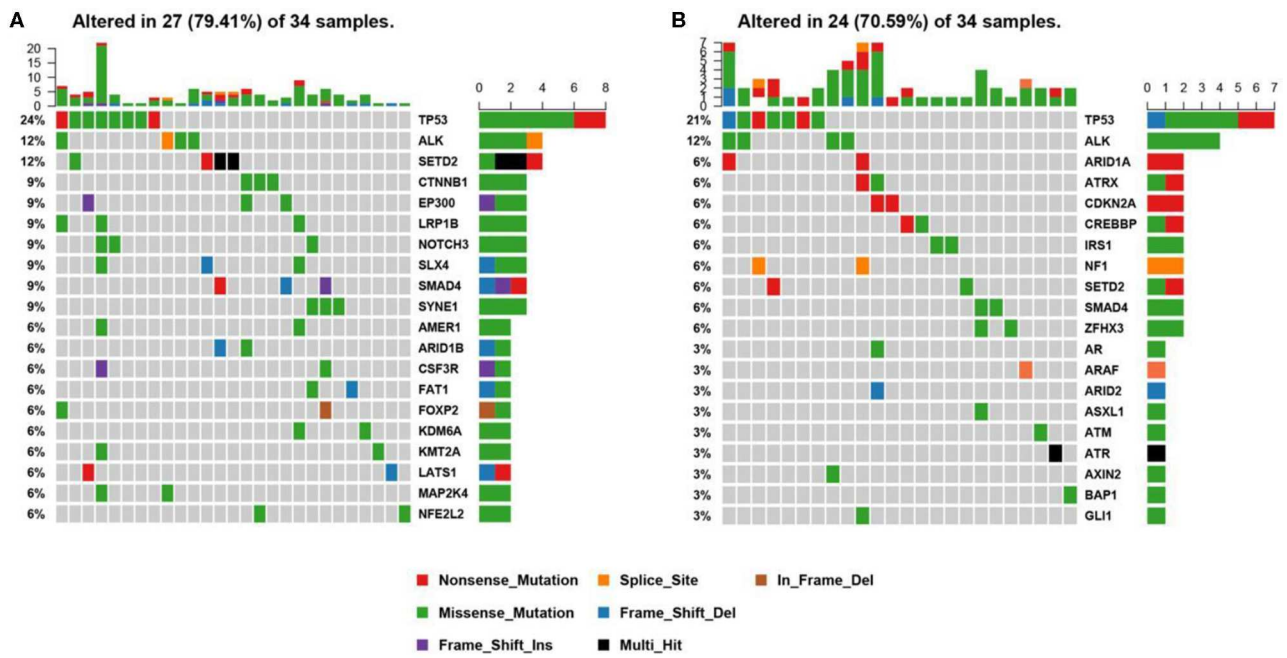


FIGURE 3 | Mutational profiles of ALK positive NSCLC patients. **(A)** This is the oncoprint of the somatic SNVs and Indels in 34 patients in our study. Somatic alterations included missense, nonsense, frameshift indel, in-frame indel, splice site, translation start site, multi_Hit mutations. The genes are ranked by the frequency of the mutations across all samples. **(B)** This is the oncoprint of the somatic SNVs and Indels in 34 patients from the MSK-IMPACT study. Somatic alterations included missense, nonsense, frameshift indel, in-frame indel, and splice-site mutations. The genes are ranked by the frequency of the mutations across all samples.

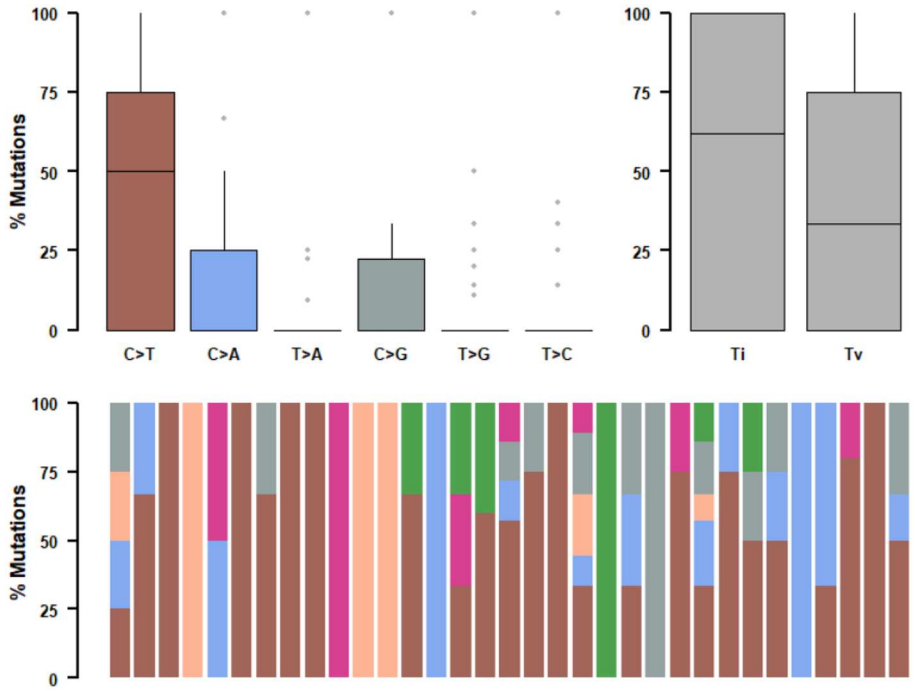


FIGURE 4 | Mutational signatures of ALK fusion positive NSCLC patients. SNPs are classified into transitions and transversions. Summarized data are visualized as a boxplot showing overall distribution of six different conversions **(Top)** and as a stacked barplot showing the fraction of conversions in each sample **(Bottom)**.

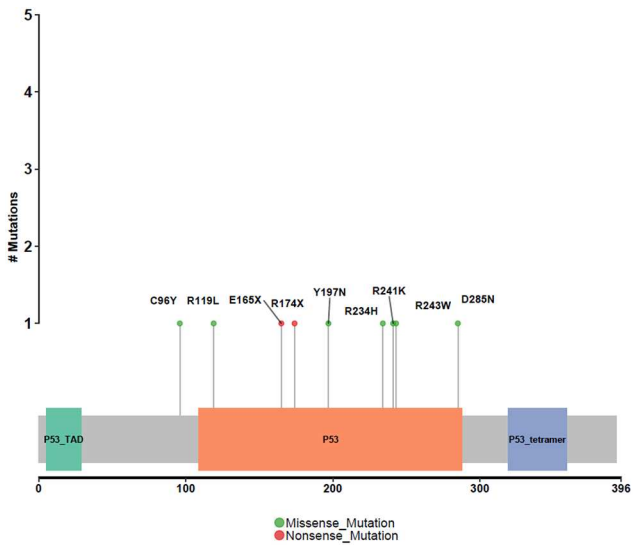


FIGURE 5 | Protein variants resulted from TP53 mutations. Protein variants caused by TP53 mutations are displayed in the lollipop plot. These sites are considered to be mutational hot-spots.

DISCUSSION

In this study, we identified ALK rearrangement events in 44 Chinese NSCLC patients using NGS technologies. Consistent with other studies, the most common ALK fusion partner is EML4, and the fusion occurs in the forms of the three most common variants. We also report a novel ALK fusion partner HMBOX1. It implied that NGS-based assessment for ALK fusions was accurate and comprehensive, having the unique advantages in detecting unknown ALK fusion partners, and identifying the exact breakpoints compared to the traditional methods, like FISH and IHC.

At the same time, we characterized the mutational profiles of the patients. The results were consistent with other studies, in terms of the relatively lower frequency of TP53 mutations, lower TMB, and fewer co-mutations compared to ALK-negative NSCLC patients (29). Besides, we identified the copy number alterations in their genome. Apart from the genes with a high frequency of copy number amplification, such as MET, MDM2, and TERT, we also identified some genes with copy number loss, such as CD274, CDKN2A, and JAK2. This information is important for guiding optimal clinical treatment. For instance, the copy number loss of CD274 probably indicates a low expression of PD-L1. MDM2 amplification had been reported to associate with a poor clinical outcome and significantly increased tumor growth rate with anti-PD-1/PD-L1 immunotherapies (30).

In conclusion, using our customized HapOncoCDx panel, we not only successfully detected the ALK fusion events in 44 Chinese NSCLC patients, but also explored their genomic mutational landscapes. To the best of our knowledge, this is the first study that exhibited the mutational landscape of Chinese NSCLC patients with ALK rearrangement. This result can provide genomic information for personalized clinical

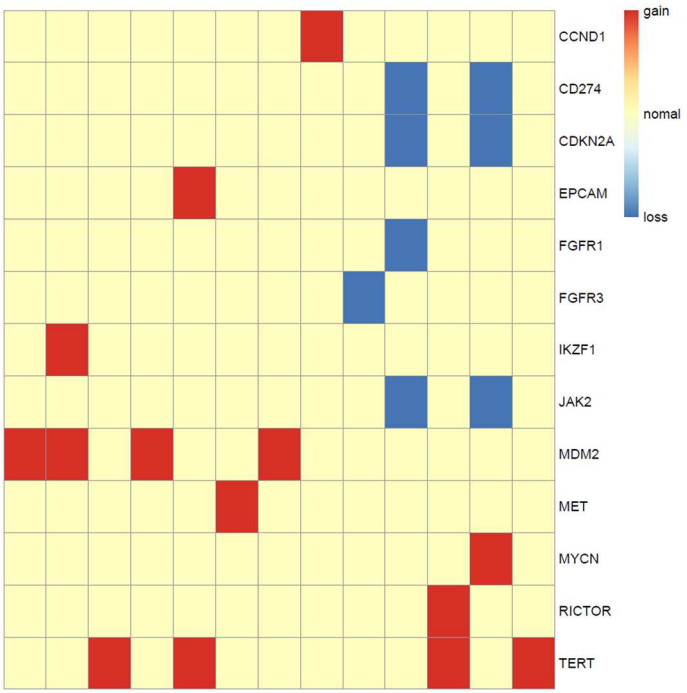


FIGURE 6 | Copy number aberrations in 13 ALK fusion positive NSCLC patients. The names of the aberrant genes are shown in the y-axis, while each column represents a patient. The type of copy number aberrations, including gain, normal, and loss are indicated by red, yellow, and blue, respectively.

management for patients with ALK fusion in the era of precision medicine.

DATA AVAILABILITY STATEMENT

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers HRA000138, HRA000138 that can be accessed at http://bigd.big.ac.cn/gsa-human.

ETHICS STATEMENT

This study was approved by the ethics committee of the Second XIANGYA Hospital of Central South University and complied with Good Clinical Practices, the principles of the Declaration of Helsinki and all applicable regulatory requirements. All patients provided written informed consent prior to any study-specific procedures.

AUTHOR CONTRIBUTIONS

SL, TH, SG, and SC designed the study. WH, YZ, LY, YLo, DZ, HZ, and YLi carried out the sequencing experiment and collected data. TH, YZ, and WL performed the bioinformatics analysis. TH, SL, and ML wrote this manuscript. JD and SC revised this manuscript. SG and SC supervised the study.

FUNDING

This work was supported by the project of Bureau of Industry and Information Technology of Shenzhen (Grant No. 20170922151538732), Shenzhen Science and Technology Innovation Committee Technical Research Project (Grant No. JSGG20180703164202084), the project of Development and Reform Commission of Shenzhen Municipality (Grant No. XMHT20190104006), and the National Key Clinical Specialty Construction Projects of China (Grant No. 2012-650).

REFERENCES

1. Cancer Genome Atlas Research Network Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* (2014) 511:543–50. doi: 10.1038/nature13385
2. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature.* (2007) 448:561–6. doi: 10.1038/nature05945
3. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell.* (2007) 131:1190–203. doi: 10.1016/j.cell.2007.11.025
4. Noh KW, Lee MS, Lee SE, Song JY, Shin HT, Kim YJ, et al. Molecular breakdown: a comprehensive view of anaplastic lymphoma kinase (ALK)-rearranged non-small cell lung cancer. *J. Pathol.* (2017) 243:307–19. doi: 10.1002/path.4950

5. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*. (2010) 363:1693–703. doi: 10.1056/NEJMoa1006448
6. Pilkington G, Boland A, Brown T, Oyee J, Bagust A, Dickson R. A systematic review of the clinical effectiveness of first-line chemotherapy for adult patients with locally advanced or metastatic non-small cell lung cancer. *Thorax*. (2015) 70:359–67. doi: 10.1136/thoraxjnl-2014-205914
7. Duruisseaux M, Besse B, Cadranel J, Pérol M, Mennecier B, Bigay-Game L, et al. Overall survival with crizotinib and next-generation ALK inhibitors in ALK-positive non-small-cell lung cancer (IFCT-1302 CLINALK): a French nationwide cohort retrospective study. *Oncotarget*. (2017) 8:21903–17. doi: 10.18632/oncotarget.15746
8. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer, *N. Engl. J. Med*. (2013) 368:2385–94. doi: 10.1056/NEJMoa1214886
9. Katayama R, Shaw AT, Khan TM, Mino-Kenudson M, Solomon BJ, Halmos B, et al. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung Cancers. *Sci Transl Med*. (2012) 4:120ra17. doi: 10.1126/scitranslmed.3003316
10. Shaw AT, Kim DW, Mehra R, Tan DSW, Felip E, Laura QM, et al. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med*. (2014) 370:1189–97. doi: 10.1056/NEJMoa1311107
11. Shaw AT, Kim TM, Crinò L, Gridelli C, Kiura K, Liu G, et al. Ceritinib versus chemotherapy in patients with ALK-rearranged non-small-cell lung cancer previously given chemotherapy and crizotinib (ASCEND-5): A randomised, controlled, open-label, phase 3 trial. *Lancet Oncol*. (2017) 18:874–86. doi: 10.1016/S1470-2045(17)30339-X
12. Ou SH, Ahn JS, De Petris L, Govindan R, Yang JC, Hughes B, et al. Alectinib in crizotinib-refractory ALK-rearranged non-small-cell lung cancer: A phase II global study. *J Clin Oncol*. (2016) 34:661–8. doi: 10.1200/JCO.2015.63.9443
13. Shaw AT, Gandhi L, Gadgeel S, Riely GJ, Cetnar J, West H, et al. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. *Lancet Oncol*. (2016) 17:234–42. doi: 10.1016/S1470-2045(15)00488-X
14. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. *N Engl J Med*. (2017) 377:829–38. doi: 10.1056/NEJMoa1704795
15. Gettinger SN, Bazhenova LA, Langer CJ, Salgia R, Gold KA, Rosell R, et al. Activity and safety of brigatinib in ALK-rearranged non-small-cell lung cancer and other malignancies: A single-arm, open-label, phase 1/2 trial. *Lancet Oncol*. (2016) 17:1683–96. doi: 10.1016/S1470-2045(16)30392-8
16. Kim DW, Tiseo M, Ahn MJ, Reckamp KL, Hansen KH, Kim SW, et al. Brigatinib in patients with crizotinib-refractory anaplastic lymphoma kinase-positive non-small-cell lung cancer: A randomized, multicenter phase II trial. *J Clin Oncol*. (2017) 35:2490–8. doi: 10.1200/JCO.2016.71.5904
17. Shaw AT, Felip E, Bauer TM, Besse B, Navarro A, Postel-Vinay S, et al. Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: An international, multicentre, open-label, single-arm first-in-man phase 1 trial. *Lancet Oncol*. (2017) 18:1590–9. doi: 10.1016/S1470-2045(17)30680-0
18. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, et al. Molecular mechanisms of resistance to first- and second- generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer Discov*. (2016) 6:1118–33. doi: 10.1158/2159-8290.CD-16-0596
19. Christopoulos P, Kirchner M, Bozorgmehr F, Endris V, Elsayed M, Budczies J, et al. Identification of a highly lethal V^{3+}TP53+ subset in ALK+ lung adenocarcinoma. *Int J Cancer*. (2019) 144:190–9. doi: 10.1002/ijc.31893
20. Chen S, Zhou Y, Chen Y, Gu J. FASTP: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. (2018) 34:i884–90. doi: 10.1093/bioinformatics/bty560
21. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. (2010) 26:589–95. doi: 10.1093/bioinformatics/btp698
22. Chen S, Zhou Y, Chen Y, Huang T, Liao W, Xu Y, et al. Gencore an efficient tool to generate consensus reads for error suppressing and duplicate removing of NGS data. *BMC Bioinform*. (2019) 20:606. doi: 10.1186/s12859-019-3280-9
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup: the sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
24. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. (2012) 22:568–76. doi: 10.1101/gr.129684.111
25. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted sequencing. *PLoS Comput Biol*. (2014) 12:e1004873. doi: 10.1371/journal.pcbi.1004873
26. Chen S, Liu M, Huang T, Liao W, Xu M, Gu J. GeneFuse: detection and visualization of target gene fusions from DNA sequencing data. *Int J Biol Sci*. (2018) 14:843–8. doi: 10.7150/ijbs.24626
27. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. (2018) 28:1747–56. doi: 10.1101/gr.239244.118
28. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. (2017) 23:703–13. doi: 10.1038/nm.4333
29. Christopoulos P, Budczies J, Kirchner M, Dietz S, Sültmann H, Thomas M, et al. Defining molecular risk in ALK+ NSCLC. *Oncotarget*. (2019) 10:3093–103. doi: 10.18632/oncotarget.26886
30. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after immunotherapy: analysis of genomic alterations associated with accelerated growth rate. *Clin Cancer Res*. (2017) 23:4242–50. doi: 10.1158/1078-0432.CCR-16-3133

Conflict of Interest: TH, ML, YZ, YLi, WL, JD, and SC were employed by the company HaploX Biotechnology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 12 May 2020
doi: 10.3389/fonc.2020.00729

Co-Occurring Alterations of ERBB2 Exon 20 Insertion in Non-Small Cell Lung Cancer (NSCLC) and the Potential Indicator of Response to Afatinib

Bo Yuan[1†], Jun Zhao[2†], Chengzhi Zhou[3], Xiumei Wang[4], Bo Zhu[5], Minglei Zhuo[2], Xilin Dong[1], Jiemei Feng[6], Cuihua Yi[7], Yunpeng Yang[8], Hua Zhang[9], Wangyan Zhou[10], Zhengtang Chen[5], Sheng Yang[11], Xinghao Ai[12], Kehe Chen[13], Xuefan Cui[14], Difa Liu[15], Chunmei Shi[11], Wei Wu[16], Yanjun Zhang[17], Lianpeng Chang[18], Jin Li[18], Rongrong Chen[18] and Shuanying Yang[1]*

[1] Department of Pulmonary and Critical Care Medicine, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Thoracic Medical Oncology-I, Peking University Cancer Hospital and Institute, Beijing, China, [3] State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [4] Department of Oncology, Inner Mongolia Autonomous Region Cancer Hospital, Hohhot, China, [5] Department of Oncology, Xinqiao Hospital, Chongqing, China, [6] Department of Respiratory Medicine, Guigang City People's Hospital, Guigang, China, [7] Department of Medical Oncology, Qilu Hospital of Shandong University, Jinan, China, [8] Department of Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China, [9] Department of Oncology, Shaanxi Provincial People's Hospital, Xi'an, China, [10] Department of Party Affairs, The First Affiliated Hospital of University of South China, Hengyang, China, [11] Department of Oncology, Fujian Medical University Union Hospital, Fuzhou, China, [12] Lung Tumor Clinical Medical Center, Shanghai Chest Hospital, Shanghai Jiaotong University, Shanghai, China, [13] Department of Oncology, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China, [14] Department of Respiratory Medicine, The First Affiliated Hospital With Nanjing Medical University, Nanjing, China, [15] Department of Oncology, Haian People's Hospital, Nantong, China, [16] Department of Thoracic Surgery, The First Hospital Affiliated to AMU (Southwest Hospital), Chongqing, China, [17] Department of Oncology, Shaanxi Provincial Cancer Hospital, Xi'an, China, [18] Geneplus-Beijing, Beijing, China*

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Yuxiang Zhao,
Biotrans Technology Co. LTD, China
Rossano Lattanzio,
Università degli Studi G. d'Annunzio Chieti e Pescara, Italy

***Correspondence:**
Shuanying Yang
18391861712@163.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *25 December 2019*
Accepted: *16 April 2020*
Published: *12 May 2020*

Citation:
Yuan B, Zhao J, Zhou C, Wang X, Zhu B, Zhuo M, Dong X, Feng J, Yi C, Yang Y, Zhang H, Zhou W, Chen Z, Yang S, Ai X, Chen K, Cui X, Liu D, Shi C, Wu W, Zhang Y, Chang L, Li J, Chen R and Yang S (2020) Co-Occurring Alterations of ERBB2 Exon 20 Insertion in Non-Small Cell Lung Cancer (NSCLC) and the Potential Indicator of Response to Afatinib. Front. Oncol. 10:729. doi: 10.3389/fonc.2020.00729

Background: Human epidermal growth factor receptor 2 (ERBB2, HER-2) exon 20 insertion (ERBB2ex20ins) remains a refractory oncogenic driver in lung cancer. So far there is limited data showing the co-occurring mutation background of ERBB2ex20ins in Chinese lung cancer and its relationship with response to afatinib.

Patients and Methods: A total of 112 Chinese patients with ERBB2ex20ins identified by next-generation sequencing from 17 hospitals were enrolled. The clinical outcomes of 18 patients receiving afatinib treatment were collected.

Results: Among the 112 patients, insertion-site subtypes comprised of A775ins (71%; 79/112), G776indel (17%; 19/112), and P780ins (12%; 14/112). There were 66.1% (74/112) of patients carrying TP53 co-mutation and FOXA1 was the most prevalent co-amplified gene (5.5%, 3/55). The co-occurring genomic feature was similar among three insertional-site subtypes and had an overall strong concordance with the western population from the MSKCC cohort ($R^2 = 0.74$, $P < 0.01$). For the prognosis, patients with co-occurring mutation in cell-cycle pathway especially TP53 showed shorter OS than patients without [median OS: 14.5 m (95% CI:12.7–16.3 m) vs. 30.3 m (95% CI: not reached), $p = 0.04$], while the OS was comparable among three subtypes. For

the response to afatinib, ERBB2ex20ins as a subclonal variant was an independent factor relating to shorter PFS [median PFS: 1.2 m (95% CI: 0.8–1.6 m) vs. 4.3 m (95% CI: 3.3–5.3 m), $p < 0.05$].

Conclusion: Our data revealed co-occurring TP53 represent an unfavorable prognosis of patients with ERBB2ex20ins, emphasizing the more valuable role of the co-mutation patterns than insertion-site subtypes in predicting prognosis of this group of patients. Moreover, the clonality status of ERBB2ex20ins was identified as a potential indicator for response to afatinib.

Keywords: non-small cell lung cancer, ERBB2 exon 20 insertion, co-occurring alterations, afatinib, clonality status

INTRODUCTION

Aberrations in human epidermal growth factor receptor 2 (HER-2, ERBB2) have emerged as oncogenic drivers and therapeutic targets in 1–4% of non-small cell lung cancer (NSCLC) and up to 6% of EGFR/KRAS/ALK-negative lung adenocarcinoma (LUAD) (1, 2). Most of ERBB2 mutations is characterized by inframe insertion occurring at exon 20 in the protein kinase domain (3). Prior studies showed that pan-ERBB family inhibitor (afatinib, dacomitinib) (4, 5), ado-trastuzumab (T-DM1) (6) as well as some new agents such as poziotinib (7), pyrotinib (8) may elicit an objective response in patients with ERBB2 exon 20 insertion (ERBB2ex20ins); however, no therapy has been approved as a standard treatment yet.

Afatinib has been demonstrated its suppressive effect on lung cancer cell lines with ERBB2ex20ins *in vivo* (9). Previous studies also revealed clinical outcomes of afatinib with a 13–19% objective response rate (ORR) and a disease control rate (DCR) around 70% in three separate cohorts (10–12); Nevertheless, there exists profound efficacy heterogeneity on them, such as patients with the same subtype displayed discordant benefits and duration of time.

Several prior studies revealed that genetic co-alterations were independent variables associated with unfavorable prognosis of EGFR-TKIs (13, 14). However, because of its low frequency, researches focused on ERBB2ex20ins have generally been limited to insufficient number of cases from single institution and prevent making a broad assessment of co-existing alteration patterns of ERBB2ex20ins, which may reflect its genomic background heterogeneity and contribute to the variable responsiveness to the targeted therapy. Therefore, making a comprehensive analysis of concomitant mutation spectrum of ERBB2ex20ins in a large cohort and correlating its co-mutation status with prognosis are urgently warranted.

Moreover, growing number of studies are paying attention to the clonality heterogeneity of targetable somatic alterations and adapting the cancer-treatment strategy to taking into account how a tumor evolves (15). It seems that therapy targeted clonal ("trunk") mutations may be more effective than targeted subclonal ("branch") ones (16, 17). Nevertheless, how the clonality status of driver aberrations modulates the efficacy of therapy is unclear.

Using the next-generation sequencing (NGS) method, we here described the co-occurring molecular spectrum of ERBB2ex20ins in a cohort of 112 NSCLC patients from 17 hospitals in China. We also compared our spectrum with the western population from Memorial Sloan Kettering Cancer Center (MSKCC) and investigated the impact of co-mutation status on the prognosis of them. Furthermore, we retrospectively assessed the efficacy and tried to identify efficacy predictive factors of afatinib in 18 patients with ERBB2ex20ins.

MATERIALS AND METHODS

Patient Cohort and Clinical Data Collection

We retrospectively screened 112 patients (from 17 hospitals) harboring ERBB2ex20ins in a College of American Pathologists (CAP) Laboratory (Geneplus-Beijing, Beijing, People's Republic of China) from July 2016 to December 2018. Samples of tumor tissue, plasma or effusion were analyzed by next-generation sequencing (NGS) assay using two versions (59 or 1,021 cancer-related genes) of capture-based targeted sequencing panel. Gene lists of two versions of sequencing panel are shown in **Table S1**. The sample type and panel for each patient are shown in **Figure 1A**. A total of 55 and 57 samples were sequenced using 1,021 or 59-gene panel, respectively. Clinicopathological features were abstracted from the accompanying pathology report submitted by the ordering physician. All patients provided written informed consent for our study. This study was approved by the institutional review board of The Second Affiliated Hospital of Xi'an Jiaotong University and all participating hospital.

The sequencing data of the Memorial Sloan Kettering Cancer Center (MSKCC) Cohort was downloaded from an open-access database named the Cancer Genome Atlas Database, which is publicly available at http://www.cbioportal.org [MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)] (18, 19). The data of overall survival (OS) was acquired from the cbioportal website directly. OS was measured from the date when the tumor specimen was collected to the date of death or last follow-up visit (20).

Response Evaluation

The clinical outcomes of 18 patients treated with afatinib were collected by each contributing doctor in charge and pooled for

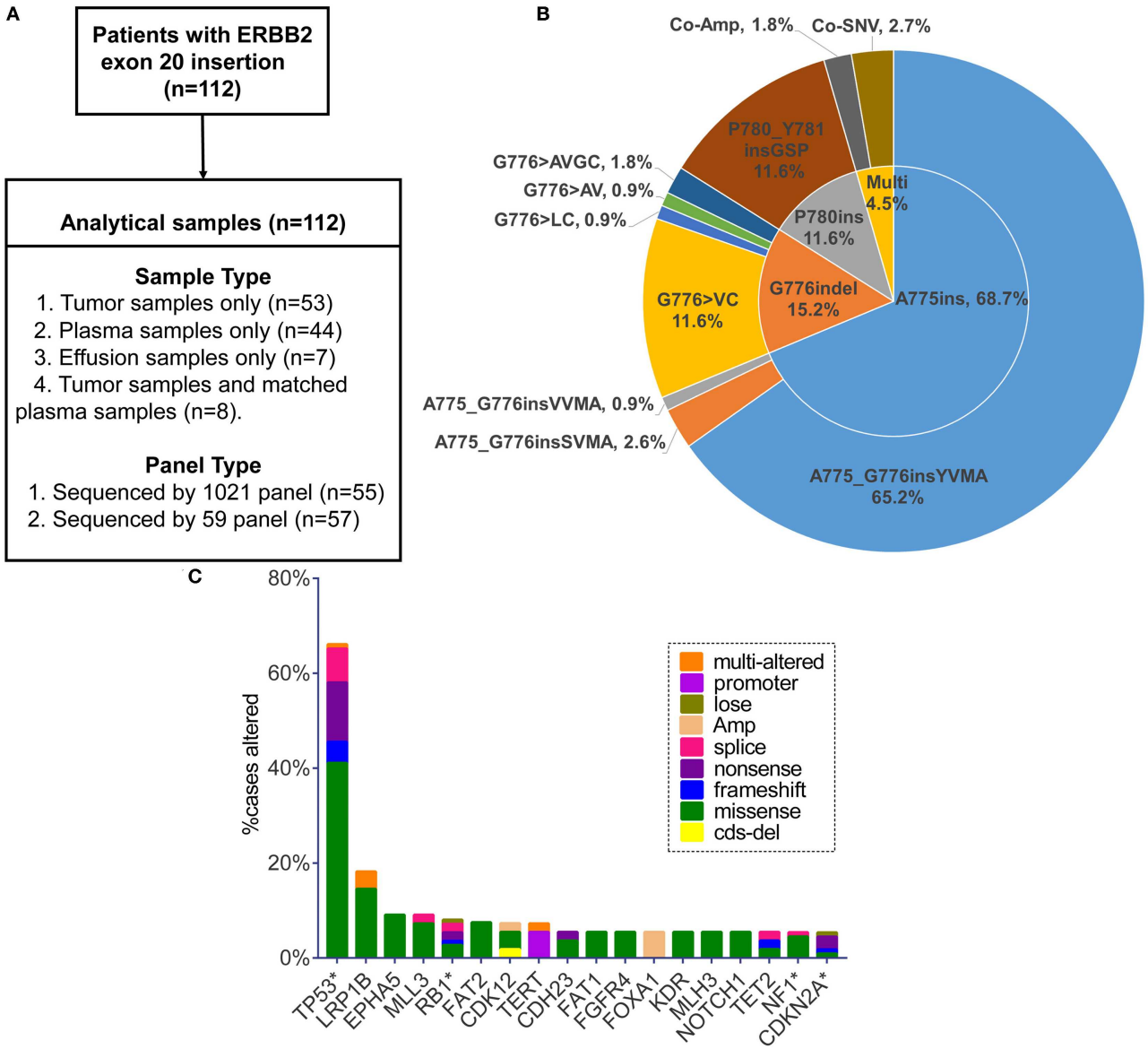


FIGURE 1 | (A) Sample types and sequencing panel details of study design. **(B)** Pie chart visualizing eight specified insertion subtypes combined with insertion site. Multi, multiple alterations; Co-Amp, co-amplification; Co-SNV, co-single nucleotide variant. **(C)** Top 18 genes in the highest co-occurring frequency with ERBB2 exon 20 insertion (no relation to the total numbers analyzed). Only the genes with concurrent frequency over 5% are shown. *The genes included in the 59-gene panel.

analysis. Patients were administered afatinib depending on their performance status and other comorbidities at a starting dose of 30, 40, or 50 mg daily. Best response evaluation was assessed according to the Response Evaluation Criteria in Solid Tumors (RECIST, v1.1). The progression free survival (PFS) for afatinib treatment was defined as the time from the start of afatinib treatment to the date of disease progression or death.

DNA Extraction

Circulating DNA and Genomic DNA for genomic testing were isolated from 3 ml of plasma or effusion and FFPE samples, respectively. Peripheral blood lymphocytes (PBL) DNA were extracted for germline reference (**Supplemental Online Methods**).

Target Capture and Next-Generation Sequencing

KAPA Library Preparation Kit (Kapa Biosystems, Wilmington, MA, USA) was applied to prepare Indexed Illumina NGS libraries from peripheral blood lymphocytes (PBL) DNA, and tumor DNA or plasma DNA according to the manufacturer's protocol. Capture probes were designed to cover coding sequences or hot exons of 59 or extended 1,021 genes that are frequently mutated in NSCLC and other common solid tumors (details of sequencing

region for each gene are uploaded in **Table S1**). Libraries were hybridized to custom-designed biotinylated oligonucleotide probes (Integrated DNA Technologies, Iowa, IA, USA). DNA sequencing was performed on the HiSeq 3000 Sequencing System (Illumina, San Diego, CA) with 2 × 101 bp paired-end reads.

Sequencing Data Analysis

Terminal adaptor sequences and low-quality data were removed from the raw data. The BWA (0.7.12-r1039) was employed to align clean reads to the reference human genome (hg19) (21). MuTect2 (3.4-46-gbc02625) and GATK was applied to call single nucleotide variants (SNVs) and small insertions and deletions (Indels), respectively. Somatic copy number variants (CNVs) were identified using CONTRA (2.0.8) (22). Moreover, we employed the NoahCare Tool Kit using NCfilter (software developed by self, version 1.5.0) for fastq data QC, NCbamInfo (version 0.2.0) for alignment QC, NCanno (version 0.1.1) for annotation with multiple databases, and NChot (version 0.1.0) for hotspot region variant review and recall. All final candidate variants were all manually verified using the Integrative Genomics Viewer (IGV) Browser.

Clonality Analysis

The subclonal architecture of all DNA samples were constructed by PyClone run with 20,000 interactions and default parameters (23). Variants were clustered as previously described (23), briefly, the copy number information of each SNV was used as input for PyClone analysis (24, 25), and the cancer cell fraction (CCF) was inferred. Variants located in the cluster with greatest mean CCF were defined as clonal, the rest were subclonal (23).

Statistical Analysis

All statistical analyses were performed using SPSS version 19.0 (SPSS Company, Chicago, IL). Categorical and continuous variables were compared by Fisher's exact test and Kruskal-Wallis *H*-test, respectively. The Pearson correlation coefficient was applied to assess the linear correlation degree of co-occurring genes' frequency appearing in Our Cohort and MSKCC Cohort. The OS and PFS were estimated using the Kaplan-Meier method and compared with the log-rank test. A multi-variant regression model was calculated for HRs and 95% CIs. All statistical tests were two-sided, and $p < 0.05$ was defined as statistical significance.

RESULT

Patients With ERBB2 exon 20 Insertion

One hundred and twelve patients carrying ERBB2ex20ins were screened from July 2016 to December 2018 (**Figure 1A**). The clinical characteristics for these patients were summarized in **Table 1**. In all, patients were predominantly in the stage IV (80/112, 72%) and had the histology of adenocarcinoma (68%,76/112). There were slightly more female (54%; 60/112) than male (46%; 52/112), with a median age of 61.5 years (range: 28–87 years).

Totally, eight specific insertion types of ERBB2ex20ins were identified. Considering the fact that certain studies discussing the efficacy of targeted therapy for ERBB2ex20ins are always based on the different insertion sites, we classified them into three subtypes. The most common subtype was four amino acids insertion at codon 775 (A775ins; 70.5%), followed by insertion combined with deletion occurring at codon 776 (G776indel; 17.0%), and three amino acids insertion at codon 780 (P780ins; 12.5%). Multi-alterations were present in five patients, with two patients harboring concurrent ERBB2 amplification and three patients carrying ERBB2ex20ins with ERBB2 single nucleotide variant (SNV) referring to p.A775_G776insYVMA+ p.R897Q, p.P780_Y781insGSP+ p.G519R, and p.G776delinsAVGC+ p.G776A (**Figure 1B**).

Moreover, insertion-subtype abundance was not significantly different among the sample types ($p = 0.41$; **Table S2**). It indicated that distinct sample type was not biased toward the detection of certain ERBB2 subtype.

TABLE 1 | Clinical characteristics of patients with ERBB2 exon 20 insertion in different positions.

Characteristics	A775ins (*n* = 79)	G776indel (*n* = 19)	P780ins (*n* = 14)	Sum	*P*-value
Age at initial diagnosis					
Median (range)	62 (28–83)	58 (29–87)	64 (48–83)	61.5 (28–87)	0.425
Unknown	6	0	2	8	
Gender					
Female	39 (49%)	11(58%)	10 (71%)	60 (54%)	0.287
Male	40 (51%)	8 (42%)	4 (29%)	52 (46%)	
Histology					
NSCLC NOS	14 (18%)	3 (16%)	1 (7%)	18 (16%)	nc
Adenocarcinoma	55 (70%)	11(58%)	10 (72%)	76 (68%)	
Squamous carcinoma	0	0	1(7%)	1(1%)	
Unknown	10 (12%)	5 (26%)	2 (14%)	17 (15%)	
Stage					
I–III	15 (19%)	2 (10%)	2 (14%)	19 (16%)	0.850
IV	55 (70%)	14 (74%)	11 (79%)	80 (72%)	
Unknown	9 (11%)	3 (16%)	1 (7%)	13 (12%)	

P-values are calculated with Fisher's exact test except for age using the Kruskal-Wallis H-test.

ins, insertion; indel, insertion and deletion; NSCLC NOS, non-small cell lung cancer not other specified; nc, not calculate.

Co-occurring Genomic Profile of ERBB2 exon 20 Insertion

Spectrum of Co-occurring Alterations and Characteristics in Different Insertion Sites

On the basis of 59 genes strongly associated with cancer, 80.4% (90/112) of patients had at least one additional alteration, with 48.9% (44/90) of them carrying one and 27.8% (25/90) carrying two. Three or more concomitant alterations were present at the rest of 23.3% (21/90) patients. TP53 was the most frequent gene co-mutant with ERBB2ex20ins, making up 66.1% (74/112) cases, with predominant alteration type of missense mutation (63.5%, 47/74), concentrating on exon 5, 8, 6, 7 (76.7%, 56/74; range: exon

4-exon 11) (**Figure S1A**, **Table S3**). The remaining prevalent co-occurring genes were LRP1B (18.2%, 10/55), EPHA5 (9.1%, 5/55), MLL3 (9.1%, 5/55), and RB1 (8.0%; 9/112) (no relation to the total numbers analyzed). FOXA1 appeared in 5.5% (3/55) of patients and became the most common co-mutant gene in the form of amplification (**Figure 1C**). Putative driver aberrations including EGFR (L858R or 19del), ROS1 fusions, ALK receptor tyrosine kinase gene (ALK) rearrangement, KRAS, BRAF (V600E) were not found in this cohort, probably mutually exclusive from ERBB2ex20ins.

Of the pathway level, we classified the co-mutant genes according to the pathway involved. 86.7% (78/90) of patients, who carried at least one additional mutation, had the co-altered genes enriched in the cell cycle, followed by receptor tyrosine kinase/growth factor signaling (RTK) (15.2%) and DNA damage/repair (8.9%) (**Figure 2B**). Furthermore, some patients had co-mutant genes involved multiple important pathways simultaneously, while some patients had more than one co-mutant genes involving one single pathway (**Table S4**).

We also explored the co-occurring alteration feature among A775ins ($n = 79$), G776indel ($n = 19$), and P780ins ($n = 14$). No substantial discrepancy was observed among the three groups at either the co-occurring somatic alterations (only TP53 was included in the analysis) or pathway enriched, the clinical details were as well (**Figures S2A,C**, **Table 1**). The location and exon distribution of TP53 mutation were comparable in three insertion-site subtypes (**Figures S1B–E**); however, when considering TP53 mutation types, there was a tendency that G776indel may be less adept at co-occurring with TP53 missense mutations, with no statistically significant ($p = 0.06$; **Figure S2B**).

Spectrum Comparison With the Western Population From the MSKCC Cohort

Next, we compared our data with the findings previously reported by the MSKCC, which included 1,563 tumor specimens from patients with NSCLC. Totally 30 patients harboring ERBB2ex20ins involved in this cohort.

Overall, both the proportion of three subtypes and the whole molecular co-occurring mutation spectrum (genes that co-altered at a frequency over 5% in each cohort) were similar between the two cohorts ($R^2 = 0.74$, $P < 0.01$), although the co-mutant frequencies of certain genes were higher in the MSKCC cohort than in ours(**Figure 2A**, **Table S5**). Notably, these genes in slightly higher frequency were in the form of copy number variant (CNV) (**Figure S2D**), and it is probably caused by the low detection rate of CNV due to the mixed plasma samples in our samples.

For the pathway analysis, the enrichment of each pathway in our cohort was in accord with the MSKCC cohort, with a slightly higher frequency of cell-cycle pathway enriched in MSKCC cohort (36.7 vs. 17.9%, $P = 0.044$) on account of the higher frequency of CDKN2A alteration (**Figure 2B**, **Table S4**).

Impact of Insertional-Site Subtypes and Co-occurring Mutational Status on OS

Based on the complete overall survival (OS) from the MSKCC cohort, prognosis impact of ERBB2 insertion-site subtypes and genes co-occurring over 5 or more cases in either cohort were evaluated. Statistic descriptive of co-occurring genes included in the analysis were summarized in **Figure S3A**. There was a trend that patients harboring co-occurring genes enriched in cell-cycle pathway showed a worse survival, with no significantly statistic difference ($p = 0.059$; **Figure S3B**). However, worse overall survival was seen in patients with co-mutations in TP53 [median OS:14.5 m (95% CI:12.7–16.3 m) vs. 30.3 m (95% CI: not reached); log-rank test], while OS was not significantly different among three subtypes ($p = 0.72$, **Figures 2C,D**).

Prior study revealed TP53 mutation in exons 5, 7, 8, and 9 sharing a better prognosis than other sites in the advanced NSCLC (26). In this regard, we investigate the prognosis value of co-occurring TP53 mutation in exons 5, 7, 8, and 9 and whether they can be even more relevant in a specific subgroup of patients with ERBB2ex20ins mutation (i.e., A775ins, G776indel, and P780ins subgroups) but found negative result (log rank, $p = 0.095$; Fisher exact test, $p = 0.427$; **Figures S3C,D**).

Clinical Outcomes of Afatinib for Patients Harboring ERBB2ex20ins

Afatinib Treatment Efficacy Overview

The basic clinical and molecular characteristics of 18 patients treated with afatinib were summarized in **Table 2**. Nearly all of patients were in the advanced stage and 61.1% (12/18) of patients receiving afatinib as 2 line or more.

Of the 18 patients, tumor remission data according to RECIST 1.1 criteria were available for 15 patients. Among them, 5 patients achieved PR (33.3%) and 4 patients achieved SD (26.7%). All PRs were, respectively, observed in 3 separate insertion subtypes, whereas the patients with PD only involved in the subtype of A775ins. The median time on treatment with afatinib was 3.7 months (95% CI: 2.1–5.3 m; range: 0.7–13.4 m). The median duration time for patients responding to afatinib was 4.5 months (95% CI: 3.6–5.4 m; range: 2.5 m−13.4 months). The response details and duration of response (DoR) of afatinib for each patient were showed in **Figure 3A**. One patient harboring G776delinsVC was treated for afatinib as first-line therapy with PR for over 13.4 months and didn't achieve disease progression until the last follow-up in this study. As for the rest of patients, 3 cases (16.7%) responding to afatinib had a DoR over 6 months.

Impact of Clonality Status and Co-occurring Mutations of ERBB2ex20ins on Afatinib Treatment Outcome

For 18 patients treated with afatinib, we identified 54 somatic SNVs, 4 CNVs and 18 somatic indels in 18 samples, for an average of 4.3 somatic variants per sample. In this regard, we applied method PyClone to evaluate whether the ERBB2ex20ins carried by the patients were clonal or subclonal mutations.

