scientific reports

Check for updates

OPEN

An accurate DNA and RNA based targeted sequencing assay for clinical detection of gene fusions in solid tumors

Gang Ji[1,2,3], Qianlan Yao[1,2,3], Min Ren[1,2,3], Qianming Bai[1,2,3], Xiaoli Zhu[1,2,3] & Xiaoyan Zhou[1,2,3]✉

Gene fusions are one of the most important molecular biomarkers for tumor diagnosis, classification and targeted therapy. How to accurately detect them is a key issue in clinical work. In this study, a custom-designed integration of DNA and RNA-based next generation sequencing (NGS) assay including 16 targeted therapy related genes was developed and validated to identify gene fusions in solid tumors. This assay accurately identified all 10 different types of fusion in 8 commercial fusion spiked-in reference standards and 29 fusions including 16 different fusion forms in 60 clinical solid tumor samples previously identified by clinical testing methods. In addition, a *TPM3*::*NTRK1* fusion was additionally identified and validated by Sanger sequencing, which showed a false-negative result for the previous result. Mutational abundance limit of detection for the assay was assessed with a series of dilution experiments. These fusions can be stably detected when the mutational abundance is down to 5% for DNA and 250–400 copies/100 ng for RNA. The intra-assay and inter-assay reproducibility was observed in three samples and three replicates. This integration of DNA and RNA-based NGS assay shows excellent performance on formalin-fixed, paraffin-embedded samples, results at different levels can complement each other, thereby facilitating precise diagnosis and treatment.

Keywords Gene fusion, DNA and RNA based, Next generation sequencing, Targeted therapy, Hospital developed test

Gene fusions are derived from genomic rearrangements, including chromosomal translocation, duplication, deletion, inversion, or altered transcription. They have been described in approximately one-third of soft tissue tumors and occur in a wide array of other solid tumor types[1,2]. Detection for gene fusions in tumors can help establish a diagnosis, provide prognostic information, and guide targeted therapy in clinical decisions. For example, some gene fusions, such as the *ALK* fusions can be used by pathologists to clarify a diagnosis due to a high specific expression in inflammatory myofibroblastic tumors[3]. Similarly, *NTRK1* fusion detection are useful to distinguish lipofibromatosis-like neural tumor from histologically similar lipofibromatosis[4]. On the other hand, non-small cell lung cancer patients with an *ALK* gene fusion may benefit from *ALK* tyrosine kinase inhibitors, such as crizotinib, ceritinib, and alectinib[5–7]. Other well-known fusions involving *ROS1*, *RET*, and all three *NTRK* genes have been well characterized, and matched targeted therapies have demonstrated remarkable clinical efficacy[8–11].

How to accurately detect gene fusions is a key issue in clinical work. Traditional methods are immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH), which have been widely used in clinical practice. However, these technologies have a common limitation, namely the poor compatibility with multiplexing, preventing these methods from interrogating multiple fusion genes simultaneously, which cannot satisfy the need when the diagnosis is unclear[12]. The emergence of next generation sequencing (NGS) technology and bioinformatics pipelines allows sequencing numerous genes in parallel and facilitates the identification of fusion. Theoretically, gene fusion can be detected at both the DNA and RNA level. Nonetheless, there are still some challenges ahead. Due to the unpredictable characteristics of genomic breakpoints and blind spots within the targeted areas, DNA based NGS requires a large range of genomes to be covered to ensure the accuracy of the results. In terms of RNA based NGS, chemical modifications and oxidative denaturation as well as RNA

[1]Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China. [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China. [3]Institute of Pathology, Fudan University, Shanghai 200032, China. ✉email: xyzhou100@163.com

degradation making downstream processes difficult or impossible, especially for formalin-fixed and paraffin-embedded (FFPE) samples[13].

Currently, commercial genetic testing reagents used in most clinical applications rely solely on either DNA or RNA based NGS to detect gene fusions, which may lead to missed detection or error detection. Meanwhile, the high cost of commercial reagents poses a serious burden on patients and healthcare finances, so we urgently need to establish an economical and accurate hospital developed test. In this study, we describe the development and validation of a custom-designed panel that simultaneously use DNA and RNA based NGS for fusion detection in solid tumor diagnosis and targeted therapy. Detailed accuracy, sensitivity, specificity, mutational abundance limit of detection (LOD), and reproducibility of the assay were also discussed.

Results

Overview of DNA and RNA sequencing assay development

The general workflow diagram for the development and validation of fusion gene detection assay is illustrated (Fig. 1). The detailed experimental procedures are described in the Materials and Methods section. This protocol needed to be validated in two ways: technical validation with reference standards and clinical validation with FFPE tumor samples.

DNA and RNA sequencing pipeline fully validated with reference standards

DNA and RNA fusion reference standards containing 10 fusions across *ALK*, *ROS1*, *RET*, and all three *NTRK* genes were purchased from GeneWell company. The detailed fusions spiked in these reference standards are listed in Supplement Table 2.

To determine how mutational abundance affected fusion detection, serial dilution experiments were performed. Four fusions with three dilution gradients, namely 2.5%, 5% and 8% for DNA mutational abundance, and 250–400 copies/100 ng, 500–800 copies/100ng and 1000–2000 copies/100 ng for RNA mutational abundance, were conducted with five repeated detections for each. The *EML4*::*ALK*, *CD74*::*ROS1* and *CCDC6*::*RET* fusions were identified in all replicates across all dilutions (Tables 1 and 2; Fig. 2). The *SLC34A2*::*ROS1* fusion were stably detected in RNA level, but 3 out of 5 repeated detections at 2.5% mutational abundance in DNA level were found with reads detected less than five (Table 1). Notably, the number of reads to support *EML4*::*ALK* and *SLC34A2*::*ROS1* differed greatly, therefore, the sensitivity of fusion detection may vary across different fusions.

Validation based on clinical samples

To assess the clinical utility of this assay, 60 samples, including 30 fusion positive and 30 fusion negative samples were selected. Their fusion status was previously determined by NGS or FISH, in which 3 samples were simultaneously detected for DNA and RNA levels, 29 samples were detected only for DNA level, and the other 28 samples were detected only for RNA level.

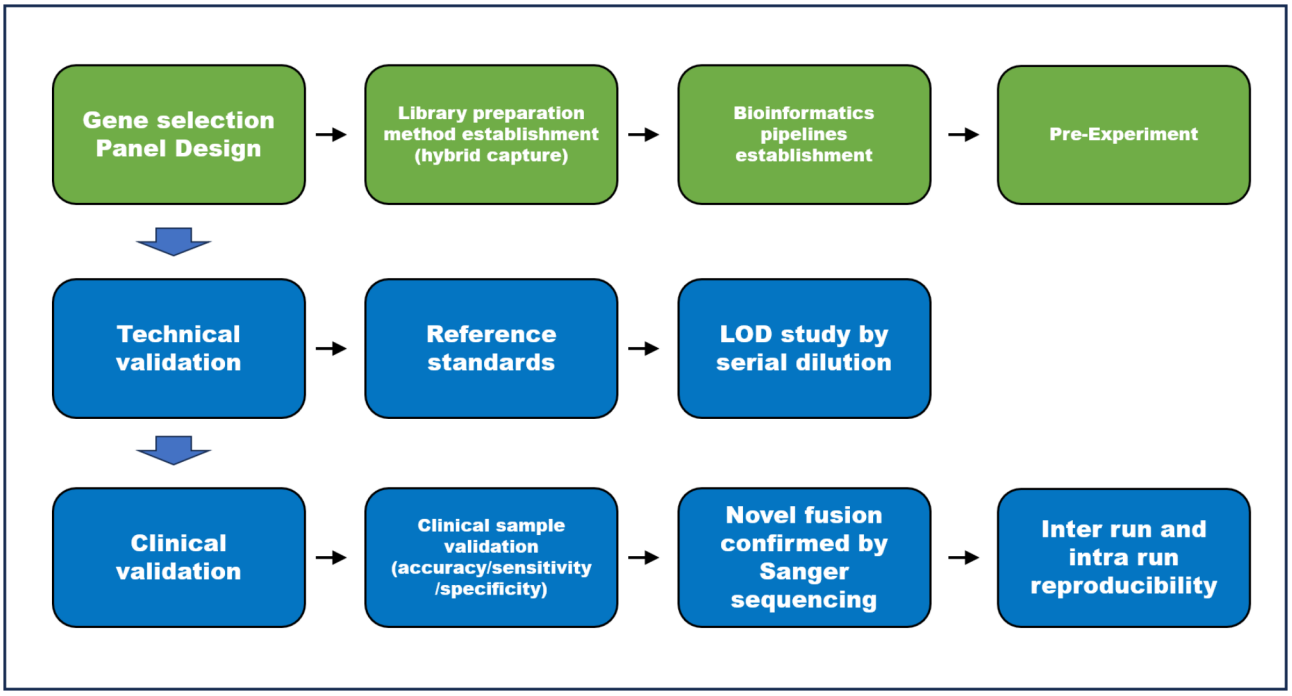

Fig. 1. The integration of DNA-based and RNA-based NGS assay validation workflow. The validation workflow consisted of technical and clinical validations. Reference standards were used for technical validation, whereas formalin-fixed, parrafin-embedded (FFPE) samples from clinical cancer patients were used for clinical validation. LOD, limit of detection.

LOD gradient (mutational abundance)	Replicates	DNA allele frequency(supporting reads/total reads)			
		EML4::ALK	*CD74::ROS1*	*CCDC6::RET*	*SLC34A2::ROS1*
2.5%	1	2.07% (32/1542)	1.6% (25/1560)	1.78% (23/1291)	0.38% (4/1045)
	2	1.7% (27/1584)	1.09% (18/1650)	1.71% (24/1399)	0.79% (9/1136)
	3	1.99% (30/1507)	0.54% (9/1638)	1.52% (22/1442)	0.38% (4/1049)
	4	1.73% (28/1612)	0.88% (15/1689)	2.03% (28/1377)	0.19% (2/1048)
	5	1.94% (30/1539)	1.07% (19/1760)	1.48% (21/1418)	0.61% (7/1144)
5%	1	4.76% (66/1386)	2.85% (42/1471)	3.74% (47/1256)	1.21% (12/985)
	2	1.1% (14/1268)	2.47% (36/1456)	3.79% (47/1237)	0.6% (6/994)
	3	4.46% (53/1188)	2.37% (31/1308)	2.7% (30/1111)	0.97% (8/821)
	4	3.91% (48/1225)	2.81% (37/1313)	3.24% (34/1048)	0.98% (8/815)
	5	0.91% (11/1206)	2.21% (33/1493)	3.03% (38/1252)	0.75% (7/931)
8%	1	6.63% (87/1312)	3.7% (50/1350)	6.05% (77/1272)	1.17% (11/937)
	2	6.27% (92/1465)	4.88% (79/1618)	4.67% (63/1348)	1.17% (12/1019)
	3	1.78% (22/1234)	4.51% (64/1416)	5.45% (68/1246)	1.76% (16/908)
	4	4.24% (62/1461)	3.27% (50/1527)	6.47% (88/1359)	1.56% (15/959)
	5	7.36% (91/1236)	2.95% (43/1457)	6.14% (71/1155)	1.46% (13/888)

Table 1. DNA reference standards detection results.

LOD gradient (mutational abundance)	Replicates	fusion fragment per million (FFPM)			
		EML4::ALK	*CD74::ROS1*	*CCDC6::RET*	*SLC34A2::ROS1*
250–400 copies/100 ng	1	4.7132	6.7181	2.7799	4.1699
	2	2.0541	5.3295	1.8272	3.1977
	3	2.3622	4.8909	0.7931	1.3219
	4	2.9539	2.194	2.7714	0.3464
	5	1.2263	3.0725	1.8703	2.0038
500–800 copies/100ng	1	7.1375	4.9507	5.658	5.9409
	2	3.7665	5.7624	5.5004	6.2862
	3	4.9274	7.9828	3.2188	5.0215
	4	9.8251	7.1936	5.7548	4.1853
	5	2.9906	4.0238	5.0622	4.787
1000–2000 copies/100 ng	1	15.0555	17.7107	12.1325	13.0083
	2	12.7402	14.5202	11.7976	16.3521
	3	11.3557	10.079	8.3992	12.7242
	4	14.5108	11.1427	8.2743	14.1214
	5	16.96	11.9572	9.0243	8.4602

Table 2. RNA reference standards detection results.

In this cohort, lung adenocarcinoma, papillary thyroid carcinoma and endometrial stromal sarcoma were the three most common tumor types, occupying 33.3%, 10%, and 5% respectively (Table 3). The proportion of other subtypes ranged from 1.7% to 3.3%, including 25 types of solid tumors such as breast carcinoma and chondrosarcoma (Fig. 3).

Among them, one malignant melanoma (PS04) sample had two fusion genes, namely *TRIM46*::*NTRK1* and *NAV1*::*NTRK1*, while the other positive samples had only one fusion gene that may dominate the disease process. In addition, two fusion genes involved in two fusion positive samples (PS56, PS57) exceed the reportable range of this assay, and were therefore considered fusion negative in subsequent analysis. Other identified gene fusions included were *ALK*, *ROS1*, *RET*, *NTRK1* and *NTRK3* with different partners. Overall, 28 positive samples containing 29 gene fusions and 32 fusion negative samples were selected in this cohort (Table 3, Supplemental Table 3).