Result revealed that ERBB2ex20ins as subclonal variants was significantly associated with shorter PFS of afatinib [median PFS: 1.2 m (95% CI: 0.8–1.6 m) vs. 4.3 m (95% CI: 3.3–5.3 m), $p < 0.01$], while co-occurring TP53 mutation and insertion-site subtypes had no significant impact on the efficacy of afatinib. This

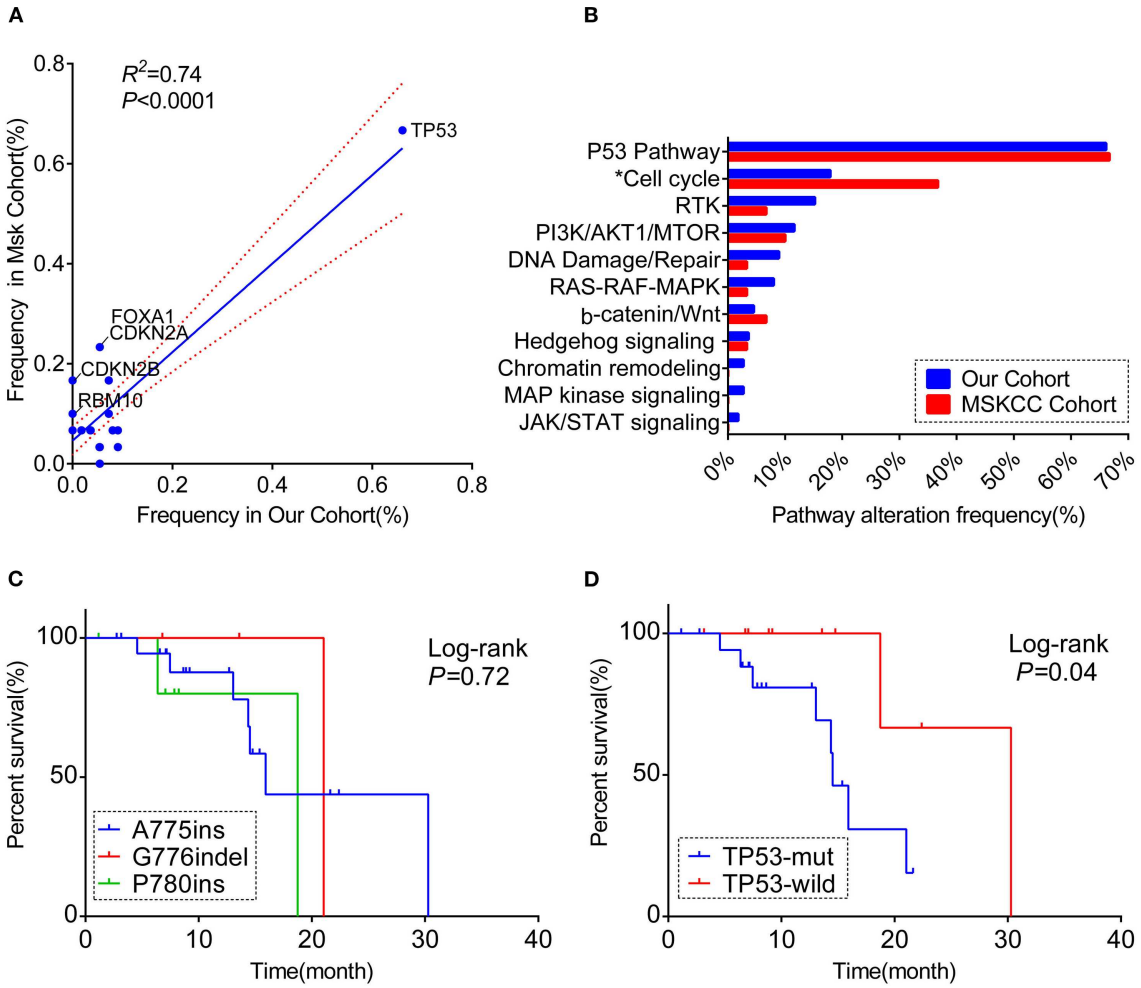


FIGURE 2 | A comprehensive comparison of the co-occurring profile between our cohort and MSKCC cohort for the frequency of **(A)** totally matched 28 genes (the genes included in the analysis were matched in both cohorts' panel with the frequency over 5%; the genes labeled are significantly different between the two cohorts). **(B)** Pathway enriched (Only genes included in 59-gene panel were analyzed; *the pathway significantly different between the two cohorts). RTK, receptor tyrosine kinase/growth factor signaling. **(C)** Kaplan–Meier curve showing the difference of median overall survival (OS) among patients harboring three different ERBB2 insertion subtypes. **(D)** Kaplan–Meier curve visualizing the effect of TP53 alteration on OS.

result remained significant when adjusted for ERBB2 insertion subtypes, TP53 mutation, TP53 missense mutation and no. line of afatinib [HR: 0.025 (0.002–0.41); $p = 0.01$)] (**Figure 3B**, **Table S6**).

Next, we divided the patients into durable clinical benefit (DCB) cohort and no durable benefit (NDB) cohort (The DCB was defined as the patients achieving PR or SD and having the duration of PFS for over 3 months; the NDB referred to the patients having the PFS <3 months). Clinical baseline parameters for two groups of patients are displayed in **Table 2** and basically similar among each parameter. The variables including concurrent TP53 mutation and TP53 missense mutation were not significantly different between the two groups (data not show, Fisher's exact test, $p > 0.05$) except for the ERBB2ex20ins clonality status (Fisher exact test, $p < 0.01$; **Figure 3C**).

Dynamic Detection and Afatinib Resistance

To gain some insight into the potential mechanism upon afatinib resistance, we analyzed two patients conducting NGS at two time points in the course of afatinib treatment. The clinical characteristics and test details for two patients were summarized in **Table S7**. Both patients were non-smoking female. For the Patient#1, ERBB2 amplification [copy number (CN) = 3.1] occurred in the repeat biopsy sample upon the progression of afatinib. Similarly, the patient#2 also presented the ERBB2 amplification (CN = 2.74) which was undetected in the initial plasma sample after taking afatinib for half of a month, unfortunately, we did not examine the ERBB2 CN status at the time of progression on afatinib. Despite this, it can be speculated that ERBB2 amplification may represent a potential resistance mechanism of afatinib.

TABLE 2 | Clinical and molecular characteristics of patients treated with afatinib.

Characteristics	NCB ($n = 8$)	DCB ($n = 10$)	Sum ($n = 18$)	*P*-value
Age at initial diagnosis, years				
Median (range)	56.5 (40–75)	54 (29–69)	55.5 (29.75)	
<65	5	9	14 (77.8%)	0.28
≥65	3	1	4 (22.2%)	
Gender				
Female	5	5	10 (55.6%)	0.66
Male	3	5	8 (44.4%)	
Tobacco use				
Never	5	4	9 (50.0%)	0.15
Former or current	3	2	5 (27.8%)	
Unknown	0	4	4 (22.2%)	
Histology				
Adenocarcinoma	8	10	18 (100.0%)	
Brain metastasis				
Yes	3	4	7 (38.9%)	0.34
No	3	6	9 (50.0%)	
NA	2	0	2 (11.1%)	
Tumor stage				
IIIa	1	0	1 (5.6%)	nc
IV	6	10	16 (88.8%)	
Unknown	1	0	1 (5.6%)	
No. line of afatinib treatment				
1	3	3	6 (27.8%)	1.00
≥2	5	7	12 (61.1%)	
ERBB2ex20ins subtypes				
A775 insertion	6	7	13 (72.2%)	1.00
G776 indel	1	2	3 (16.7%)	
P780 insertion	1	1	2 (11.1%)	
Concurrent TP53 alteration				
Yes	7	6	13 (72.2%)	0.31
No	1	4	5 (27.8%)	
Concurrent TP53 missense mutation				
Yes	6	4	10 (55.6%)	0.19
No	2	6	8 (44.4%)	

DCB, durable clinical benefit; NCB, no durable benefit; nc, not calculate.

DISCUSSION

In this study, we delineate the co-occurring alterations and common pathway involved addicted to ERBB2ex20ins in a representative NSCLC cohort of 112 patients and correlate co-mutation patterns with the prognosis of patients harboring ERBB2ex20ins. Moreover, to our knowledge, we present the first time to examine the impact of clonality status of oncogenic drivers in relation to the efficacy of targeted therapy.

The recent widespread use of NGS enables us to move researches from concentrating solely on the driver gene to the full view of genomic co-alterations, which may have prognostic implications. To date, somatic mutations in TP53 are the most prevalent co-mutation in EGFR-mutant lung adenocarcinoma (LUAD) with a frequency of 54.6–64.6% and several studies have identified TP53 co-alteration as a negative prognosis marker, with consistently predicting worse clinical outcomes receiving EGFR-TKI therapy (27). In our study, TP53 ranked as the most common accompanying somatic altered gene with a frequency of 66%; this frequency was slightly higher than the previously reported 51.6% (10), possibly due to the fewer proportion of female in our cohort (28). Our results showed that patients had a worse OS when co-occurring mutation in TP53, which is also validated in a previous study (29). Recently, the different prognosis value was recognized in distinct exons and alteration types of TP53 mutation and the results were inconsistent. Exons 5, 7, 8 and 9 were reported to share a better prognosis than other sites (26); it is worth mentioning that the study referred here sought to reveal the prognostic value of TP53 alterations in advanced NSCLC compared to most of studies limited to the early stage or EGFR-mutant background (30, 31). Unfortunately, TP53 mutations in exons 5, 7, 8, and 9 did not produce more favorable prognosis than other sites in the advanced NSCLC patients carrying ERBB2 mutation. Clinicopathological characteristics and treatment status should be well-defined to clearly investigate the prognosis impact of various TP53 exons. Moreover, there is a tendency that G776indel subtype may be less adept at co-occurring with TP53 missense mutations, however, whether this characteristic will have a beneficial effect on the prognosis for them remains to be explored. Interestingly, we found neither the co-mutant frequency of TP53 nor pathway enrichment was significantly different among three insertion subtypes, and the OS was comparable as well. For the clinical practice, we suppose that the co-occurring mutation status may have greater impact on the prognosis for this subset of patients than the insertion subtypes itself. Moving forward, the study highlighted multiple concurrent mutations besides ERBB2 insertion subtypes should be tested prospectively in order to provide better predictions of prognosis for them.

In order to systematically understand the co-mutation profile of ERBB2ex20ins, we cataloged co-altered genes based on existing biological pathway knowledge and the cell-cycle (86.7%), receptor tyrosine kinase/growth factor signaling (RTK) (15.2%) and DNA Damage/Repair (8.9%) showed predominance among all the involved pathways. Prior study found that cell-cycle and DNA-damage response pathway are involved in leptomeningeal metastasis of NSCLC (32). This finding was somehow inter-correspondent with the likely unfavorable prognosis for the patients with the co-occurring genes enriched in the cell-cycle pathway in our study, although the survival discrepancy was not significant maybe on account of the insufficient follow-up time or limited sub-group sample size. Unfortunately, we cannot collect the detailed metastasis status for each patient; however, for the patients treated with afatinib, we found seven in 16 of patients having brain metastasis. Moreover, a recent retrospective study found patients carrying ERBB2 mutations in lung cancer developed more brain metastases on treatment than patients with KRAS mutations [28 vs. 8%; hazard ratio (HR),

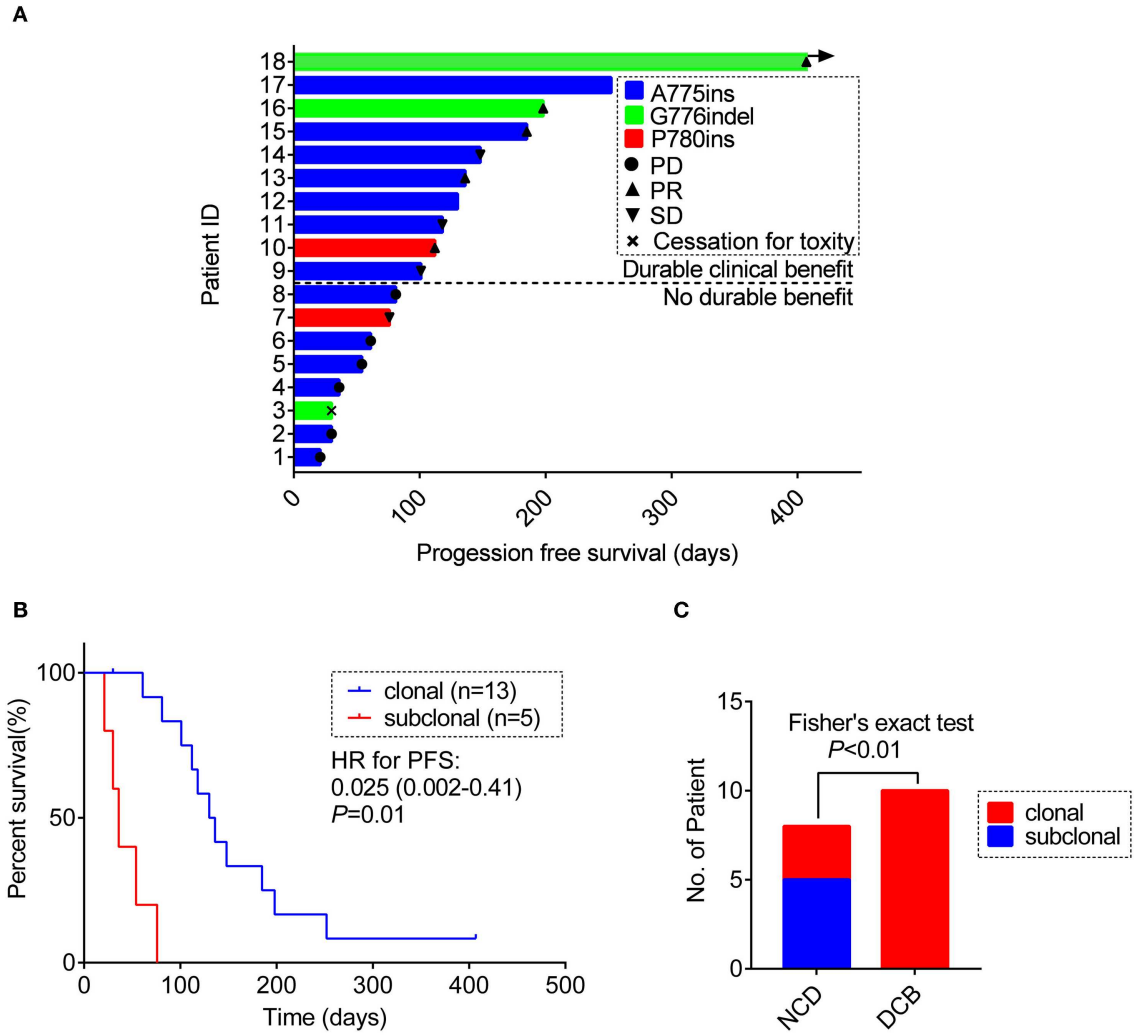


FIGURE 3 | (A) Swimming plot visualizing the response details for afatinib in each patient ($n = 18$). **(B)** Kaplan-Meier comparing PFS for ERBB2ex20ins as clonal or subclonal variant (p-values determined by multi-variant regression analysis and HR with 95% CI are shown). **(C)** The distribution of ERBB2 clonality status between the DCB group and the NDB group. DCB, durable clinical benefit; NDB, no durable benefit.

5.2; $p < 0.001$] and trended more than patients with EGFR mutations [28 vs. 16%; HR, 1.7; $p = 0.06$; (33)]. These findings may underline the central nervous system (CNS) surveillance practices in patients with ERBB2 alterations and the urgent need for the development of novel HER2-targeted agents with active efficacy in the CNS.

The co-occurring genomic spectrum of ERBB2ex20ins in our cohort of Chinese people had an overall strong concordance with the MSKCC cohort from the United States ($R^2 = 0.74$, $p < 0.01$). In a retrospective study collated two cohorts of patients with ERBB2 alteration from the MSKCC and Guangdong General Hospital, they also found great consistency with each other in the aspect of the prevalence and baseline clinical parameters of patients possessing ERBB2 mutation (34). These findings, on the one hand, can be supporting evidence for U.S.-China collaborations in clinical trials to accelerate new drug development for this infrequent mutation; on the one hand, highlight the robustness of our results.

An important aspect of our study is that we found the clonality status of ERBB2ex20ins was an independent potential indicator for response to afatinib. It is well-known that there exists substantial intratumor heterogeneity and tumor evolves in a trunk-branch model. The "trunk" mutation (clonal mutation) was known as taking place in the early development of cancer and expected to present in every tumor subclone and region, whereas the mutation defined as "branch" would present in a certain fraction of tumor cells and regions (17). Thus, alterations closer to the clonal variant were associated with numerical greater variant allele frequency (VAF). Driver mutations in lung cancer can occur both clonally and subclonally (35). Rachiglio et al. (36) reported that patients harboring EGFR alteration in a lower VAF presented shorter PFS than not,

to some extent, this reflecting the clonality status may affect the efficacy of TKIs. These lead us to speculate that other small molecule TKIs in other molecularly defined cohorts may be even more efficacious when targeting the driver mutation as the clonal variants. In this study, however, we should make this result conclusive with caution due to the limited sample size, further exploration in a large cohort named DARWIN trial (Deciphering Anti-tumor Response With intratumor Heterogeneity; Clinical Trials No. NCT02183883) is ongoing (37, 38).

ERBB2 amplification has been identified as a resistance mechanism induced upon treatment with erlotinib or gefitinib (39, 40) and was observed in 12% of tumor samples obtained from patients at resistance to EGFR TKI therapy (41); however, its role in afatinib resistance is unclear. Chuang et al. (42) reported a patient carrying ERBB2ex20ins, whose plasma samples were obtained upon progression on her initial chemotherapy, erlotinib and afatinib, and results showed that the ERBB2 copy number (CN) level increased over time. In our study, we also found two patients treated with afatinib acquired ERBB2 CN gain after taking afatinib for half of a month and upon gained resistance; notably, for the first patient, the biopsy from the initial lesion was taken, respectively, before and upon progression on afatinib, which makes the result more reliable. This makes us speculate that the patients carrying combined ERBB2 mutation and amplification may be less benefit from afatinib. Further basic research may explore this hypothesis.

Admittedly, our study exists several limitations. Firstly, the sequencing panel in the cohort is not uniform, which impeded us making the deep understanding of co-occurring landscape of ERBB2ex20ins, and we can only analyze co-mutant feature of TP53 among the three insertion-site subgroups. Large-panel NGS should be conducted uniformly in further studies when enrolling patients for the research. Another limitation is due to its retrospective nature, a small sample size of the study, selected bias and various imaging intervals are inevitable in the process of assessing the clinical outcomes of afatinib; nonetheless, the ORR and PFS of afatinib treatment was almost correlated well with prior studies. Furthermore, since we used single tumor sample taken at a one-time point in the disease course, we may not verify the true clonality status of each mutation; however, this single-point samples may be more likely to underestimate the true extent of heterogeneity within tumors rather than distinguishing clonal from subclonal variants. Importantly, although multifocal or repeated tumor biopsies is better for tracking the true evolution process of tumor development, single sampling may be easier to achieve in the clinical practice.

In summary, our data revealed co-occurring TP53 represent an unfavorable prognosis of patients with ERBB2ex20ins, highlighting the greater impact of the co-mutation patterns than insertion-site subtypes on the prognosis of this group of patients. Furthermore, our clinical outcome data for afatinib confirmed its certain efficacy for patients with ERBB2ex20ins and suggested the clonality status of ERBB2 mutation may be a potential indicator of response to afatinib.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: European Variation Archive (EVA) (www.ebi.ac.uk/eva) (PRJEB37583).

ETHICS STATEMENT

The study involving human participants were reviewed and approved by the institutional review board of The Second Affiliated Hospital of Xi'an Jiaotong University and all participating hospitals (including Peking University Cancer Hospital and Institute, The First Affiliated Hospital of Guangzhou Medical University, Inner Mongolia Autonomous Region Cancer Hospital, Xinqiao Hospital, Guigang City People's Hospital, Qilu Hospital of Shandong University, Sun Yat-sen University Cancer Center, Shaanxi Provincial People's Hospital, The First Affiliated Hospital of University of South China, Fujian Medical University Union Hospital, Shanghai Chest Hospital, The People's Hospital of Guangxi Zhuang Autonomous Region, The First Affiliated Hospital with Nanjing Medical University, Haian People's Hospital, The First Hospital Affiliated to AMU (Southwest Hospital), Shaanxi Provincial Cancer Hospital).

AUTHOR CONTRIBUTIONS

ShuY, LC, BY, and JZ designed the study. BY, JZ, JL, and RC contributed to data analysis. BY and JZ prepared the manuscript of the study. CZ, XW, BZ, MZ, XD, JF, CY, YY, HZ, WZ, ZC, SheY, XA, KC, XC, DL, CS, WW, and YZ collected the samples and assembled the clinical data. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 81672300) and the Basic Research Plan of Natural Science of Shaanxi Province (Grant No. 2018ZDXM-SF-042).

ACKNOWLEDGMENTS

We thank the Geneplus-Beijing Institute (Beijing, People's Republic of China) for technical support and bioinformatic analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00729/full#supplementary-material

REFERENCES

1. Arcila ME, Chaft JE, Nafa K, Roy-Chowdhuri S, Lau C, Zaidinski M, et al. Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. *Clin Cancer Res*. (2012) 18:4910–8. doi: 10.1158/1078-0432.CCR-12-0912
2. Mazieres J, Peters S, Lepage B, Cortot AB, Barlesi F, Beau-Faller M, et al. Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. *J Clin Oncol*. (2013) 31:1997–2003. doi: 10.1200/JCO.2012.45.6095
3. Shigematsu H, Takahashi T, Nomura M, Majmudar K, Suzuki M, Lee H, et al. Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res*. (2005) 65:1642–6. doi: 10.1158/0008-5472.CAN-04-4235
4. Mazieres J, Barlesi F, Filleron T, Besse B, Monnet I, Beau-Faller M, et al. Lung cancer patients with HER2 mutations treated with chemotherapy and HER2-targeted drugs: results from the European EUHER2 cohort. *Ann Oncol*. (2016) 27:281–6. doi: 10.1093/annonc/mdv573
5. Kris MG, Camidge DR, Giaccone G, Hida T, Li BT, O'Connell J, et al. Targeting HER2 aberrations as actionable drivers in lung cancers: phase II trial of the pan-HER tyrosine kinase inhibitor dacomitinib in patients with HER2-mutant or amplified tumors. *Ann Oncol*. (2015) 26:1421–7. doi: 10.1093/annonc/mdv186
6. Li BT, Shen R, Buonocore D, Olah ZT, Ni A, Ginsberg MS, et al. Ado-trastuzumab emtansine for patients with HER2-mutant lung cancers: results from a phase II basket trial. *J Clin Oncol*. (2018) 36:2532–7. doi: 10.1200/JCO.2018.77.9777
7. Robichaux JP, Elamin YY, Tan Z, Carter BW, Zhang S, Liu S, et al. Mechanisms and clinical activity of an EGFR and HER2 exon 20-selective kinase inhibitor in non-small cell lung cancer. *Nat Med*. (2018) 24:638–46. doi: 10.1038/s41591-018-0007-9
8. Wang Y, Jiang T, Qin Z, Jiang J, Wang Q, Yang S, et al. HER2 exon 20 insertions in non-small-cell lung cancer are sensitive to the irreversible pan-HER receptor tyrosine kinase inhibitor pyrotinib. *Ann Oncol*. (2019) 30:447–55. doi: 10.1093/annonc/mdy542
9. Suzawa K, Toyooka S, Sakaguchi M, Morita M, Yamamoto H, Tomida S, et al. Antitumor effect of afatinib, as a human epidermal growth factor receptor 2-targeted therapy, in lung cancers harboring HER2 oncogene alterations. *Cancer Sci*. (2016) 107:45–52. doi: 10.1111/cas.12845
10. Liu Z, Wu L, Cao J, Yang Z, Zhou C, Cao L, et al. Clinical characterization of ERBB2 exon 20 insertions and heterogeneity of outcomes responding to afatinib in Chinese lung cancer patients. *Onco Targets Ther*. (2018) 11:7323–31. doi: 10.2147/OTT.S173391
11. Peters S, Curioni-Fontecedro A, Nechushtan H, Shih JY, Liao WY, Gautschi O, et al. Activity of afatinib in heavily pretreated patients with ERBB2 mutation-positive advanced NSCLC: findings from a global named patient use program. *J Thorac Oncol*. (2018) 13:1897–905. doi: 10.1016/j.jtho.2018.07.093
12. Lai WV, Lebas L, Barnes TA, Milia J, Ni A, Gautschi O, et al. Afatinib in patients with metastatic or recurrent HER2-mutant lung cancers: a retrospective international multicentre study. *Eur J Cancer*. (2019) 109:28–35. doi: 10.1016/j.ejca.2018.11.030
13. Blakely CM, Watkins TBK, Wu W, Gini B, Chabon JJ, McCoach CE, et al. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. *Nat Genet*. (2017) 49:1693–704. doi: 10.1038/ng.3990
14. Hong S, Gao F, Fu S, Wang Y, Fang W, Huang Y, et al. Concomitant genetic alterations with response to treatment and epidermal growth factor receptor tyrosine kinase inhibitors in patients with egfr-mutant advanced non-small cell lung cancer. *JAMA Oncol*. (2018) 4:739–42. doi: 10.1001/jamaoncol.2018.0049
15. Willyard C. Cancer therapy: an evolved approach. *Nature*. (2016) 532:166–8. doi: 10.1038/532166a
16. Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*. (2014) 25:91–101. doi: 10.1016/j.ccr.2013.12.015
17. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med*. (2012) 4:127ps10. doi: 10.1126/scitranslmed.3003854
18. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Dis*. (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095
19. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. (2013) 6:pl1. doi: 10.1126/scisignal.2004088
20. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. (2017) 23:703–13. doi: 10.1038/nm.4333
21. Li H, Durbin R. Fast and accurate short read alignment with burrows-Wheeler transform. *Bioinformatics*. (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
22. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics*. (2012) 28:1307–13. doi: 10.1093/bioinformatics/bts146
23. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*. (2014) 11:396–8. doi: 10.1038/nmeth.2883
24. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med*. (2017) 376:2109–21. doi: 10.1056/NEJMoa1616288
25. Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun*. (2015) 6:8760. doi: 10.1038/ncomms9760
26. Jiao XD, Qin BD, You P, Cai J, Zang YS. The prognostic value of TP53 and its correlation with EGFR mutation in advanced non-small cell lung cancer, an analysis based on cBioPortal data base. *Lung Cancer*. (2018) 123:70–5. doi: 10.1016/j.lungcan.2018.07.003
27. Skoulidis F, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer*. (2019) 19:495–509. doi: 10.1038/s41568-019-0179-8
28. Ma X, Le Teuff G, Lacas B, Tsao MS, Graziano S, Pignon JP, et al. Prognostic and predictive effect of TP53 mutations in patients with non-small cell lung cancer from adjuvant cisplatin-based therapy randomized trials: a LACE-bio pooled analysis. *J Thorac Oncol*. (2016) 11:850–61. doi: 10.1016/j.jtho.2016.02.002
29. Tomizawa K, Suda K, Onozato R, Kosaka T, Endoh H, Sekido Y, et al. Prognostic and predictive implications of HER2/ERBB2/neu gene mutations in lung cancers. *Lung Cancer*. (2011) 74:139–44. doi: 10.1016/j.lungcan.2011.01.014
30. Canale M, Petracci E, Delmonte A, Chiadini E, Dazzi C, Papi M, et al. Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. *Clin Cancer Res*. (2017) 23:2195–202. doi: 10.1158/1078-0432.CCR-16-0966
31. Labbe C, Cabanero M, Korpanty GJ, Tomasini P, Doherty MK, Mascaux C, et al. Prognostic and predictive effects of TP53 co-mutation in patients with EGFR-mutated non-small cell lung cancer (NSCLC). *Lung Cancer*. (2017) 111:23–9. doi: 10.1016/j.lungcan.2017.06.014
32. Fan Y, Zhu X, Xu Y, Lu X, Xu Y, Wang M, et al. Cell-cycle and DNA-damage response pathway is involved in leptomeningeal metastasis of non-small cell lung cancer. *Clin Cancer Res*. (2018) 24:209–16. doi: 10.1158/1078-0432.CCR-17-1582
33. Offin M, Feldman D, Ni A, Myers ML, Lai WV, Pentsova E, et al. Frequency and outcomes of brain metastases in patients with HER2-mutant lung cancers. *Cancer*. (2019) 125:4380–87. doi: 10.1002/cncr.32461
34. Xu C, Zhang Y, Liu D, Shen R, Razavi P, Liu S, et al. P1.01-99 detecting HER2 alterations by Next Generation Sequencing (NGS) in patients with advanced NSCLC from the United States and China. *J Thora Oncol*. (2018) 13:S502. doi: 10.1016/j.jtho.2018.08.655
35. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med*. (2015) 7:283ra54. doi: 10.1126/scitranslmed.aaa1408
36. Rachiglio AM, Fenizia F, Piccirillo MC, Galetta D, Crino L, Vincenzi B, et al. The presence of concomitant mutations affects the activity of EGFR tyrosine

kinase inhibitors in EGFR-mutant non-small cell lung cancer (NSCLC) patients. *Cancers*. (2019) 11:341. doi: 10.3390/cancers11030341
37. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*. (2015) 27:15–26. doi: 10.1016/j.ccell.2014.12.001
38. Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol*. (2014) 12:e1001906. doi: 10.1371/journal.pbio.1001906
39. van der Wekken AJ, Saber A, Hiltermann TJ, Kok K, van den Berg A, Groen HJ. Resistance mechanisms after tyrosine kinase inhibitors afatinib and crizotinib in non-small cell lung cancer, a review of the literature. *Crit Rev Oncol Hematol*. (2016) 100:107–16. doi: 10.1016/j.critrevonc.2016.01.024
40. Takezawa K, Pirazzoli V, Arcila ME, Nebhan CA, Song X, de Stanchina E, et al. HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFRT790M mutation. *Cancer Dis*. (2012) 2:922–33. doi: 10.1158/2159-8290.CD-12-0108
41. Rotow J, Bivona TG. Understanding and targeting resistance mechanisms in NSCLC. *Nat Rev Cancer*. (2017) 17:637–58. doi: 10.1038/nrc.2017.84
42. Chuang JC, Stehr H, Liang Y, Das M, Huang J, Diehn M, et al. ERBB2-mutated metastatic non-small cell lung cancer: response and resistance to targeted therapies. *J Thorac Oncol*. (2017) 12:833–42. doi: 10.1016/j.jtho.2017.01.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

REVIEW
published: 15 May 2020
doi: 10.3389/fonc.2020.00798

Next Generation Sequencing and Machine Learning Technologies Are Painting the Epigenetic Portrait of Glioblastoma

*Ivana Jovčevska**

Medical Centre for Molecular Biology, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

Even with a rare occurrence of only 1.35% of cancer cases in the United States of America, brain tumors are considered as one of the most lethal malignancies. The most aggressive and invasive type of brain tumor, glioblastoma, accounts for 60–70% of all gliomas and presents with life expectancy of only 12–18 months. Despite trimodal treatment and advances in diagnostic and therapeutic methods, there are no significant changes in patient outcome. Our understanding of glioblastoma was significantly improved with the introduction of next generation sequencing technologies. This led to the identification of different genetic and molecular subtypes, which greatly improve glioblastoma diagnosis. Still, because of the poor life expectancy, novel diagnostic, and treatment methods are broadly explored. Epigenetic modifications like methylation and changes in histone acetylation are such examples. Recently, in addition to genetic and molecular characteristics, epigenetic profiling of glioblastomas is also used for sample classification. Further advancement of next generation sequencing technologies is expected to identify in detail the epigenetic signature of glioblastoma that can open up new therapeutic opportunities for glioblastoma patients. This should be complemented with the use of computational power i.e., machine and deep learning algorithms for objective diagnostics and design of individualized therapies. Using a combination of phenotypic, genotypic, and epigenetic parameters in glioblastoma diagnostics will bring us closer to precision medicine where therapies will be tailored to suit the genetic profile and epigenetic signature of the tumor, which will grant longer life expectancy and better quality of life. Still, a number of obstacles including potential bias, availability of data for minorities in heterogeneous populations, data protection, and validation and independent testing of the learning algorithms have to be overcome on the way.

Keywords: glioblastoma, next generation sequencing, diagnosis, therapy, methylation, epigenetics, machine learning, deep learning

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Sevtap Savas,
Memorial University of Newfoundland, Canada
Zhenhua Xu,
Children's National Hospital, United States

***Correspondence:**
Ivana Jovčevska
ivana.jovcevska@mf.uni-lj.si

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *03 November 2019*
Accepted: *23 April 2020*
Published: *15 May 2020*

Citation:
Jovčevska I (2020) Next Generation Sequencing and Machine Learning Technologies Are Painting the Epigenetic Portrait of Glioblastoma. Front. Oncol. 10:798. doi: 10.3389/fonc.2020.00798

INTRODUCTION

This review starts with outlining the complex nature of glioblastoma by providing brief information about its occurrence, mortality rate, molecular features, and heterogeneity. Our level of understanding of glioblastoma genetics has remarkably increased since the introduction of next generation sequencing. However, lack of effective patient treatment necessitates implementation

of modern diagnostic and treatment options using newer technological developments. One such example is exploration of epigenetic markers for glioblastoma diagnosis and treatment. Although epigenetics in glioblastoma is still at its infancy, it shows potential for development of novel therapies. Moreover, development of machine learning and deep learning algorithms for glioblastoma patient care may improve objective disease diagnosis and can contribute to tailoring the most effective treatment based on patient molecular profile i.e., precision medicine. At last, understanding the molecular background of each patient will raise the quality of clinical care from the current supportive classical treatment to the level of significantly improving patient life expectancy and quality of life.

Glioblastoma

In the United States, cancer is the second leading cause of death in both genders with the four most prevalent types being lung, breast, prostate, and colorectal, while brain cancers account for only about 1.35% of the cases (1, 2). However, contrary to their rare occurrence, in the year 2016 brain tumors were the major cause of cancer-related death among men younger than 40 years of age and women younger than 20 years of age (2). With 26% and 21% of the cases, brain and other nervous system tumors represent one of the most commonly diagnosed tumors in children and adolescents, respectively (2). Among adults, gliomas account for almost 80% of the primary malignant brain tumors (3). Gliomas can be classified based on location, differentiation pattern, anaplasia, mitotic activity, and necrosis. Moreover, according to the World Health Organization (WHO), histologically they progress from benign (WHO grade I and II) to malignant (WHO grade III and IV) (3). The most malignant type is the grade IV glioblastoma which accounts for 60–70% of all gliomas (4) and 16% of all primary brain tumors (5). The age-adjusted annual incidence of glioblastoma is 3.19 per 100,000 people in the United States (6). Glioblastoma is most commonly diagnosed in elderly Caucasian men with mean age of 64 years (6–8). Genetically there are two main glioblastoma subtypes—isocitrate dehydrogenase (IDH) wild type (or primary) and mutant (or secondary) that are histo-pathologically the same, but with different clinical progression (9, 10). *IDH*-mutant glioblastomas tend to develop from previously diagnosed WHO grade II or III gliomas and appear in patients younger than 50 years of age, while *IDH*-wild type glioblastomas appear *de novo* in patients with median age of 60 years (3, 11). In general, patients with *IDH*-mutant glioblastomas show better overall survival than patients with *IDH*-wild type tumors (11).

Glioblastoma Diagnosis and Therapy

Due to unspecific symptoms like numbness, mood swings, fatigue, and mild memory loss (12), glioblastoma is usually diagnosed at an advanced stage, when little can be done for the patient. Definitive diagnosis can only be done histologically, but needs to be complemented with the recent advances in the molecular classification. The latest WHO classification of brain tumors (13) constitutes a combination of phenotypic and molecular parameters of brain tumors, and that leads to greater diagnostic accuracy. Still, in cases where molecular diagnostic testing is not available or is inconclusive, brain tumors are labeled as "not otherwise specified (NOS)." Standard treatment comprises maximal surgical resection, followed by concomitant chemotherapy with temozolomide and radiation, and then adjuvant chemotherapy (12). Temozolomide was approved by the US Food and Drug Administration (FDA) for treatment of glioblastoma in the year 2005 (14). It is a small lipophilic molecule that is absorbed completely after oral administration, minimally binds to plasma proteins and is able to penetrate the blood brain barrier (15). Still, this aggressive treatment gives patients only 12–18 months *post* diagnosis (16, 17), while the 5-year survival is only 9.8% (17). High mortality rate is a result of the localization and rapid tumor growth (3). In order to improve patient care and life expectancy, numerous alternative treatments such as tumor treating fields (18–20), gamma knife radiosurgery (21), and immunotherapy (22–25) are currently being explored.

DNA SEQUENCING

Sanger Sequencing

The first commercialized method for DNA sequencing named Sanger sequencing (26) was extensively used for almost three decades. Sanger sequencing or chain-termination sequencing is based on the use of 2′-deoxynucleotides (dNTPs) and 2′,3′-dideoxynucleotides (ddNTPs) for synthesis and termination of synthesis of the complementary DNA template, respectively. This leads to generation of fragments with different sizes which are separated by high-resolution gel electrophoresis and analyzed to reveal the DNA sequence. Automated Sanger sequencing used fluorescently labeled primers or terminating ddNTPs. Excitation of the fluorophores produced fluorescent emission in different colors that that were used for revealing the DNA sequence. One of the greatest accomplishments of automated Sanger sequencing was sequencing the complete human genome (27) that is considered the largest project of today's mankind (28). Still, its limitations in terms of cost, time, low throughput, efficiency and sensitivity, drove the development of newer sequencing technologies collectively named "next generation sequencing" (NGS).

"Next Generation Sequencing" Boom

NGS methods are based on the same principle as Sanger sequencing: they both use polymerases for synthesis, modified nucleotides, and fluorescent detection (29). However, for some NGS platforms like Illumina, Life Technologies SOLiD, Ion Torrent Personal Genome Machine (PGM), and Roche 454 systems, the DNA template has to be clonally amplified prior to sequencing, while for the more sensitive Heliscope and Pacific Biosciences (PacBio) single molecule real-time (SMRT) systems pre-amplification is not needed (30). Different NGS platforms use different chemistry for library preparation and sequencing (31). For example, Illumina sequencers are based on the "sequencing by synthesis" approach with fluorescently labeled reversible nucleotide terminator chemistry (32). On the other hand, Ion torrent technology generates sequence templates on a bead or sphere via emulsion PCR with sequencing-by-ligation approach and proton release detection. At last, PacBio sequencers

are based on SMRT sequencing with fluorescent detection (30). One of the major advantages of NGS is increased throughput at decreased expenditure i.e., its ability to generate large amount of data at reasonable costs. As an example, the standard Sanger sequencing yielded ∼6 Mb DNA sequence per day at a cost of $500/1 Mb while NGS sequencers like Illumina GA (San Diego, CA, USA), yield ∼5,000 Mb DNA sequence per day at a cost of $0.50/1 Mb (33). Still, potential problems that arise are setting the necessary infrastructure for NGS including machinery, costs for reagents, space for sample processing, and data storage (34). Moreover, trained personnel with adequate understanding of the software for data analysis and interpretation is a necessity. A more complex problem that should not be neglected is who owns the genetic information obtained from such analysis, and to what extent the raw data can be used for other pathologies besides the one originally intended? These limitations of NGS are issues that still need to be addressed together with institutional ethics boards, researchers and participants as the technology develops.

How did NGS become so popular? After the introduction of commercial next generation sequencers in the year 2005 (29), a new age in nucleic acid research was started. NGS is suitable for a wide range of applications in particular for analysis of genetic variations and mutations, mRNA, non-coding RNA, methylation studies as well as chromatin immunoprecipitation (ChIP)-sequencing (35). By generating genetic information from different individuals, NGS made it possible to perform large scale, comparative, and evolutionary studies (36), and also helped in the development of pharmaceuticals (37). Moreover, NGS started the era of genomic medicine—incorporating patient's tumor genetic information for diagnosis, treatment, and prevention of diseases. NGS can help in overcoming treatment challenges by identifying druggable genetic targets. At last, with NGS the human genome can be resequenced in order to get deeper understanding of how genetic changes affect health and disease (38). NGS offers enormous possibilities for increasing our understanding of human genetics of health and disease, which will change the way we diagnose disease and treat individuals.

Third Generation Sequencing

The best method for identification of genetic variations crucial for disease development is DNA sequencing. The right sequencing method is desired to be high-throughput, low-cost and able to sequence long reads with high accuracy (39). Despite all the advantages that next generation sequencing offers, the short length of the obtained reads is its weak spot. This led to the emergence of a third generation of sequencing that enables single molecule long reads (39) such as SMRT sequencing by PacBio (40) and nanopore sequencing originally introduced in the year 1996 (41). The authors showed that with an electrical field, single-stranded DNA (ssDNA), or RNA molecules can be driven through a 1–10 nm ion channel, i.e., nanopore, in a lipid bilayer or membrane. When passing through the nanopore, different bases of the DNA strand will cause specific fluctuations in the electrical current; these signals can be converted to DNA sequence information (39). The advantage of this method is the short time for sample preparation, electrical, or fluorescent readout and reads in length of several kilobases of single DNA molecules in real time (29). However, there are two possible issues that have to be resolved: (1) Length of the recognition region of the nanopore should not be larger than 0.5 nm—size that is equivalent to the base-spacing in a ssDNA, otherwise signal interference from adjacent bases will be observed (39); and (2) The current speed of DNA translocation, 300 bases/ms, is too fast for individual bases to be identified—ideally it should be adjusted to 1 base/ms (39). A solution has been proposed by IBM (New York, NY, USA) by creating a nanopore matrix i.e., a transistor with alternating fields of metal and dielectric materials which can control the speed of DNA translocation (42, 43).

New technological developments in the sequencing fields offer different techniques for establishing patients' tumors' molecular profiles which are expected to accelerate the development of personalized medicine. For example, targeted sequencing will allow detection of specific genetic changes of a predefined set of genes; whole exome sequencing will provide information about the coding regions of genes of interest; while RNA sequencing will give information about the post-transcriptional changes (44). At last, whole genome sequencing will provide a complete in depth genetic picture for each patient, but at the cost of great computational power, time, and resources. Because of the high molecular diversity of glioblastoma, these technological advancements are expected to deepen our knowledge of its mechanisms of development and progression. At last, by understanding how the disease works at different molecular levels (transcriptomic, genetic, epigenetic, and protein), new more powerful drugs can be designed that will be of a great benefit for the patients.

PAVING THE ROAD TO PRECISION MEDICINE

Cancer is a complex disease which arises as a result of combination of hereditary i.e., genetic and environmental factors such as physical and chemical agents, diet, lifestyle, tobacco, and alcohol use (35). These interactions leave footprints in the genome either as mutations or as epigenetic modifications. Genetic changes range from single base substitutions to major chromosomal losses, while the epigenetic modifications influence gene expression as well as DNA replication and repair (45, 46).

Glioblastoma is a disease that is characterized with heterogeneity at both intra- and inter-tumoral level. As a response to its complexity, the scientific society moves away from identification of a single gene as a cause of a disease, and toward identification of combination of molecular changes that eventually lead to tumor development (35). Such changes can be commonly observed in different individuals with the same disease. Personalized medicine implies development of drugs for the needs of a single patient, while precision medicine refers to the classification (or diagnosing) of individuals into genomic subclasses which can be treated in more targeted i.e., precise ways (47). The advantage of using NGS for diagnostics is the simultaneous detection of a number of different markers, which otherwise requires separate consecutive tests and prolongs the diagnostic process.

Genetics of Glioblastoma

During the course of The Cancer Genome Atlas (TCGA) project glioblastomas were genetically characterized. The most important findings included changes in three core pathways receptor tyrosine kinase (RTK)/rat sarcoma (RAS)/PI3K, p53, and retinoblastoma (RB) with alterations in 88, 78, and 87% of the analyzed cases, respectively (48). The most frequent gene alterations were found in epidermal growth factor receptor (*EGFR*—mutation in 45% of the cases), phosphatase and tensin homolog (*PTEN*—inactivated in 36% of the cases), cyclin-dependent kinase inhibitor 2A (*CDKN2A*—inactivated in 52% of the cases), cyclin-dependent kinase inhibitor 2B (*CDKN2B*—inactivated in 52% of the cases), tumor protein p53 (*TP53*—mutated in 35% of the cases) and *RB* (homozygote deletion in 11% of the cases) (11, 48). *IDH* mutations are rare in primary glioblastoma patients with *EGFR* and *PTEN* alterations, but are commonly found in low grade gliomas and together with *TP53* mutations in high grade gliomas that evolved from low grade gliomas (49). Later, Verhaak et al. established the molecular profile of glioblastoma (50). In their study, using molecular profiling they defined four glioblastoma subtypes: classical, proneural, neural, and mesenchymal with different molecular properties. This illustrates high glioblastoma heterogeneity at the molecular level that is present both within (intra) and among (inter) tumors.

Examining intratumor heterogeneity can be precisely performed with method that allows for analysis at individual cell resolution level, such as single-cell RNA sequencing (scRNA-seq) (51). scRNA-seq can be used not only to analyze tumor cells, but also non-tumor cells that shape the microenvironment and aid in tumor progression (52, 53). Numerous research groups are already using this technology to "shred glioblastomas to single-cells" and contribute to their molecular understanding (51, 54–58). One such study identified presence of different cells within the tumor compared to cells from the surrounding based on macrophage and microglia gene expression profiles (52). In another study (59) the authors provided evidence that glioblastoma stem cells shape the transcriptional and cellular landscapes of the tumor. In a different study (51), the authors proposed potential role of the expression levels of rare genes in glioblastoma tumorigenesis. Using the knowledge about molecular heterogeneity of glioblastoma, institutes already implemented next generation sequencing panels containing a specific set of glioblastoma-specific genes for patient diagnosis (60). It is reasoned that the intratumoral heterogeneity reflects the existence of different cellular subclones within the same tumor—this is why deciding the therapeutic strategy from a single biopsy specimen may not be enough for successful therapy (61). Moreover, the molecular diversity of glioblastoma is further increased during treatment—namely, tumors of patients treated with temozolomide present with a hypermutation phenotype (62, 63) which is associated with promotion of tumor growth and metastasis (64). At last, transcriptome analyses are also used for defining glioblastoma signatures that will help in precise disease diagnosis, as well as to anticipate therapy response and patient outcome (53).