Analytic sensitivity and specificity

Of the 28 fusion positive and 32 negative samples, all fusion transcripts were detected using this assay, demonstrating 100% sensitivity. As shown in Table 4, the specificity of fusions detected by the assay was 96.9% (31/32) compared with previous results. Notably, a *TPM3*::*NTRK1* fusion was additionally identified in one fusion-negative thyroid papillary carcinoma sample (PS36) as determined by our laboratory's previous test. This *NTRK1* fusion was confirmed by Sanger sequencing, which found that this is a false-negative result for the previous result (Supplemental Fig. 1). After calibration, the specificity of the assay should be 100%.

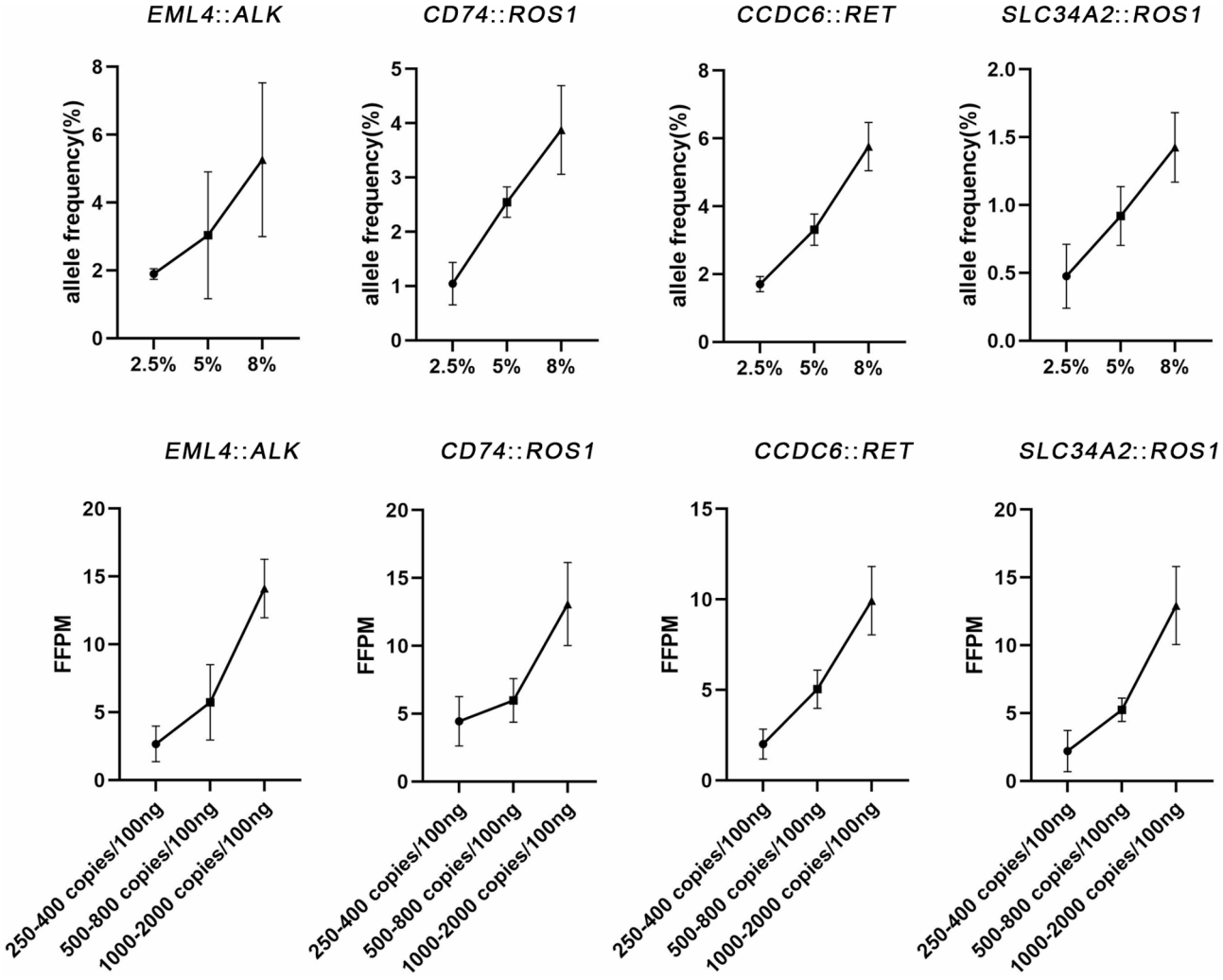

Fig. 2. Fusions detection for serial dilution experiments. Scatterplot of allele frequency or FFPM versus mutational abundance for *EML4*::*ALK*, *CD74*::*ROS1*, *CCDC6*::*RET*, and *SLC34A2*::*ROS1* fusions. Each point indicates mean allele frequency or FFPM of five replicates under specific mutational abundance, and the line represents the SD. FFPM, fusion fragment per million.

From the perspective of sole result based on DNA or RNA, fusion status detected by the DNA assay were 93.4% (57/61) concordant to those by the previous results (Supplemental Table 4). Except for a *TPM3*::*NTRK1* being additionally detected at two levels in the papillary thyroid carcinoma (PS36) sample, the *ETV6*::*NTRK3* in the papillary thyroid carcinoma (PS37) sample, *CCDC6*::*RET* in the papillary thyroid carcinoma (PS38) sample and *SORBS1*::*ALK* in the pelvic spindle cell tumor (PS39) sample were missed by this DNA assay. The RNA assay was used to validate the inconsistency between the two different test and could complement the DNA based result. Meanwhile, fusions status detected by the RNA assay were 86.9% (53/61) concordant to those by the previous results (Supplemental Table 5). The *TRIM46*::*NTRK1* in the malignant melanoma (PS04) sample, *ECHDC1*::*ROS1* in the epithelioid hemangioendothelioma (PS24) sample, and one *KIF5B*::*RET*, one *EML4*::*ALK*, one *KIF5B*::*ALK*, two *CD74*::*ROS1* in 5 lung adenocarcinoma (from PS19 to PS23) samples were missed by this RNA assay. The DNA assay was used to validate the inconsistency between the two different test and could complement the RNA based result.

High precision in detecting fusions by the assay

To evaluate the precision of this assay, three samples containing one standard FFPE sample with six NTRK fusions (*BTBD1*::*NTRK3*, *ETV6*::*NTRK3*, *LMNA*::*NTRK1*, *NACC2*::*NTRK2*, *QKI*::*NTRK2* and *TPM3*::*NTRK1*), one clinical FFPE sample with *EML4*::*ALK* and one clinical negative FFPE sample were assessed (Table 5). The intra run reproducibility was performed in triplicate within one sequencing run (RUN1-1,1–2 and 1–3). Both fusion positive samples had the expected fusion(s) detected, and no fusions were identified in the fusion negative sample in all three replicates. The inter run reproducibility was performed in triplicate across three different sequencing run (RUN1, 2and 3). There was complete concordance of gene fusion results for all samples across different runs. Further quantitative analysis revealed that the CV of allele frequency detected by this DNA assay

Sample no.	Known fusion(s)	Reference method(s)	Fusion detected with this assay (Supporting reads)		Status (known fusion detected or not)	Concordance
			DNA	RNA		
PS01	*EML4*:: *ALK*	RNA NGS + FISH	*EML4*::*ALK*(25)	*EML4*::*ALK*(14)	Yes	Yes
PS02	/	DNA + RNA NGS	/	/	/	Yes
PS03	/	DNA + RNA NGS	/	/	/	Yes
PS04	*TRIM46*::*NTRK1* *NAV1*::*NTRK1*	DNA NGS	*TRIM46*::*NTRK1*(141) *NAV1*::*NTRK1*(134)	*NAV1*::*NTRK1*(299)	Yes	Yes
PS05	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK*(65)	*EML4*::*ALK*(185)	Yes	Yes
PS06	*ERC1*::*RET*	DNA NGS	*ERC1*::*RET*(95)	*ERC1*::*RET*(37)	Yes	Yes
PS07	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK*(230)	*EML4*::*ALK*(111)	Yes	Yes
PS08	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK*(49)	*EML4*::*ALK*(63)	Yes	Yes
PS09	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK*(74)	*EML4*::*ALK* (172)	Yes	Yes
PS10	*CD74*::*ROS1*	DNA NGS	*CD74*::*ROS1*(25)	*CD74*::*ROS1*(7)	Yes	Yes
PS11	*CD74*::*ROS1*	DNA NGS	*CD74*::*ROS1*(86)	*CD74*::*ROS1*(5)	Yes	Yes
PS12	*SDC4*::*ROS1*	DNA NGS	*SDC4*::*ROS1*(55)	*SDC4*::*ROS1*(319)	Yes	Yes
PS13	*CD74*::*ROS1*	DNA NGS	*CD74*::*ROS1*(34)	*CD74*::*ROS1*(62)	Yes	Yes
PS14	*CD74*::*ROS1*	DNA NGS	*CD74*::*ROS1*(22)	*CD74*::*ROS1*(24)	Yes	Yes
PS15	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK*(101)	*EML4*::*ALK*(23)	Yes	Yes
PS16	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK*(18)	*EML4*::*ALK*(11)	Yes	Yes
PS17	*TPM3*::*NTRK1*	DNA NGS	*TPM3*::*NTRK1*(152)	*TPM3*::*NTRK1*(7)	Yes	Yes
PS18	*EZR*::*ROS1*	DNA NGS	*EZR*::*ROS1*(22)	*EZR*::*ROS1*(128)	Yes	Yes
PS19	*CD74*::*ROS1*	DNA NGS	*CD74*::*ROS1*(42)	/	Yes	Yes
PS20	*CD74*::*ROS1*	DNA NGS	*CD74*::*ROS1*(157)	/	Yes	Yes
PS21	*KIF5B*::*RET*	DNA NGS	*KIF5B*::*RET*(167)	/	Yes	Yes
PS22	*EML4*::*ALK*	DNA NGS	*EML4*::*ALK* (64)	/	Yes	Yes
PS23	*KIF5B*::*ALK*	DNA NGS	*KIF5B*::*ALK*(51)	/	Yes	Yes
PS24	*ECHDC1*::*ROS1*	DNA NGS	*ROS1*::*ECHDC1*(66)	/	Yes	Yes
PS25	/	DNA NGS	/	/	Yes	Yes
PS26	/	DNA NGS	/	/	/	Yes
PS27	/	FISH	/	/	/	Yes
PS28	/	DNA NGS	/	/	/	Yes
PS29	/	DNA NGS	/	/	/	Yes
PS30	/	DNA NGS	/	/	/	Yes
PS31	/	DNA NGS	/	/	/	Yes
PS32	/	DNA NGS	/	/	Yes	Yes
PS33	*SPECC1L*::*NTRK3*	RNA NGS	*SPECC1L*::*NTRK3*(135)	*SPECC1L*::*NTRK3*(515)	Yes	Yes
PS34	*MLPH*::*ALK*	RNA NGS	*MLPH*::*ALK*(7)	*MLPH*::*ALK*(8)	Yes	Yes
PS35	*CCDC6*::*RET*	RNA NGS	*CCDC6*::*RET*(311)	*CCDC6*::*RET*(96)	Yes	Yes
PS36	/	RNA NGS	*TPM3*::*NTRK1*(265)	*NTRK1*::*TPM3*(27)	a new fusion was detected	NO
PS37	*ETV6*::*NTRK3*	RNA NGS	/	*ETV6*::*NTRK3*(47)	Yes	Yes
PS38	*CCDC6*::*RET*	RNA NGS	/	*CCDC6*::*RET*(47)	Yes	Yes
PS39	*SORBS1*::*ALK*	RNA NGS	/	*SORBS1*::*ALK*(128)	Yes	Yes
PS40	/	RNA NGS	/	/	/	Yes
PS41	/	RNA NGS	/	/	/	Yes
PS42	/	RNA NGS	/	/	/	Yes
PS43	/	RNA NGS	/	/	/	Yes
PS44	/	RNA NGS	/	/	/	Yes
PS45	/	RNA NGS	/	/	/	Yes
PS46	/	RNA NGS	/	/	/	Yes
PS47	/	RNA NGS	/	/	/	Yes
PS48	/	RNA NGS	/	/	/	Yes
PS49	/	RNA NGS	/	/	/	Yes
PS50	/	RNA NGS	/	/	/	Yes
PS51	/	RNA NGS	/	/	/	Yes
PS52	/	RNA NGS	/	/	/	Yes
PS53	/	RNA NGS	/	/	/	Yes
PS54	/	RNA NGS	/	/	/	Yes

Continued

Sample no.	Known fusion(s)	Reference method(s)	Fusion detected with this assay (Supporting reads)		Status (known fusion detected or not)	Concordance
			DNA	RNA		
PS55	/	RNA NGS	/	/	/	Yes
PS56	*CEP170::RAD51B*	RNA NGS	/	/	NO	Yes
PS57	*HGSNAT::KAT6A*	RNA NGS	/	/	NO	Yes
PS58	/	RNA NGS	/	/	/	Yes
PS59	/	RNA NGS	/	/	/	Yes
PS60	/	RNA NGS	/	/	/	Yes

Table 3. Clinical samples detection results.