Epigenetics of Glioblastoma

Epigenetic modifications are heritable changes that affect gene expression, but do not change the DNA sequence itself (65, 66). Such changes are DNA methylation, histone modifications, and chromatin remodeling (67). Histones are positively charged proteins H1, H2A, H2B, H3, and H4 (68). Chromatin refers to the complex of negatively charged DNA and positively charged histone proteins, or the fundamental subunit "nucleosome" in the nucleus. Every nucleosome consists of about 146–147 bp DNA associated with octameric core of histone proteins—two H3-H4 histone dimers surrounded by two H2A-H2B dimers (69). Histone acetylation i.e., addition of acetyl groups to lysines of H3 or H4, weakens the interactions between histones and DNA which opens the accessibility to the transcription apparatus, while histone deacetylation removes the acetyl groups which results in histone condensation and gene inactivation (70). These dynamic processes are maintained by histone acetyltransferases (HAT) and deacetylases (HDAC). Histone modifications are different in pediatric and adult gliomas. In pediatric gliomas, the most common mutations are K27M and G34R/V on histone variant H3.3 (71). Mutations in *H3F3A* show specificity for pediatric high grade gliomas, and are not found in pediatric low grade gliomas, embryonic tumors, or ependymomas nor in adult glioblastoma (72). In adult gliomas, *IDH1* mutations indirectly affect H3K27 or H3K36 methylation (73). Lysine (K) methylation at positions K4, K36, and K39 on H3 marks active chromatin regions, while at positions K9 and K27 at H3 it marks inactive chromatin regions (74, 75); still, lysine methylation does not change the net charge of the histone tail (76). Another epigenetic modification is chromatin remodeling that causes conformation changes in chromatin which regulate the DNA-dependent processes, replication, and repair systems as well as centromere and telomere maintenance (67, 77). These 3D conformational chromatin changes can affect gene expression by regulating access to RNA polymerases and transcription factors (77). Examples of the involvement of chromatin remodeling in glioblastoma pathogenesis are switch/sucrose non-fermenting (SWI/SNF) core complex (78) and Brahma-related gene 1 (BRG1), one of the catalytic subunits of the SWI/SNF chromatin remodeling complex (79) that regulate stemness and tumorigenic potential of glioma initiating cells.

DNA Methylation

From the four DNA bases, only cysteine and adenine can be methylated. Yet, DNA methylation usually refers to the covalent transfer of methyl groups to the C-5 position of the cytosine ring to create 5-methylcytosine as shown in **Figure 1**. In mammals, DNA methylation occurs on any cytosine of the genome; however, in the majority of the cases it occurs in a 5′-CpG-3′ dinucleotide context of somatic cells, and up to 25% of methylation occurs in non-CpG context of embryonic cells (82). Typically, CpG islands belong to or are near promoter regions of genes where transcription starts (74). Adenine methylation is observed in bacterial, plant, and lately also in mammalian DNA, but is not that much explored and its role is largely unknown (83–85). Methylation is needed for silencing transposable elements and genes on the inactive X-chromosome, as well as maintaining

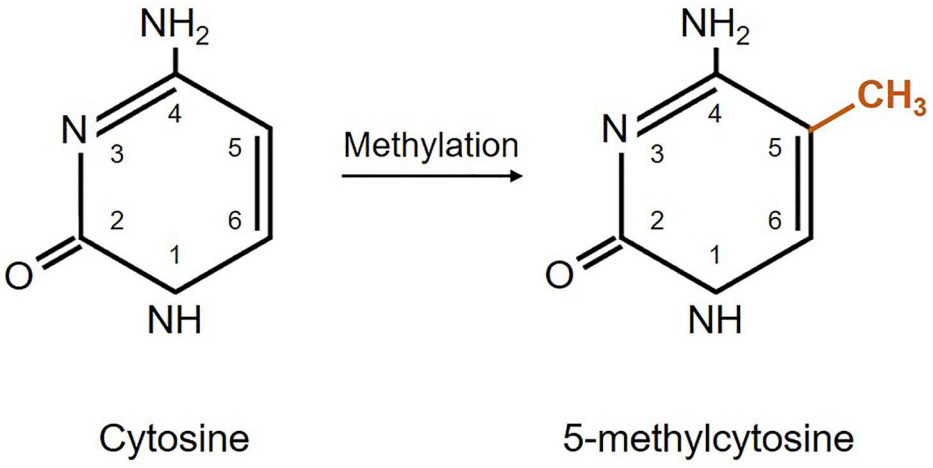


FIGURE 1 | Schematic representation of DNA methylation. Cytosine methylation is mediated by a family of DNA methyltransferase enzymes DNMT1, DNMT3A, and DNMT3B (65, 80, 81).

genome stability (86). Besides, DNA methylation plays an important role in regulation of gene expression that has an impact on the clinical outcome of glioblastoma patients (65, 87, 88).

DNA Methylation in Glioblastoma

Cancers in general are characterized by global hypomethylation which is associated with gene expression, activation, and chromosomal rearrangements of oncogenes which leads to genomic instability, oncogene activation, and tumor progression, as well as locus-specific hypermethylation which results in heritable transcriptional silencing of tumor suppressor genes (82, 89). Global hypomethylation occurs in 80% of glioblastomas (90). DNA methylation that occurs in CpG islands in gene promoter regions inversely correlates with gene transcription. In glioblastomas, besides at genetic, intratumor heterogeneity is present also at DNA methylation level. Wenger et al. analyzed multiple spatially separated tumor specimens from 12 glioblastoma patients (38 samples total) and observed existence of different methylation subclasses intertumorally—three samples presented with combined existence of mesenchymal and receptor tyrosine kinase (RTK) II subclasses, while two contained both RTKI and RTKII at once (91). Using clustering of *IDH*-mutant gliomas, Ceccarelli et al. observed existence of three major glioma groups: Codel, *IDH*-mutant 1p/19q codeleted low grade gliomas; glioma CpG island methylator phenotype (G-CIMP)-low, *IDH*-mutant non-1p/19q codeleted low and high grade gliomas with low genome-wide DNA methylation; and G-CIMP-high, *IDH*-mutant non-1p/19q codeleted low and high grade gliomas with higher level of genome-wide DNA methylation (92). Among these, the G-CIMP low group was found to have worst prognosis, while G-CIMP and Codel presented with similar favorable prognoses. The authors also analyzed *IDH*-wild type gliomas and observed existence of three DNA methylation groups: classic-like, mesenchymal-like, and *IDH*-wildtype low and high grade gliomas. An interesting observation was that only patients with low grade gliomas from the *IDH*-wildtype group experienced significantly longer survival. In a different study, Pangeni et al. performed methylation array profiling on a panel of 23 patient-derived glioblastoma stem cell lines and data for TCGA patients with *IDH*-wildtype glioblastomas (89). Different glioblastoma subtypes were included in the analysis. The authors observed similar bi-modal distribution of CpG methylation in glioblastoma stem cells and glioblastoma tumors. They also found more mesenchymal-hypermethylated than hypomethylated genes in both glioblastoma stem cells and glioblastoma tumors. They observed that high expression levels of proneural-methylated genes CASP8 and FADD-like apoptosis regulator (*CFLAR*) and Sp100 nuclear antigen (*SP100*), and low expression levels of the mesenchymal-methylated genes transmembrane and coiled-coil domain family 1 (*TMCC1*), Rho guanine nucleotide exchange factor 7 (*ARHGEF7*), Notch homolog 1, translocation-associated (*NOTCH1*), midnolin (*MIDN*), potassium voltage-gated channel subfamily Q member 2 (*KCNQ2*), ataxin 10 (*ATXN10*), ubiquitin specific peptidase 54 (*USP54*), and TUB bipartite transcription factor (*TUB*) correlate to poorer patient prognosis (89).

In glioblastoma, DNA methylation is closely correlated to response to temozolomide treatment. So far, methylation of the O^6-methylguanine-DNA methyltransferase (MGMT) is the only predictive biomarker for patient response to first-line temozolomide chemotherapy (93, 94). MGMT is a DNA repair protein that reverses the damage done by alkylating agents such as temozolomide. Temozolomide adds methyl groups at N^7 and O^6 sites on guanines and O^3 sites of adenines in genomic DNA and this is why it is cytotoxic to cells (14). *MGMT*-promoter methylation causes gene silencing, therefore less protein is expressed which leads to sensitivity to chemotherapy with alkylating agents (95). In a study by Smrdel et al. the authors compared overall survival and time to progression in patients with and without methylated MGMT (96). Their results show

longer overall survival (43 vs. 16 months) and time to progression (36 vs. 11 months) in patients with methylated MGMT compared to patients without methylated MGMT, respectively. The authors also observed MGMT methylation in 36 out of 38 (95%) patients who present with long survival (more than 30 months after diagnosis) and in only 12 out of 33 (36%) patients in the control group (short term survival patients i.e., <30 months after diagnosis). In general, patients with methylated *MGMT* promoter respond better to temozolomide therapy and present with longer survival (3, 63, 97, 98). These findings are in concordance with another study where Felsberg et al. took in consideration 67 adult patients diagnosed with glioblastoma (99). The authors conclude that *MGMT* promoter hypermethylation is associated with longer time to progression after initiation of chemotherapy (245 vs. 100 days) and longer overall survival (692 vs. 474 days).

On the other hand, one study reports identification of a new histone deacetylase inhibitor, 7-ureido-N-hydroxyheptanamide derivative—CKD5, that shows strong anticancer effect in glioblastoma *in vitro* and *in vivo* (100). The advantages of this inhibitor are its water solubility (>1,000 mg/mL), negative result on Ames test which indicated that CKD5 is not genotoxic and does not introduce DNA mutations, and it showed no signs of inducing cardiac toxicity in pre-clinical trials. However, for glioblastoma treatment delivery methods have to be developed as CKD5 cannot penetrate the blood brain barrier. The use of H3K4 and H3K9 me1/2 demethylase KDM1 and H3K4 me2/3 demethylase KDM5A as potential therapeutic targets was also tested (76). The authors of the study show that inhibiting KDM1 and KDM5A enzyme activity presents with strong antitumor affect in wild-type and temozolomide-resistant glioblastoma cells. Romani et al. used a multi-KDM inhibitor, JIB 04, that has strong anti-clonogenic activity in wild-type and temozolomide-resistant glioblastoma cells after only 4 h incubations at low JIB 04 concentrations (0.5 and 1 μM). The authors also tested different drug combinations and showed synergistic effect of JIB 04 and GSK J4 (selective inhibitor of KDM6A/B) on cell proliferation and reduction of the clonogenic potential of temozolomide-resistant glioblastoma cells. These studies show that with targeting epigenetic changes non-traditional treatment methods for glioblastoma patients whose tumors are resistant to the temozolomide chemotherapy can be developed. Knowing the aggressiveness of the tumor and its poor response to current available treatment options, these findings give new hope for glioblastoma patients.

Clinical Trials

Since the discovery of the altered epigenetic profiles of cancers, epigenetics is getting more and more attention in the research community. Histone deacetylases, the enzymes that remove the acetyl group from histones which is initially associated with gene repression, are often overexpressed in cancer (100). Due to their reversible nature and role in gene expression, epigenetic changes, especially changes in histone acetylation, are also considered as possible therapeutic targets (86) which can be seen from the growing number of clinical trials based on the use of different enzyme inhibitors. Data was obtained from the database of publicly and privately funded clinical studies worldwide, https://clinicaltrials.gov, using the following keywords: "glioblastoma," "Vorinostat," "Romidepsin," "Belinostat," "Panobinostat," "LHB589," "Valproic acid," "Olaparib," "Veliparib," "ABT-888," "Iniparib," "BSI-201," and "CBL0137." The search conducted on the 8th of August 2019 gave the results presented in **Table 1**. Still, an obstacle in these clinical trials are our current lack of knowledge about the mechanism of HDAC inhibitors and how they affect cellular signaling pathways; moreover, methods for improved penetration of the HDAC inhibitors into the brain and across the blood brain barrier are still to be found (77).

Novel Next Generation Sequencing-Based Diagnostic Approaches

Although the work of Verhaak et al. changed the way we diagnose glioblastoma and increased our understanding of the disease (50), there are still a number of patients whose tumors cannot be classified according to the currently defined subtypes (13). Anyway, understanding the molecular background of disease requires availability of sets of reference samples from healthy donors (126) for comparison.

Molecular Re-classification of Glioblastomas

Using next generation sequencing data available from TCGA, Bolouri et al. report clustering of glioblastomas based on their methylation profiles (127). The authors of the study used genome wide and methylation data from a merged cohort of glioblastomas and lower grade gliomas and obtained three big glioma clusters: (1) oligodendrogliomas, (2) astrocytomas and oligoastrocytomas, and (3) mostly glioblastomas with a few astrocytomas and oligoastrocytomas (127).

More recently, Capper et al. aided in the classification of glioblastomas that do not belong to the known entities i.e., WHO subclasses (128). The authors established a system for classification of brain tumors based on their epigenetic profiles. They generated genome-wide DNA methylation profiles of 76 histopathological entities and seven entity variants that occur in the central nervous system. Unsupervised clustering within each entity and across histologically similar tumor entities led to obtaining 82 tumor classes with different DNA methylation profiles. Of these, 29 classes were equivalent to single WHO entities, 29 classes were subclasses of WHO entities, in eight classes WHO entities could not be recapitulated, in 11 classes were not identical to WHO entities, while the remaining five classes presented with methylation profiles that are not described by the WHO. To test clinical implementation of this system, Capper et al. analyzed 1,104 samples or 64 different histopathological entities of adult and pediatric brain tumors. The majority of the samples (88%) matched an established DNA methylation profile. From these, in 76% of the cases pathological and methylation profiling were concordant. In the remaining 12% of the cases, methylation and pathological diagnoses were not concordant; samples were molecularly re-evaluated and diagnosis was changed as it was suggested by the methylation subclassifier; diagnoses were changed into both lower (30%) and higher (41%) WHO grades. The study demonstrates that

TABLE 1 | List of selected glioblastoma clinical trials using drugs targeted against enzymes involved in epigenetic modifications.

Drug	Role	Clinical trial number	Clinical trial phase	Selected references
Vorinostat	HDAC inhibitor	NCT00762255	I (completed)	(101–107)
		NCT01266031	I (completed)	
		NCT01110876	I (terminated)	
		NCT03426891	I (recruiting)	
		NCT01342757	N/A	
		NCT00555399	I/II	
		NCT00731731	I/II	
		NCT00641706	II (completed)	
		NCT00238303	II (completed)	
		NCT00939991	I/II	
		NCT01738646	II (completed)	
		NCT01189266	I/II	
		NCT00268385	I	
		NCT01378481	I (terminated)	
Romidepsin	HDAC inhibitor	NCT00085540	I/II (completed)	(108)
Belinostat	HDAC inhibitor	NCT02137759	II	(109)
Panobinostat (LBH589)	HDAC inhibitor	NCT01115036	II (withdrawn)	(110)
		NCT00848523	II (terminated)	
Valproic acid	HDAC inhibitor	NCT02648633	I (terminated)	(111–117)
		NCT02758366	II	
		NCT01817751	II	
		NCT01861990	I (withdrawn)	
		NCT03243461	III	
		NCT00302159	II (completed)	
Olaparib	PARP inhibitor	NCT02974621	II	(118, 119)
		NCT01390571	I (completed)	
		NCT03212274	II (recruiting)	
Veliparib (ABT-888)	PARP inhibitor	NCT02152982	II/III	(120, 121)
		NCT00770471	I (completed)	
		NCT01026493	I/II (completed)	
		NCT03581292	II (recruiting)	
		NCT01514201	I/II (completed)	
Iniparib (BSI-201)	PARP inhibitor	NCT00687765	I (completed)	(122)
CBL0137	Histone chaperone FACT inhibitor	NCT01905228	I (recruiting)	(123–125)

HDAC, histone deacetylase; PARP, poly(ADP-ribose) polymerase; FACT, facilitates chromatin transcription; N/A, not applicable.

variability within and among tumors can be detected with deeper molecular analysis which can lead to more precise diagnosis and better treatment.

Cancer methylome is a combination of DNA methylation changes and characteristics of the cells of origin; in heterogeneous metastatic tumors this can aid in defining the primary cancer site (128). Although adding molecular characteristics into the histological diagnosis of glioblastoma is beneficial for precise diagnosis, single-gene tests based on DNA methylation analysis like *MGMT* methylation status, fluorescence *in situ* hybridization [1p/19q codeletion, *EGFR*, proto-oncogene C-Myc (*MYC*), class E basic helix-loop-helix protein 37 (*MYCN*), platelet derived growth factor receptor alpha (*PDGFRA*) and 19q13.42), or immunohistochemistry (catenin beta-1 (CTNNB1) and Lin-28 homolog A (LIN28A)], have proven difficult to be standardized (128). The studies from Bolouri et al. and Capper et al. only illustrate the importance of implementation of methylation profiling in glioblastoma diagnosis. Due to tumor heterogeneity it is important for all variables to be taken in consideration for fully informed decision about patient's therapeutic course to be made.

Nanopore Sequencing for Same-Day Patient Diagnosis

Further development of next generation sequencing techniques may lead to same-day patient diagnosis with nanopore-based systems. Similar to a small 100 mV electrical current passing through a nanopore placed in a membrane separating two chambers with aqueous electrolytes that can be measured with standard electrophysiological techniques, nucleobases of electrophoretically driven ssDNA or RNA molecules would modulate the current when passing through a suitable nanopore (129). The ideal nanopore sequencer is characterized by inexpensive sample preparation complemented with disposable chip with integrated microfluidics and probes, and a portable benchtop instrument for processing of long reads (>10,000–50,000 nt). Using a nanopore sequencer, Jain et al. reported sequencing and assembly of a reference human genome from ultra-long reads up to 882 kb in length with 5x coverage (130). The advantage of this benchtop method over next generation sequencing is its simplicity, speed, size, and reduced cost—nanopore can provide sequencing results faster and in

resource-restricted areas (44). One big concern however, is the amount of starting material; namely, ∼700 μg of human diploid genetic material will be needed to provide adequate throughput for the nanopore array, which, in theory, can be obtained with routine extraction procedures (129). Still, in the time when single-cell sequencing research is escalating (52, 55, 131) this can be seen as a pitfall.

Meanwhile, nanopore sequencing is already being tested for implementation in cancer diagnostics. In their retrospective proof-of-principle study, Euskirchen et al. examined the utility of nanopore sequencing (Oxford Nanopore Technologies, Oxford, UK) for multimodal molecular diagnosis using previously characterized brain tumors (132). Using deep amplicon sequencing, the authors were able to identify single nucleotide variants in *IDH1*, *IDH2*, *TP53*, *H3F3A,* and *TERT* promoter, and diagnostically relevant alterations like 1p/19q codeletion within minutes of sequencing. Moreover, they obtained 0.1X genome coverage within 6 h where copy number and epigenetic profiles matched the ones from microarray data. Because it can detect base modifications such as 5-methylation of cytosines, nanopore sequencing is also suitable for methylation analysis without the need of bisulfite conversion of the DNA. This will increase consistency during nucleic acid processing as well as significantly reduce the time needed for sample preparation. In the study, the authors observed good correlation between the methylation events in CpG sites obtained with nanopore sequencing and the corresponding microarray data.

The study by Euskirchen et al. shows the potential that nanopore sequencing offers for timely cancer diagnosis. In addition, due to small size and inexpensive technology, this method of histomolecular disease classification can be used worldwide even in regions with restricted clinical resources. However, further optimization like for use with highly fragmented nucleic acids originating from improperly stored or formalin-fixed paraffin-embedded tissue samples is still needed. Furthermore, as the method is still in developmental stages frequent improvements of the chemistry and software challenge its clinical implementation at this time.

Artificial Intelligence in Biomedical Sciences

In the light of new technological developments, the use of artificial intelligence (AI) in biomedical research will bring glioblastoma diagnostics on an advanced level. Machine learning is an application of artificial intelligence that allows for computers to work on tasks, learn from the data and improve their performance based on the gained experience (133, 134). It is a combination of mathematics and computer science that is based on building statistical models from large datasets i.e., billions or trillions data points (133). Classical statistical models describe the relationship between covariates (e.g., clinical factors) and a single dependent variable (e.g., outcome) obtained from a sample population and projected to a larger population. For instance, statistical models are suitable for deciding on treatment strategies based on survival, while machine learning models seek to predict the outcome using clinical factors as input features (135). Machine learning is defined as "the study of algorithmically built mathematical models that have been fitted for the training dataset to make predictions for the similarly obtained and structured validation dataset" (136). Extensive use of machine learning in biomedical fields, either for diagnostic, or therapeutic purposes, is conditioned by the availability of large data sets and appropriate IT infrastructure. Large datasets containing genetic information are generated by sequencing the human genome—a method that became routine with the wide implementation of next generation sequencing in research.

Machine Learning for Disease Diagnosis and Therapy

In cancer diagnostics, microscopic examination of patient samples is crucial for determining cancer staging. However, microscopy is based on the image interpretation of an expert individual which can also be subjective; lately quantitative examination of microscopy samples is required; and lastly, the availability of such expert individuals can be limited (137). AI can help in automated image analysis for pathological purposes with improved diagnostic accuracy, quantification, and efficiency. One such example is the augmented reality microscope—optical light microscope that enables real-time integration of AI (137).

Generation of large amount of data that contains information about human genetics allows for the development of machine learning techniques whose algorithms are either supervised or unsupervised clustering type. In supervised learning algorithms learn from labeled data, while in unsupervised learning the algorithms try to understand relationships from unlabelled data (134).

Machine learning for therapeutic purposes will be additionally enriched by the implementation of *in silico* drug discovery and design systems. One such example is the DrugBank database that contains quantitative, analytic and molecular information about drugs, and drug targets (138). DrugBank is organized into four major groups: (1). FDA approved small molecule drugs (>700 entries), (2). FDA approved bio-tech (protein/peptide) drugs (>100 entries), (3). Nutraceuticals or micronutrients such as vitamins and metabolites (>60 entries), and (4). Experimental drugs such as unapproved, de-listed and illicit drugs, enzyme inhibitors, and potential toxins (3,200 entries). Machine learning for drug discovery will offer a cost-effective and timely alternative to current experimental procedures (139). Another perspective is applying machine learning technology for predicting clinical efficiency of drugs and individualized treatment methods (140). This method which is named "drug scoring" or "personalized (individual) medicine" will take into account features that describe activation of cell signaling and metabolic pathways which will distinguish patients who will respond to the treatment from those who will not benefit from it. There are two principles for drug scoring: *a priori*—evaluating the ability of a drug to restore normal status or stop a physiological process that is considered pathogenic; and *a posteriori*—resulting from a machine learning process on a training dataset containing information from patients treated with the drug in question (140). The authors have developed a method for translating drug efficiency results obtained using cell lines to predict clinical effect

on humans. The use of this method can potentially reduce the costs of drug screening.

Even though machine learning shows potential to improve disease diagnosis and therapy, it still possesses a few limitations such as separating causation from correlation, removing biased data, and regulating predictive analytics (141). For machine learning to be safely used in disease diagnosis and/or treatment, the data which is taken into consideration has to be thoroughly examined to ensure it is appropriate for the specific problem. For correct identification and analysis, data must be equally collected and annotated, and it must be representative even for minorities in heterogeneous populations. In addition, self-implemented risk factors like smoking should be taken in consideration but should not be a limiting factor when deciding on future healthcare measures. An ethical issue in deciding treatment with machine learning can be existence of a permanent health condition, or chronic infection, like an individual being HIV positive. Another possible problem is bias toward populations that provide the most data, and in some societies, toward those that are able to afford medical procedures. As machine learning algorithms are trained on retrospective data it is possible for human subjectivity to influence the results; however, this can be improved by introducing new raw data (135). Before applying machine learning into clinical care, researchers must also consider protecting the privacy of the patients, ensure protection of data and patient information, and allow for equal treatment of all affected parties (141). At last, validation and independent testing of the machine learning algorithms must be performed in order to exclude mistakes due to technical differences. With all the advantages that it offers, machine learning in biomedicine is still at the beginning of its development and requires a multidisciplinary team to answer questions about ethical, legal, moral, and technological issues before it can objectively aid in better patient care.

Deep Learning

Machine learning works only with structured data which means reduction of amount of data in the raw format, significant time input from a medical professional to structure the data and introduction of human subjectivity (135). On the other hand, deep learning can use a wide range of different parameters which can be optimized by training on labeled data for prediction. While machine learning has already been used to determine gene expression patterns relevant for glioblastoma patient survival (50, 142), the use of deep learning for prognostic gene discovery it at its beginnings (143). The advantage of deep learning is that it

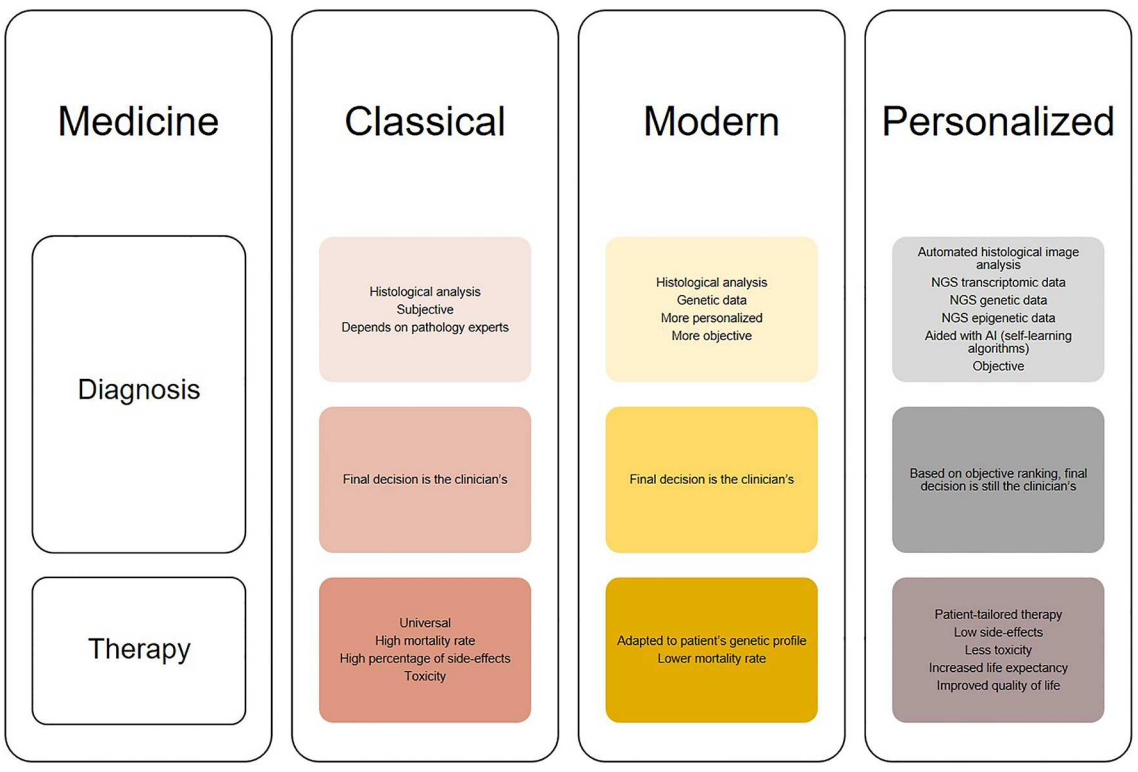


FIGURE 2 | Illustration of the changes in diagnosis and treatment of glioblastomas in different medical approaches (classical, modern, and personalized). Classical medicine relied on histological analysis of tissues, which is merely subjective, while therapy is universal for the patients which does not show much clinical success. On the other hand, modern medicine takes in consideration both histologic and genetic components which leads to greater diagnostic accuracy [examples: "glioblastoma, IDH-mutant" and "oligodendroglioma, IDH-mutant and 1p/19q-codeleted" (13)], while therapy is modified to suit tumor genetic profile. The trend is moving toward personalized medicine, where diagnosis will be thorough and objective aided by automated histological image analysis, next generation sequencing (NGS) and artificial intelligence (AI) algorithms, and therapy will be adapted not only to the genetic but also transcriptomic and epigenetic patient profile. This will result in increased overall survival and better quality of life.

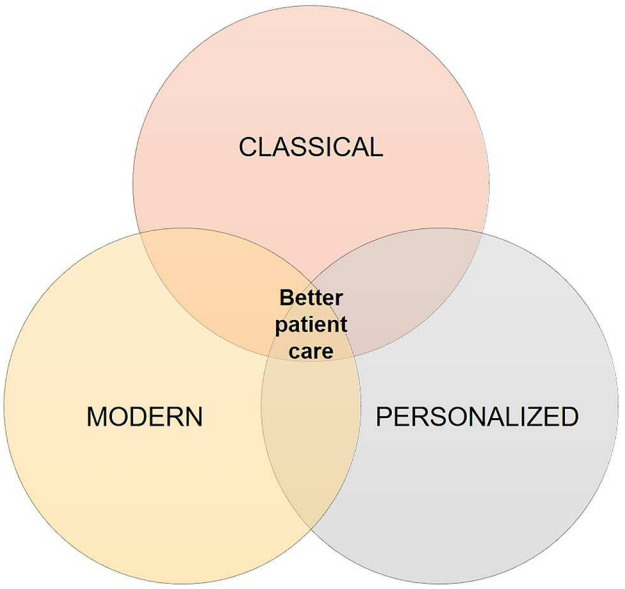


FIGURE 3 | Glioblastoma diagnosis. This figure illustrates what all healthcare systems have in common and aim for—which is better patient care. Starting from the classical, to the modern, where we are now, and going toward personalized medicine the goal is still improving patient quality of life i.e., both disease diagnosis and patient treatment.

can model a large number of differentially expressed genes. Using TCGA data as input, Wong et al. used deep learning to model the relationship between genes and their corresponding proteins on survival prognosis (143). Their model identified different genes associated with glioblastoma survival, glioblastoma cancer cell migration, or glioblastoma stem cells. In a different study, Young et al. used deep learning to classify glioblastomas into six subtypes which corresponded with significant differences in patient survival (144). These findings are in concordance with those from Brennan et al. who used DNA methylation data to classify glioblastomas into six subtypes: mesenchymal, proneural, neural and classical, as identified by Verhaak et al., supplemented with G-C island methylation phenotype (G-CIMP) and non-G-CIMP subtypes within the proneural subtype (145).

Although still at its infancy, the use of deep learning may open up new possibilities for glioblastoma diagnosis. Due to its ability to analyse large amount of data, deep learning can aid identifying features with biological significance which are currently unknown or too complex to be understood. Moreover, novel applications of deep learning may reveal hidden structures of cellular pathways and disease mechanisms. Glioblastoma diagnosis has significantly changed through the years as it can be seen in **Figure 2**. It started from histological analysis (classical) through addition of molecular and genetic characteristics of the individuals (modern) and is moving toward implementation of self-learning algorithms (personalized) which will eventually lead to the next presently unknown level. However, all these "medicines" have a common goal that is longer life expectancy and greater quality of life or better patient care (**Figure 3**). This transition from the classical through modern to personalized medicine was greatly aided by the massive use of NGS methods and is able to further develop also because of them.

CONCLUSIONS

With the use of NGS researchers generate large amount of data about transcriptomic, genomic, and epigenomic characteristics of humans. However, so far, only a small fraction has been proven to have clinical implementation. Understanding the rest of the "unlocked" data will only be possible with the development of more powerful analytic objective systems. The potential that NGS holds for glioblastoma research and clinical implementation is massive. Intelligent and careful use of NGS data can change the way we diagnose and treat glioblastomas. Studying epigenetic modifications in glioblastoma offers potential for identification of clinical biomarkers either for patient diagnosis or discovering drug targets. Rapid development of different methodologies for analysis of big data may lead to the development of individual diagnosis and patient-tailored treatment. It is expected for *in silico* diagnosis to be comparable and consistent, less variable, objective, and without human error. However, machine learning and deep learning algorithms have still a lot to learn before this diagnostic model can be implemented in everyday clinical practice. At last, these models need to be trained to understand biological systems so they can have an "insight" into the disease biology. This way, machine learning and deep learning models can adapt their findings in concordance to the nature of the analyzed disease, and simultaneously learn and change as the disease evolves.

AUTHOR CONTRIBUTIONS

IJ performed literature search, collected data and information, prepared figures and table, and wrote the manuscript.

FUNDING

IJ acknowledges financing from the Interreg EC Project 2014-2020 Ref. No. 146 Acronym: TRANS-GLIOMA, the Research Programme Grant P1-0390 and the Z3-1869 project from the Slovenian Research Agency (SRA). The funders had no role in manuscript design, writing or decision to publish.

ACKNOWLEDGMENTS

IJ would like to thank Alja Zottel and Neja Šamec (both from MCMB, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana) for critical reading of the manuscript, constructive suggestions, and comments.

REFERENCES

1. Kanu OO, Mehta A, Di C, Lin N, Bortoff K, Bigner DD, et al. Glioblastoma multiforme: a review of therapeutic targets. *Expert Opin Ther Targets.* (2009) 13:701–18. doi: 10.1517/14728220902942348
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* (2019) 69:7–34. doi: 10.3322/caac.21551
3. Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, et al. Glioma. *Nat Rev Dis Primers.* (2015) 1:15017. doi: 10.1038/nrdp.2015.17
4. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med.* (2008) 359:492–507. doi: 10.1056/NEJMra0708126
5. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro Oncol.* (2013) 15 (Suppl. 2):ii1–56. doi: 10.1093/neuonc/not151
6. Tamimi, A. F., and Juweid, M. (2017). Epidemiology and outcome of glioblastoma. In: De Vleeschouwer S, editor. *Glioblastoma.* Brisbane, QLD: Codon Publications. doi: 10.15586/codon.glioblastoma.2017.ch8
7. Aldape K, Zadeh G, Mansouri S, Reifenberger G, von Deimling A. Glioblastoma: pathology, molecular mechanisms and markers. *Acta Neuropathol.* (2015) 129:829–48. doi: 10.1007/s00401-015-1432-1
8. Movassaghi M, Shabihkhani M, Hojat SA, Williams RR, Chung LK, Im K, et al. Early experience with formalin-fixed paraffin-embedded (FFPE) based commercial clinical genomic profiling of gliomas-robust and informative with caveats. *Exp Mol Pathol.* (2017) 103:87–93. doi: 10.1016/j.yexmp.2017.06.006
9. Ohgaki H, Dessen P, Jourde B, Horstmann S, Nishikawa T, Di Patre PL, et al. Genetic pathways to glioblastoma: a population-based study. *Cancer Res.* (2004) 64:6892–9. doi: 10.1158/0008-5472.CAN-04-1337
10. Olar A, Aldape KD. Using the molecular classification of glioblastoma to inform personalized treatment. *J Pathol.* (2014) 232:165–77. doi: 10.1002/path.4282
11. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science.* (2008) 321:1807–12. doi: 10.1126/science.1164382
12. Alexander BM, Cloughesy TF. Adult Glioblastoma. *J Clin Oncol.* (2017) 35:2402–9. doi: 10.1200/JCO.2017.73.0119
13. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
14. Lee SY. Temozolomide resistance in glioblastoma multiforme. *Genes Dis.* (2016) 3:198–210. doi: 10.1016/j.gendis.2016.04.007
15. Thomas A, Tanaka M, Trepel J, Reinhold WC, Rajapakse VN, Pommier Y. Temozolomide in the era of precision medicine. *Cancer Res.* (2017) 77:823–6. doi: 10.1158/0008-5472.CAN-16-2983
16. Zhong J, Paul A, Kellie SJ, O'Neill GM. Mesenchymal migration as a therapeutic target in glioblastoma. *J Oncol.* (2010) 2010:430142. doi: 10.1155/2010/430142
17. Noroxe DS, Poulsen HS, Lassen U. Hallmarks of glioblastoma: a systematic review. *ESMO Open.* (2016) 1:e000144. doi: 10.1136/esmoopen-2016-000144
18. Rulseh AM, Keller J, Klener J, Sroubek J, Dbaly V, Syrucek M, et al. Long-term survival of patients suffering from glioblastoma multiforme treated with tumor-treating fields. *World J Surg Oncol.* (2012) 10:220. doi: 10.1186/1477-7819-10-220
19. Stupp R, Taillibert S, Kanner A, Read W, Steinberg D, Lhermitte B, et al. Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA.* (2017) 318:2306–16. doi: 10.1001/jama.2017.18718
20. Rick J, Chandra A, Aghi MK. Tumor treating fields: a new approach to glioblastoma therapy. *J Neurooncol.* (2018) 137:447–53. doi: 10.1007/s11060-018-2768-x
21. Paolillo M, Boselli C, Schinelli S. Glioblastoma under siege: an overview of current therapeutic strategies. *Brain Sci.* (2018) 8:15. doi: 10.3390/brainsci8010015
22. Finocchiaro G, Pellegatta S. Perspectives for immunotherapy in glioblastoma treatment. *Curr Opin Oncol.* (2014) 26:608–14. doi: 10.1097/CCO.0000000000000135
23. Xu LW, Chow KK, Lim M, Li G. Current vaccine trials in glioblastoma: a review. *J Immunol Res.* (2014) 2014:796856. doi: 10.1155/2014/796856
24. Lemee JM, Clavreul A, Menei P. Intratumoral heterogeneity in glioblastoma: don't forget the peritumoral brain zone. *Neuro Oncol.* (2015) 17:1322–32. doi: 10.1093/neuonc/nov119
25. Artene SA, Tuta C, Dragoi A, Alexandru O, Stefana Oana P, Tache DE, et al. Current and emerging EGFR therapies for glioblastoma. *J Immunoassay Immunochem.* (2018) 39:1–11. doi: 10.1080/15321819.2017.1411816
26. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA.* (1977) 74:5463–7. doi: 10.1073/pnas.74.12.5463
27. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature.* (2004) 431:931–45. doi: 10.1038/nature03001
28. Metzker ML. Emerging technologies in DNA sequencing. *Genome Res.* (2005) 15:1767–76. doi: 10.1101/gr.3770505
29. Stranneheim H, Lundeberg J. Stepping stones in DNA sequencing. *Biotechnol J.* (2012) 7:1063–73. doi: 10.1002/biot.201200153
30. Buermans HP, den Dunnen JT. Next generation sequencing technology: advances and applications. *Biochim Biophys Acta.* (2014) 1842:1932–41. doi: 10.1016/j.bbadis.2014.06.015
31. Hodzic J, Gurbeta L, Omanovic-Miklicanin E, Badnjevic A. Overview of next-generation sequencing platforms used in published draft plant genomes in light of genotypization of immortelle plant (Helichrysium Arenarium). *Med Arch.* (2017) 71:288–92. doi: 10.5455/medarh.2017.71.288-292
32. Hert DG, Fredlake CP, Barron AE. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis.* (2008) 29:4618–26. doi: 10.1002/elps.200800456
33. Kircher M, Kelso J. High-throughput DNA sequencing–concepts and limitations. *Bioessays.* (2010) 32:524–36. doi: 10.1002/bies.200900181
34. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics. *J Pathol Inform.* (2012) 3:40. doi: 10.4103/2153-3539.103013
35. Pfeifer GP, Hainaut P. Next-generation sequencing: emerging lessons on the origins of human cancer. *Curr Opin Oncol.* (2011) 23:62–8. doi: 10.1097/CCO.0b013e3283414d00
36. Uddin M, Goodman M, Erez O, Romero R, Liu G, Islam M, et al. Distinct genomic signatures of adaptation in pre- and postnatal environments during human evolution. *Proc Natl Acad Sci USA.* (2008) 105:3215–20. doi: 10.1073/pnas.0712400105
37. Doostparast Torshizi A, Wang K. Next-generation sequencing in drug development: target identification and genetically stratified clinical trials. *Drug Discov Today.* (2018) 23:1776–83. doi: 10.1016/j.drudis.2018.05.015
38. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* (2010) 11:31–46. doi: 10.1038/nrg2626
39. Wang Y, Yang Q, Wang Z. The evolution of nanopore sequencing. *Front Genet.* (2014) 5:449. doi: 10.3389/fgene.2014.00449
40. Makalowski W, Shabardina V. Bioinformatics of nanopore sequencing. *J Hum Genet.* (2020) 65:61–7. doi: 10.1038/s10038-019-0659-4
41. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA.* (1996) 93:13770–3. doi: 10.1073/pnas.93.24.13770
42. Luan B, Aksimentiev A. Control and reversal of the electrophoretic force on DNA in a charged nanopore. *J Phys Condens Matter.* (2010) 22:454123. doi: 10.1088/0953-8984/22/45/454123
43. Luan B, Peng H, Polonsky S, Rossnagel S, Stolovitzky G, Martyna G. Base-by-base ratcheting of single stranded DNA through a solid-state nanopore. *Phys Rev Lett.* (2010) 104:238103. doi: 10.1103/PhysRevLett.104.238103
44. Young JS, Prados MD, Butowski N. Using genomics to guide treatment for glioblastoma. *Pharmacogenomics.* (2018) 19:1217–29. doi: 10.2217/pgs-2018-0078