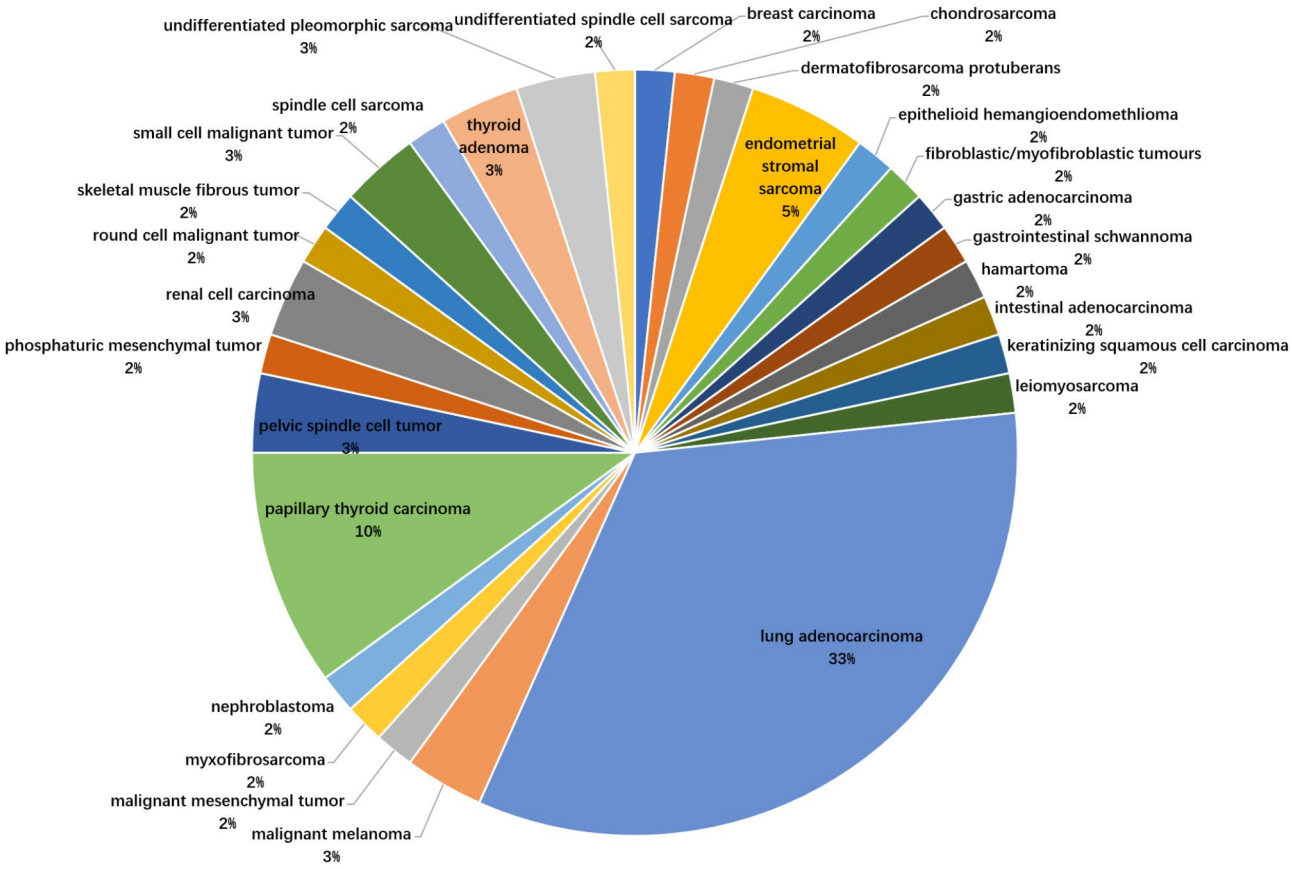

Fig. 3. The tumor types and proportions of 60 clinical validation samples.

This assay (DNA + RNA)	Previous results		Performance
	Positive	Negative	
Positive	29	1	96.7% (PPV)
Negative	0	31	100% (NPV)
Performance	100% (sensitivity)	96.9% (specificity)	98.4% (accuracy)

Table 4. Assay performance characteristics for 60 clinical samples.

was consistent for any fusion among intra and inter runs. Meanwhile, the CV of fusion fragment per million (FFPM) values measured by this RNA assay among different repeated experiments also showed similar and consistent results. These findings indicated that this integration of DNA-based and RNA-based NGS assay could be stably used to detect gene fusions for FFPE samples.

Sample no.	Known fusion(s)	Replicates	Fusion detected with this assay									
			DNA					RNA				
			Supporting reads	Total reads	Allele frequency	Intra CV (100%)	Overall CV (100%)	Supporting reads	Total reads	FFPM	Intra CV (100%)	Overall CV (100%)
PS01	*EML4*::*ALK*	RUN1-1	13	307	0.042	5.0%	6.7%	15	9599	5.5048	18.5%	17.2%
		RUN1-2	43	1218	0.035			13	6440	3.9968		
		RUN1-3	45	1171	0.038			8	5318	5.7127		
		RUN2	46	1199	0.038			10	8028	6.2707		
		RUN3	41	1047	0.039			15	6041	4.666		
SD4	*BTBD1*::*NTRK3*	RUN1-1	25	942	0.027	14.4%	16.5%	67	593	55.7755	1.6%	6.7%
		RUN1-2	28	1215	0.023			91	733	55.9022		
		RUN1-3	37	1339	0.028			63	487	54.3125		
		RUN2	26	910	0.029			104	724	49.1779		
		RUN3	33	932	0.035			71	563	48.3127		
	ETV6::*NTRK3*	RUN1-1	26	1046	0.025	26.2%	18.3%	77	958	43.5839	4.9%	3.6%
		RUN1-2	37	1200	0.031			78	1011	39.5821		
		RUN1-3	24	1236	0.019			101	1138	42.3943		
		RUN2	28	1305	0.021			73	879	41.4758		
		RUN3	22	877	0.025			93	1092	42.4153		
	LMNA::*NTRK1*	RUN1-1	28	685	0.041	19.5%	23.7%	56	1485	39.1137	2.6%	3.5%
		RUN1-2	20	803	0.025			58	1478	37.903		
		RUN1-3	28	847	0.033			54	1330	37.1316		
		RUN2	19	807	0.024			79	1398	40.3067		
		RUN3	24	687	0.035			63	1299	39.9716		
	NACC2::*NTRK2*	RUN1-1	30	553	0.054	9.7%	10.0%	44	415	27.951	3.1%	13.4%
		RUN1-2	39	688	0.057			64	575	26.546		
		RUN1-3	46	714	0.064			58	489	26.4407		
		RUN2	34	671	0.051			67	595	23.5094		
		RUN3	27	508	0.053			59	557	19.5891		
	QKI::*NTRK2*	RUN1-1	24	531	0.045	17.8%	13.2%	52	762	34.9229	4.1%	3.8%
		RUN1-2	40	635	0.063			68	899	36.4636		
		RUN1-3	39	655	0.060			44	599	33.6231		
		RUN2	40	731	0.055			70	955	33.4049		
		RUN3	23	442	0.052			66	905	35.8051		
	TPM3::*NTRK1*	RUN1-1	20	879	0.023	21.3%	16.2%	68	1176	47.2158	5.8%	4.7%
		RUN1-2	31	1166	0.027			93	1382	43.1805		
		RUN1-3	27	1186	0.023			72	1032	48.3128		
		RUN2	16	852	0.019			77	1228	43.9586		
		RUN3	21	738	0.028			79	1134	46.2712		

Continued

Sample no.	Known fusion(s)	Replicates	Fusion detected with this assay									
			DNA					RNA				
			Supporting reads	Total reads	Allele frequency	Intra CV (100%)	Overall CV (100%)	Supporting reads	Total reads	FFPM	Intra CV (100%)	Overall CV (100%)
PS47	negative	RUN1-1	/	/	/	/	/	/	/	/	/	/
		RUN1-2	/	/	/			/	/	/		
		RUN1-3	/	/	/			/	/	/		
		RUN2	/	/	/			/	/	/		
		RUN3	/	/	/			/	/	/		

Table 5. Intra run and inter run reproducibility.

Discussion

According to the latest Chinese expert consensus on clinical practice of fusion genes detection, a unified DNA and RNA based NGS strategy would maximize the benefits of gene fusion detection for lung cancer patients[14,15]. To the best of our knowledge, there is no approved commercial reagent kit that can simultaneously detect gene fusion at the DNA and RNA levels so far in China. Therefore, clinical genetic testing poses requirements for the development of in-hospital testing processes. In this study, we designed, developed, and verified an integration of DNA-based and RNA-based NGS assay that can be used to detect gene fusions in solid tumors. Our panel mainly focuses on kinase genes, such as *ALK*, *ROS1*, *RET*, and *NTRK* genes, which can identify a therapeutic target that can be compatible with agents that are approved or available in clinical trials. Meanwhile, this assay can also be used to aid in differential diagnosis, especially for some diagnosis of soft tissue tumors.

Since the routine testing of molecular oncology typically using FFPE samples, the deterioration of nucleic acid quality caused by formalin fixation brings on difficulties to genetic testing, especially for RNA[16]. We optimized our workflow to use FFPE samples for the generation of high-quality NGS data from relatively low input and neoplastic content. In practical terms, the neoplastic content in sample PS24 only accounts for 20%, and the extracted DNA and RNA concentrations are only 1.69 ng/μL and 26 ng/μL, respectively (Supplement Table 3). We then used 80 ng DNA and 370 ng RNA for library construction, and found all experimental quality control indicators were qualified and a *ECHDC1*::*ROS1* rearrangement was found. In addition, we used 50 ng DNA and 100 ng RNA as initial input for library construction in all tests using reference standards during our development and validation phase. This assay can stably detect fusions even when the LOD of mutational abundance is as low as 5% and 250–400 copies/100 ng for DNA and RNA input, respectively. Additionally, the assay demonstrated good intra-assay and inter-assay reproducibility not only for the presence or absence of fusions but also for number of supporting reads and allele frequency or FFPM. In clinical validation, the assay could detect the fusions in FFPE samples among solid tumors using this assay, with a 100% sensitivity and 100% specificity. There was an additionally identified *TPM3*::*NTRK1* in one fusion-negative thyroid papillary carcinoma sample as determined by our laboratory's previous test, which was then verified by Sanger sequencing, indicating the high accuracy of this assay.

Although simultaneously use DNA and RNA sequencing for fusion detection may increase the costs and turnaround time, a series of studies have shown the mutual complementation of results based on DNA and RNA, making the molecular features clearer and more reliable. MSKCC has reported a large retrospective study that newly identified 33 actionable fusions or *MET* skipping from 254 DNA level driver-negative lung cancer patients using RNA sequencing, indicating a false-negative value of 13% for DNA sequencing[17]. Similarly, another large-scale study in China investigated 376 cases of *ALK*, *ROS1* and *RET* fusions using DNA sequencing among 3787 non-small cell lung cancer samples[18]. They classified these fusions into three types, including canonical, non-canonical and primary/reciprocal subtypes based on genomic breakpoint position, and found 12.8% (6/47) of these fusions detected in the latter two subtypes were actually non-productive rearrangements that generated no aberrant transcripts/proteins according to further RNA sequencing/IHC. Interestingly, both studies mentioned above also found that *ROS1* fusions were the most prevalent fusions commonly missed by DNA NGS, followed by *NRG1*, *ALK*, and *RET* fusions[17,18]. In addition, there are other two studies involving non-canonical fusion are worthy of note as well. One of the studies identified and validated noncanonical *RET* fusion in non-small cell lung cancer through DNA and RNA sequencing. They found that 72.7% (8/11) of cases presumed out-of-frame fusion or 5'-end fusion by DNA sequencing was proved to be canonical *RET* fusion transcripts by RNA sequencing, suggesting a low concordance between the two levels[19]. Another study firstly identified seven *NTRK1* rearrangements and other 432 pan-negative cases among 4619 lung adenocarcinoma patients using DNA sequencing, and then verified them with total nucleic acid based NGS and IHC. They found that five out of seven *NTRK1* rearrangements were not transcribed, and two *NTRK2* fusion cases were newly found in the other qualified pan-negative samples[20]. In summary, the discordance of fusion detection between DNA and RNA sequencing has many reasons. Except for the well-known fact that low tumor content and nucleic acid degradation can lead to false negative results, missed detections of DNA NGS may be related to the difficulty for comprehensively cover intronic or repetitive sequence regions in probe design, while RNA NGS can directly detect functional transcripts and more clearly reflect the situation of transcriptional modification or transcriptional silencing involved in unknown biological mechanisms[18–23]. In our validation set, the inconsistency rate between DNA and RNA results is 19.7% (12/61). Except for one sample (PS36) that was additionally identified with a *TPM3*::*NTRK1* fusion at both the DNA and RNA levels at a detection locus not included in the laboratory's previous detection method, the remaining samples showed inconsistencies, including seven samples with DNA level fusion but negative in RNA level, and three cases with the opposite situations. The PS04 sample simultaneously harbor *NAV1*::*NTRK1* and *TRIM46*::*NTRK1* fusions at the DNA level, but only *NAV1*::*NTRK1* undergoes transcription at the RNA level, indicating the presence of complex fusions (Supplemental Fig. 2). Also, the PS24 sample had complex rearrangements at the DNA level, where both *ECHDC1* and *ROS1* gene only retained the 3 'end and could not be further expressed (Supplemental Fig. 3). The other five samples (PS19-23) with no fusion detected at the RNA level may be due to sample quality issues or the presence of complex biological mechanisms. As for the three fusions (PS37-39) that were missed at the DNA level, the reason may be more inclined towards that they have a very low fusion abundance or breakpoints outside the probe coverage range.