45. Davalos V, Esteller M. MicroRNAs and cancer epigenetics: a macrorevolution. *Curr Opin Oncol.* (2010) 22:35–45. doi: 10.1097/CCO.0b013e328333dcbb
46. Espada J, Esteller M. DNA methylation and the functional organization of the nuclear compartment. *Semin Cell Dev Biol.* (2010) 21:238–46. doi: 10.1016/j.semcdb.2009.10.006
47. Ashley EA. Towards precision medicine. *Nat Rev Genet.* (2016) 17:507–22. doi: 10.1038/nrg.2016.86
48. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* (2008) 455:1061–8. doi: 10.1038/nature07385
49. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med.* (2009) 360:765–73. doi: 10.1056/NEJMoa0808710
50. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* (2010) 17:98–110. doi: 10.1016/j.ccr.2009.12.020
51. Pang L, Hu J, Li F, Yuan H, Yan M, Liao G, et al. Discovering Rare genes contributing to cancer stemness and invasive potential by GBM Single-cell transcriptional analysis. *Cancers.* (2019) 11:2025. doi: 10.3390/cancers11122025
52. Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, et al. Single-cell RNA-Seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* (2017) 21:1399–410. doi: 10.1016/j.celrep.2017.10.030
53. Lopes MB, Vinga S. Tracking intratumoral heterogeneity in glioblastoma via regularized classification of single-cell RNA-Seq data. *BMC Bioinformatics.* (2020) 21:59. doi: 10.1186/s12859-020-3390-4
54. Francis JM, Zhang CZ, Maire CL, Jung J, Manzo VE, Adalsteinsson VA, et al. EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* (2014) 4:956–71. doi: 10.1158/2159-8290.CD-13-0879
55. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science.* (2014) 344:1396–401. doi: 10.1126/science.1254257
56. Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, et al. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc Natl Acad Sci USA.* (2015) 112:851–6. doi: 10.1073/pnas.1320611111
57. Muller S, Liu SJ, Di Lullo E, Malatesta M, Pollen AA, Nowakowski TJ, et al. Single-cell sequencing maps gene expression to mutational phylogenies in PDGF- and EGF-driven gliomas. *Mol Syst Biol.* (2016) 12:889. doi: 10.15252/msb.20166969
58. Abou-El-Ardat K, Seifert M, Becker K, Eisenreich S, Lehmann M, Hackmann K, et al. Comprehensive molecular characterization of multifocal glioblastoma proves its monoclonal origin and reveals novel insights into clonal evolution and heterogeneity of glioblastomas. *Neuro Oncol.* (2017) 19:546–57. doi: 10.1093/neuonc/now231
59. Kim EL, Sorokin M, Kantelhardt SR, Kalasauskas D, Sprang B, Fauss J, et al. Intratumoral heterogeneity and longitudinal changes in gene expression predict differential drug sensitivity in newly diagnosed and recurrent glioblastoma. *Cancers.* (2020) 12:520. doi: 10.3390/cancers12020520
60. Synhaeve NE, van den Bent MJ, French PJ, Dinjens WNM, Atmodimedjo PN, Kros JM, et al. Clinical evaluation of a dedicated next generation sequencing panel for routine glioma diagnostics. *Acta Neuropathol Commun.* (2018) 6:126. doi: 10.1186/s40478-018-0633-y
61. Parker NR, Hudson AL, Khong P, Parkinson JF, Dwight T, Ikin RJ, et al. Intratumoral heterogeneity identified at the epigenetic, genetic and transcriptional level in glioblastoma. *Sci Rep.* (2016) 6:22477. doi: 10.1038/srep22477
62. Kim J, Lee IH, Cho HJ, Park CK, Jung YS, Kim Y, et al. Spatiotemporal evolution of the primary glioblastoma genome. *Cancer Cell.* (2015) 28:318–28. doi: 10.1016/j.ccell.2015.07.013
63. Choi S, Yu Y, Grimmer MR, Wahl M, Chang SM, Costello JF. Temozolomide-associated hypermutation in gliomas. *Neuro Oncol.* (2018) 20:1300–9. doi: 10.1093/neuonc/noy016
64. Muscat AM, Wong NC, Drummond KJ, Algar EM, Khasraw M, Verhaak R, et al. The evolutionary pattern of mutations in glioblastoma reveals therapy-mediated selection. *Oncotarget.* (2018) 9:7844–58. doi: 10.18632/oncotarget.23541
65. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology.* (2013) 38:23–38. doi: 10.1038/npp.2012.112
66. Moorcraft SY, Gonzalez D, Walker BA. Understanding next generation sequencing in oncology: a guide for oncologists. *Crit Rev Oncol Hematol.* (2015) 96:463–74. doi: 10.1016/j.critrevonc.2015.06.007
67. Romani M, Pistillo MP, Banelli B. Epigenetic targeting of glioblastoma. *Front Oncol.* (2018) 8:448. doi: 10.3389/fonc.2018.00448
68. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell.* (2007) 128:707–19. doi: 10.1016/j.cell.2007.01.015
69. Marino-Ramirez L, Kann MG, Shoemaker BA, Landsman D. Histone structure and nucleosome stability. *Expert Rev Proteomics.* (2005) 2:719–29. doi: 10.1586/14789450.2.5.719
70. Jenuwein T, Allis CD. Translating the histone code. *Science.* (2001) 293:1074–80. doi: 10.1126/science.1063127
71. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature.* (2012) 482:226–231. doi: 10.1038/nature10833
72. Gielen GH, Gessi M, Hammes J, Kramm CM, Waha A, Pietsch T. H3F3A K27M mutation in pediatric CNS tumors: a marker for diffuse high-grade astrocytomas. *Am J Clin Pathol.* (2013) 139:345–9. doi: 10.1309/AJCPABOHBC33FVMO
73. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature.* (2009) 462:739–44. doi: 10.1038/nature08617
74. Baylin SB. (2005). DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol.* 2 (Suppl. 1):S4–11. doi: 10.1038/ncponc0354
75. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* (2011) 21:381–95. doi: 10.1038/cr.2011.22
76. Romani M, Daga A, Forlani A, Pistillo MP, Banelli B. Targeting of histone demethylases KDM5A and KDM6B inhibits the proliferation of temozolomide-resistant glioblastoma cells. *Cancers.* (2019) 11:878. doi: 10.3390/cancers11060878
77. Williams MJ, Singleton WG, Lowis SP, Malik K, Kurian KM. Therapeutic targeting of histone modifications in adult and pediatric high-grade glioma. *Front Oncol.* (2017) 7:45. doi: 10.3389/fonc.2017.00045
78. Hiramatsu H, Kobayashi K, Kobayashi K, Haraguchi T, Ino Y, Todo T, et al. The role of the SWI/SNF chromatin remodeling complex in maintaining the stemness of glioma initiating cells. *Sci Rep.* (2017) 7:889. doi: 10.1038/s41598-017-00982-3
79. Ganguly D, Sims M, Cai C, Fan M, Pfeffer LM. Chromatin remodeling factor BRG1 regulates stemness and chemosensitivity of glioma initiating cells. *Stem Cells.* (2018) 36:1804–15. doi: 10.1002/stem.2909
80. McCabe MT, Brandes JC, Vertino PM. Cancer DNA methylation: molecular mechanisms and clinical implications. *Clin Cancer Res.* (2009) 15:3927–37. doi: 10.1158/1078-0432.CCR-08-2784
81. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell.* (2014) 156:45–68. doi: 10.1016/j.cell.2013.12.019
82. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer.* (2011) 2:607–17. doi: 10.1177/1947601910393957
83. Ratel D, Ravanat JL, Berger F, Wion D. N6-methyladenine: the other methylated base of DNA. *Bioessays.* (2006) 28:309–15. doi: 10.1002/bies.20342
84. Iyer LM, Zhang D, Aravind L. Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays.* (2016) 38:27–40. doi: 10.1002/bies.201500104
85. Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature.* (2016) 532:329–33. doi: 10.1038/nature17640
86. Nagarajan RP, Costello JF. Epigenetic mechanisms in glioblastoma multiforme. *Semin Cancer Biol.* (2009) 19:188–97. doi: 10.1016/j.semcancer.2009.02.005

87. Etcheverry A, Aubry M, de Tayrac M, Vauleon E, Boniface R, Guenot F, et al. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics.* (2010) 11:701. doi: 10.1186/1471-2164-11-701
88. Smith AA, Huang YT, Eliot M, Houseman EA, Marsit CJ, Wiencke JK, et al. A novel approach to the discovery of survival biomarkers in glioblastoma using a joint analysis of DNA methylation and gene expression. *Epigenetics.* (2014) 9:873–83. doi: 10.4161/epi.28571
89. Pangeni RP, Zhang Z, Alvarez AA, Wan X, Sastry N, Lu S, et al. Genome-wide methylomic and transcriptomic analyses identify subtype-specific epigenetic signatures commonly dysregulated in glioma stem cells and glioblastoma. *Epigenetics.* (2018) 13:432–48. doi: 10.1080/15592294.2018.1469892
90. Cadieux B, Ching TT, VandenBerg SR, Costello JF. Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. *Cancer Res.* (2006) 66:8469–76. doi: 10.1158/0008-5472.CAN-06-1547
91. Wenger A, Ferreyra Vega S, Kling T, Bontell TO, Jakola AS, Caren H. Intratumor DNA methylation heterogeneity in glioblastoma: implications for DNA methylation-based classification. *Neuro Oncol.* (2019) 21:616–27. doi: 10.1093/neuonc/noz011
92. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell.* (2016) 164:550–63. doi: 10.1016/j.cell.2015.12.028
93. Westphal M, Lamszus K. Circulating biomarkers for gliomas. *Nat Rev Neurol.* (2015) 11:556–66. doi: 10.1038/nrneurol.2015.171
94. Daniel P, Sabri S, Chaddad A, Meehan B, Jean-Claude B, Rak J, et al. Temozolomide induced hypermutation in glioma: evolutionary mechanisms and therapeutic opportunities. *Front Oncol.* (2019) 9:41. doi: 10.3389/fonc.2019.00041
95. Thomas L, Di Stefano AL, Ducray F. Predictive biomarkers in adult gliomas: the present and the future. *Curr Opin Oncol.* (2013) 25:689–94. doi: 10.1097/CCO.0000000000000002
96. Smrdel U, Popovic M, Zwitter M, Bostjancic E, Zupan A, Kovac V, et al. Long-term survival in glioblastoma: methyl guanine methyl transferase (MGMT) promoter methylation as independent favourable prognostic factor. *Radiol Oncol.* (2016) 50:394–401. doi: 10.1515/raon-2015-0041
97. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med.* (2005) 352:997–1003. doi: 10.1056/NEJMoa043331
98. Suva ML. Genetics and epigenetics of gliomas. *Swiss Med Wkly.* (2014) 144:w14018. doi: 10.4414/smw.2014.14018
99. Felsberg J, Rapp M, Loeser S, Fimmers R, Stummer W, Goeppert M, et al. Prognostic significance of molecular markers and extent of resection in primary glioblastoma patients. *Clin Cancer Res.* (2009) 15:6683–93. doi: 10.1158/1078-0432.CCR-08-2801
100. Choi SA, Kwak PA, Park CK, Wang KC, Phi JH, Lee JY, et al. A novel histone deacetylase inhibitor, CKD5, has potent anti-cancer effects in glioblastoma. *Oncotarget.* (2017) 8:9123–33. doi: 10.18632/oncotarget.13265
101. Galanis E, Jaeckle KA, Maurer MJ, Reid JM, Ames MM, Hardwick JS, et al. Phase II trial of vorinostat in recurrent glioblastoma multiforme: a north central cancer treatment group study. *J Clin Oncol.* (2009) 27:2052–8. doi: 10.1200/JCO.2008.19.0694
102. Chinnaiyan P, Chowdhary S, Potthast L, Prabhu A, Tsai YY, Sarcar B, et al. Phase I trial of vorinostat combined with bevacizumab and CPT-11 in recurrent glioblastoma. *Neuro Oncol.* (2012) 14:93–100. doi: 10.1093/neuonc/nor187
103. Lee EQ, Puduvalli VK, Reid JM, Kuhn JG, Lamborn KR, Cloughesy TF, et al. Phase I study of vorinostat in combination with temozolomide in patients with high-grade gliomas: North American Brain Tumor Consortium Study 04-03. *Clin Cancer Res.* (2012) 18:6032–9. doi: 10.1158/1078-0432.CCR-12-1841
104. Ellingson BM, Abrey LE, Nelson SJ, Kaufmann TJ, Garcia J, Chinot O, et al. Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma. *Neuro Oncol.* (2018) 20:1240–50. doi: 10.1093/neuonc/noy053
105. Galanis E, Anderson SK, Miller CR, Sarkaria JN, Jaeckle K, Buckner JC, et al. Phase I/II trial of vorinostat combined with temozolomide and radiation therapy for newly diagnosed glioblastoma: results of alliance N0874/ABTC 02. *Neuro Oncol.* (2018) 20:546–56. doi: 10.1093/neuonc/nox161
106. Ghiaseddin A, Reardon D, Massey W, Mannerino A, Lipp ES, Herndon JEII, et al. Phase II study of bevacizumab and vorinostat for patients with recurrent World Health Organization grade 4 malignant glioma. *Oncologist.* (2018) 23:157–e121. doi: 10.1634/theoncologist.2017-0501
107. Peters KB, Lipp ES, Miller E, Herndon JE II, McSherry F, Desjardins A, et al. Phase I/II trial of vorinostat, bevacizumab, and daily temozolomide for recurrent malignant gliomas. *J Neurooncol.* (2018) 137:349–56. doi: 10.1007/s11060-017-2724-1
108. Iwamoto FM, Lamborn KR, Kuhn JG, Wen PY, Yung WK, Gilbert MR, et al. A phase I/II trial of the histone deacetylase inhibitor romidepsin for adults with recurrent malignant glioma: North American Brain Tumor Consortium Study 03-03. *Neuro Oncol.* (2011) 13:509–16. doi: 10.1093/neuonc/nor017
109. Gurbani SS, Yoon Y, Weinberg BD, Salgado E, Press RH, Cordova JS, et al. Assessing treatment response of glioblastoma to an HDAC inhibitor using whole-brain spectroscopic MRI. *Tomography.* (2019) 5:53–60. doi: 10.18383/j.tom.2018.00031
110. Lee EQ, Reardon DA, Schiff D, Drappatz J, Muzikansky A, Grimm SA, et al. Phase II study of panobinostat in combination with bevacizumab for recurrent glioblastoma and anaplastic glioma. *Neuro Oncol.* (2015) 17:862–7. doi: 10.1093/neuonc/nou350
111. Kerkhof M, Dielemans JC, van Breemen MS, Zwinkels H, Walchenbach R, Taphoorn MJ, et al. Effect of valproic acid on seizure control and on survival in patients with glioblastoma multiforme. *Neuro Oncol.* (2013) 15:961–7. doi: 10.1093/neuonc/not057
112. Krauze AV, Myrehaug SD, Chang MG, Holdford DJ, Smith S, Shih J, et al. A Phase 2 study of concurrent radiation therapy, temozolomide, and the histone deacetylase inhibitor valproic acid for patients with glioblastoma. *Int J Radiat Oncol Biol Phys.* (2015) 92:986–92. doi: 10.1016/j.ijrobp.2015.04.038
113. Tinchon A, Oberndorfer S, Marosi C, Gleiss A, Geroldinger A, Sax C, et al. Haematological toxicity of valproic acid compared to Levetiracetam in patients with glioblastoma multiforme undergoing concomitant radio-chemotherapy: a retrospective cohort study. *J Neurol.* (2015) 262:179–86. doi: 10.1007/s00415-014-7552-z
114. Happold C, Gorlia T, Chinot O, Gilbert MR, Nabors LB, Wick W, et al. Does valproic acid or levetiracetam improve survival in glioblastoma? a pooled analysis of prospective clinical trials in newly diagnosed glioblastoma. *J Clin Oncol.* (2016) 34:731–9. doi: 10.1200/JCO.2015.63.6563
115. Watanabe S, Kuwabara Y, Suehiro S, Yamashita D, Tanaka M, Tanaka A, et al. Valproic acid reduces hair loss and improves survival in patients receiving temozolomide-based radiation therapy for high-grade glioma. *Eur J Clin Pharmacol.* (2017) 73:357–63. doi: 10.1007/s00228-016-2167-1
116. Krauze AV, Mackey M, Rowe L, Chang MG, Holdford DJ, Cooley T, et al. Late toxicity in long-term survivors from a phase 2 study of concurrent radiation therapy, temozolomide and valproic acid for newly diagnosed glioblastoma. *Neurooncol Pract.* (2018) 5:246–50. doi: 10.1093/nop/npy009
117. Valiyaveettil D, Malik M, Joseph DM, Ahmed SF, Kothwal SA, Vijayasaradhi M. Effect of valproic acid on survival in glioblastoma: a prospective single-arm study. *South Asian J Cancer.* (2018) 7:159–62. doi: 10.4103/sajc.sajc_188_17
118. Fulton B, Short SC, James A, Nowicki S, McBain C, Jefferies S, et al. PARADIGM-2: two parallel phase I studies of olaparib and radiotherapy or olaparib and radiotherapy plus temozolomide in patients with newly diagnosed glioblastoma, with treatment stratified by MGMT status. *Clin Transl Radiat Oncol.* (2018) 8:12–6. doi: 10.1016/j.ctro.2017.11.003
119. Lesueur P, Lequesne J, Grellard JM, Dugue A, Coquan E, Brachet PE, et al. Phase I/IIa study of concomitant radiotherapy with olaparib and temozolomide in unresectable or partially resectable glioblastoma: OLA-TMZ-RTE-01 trial protocol. *BMC Cancer.* (2019) 19:198. doi: 10.1186/s12885-019-5413-y
120. Su JM, Thompson P, Adesina A, Li XN, Kilburn L, Onar-Thomas A, et al. A phase I trial of veliparib (ABT-888) and temozolomide in children with recurrent CNS tumors: a pediatric brain tumor consortium report. *Neuro Oncol.* (2014) 16:1661–8. doi: 10.1093/neuonc/nou103
121. Robins HI, Zhang P, Gilbert MR, Chakravarti A, de Groot JF, Grimm SA, et al. A randomized phase I/II study of ABT-888 in combination with temozolomide in recurrent temozolomide resistant glioblastoma: an

NRG oncology RTOG group study. *J Neurooncol.* (2016) 126:309–16. doi: 10.1007/s11060-015-1966-z
122. Blakeley JO, Grossman SA, Chi AS, Mikkelsen T, Rosenfeld MR, Ahluwalia MS, et al. Phase II Study of Iniparib with Concurrent Chemoradiation in Patients with newly diagnosed glioblastoma. *Clin Cancer Res.* (2019) 25:73–9. doi: 10.1158/1078-0432.CCR-18-0110
123. Dermawan JK, Hitomi M, Silver DJ, Wu Q, Sandlesh P, Sloan AE, et al. Pharmacological targeting of the histone chaperone complex fact preferentially eliminates glioblastoma stem cells and prolongs survival in preclinical models. *Cancer Res.* (2016) 76:2432–42. doi: 10.1158/0008-5472.CAN-15-2162
124. Barone TA, Burkhart CA, Safina A, Haderski G, Gurova KV, Purmal AA, et al. Anticancer drug candidate CBL0137, which inhibits histone chaperone FACT, is efficacious in preclinical orthotopic models of temozolomide-responsive and -resistant glioblastoma. *Neuro Oncol.* (2017) 19:186–96. doi: 10.1093/neuonc/now141
125. Jin MZ, Xia BR, Xu Y, Jin WL. Curaxin CBL0137 exerts anticancer activity via diverse mechanisms. *Front Oncol.* (2018) 8:598. doi: 10.3389/fonc.2018.00598
126. Suntsova M, Gaifullin N, Allina D, Reshetun A, Li X, Mendeleeva L, et al. Atlas of RNA sequencing profiles for normal human tissues. *Sci Data.* (2019) 6:36. doi: 10.1038/s41597-019-0043-4
127. Bolouri H, Zhao LP, Holland EC. Big data visualization identifies the multidimensional molecular landscape of human gliomas. *Proc Natl Acad Sci USA.* (2016) 113:5394–9. doi: 10.1073/pnas.1601591113
128. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. *Nature.* (2018) 555:469–74. doi: 10.1038/nature26000
129. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol.* (2008) 26:1146–53. doi: 10.1038/nbt.1495
130. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* (2018) 36:338–45. doi: 10.1038/nbt.4060
131. Sen R, Dolgalev I, Bayin NS, Heguy A, Tsirigos A, Placantonakis DG. Single-Cell RNA Sequencing of Glioblastoma Cells. *Methods Mol Biol.* (2018) 1741:151–70. doi: 10.1007/978-1-4939-7659-1_12
132. Euskirchen P, Bielle F, Labreche K, Kloosterman WP, Rosenberg S, Daniau M, et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta Neuropathol.* (2017) 134:691–703. doi: 10.1007/s00401-017-1743-5
133. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
134. Tandel GS, Biswas M, Kakde OG, Tiwari A, Suri HS, Turk M, et al. A review on a deep learning perspective in brain cancer classification. *Cancers.* (2019) 11:111. doi: 10.3390/cancers11010111
135. Senders JT, Harary M, Stopa BM, Staples P, Broekman MLD, Smith TR, et al. Information-based medicine in glioma patients: a clinical perspective. *Comput Math Methods Med.* (2018) 2018:8572058. doi: 10.1155/2018/8572058
136. Borisov N, Buzdin A. New paradigm of machine learning (ML) in personalized oncology: data trimming for squeezing more biomarkers from clinical datasets. *Front Oncol.* (2019) 9:658. doi: 10.3389/fonc.2019.00658
137. Chen PC, Gadepalli K, MacDonald R, Liu Y, Kadowaki S, Nagpal K, et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat Med.* (2019) 25:1453–57. doi: 10.1038/s41591-019-0539-7
138. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* (2006) 34:D668–672. doi: 10.1093/nar/gkj067
139. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today.* (2017) 22:1680–5. doi: 10.1016/j.drudis.2017.08.010
140. Borisov N, Tkachev V, Suntsova M, Kovalchuk O, Zhavoronkov A, Muchnik I, et al. A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency. *Cell Cycle.* (2018) 17:486–91. doi: 10.1080/15384101.2017.1417706
141. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med.* (2019) 25:1337–40. doi: 10.1038/s41591-019-0548-6
142. Kim YW, Koul D, Kim SH, Lucio-Eterovic AK, Freire PR, Yao J, et al. Identification of prognostic gene signatures of glioblastoma: a study based on TCGA data analysis. *Neuro Oncol.* (2013) 15:829–39. doi: 10.1093/neuonc/not024
143. Wong KK, Rostomily R, Wong STC. Prognostic gene discovery in glioblastoma patients using deep learning. *Cancers.* (2019) 11:53. doi: 10.3390/cancers11010053
144. Young JD, Cai C, Lu X. Unsupervised deep learning reveals prognostically relevant subtypes of glioblastoma. *BMC Bioinformatics.* (2017) 18(Suppl. 11):381. doi: 10.1186/s12859-017-1798-2
145. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell.* (2013) 155:462–77. doi: 10.1016/j.cell.2013.09.034

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 29 May 2020
doi: 10.3389/fonc.2020.00666

Lynch Syndrome Germline Mutations in Breast Cancer: Next Generation Sequencing Case-Control Study of 1,263 Participants

Aleksey G. Nikitin[1†], Daria A. Chudakova[2,3†], Rafael F. Enikeev[4], Dina Sakaeva[5], Maxim Druzhkov[6], Leyla H. Shigapova[7], Olga I. Brovkina[8], Elena I. Shagimardanova[9], Oleg A. Gusev[9,10] and Marat G. Gordiev[4,11*]*

[1] Pulmonology Research Institute, Federal Medical-Biological Agency of Russia, Moscow, Russia, [2] School of Biological Sciences, University of Auckland, Auckland, New Zealand, [3] Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand, [4] Tatarstan Cancer Centre, Kazan, Russia, [5] Department of Pharmacology, Bashkir State Medical University, Ufa, Russia, [6] Druzhkov Clinic Ltd., Kazan, Russia, [7] Extreme Biology Lab, Scientific and Clinical Center for Precision and Regenerative Medicine, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia, [8] Federal Research and Clinical Centre, Federal Medical-Biological Agency of Russia, Moscow, Russia, [9] Kazan (Volga Region) Federal University, Kazan, Russia, [10] KFU-RIKEN Translational Genomics Unit, RIKEN Cluster for Science, Technology and Innovation Hub, RIKEN, Yokohama, Japan, [11] National Bioservice, Saint Petersburg, Russia

Genome instability—the increased tendency of acquiring mutations in the genome and ability of a cell to tolerate high mutation burden—is one of the drivers of cancer. Genome instability results from many causes including defects in DNA repair systems. Previously, it has been shown that germline pathogenic mutations in DNA Mismatch Repair (MMR) pathway cause cancer-predisposing Lynch Syndrome. We proposed that Lynch Syndrome-related germline mutations (LS-mutations) are associated with breast cancer (BC). In this study, we performed Targeted Next-Generation Sequencing of MMR pathway genes *MLH1, MSH2, MSH6, EPCAM,* and *PMS2* in a cohort of 711 patients with hereditary BC, 60 patients with sporadic BC, and 492 healthy donors. Sixty-nine patients (9.7%) with hereditary BC harbored at least one germline mutation in the MMR pathway genes, of them 32 patients (4.5%) harbored mutations in MMR pathway genes which we define as pathogenic or likely pathogenic, and of them 26 patients (3.6%) did not have any pathogenic mutations in DDR pathway genes, compared to two mutations in MMR pathway genes (0.4%) detected in a group of 492 healthy donors [$p = 0.00013$, $OR = 8.9$ (CI 95% 2.2–78.4)]. Our study demonstrates that LS-mutations are present in patients with hereditary BC more frequently than in healthy donors, and that there is an association of hereditary BC and mutations c.1321G>A in *MLH1*, c.260C>G and c.2178G>C in *MSH2*, c.3217C>T in *MSH6*, c.1268C>G and c.86G>C in *PMS2* genes. This finding provides a rationale for including pathogenic LS-mutations into genetic counseling tests for patients with hereditary BC.

Keywords: breast cancer, *EPCAM*, Lynch syndrome, *MLH1, MSH2, MSH6, PMS2*, Targeted Next-Generation Sequencing

OPEN ACCESS

Edited by:
Ye Wang,
Qingdao University Medical College, China

Reviewed by:
Yuxiang Zhao,
Biotrans Technology Co., LTD, China
Paola Parrella,
Casa Sollievo della Sofferenza (IRCCS), Italy

****Correspondence:***
Oleg A. Gusev
oleg.gusev@riken.jp
Marat G. Gordiev
marat7925@gmail.com

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *02 November 2019*
Accepted: *09 April 2020*
Published: *29 May 2020*

Citation:
Nikitin AG, Chudakova DA, Enikeev RF, Sakaeva D, Druzhkov M, Shigapova LH, Brovkina OI, Shagimardanova EI, Gusev OA and Gordiev MG (2020) Lynch Syndrome Germline Mutations in Breast Cancer: Next Generation Sequencing Case-Control Study of 1,263 Participants. Front. Oncol. 10:666. doi: 10.3389/fonc.2020.00666

INTRODUCTION

Genome instability is one of the key hallmarks of cancer (1). The stability of genome is maintained in a cell by many mechanisms including DNA repair. The repair of DNA single-base mismatch and insertion/deletion loops occurring during DNA replication is executed by the Mismatch Repair (MMR) pathway ensuring mutation avoidance and precision of DNA replication (2). The MMR pathway proteins also take part in other cellular processes, and the whole spectrum of their diverse roles is yet to be understood (3).

The cancer-predisposing Lynch syndrome (LS) is an autosomal dominant disorder caused by germline mutations in the MMR pathway genes, mainly mutL homolog 1 (*MLH1*), mutS homolog 2 (*MSH2*), mutS homolog 6 (*MSH6*), epithelial cell adhesion molecule (*EPCAM*), and post-meiotic segregation increased 2 (*PMS2*) (3, 4). The LS can also result from mutations located in flanking regions of MMR genes (5, 6). Predominantly, the LS is caused by the presence of loss-of-function germline mutations in *MLH1* and *MSH2* genes (7), mutations in *MSH6* and *PMS2* are less frequent, and *EPCAM* is the less frequently mutated gene in the LS. The individuals with the LS tend to exhibit nucleotide loss or gain within the DNA microsatellite loci (microsatellite instability, MSI) (8), and their cells have a "mutator phenotype" which is causative to many types of malignancies.

The LS, originally identified as a disorder associated with colorectal cancer and previously known as hereditary non-polyposis colorectal cancer, is currently defined as a multi-tumor syndrome. The LS is found to be related to plethora of extracolonic malignancies including cancers of urinary tract, endometrial, small bowel and others (9).

Whether LS-associated malignancies include both ovarian cancer (OC) and breast cancer (BC) is yet an open topic of discussion. The link between the LS-associated germline mutations and hereditary OC has been demonstrated in several studies, and it is estimated that 10–15% of hereditary OC are LS-related (10). Recently, germline mutations in *MSH2*, *MSH6,* and *PMS2* genes have been found associated with BC (11, 12). However, neither the revised Amsterdam criteria for LS diagnosis nor the revised Bethesda criteria for MSI tests include BC (13), despite the data suggesting the link between BC and LS. In the recent publication "Lynch syndrome: five unanswered questions" the authors suggest that whether BC should be included or excluded from LS-related tumors is a perhaps the most important question in a field of "LS tumor spectrum" (14).

BC has a strong hereditary component and in many cases is caused by germline mutations in the predisposition genes such as DNA damage recognition and repair (DDR) genes *BRCA1*, *BRCA2,* and others, which are currently included into the multi-gene panels for BC risk assessment (15, 16). Nevertheless, a sizeable proportion of the patients with familial history of BC do not carry germline mutations in the currently known genes. Although some of such cases might be explained in part by the presence of heritable epigenetic marks ("epimutations") leading to the disease (17), it's possible that germline pathogenic predisposing mutations in other, yet unknown genes exist, but are not identified yet (the "missing heritability" phenomena) (18). If BC is a part of LS, then pathogenic mutations in MMR pathway genes associated with BC should be included into the clinical genetic testing panels.

The Next Generation Sequencing (NGS) technologies are instrumental tools in molecular diagnostic allowing rapid and simultaneous analysis of broad panels of disease-associated germline mutations within multiple genes. The results of NGS-based tests for genetic risk assessment are concordant with conventional diagnostic methods, as demonstrated for hereditary BC and/or OC (19). Based on our pilot study indicating the presence of germline mutations within the MMR genes in patients with BC and familial history of cancer (20), we proposed that BC is a part of LS. To test this, we performed Targeted NGS of *MLH1, MSH2, MSH6, EPCAM,* and *PMS2* genes in a big cohort comprising of 711 patients with hereditary BC, 60 patients with sporadic cancer, and 492 healthy donors from Volga and Central Federal Districts, Russian Federation.

Our study demonstrates that the frequencies of the most of causative LS germline mutations are higher in patients with hereditary BC compared to healthy population control, and finds the association of several germline mutations within the genes *MLH1, MSH2, MSH6,* and *PMS2* with hereditary BC. This finding provides insights into the biology of LS and BC and supports including LS-associated mutations in genetic tests for the patients with hereditary BC in our study population.

MATERIALS AND METHODS

Study Population

The study included 711 participants with hereditary BC and 60 participants with sporadic BC receiving a treatment for BC (chemotherapy and/or surgical treatment) at several medical centers in Volga Federal District, Russian Federation. The control comprised 492 healthy donors from Volga and Central Federal Districts, Russian Federation. The study participants self-identified with Slavic, Tatar or Bashkir ethnicities, and we ensured that both case and control group similarly reflect the ethnic diversity of the population from this geographical region. The criteria for inclusion into the patient's cohort, based on age of cancer manifestation, familial history of cancer, and clinical-pathological characteristics of the disease were previously described in our smaller scale pilot study (20). In particular, patients were included into the hereditary BC group if they had (1) BC diagnosis and familial history of any cancer (including kidney cancer, esophageal cancer, stomach cancer, lung cancer, sarcoma, colon cancer, leukemia, breast cancer, and ovarian cancer) in the first-, or second-, or third-degree relative or (2) BC manifestation at early age (before 30 y.o.) or (3) manifestation of triple negative BC at early age (before 35 y.o.). The estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67 status of the patients was established by clinical pathologists as part of the routine patient care. The clinical and demographical characteristics for the patients are shown

in **Supplementary Table 1**. All study participants provided informed consent prior enrolling to the study in accordance with Declaration of Helsinki.

Targeted Next-Generation Sequencing

Genomic DNA was isolated from the whole blood samples using DNA Blood Mini QIAcube Kit (Qiagen), and 100 ng of DNA was used to generate Targeted NGS libraries. The target enrichment, sequencing and analysis were performed as described previously (20, 21) with slight modifications. In particular, KAPA HyperPlus (Roche) was used for library preparation and DNA enzymatic fragmentation, DNA was fragmented to the size of 150–200 b.p. Concentration of the DNA in the library was measured by Qubit (ThermoFisher Scientific) following manufacturer's instructions, the presence of the primer dimers was assessed using Agilent High Sensitivity DNA Kit (Agilent), the optimal length of the fragment including adapter was 290–330 b.p. Next, libraries were combined and hybridized with SeqCap EZ Choice (Roche), following manufacturer's recommendations. Hybridization was performed at +47°C for 16 h. SeqCap Capture beads were used for enrichment, and amplification was performed using KAPA HiFi HS MasterMix (Roche). Sequencing was performed using MiSeq (Illumina). The gene panel included *MLH1, MSH2, MSH6, EPCAM,* and *PMS2* genes (MMR pathway genes). In the study participants carrying mutations in the MMR pathway mutational status of the other genes associated with BC, namely namely *ATM, BRCA1, BRCA2, APC,* and *TP53* genes (DDR pathway genes), was also determined by NGS. Patients carrying pathogenic mutations in DDR pathway genes were excluded from analysis. *In-silico* tools SIFT, PolyPhen2, MutationTaster, CADD, DANN, M-CAP, and REVEL were used for the prediction of pathogenicity. All sequencing data were submitted to SRA database and can be accessed at https://www.ncbi.nlm.nih.gov/sra/PRJNA588789.

Statistical Analysis

The data was analyzed using standard statistical tests as described previously (21). In particular, a two-tailed Fisher exact test was performed using the R software v.3.3 (*fisher.test* function). Statistical as significance was defined a $p < 0.05$.

RESULTS AND DISCUSSION

While the role of LS in hereditary OC is established and widely accepted, it is a long-standing question whether BC should also be classified under an umbrella of LS (22), as results of previous studies are inconsistent and contradictory. This lack of consistency might be explained by the inter-population and inter-ethnic differences and result from the unique ethnic-specific genetic traits within the study cohorts (23). It is apparent now that the epidemiology and distribution of pathogenic germline mutations in BC are population-specific (24, 25). Thus, the population- and ethnic background of the patient should be considered at the stage of the cancer risk genetic evaluation (23, 26), as genetic risks might be mis-estimated if based on a data obtained in a study population with different ancestral (and, thus, genetic) background. Moreover, genetic studies of the complex hereditary diseases in understudied populations provide a unique opportunity to identify novel genetic markers. Currently, a large body of data on genetics of familial BC exists for some well-studied populations and ethnic groups, while some populations and ethnicities remain understudied, resulting in a so-called "social inequity in cancer." Hence, further studies focusing on ethnic-specific and population-specific aspects of hereditary BC are of definite clinical value.

Previously, we demonstrated that the spectrum and frequencies of pathogenic nucleotide variants in DDR pathway genes in Tatar patients with familial OC and/or BC from Kazan region of the Volga Federal District of the Russian Federation are different from ones reported for the patients of Slavic descent from Moscow (21). Here, we applied Targeted NGS to determine the prevalence and spectrum of germline mutations in MMR pathway genes in the patients with hereditary BC from the Volga and Central Federal Districts of the Russian Federation and in healthy donors of the similar ethnical backgrounds. We performed Targeted NGS and, in a cohort comprising 711 participants with hereditary BC, identified in 17 participants 10 mutations in *MLH1* gene (c.945C>G, c.1637A>G, c.803A>G, c.1321G>A, c.1937A>G, c.-7C>T, c.2194A>G, c.472A>G, c.2194A>G, andc.1090A>C), in 9 participants 5 mutations in *MSH2* gene (c.260C>G, c.2178G>C, c.2178G>C, c.2197G>A, c.815C>T), in 19 participants 19 mutations in *MSH6* gene (c.4004A>C, c.2291C>A, c.2156C>T, c.2673C>G, c.893G>A, c.2554_2556del, c.3674C>T, c.3674C>T, c.3986C>T, c.3674C>T, c.2503C>G, c.3217C>T, c.3254dupC, c.3259C>T, c.1063G>A, c.3951T>G, c.968C>G, c.1481C>T, c.3151G>A), in 6 participants 3 mutations in *EPCAM* gene (c.557A>C, c.859-3C>G, c.272A>T), and in 20 participants 14 mutations in *PMS2* gene (c.1642G>A, c.1268C>G, c.1268C>G, c.86G>C, c.944G>A, c.1567T>A, c.2438G>A, c.1399G>A, c.2149G>A, c.1630G>A,c.1753C>T, c.1595A>G, c.1630G>A, c.1901A>G). In a group of 60 patients with sporadic BC we found no germline mutations in MMR pathway genes. In a group of 462 healthy participants, one germline mutation in *EPCAM* gene (c.859-3C>G) was found in 2 participants. All mutations were heterozygous. The mutations included variants of unknown/uncertain significance (VUS), Likely pathogenic and Pathogenic mutations, based on The Human Mutations Database (HGMD) and the ClinVar database.

The presence of the mutations in the DDR pathway genes was also assessed in the study participants currying mutations in the MMR pathway. Among the patients with mutations in MMR pathway, 17 (2.4%) also harbored pathogenic mutations in the DDR pathway genes (**Supplementary Table 2**). If such mutations in DDR were present, the OR for mutation in MMR pathway genes was calculated twice, including and excluding cases with mutations in DDR. Only OR calculated for cases with germline mutations in MMR pathway genes and without pathogenic germline mutations in DDR genes was used to

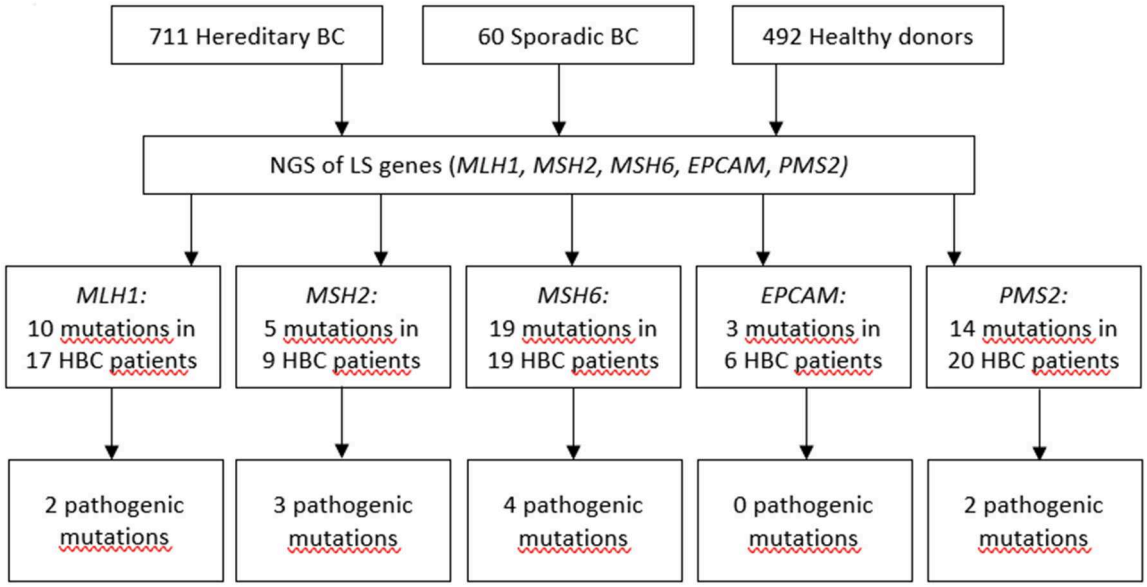


FIGURE 1 | Pathogenic mutations detected in LS genes.

assess the pathogenicity of the mutation. Some of mutations were detected more than once in the patient cohort, the most recurrent ones were c.1321G>A in *MLH1*, c.260C>G in *MSH2,* and c.86G>C in *PMS2* genes. The spectrum and frequencies of the mutations in the study cohort are shown in **Supplementary Table 2**. Sixty-nine patients (9.7%) harbored at least one germline mutation in the MMR pathway genes, of them thirty-two patients harbored mutations in MMR pathway genes which we define as pathogenic or likely pathogenic, and of them twenty-six patients (3.6%) did not have any pathogenic mutations in DDR pathway genes, compared to two mutations in MMR pathway genes (0.4%) detected in a group of 492 healthy donors [$p = 0.00013$, $OR = 8.9$ (CI 95% 2.2–78.4] (**Figure 1**, **Supplementary Tables 2, 3**). The age of occurrence of the clinical manifestations of disease in hereditary BC patients with and without LS-mutations was 45.3 ± 9.7 and 47.1 ± 11.3 years, respectively, compared to 58.9 ± 9.3 in a group of patients with sporadic BC. The percentage of HER2+ patients in hereditary BC patients with and without LS-mutations was 34 and 30%, respectively.

We detected 10 different mutations in *MLH1, 5* in *MSH2, 19* in *MSH6, 3* in *EPCAM,* and 14 in *PMS2* genes in a cohort of participants with hereditary BC, while only one of them was found in the reference healthy control (**Figure 1**, **Supplementary Table 2**). In the samples from the Exome Aggregation Consortium database (http://gnomad.broadinstitute.org/) the frequencies of these mutations are extremely low and similar to the data obtained in our study population in the control group (**Supplementary Table 2**). In other populations, the frequency of *MSH6* gene mutations was determined as 0.2% in a study performed in Germany in a group of patients with BC and/or OC (27), and in a study performed in USA in a group of females with stages I to III of BC (28). The rate of Ashkenazi Jewish founder mutations c.3984_3987dupGTCA and c.3959_3962delCAAG within *MSH6* gene in a study population comprising 1016 participants with familial history of BC and/or OC was 0.49% (29). Additionally, the recent case report suggests an association between sporadic BC and biallelic mutations in *MSH6* (30). In a small cohort of triple-negative BC patients with early onset or familial history of cancer the frequency of *MSH2* gene mutations was 4% (31). The cohort study of several families in UK suggested germline mutation carriers in MLH1 gene are at moderate risk of BC (32). Recently, study by Roberts et al. reported that *MSH6* and *PMS2* germline pathogenic variants are associated with increased risk of BC (12). However, this association was not confirmed in the other study (33). It was suggested that the ascertainment bias might has affected the interpretation of the data by Roberts et al., as the study cohort was enriched with participants with BC diagnosis, and the BC risk in the total study cohort doubled BC risk in the reference cohort (general population) (33). Other studies found either no association of *PMS2* with BC (34), or demonstrated that carriers of mutations in *PMS2* gene had significant standardized incidence ratios for OC and BC (35).

The results of our study are in line with the previous studies demonstrating an association of germline mutations in *MLH1, MSH2, MSH6,* and *PMS2* genes with hereditary BC. There was also a tendency for the carriers of LS-mutations to have earlier manifestation of the disease (45.3 ± 9.7 y.o. with LS-mutations and 47.1 ± 11.3 y.o. without). However, given the importance of clinical decisions in BC risk assessment,

we agree with Evans et al. (34) that the decision of the genetic counseling specialist should be based "on the overall evidence available."

We also suggest changing the classification of several mutations characterized in this study, based on variant interpretation standards and guidelines of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. In particular, currently the c.1321G>A in *MLH1* gene is defined in databases as VUS, but we suggest that it is a Likely pathogenic mutation [$OR = 10.9$ (3.4–26.6), $p = 0.0001$, **Supplementary Table 2**]. The 260C>G in *MSH2* gene is also VUS, but we suggest it should be classified as a Pathogenic mutation [$OR = 361.4$ (51.9–4387.9), $p = 1.359e\text{-}08$, **Supplementary Table 2**]. The c.2178G>C in *MSH2* is classified as VUS, but based on our study we suggest that it is a Likely pathogenic mutation [$OR = 53.4$ (5.2–297.8), $p = 0.001$, **Supplementary Table 2**]. The c.3217C>T in *MSH6* gene is classified as VUS, but we suggest it is a Likely pathogenic mutation [$OR = 19.5$ (3.8–63.2), $p = 0.0007$, **Supplementary Table 2**]. Finally, the c.859-3C>G in *EPCAM* gene is classified as VUS, but we propose that it is a Likely Benign mutation, based on its relatively high frequency in our healthy control group (much higher than in the gnomAD).

The large number of the mutations within the *MSH6* gene among the patients in our study cohort is of particular interest. The currently available MSI tests based on mononucleotide repeat markers BAT25, BAT26, NR21, NR24, and MONO27 often give false-negative results in case of *MSH6* gene (36), thus other methods of MSI detection, such as IHC, are more sensitive. Given that *MSH6* is the most frequently mutated LS gene in BC in our study population, we suggest using IHC as a preferential method of MSI assessment in hereditary BC.

There are several limitations of our study. Firstly, the populations of the Volga and Central Federal District comprise several ethnicities, thus differences in mutation frequencies between control and case group might be attributed to specific genetic traits of the participants of different descent, rather than biology of LS and BC. This highlights an importance of ensuring that ethnic groups are equally represented in both case and control groups. We tried to address this by recruiting a large number of participants and ensuring that both case and control group similarly reflect the ethnic diversity of the population. Secondly, some patients with germline mutations in MMR pathway genes may still have a tumor with functional MMR proteins, as previously described (37). This advocates for more rigorous testing, especially when choosing tailored therapy. In future studies, this should be addressed by applying Targeted NGS for analysis of both tumor biopsy and whole blood tissue samples, which was beyond the scope of the current study.

The fundamental step in cancer risk evaluation, prevention and clinical surveillance of hereditary cancers is a detection of predisposing germline mutations in individuals with familial history of cancer. In many cases the therapy decisions are also guided by the genotype of a patient (the clinical approach known as personalized therapy) as carriers of different allelic variants may respond to the treatment differently. For example, *MSH2*-deficient cancer cells are selectively sensitive to Methotrexate, and it's been proposed that patients with *MSH2* deficiency will respond to the Methotrexate therapy (38). Another compound selectively targeting MMR-deficient cancer cells is FDA approved drug Triamterene (39). It has also been shown that patients with high MSI and mutations in MMR genes have favorable response to the PD-1 blockade immunotherapy in a broad spectrum of cancers (40). Finally, several mutations in MMR pathway genes have been found associated with radiosensitivity in BC patients (41). Thus, determining mutation status of the MMR pathway genes can guide personalized therapy. Additionally, the carriers of LS-related germline mutations identified by the genetic tests benefit from the chemoprevention therapies (42), that would not otherwise be subscribed in the absence of a suggestive clinical evidence and prior to the manifestation of the disease.

The pathogenic mutations in genes from MMR pathway result in compromised DNA repair. The defects in DNA repair are associated with increased neoantigen load and linked to the elevated expression of immunosuppressive PD-L1 by the cancer cells (43). Patients with tumors expressing high level of PD-L1 benefit from the immune checkpoint blockade therapy, thus identification of such patients has important clinical implications. The correlation between the level of PD-L1 and both methylation of *BRCA1* gene and its mutation status has been found in OC (44, 45). We propose that in hereditary BC the PD-L1 level may correlate with the presence of pathogenic mutations in the genes from both MMR and DDR pathways (such as *BRCA1/2* and others), and suggest that in future studies such correlation should be assessed as a potential clinical biomarker.

In conclusion, our study demonstrates the relatively frequent presence of the germline LS-mutations in the patients with hereditary BC, and association of hereditary BC with c.1321G>A in *MLH1*, c.260C>G and c.2178G>C in *MSH2*, c.3217C>T in *MSH6*, and c.1268C>G and c.86G>C in *PMS2* genes. We recommend including MMR pathway genes into the multi-gene panels for risk assessment of hereditary BC, based on the overall clinical picture.

DATA AVAILABILITY STATEMENT

All sequencing data were submitted to SRA database and can be accessed at https://www.ncbi.nlm.nih.gov/sra/PRJNA588789.

ETHICS STATEMENT

This studies involving human participants were reviewed and approved by the study was carried out in accordance with the recommendations of International code of medical ethics and local ethical committees of Kazan Federal University and Tatarstan Cancer Center. In accordance with the Declaration of Helsinki all study participants gave written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AN, DS, ES, LS, MG, MD, OB, and RE: data collection, analysis, and interpretation. AN, DC, ES, MG, and OG: study conception and design. AN, DC, OB, and OG: drafting and critical revision of the manuscript.

FUNDING

This work was supported by the Russian Foundation for Basic Research 18415-160009 r_a, the Russian Government Program of Competitive Growth of Kazan, Federal University and FMBA of Russia Government Order. The authors declare that this study received funding from AstraZeneca Russia ESR-17-12934. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

The authors express deepest and heartfelt gratitude to all participants of the study. DC thanks the City of Melbourne's Library Service.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00666/full#supplementary-material

REFERENCES

1. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. (2000) 100:57–70. doi: 10.1016/S0092-8674(00)81683-9
2. Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res*. (2008) 18:85–98. doi: 10.1038/cr.2007.115
3. Peltomäki P. Update on lynch syndrome genomics. *Fam Cancer*. (2016) 15:385–93. doi: 10.1007/s10689-016-9882-8
4. Tutlewska K, Lubinski J, Kurzawski G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. *Hered Cancer Clin Pract*. (2013) 11:9. doi: 10.1186/1897-4287-11-9
5. Hirata K, Kanemitsu S, Nakayama Y, Nagata N, Itoh H, Ohnishi H, et al. HNPCC registry and genetic testing project of the Japanese Society for Cancer of the Colon and Rectum (JSCCR). *J Fam Tumour*. (2003) 3:68–75. doi: 10.18976/jsft.3.2_68
6. Hirata K, Kanemistsu S, Nakayama Y, Nagata N, Itoh H, Ohnishi H, et al. A novel germline mutation of MSH2 in a hereditary nonpolyposis colorectal cancer patient with liposarcoma. *Am J Gastroenterol*. (2006) 101:193–96. doi: 10.1111/j.1572-0241.2005.00308.x
7. Balmaña J, Stockwell DH, Steyerberg EW, Stoffel EM, Deffenbaugh AM, Reid JE, et al. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. *JAMA*. (2006) 296:1469–78. doi: 10.1001/jama.296.12.1469
8. Latham A, Srinivasan P, Kemel Y, Shia J, Bandlamudi C, Mandelker D, et al. Microsatellite instability is associated with the presence of lynch syndrome pan-cancer. *J Clin Oncol*. (2019) 37:286–95. doi: 10.1200/JCO.18.00283
9. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer*. (2015) 15:181–94. doi: 10.1038/nrc3878
10. Bewtra C, Watson P, Conway T, Read-Hippee C, Lynch HT. Hereditary ovarian cancer: a clinicopathological study. *Int J Gynecol Pathol*. (1992) 11:180–7.
11. Goldberg M, Bell K, Aronson M, Semotiuk K, Pond G, Gallinger S, et al. Association between the Lynch syndrome gene MSH2 and breast cancer susceptibility in a Canadian familial cancer registry. *J Med Genet*. (2017) 54:742–6. doi: 10.1136/jmedgenet-2017-104542
12. Roberts ME, Jackson SA, Susswein LR, Zeinomar N, Ma X, Marshall ML, et al. MSH6 and PMS2 germ-line pathogenic variants implicated in Lynch syndrome are associated with breast cancer. *Genet Med*. (2018) 20:1167–74. doi: 10.1038/gim.2017.254
13. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. (2004) 96:261–8. doi: 10.1093/jnci/djh034
14. Castellsagué E, Foulkes WD. Lynch syndrome: five unanswered questions. *Clin Genet*. (2015) 87:503–6. doi: 10.1111/cge.12580
15. Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med*. (2015) 372:2243–57. doi: 10.1056/NEJMsr1501341
16. Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. *Ann Oncol*. (2015) 26:1291–9. doi: 10.1093/annonc/mdv022
17. Joo JE, Dowty JG, Milne RL, Wong EM, Dugué P-A, English D, et al. Heritable DNA methylation marks associated with susceptibility to breast cancer. *Nat Commun*. (2018) 9:867. doi: 10.1038/s41467-018-03058-6
18. Winship I, Southey MC. Gene panel testing for hereditary breast cancer. *Med J Aust*. (2016) 204: 188–90. doi: 10.5694/mja15.01335
19. Lincoln SE, Kobayashi Y, Anderson MJ, Yang S, Desmond AJ, Mills MA, et al. A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. *J Mol Diagn*. (2015) 17:533–44. doi: 10.1016/j.jmoldx.2015.04.009
20. Gordiev M, Shigapova LH, Brovkina O, Enikeev RF, Druzhkov M, Nikitin A, et al. Lynch syndrome-associated hereditary mutations cause breast and ovarian cancer: results from Russian Hereditary Oncogenomics project. *Ann Oncol*. (2018) 29. doi: 10.1093/annonc/mdy270.236
21. Brovkina OI, Shigapova L, Chudakova DA, Gordiev MG, Enikeev RF, Druzhkov MO, et al. The ethnic-specific spectrum of germline nucleotide variants in DNA damage response and repair genes in hereditary breast and ovarian cancer patients of Tatar descent. *Front Oncol*. (2018) 8:421. doi: 10.3389/fonc.2018.00421
22. Ford JM. Is breast cancer a part of Lynch syndrome? *Breast Cancer Res*. (2012) 14:110. doi: 10.1186/bcr3241
23. Jing L, Su L, Ring BZ. Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. *PLoS ONE*. (2014) 9:e97522. doi: 10.1371/journal.pone.0097522
24. Bhaskaran SP, Chandratre K, Gupta H, Zhang L, Wang X, Cui J, et al. Germline variation in BRCA1/2 is highly ethnic-specific: evidence from over 30,000 Chinese hereditary breast and ovarian cancer patients. *Int J Cancer*. (2019) 145:962–73. doi: 10.1002/ijc.32176
25. Sokolenko AP, Iyevleva AG, Mitiushkina NV, Suspitsin EN, Preobrazhenskaya EV, Kuligina ES, et al. Hereditary breast-ovarian cancer syndrome in Russia. *Acta Naturae*. (2010) 2:31–5.
26. Stepanov VA. Genomes, populations and diseases: ethnic genomics and personalized medicine. *Acta Naturae*. (2010) 2:15–30.
27. Kraus C, Hoyer J, Vasileiou G, Wunderle M, Lux MP, Fasching PA, et al. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int J Cancer*. (2017) 140:95–102. doi: 10.1002/ijc.30428
28. Tung N, Lin NU, Kidd J, Allen BA, Singh N, Wenstrup RJ, et al. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. *J Clin Oncol*. (2016) 34:1460–8. doi: 10.1200/JCO.2015.65.0747

29. Bernstein-Molho R, Laitman Y, Schayek H, Iomdin S, Friedman E. The rate of the recurrent MSH6 mutations in Ashkenazi Jewish breast cancer patients. *Cancer Causes Control*. (2019) 30:97–101. doi: 10.1007/s10552-018-1106-0
30. Bush L, Aronson M, Tabori U, Campbell BB, Bedgood RB, Jasperson K. Delineating a new feature of constitutional mismatch repair deficiency (CMMRD) syndrome: breast cancer. *Fam Cancer*. (2019) 18:105–8. doi: 10.1007/s10689-018-0088-0
31. Lin PH, Kuo WH, Huang AC, Lu YS, Lin CH, Kuo SH, et al. Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget*. (2016) 7:8310–20. doi: 10.18632/oncotarget.7027
32. Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat*. (2008) 29:1412–24. doi: 10.1002/humu.20796
33. Wang X, Brzosowicz JP, Park JY. Response to Roberts et al. Cohort ascertainment and methods of analysis impact the association between cancer and genetic predisposition - the tale of breast cancer risk and Lynch syndrome genes MSH6/PMS2. *Genet Med*. (2019) 21:2156–7. doi: 10.1038/s41436-019-0471-8
34. Evans DG, Woodward ER, Lalloo F, Møller P, Sampson J, Burn J, et al. Are women with pathogenic variants in PMS2 and MSH6 really at high lifetime risk of breast cancer? *Genet Med*. (2019) 21:1878–9. doi: 10.1038/s41436-018-0401-1
35. Ten Broeke SW, Brohet RM, Tops CM, van der Klift HM, Velthuizen ME, Bernstein I, et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. *J Clin Oncol*. (2015) 33:319–25. doi: 10.1200/JCO.2014.57.8088
36. Takehara Y, Nagasaka T, Nyuya A, Haruma T, Haraga J, Mori Y, et al. Accuracy of four mononucleotide-repeat markers for the identification of DNA mismatch-repair deficiency in solid tumors. *J Transl Med*. (2018) 16:5. doi: 10.1186/s12967-017-1376-4
37. Sorscher S. The importance of distinguishing sporadic cancers from those related to cancer predisposing germline mutations. *Oncologist*. (2018) 23:1266–8. doi: 10.1634/theoncologist.2017-0681
38. Martin SA, McCarthy A, Barber LJ, Burgess DJ, Parry S, Lord CJ, et al. Methotrexate induces oxidative DNA damage and is selectively lethal to tumour cells with defects in the DNA mismatch repair gene MSH2. *EMBO Mol Med*. (2009) 1:323–37. doi: 10.1002/emmm.200900040
39. Guillotin D, Austin P, Begum R, Freitas MO, Merve A, Brend T, et al. Drug-Repositioning screens identify triamterene as a selective drug for the treatment of DNA mismatch repair deficient cells. *Clin Cancer Res*. (2017) 23:2880–90. doi: 10.1158/1078-0432.CCR-16-1216
40. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. (2017) 357:409–13. doi: 10.1126/science.aan6733
41. Mangoni M, Bisanzi S, Carozzi F, Sani C, Biti G, Livi L, et al. Association between genetic polymorphisms in the XRCC1, XRCC3, XPD, GSTM1, GSTT1, MSH2, MLH1, MSH3, and MGMT genes and radiosensitivity in breast cancer patients. *Int J Radiat Oncol Biol Phys*. (2011) 81:52–8. doi: 10.1016/j.ijrobp.2010.04.023
42. Heijink DM, de Vries EGE, Koornstra JJ, Hospers GAP, Hofstra RMW, van Vugt MATM, et al. Perspectives for tailored chemoprevention and treatment of colorectal cancer in Lynch syndrome. *Crit Rev Oncol Hematol*. (2011) 80:264–77. doi: 10.1016/j.critrevonc.2010.11.009
43. Chae YK, Anker JF, Bais P, Namburi S, Giles FJ, Chuang JH. Mutations in DNA repair genes are associated with increased neo-antigen load and activated T cell infiltration in lung adenocarcinoma. *Oncotarget*. (2018) 9:7949–60. doi: 10.18632/oncotarget.23742
44. Wieser V, Gaugg I, Fleischer M, Shivalingaiah G, Wenzel S, Sprung S, et al. BRCA1/2 and TP53 mutation status associates with PD-1 and PD-L1 expression in ovarian cancer. *Oncotarget*. (2018) 9:17501–11. doi: 10.18632/oncotarget.24770
45. Zhu X, Zhao L, Lang J. The BRCA1 methylation and PD-L1 expression in sporadic ovarian cancer. *Int J Gynecol Cancer*. (2018) 28:1514–9. doi: 10.1097/IGC.0000000000001334

Conflict of Interest: MD was employed by Druzhkov Clinic Ltd., Russia. MG was employed by National Bioservice LLC, Russia.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 June 2020
doi: 10.3389/fgene.2020.00560

Novel Genetic Variations in Acute Myeloid Leukemia in Pakistani Population

Saba Shahid[1†], Muhammad Shakeel[2*†], Saima Siddiqui[3], Shariq Ahmed[1], Misha Sohail[1], Ishtiaq Ahmad Khan[2], Aiysha Abid[4] and Tahir Shamsi[3]*

[1] Department of Genomics, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan, [2] Jamil-ur-Rahman Center for Genome Research, Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan, [3] Department of Hematology, National Institute of Blood Diseases and Bone Marrow Transplantation Karachi, Karachi, Pakistan, [4] Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan

OPEN ACCESS

Edited by:
Ira Ida Skvortsova, Medical University of Innsbruck, Austria

Reviewed by:
Muhammad Ajmal, COMSATS University Islamabad, Pakistan
Amy P. Hsu, National Institute of Allergy and Infectious Diseases (NIH), United States

***Correspondence:**
Saba Shahid
sabashahid_dbt@yahoo.com
Muhammad Shakeel
mshakeel_211@yahoo.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Genetics

Received: *25 October 2019*
Accepted: *07 May 2020*
Published: *23 June 2020*

Citation:
Shahid S, Shakeel M, Siddiqui S, Ahmed S, Sohail M, Khan IA, Abid A and Shamsi T (2020) Novel Genetic Variations in Acute Myeloid Leukemia in Pakistani Population. Front. Genet. 11:560. doi: 10.3389/fgene.2020.00560

Acute myeloid leukemia (AML) is a hematological malignancy characterized by clonal expansion of blast cells that exhibit great genetic heterogeneity. In this study, we describe the mutational landscape and its clinico-pathological significance in 26 myeloid neoplasm patients from a South Asian population (Pakistan) by using ultra-deep targeted next-generation DNA sequencing of 54 genes (~5000×) and its subsequent bioinformatics analysis. The data analysis indicated novel non-silent somatic mutational events previously not reported in AML, including nine non-synonymous and one stop-gain mutations. Notably, two recurrent somatic non-synonymous mutations, i.e., *STAG2* (causing p.L526F) and *BCORL1* (p.A400V), were observed in three unrelated cases each. The *BCOR* was found to have three independent non-synonymous somatic mutations in three cases. Further, the *SRSF2* with a protein truncating somatic mutation (p.Q88X) was observed for the first time in AML in this study. The prioritization of germline mutations with ClinVar, SIFT, Polyphen2, and Combined Annotation Dependent Depletion (CADD) highlighted 18 predicted deleterious/pathogenic mutations, including two recurrent deleterious mutations, i.e., a novel heterozygous non-synonymous SNV in *GATA2* (p.T358P) and a frameshift insertion in *NPM1* (p.L258fs), found in two unrelated cases each. The *WT1* was observed with three independent potential detrimental germline mutations in three different cases. Collectively, non-silent somatic and/or germline mutations were observed in 23 (88.46%) of the cases (0.92 mutation per case). Furthermore, the pharmGKB database exploration showed a missense SNV rs1042522 in *TP53*, exhibiting decreased response to anti-cancer drugs, in 19 (73%) of the cases. This genomic profiling of AML provides deep insight into the disease pathophysiology. Identification of pharmacogenomics markers will help to adopt personalized approach for the management of AML patients in Pakistan.

Keywords: genomic screening, AML, next generation sequencing, myeloid sequencing panel, novel no-silent somatic mutation

INTRODUCTION

Acute myeloid leukemia (AML) is the most frequent form of acute leukemia in adults with a poor survival rate of about 5 years only (Horton and Huntly, 2012; Cancer Genome Atlas Research Network et al., 2013). It is caused by pathogenic variations in normal progenitor myeloid hematopoietic cells, leading to altered differentiation, proliferation, and self-renewal capability of the cells (Papaemmanuil et al., 2016). In the last decade, there has been significant increase in understanding of underlying mutational landscape of AML (Arber et al., 2016; Papaemmanuil et al., 2016). Consequently, the prognosis, diagnosis, and treatment have been transformed from histological findings to cytogenetic and genomic testing (Grimwade et al., 2010). Analyzing the genetic alterations in AML can be helpful to reduce ambiguities in further characterization of the molecular heterogeneity of normal karyotype AML (Renneville et al., 2008).

Recent studies on the molecular pathogenesis have identified prognostic significance of genetic variation and their contribution in the pathogenesis of AML (Papaemmanuil et al., 2016). The improved AML prognosis associated with mutated *NPM1* and biallelic mutations in the *CEBPA* have resulted in a change in the disease definition (May Green et al., 2010; Hollink et al., 2011). These recent advances have changed the classification and introduced molecular subtypes of the AML with gene mutations (*NPM1* and *CEPBA*) by the recommendation of WHO classification of hematopoietic tumors in 2008 (Campo et al., 2008). In the revised version of 2016, WHO classification introduced additional germline predisposition associated with genetic alterations in the genes *CEBPA*, *DDX41*, *RUNX1*, *ANKRD26*, *ETV6*, and *GATA2* (Cazzola, 2016; Swerdlow et al., 2016). Further studies on genetic landscape of AML have expanded the mutational spectrum where *TET2*, *DNMT3A*, *NPM1*, *SRSF2*, and *ASXL1* genes are mutated frequently in elderly people (Prassek et al., 2018).

Mutational profiling plays an important role in the diagnosis of AML and is now routinely available as a part of the diagnostic workup. It provides diagnostic accuracy, which increases the precision in risk stratification and helps in adopting therapeutic options (Papaemmanuil et al., 2013; Kuo and Dong, 2015). The development of *FLT3* and *IDH2* inhibitors (Lee et al., 2017; Stein et al., 2017) is achieved only by extensive genomic studies. With the advent of next-generation DNA sequencing (NGS), the cost of genome sequencing has decreased significantly. Amplicon-based targeted sequencing represents an attractive mutation detection method in selected gene panels (Harismendy et al., 2011; Jünemann et al., 2013). This strategy needs less amount of DNA and provides large data of multiple genes in a short turnaround time. Therefore, the genomic tractability of AML makes it a feasible option for targeted NGS testing clinically. The aim of this study was to assess the frequency and clinico-pathological significance of frequently mutated genes by targeting sequencing in AML cases. The targeted sequencing panel comprises of genes involved in various biological functions such as epigenetic regulator genes, the cohesin complex protein encoding genes, genes of activated signaling, tumor repressor genes, and spliceosome genes. This is the first study on molecular characterization of AML patients from South Asia using myeloid sequencing panel, which will be helpful in early diagnosis as well as risk management.

MATERIALS AND METHODS

Ethical and Consent Statement

For this study, 26 AML patients were recruited and sequenced for TruSight myeloid sequencing panel between December 2015 and 2018. These patients included 15 males and 11 females with a median age of 35 years (range: 7–51 years). The clinical presentation of the cases, chromosomal abnormalities, and percentage of circulating blast cells are given in **Supplementary Table S1**. The study design was approved by the Research Ethics Committee and Review Board of NIBD, and in accordance with the tenets of the Declaration of Helsinki. A written informed consent was obtained from patients and their legal guardians for participation in this study and publication of the findings. Peripheral venous blood specimens of all the recruited patients were collected in EDTA tubes, and stored at 4°C till DNA isolation and subsequent analysis.

DNA Extraction

Genomic DNA was isolated from peripheral blood by using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The quality of the extracted DNA was assessed by 2% agarose gel and quantified by Qubit DNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, United States).

Myeloid Sequencing Panel

TruSight myeloid sequencing panel (Illumina, San Diego, CA, United States) is designed to sequence targeted regions of 54 genes frequently reported for somatic mutations (complete coding exons of 15 genes and exonic hotspots of 39 genes). The genes whose complete coding exons were sequenced include *BCOR, BCORL1, CDKN2A, CEBPA, CUX1, DNMT3A, ETV6/TEL, EZH2, KDM6A, IKZF1, PHF6, RAD21, RUNX1/AML1, STAG2,* and *ZRSR2,* and exonic hotspots of 39 genes include *ABL1, ASXL1, ATRX, BRAF, CALR, CBL, CBLB, CBLC, CSF3R, FBXW7, FLT3, GATA1, GATA2, GNAS, HRAS, IDH1, IDH2, JAK2, JAK3, KIT, KRAS, KMT2A/MLL, MPL, MYD88, NOTCH1, NPM1, NRAS, PDGFRA, PTEN, PTPN11, SETBP1, SF3B1, SMC1A, SMC3, SRSF2, TET2, TP53, U2AF1,* and *WT1*. The panel consists of 568 amplicons (length range: 225–275 bp) and covers ∼141 kb of genomic region of ∼250-bp fragment lengths.

DNA Libraries Preparation

The sequencing libraries were prepared from 50 ng of genomic DNA per sample using TruSight myeloid sequencing panel according to the manufacturer's protocol. Briefly, the libraries were prepared by annealing uniquely targeted specific oligos at upstream and downstream to the region of interest (ROI), followed by the removal of unbound oligos in subsequent washing steps by using a filter plate. In extension and

ligation step, DNA polymerase was used to connect the hybridized upstream and downstream oligos resulting in the formation of products containing the targeted regions of interest flanked by sequences required for amplification. Next, the amplification step added indexes adapters and prepared for cluster generation. Then, the libraries were cleaned up by using AMPure XP beads to purify PCR products. After the purification, libraries were quantified by Qubit DNA HS Assay kit (Life Technologies, United States).

Libraries were normalized to attain equal library representation that pooled in batches of four samples as per the given guideline. A Pooled Amplicon Library (PAL) was prepared by mixing 5 μl of each of uniquely indexed library. Then, libraries were diluted by taking 6 μl of PAL and 594 μl of ice-cold HT1 incorporation buffer and heat-denatured at 92°C for 2 min. The diluted amplicon libraries were placed on the ice water bath for 5 min, and then 600 μl of the final sample was loaded into the sequencing reagent cartridge kit V2 (MS-102-2002). Workflow for DNA library preparation using Illumina TruSight myeloid sequencing panel is given in **Supplementary Figure S1**. The DNA sequencing was performed on a MiSeq instrument with standard V2 flow cells with paired end sequencing (150 bp × 2), as per the manufacturer's instruction.

Sanger Sequencing

Sanger Sequencing was performed to confirm the variants that were identified as pathogenic through standard protocol (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems®). The status of known mutations in *NPM1* and *FLT3* genes were checked by Sanger sequencing and later by allele-specific polymerase chain reaction (PCR) and PCR-restriction fragment length polymorphism (PCR-RFLP) analysis. Electropherogram of identified mutation in AML cases are given in **Supplementary Figure S2**.

Data Analysis

For data analysis, variants calling was performed using the standard pipeline, as described elsewhere (Lek et al., 2016). The alignment of short DNA sequences with human reference genome hg19 (UCSC) was performed by using Burrows–Wheeler Aligner (BWA-MEM) algorithm (Li and Durbin, 2009). The sequence alignment files (SAM) were converted into binary format (BAM) files using SAMtools (Li et al., 2009); and the removal of duplicates (PCR artifacts) was performed using the PICARD tool[1]. The base quality score recalibration (BQSR), realignment around small insertions and deletions, and variants calling were carried out by using on-instrument pipeline with Genome Analysis Tool Kit (GATK) best practices (DePristo et al., 2011). The variants with QUAL < 50, GQ < 20, and population variant allele frequency ≥1% in either gnomAD_genome or 1000 Genomes Project were filtered out as recommended previously (Tyner et al., 2018). Given that no matching normal tissue samples were sequenced, a bit stringent criterion was applied for somatic variants; i.e., the variants with allelic fraction (VAF) less than half of the percent circulating blast cells in each patient (VAF < 1/2 × %circulating blasts) were considered as somatic. To find possibly pathogenic and/or deleterious somatic associated with AML, a multi-tool prioritization approach was adopted, as recommended by American College of Medical Genetics and Genomics (Richards et al., 2015; **Figure 1**). The identified variants were annotated with ANNOVAR (Yang and Wang, 2015) and Variants Effect Predictor (McLaren et al., 2016) tools to determine their functional consequences. The deleterious impact of non-synonymous variants was assessed with SIFT, Polyphen2, and Combined Annotation Dependent Depletion (CADD), as described previously (Shakeel et al., 2018).

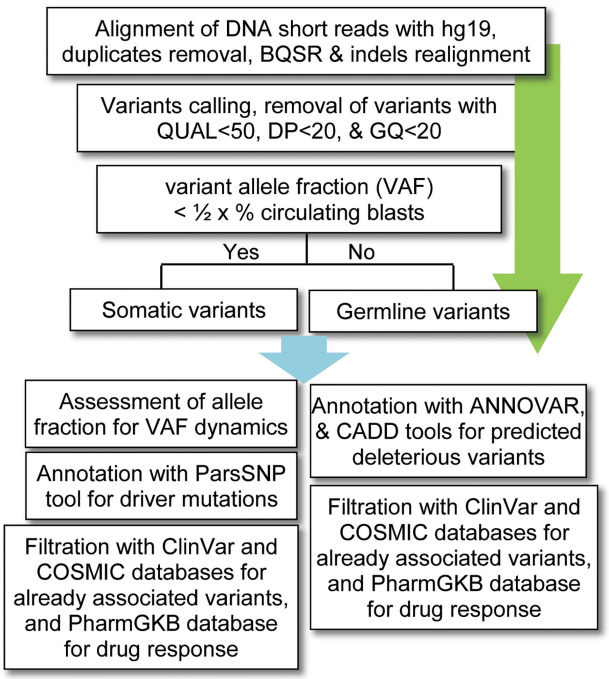


FIGURE 1 | Workflow for bioinformatics analysis of the DNA sequencing data. BQSR, base quality score recalibration, QUAL, variant quality score, DP, depth of coverage, GQ, genotype quality, VAF, variants allele fraction, CADD, combined annotation dependent depletion.

To prioritize biologically active driver mutations over inactive passengers, the parsimony-guided unsupervised functional impact predictor ParsSNP tool was used. This tool uses an expectation maximization framework to find mutations that explain tumor incidence broadly, without using predefined training labels that can introduce biases (Kumar et al., 2016). The identified variants were also searched in ClinVar database (Landrum et al., 2014) for pathogenic/likely pathogenic association with myeloid malignancies. The interaction between the proteins with deleterious variants in the same samples was determined using STRING database (von Mering et al., 2005). The curation from pharmGKB database (Hewett et al., 2002) was performed to determine variants that likely have a role in leukemic chemotherapy.

[1]http://picard.sourceforge.net

TABLE 1 | Total variants pertaining to various genomic regions and their functional distribution.

Genomic region	No. of variants
Exonic	146
Intronic	133
Upstream	0
Downstream	3
UTR5	2
UTR3	9
Functional Impact	
Non-synonymous	71
Synonymous	61
Stop-gain	3
Splicing	2
Frameshift insertion	3
Frameshift deletion	2
Non-frameshift insertion	2
Non-frameshift deletion	3

RESULTS

This study involves determination and assessment of genetic variations in 26 AML cases of a South Asian population (Pakistan) through Illumina TruSight myeloid sequencing panel. This panel was designed to identify somatic mutations in myeloid malignancies. The median depth of coverage for coding variants was 4979×, and average coverage was 15,477×. Likewise, for non-coding regions, the median depth of coverage was 9558× and average coverage was 24,348×. After filtering out the variants with QUAL < 50, DP < 20, and GQ < 20, there were 293 variants in 54 genes, where each patient contained on average 80 variants (SD $\pm$ 8.5). The variants allele fraction distribution revealed the median of 0.51 across all 26 samples (**Supplementary Figure S3**).

The ANNOVAR annotation was performed to evaluate the genetic variants corresponding to different genomic locations and their functional impact, as detailed in **Table 1**. The number of non-synonymous sites was observed to be higher than that of synonymous sites, with a nonsyn/syn ratio of 1.16. This ratio is higher than the reported overall nonsyn/syn ratio for germline variants in South Asian populations (1000 Genomes Project Consortium et al., 2015). For normalization, variants within the targeted genomic regions studied in this research were a subset from 1000 Genomes PJL (Punjabi Lahori, Pakistan) individuals, and nonsyn/syn ratio was determined. The PJL healthy individuals showed a ratio of 0.88. The nonsyn/syn ratio in targeted regions was higher in the present study AML cases than in healthy individuals due to the higher proportion of novel non-synonymous variants in the patients, which is persistent with previous reports (Liu et al., 2012).

The Landscape of Somatic Mutations

Considering the variants with allelic fraction less than half of the %circulating blasts in each case, there were 38 somatic mutations as a whole, including 31 SNVs and 7 insertions/deletions (1.46 mutation/case). These somatic variations comprised 23 non-silent mutations including 18 non-synonymous SNVs, 2 stop-gain SNVs, 1 splicing SNV, and 2 frameshift deletions, and 17 silent mutations including 2 synonymous SNVs, 2 downstream SNVs, 1 3′untranslated region SNV, 7 intronic SNVs, 2 non-frameshift insertions, and 1 non-frameshift deletion (**Figure 2** and **Supplementary Table S2**). Further, it was observed that some cases contained higher number of somatic mutations in different genes. A Kruskal-Wallis test and post hoc Dunn test of multiple comparisons among all the cases showed a significantly higher number of somatic mutations in two cases, AM01 and AM03 ($p < 0.01$ after multiple corrections).

Strikingly, three recurrent non-silent, and five recurrent silent somatic mutations were observed in more than one case. The non-silent recurrent somatic mutations included a non-synonymous SNV in *CDKN2A* (p.R90C) in four cases, a non-synonymous SNV in *STAG2* (p.L526F) in three cases, and a non-synonymous SNV in *BCORL1* (p.A400V) in three cases. Notably, p.L526F(*STAG2*) and p.A400V(*BCORL1*) affected all the transcripts of respective genes, whereas p.R90C(*CDKN2A*) affected only one out of six transcripts. The VAF of p.R90C (*CDKN2A*) was double in one case AM21 (VAF 0.214) than in the other three cases who carried similar burden of this variant (VAF 0.084–0.105). The mutational burden of two other recurrent mutations, i.e., p.L526F(*STAG2*) and p.A400V(*BCORL1*), was similar among the cases, i.e., 0.102–0.111 and 0.08–0.106, respectively. Further, it was noted that five genes, *CSF3R*, *NOTCH1*, *CBL*, *RUNX1*, and *BCOR*, were found to have independent non-silent mutational events in two cases each, with VAF 0.098–0.108, 0.158–0.278, 0.1448–0.203, 0.159–0.25, and 0.135–0.167, respectively. Two genes, *KIT* and *EZH2*, had two coexisting mutations each (VAF 0.418–0.42 and 0.043–0.045, respectively) in the same cases (AM03 and AM01, respectively), affecting all the transcripts of their genes.

Curation of somatic mutations in COSMIC database highlighted 10 mutational events not observed in the database, whereas four mutations had been cataloged with a different variation type at the sites than observed in this study (detailed in **Supplementary Table S2**). The non-cataloged 10 mutations also included the stop-gain SNV (p.Q88X) in *SRSF2*, affecting all two transcripts of this gene, and the two recurrent non-synonymous SNVs (p.L526F of *STAG2* and p.A400V of *BCORL1*). Filtration of somatic mutations with ClinVar database highlighted three pathogenic SNVs already associated with hematological disorders. We also assessed conservation status of the non-silent variants sites using PhyloP conservation scores of non-neutral substitution rates based on alignment with 100 vertebrates (Pollard et al., 2010). This analysis revealed 13 mutations in comparatively high conserved regions (PhyloP score >4), 5 mutations in moderately conserved regions ($4 \leq$ PhyloP score ≥ 1), and 5 mutations in non-conserved regions (PhyloP < 1) of proteins. Among the recurrent non-silent mutations, p.L526F was observed in the highly conserved region of *STAG2*, suggesting its more profound deleterious effect; p.A400V occurred in moderately conserved region of *BCORL1*, whereas p.R90C occurred in the non-conserved region of *CDKN2A*. The assessment of somatic mutations for potentially driver role through ParsSNP tool highlighted a

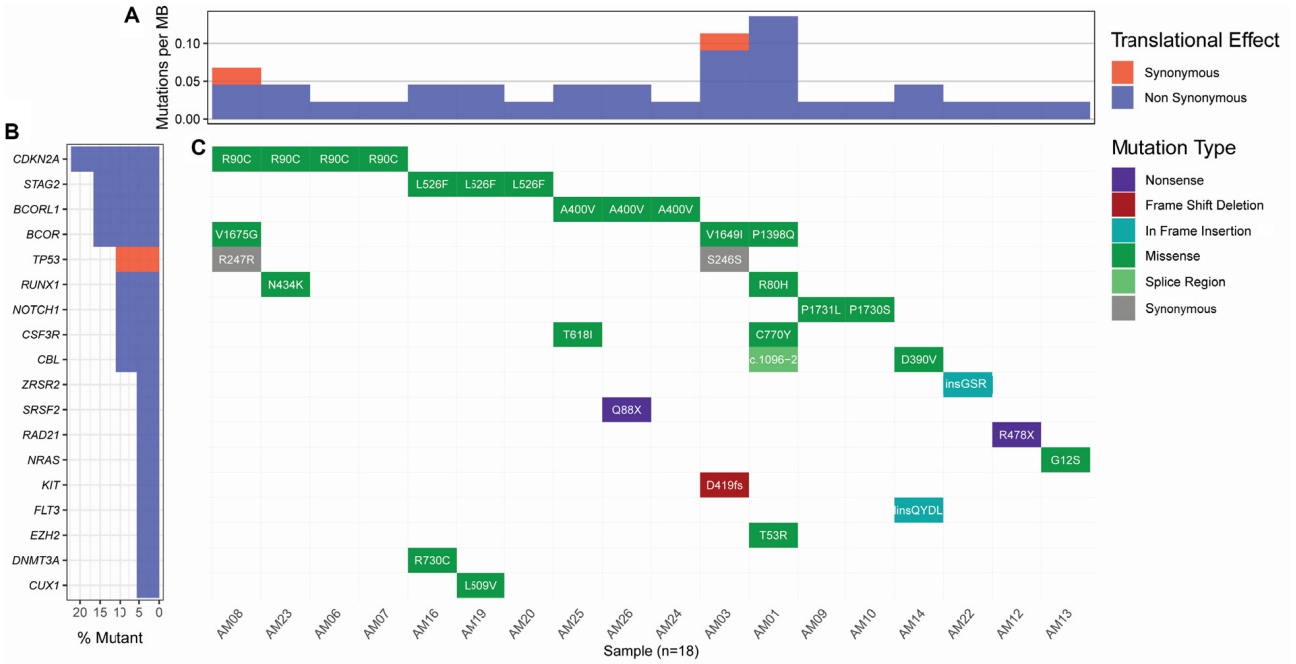


FIGURE 2 | Landscape of somatic mutations in AML. All the non-silent and synonymous somatic mutations are shown here. **(A)** The numbers of non-synonymous mutations are dominantly higher than synonymous mutations at the sequenced targeted regions. **(B)** Percent of samples with the somatic mutation in a gene. **(C)** Detail of somatic mutations in AML cases.

driver mutation p.G12S in *NRAS* (rs121913250), which is well characterized and already reported recurrently in COSMIC database (COSV54736621 and COSM563).

Co-existence of Somatic Mutations

It was also noted that seven cases carried more than one non-silent variants; however, the genes harboring co-existing mutations were different among all the cases. The difference in VAF of co-existing mutations gave a clue to define the clonal composition, i.e., a founding clone (the clone with the highest VAF values) and the subclone (**Figure 3**). In AM01, the novel p.R80H mutation in the conserved region of Runt-related transcription factor 1 (*RUNX1*), might be the somatic event (VAF 0.25) followed by disrupting splice site (c.1096-2) in Cbl proto-oncogene (*CBL*) (VAF 0.203) in founding clones leading to the abnormal proliferation of hematopoietic stem cells. Analysis of protein interaction between RUNX1 and CBL through STRING database revealed no interaction between these two proteins, indicating independent mutational events. Previously, a different mutation, p.R80A, at same position of *RUNX1*, has been shown to strongly reduce its binding with DNA (Bravo et al., 2001). In AM03, the co-existing p.D419fs and p.R420fs deletions in *KIT* (VAF 0.418 and 0.42) originated more probably in founding clone, prior to the p.V1649I of *BCOR* (0.135) in subclone. In AM19, the novel p.L509V mutation in *CUX1* (VAF 0.366) might be originated in founding clone, followed by p.L526F of *STAG2* (0.126) in subclone; in AM26, the p.Q88X in *SRSF2* (VAF 0.214) in founding clone and p.A400V in *BCORL1* (VAF 0.08) in subclone. The co-existing somatic mutations in three other cases, AM16 [p.R730C of *DNMT3A* (VAF 0.112) and p.L526F of *STAG2* (VAF 0.102)], AM23 [p.R90C of *CDKN2A* (VAF 0.214) and p.N434K of *RUNX1* (VAF 0.159)], and AM25 [p.T618I of *CSF3R* (VAF 0.108) and p.A400V of *BCORL1* (VAF 0.097)], more probably originated in the same clones. It was noteworthy that all the coexisting mutational events happened in genes belonging to different biological functional categories previously described in myeloid leukemias (Cancer Genome Atlas Research Network et al., 2013), indicating that different underlying processes were involved in the pathophysiology of AML in this cohort.

Germline Mutation Predisposition

In addition to the somatic mutations, the predisposition due to germline mutations was also assessed. For this, the germline variants with ClinVar pathogenic/likely pathogenic significance, producing a stop-gain or stop-loss site, disrupting splicing sites, frameshift insertions, or deletions, and non-synonymous alterations predicted as deleterious by SIFT and Polyphen2 tools, were brought into subsequent analysis, as described previously (Bertelsen et al., 2019). This analysis prioritized 18 germline variants pertaining to 15 genes including 13 non-synonymous SNVs, a stop-gain SNV, a splice-site SNV, and 3 frameshift insertions (**Supplementary Table S3**). Two recurrent mutations, p.T358P in *GATA2* affecting all three transcripts, and p.L258fs insertion in *NPM1* affecting two out of seven transcripts, were observed in two non-related cases each. Further, *WT1* was observed recruiting three independent germline mutations in three different unrelated cases. Filtration with ClinVar database highlighted six non-synonymous pathogenic SNVs

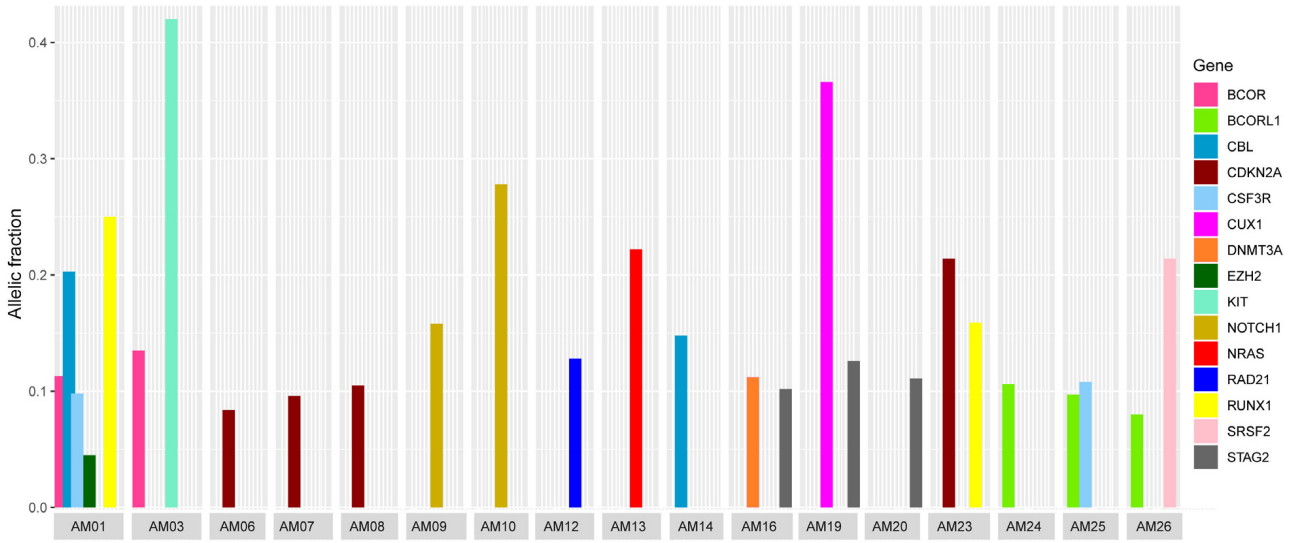


FIGURE 3 | Comparison of variants allelic fraction (VAF) of co-existing non-silent somatic mutations in various cases. The mutations in *RUNX1* and *CBL* in AM01, the *KIT* mutation in AM03, the *CUX1* mutation in AM19, and *SRSF2* mutation in AM26 originated more likely in founding clones, followed by mutations in other genes in subclones.

and a frameshift pathogenic insertion already associated with hematological neoplasms. The germline variants were filtered with COSMIC database, which revealed six novel variants not cataloged in this database. The assessment of PhyloP scores revealed 11 mutations in comparatively high conserved regions, 2 mutations in moderately conserved regions, and 5 mutations in non-conserved regions of respective proteins. The variants affecting highly conserved regions included the recurrent p.T358P in *GATA2*, protein truncating p.R441X in *WT1*, and splicing c.418+1 in *PHF6*. Exploration of protein truncating p.R441X (rs121907909) in Ensembl genome browser[2] revealed that it affects eight protein coding transcripts introducing a premature stop codon, whereas two protein coding transcripts are protected through NMD pathway (**Supplementary Table S4**).

To explore the role of identified genetic variants in drug response, pharmGKB database was searched. This analysis showed a missense SNV rs1042522 (G > C) in *TP53*, with GG and GC genotypes in 19 (73%) cases. These genotypes have been found to show decreased response to cisplatin, paclitaxel, capecitabine, and oxaliplatin anti-cancer drugs as compared to the CC genotype.

DISCUSSION

Next-generation sequencing analysis of myeloid neoplasm including AML and other related disorders has yielded several significant advances in the identification of diagnostic, prognostic, and therapeutic markers for these disorders (Arber et al., 2016; Papaemmanuil et al., 2016). This study was designed to screen AML patients in a clinical diagnostic setup by using a specifically designed myeloid sequencing panel and provides clinico-pathological significance of identified deleterious/non-silent mutations in the Pakistani population. We identified 293 variants including single-nucleotide variants, and small insertions and deletions in coding as well as in non-coding regions in a small cohort of 26 AML patients. Sequence variants not observed in ClinVar, dbSNP, and gnomAD were considered as novel variants. The pathogenicity of sequence variants with a global minor allele frequency (GMAF) of <0.1 was assessed by using several *in silico* bioinformatics tools and the variants were classified according to the ACMG criteria (Richards et al., 2015). The deleterious impact of non-synonymous variants was assessed with SIFT, Polyphen2, and CADD, as described previously (Shakeel et al., 2018). Although variants were not functionally validated using any in vitro system, *in silico* analyses have generated strong and convincing scores that suggest the possible pathogenicity of the identified variants in respective cases. To the best of our knowledge, this is the first study to report genetic variations in myeloid malignancies from this South Asian population using NGS technology.

The higher nonsyn/syn ratio in AML cohort represents higher mutation rate and/or positive selection on non-synonymous sites, as indicated in various cancers previously (Greenman et al., 2007; Pengyuan et al., 2012). By applying the multi-tool prioritization approach, we were able to find at least one pathogenic/deleterious non-silent somatic or predisposing germline mutation in 23 of the 26 cases (88.46%), where 9 cases had both the somatic and germline mutations, 8 cases had somatic mutation only, and 6 cases had germline mutation only. In order to explore possible biological relationship between a predisposing germline mutation and somatic mutational events in the cases having coexisting germline and somatic mutations, a circos plot was constructed, which revealed that the two cases with germline non-synonymous mutation p.T358P in *GATA2* (AM01 and

[2]www.ensembl.org/index.html

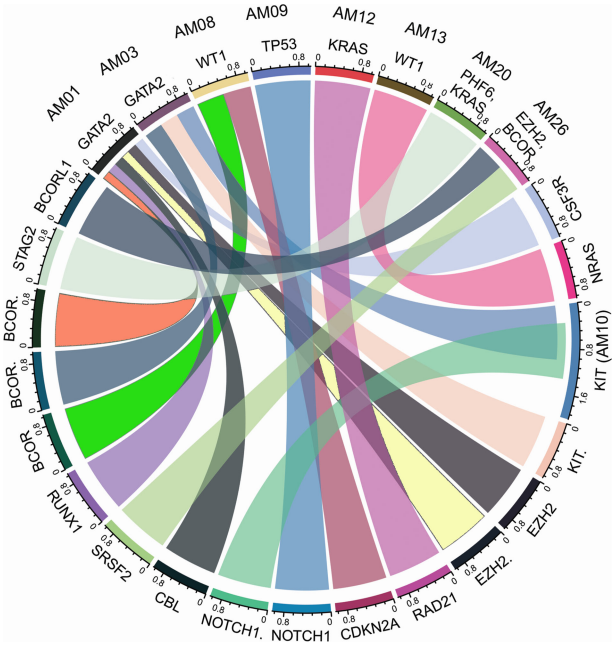


FIGURE 4 | Co-existence of prioritized/pathogenic germline variants with the non-silent somatic mutations in cases with both germline and somatic events. The names of cases are mentioned above the genes with germline mutations.