Generally, our assay simultaneously uses DNA and RNA sequencing for fusion detection, results at different levels can complement each other. In addition, the probe capture method ensures that the results are as comprehensive and accurate as possible, thereby facilitating precise diagnosis and treatment. In addition, there are also many advantages to establish a hospital developed test. First, our assay is more in line with clinical needs. Second, because our assay is developed and validated internally, we have a clearer understanding of the details involved, which is also more helpful for quality control. Third, the raw material cost of the reagents used

in our assay is much cheaper than commercial reagents. The cost of our assay simultaneously uses DNA and RNA sequencing may be lower than the cost of single level testing with commercial reagents, making it more cost-effective.

Conclusion

This integration of DNA and RNA-based NGS assay developed in our study shows excellent performance on FFPE samples. The assay was indeed a significant improvement overall from a quality and cost perspective within our in-hospital testing processes for the detection of fusions in solid tumors.

Materials and methods

Samples

A total of 68 independent samples, including 8 commercial fusion spiked-in reference standards (GeneWell, Shenzhen, China; OrigiMed, Shanghai, China) and 60 clinical FFPE samples were used for the development and clinical validation of the sequencing assay. The 60 FFPE samples were collected from the Department of Pathology of Fudan University Shanghai Cancer Center (Shanghai, China) during 2021–2022. The study was conducted in accordance with the Declaration of Helsinki. The study was approved by Medical Ethics Committee of Fudan University Shanghai Cancer Center and informed consent was taken from all individual participants (2103232-2). The validation was performed in a double-blinded manner to eliminate the effect from personnel factors.

Panel design

A total of 16 genes known to be involved in cancer associated fusion genes were manually screened from the presence of clinically targeted therapy relevant fusions present in solid tumors. The genes targeted in this panel and probe region are listed in Supplement Table 1. The overall design principle is that for DNA probes, coverage of all coding regions, UTR regions, and fusion related intron regions were required, while for RNA probes, coverage of all coding regions and UTR regions were required. The gene list was sent to Intergrated DNA Technologies (IDT) for Target Capture Probe Design& Ordering Tool design layout. For the canonical protein-coding genes, biotinylated DNA probes were tiled across all hg19 annotated exons from all isoforms with limited trimming of regions containing repetitive sequences or strong homology to other genes to minimize off-target results. Target-specific probes produced from IDT Company were typically 120 bp in length. They were custom designed to identify known fusion transcripts and potential novel fusion transcripts.

Genomic DNA and total RNA isolation

FFPE tissue samples were collected and retrieved from our biobank and the Department of Pathology for NGS testing. Genomic DNA and total RNA extracted from tumor tissue sections were performed using QIAamp DNA MiniKit and RNeasy FFPE Kit (QIAgen, Valencia, CA), respectively. Then the amount of DNA and RNA was quantified by the fluorescence method of the Qubit DNA/RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA), and the OD260/280 was measured by Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA). Samples were considered of satisfactory quality if DNA ≥ 200ng, OD260/280 ranged 1.8-2.0 and RNA ≥ 1 μg, OD260/280 ranged 1.8–2.2.

Library Preparation and sequencing

Complementary DNA (cDNA) was synthetized from 100ng to 1 μg of total RNA with the First-Strand cDNA Synthesis System (Promega Corporation, Fitchburg, Wisconsin) and the Next Ultra II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs, Beverly, Massachusetts). 50 μL purified cDNA or 50-200ng gDNA was fragmented to a size range from 100 to 250 bp with Focused-Ultrasonicator (Covaris Inc, Woburn, Massachusetts). Libraries were prepared by the KAPA Hyper Prep Kit (for Illumina) (KAPA Biosystems, Wilmington, Massachusetts) and custom targeted capture probes (Integrated DNA Technologies, Coralville, Iowa) following the manufacturer's instructions with several modifications in the steps of amplification and library size selection. Briefly, end repair and A-tailing was performed using 3 μL enzyme and 7 μL buffer with the following conditions: 20 °C×30 min and 65 °C×30 min. Adapter ligation was performed using 2 μL adapter, 8 μL PCR-grade water, 10 μL DNA ligase and 30 μL ligation buffer with the following conditions: 20 °C×20 min. Post-ligation cleanup was performed with a 0.8X beads. Library amplification was performed using 20 μL adapter-ligated library, 5 μL primer mix and 25 μL HiFi HotStart ReadyMix buffer with the following cycling conditions: 98 °C×45 s, 10–14 cycles of 98 °C×15 s, 60 °C×30 s and 72 °C×30 s, followed with 72 °C×1 min. Post-amplification cleanup was performed with a 1X beads. According to the quality of different libraries, 1–6 equally pooled samples (500ng for each) were mixed with 5 μL human Cot DNA and 2 μL blocker, and vacuum concentrated to dryness. Hybridization was performed using 8.5 μL hybridization buffer, 2.7 μL enhancer and 5.8 μL custom probes with the following conditions: 95 °C×30 s and 65 °C×16 h. Capture was performed using 50 μL streptavidin beads. Post-capture PCR was performed with the same conditions as the previous steps. The library was quantified and analyzed with the Qubit 4.0 Fluorometer (Life Technologies) and Agilent 2100 Bioanalyzer assay (Agilent Technologies, Santa Clara, California), respectively. Paired-end 2 × 150 bp sequencing was performed on Illumina Next550 DX platform (Illumina, San Diego, California).

Sequencing data analysis

Raw sequencing reads were mapped to the hg19 reference sequence with BWA (version 0.7.12). PCR duplicates were removed by Picard (version 2.5.0), and recalibrated by the BaseRecalibrator tool from GATK (version 3.1.1). An in-house developed algorithm was used for DNA fusion detection by chimeric sequence[24]. The minimum of 1,500,000 unique, mapped reads was required to pass quality control for downstream analysis. Chimeric

sequences were extract from the BAM file based on the insertion length between paired reads greater than 2 kb. The chimeric sequences were clustered into small clusters based on a clustering distance of less than or equal to 1000 bp. Then merge them into large clusters based on a distance of less than or equal to 1000 bp between small clusters, with a predetermined number of regions within the large cluster of less than or equal to 6. Determine the rearrangement region based on the large cluster, annotate the gene results of the rearrangement region, and finally determine the rearrangement type. The reads of mapping score less than 60 was removed and a minimum of 5 support reads of fusion pair was required.

For RNA-Seq data, the minimum of 500× median depth was required to pass quality control for downstream analysis. RNA-seq reads were using STAR (version 2.7.9) algorithm for mapping and Arriba (version 2.1.0) for fusion detection. The support reads less than 5, multimappers greater than 5 were removed. Besides, the inframe pairs were choosed and will be considered pseudogenes, gene family databases, and public databases of rearrangement/fusion for further positive determination.

Data availability

The datasets generated and/or analysed during the current study are available in the NODE (http://www.biosino.org/node) repository, http://www.biosino.org/node/project/detail/OEP00005920.