AM03) had higher number of non-silent somatic mutations (**Figure 4**). *GATA2* encodes an endothelial transcription factor GATA-2 that plays an essential role in gene regulation during vascular development and hematopoietic differentiation. The observed mutation is within the highly conserved region of GATA2 (a zinc finger domain). Although this mutation is novel and not cataloged in dbSNP or COSMIC databases, a p.R361P change near the observed p.T358P mutation in the same zinc finger domain has been shown to be associated with Emberger syndrome (lymphedema with predisposition to AML) (Ostergaard et al., 2011). Further, search in STRING database showed experimental and curated pathway interaction between GATA2 and RUNX1, the two mutated genes in AM01 (**Supplementary Figure S4**). It has been shown previously through Chip-seq analysis that there is concurrent binding of GATA2 and RUNX1 along with GATA1, FLI1, and SCL transcription factors on promoters of a set of genes, e.g., *CEBPA*, which are highly enriched for known regulators of hematopoiesis (Timchenko et al., 1995; Tijssen et al., 2011).

The genetic heterogeneity and the complex interaction among different oncogenic pathways in AML have been focused in previous studies to explore its prognostic significance (Estey, 2014). The most recurrent non-silent somatic mutation p.R90C in cyclin-dependent kinase inhibitor 2A gene (*CDKN2A*) was found in four cases. The observed variant belongs to non-conserved region of *CDKN2A*, yet it has been cataloged in ClinVar associated with hereditary cancer-predisposing syndrome with uncertain significance. The kinase inhibitor arrests cell cycle at G1 and G2 stages and acts as a tumor suppressor (Genecards database, 2019). The arginine-to-cysteine substitution as a result of this variation may hamper its ability to arrest cell cycle, leading to accelerated cell proliferation. Although this variant is ultra-rare (alternate allele frequency of 4.71×10^{-6} in the gnomAD database and 8.42×10^{-6} in the ExAC database), its recurrence in 15% of our cases suggests its likely prognostic role in AML in this population. The second recurrent non-silent somatic mutation p.L526F (*STAG2*) occurs in a conserved domain of cohesin subunit SA-2, which is a component of the cohesin complex required for the cohesion of sister chromatids after DNA replication. Previously, non-silent mutations at different sites in *STAG2* were found in 1.3% of AML cases (Thol et al., 2014), whereas, in this study, the observed mutation was found in 11.5% cases. The third novel recurrent mutation p.A400V was observed in *BCORL1*, affecting three cases. *BCORL1* encodes BCL6 corepressor like 1 protein, which specifically inhibits gene expression when recruited to promoter regions by sequence-specific DNA-binding proteins such as BCL6 (Pagan et al., 2007). The concurrence of non-silent mutations in genes belonging to different functional categories represents the heterogeneity of AML in this cohort, which is persistent with previous reports (Cancer Genome Atlas Research Network et al., 2013).

The novel non-synonymous somatic mutations p.V1675G, p.V1649I, and p.P1398Q in conserved regions of the BCL6 co-repressor (*BCOR*) were found independently in 11.5% cases (three of this cohort), which is three times higher compared to those reported by Grossmann et al., where *BCOR* gene mutations were identified in 3.8% (10 of 262) of cytogenetically normal (CN) AML cases with poor response (Grossmann et al., 2011). It is noteworthy that *BCOR* also contained a germline deleterious non-synonymous mutation p.S1582G. Strikingly, this mutation was also not found in COSMIC database. Together with the germline mutation, the frequency of *BCOR* non-silent/deleterious mutations becomes 15.4% (four cases). This depicts *BCOR* as a high-risk gene in the South Asian population. The *BCOR* encoded protein, BCL6 co-repressor, is a component of a variant Polycomb group repressive complex 1, and has the ability to specifically repress gene transcription when recruited to promoter regions by sequence-specific DNA-binding proteins such as BCL6 and MLLT3 (Huynh et al., 2000; Sanchez et al., 2007). It contributes as a major player in the embryonic differentiation and mesenchymal stem cell function (Wamstad et al., 2008; Fan et al., 2009). Recently, BCOR mutant bone marrow cells showed significantly higher proliferation and differentiation rates with upregulated expression of HOX genes (Cao et al., 2016).

The novel protein truncating somatic mutation in *SRSF2* generates alteration of c.C262T in exon1 of the resulting transcript, leading to the premature termination at p.Q88X, and causing inactivation of the RNA-binding domain (residues 1–101) of the protein. This mutation affects five protein-coding transcripts, whereas two transcripts undergo non-sense-mediated decay. The *SRSF2* is a member of the serine/arginine

rich (SR) class of splicing factors involved in both constitutive and alternative mRNA splicing. Previously, dysfunctional SRSF due to sequence variations at p.P95H position have been found to activate aberrant alternative splicing in hematopoietic cells, whereby, having its role in onset of myelodysplastic syndromes (MDS) and AML (Liang et al., 2018; Masaki et al., 2019). In this context, the truncating mutation observed in this study would more likely result in non-functional SRSF, which would lead to malignancy due to hampered alternative splicing in hematopoietic cells. This study reports a first truncating mutation in the SRSF2 RNA-binding domain in an AML case. The other stop-gain somatic mutation in *RAD21*, causing a protein truncation at p.R478X, affects two protein coding transcripts. This variation has been cataloged in the COSMIC database with four recurrences (COSM1735718) associated with AML. The third and germline stop-gain mutation in *WT1* causes p.R441X truncation. Previously, the protein truncating variations in *WT1* have been shown to attenuate the TP53-induced DNA damage response in T-cell acute lymphoblastic leukemia (Bordin et al., 2018).

CONCLUSION

In conclusion, this is the first report of a comprehensive analysis of somatic as well as germline mutations in AML from Pakistan using next-generation DNA sequencing technology. Our data strongly support and extend the spectrum of detrimental mutations identified in previous studies employing targeted resequencing approach for the diagnosis of AML. This study also highlights the usefulness of panel sequencing in cases where prognosis becomes challenging. The small cohort size and retrospective nature, i.e., sample collection from a single medical center, are the limiting factors of the study. Furthermore, the novel findings of this preliminary study require validation in a larger cohort with different time scales. Nevertheless, the findings provide an assessment of predisposing detrimental mutations of AML in this region and its utility in clinical settings.

DATA AVAILABILITY STATEMENT

The raw datasets presented in this article have been deposited in BioProject – accession PRJNA627793.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee and Review Board of NIBD. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SSh: study design and execution, and manuscript writing. MSh: data analysis, result interpretation, and manuscript writing. SSi: clinical examination and evaluation of AML patients. SA: library preparation for myeloid sequencing panel. MSo: DNA extraction and quantification. AA: review of manuscript. IK and TS: involving in study design, patient recruitment, review of manuscript, and supervision throughout the study.

ACKNOWLEDGMENTS

We appreciate the patients who participated in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00560/full#supplementary-material

FIGURE S1 | Workflow for DNA library preparation using Illumina TruSight myeloid sequencing panel.

FIGURE S2 | Electropherograms of Sanger sequencing of identified mutations in AML cases.

FIGURE S3 | Density plot to show the distribution/spectrum of variants allelic fraction (VAF) of all variants across the 26 AML patients of this study. The perpendicular at 0.51 allele fraction represents the median percentage of circulating blasts cells across all cases.

FIGURE S4 | Protein-protein interaction between GATA2 and the proteins with somatic mutations in AM01.

TABLE S1 | Clinical features of 26 AML cases.

TABLE S2 | SOMATIC Mutations.

TABLE S3 | Germline Mutations in AML.

TABLE S4 | Effect of stopgain and frameshift variants (germline and somatic) on transcripts.

REFERENCES

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., and Le Beau, M. M. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 20, 2391–2405. doi: 10.1182/blood-2016-03-643544

Bertelsen, B., Tuxen, I. V., Yde, C. W., Gabrielaite, M., Torp, M. H., Kinalis, S., et al. (2019). High frequency of pathogenic germline variants within homologous recombination repair in patients with advanced cancer. *Genom. Med.* 4:13. doi: 10.1038/s41525-019-0087-6

Bordin, F., Piovan, E., Masiero, E., Ambesi-Impiombato, A., Minuzzo, S., Bertorelle, R., et al. (2018). WT1 loss attenuates the TP53-induced DNA damage response in T-cell acute lymphoblastic leukemia. *Haematologica* 103, 266–277. doi: 10.3324/haematol.2017.170431

Bravo, J., Li, Z., Speck, N. A., and Warren, A. J. (2001). The leukemia-associated AML1 (Runx1)–CBF beta complex functions as a DNA-induced molecular clamp. *Nat. Struct. Biol.* 8, 371–378. doi: 10.1038/86264

Campo, E., Swerdlow, S. H., Harris, N. L., Pileri, S., Stein, H., and Jaffe, E. S. (2008). The 2008 WHO classification of lymphoid neoplasms and beyond: evolving

concepts and practical applications. *Blood* 12, 5019–5032. doi: 10.1182/blood-2011-01-293050

Cancer Genome Atlas Research Network, Ley, T. J., Miller, C., Ding, L., Raphael, B. J., Mungall, A. J., et al. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 368, 2059–2074. doi: 10.1056/NEJMoa1301689

Cao, Q., Gearhart, M., Gery, S., Shojaee, S., Yang, H., Sun, H., et al. (2016). BCOR regulates myeloid cell proliferation and differentiation. *Leukemia* 30, 1155–1165. doi: 10.1038/leu.2016.2

Cazzola, M. (2016). Introduction to a review series: the 2016 revision of the WHO classification of tumors of hematopoietic and lymphoid tissues. *Blood* 127, 2361–2364. doi: 10.1182/blood-2016-03-657379

DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., and Hartl, C. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.

Estey, E. H. (2014). Acute myeloid leukemia: update on risk stratification and management. *Am. J. Hematol.* 89, 1063–1081. doi: 10.1002/ajh.23834

Fan, Z., Yamaza, T., Lee, J. S., Yu, J., Wang, S., Fan, G., et al. (2009). BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. *Nat. Cell Biol.* 11, 1002–1009. doi: 10.1038/ncb1913

Genecards database (2019). *CDKN2A Gene (Protein Coding): Cyclin Dependent Kinase Inhibitor 2A*. Available online at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CDKN2A(accessed March 15, 2019).

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.

Grimwade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., et al. (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* 116, 354–365. doi: 10.1182/blood-2009-11-254441

Grossmann, V., Tiacci, E., Holmes, A. B., Kohlmann, A., Martelli, M. P., Kern, W., et al. (2011). Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* 118, 6153–6163. doi: 10.1182/blood-2011-07-365320

Harismendy, O., Schwab, R. B., Bao, L., Olson, J., Rozenzhak, S., Kotsopoulos, S. K., et al. (2011). Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome Biol.* 12:R124.

Hewett, M., Oliver, D. E., Rubin, D. L., Easton, K. L., Stuart, J. M., Altman, R. B., et al. (2002). PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res.* 30, 163–165. doi: 10.1093/nar/30.1.163

Hollink, I. H., Van den Heuvel-Eibrink, M. M., Arentsen-Peters, S. T., Zimmermann, M., Peeters, J. K., and Valk, P. J. (2011). Characterization of CEBPA mutations and promoter hypermethylation in pediatric acute myeloid leukemia. *Haematologica* 96, 384–392. doi: 10.3324/haematol.2010.031336

Horton, S. J., and Huntly, B. J. (2012). Recent advances in acute myeloid leukemia stem cell biology. *Haematologica* 97, 966–974. doi: 10.3324/haematol.2011.054734

Huynh, K. D., Fischle, W., Verdin, E., and Bardwell, V. J. (2000). BCOR, a novel corepressor involved in BCL-6 repression. *Genes Dev.* 14, 1810–1813.

Jünemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating bench top sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522

Kumar, R. D., Swamidass, S. J., and Bose, R. (2016). Unsupervised detection of cancer driver mutations with parsimony-guided learning. *Nat. Genet.* 48:1288. doi: 10.1038/ng.3658

Kuo, F. C., and Dong, F. (2015). Next-generation sequencing-based panel testing for myeloid neoplasms. *Curr. Hematol. Malig. Rep.* 10, 104–111. doi: 10.1007/s11899-015-0256-3

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., et al. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 42, D980–D985.

Lee, L. Y., Hernandez, D., Rajkhowa, T., Smith, S. C., Raman, J. R., Nguyen, B., et al. (2017). Preclinical studies of gilteritinib, a next-generation FLT3 inhibitor. *Blood* 129, 257–260. doi: 10.1182/blood-2016-10-745133

Lek, M., Karczewski, K., Minikel, E., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.1.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheelertransform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). 1000 Genome project data processing subgroup. The sequence alignment/map format and samtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liang, Y., Tebaldi, T., Rejeski, K., Joshi, P., Stefani, G., and Taylor, A. (2018). SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. *Leukemia* 32, 2659–2671. doi: 10.1038/s41375-018-0152-7

Liu, P., Morrison, C., Wang, L., Xiong, D., Vedell, P., Cui, P., et al. (2012). Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* 33, 1270–1276. doi: 10.1093/carcin/bgs148

Masaki, S., Ikeda, S., Hata, A., Shiozawa, Y., Kon, A., Ogawa, S., et al. (2019). Myelodysplastic syndrome-associated SRSF2 mutations cause splicing changes by altering binding motif sequences. *Front. Genet.* 10:338. doi: 10.3389/fgene.2019.00338

May Green, C. L., Koo, K. K., Hills, R. K., Burnett, A. K., Linch, D. C., and Gale, R. E. (2010). Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations. *J. Clin. Oncol.* 16, 2739–2747. doi: 10.1200/JCO.2009.26.2501

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., and Thormann, A. (2016). The ensembl variant effect predictor. *Genome Biol.* 17, 1–14.

Ostergaard, P., Simpson, M., Connell, F. C., Steward, C. G., Brice, G., Woollard, W. J., et al. (2011). Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat. Genet.* 43, 929–931. doi: 10.1038/ng.923

Pagan, J. K., Arnold, J., Hanchard, K. J., Kumar, R., Bruno, T., Jones, M. J. K., et al. (2007). A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP. *J. Biol. Chem.* 282, 15248–15257. doi: 10.1074/jbc.m700246200

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., et al. (2016). Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* 374, 2209–2221. doi: 10.1056/NEJMoa1516192

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122, 3616–36127. doi: 10.1182/blood-2013-08-518886

Pengyuan, L., Carl, M., Liang, W., Donghai, X., Peter, V., Peng, C., et al. (2012). Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* 33, 1270–1276.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121. doi: 10.1101/gr.097857.109

Prassek, V. V., Rothenberg-Thurley, M., Sauerland, M. C., Herold, T., Janke, H., Ksienzyk, B., et al. (2018). Genetics of acute myeloid leukemia in the elderly: mutation spectrum and clinical impact in intensively treated patients aged 75 years or older. *Haematologica* 11, 1853–1861. doi: 10.3324/haematol.2018.191536

Renneville, A., Roumier, C., Biggio, V., Nibourel, O., Boissel, N., Fenaux, P., et al. (2008). Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *Leukemia* 22, 915–931. doi: 10.1038/leu.2008.19

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., and Gastier-Foster, J. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the association for molecular pathology. *Genet. Med.* 17, 405–423. doi: 10.1038/gim.2015.30

Sanchez, C., Sanchez, I., Demmers, J. A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. *Mol. Cell. Proteomics* 6, 820–834. doi: 10.1074/mcp.m600275-mcp200

Shakeel, M., Irfan, M., and Khan, I. A. (2018). Estimating the mutational load for cardiovascular diseases in Pakistani population. *PLoS One* 13:e0192446. doi: 10.1371/journal.pone.0192446

Stein, E. M., DiNardo, C. D., Pollyea, D. A., Fathi, A. T., Roboz, G. J., Altman, J. K., et al. (2017). Enasidenib in mutant-IDH2relapsed or refractory acute myeloid leukemia. *Blood* 130, 722–731. doi: 10.1182/blood- 2017-04-779405

Swerdlow, S. H., Campo, E., Pileri, S. A., Harris, N. L., Stein, H., and Siebert, R. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127, 2375–2390. doi: 10.1182/blood-2016-01-643569

Thol, F., Bollin, R., Gehlhaar, M., Walter, C., Dugas, M., Suchanek, K. J., et al. (2014). Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood* 123, 914–920. doi: 10.1182/blood-2013-07-518746

Tijssen, M. R., Cvejic, A., Joshi, A., Hannah, R. L., Ferreira, R., Forrai, A., et al. (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev. Cell* 20, 597–609. doi: 10.1016/j.devcel.2011.04.008

Timchenko, N., Wilson, D. R., Taylor, L. R., Abdelsayed, S., Wilde, M., Sawadogo, M., et al. (1995). Autoregulation of the human C/EBP alpha gene by stimulation of upstream stimulatory factor binding. *Mol. Cell. Biol.* 15, 1192–1202. doi: 10.1128/mcb.15.3.1192

Tyner, J. W., Tognon, C. E., Bottomly, D., Wilmot, B., Kurtz, S. E., Savage, S. L., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. *Nature* 7728, 526–531. doi: 10.1038/s41586-018-0623-z

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33(Database issue), D433–D437. doi: 10.1093/nar/gki005

Wamstad, J. A., Corcoran, C. M., Keating, A. M., and Bardwell, V. J. (2008). Role of the transcriptional corepressor Bcor in embryonic stem cell differentiation and early embryonic development. *PLoS One* 3:e2814. doi: 10.1371/journal.pone.0002814

Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and WANNOVAR. *Nat. Protoc.* 10, 1556–1566. doi: 10.1038/nprot.2015.105

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fgene.2020.507816

Whole Genome Sequencing Identifies Key Genes in Spinal Schwannoma

Xin Gao[1†], Li Zhang[2,3†], Qi Jia[1†], Liang Tang[1], Wen Guo[1,4], Tao Wang[1], Zheyu Wu[1,5], Wang Zhou[1], Zhenxi Li[1]* and Jianru Xiao[1]**

[1] Orthopedic Oncology Center, Department of Orthopedics, Changzheng Hospital, Second Military Medical University, Shanghai, China, [2] Key Laboratory of Advanced Theory and Application in Statistics and Data Science – MOE, School of Statistics, East China Normal University, Shanghai, China, [3] Center for Bioinformatics and Computational Biology, School of Life Sciences, Institute of Biomedical Sciences, East China Normal University, Shanghai, China, [4] Department of Orthopedics, Taizhou People's Hospital, Taizhou, China, [5] Department of Orthopedics, Zhongnan Hospital of Wuhan University, Wuhan, China

OPEN ACCESS

Edited by:
Ira Ida Skvortsova,
Medical University of Innsbruck, Austria

Reviewed by:
Miguel Torres-Martin,
Mount Sinai Medical Center, United States
Bárbara Meléndez,
Hospital Virgen de la Salud, Spain
Netta Mäkinen,
Dana–Farber Cancer Institute, United States
Nives Pecina-Slaus,
University of Zagreb, Croatia

***Correspondence:**
Wang Zhou
brilliant212@163.com
Zhenxi Li
zhenxili.ecnu@gmail.com
Jianru Xiao
xiaojianruvvip@163.com

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Genetics

Received: 28 October 2019
Accepted: 10 September 2020
Published: 30 October 2020

Citation:
Gao X, Zhang L, Jia Q, Tang L, Guo W, Wang T, Wu Z, Zhou W, Li Z and Xiao J (2020) Whole Genome Sequencing Identifies Key Genes in Spinal Schwannoma. Front. Genet. 11:507816. doi: 10.3389/fgene.2020.507816

Spinal schwannoma is the most common primary spinal tumor but its genomic landscape and underlying mechanism driving its initiation remain elusive. The aim of the present study was to gain further insights into the molecular mechanisms of this kind of tumor through whole genome sequencing of nine spinal schwannomas and paired blood samples. The results showed that *ATM, CHD4, FAT1, KMT2D, MED12, NF2,* and *SUFU* were the most frequently mutated cancer-related genes. In addition, the somatic copy number alterations (CNA) was potentially associated with spinal schwannoma, among which *NF2* was found to be frequently deleted in schwannoma samples. Only a few genes were located within the amplified regions. In contrast, the deleted regions in 15q15.1 and 7q36.1 contained most of these genes. With respect to tumorigenesis, *NF2* had the highest variant allele frequency (VAF) than other genes, and homozygous deletion was observed in *NF1*, *NF2,* and *CDKN2C*. Pathway-level analysis suggested that Hippo signaling pathway may be a critical pathway controlling the initiation of spinal schwannoma. Collectively, this systematic analysis of DNA sequencing data revealed that some key genes including *NF1*, *NF2,* and *CDKN2C* and Hippo signaling pathway were associated with spinal schwannoma, which may help improve our understanding about the genomic landscape of spinal schwannoma.

Keywords: hippo signaling pathway, copy number alterations, frequently mutated genes, whole genome sequencing, spinal schwannoma

INTRODUCTION

Spinal schwannoma is the most common primary spinal tumor, accounting for almost one-third of all spinal tumors (Seppala et al., 1995; Abul-Kasim et al., 2008), with an annual incidence of 0.3 to 0.4 per 100,000 (Seppala et al., 1995). Patients with spinal schwannoma usually have to endure pain, spinal root deficits, pyramidal tract compression, and sphincter disorders, which negatively affect patients' quality of life (Lenzi et al., 2017). The gold standard treatment for spinal schwannomas is complete surgical resection (Jinnai and Koyama, 2005; Safaee et al., 2017). However, complete tumor removal is associated with a high incidence of complications,

in which cutting the involved functionally relevant nerve root is likely to result in a permanent and significant neurological deficit (Celli et al., 2005), with up to 30% of patients developing postoperative neurologic deteriorations (Safaee et al., 2017). Unfortunately, there are no medical therapies available for spinal schwannomas, while a safe and effective treatment for these patients is urgently needed (Campian and Gutmann, 2017). Therefore, it is of great importance to gain a comprehensive insight into the genetic landscape of schwannoma and to identify the potential genes driving its initiation, thereby seeking therapeutic targets.

Schwannomas are derived from Schwann cells, the myelin-producing cells of the peripheral nervous system. Schwannomas can occur either spontaneously or as the hallmark tumor of neurofibromatosis type 2. Previous studies have identified several schwannoma-predisposing mutations. Specifically, mutation of the *NF2* gene is the most characteristic genetic risk factor for schwannoma (Rouleau et al., 1993; Trofatter et al., 1993; Lin and Gutmann, 2013; Pathmanaban et al., 2017; Carlson et al., 2018; Roberts et al., 2019). Germline mutations in *LZTR1* predispose to an inherited disorder of multiple schwannomas (Piotrowski et al., 2014). Mutation of the *SMARCB1* gene was found in 45% of familial and 7% of sporadic schwannomatosis cases (Smith et al., 2012), and it was also reported to play a role in the development of some sporadic spinal schwannomas (Paganini et al., 2018). In order to determine the frequency of these predisposing mutations in solitary schwannomas, Pathmanaban et al. screened 135 schwannoma cases using Sanger sequencing or next-generation sequencing, and the results showed that *NF2*, *LZTR1*, or *SMARCB1* mutations were found in 54.5% of them (Pathmanaban et al., 2017). Furthermore, in an integrative analysis performed by Agnihotri et al. (2016) whole exome sequencing analysis identified recurrent mutations in *NF2*, *ARID1A*, *ARID1B*, and *DDR1*, and RNA sequencing identified a recurrent in-frame *SH3PXD2A-HTRA1* fusion in 10% cases. Whole exome sequencing was also performed on vestibular schwannomas. *NF2*, *CDC27*, and *USP8* were identified to be the most common tumor-specific mutations (Havik et al., 2018). Nevertheless, a focused attention on the genomic landscape of schwannomas in the spine is still needed. Furthermore, mutations in non-coding regions, such as introns, regulatory elements, and non-coding RNA, remain widely unexplored.

In this study, whole genome sequencing analysis of nine paired spinal schwannomas (ICD-O of 9560/0) and blood samples was performed to provide some useful information regarding the overall landscape of driver mutations and mutated pathways in schwannomas genomes.

MATERIALS AND METHODS

DNA Extraction and Whole Genome Sequencing

This study was approved by the Changzheng Hospital medical ethics committee, and genomic DNA samples were obtained from fresh tissues and blood samples after being resected during surgery. All the patients were recruited with written informed consent, in accordance with the Declaration of Helsinki. TIAN amp Blood DNA Kit (TIANGEN Biotech Co., Ltd., Beijing, China) was used for extracting genomic DNA from paired schwannoma and blood samples according to the manufacturer's instructions. The whole genome DNA were sequenced by Illumina X-Ten platform in Shanghai, China, and 150 bp paired-end reads were generated.

Reads Mapping and Variants Calling and Annotation

The whole genome paired reads of 300bp (150bp at each end) were mapped to human reference genome (UCSC hg19 assembly) using BWA 0.7.12 "mem" mode with default options (Li and Durbin, 2009). The PCR duplicates of the mapped reads and low-quality reads (BaseQ < 20) were then removed by SAMtools "rmdup" with version 0.1.19. The resulting bam files were sorted and indexed by SAMtools sort and index, respectively. Somatic mutations were called by Strelka 2.8.4 software (Saunders et al., 2012) with default options. The somatic mutations should have a minimal of 4 read counts supporting the variant and over 20 reads covering the locus. The somatic copy number alteration (CNA)s were called by SAASCNV 0.3.4 (Zhang and Hao, 2015) with P-value < 0.05. We used the ANNOVAR software for variant annotation (Wang et al., 2010). It should be noted that all the somatic mutations were identified in the tumor tissues but absent in the corresponding normal tissues.

Analyses of the Somatic Mutations in Schwannoma

The analyses included identification of potential genes driving schwannoma initiation, and identification of frequently mutated pathways. The potential genes driving schwannoma initiation were identified based on the variant allele frequency (VAF), and the frequently mutated pathways were identified based on the number of mutated genes in the oncogenic pathways. All the analyses were implemented in R maftools package (Mayakonda et al., 2018). The clonality analysis was conducted in R package CLONETv2 with default options (Carreira et al., 2014).

Significantly Amplified and Deleted Regions

The somatic CNAs were first segmented by SAASCNV (Zhang and Hao, 2015). The significantly amplified and deleted regions were identified by GISTIC 2.0 on the Gene Pattern webserver (Mermel et al., 2011). The CNAs with q-value < 0.05 were deemed as the significantly amplified and deleted regions.

Gene Set Enrichment Analysis

The gene sets collected from KEGG pathways were used in the enrichment analysis. The gene set enrichment analysis (GSEA) was conducted by hypergeometric test. The GSEA was implemented in R clusterProfiler package (Yu et al., 2012). The pathways were considered statistically significant if the q-value < 0.05.

RESULTS

Clinical Characteristics and Outcomes of the Patients

The characteristics of 9 patients are shown in **Table 1**. This cohort was comprised of 4 men and 5 women, with a mean age of 57.6 years (median 63, range 27–69). The most common symptom was pain and hypesthesia. One patient (case 8) was diagnosed with the recurrent schwannoma which was totally resected 15 years ago. Tumor size ranged from 1 to 6.5 cm in the maximum diameter, with 2 of them larger than 5 cm. Schwannomatosis were diagnosed in 3 patients who had multiple spinal schwannomas without bilateral vestibular schwannoma. No pathogenic germline NF2 mutations were found in these patients. Complete resection of the tumor(s) and posterior stabilization of the spine was performed for all patients. According to postoperative pathological examination, the tumor cell percentages of all these samples were above 88% (Range 88–95%). Postoperatively, all patients recovered well without surgical complications. The mean follow-up duration was 42.9 (median 45, range 33–48) months. All patients were alive with no evidence of disease at the last follow-up in July 2019.

Whole Genome Sequencing of Nine Spinal Schwannomas and Paired Blood Samples

To explore the alterations in the genome of spinal schwannoma, we performed whole genome sequencing (WGS) to a median depth of 36.77X (range, 32.34X to 40.59X), and identified 832 somatic mutations, which were present in tumors but absent in paired blood samples, including 763 single nucleotide variants (SNVs) and 69 insertion or deletions (InDels) across the 9 patients (**Figures 1A,B**), with a median of 87 variants (**Figure 1C**). Based on the RefSeq gene annotation, we identified *TTN*, *MUC4*, *FLG2*, *MUC17*, *OR2T4*, *ZNF850*, *FAM186A*, *ALMS1*, *FAM47C*, and *ATM* as the top ten mutated genes (**Figure 1D**), however, only *ATM* was previously reported to be implicated in cancer (Kim et al., 2014; Chen et al., 2015; Feng et al., 2015).

Genetic Landscape of Spinal Schwannoma

To identify the genes responsible for schwannoma, we further investigated whether the mutated genes could be found in the COSMIC Cancer Gene Census database (Forbes et al., 2008), a database curating genes causally implicated in cancer. As shown **Figure 2A**, *ATM*, accounting for 33% of the samples, was the most frequently mutated cancer-related gene in schwannoma, followed by *CHD4* (22%), *FAT1* (22%), *KMT2D* (22%), *MED12* (22%), *NF2* (22%), and *SUFU* (22%). Most of them were predicted as pathogenic by SIFT, PolyPhen-2, or MutationTaster (**Table 2**). Particularly, *NF2* (Neurofibromin 2) has been widely identified as the pathogenic genes for both sporadic and familial schwannoma (Jacoby et al., 1994; Rm et al., 2018).

Moreover, we also profiled the somatic CNA in schwannoma based on the GISTIC algorithm (q-value < 0.05). The identified significantly amplified regions were 1p36.33, 2q37.3, 4q35.2, 9p24.3, 10q26.3, 11p11.12, 12p13.33, 15q26.3, 16p13.3, and 17q25.3, while the significantly deleted regions were 12q24.31, 15q15.1, 19q13.42, 22q11.1, 22q11.21, and 22q11.23 (**Figure 2B**). In accordance with the somatic mutations and InDels, *NF2*, harbored in the cytoband 22q11, was also identified to be frequently deleted in the samples of schwannoma (**Supplementary Figure S1**). These results indicated that *NF2* played a key role in the initiation of schwannoma.

Furthermore, we also investigated the genes within these CNA regions potentially involved in spinal schwannoma. Only a few genes were located within the amplified regions. whereas the deleted regions contained most of these genes

TABLE 1 | Clinical characteristics of nine patients with spinal schwannoma.

No.	Sex	Age	Location	Single/ multiple*	Primary/ recurrent	Histological grade	Tumor size (cm)	Adjacent relations to spinal canal	Shape	Bone structure destruction	Resection mode	Follow-up (M)	Final Status
1	M	47	L1-2	Multiple	Primary	Grade I	2.5; 1.5; 1	Intraspinal	Round	No	Total	35	NED
2	M	66	L2; S1	Multiple	Primary	Grade I	2.4; 1.8	Intraspinal	Round	No	Total	47	NED
3	F	65	C6-7	Single	Primary	Grade I	6	Intra-extraspinal	Irregular	Yes	Total	33	NED
4	F	53	L2	Single	Primary	Grade I	2.4	Intraspinal	Ellipse	No	Total	48	NED
5	M	62	S1-2	Single	Primary	Grade I	3.5	Intra-extraspinal	Round	yes	Total	43	NED
6	F	63	T1-2	Single	Primary	Grade I	4.3	Intra-extraspinal	Dumbbell	No	Total	46	NED
7	M	27	L4-5	Single	Primary	Grade I	6.5	Intra-extraspinal	Dumbbell	Yes	Total	45	NED
8	F	66	L1-2	Single	Recurrent	Grade I	4.2	Intraspinal	Ellipse	No	Total	44	NED
9	F	69	L3-4	Multiple	Primary	Grade I	2; 1.1	Intraspinal	Round	No	Total	45	NED

*NED, no evidence of disease. *The three patients with multiple spinal schwannomas were diagnosed as schwannomatosis. They were sporadic schwannoma(s) without family history.*

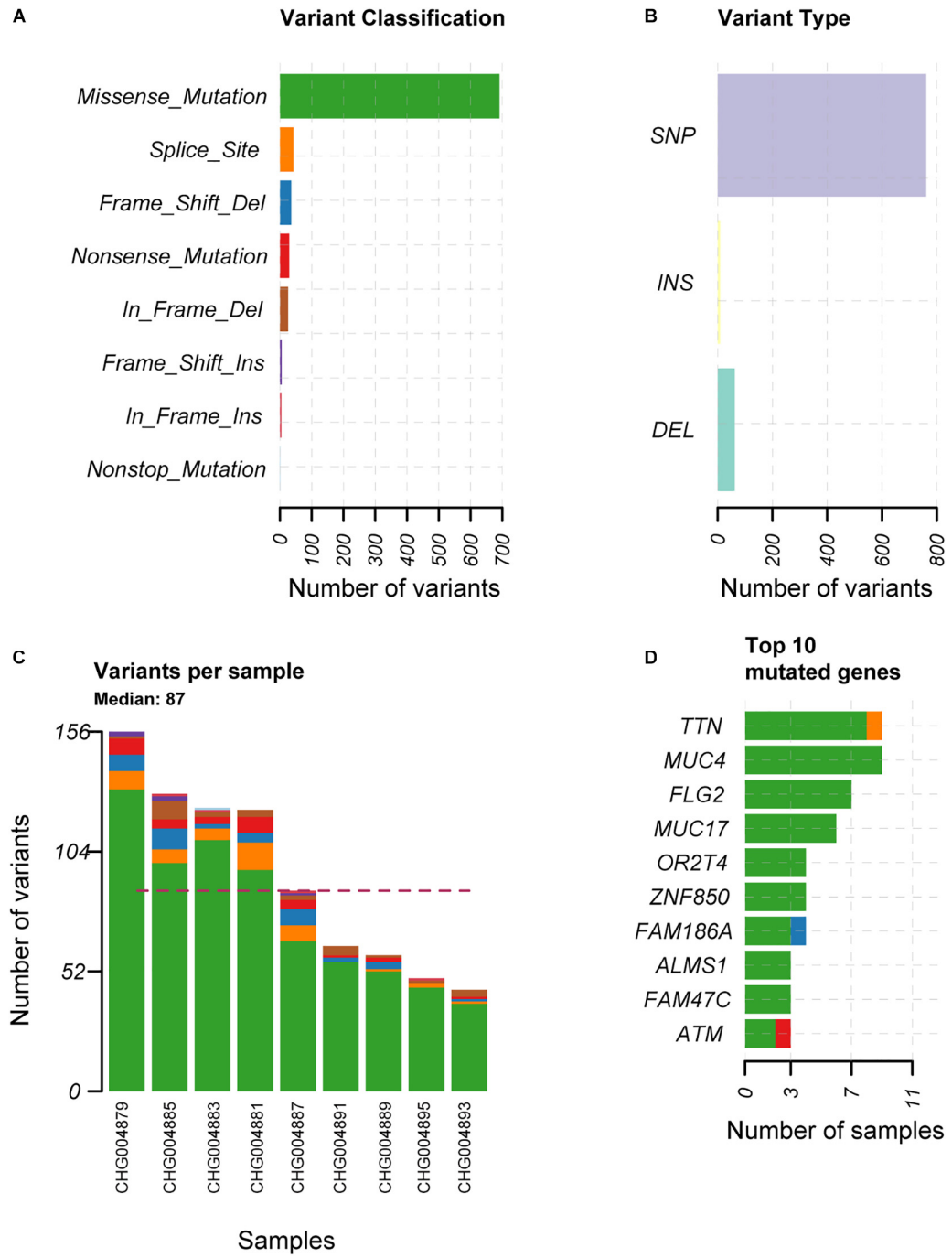


FIGURE 1 | The overview of the somatic mutations in spinal schwannoma. The variant classification, variant type, the number of variants per sample, and the top-ten mutated genes in the nine spinal schwannomas are displayed in **(A–D)**, respectively. The *x*-axis in **(A,B,D)** represent the number of mutations or samples. The *x*-axis in **(C)** represents the nine samples.

(**Supplementary Tables S1, S2**). Specifically, *ITPKA*, *LTK*, *CHP*, *OIP5*, *RTF1*, *RPAP1*, *NDUFAF1*, *NUSAP1*, *INO80*, *EXD1*, and *OIP5-AS1*, fell within the cytoband of 15q15.1 (n = 6, 66.7%). The gene set enrichment analysis revealed that these genes might participate in the maintenance of chromosomal structure, such as DNA conformation change, chromosome segregation, and nucleosome organization (q-value < 0.05). In addition, we observed that *XRCC2* and *MLL3/KMT2C* were located within 7q36.1 (n = 5, 55.6%), which had a q-value 0.057, slightly higher than the threshold 0.05, however, the two genes were involved in the tumorigenesis or progression of several malignant tumors (**Figure 2B** and

TABLE 2 | The pathogenicity of the somatic mutations in spinal schwannoma by SIFT, PolyPhen-2, and MutationTaster.

Gene Symbol	Effect	Transcript:exon:CDS	SIFT	Polyphen2-HDIV	Polyphen2-HVAR	Mutation Taster	Sample	VAF
ATM	Non-sense	NM_000051:exon46:c.C6725A:p.S2242X	NA	NA	NA	D	CHG004878	0.13
ATM	Missense	NM_000051:exon60:c.T8699C:p.L2900P	D	D	D	D	CHG004882	0.13
ATM	Missense	NM_000051:exon24:c.T3458C:p.V1153A	T	B	B	N	CHG004884	0.10
CHD4	Missense	NM_001297553:exon9:c.C1460T	T	D	B	D	CHG004878	0.13
CHD4	Missense	NM_001297553:exon18:c.G2831A:p.G944E	D	D	D	D	CHG004884	0.34
FAT1	Non-sense	NM_005245:exon2:c.G370T:p.E124X				A	CHG004880	0.14
FAT1	Missense	NM_005245:exon10:c.C6110T:p.T2037M	D	B	B	N	CHG004888	0.18
KMT2D	Missense	NM_003482:exon34:c.G9484A:p.G3162S	D	B	B	N	CHG004878	0.10
KMT2D	Missense	NM_003482:exon48:c.G15713A:p.R5238Q	T	D	P	D	CHG004892	0.14
MED12	Missense	NM_005120:exon3:c.C385A:p.L129I	D	D	D	D	CHG004886	0.25
MED12	Missense	NM_005120:exon28:c.G4037A:p.R1346H	D	P	B	D	CHG004890	0.25
NF2	Missense	NM_181830:exon14:c.G1517A:p.C506Y	D	D	D	D	CHG004880	0.22
NF2	Non-sense	NM_181830:exon10:c.C979T:p.Q327X				A	CHG004884	0.50
NF2	Inframe deletion	NM_181830:exon6:c.513_557del:p.171_186del	NA	NA	NA	NA	CHG004880	0.53
SUFU	Missense	NM_001178133:exon10:c.C1177T:p.R393W	D	D	D	D	CHG004880	0.49
SUFU	Missense	NM_001178133:exon6:c.G691A:p.G231S	D	D	D	D	CHG004886	0.11

D, deleterious; P, probably deleterious; B, benign; T, tolerant; A, disease causing automatic; N, polymorphism; NA, not available.

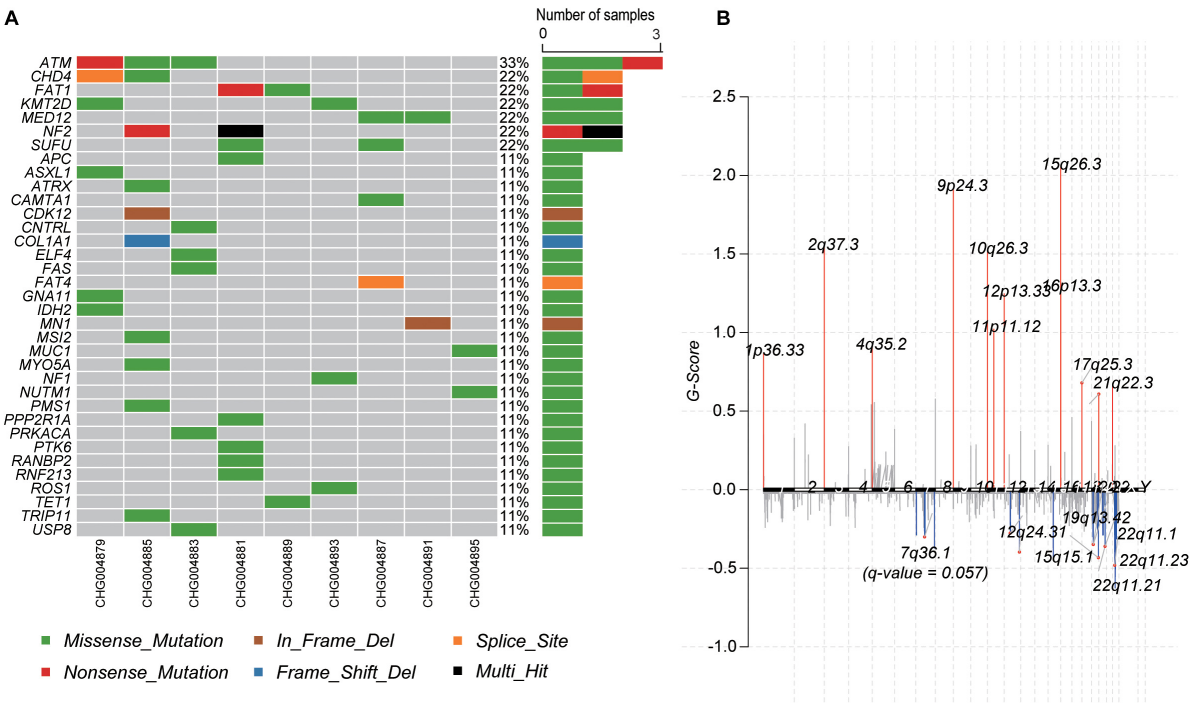


FIGURE 2 | The mutational landscape of spinal schwannoma. **(A)** The somatic mutations of cancer driver genes across the nine spinal schwannomas. **(B)** The somatic copy number alterations along the chromosomes. The G-score was to evaluate the significance of the gains and losses.

Supplementary Table S1), suggesting that these genes might play key roles in schwannoma.

The Potential Genes Driving Schwannoma Initiation

With the whole genome sequencing data, we aimed to identify the genes potentially driving schwannoma initiation. In tumorigenesis, the more the VAF was close to 50%, the earlier the mutation may occur. We then ranked the genes based on the median of VAF across the samples (**Figure 3A**). Among the genes mutated in more than one sample, *NF2* had a higher VAF than other genes, followed by *DEPDC5*, which were involved in mTOR signaling pathway. Furthermore, we also conducted the subclonality analysis, and identified

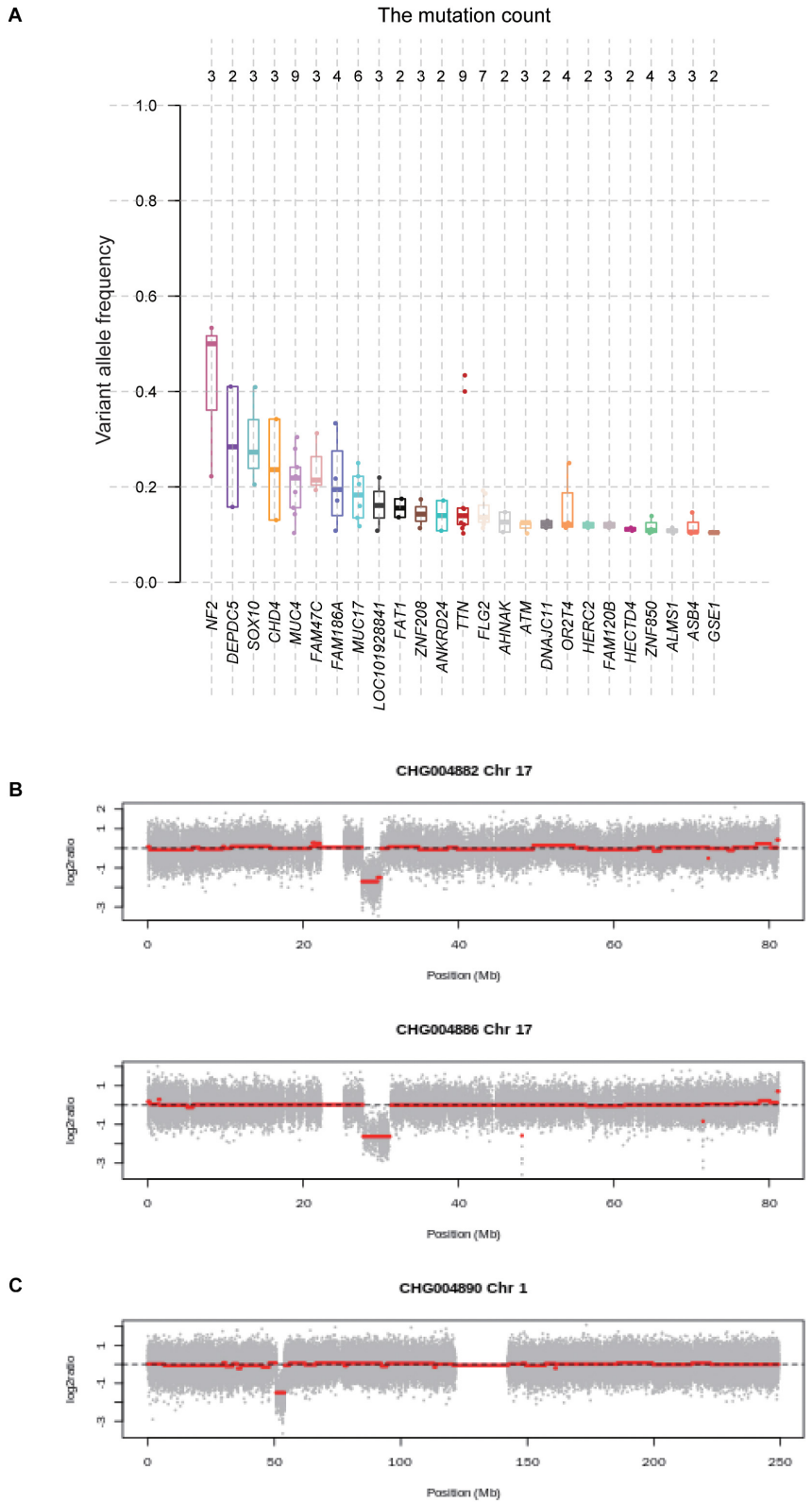


FIGURE 3 | The variant allele frequency (VAF) of the genes mutated in spinal schwannomas. **(A)** The numbers on the top represent the number of mutations. The genes are ordered by the median of VAF across the mutations. **(B,C)** The log2 copy number ratio of chromosomes 17 and 1 in the spinal schwannoma samples.

that *NF2* gene mutations were present in all subclones of CHG004880 and CHG004884, suggesting that *NF2* was a candidate driver in spinal schwannoma (**Supplementary Figure S2**). Particularly, *SOX10* and *CHD4* also showed relatively high VAF among those genes (VAF > 20%). In addition, we also examined the five samples without *NF2* gene mutations. Interestingly, the spinal schwannoma samples of CHG004882 and CHG004886 harbored homozygous deletion in *NF1* (**Figure 3B**). Homozygous deletion of CDKN2C was observed in CHG004890 (**Figure 3C**). These results indicated that the CNV loss in *NF1* and *CDKN2C* might also contribute to tumorigenesis of spinal schwannoma.

Identification of Frequently Mutated Pathways in Schwannoma

To further investigate the critical pathways that controlled the initiation of schwannoma, we mapped the genes with somatic mutations to the oncogenic pathways. The Hippo signaling pathway was observed to be frequently mutated in schwannoma (**Figure 4A**). Combining the somatic mutations with CNAs, we found that *NF2*, *SAV1*, *LLGL1/2*, and *CSNK1D/E* were key regulators with mutations in Hippo signaling pathway (**Figure 4B**). Particularly, *NF2*, *SAV1*, and *LLGL1/2*, the upstream regulators of YAP/TAZ transcription factors, were frequently deleted or mutated with loss-of-function patterns, which may promote the transcription of YAP/TAZ target genes (**Figure 4B**). Among the 9 samples, 6 were detected to have mutations in the Hippo signaling pathway (**Figure 4C**), accounting for 67% of the samples. These results suggested that the Hippo signaling pathway was a critical pathway for schwannoma.

DISCUSSION

Spinal schwannoma is the most common primary spinal tumor, typically arising from spinal nerve roots. To investigate the genomic landscape of the spinal tumor, we performed whole genome sequencing of nine tumors and paired blood samples with high coverage (**Supplementary Table S3**). Specifically, we identified *TTN*, *MUC4*, *FLG2*, *MUC17*, *OR2T4*, *ZNF850*, *FAM186A*, *ALMS1*, *FAM47C*, and *ATM* as the top ten mutated genes. However, further analysis of those frequently mutated genes revealed that most of them were long genes and had lower VAF, indicating that most of these genes were passengers following the driver genes. Particularly, *ATM*, accounting for 33% of the samples, was the most frequently mutated cancer-related gene in schwannoma, followed by *CHD4*, *FAT1*, *KMT2D*, *MED12*, *NF2*, and *SUFU* (>20%). Notably, *SUFU* was reported to be responsible for meningioma (Aavikko et al., 2012), a type of tumor from central nervous system. In accordance with previous studies of sporadic schwannoma by next generation sequencing (Agnihotri et al., 2016; Havik et al., 2018), *NF2* was identified as the pathogenic gene in both spinal and vestibular schwannomas. As the other genes like *CHD4*, *FAT1*, *KMT2D*, and *MED12* were not detected in vestibular schwannoma, we thus speculated that some of them might be specific in spinal schwannoma.

Moreover, we also profiled the somatic CNA in schwannoma. *NF2* gene, harbored in the cytoband 22q11, was also identified to be frequently deleted in the samples of spinal schwannoma. The integrative analysis of the *NF2* point mutations, InDels and copy number deletions revealed that CHG004880 and CHG004884 had bi-allelic mutations in *NF2*, resulting in bi-allelic NF2 inactivation. Consistently, the occurrence of both somatic mutations and CNAs in *NF2* has been reported to result in schwannoma by previous studies (Jacoby et al., 1994; Rm et al., 2018).

Furthermore, we also investigated the genes within these CNAs in spinal schwannoma. We found that there were only a few genes located within the amplified regions. In contrast, the deleted regions contained a majority of genes (**Supplementary Tables S1, S2**). The higher frequency in deletions than gains suggested that the tumorigenesis of spinal schwannoma might be caused by these deletions. Particularly, losses of *XRCC2* and *MLL3/KMT2C*, which were located within 7q36.1, were also observed in the spinal schwannomas. *XRCC2* and *MLL3/KMT2C* were involved in homologous recombination repair and histone modification, suggesting that loss of DNA damage repair and epigenetic alterations might also play key roles in schwannoma.

Besides, among the somatic mutations and CNAs in the nine spinal schwannomas, *SMARCB1*, *SMARCE1* or *LZTR1* were absent in these spinal schwannoma tissues. As the familial spinal schwannoma usually have germline mutations in *SMARCB1*, *SMARCE1* or *LZTR1*, to our knowledge, these mutations are not prevalent in sporadic spinal schwannoma patients.

In addition, to identify genes potentially initiating the schwannoma, we ranked the genes with somatic mutations by the median of VAF across the samples (**Figure 3A**). Combined with the clonality analysis, *NF2* gene was a candidate driver in spinal schwannoma. Besides *NF2*, *SOX10* and *CHD4* also showed relatively high VAF among those genes (VAF > 20%). Notably, a previous study reported that loss of SOX10 function contributed to the phenotype of human Merlin-null schwannoma cells, indicating that the mutation of SOX10 might be associated with the initiation of schwannoma (Doddrell et al., 2013). Moreover, the different location of CHD4 staining has been reported to be used as a potential biomarker to differentiate cellular schwannoma from malignant peripheral sheath tumor (MPNST) (Wu et al., 2018). In addition, *NF1* and *CDKN2C* were also found to be homozygously deleted in spinal schwannoma. To our knowledge, *NF1* mutations were rarely reported in spinal schwannoma. Exceptionally, lack of NF1 expression has been observed in a sporadic schwannoma from a patient without neurofibromatosis (Gutmann et al., 1995), suggesting that *NF1* might be a novel driver in spinal schwannoma. Similarly, loss of *CDKN2C*, a cell growth regulator that controls cell cycle G1 progression, might also contribute to spinal schwannoma by the manner of CDKN2A (Koutsimpelas et al., 2011; Rohrich et al., 2016). Moreover, the pathway-level analysis revealed that Hippo signaling pathway was one of the frequently mutated pathways (6/9, 67%), in which, *NF2*, *SAV1*, and *LLGL1/2* were frequently deleted or mutated with loss-of-function patterns. *SAV1* and *LLGL1/2* have not been reported to cause the tumorigenesis of spinal schwannoma. However, the Hippo

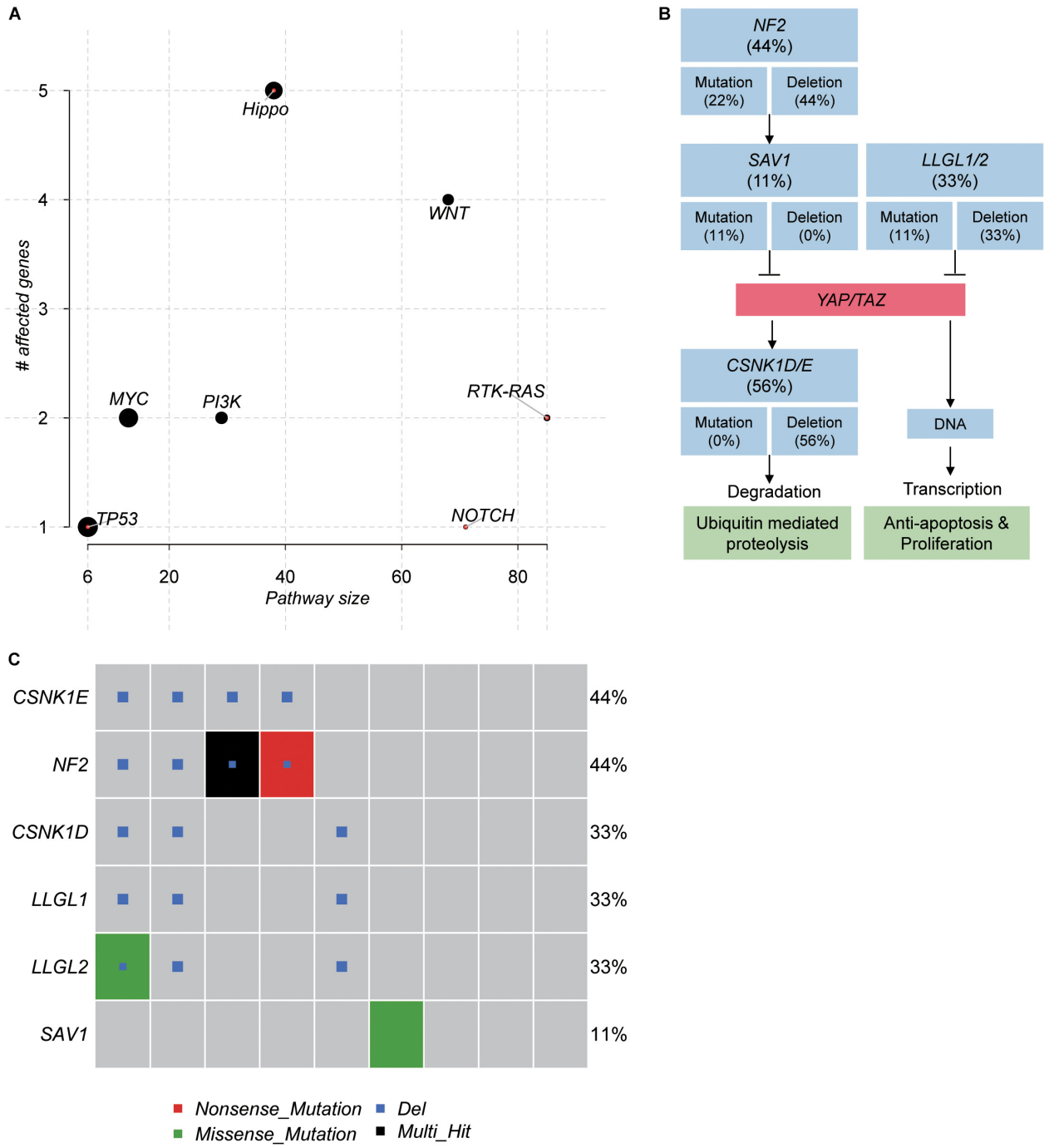


FIGURE 4 | The pathways frequently mutated in spinal schwannoma. **(A)** The significance of the pathways frequently mutated in spinal schwannoma. The *y*-axis and *x*-axis represent the total number of genes and the number of mutated genes in the pathways. **(B)** The mutated genes and frequency in Hippo signaling pathway. **(C)** The number of mutated genes in Hippo signaling pathway across the nine spinal schwannomas. The *x*-axis represents the spinal schwannoma samples. The Multi_Hit indicates multiple variants of a gene were identified in the same sample.

signaling pathway has been widely reported to be implicated in schwannoma by previous studies (Nikuseva-Martic et al., 2007; Oh et al., 2015; Brodhun et al., 2017; Zhao et al., 2018). *SAV1* promotes activation of MST-LATS kinase cascade in Hippo signaling, which suppresses the activities of YAP/TAZ by phosphorylation (Bae et al., 2017). *LLGL1/2* have been recognized as direct negative regulators of YAP/TAZ via phosphorylation (Cordenonsi et al., 2011). The loss of these tumor suppressors might lead to lower phosphorylation of YAP/TAZ, thereby promoting YAP/TAZ to translocate to the cell nucleus.

Admittedly, the present study still had some limitations. First, the sample size is small, which limited the accuracy of the conclusions. Second, the pathogenicity of the mutated genes in schwannoma need to be further validated. Third, the consequences of these potentially pathogenic genes and the underlying mechanisms should be further investigated. However, we aimed to identify the potentially pathogenic genes responsible for schwannoma, and the systematic analysis of DNA sequencing data improved our understanding of the genomic landscape in spinal schwannoma.

DATA AVAILABILITY STATEMENT

The raw data have been uploaded to the National Omics Data Encyclopedia (NODE) database (https://www.biosino.org/node/) with accession number OEP000894.

AUTHOR CONTRIBUTIONS

XG, WZ, ZL, and JX conceived and designed the study. LZ conducted the analyses. XG and QJ drafted the manuscript. LT, WG, TW, and ZW assisted with the analysis and the interpretation of the results. All authors read and approved the manuscript before submission.

FUNDING

This work was supported by Key Project funding in the Basic Research Field of the Shanghai Municipal Science and Technology Commission (17JC1400903), National Natural Science Foundation of China (82002838), and Shanghai Youth Science and Technology Talent Sailing Program (20YF1449100).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.507816/full#supplementary-material

Supplementary Figure 1 | The coverage and B allele frequency of chromosome 22 in samples of CHG004878, CHG004880, CHG004884, and CHG004888. The whole chromosome loss in chr22 was observed in samples CHG004878, CHG004880, and CHG004884, but the CHG004888 was found to have focal deletion adjacent to NF2. The position of NF2 in chr22 was highlighted by blue boxes.

Supplementary Figure 2 | The read alignment of NF2 mutations in CHG004880 and CHG004884 by integrative genomic viewer (IGV).

Supplementary Figure 3 | The high resolution version of **Figure 3B**.

REFERENCES

Aavikko, M., Li, S. P., Saarinen, S., Alhopuro, P., Kaasinen, E., Morgunova, E., et al. (2012). Loss of SUFU function in familial multiple meningioma. *Am. J. Hum. Genet.* 91, 520–526. doi: 10.1016/j.ajhg.2012.07.015

Abul-Kasim, K., Thurnher, M. M., McKeever, P., and Sundgren, P. C. (2008). Intradural spinal tumors: current classification and MRI features. *Neuroradiology* 50, 301–314. doi: 10.1007/s00234-007-0345-7

Agnihotri, S., Jalali, S., Wilson, M. R., Danesh, A., Li, M., Klironomos, G., et al. (2016). The genomic landscape of schwannoma. *Nat. Genet.* 48, 1339–1348.

Bae, S. J., Ni, L., Osinski, A., Tomchick, D. R., Brautigam, C. A., and Luo, X. (2017). SAV1 promotes Hippo kinase activation through antagonizing the PP2A phosphatase STRIPAK. *Elife* 6:e30278.

Brodhun, M., Stahn, V., and Harder, A. (2017). Pathogenesis and molecular pathology of vestibular schwannoma. *HNO* 65, 362–372.

Campian, J., and Gutmann, D. H. (2017). CNS Tumors in Neurofibromatosis. *J. Clin. Oncol.* 35, 2378–2385. doi: 10.1200/jco.2016.71.7199

Carlson, M. L., Smadbeck, J. B., Link, M. J., Klee, E. W., Vasmatzis, G., and Schimmenti, L. A. (2018). Next Generation Sequencing of Sporadic Vestibular Schwannoma: Necessity of Biallelic NF2 Inactivation and Implications of Accessory Non-NF2 Variants. *Otol. Neurotol.* 39, 860–871e.

Carreira, S., Romanel, A., Goodall, J., Grist, E., Ferraldeschi, R., Miranda, S., et al. (2014). Tumor clone dynamics in lethal prostate cancer. *Sci. Transl. Med.* 6:254ra125. doi: 10.1126/scitranslmed.3009448

Celli, P., Trillo, G., and Ferrante, L. (2005). Spinal extradural schwannoma. *J. Neurosurg. Spine* 2, 447–456. doi: 10.3171/spi.2005.2.4.0447

Chen, W. T., Ebelt, N. D., Stracker, T. H., Xhemalce, B., Van Den, Berg CL, and Miller, K. M. (2015). ATM regulation of IL-8 links oxidative stress to cancer cell migration and invasion. *Elife* 4:e07270.

Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., et al. (2011). The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell* 147, 759–772. doi: 10.1016/j.cell.2011.09.048

Doddrell, R. D., Dun, X. P., Shivane, A., Feltri, M. L., Wrabetz, L., Wegner, M., et al. (2013). Loss of SOX10 function contributes to the phenotype of human Merlin-null schwannoma cells. *Brain* 136(Pt 2), 549–563. doi: 10.1093/brain/aws353

Feng, X., Li, H., Dean, M., Wilson, H. E., Kornaga, E., Enwere, E. K., et al. (2015). Low ATM protein expression in malignant tumor as well as cancer-associated stroma are independent prognostic factors in a retrospective study of early-stage hormone-negative breast cancer. *Breast Cancer Res.* 17:65.

Forbes, S. A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., et al. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.* 2008:18428421

Gutmann, D. H., Silos-Santiago, I., Geist, R. T., Daras, M., and Rutkowski, J. L. (1995). Lack of NF1 expression in a sporadic schwannoma from a patient without neurofibromatosis. *J. Neurooncol.* 25, 103–111. doi: 10.1007/bf01057754

Havik, A. L., Bruland, O., Myrseth, E., Miletic, H., Aarhus, M., Knappskog, P. M., et al. (2018). Genetic landscape of sporadic vestibular schwannoma. *J. Neurosurg.* 128, 911–922. doi: 10.3171/2016.10.jns161384

Jacoby, L. B., MacCollin, M., Louis, D. N., Mohney, T., Rubio, M. P., Pulaski, K., et al. (1994). Exon scanning for mutation of the NF2 gene in schwannomas. *Hum. Mol. Genet.* 3, 413–419. doi: 10.1093/hmg/3.3.413

Jinnai, T., and Koyama, T. (2005). Clinical characteristics of spinal nerve sheath tumors: analysis of 149 cases. *Neurosurgery* 56, 510–515. doi: 10.1227/01.neu.0000153752.59565.bb

Kim, H., Saka, B., Knight, S., Borges, M., Childs, E., Klein, A., et al. (2014). Having pancreatic cancer with tumoral loss of ATM and normal TP53 protein expression is associated with a poorer prognosis. *Clin. Cancer Res.* 20, 1865–1872. doi: 10.1158/1078-0432.ccr-13-1239

Koutsimpelas, D., Felmeden, U., Mann, W. J., and Brieger, J. (2011). Analysis of cytogenetic aberrations in sporadic vestibular schwannoma by comparative genomic hybridization. *J. Neurooncol.* 103, 437–443. doi: 10.1007/s11060-010-0412-5

Lenzi, J., Anichini, G., Landi, A., Piciocchi, A., Passacantilli, E., Pedace, F., et al. (2017). Spinal Nerves Schwannomas: Experience on 367 Cases-Historic Overview on How Clinical, Radiological, and Surgical Practices Have Changed over a Course of 60 Years. *Neurol. Res. Int.* 2017:3568359.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lin, A. L., and Gutmann, D. H. (2013). Advances in the treatment of neurofibromatosis-associated tumours. *Nat. Rev. Clin. Oncol.* 10, 616–624. doi: 10.1038/nrclinonc.2013.144

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome. Res.* 28, 1747–1756. doi: 10.1101/gr.239244.118

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome. Biol.* 12:R41.

Nikuseva-Martic, T., Beros, V., Pecina-Slaus, N., Pecina, H. I., and Bulic-Jakus, F. (2007). Genetic changes of CDH1, APC, and CTNNB1 found in human brain tumors. *Pathol. Res. Pract.* 203, 779–787. doi: 10.1016/j.prp.2007.07.009

Oh, J. E., Ohta, T., Satomi, K., Foll, M., Durand, G., McKay, J., et al. (2015). Alterations in the NF2/LATS1/LATS2/YAP Pathway in Schwannomas. *J. Neuropathol. Exp. Neurol.* 74, 952–959. doi: 10.1097/nen.0000000000000238

Paganini, I., Capone, G. L., Vitte, J., Sestini, R., Putignano, A. L., Giovannini, M., et al. (2018). Double somatic SMARCB1 and NF2 mutations in sporadic spinal schwannoma. *J. Neurooncol.* 137, 33–38. doi: 10.1007/s11060-017-2711-6

Pathmanaban, O. N., Sadler, K. V., Kamaly-Asl, I. D., King, A. T., Rutherford, S. A., Hammerbeck-Ward, C., et al. (2017). Association of Genetic Predisposition With Solitary Schwannoma or Meningioma in Children and Young Adults. *JAMA Neurol.* 74, 1123–1129. doi: 10.1001/jamaneurol.2017.1406

Piotrowski, A., Xie, J., Liu, Y. F., Poplawski, A. B., Gomes, A. R., Madanecki, P., et al. (2014). Germline loss-of-function mutations in LZTR1 predispose to an inherited disorder of multiple schwannomas. *Nat. Genet.* 46, 182–187. doi: 10.1038/ng.2855

Rm, D. E. C., Decsa, C., Pinto, G. R., Paschoal, E. H. A., Tuji, F. M., Donb, B., et al. (2018). Frequency of the Loss of Heterozygosity of the NF2 Gene in Sporadic Spinal Schwannomas. *Anticancer Res.* 38, 2149–2154.

Roberts, D. S., Maurya, R., Takemon, Y., Vitte, J., Gong, L., Zhao, J., et al. (2019). Linked-read Sequencing Analysis Reveals Tumor-specific Genome Variation Landscapes in Neurofibromatosis Type 2 (NF2) Patients. *Otol. Neurotol.* 40, 150–159e.

Rohrich, M., Koelsche, C., Schrimpf, D., Capper, D., Sahm, F., Kratz, A., et al. (2016). Methylation-based classification of benign and malignant peripheral nerve sheath tumors. *Acta Neuropathol.* 131, 877–887.

Rouleau, G. A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., et al. (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* 363, 515–521. doi: 10.1038/363515a0

Safaee, M. M., Lyon, R., Barbaro, N. M., Chou, D., Mummaneni, P. V., Weinstein, P. R., et al. (2017). Neurological outcomes and surgical complications in 221 spinal nerve sheath tumors. *J. Neurosurg. Spine* 26, 103–111. doi: 10.3171/2016.5.spine15974

Saunders, C. T., Wong, W. S., Swamy, S., Becq, J., Murray, L. J., and Cheetham, R. K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28, 1811–1817. doi: 10.1093/bioinformatics/bts271

Seppala, M. T., Haltia, M. J., Sankila, R. J., Jaaskelainen, J. E., and Heiskanen, O. (1995). Long-term outcome after removal of spinal schwannoma: a clinicopathological study of 187 cases. *J. Neurosurg.* 83, 621–626. doi: 10.3171/jns.1995.83.4.0621

Smith, M. J., Wallace, A. J., Bowers, N. L., Rustad, C. F., Woods, C. G., Leschziner, G. D., et al. (2012). Frequency of SMARCB1 mutations in familial and sporadic schwannomatosis. *Neurogenetics* 13, 141–145. doi: 10.1007/s10048-012-0319-8

Trofatter, J. A., MacCollin, M. M., Rutter, J. L., Murrell, J. R., Duyao, M. P., Parry, D. M., et al. (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* 72, 791–800. doi: 10.1016/0092-8674(93)90406-g

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucl. Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

Wu, C. C., Pan, M. R., Wei, Y. C., Lin, C. H., Yang, S. F., Tsai, H. P., et al. (2018). CHD4 as a Potential Biomarker in Differentiating Between Cellular Schwannoma and Malignant Peripheral Nerve Sheath Tumor. *Appl. Immunohistochem. Mol. Morphol.* 26, 775–780.

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, Z., and Hao, K. (2015). SAAS-CNV: A Joint Segmentation Approach on Aggregated and Allele Specific Signals for the Identification of Somatic Copy Number Alterations with Next-Generation Sequencing Data. *PLoS Comput. Biol.* 11:e1004618. doi: 10.1371/journal.pcbi.1004618

Zhao, F., Yang, Z., Chen, Y., Zhou, Q., Zhang, J., Liu, J., et al. (2018). Deregulation of the Hippo Pathway Promotes Tumor Cell Proliferation Through YAP Activity in Human Sporadic Vestibular Schwannoma. *World Neurosurg.* 117, 269–279e.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 December 2020
doi: 10.3389/fgene.2020.521004

Identification and Validation of Potential Pathogenic Genes and Prognostic Markers in ESCC by Integrated Bioinformatics Analysis

*Lu Tang[1], Yuqiao Chen[1], Xiong Peng[2], Yuan Zhou[1], Hong Jiang[3], Guo Wang[4] and Wei Zhuang[1]**

[1] Department of Thoracic Surgery, Xiangya Hospital, Central South University, Changsha, China, [2] Department of Thoracic Surgery, The Second Xiangya Hospital, Central South University, Changsha, China, [3] Department of Neurology, Xiangya Hospital, Central South University, Changsha, China, [4] Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Anton A. Buzdin,
I.M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Marco Trerotola,
University of Studies G. d'Annunzio Chieti–Pescara, Italy
Nikolay Mikhaylovich Borisov,
Moscow Institute of Physics and Technology, Russia

***Correspondence:**
Wei Zhuang
zhuangwei@csu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Genetics

Received: 17 December 2019
Accepted: 26 October 2020
Published: 10 December 2020

Citation:
Tang L, Chen Y, Peng X, Zhou Y, Jiang H, Wang G and Zhuang W (2020) Identification and Validation of Potential Pathogenic Genes and Prognostic Markers in ESCC by Integrated Bioinformatics Analysis. Front. Genet. 11:521004. doi: 10.3389/fgene.2020.521004

Esophageal squamous cell carcinoma (ESCC) is one of the most fatal malignancies of the digestive tract, but its underlying molecular mechanisms are not known. We aim to identify the genes involved in ESCC carcinogenesis and discover potential prognostic markers using integrated bioinformatics analysis. Three pairs of ESCC tissues and paired normal tissues were sequenced by high-throughput RNA sequencing (RNA-seq). Integrated bioinformatics analysis was used to identify differentially expressed coding genes (DECGs) and differentially expressed long non-coding RNA (lncRNA) genes (DELGs). A protein–protein interaction (PPI) network of DECGs was established using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) website and visualized with Cytoscape. Survival analysis was conducted by log-rank tests to identify "hub" genes with potential prognostic value, and real-time reverse transcription-quantitative polymerase chain reaction (RT-qPCR) was conducted to assess expression of these genes in ESCC tissues. Transwell^{TM} assays were employed to examine the migration ability of cells after knockdown of *LINC01614* expression, followed by investigation of epithelial–mesenchymal transition (EMT) by western blotting (WB). A total of 106 upregulated genes and 42 downregulated genes were screened out from the ESCC data sets. Survival analysis showed two hub protein-coding genes with higher expression in module 1 of the PPI network (*SPP1* and *BGN*) and another three upregulated lncRNAs (*LINC01614*, *LINC01415*, *NKILA*) that were associated with a poor prognosis. High expression of *SPP1*, *BGN*, *LINC01614*, and *LINC01415* in tumor samples was validated further by RT-qPCR. *In vitro* experiments show that knockdown of *LINC01614* expression could significantly inhibit the migration of ESCC cells by regulating EMT, which was confirmed by WB. These results indicate that *BGN*, *SPP1*, *LINC01614*, and *LINC01415* might be critical genes in ESCC and potential prognostic biomarkers.

Keywords: expression, long non-coding RNA, esophageal squamous cell carcinoma, next-generation sequencing, RNA-seq, bioinformatics analysis

INTRODUCTION

Esophageal cancer is the sixth most fatal malignancy worldwide with an overall survival rate ranging from 15 to 25% (Lagergren et al., 2017). The two major histologic types of esophageal cancer are ESCC and esophageal adenocarcinoma (EAC), and they have distinct genetic profiles (Cancer Genome Atlas Research Network, 2017). In China, >90% of cases of esophageal cancer are ESCC (Zeng et al., 2015). A deeper understanding of the transcriptional dysregulation of ESCC is critical for predicting the prognosis, providing appropriate treatment, and improving clinical outcomes (Peng et al., 2012; Wang et al., 2012; Zhang et al., 2013; Chen et al., 2016; Huang et al., 2017).

The advent of next-generation sequencing has enabled studies on the transcriptional features of ESCC and identification of potential target genes (Tong et al., 2012; Wang et al., 2018). In Tong et al. (2012) used transcriptome data from seven ESCC samples and five non-tumor specimens to profile the transcriptional features of ESCC. Subsequently, Wang et al. (2018) depicted the "landscape" of lncRNAs and messenger (m)RNAs in ESCC using RNA sequencing (RNA-seq) data from seven pairs of tumor samples and matched normal tissues. However, RNA-seq results from different studies are often inconsistent owing to sample heterogeneity. Furthermore, the small sample sizes of those studies limits the reproducibility and reliability of their results.

In the present study, we use RNA-seq to investigate the profiles of transcriptional features of three pairs of ESCC tissues and paired normal mucosal tissues from Xiangya Hospital within Central South University (Changsha, China). Furthermore, we integrate all the public RNA-seq data from the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA), including the GSE111011, GSE32424, and TCGA_ESCC data sets, to identify potential pathogenic genes in ESCC. A microarray data set for ESCC (GSE53625) and TCGA data set for head and neck squamous cell carcinoma (HNSCC) (TCGA_HNSCC) were used to explore the prognostic value of these "hub" genes in discovery data sets. Expression of hub genes with potential prognostic value was confirmed further by real-time reverse RT-qPCR. Further *in vitro* studies were undertaken to explore the biological function and underlying mechanism of *LINC01614*, expression of which was upregulated in ESCC and which is considered to be a potential prognostic marker.

MATERIALS AND METHODS

Ethical Approval of the Study Protocol

Ethical approval for the collection and use of all tissues was obtained from the ethics committee of the Xiangya Hospital of Central South University. Written informed consent was obtained from each patient to use his/her material.

Patients and RNA-seq Data

Three samples of ESCC tumor tissue and paired normal mucosa tissues for RNA-seq were collected from patients who had undergone esophagectomy without neoadjuvant chemotherapy or radiotherapy at Xiangya Hospital. Specimens were taken from the center of the tumor. Paired normal tissues were taken from surgically dissected tissues ∼5 cm away from the tumor. These three pairs of tissues were snap-frozen in liquid nitrogen after surgery and before RNA extraction. The process used for RNA-seq is described in **Supplementary File 1**. Another 65 ESCC tumor specimens and 20 non-cancerous specimens were obtained from Xiangya Hospital for use in RT-qPCR.

Another two RNA-seq data sets, GSE111011 and GSE32424, obtained with the Illumina HiSeq 2500 and Illumina Genome Analyzer IIx platforms, respectively, were downloaded from the National Center for Biotechnology Information Sequence Read Archive[1] with the identifiers SRP133303 (Wang et al., 2018) and SRP008496 (Tong et al., 2012). GSE111011 contained seven normal samples and seven tumor samples. GSE32424 contained five normal samples and seven tumor samples. More information about these datasets are shown in **Table 1**.

Data Processing

The Xiangya Hospital data set, GSE111011, and GSE32424 were analyzed using a particular workflow. Briefly, clean reads were obtained from raw reads by removing adaptor sequences, reads with >5% ambiguous bases, and low-quality reads, and, they were then mapped and aligned to the human genome (GRCH38) using HISAT2 (Kim et al., 2015). RNA-seq data from TCGA_ESCC were downloaded from Firehose[2]. The GSE53625 data set (which was based on GPL18109 and contained 179 ESCC and 179 paired normal control samples) was downloaded from GEO. GSE32424 contained five normal samples and seven tumor samples. More information about these datasets are shown in **Table 1**. The RNA-seq data sets (Xiangya Hospital, GSE111011, GSE32424, and TCGA ESCC) were defined as "discovery data sets." The "DEseq2" R package (R Project for Statistical Computing, Vienna, Austria) was used to screen out differentially expressed genes (DEGs) between ESCC and non-cancerous controls in all data sets. Then, we screened out differentially expressed coding genes (DECGs) and differentially expressed lncRNA genes (DELGs) using the criteria of $|\log2(\text{fold change})| > 1$ and false discovery rate (FDR) <0.01. Subsequently, the DECGs and DELGs were used to draw volcano plots using R.

To investigate the molecular function of DECGs, we used the "clusterProfiler" R package for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and for Gene Ontology (GO) analysis with respect to three domains: cellular component, biological process, and molecular function (Yu et al., 2012).

Construction of a Protein–Protein Interaction Network and Module Analysis

The STRING database[3] was used to construct a PPI network of DECGs and to investigate the relationships among them (Szklarczyk et al., 2015) with a medium confidence of 0.400. Cytoscape was employed to visualize the PPI network. The MCODE (Bandettini et al., 2012) Cytoscape plugin was used to identify highly interacted nodes in the subnetworks. The

[1]https://www.ncbi.nlm.nih.gov/sra

[2]https://gdac.broadinstitute.org

[3]https://string-db.org

following parameters were set to their default values: maximum depth = 100, degree = 2, node score = 0.2, and k-core = 2.

Survival Analysis and Validation of Hub Genes

To explore the prognostic value of DELGs and DECGs from the PPI network, log-rank survival analysis was carried out by the Kaplan–Meier method in the GSE53625 and TCGA_HNSCC data sets. X-Tile (Camp et al., 2004) was used to determine optimal cutoff points for the log-rank test.

Reverse transcription-quantitative polymerase chain reaction was undertaken using FastStart Universal SYBR Green Master (ROX) (Roche, Basel, Switzerland). Each sample was standardized by β-actin as an internal control gene. RT-qPCR parameters were 95°C for 10 min (holding stage), 40 cycles of 95°C for 15 s, and 60°C for 1 min (PCR stage), and then 95°C for 15 s and 75°C for 1 min (melt-curve stage). Data were analyzed by the comparative cycle threshold ($2^{-\Delta\Delta Ct}$) method.

The sense and antisense primer sequences encoding BGN, SPP1, LINC01614, INC01415, NKILA, and β-actin mRNA were as in **Supplementary Table 1**.

Gene Set Enrichment Analysis

We wished to further explore the different biological pathways in patients with high and low expression of target genes. GSEA was done using R 3.6.2[4] employing TCGA_ESCC data. The annotated gene set of "H: Hallmark gene sets" downloaded from the MSigDB was used in the analysis. $P < 0.05$ was considered to denote significant enrichment.

Cell Lines and Culture Conditions

Five ESCC cell lines (KYSE30, KYSE150, KYSE410, Eca109, and TE-1) were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). All cell lines were cultured at 37°C with 5% CO_2 in RPMI 1640 medium (Gibco, Carlsbad, CA, United States) with 10% fetal bovine serum (FBS; Gibco).

Transfection of siRNAs

Two small interfering RNAs (siRNA-1, siRNA-2) against *LINC01614* and a scrambled control siRNA (siRNA-NC) were purchased from Suzhou Genepharma Co., Ltd. The sequences of the siRNAs are given in **Supplementary Table 2**. The knockdown efficiency was tested by RT-qPCR 48–72 h after transfection.

Transwell Assays

Transwell assays were carried out in Transwell chambers (pore size, 8 μm; Costar, Washington, DC, United States) according to the manufacturer's instructions. In brief, cells were harvested after treatment with siRNA for 24 h or after no treatment. Then, 200 μL of serum-free medium (containing 5×10^4 cells) was placed in the upper chamber of each insert, and 800 μL of RPMI 1640 medium with 20% FBS was placed in the lower chamber. After incubation for 48 h at 37°C, the remaining tumor cells inside the upper chamber were removed with cotton swabs before fixation and staining.

[4]http://www.r-project.org/

Western Blotting

Cells were lysed using total protein extraction buffer (Beyotime Biotechnology, Shanghai, China). Equal amounts of lysate samples were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and then immunoblotted with primary antibodies and the corresponding horseradish peroxidase-labeled secondary antibodies. The procedure is described in more detail in **Supplementary File 1**.

Statistical Analysis

Bioinformatics analysis was carried out using R 3.6.2. The results of the cytology experiments were analyzed using Prism 6 (GraphPad, San Diego, CA, United States). Student's *t* test was used to compare two independent continuous variables. The unpaired *t* test was used to detect clinical samples of ESCC. The log-rank test was employed for survival analyses. $P < 0.05$ was regarded as significant.

RESULTS

Identification of DEGs

A flowchart of the processing and analyses of data undertaken in our study is shown in **Figure 1**. A total of 99 ESCC tissues and 27 normal tissues from different resources (Xiangya Hospital, GSE111011, GSE32424, and TCGA_ESCC) were analyzed by RNA-seq.

There were 816 DECGs and DELGs (377 upregulated and 439 downregulated) in the samples from Xiangya Hospital, 4754 DECGs, and DELGs (2569 upregulated and 2185 downregulated) in the GSE111011 data set, 4814 DECGs, and DELGs (2671 upregulated and 2143 downregulated) in the GSE32424 data set, and 7322 DECGs and DELGs (3347 upregulated and 3795 downregulated) in the TCGA data set. These are illustrated in the volcano plots in **Figures 2A–D**. As shown in the Venn diagrams in **Figures 2E,F**, 148 genes (106 upregulated and 42 downregulated) were consistently differentially expressed in all four databases. These commonly expressed genes are listed in **Supplementary Table 3**, including four upregulated genes encoding lncRNAs: *LINC01614*, *LINC01415*, *NKILA*, and *HMGA2-AS1*.

Functional Analysis of DEGs

Analysis of GO annotations showed that the upregulated DECGs were enriched in extracellular structure organization (ontology: biological process), collagen-containing extracellular matrix (ontology: cellular component), and extracellular matrix structural constituent (ontology: molecular function) (**Figures 3A–C**). In analysis of the KEGG pathway, the upregulated genes were enriched significantly in proteoglycans and microRNAs in cancer (**Figure 3D**). Moreover, the downregulated genes were enriched in the fatty acid metabolic process (ontology: biological process), actin cytoskeleton (ontology: cellular component), and arachidonic acid monooxygenase activity (ontology: molecular function) (**Figures 3E–G**). According to analysis of the

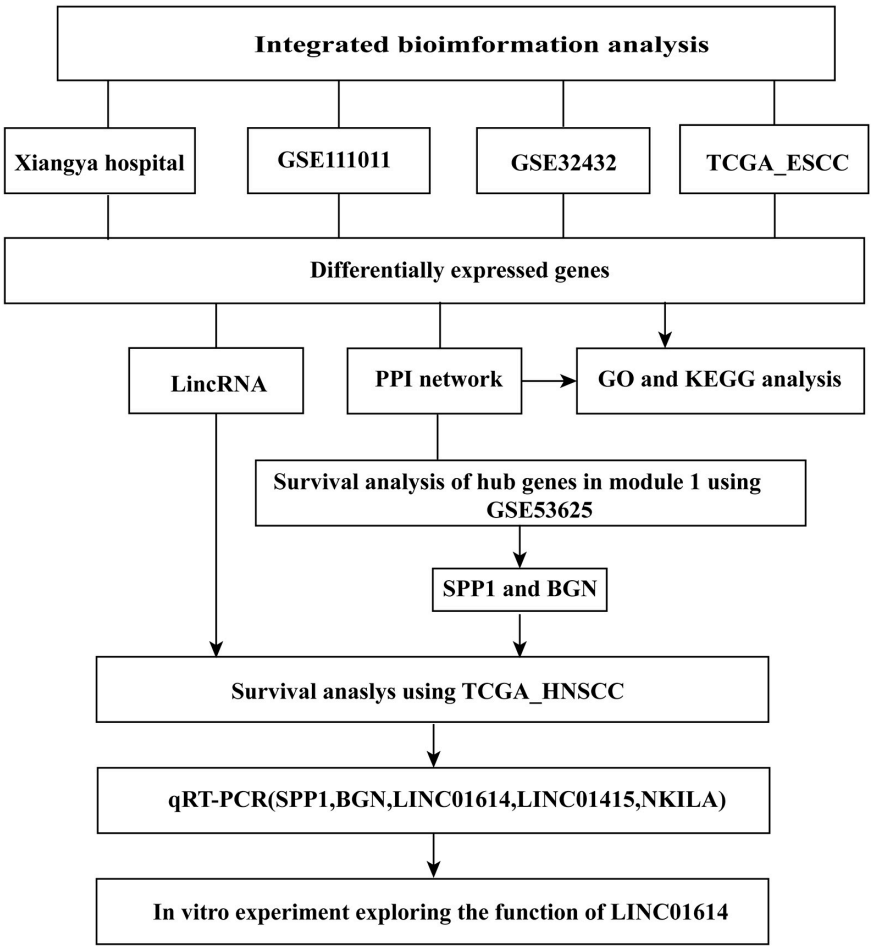


FIGURE 1 | Flowchart showing the processing and analysis of data in this study.

KEGG pathway, the downregulated genes were enriched significantly in arachidonic acid metabolism and chemical carcinogenesis (**Figure 3H**).

Construction of a Protein–Protein Interaction Network, Module Analysis, and Selection of Hub Genes

A PPI network containing 71 nodes and 172 edges was constructed by uploading 127 protein-coding genes to the STRING online database and visualized using Cytoscape (**Figure 4A**). In **Figure 4**, rectangles represent genes that were highly expressed in ESCC tissues, whereas diamonds represent genes with low expression in ESCC tissues. Subsequently, two modules were extracted using the MCODE Cytoscape plugin. Module 1 consists of 10 nodes and 29 edges (**Figure 4B**), and module 2 comprises nine nodes and 31 edges (**Figure 4C**). According to analysis of the KEGG pathway, module 1 was mainly enriched in extracellular matrix–receptor interaction and focal adhesion (**Figure 4D**), whereas module 2 was enriched primarily in DNA replication (**Figure 4E**).

Validation of Hub Genes in Modules 1 and 2

All the hub genes in the two modules were evaluated using the GSE53625 and TCGA_HNSCC data sets. Results show that the hub genes had high expression in tumor samples in both data sets which were consistent with the results from the discovery data set (**Supplementary Figure 1**).