Received: 12 November 2024; Accepted: 21 February 2025
Published online: 28 February 2025

References

1. Mertens, F., Antonescu, C. R. & Mitelman, F. Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. *Genes Chromosomes Cancer*. **55**, 291–310. https://doi.org/10.1002/gcc.22335 (2016).
2. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*. **7**, 233–245. https://doi.org/10.1038/nrc2091 (2007).
3. Yao, Q. et al. Assessment of ALK fusions in uncommon inflammatory myofibroblastic tumors with ALK IHC positivity but FISH-Equivocal findings by targeted RNA sequencing. *Arch. Pathol. Lab. Med.* **146**, 1234–1242. https://doi.org/10.5858/arpa.2021-0230-OA (2022).
4. Higaki-Mori, H. et al. Infantile Lipofibromatosis-like neural tumour investigated by a fusion gene detection assay. *Acta Derm. Venereol.* **100**, adv00180. https://doi.org/10.2340/00015555-3542 (2020).
5. Kwak, E. L. et al. Anaplastic lymphoma kinase Inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703. https://doi.org/10.1056/NEJMoa1006448 (2010).
6. Shaw, A. T. & Engelman, J. A. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N. Engl. J. Med.* **370**, 2537–2539. https://doi.org/10.1056/NEJMc1404894 (2014).
7. Gadgeel, S. M. et al. Safety and activity of alectinib against systemic disease and brain metastases in patients with crizotinib-resistant ALK-rearranged non-small-cell lung cancer (AF-002JG): results from the dose-finding portion of a phase 1/2 study. *Lancet Oncol.* **15**, 1119–1128. https://doi.org/10.1016/S1470-2045(14)70362-6 (2014).
8. Shaw, A. T. et al. Crizotinib in ROS1-rearranged advanced non-small-cell lung cancer (NSCLC): Updated results, including overall survival, from PROFILE 1001. *Ann. Oncol.* **30**, 1121–1126. https://doi.org/10.1093/annonc/mdz131 (2019).
9. Drilon, A. et al. Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov.* **3**, 630–635. https://doi.org/10.1158/2159-8290.CD-13-0035 (2013).
10. Laetsch, T. W. et al. Larotrectinib for paediatric solid tumours harbouring NTRK gene fusions: Phase 1 results from a multicentre, open-label, phase 1/2 study. *Lancet Oncol.* **19**, 705–714. https://doi.org/10.1016/S1470-2045(18)30119-0 (2018).
11. Doebele, R. C. et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: Integrated analysis of three phase 1–2 trials. *Lancet Oncol.* **21**, 271–282. https://doi.org/10.1016/S1470-2045(19)30691-6 (2020).
12. Hamard, C. et al. IHC, FISH, CISH, NGS in non-small cell lung cancer: What changes in the biomarker era? *Rev. Pneumol Clin.* **74**, 327–338. https://doi.org/10.1016/j.pneumo.2018.09.013 (2018).
13. Mittempergher, L. et al. Gene expression profiles from formalin fixed paraffin embedded breast cancer tissue are largely comparable to fresh frozen matched tissue. *PLoS One*. **6**, e17163. https://doi.org/10.1371/journal.pone.0017163 (2011).
14. Molecular Pathology Collaboration Group of Tumor Pathology Committee of China Anti-Cancer. Chinese medical association Chinese society of, O. & pathology quality control, C. [Expert consensus on clinical practice of fusion genes detection in non-small cell lung cancer in China (2023 version)]. *Zhonghua Bing Li Xue Za Zhi*. **52**, 565–573. https://doi.org/10.3760/cma.j.cn112151-20221111-00946 (2023).
15. China Primary Health Care Foundation Tumor, Precision, D. & Treatment, C. Chinese Anti-Cancer association, L. C. S. G. O. C. O. O. [Chinese expert consensus O. the clinical practice O. Non-small cell lung Cancer fusion gene detection based O. RNA-based NGS]. *Zhongguo Fei Ai Za Zhi*. **26**, 801–812. https://doi.org/10.3779/j.issn.1009-3419.2023.102.43 (2023).
16. Wimmer, I. et al. Systematic evaluation of RNA quality, microarray data reliability and pathway analysis in fresh, fresh frozen and formalin-fixed paraffin-embedded tissue samples. *Sci. Rep.* **8**, 6351. https://doi.org/10.1038/s41598-018-24781-6 (2018).
17. Benayed, R. et al. High yield of RNA sequencing for targetable kinase fusions in lung adenocarcinomas with no mitogenic driver alteration detected by DNA sequencing and low tumor mutation burden. *Clin. Cancer Res.* **25**, 4712–4722. https://doi.org/10.1158/1078-0432.CCR-19-0225 (2019).
18. Li, W. et al. Potential unreliability of uncommon ALK, ROS1, and RET genomic breakpoints in predicting the efficacy of targeted therapy in NSCLC. *J. Thorac. Oncol.* **16**, 404–418. https://doi.org/10.1016/j.jtho.2020.10.156 (2021).
19. Xiang, C. et al. Identification and validation of noncanonical RET fusions in Non-Small-Cell lung Cancer through DNA and RNA sequencing. *J. Mol. Diagn.* **24**, 374–385. https://doi.org/10.1016/j.jmoldx.2021.12.004 (2022).
20. Zhao, R. et al. Identification of NTRK gene fusions in lung adenocarcinomas in the Chinese population. *J. Pathol. Clin. Res.* **7**, 375–384. https://doi.org/10.1002/cjp2.208 (2021).
21. Peng, H. et al. Development and validation of an RNA sequencing assay for gene fusion detection in Formalin-Fixed, Paraffin-Embedded tumors. *J. Mol. Diagn.* **23**, 223–233. https://doi.org/10.1016/j.jmoldx.2020.11.005 (2021).
22. Bergeron, D. et al. RNA-Seq for the detection of gene fusions in solid tumors: development and validation of the JAX fusionseq 2.0 assay. *J. Mol. Med. (Berl).* **100**, 323–335. https://doi.org/10.1007/s00109-021-02149-0 (2022).
23. Hu, W. et al. Development and validation of an RNA sequencing panel for gene fusions in soft tissue sarcoma. *Cancer Sci.* **113**, 1843–1854. https://doi.org/10.1111/cas.15317 (2022).
24. Cao, J. et al. An accurate and comprehensive clinical sequencing assay for Cancer targeted and immunotherapies. *Oncologist* **24**, e1294–e1302. https://doi.org/10.1634/theoncologist.2019-0236 (2019).

Acknowledgements
This study was supported by the Innovation Group Project of Shanghai Municipal Health Commission (No. 2019CXJQ03), Innovation Program of Shanghai Science and Technology Committee (No. 20Z11900300).

Author contributions
Conception and design of the investigation: X.Z., Q.B., X.Z. Data generation: G.J., Q.Y., M.R. Data analysis: Q.Y. Manuscript preparation: G.J. Reviewed the manuscript and Funding: X.Z. All authors gave final approval for publication.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-91640-6.

Correspondence and requests for materials should be addressed to X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025