Survival Analysis of Hub Genes in Module 1 and Differentially Expressed lncRNAs

The prognostic value of the 10 hub genes in module 1 and the four upregulated lncRNAs was determined using the log-rank test. The best cutoff value for BGN and SPP1 in GSE53625 was 14.8 and 15.9 RPKM, respectively. The best cutoff value (in RPKM) for BGN, SPP1, LINC01614, LINC01415, and NKILA in TCGA_HNSCC was 14.1, 9.72, 3.90, 5.9, and 5.1, respectively. The results obtained using GSE53625 and TCGA_HNSCC show that higher expression of *SPP1* and *BGN* was related to worse overall survival in ESCC (**Figures 5A–D**) as was higher

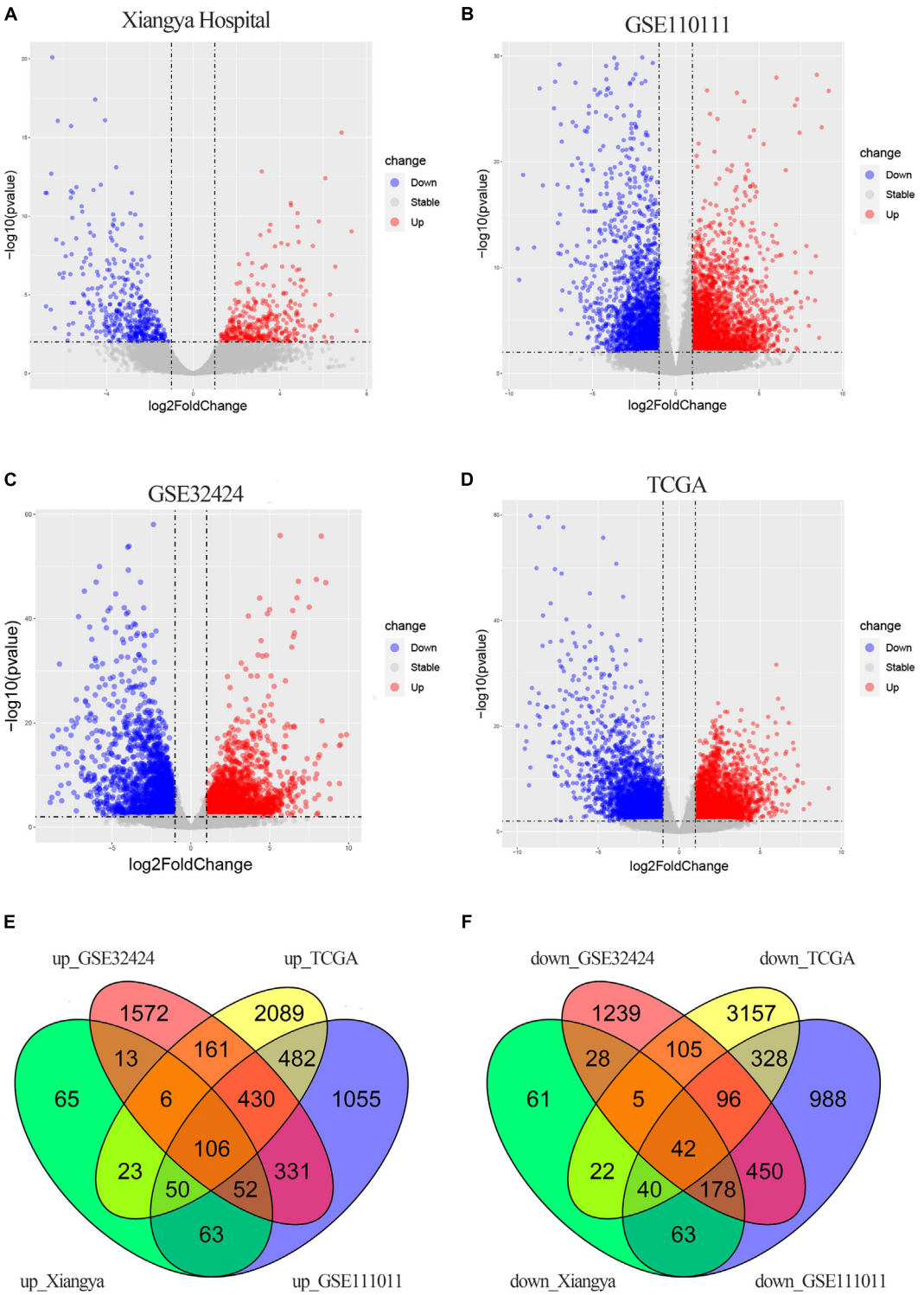


FIGURE 2 | DEGs in ESCC. **(A–D)** Volcano plots showing DEGs in the discovery data sets: Xiangya Hospital **(A)**, GSE111011 **(B)**, GSE32424 **(C)**, and TCGA_ESCC **(D)**. Dark circles represent genes without significant differential expression (FDR > 0.01), red circles represent upregulated mRNAs with significant differential expression [log2(fold change) >1 and FDR < 0.01], and blue circles represent downregulated mRNAs with significant differential expression [log2(fold change) <1 and FDR < 0.01]. The top 20 upregulated and downregulated genes are listed. **(E,F)** Venn diagrams of the overlapping DEGs, including 106 upregulated **(E)** and 42 downregulated **(F)** genes, from the four data sets.

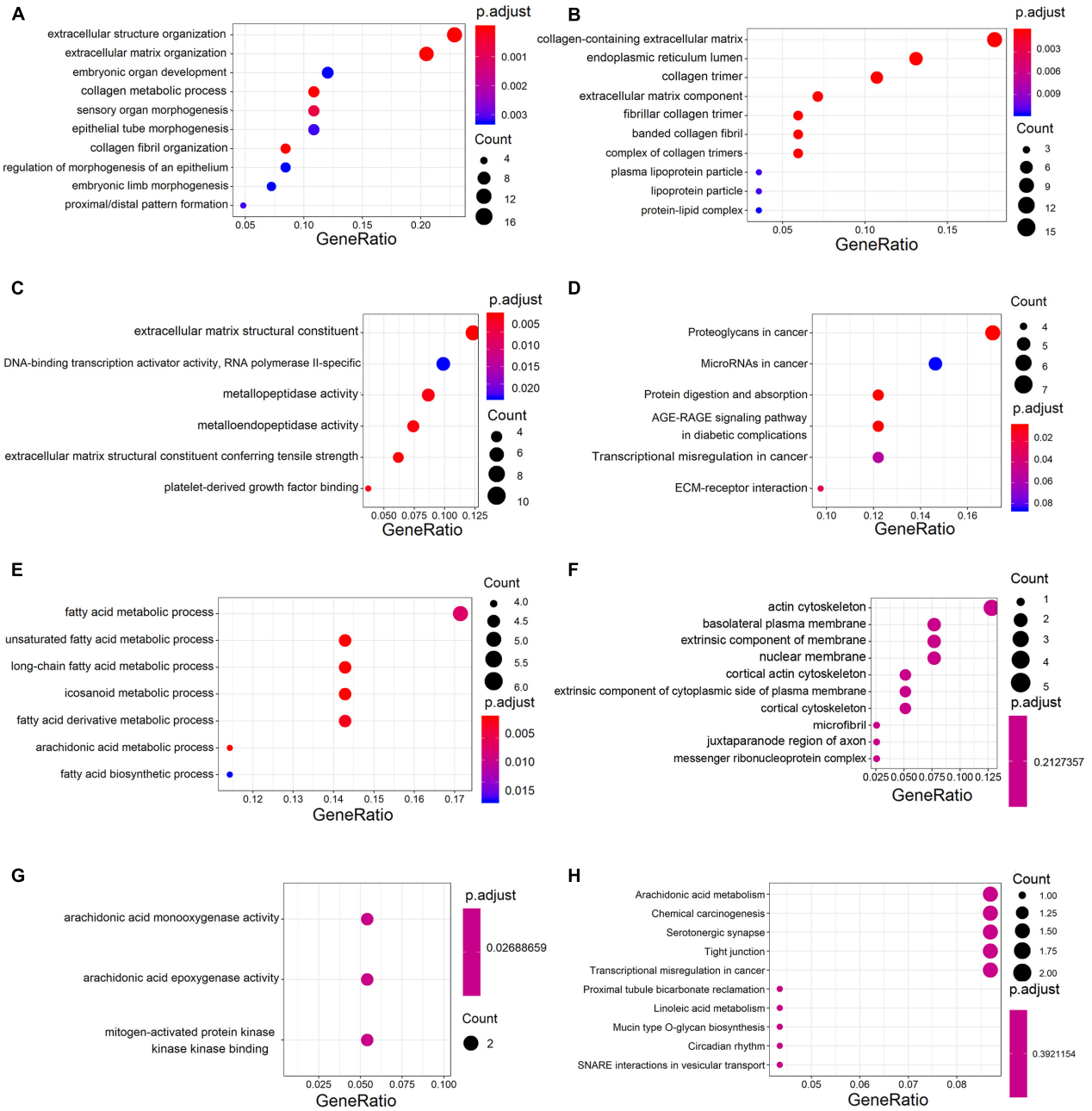


FIGURE 3 | Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of significant DECGs. **(A–C)** Analysis of GO annotations of upregulated DECGs with respect to three domains: biological process **(A)**, cellular component **(B)**, and molecular function **(C)**. **(D)** KEGG pathway analysis of upregulated DECGs. **(E–G)** Analysis of GO annotations of downregulated DECGs with respect to three domains: biological process **(E)**, cellular component **(F)**, and molecular function **(G)**. **(H)** KEGG pathway analysis of downregulated DECGs. The size of a dot represents the number of genes enriched for each GO term and KEGG pathway; colors from red to blue represent the adjusted *P*-value.

expression of three differentially expressed lncRNAs (*LINC01614*, *LINC01415*, and *NKILA*) (**Figures 5E–G**).

Validation of Hub Genes in ESCC Tissue

To further investigate expression of hub genes with prognostic value in ESCC, we undertook RT-qPCR screening of the expression of *BGN*, *SPP1*, *LINC01415*, *LINC01614*, and *NKILA* in 65 ESCC specimens and 20 non-cancerous specimens of esophageal tissue. Expression of *BGN*, *SPP1*, *LINC01415*, and *LINC01614* was much higher in tumor samples than in normal esophageal tissues (**Figures 6A–D**). However, there was no difference in expression of *NKILA* between tumor samples and normal samples (**Figure 6E**).

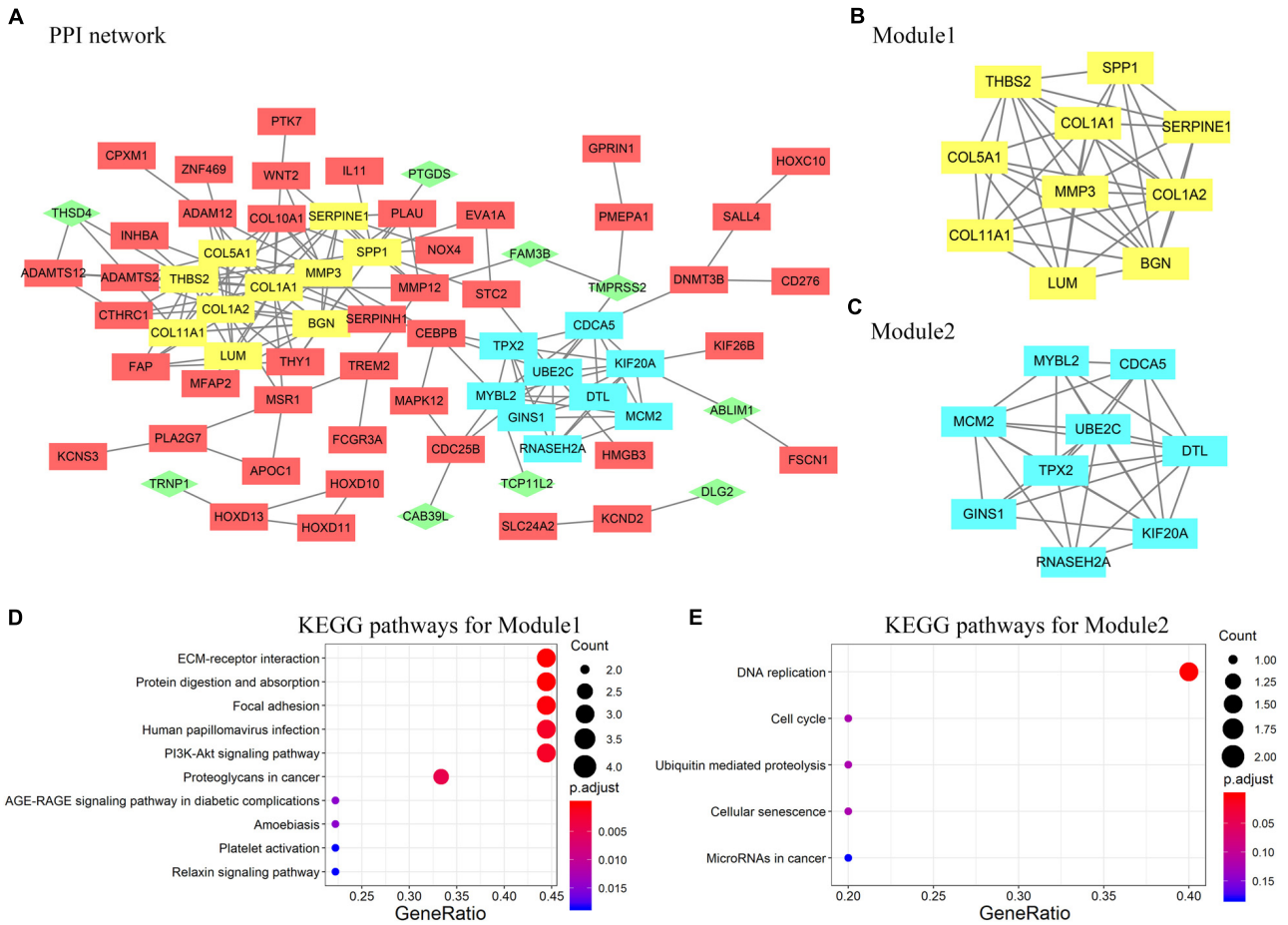


FIGURE 4 | (A) Protein–protein interaction (PPI) network of DEGs constructed in STRING and visualized by Cytoscape. Rectangles represent genes highly expressed in ESCC tissues; diamonds and green color represent genes with low expression in ESCC tissues. **(B,C)** Two modules selected by modular analysis with the MCODE Cytoscape plugin (yellow, module 1; blue, module 2). **(D)** KEGG pathway analysis showing enrichment of module 1 genes in ECM–receptor interaction. **(E)** KEGG pathway analysis showing enrichment of module 2 genes in DNA replication. Nodes that did not interact with other nodes are hidden.

Knockdown of *LINC01614* Expression Inhibits Metastasis of Esophageal Squamous Cell Carcinoma via Regulation of EMT

Gene set enrichment analysis revealed that the EMT gene set was positively correlated with LINC01614 expression (NES 3.894, P = 0.006; **Supplementary Figure 2**). First, we screened *LINC01614* expression in five ESCC cell lines (KYSE150, KYSE410, KYSE30, Eca109, and TE-1) using RT-qPCR (**Figure 7A**). Two cell lines (Eca109 and KYSE410) were selected for subsequent experiments because they showed relatively high expression. Eca109 and KYSE410 cells were transfected with two siRNAs against *LINC01614*: si-LINC01614#1 and si-LINC01614#2, respectively. Scrambled siRNA-transfected cells were used as negative controls. si-LINC01614#1 and si-LINC01614#2 showed significant knockdown efficiency (**Figures 7B,C**). The results of the Transwell assay showed that the migration rate of *LINC01614* knockdown cells was less than that of control cells (P < 0.001, **Figures 7D,E**). Western blotting (WB) showed that knockdown of *LINC01614* expression reduced the expression of N-cadherin and *ZEB1* in Eca109 and KYSE410 cells (**Figures 7F,G**). We also explored the role of LINC01415 in ESCC: knockdown of *LINC01415* expression reduced the migration of ESCC cells without affecting EMT-related markers (**Supplementary Figure 3**).

DISCUSSION

Exploring the potential mechanisms underlying ESCC development would be of considerable benefit for prognosis prediction. In this study, 106 upregulated and 42 downregulated genes were identified in discovery data sets, including four genes encoding lncRNAs whose functions were evaluated by *in vitro* studies. The MCODE plugin of Cytoscape was used to screen out two significant modules from the PPI network of 127 protein-coding genes.

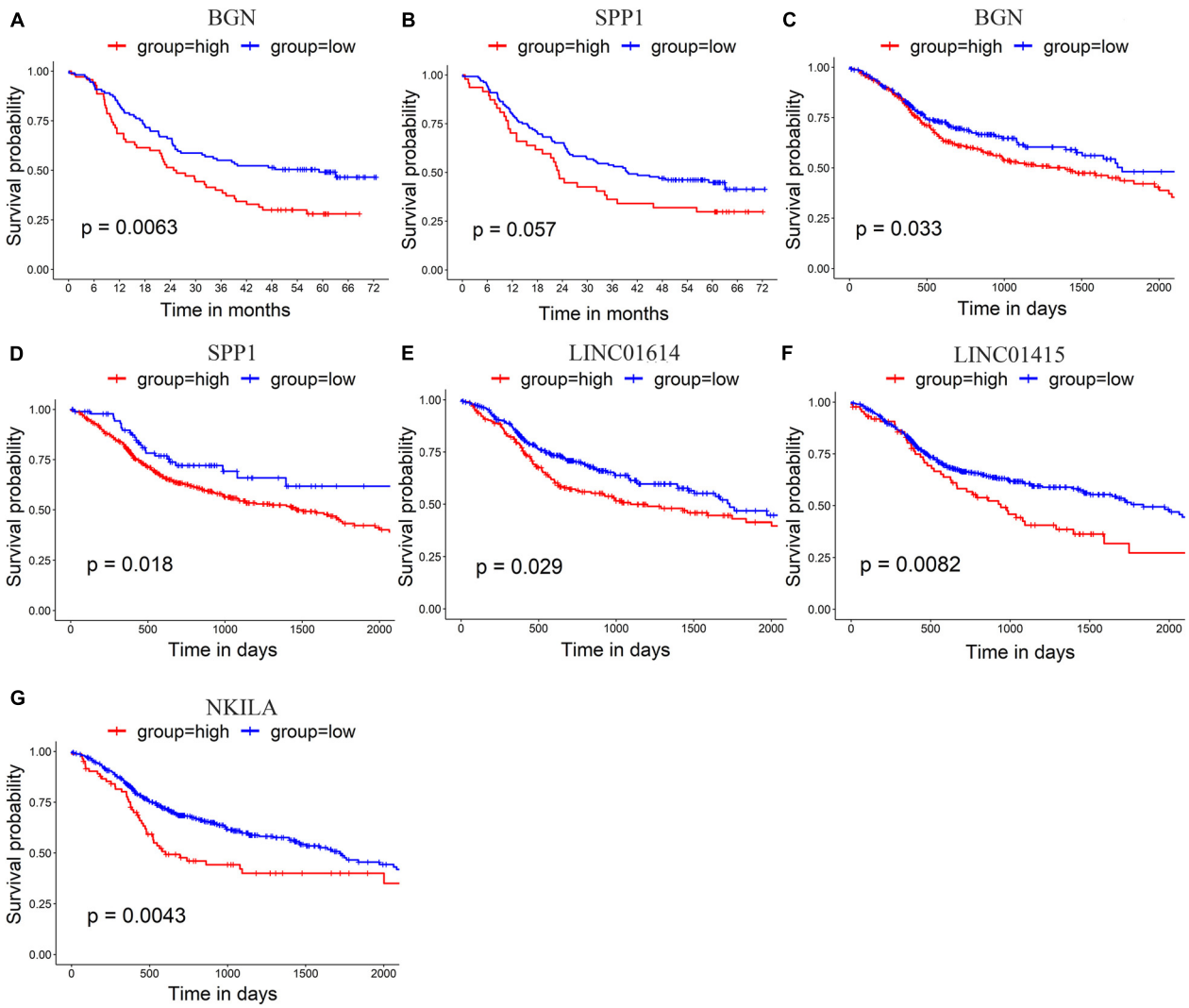


FIGURE 5 | (A,B) Survival analysis of GSE53625 by log-rank test shows that higher expression of BGN and SPP1 is related to poor survival. **(C–G)** Survival analysis of TCGA _HNSCC by log-rank test shows that higher expression of BGN **(C)**, SPP1 **(D)**, LINC01614 **(E)**, LINC01415 **(F)**, and NKILA **(G)** is related to poor survival in HNSCC.

TABLE 1 | Characteristics of the data sets used in this study.

Dataset	Platform	Sample size		Tumor type	Purpose
		Normal	Tumor		
Xiangya	Illumina HiSeq 3000	3	3	ESCC	Discovery
GSE111011	Illumina HiSeq 2500	7	7	ESCC	Discovery
GSE32424	Illumina Genome Analyzer IIx	5	7	ESCC	Discovery
TCGA_ESCC	Illumina HiSeq	11	82	ESCC	Discovery
GSE53625	GPL18109 (microarray)	179	179	ESCC	Validation of DECGs Survival analysis
TCGA_HNSCC	Illumina HiSeq	44	502	HNSCC	Validation of survival analysis

The microarray data set GSE53625 was employed to explore the prognostic value of hub genes. Of the 10 hub protein-coding genes in module 1 of the PPI, high expression of *SPP1* and *BGN* indicated a poor prognosis. BGN and SPP1 are essential components of the extracellular matrix, which has a critical role during the migration and progression of tumor cells (Rangaswami et al., 2006; Hu et al., 2016). Extensive studies have elucidated the crucial role of *BGN* in regulating the progression

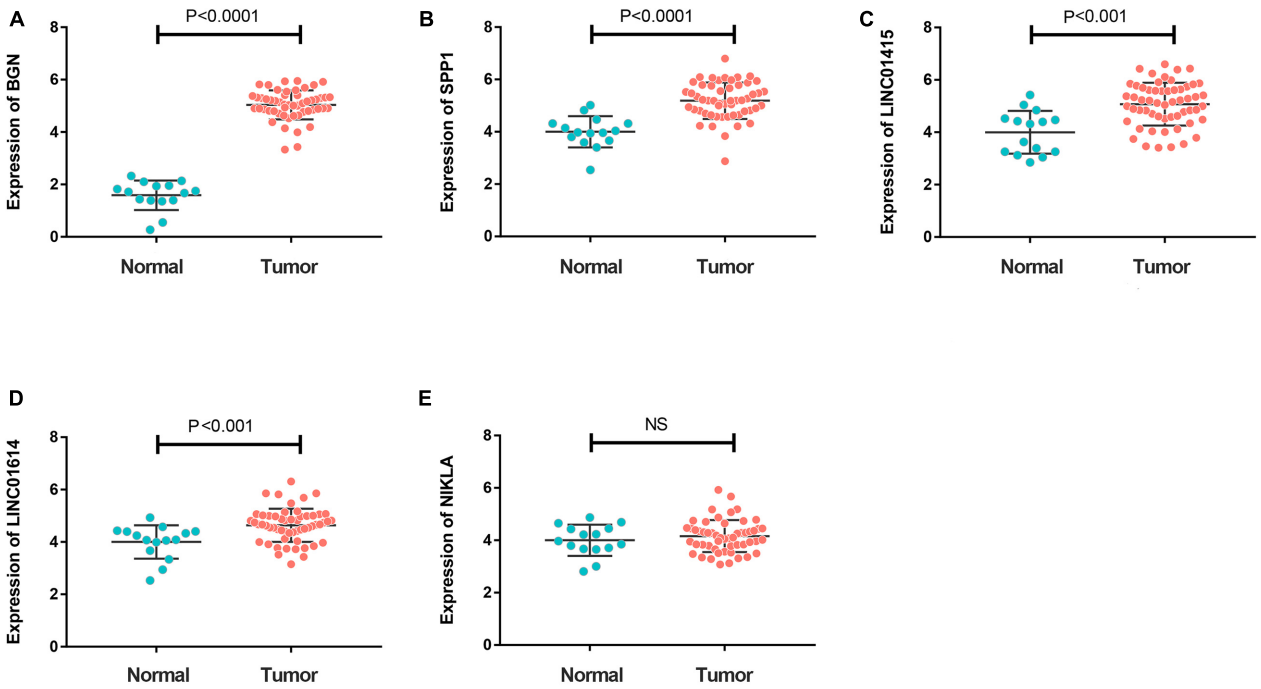


FIGURE 6 | (A–D) Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) showing much higher expression of *BGN* **(A)**, *SPP1* **(B)**, *LINC01415* **(C)**, and *LINC01614* **(D)** in ESCC tumor tissues compared with that in normal tissues ($P < 0.05$). **(E)** A similar expression of *NKILA* was observed in tumor and normal samples. The mean gene expression in normal tissue was used as a reference to calculate the value of $2^{-\Delta\Delta Ct}$. NC, negative control.

and metastasis of various malignancies, including prostate (Jacobsen et al., 2017), gastric (Hu et al., 2016), endometrial (Sun et al., 2016), and colon cancers (Jacobsen et al., 2017; Liu B. et al., 2018). Similarly, osteopontin (which is encoded by *SPP1*) is correlated significantly with tumor metastasis and a poor prognosis in cancers (Briones-Orta et al., 2017; Cabiati et al., 2017; Xu et al., 2017; Li et al., 2018; Zhao et al., 2019). However, the prognostic value of the four upregulated lncRNAs could not be determined owing to the deficiency of lncRNA probes in the GSE53625 microarray platform.

The TCGA_ESCC database had limited testing power because it contains data for 82 patients, only 31 of whom reached the endpoint in the follow-up. Moreover, ESCC is distinct from EAC in its genetic and epigenetic features (Cancer Genome Atlas Research Network, 2017). Therefore, we sought another approach for our ESCC research.

Moreover, a previous study suggests that the unmatched norms from healthy individuals are different from paired normal tissue, which is obtained from patients (Buzdin et al., 2018). In this study, we initially screened the DEGs between tumor tissue and paired normal tissues in the TCGA data set. Then, we also profiled the DEGs between the TCGA and GTEx databases; the latter has gene expression data obtained from healthy individuals (**Supplementary Figure 4A**). Results show that, when compared to normal tissue, the number of DEGs is markedly higher (**Supplementary Figure 4B**). However, there are still 3521 overlapped genes. Further, we detected the overlapped genes with these DEGs, which revealed that there were 110 genes that shared the common expression pattern. Notably, the five genes (LINC01415, LINC01614, NKILA, SPP1, and BGN) we selected for analysis were within the 110 genes.

According to previous studies, ESCC and HNSCC can be considered almost a single disease entity with similar molecular characteristics according to multiplatform data, including data on somatic copy number alterations, DNA methylation, and transcription (Cancer Genome Atlas Research Network, 2017). They also have the same histology type (i.e., squamous cell carcinoma) and field cancerization (i.e., upper gastrointestinal tract) and have common risk factors, including use of tobacco and alcohol (Onochi et al., 2019). Furthermore, similar expression of hub genes to that found in ESCC was detected in the TCGA_HNSCC database. Thus, we believe that combined analysis of the data for ESCC and HNSCC is a promising approach for exploring ESCC features. Furthermore, studies using TCGA_HNSCC have shown the same results as those using GSE53625 (i.e., higher expression of *SPP1* and *BGN* indicate poor overall survival in both data sets). Besides this, overexpression of *LINC01614*, *LINC01415*, and *NKILA* was related to a poor prognosis in HNSCC and ESCC.

To verify the reliability of these bioinformatics evaluations, we carried out RT-qPCR to assess expression of hub genes (*BGN*, *SPP1*, *LINC01614*, and *LINC01415*) in clinical tumor samples. These genes had much higher expression in tumor tissues than that in normal tissues data that were consistent with the bioinformatics results.

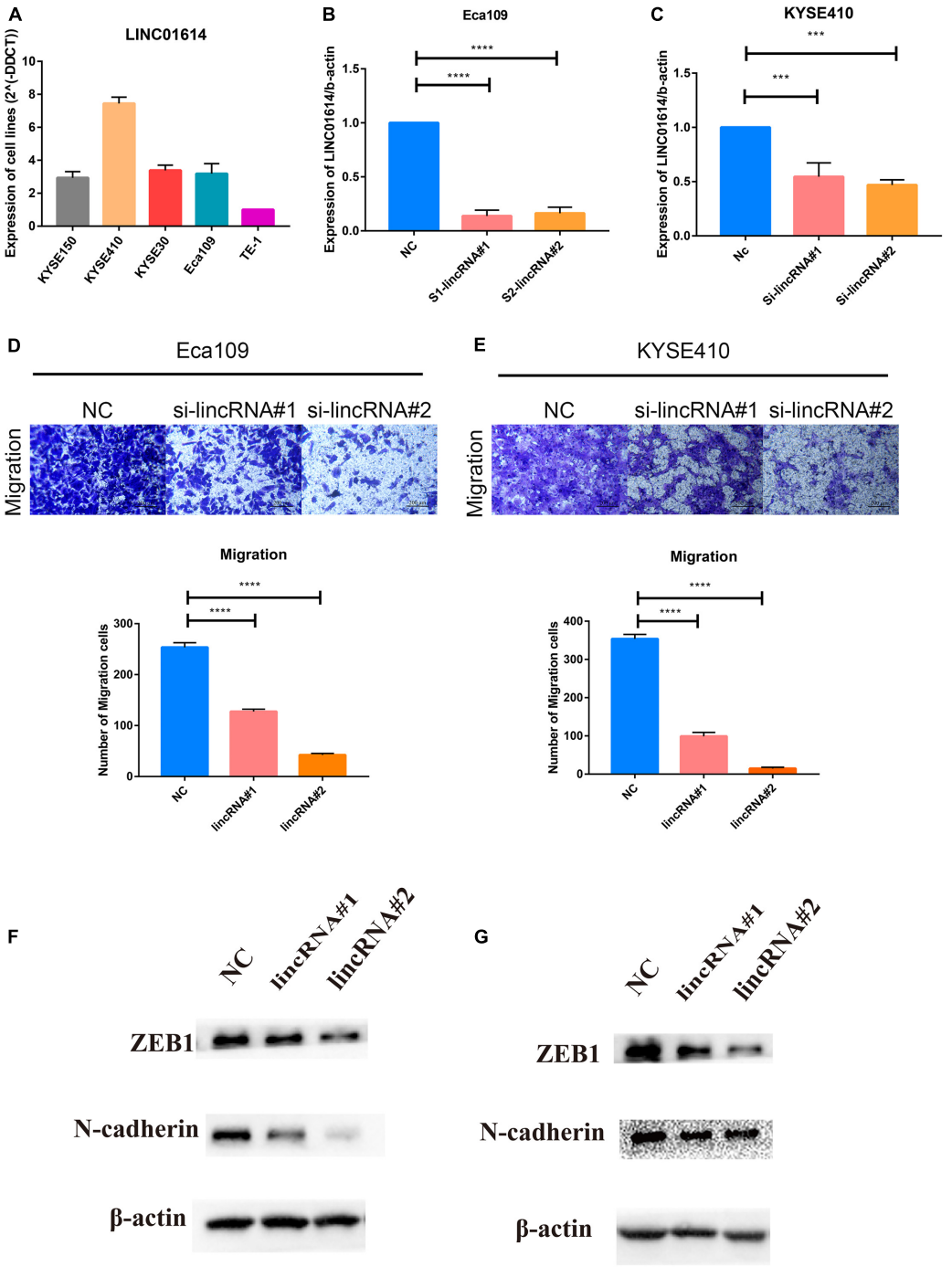


FIGURE 7 | (A) Two cell lines (Eca109 and KYSE410) were selected for subsequent experiments due to their relatively high expression among the five candidate ESCC cell lines (KYSE150, KYSE410, KYSE30, Eca109, and TE-1). **(B,C)** si-LINC01614#1 and si-LINC01614#2 show significant knockdown efficiency. **(D,E)** Downregulation of LINC01614 expression inhibited the migration ability of ESCC cell lines (Eca109 and KYSE410). **(F,G)** Knockdown of LINC01614 expression reduced expression of N-cadherin and ZEB1 in Eca109 and KYSE410 cells. Data are the mean ± SD from three independent experiments. ***$P < 0.001$; ****$P < 0.0001$. NC, negative control.

NKILA is a nuclear factor kappa light-chain enhancer of activated B cell (NF-κB) interacting lncRNA. It has been shown to function as a tumor suppressor by inhibiting the NF-κB pathway in various cancers, including breast cancer (Liu et al., 2015; Wu et al., 2018), melanoma (Bian et al., 2017), non-small cell lung cancer (Lu et al., 2017), nasopharyngeal carcinoma

(Zhang et al., 2019), laryngeal cancer (Yang et al., 2018), and ESCC (Lu et al., 2018).

Recent studies show that *LINC01614* has high expression in tumor tissues and is associated with a poor prognosis in breast cancer (Wang et al., 2019) and non-small cell lung cancer (Sun and Ling, 2019). *LINC01614* also promotes the carcinogenesis of lung adenocarcinoma by downregulating expression of microRNA-217 and upregulating expression of Forkhead Box Protein P1 (FOXP1) (Liu A. N. et al., 2018). However, the role of *LINC01614* in ESCC is not known. We show that inhibition of *LINC01614* expression (i) by siRNA could significantly suppress migration of ESCC cells and (ii) in ESCC cell lines could decrease expression of EMT markers, including N-cadherin and ZEB1. In summary, *LINC01614* is a novel oncogene in ESCC with a critical role in the metastasis of ESCC cells.

Overall, by integrated bioinformatic analysis of transcriptome data, we identified two upregulated lncRNAs (*LINC01614* and *LINC01415*) and two hub protein-coding genes (*SPP1* and *BGN*) as potential pathogenic genes and prognostic markers in ESCC. Moreover, GSEA revealed that the EMT gene set was positively correlated with LINC01614 expression. *In vitro* experiments revealed that knockdown of *LINC01614* expression suppressed the migration ability of ESCC cell lines via EMT regulation.

However, our study is limited by an insufficient number of samples and loss of patients to follow-up. Besides this, the regulatory mechanism of *LINC01614* in ESCC is not known. Consequently, further clinical data and additional basic research are required to explore the role of *LINC01614* in ESCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/sra/PRJNA594797, https://www.ncbi.nlm.nih.gov/sra/SRP133303, https://www.ncbi.nlm.nih.gov/sra/SRP008496, https://gdac.broadinstitute.org, and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53625.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Xiangya Hospital of Central South University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GW and WZ contributed to conception and design of the study. YZ organized the database. XP performed the statistical analysis. YC and LT wrote the first draft of the manuscript. HJ wrote the sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

ACKNOWLEDGMENTS

We thank Guangzhou RiboBio Co., Ltd., for assistance with sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.521004/full#supplementary-material

Supplementary Figure 1 | All the hub genes in modules 1 and 2 show high expression in tumor samples in GSE53625 **(A,B)** and TCGA_HNSCC **(C,D)**. These data are consistent with the results in the discovery data set. $^{****}P < 0.001$.

Supplementary Figure 2 | Enrichment plots of GSEA correlation analyses for LINC01614 with EMT-associated gene sets using TCGA_ESCC data.

Supplementary Figure 3 | (A) Two cell lines (Eca109 and KYSE30) were selected for subsequent experiments owing to their relatively high expression among the five candidate ESCC cell lines (KYSE150, KYSE410, KYSE30, Eca109, and TE-1). **(B,C)** si-LINC01415#1 and si-LINC01415#2 show significant knockdown efficiency. **(D,E)** Downregulation of LINC01415 expression inhibited the migration ability of esophageal squamous cell lines (Eca109 and KYSE30). **(F,G)** Further investigation showed that knockdown of LINC01415 expression could reduce expression of N-cadherin and ZEB1 in both cell lines (Eca109 and KYSE30). Data are the mean ± SD from three independent experiments. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; NC, negative control.

Supplementary Figure 4 | (A) Volcano plots of TCGA and GTEx databases. **(B,C)** Venn diagrams of the overlapping DEGs.

REFERENCES

Bandettini, W. P., Kellman, P., Mancini, C., Booker, O. J., Vasu, S., Leung, S. W., et al. (2012). MultiContrast delayed enhancement (Mcode) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J. Cardiovasc. Magn. Reson.* 14:83.

Bian, D., Gao, C., Bao, K., and Song, G. (2017). The long non-coding RNA NKILA inhibits the invasion-metastasis cascade of malignant melanoma via the regulation of NF-kB. *Am. J. Cancer Res.* 7, 28–40.

Briones-Orta, M. A., Avendano-Vazquez, S. E., Aparicio-Bautista, D. I., Coombes, J. D., Weber, G. F., and Syn, W. K. (2017). Osteopontin splice variants and polymorphisms in cancer progression and prognosis. *Biochim. Biophys. Acta Rev. Cancer* 1868, 93–108.A. doi: 10.1016/j.bbcan.2017.02.005

Buzdin, A., Sorokin, M., Garazha, A., Sekacheva, M., Kim, E., Zhukov, N., et al. (2018). Molecular pathway activation - New type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin. Cancer Biol.* 53, 110–124. doi: 10.1016/j.semcancer.2018.06.003

Cabiati, M., Gaggini, M., Cesare, M. M., Caselli, C., De Simone, P., Filipponi, F., et al. (2017). Osteopontin in hepatocellular carcinoma: a possible biomarker for diagnosis and follow-up. *Cytokine* 99, 59–65. doi: 10.1016/j.cyto.2017.07.004

Camp, R. L., Dolled-Filhart, M., and Rimm, D. L. (2004). X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin. Cancer Res.* 10, 7252–7259.

Cancer Genome Atlas Research Network (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature* 541, 169–175. doi: 10.1038/nature20805

Chen, C., Peng, H., Huang, X., Zhao, M., Li, Z., Yin, N., et al. (2016). Genome-wide profiling of DNA methylation and gene expression in esophageal squamous cell carcinoma. *Oncotarget* 7, 4507–4521. doi: 10.18632/oncotarget.6607

Hu, L., Zang, M. D., Wang, H. X., Li, J. F., Su, L. P., Yan, M., et al. (2016). Biglycan stimulates VEGF expression in endothelial cells by activating the TLR signaling pathway. *Mol. Oncol.* 10, 1473–1484. doi: 10.1016/j.molonc.2016.08.002

Huang, J., Wang, G., Tang, J., Zhuang, W., Wang, L. P., Liou, Y. L., et al. (2017). DNA methylation status of PAX1 and ZNF582 in esophageal squamous cell carcinoma. *Int. J. Environ. Res. Public Health* 14:216. doi: 10.3390/ijerph14020216

Jacobsen, F., Kraft, J., Schroeder, C., Hube-Magg, C., Kluth, M., Lang, D. S., et al. (2017). Up-regulation of Biglycan is Associated with Poor Prognosis and PTEN Deletion in Patients with Prostate Cancer. *Neoplasia* 19, 707–715. doi: 10.1016/j.neo.2017.06.003

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360.

Lagergren, J., Smyth, E., Cunningham, D., and Lagergren, P. (2017). Oesophageal cancer. *Lancet* 390, 2383–2396.

Li, S., Yang, R., Sun, X., Miao, S., Lu, T., Wang, Y., et al. (2018). Identification of SPP1 as a promising biomarker to predict clinical outcome of lung adenocarcinoma individuals. *Gene* 679, 398–404. doi: 10.1016/j.gene.2018.09.030

Liu, A. N., Qu, H. J., Yu, C. Y., and Sun, P. (2018). Knockdown of LINC01614 inhibits lung adenocarcinoma cell progression by up-regulating miR-217 and down-regulating FOXP1. *J. Cell Mol. Med.* 22, 4034–4044. doi: 10.1111/jcmm.13483

Liu, B., Xu, T., Xu, X., Cui, Y., and Xing, X. (2018). Biglycan promotes the chemotherapy resistance of colon cancer by activating NF-kappaB signal transduction. *Mol. Cell Biochem.* 449, 285–294. doi: 10.1007/s11010-018-3365-1

Liu, B., Sun, L., Liu, Q., Gong, C., Yao, Y., Lv, X., et al. (2015). A cytoplasmic NF-kappaB interacting long noncoding RNA blocks IkappaB phosphorylation and suppresses breast cancer metastasis. *Cancer Cell* 27, 370–381. doi: 10.1016/j.ccell.2015.02.004

Lu, Z., Chen, Z., Li, Y., Wang, J., Zhang, Z., Che, Y., et al. (2018). TGF-beta-induced NKILA inhibits ESCC cell migration and invasion through NF-kappaB/MMP14 signaling. *J. Mol. Med.* 96, 301–313. doi: 10.1007/s00109-018-1621-1

Lu, Z., Li, Y., Wang, J., Che, Y., Sun, S., Huang, J., et al. (2017). Long non-coding RNA NKILA inhibits migration and invasion of non-small cell lung cancer via NF-kappaB/Snail pathway. *J. Exp. Clin. Cancer Res.* 36:54. doi: 10.1186/s13046-017-0518-0

Onochi, K., Shiga, H., Takahashi, S., Watanabe, N., Fukuda, S., Ishioka, M., et al. (2019). Risk factors linking esophageal squamous cell carcinoma with head and neck cancer or gastric cancer. *J. Clin. Gastroenterol.* 53, e164–e170. doi: 10.1097/MCG.0000000000001019

Peng, H. H., Zhang, X., and Cao, P. G. (2012). MMP-1/PAR-1 signal transduction axis and its prognostic impact in esophageal squamous cell carcinoma. *Braz. J. Med. Biol. Res.* 45, 86–92. doi: 10.1590/s0100-879x2011007500152

Rangaswami, H., Bulbule, A., and Kundu, G. C. (2006). Osteopontin: role in cell signaling and cancer progression. *Trends Cell Biol.* 16, 79–87. doi: 10.1016/j.tcb.2005.12.005

Sun, H., Wang, X., Zhang, Y., Che, X., Liu, Z., Zhang, L., et al. (2016). Biglycan enhances the ability of migration and invasion in endometrial cancer. *Arch. Gynecol. Obstet.* 293, 429–438. doi: 10.1007/s00404-015-3844-5

Sun, Y., and Ling, C. (2019). Analysis of the long non-coding RNA LINC01614 in non-small cell lung cancer. *Medicine* 98:e16437. doi: 10.1097/md.0000000000016437

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003

Tong, M., Chan, K. W., Bao, J. Y., Wong, K. Y., Chen, J. N., Kwan, P. S., et al. (2012). Rab25 is a tumor suppressor gene with antiangiogenic and anti-invasive activities in esophageal squamous cell carcinoma. *Cancer Res.* 72, 6024–6035. doi: 10.1158/0008-5472.can-12-1269

Wang, B., Yin, B. L., He, B., Chen, C., Zhao, M., Zhang, W., et al. (2012). Overexpression of DNA damage-induced 45 alpha gene contributes to esophageal squamous cell cancer by promoter hypomethylation. *J. Exp. Clin. Cancer Res.* 31:11. doi: 10.1186/1756-9966-31-11

Wang, W., Wei, C., Li, P., Wang, L., Li, W., Chen, K., et al. (2018). Integrative analysis of mRNA and lncRNA profiles identified pathogenetic lncRNAs in esophageal squamous cell carcinoma. *Gene* 661, 169–175. doi: 10.1016/j.gene.2018.03.066

Wang, Y., Song, B., Zhu, L., and Zhang, X. (2019). Long non-coding RNA, LINC01614 as a potential biomarker for prognostic prediction in breast cancer. *PeerJ* 7:e7976. doi: 10.7717/peerj.7976

Wu, W., Chen, F., Cui, X., Yang, L., Chen, J., Zhao, J., et al. (2018). LncRNA NKILA suppresses TGF-beta-induced epithelial-mesenchymal transition by blocking NF-kappaB signaling in breast cancer. *Int. J. Cancer* 143, 2213–2224. doi: 10.1002/ijc.31605

Xu, C., Sun, L., Jiang, C., Zhou, H., Gu, L., Liu, Y., et al. (2017). SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. *Biomed. Pharmacother.* 91, 1167–1177. doi: 10.1016/j.biopha.2017.05.056

Yang, T., Li, S., Liu, J., Yin, D., Yang, X., and Tang, Q. (2018). lncRNA-NKILA/NF-kappaB feedback loop modulates laryngeal cancer cell proliferation, invasion, and radioresistance. *Cancer Med.* 7, 2048–2063. doi: 10.1002/cam4.1405

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.

Zeng, H., Zheng, R., Guo, Y., Zhang, S., Zou, X., Wang, N., et al. (2015). Cancer survival in China, 2003-2005: a population-based study. *Int. J. Cancer* 136, 1921–1930. doi: 10.1002/ijc.29227

Zhang, H., Chen, W., Duan, C. J., and Zhang, C. F. (2013).). Overexpression of HSPA2 is correlated with poor prognosis in esophageal squamous cell carcinoma. *World J. Surg. Oncol.* 11:141. doi: 10.1186/1477-7819-11-141

Zhang, W., Guo, Q., Liu, G., Zheng, F., Chen, J., Huang, D., et al. (2019). NKILA represses nasopharyngeal carcinoma carcinogenesis and metastasis by NF-kappaB pathway inhibition. *PLoS Genet* 15:e1008325. doi: 10.1371/journal.pgen.1008325

Zhao, L., Chi, W., Cao, H., Cui, W., Meng, W., Guo, W., et al. (2019). Screening and clinical significance of tumor markers in head and neck squamous cell carcinoma through bioinformatics analysis. *Mol. Med. Rep.* 19, 143–154. doi: 10.3892/mmr.2018.9639

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership