Genetic Diagnostics

A Smarter Way to Fish For Fusions

Application Note

Tumor Development

Cancer is the second most frequent cause of death worldwide. Tumor development begins when the DNA of a single cell changes, causing the cell to multiply abnormally. DNA changes, called mutations, happen frequently and are typically corrected by the cell's DNA repair systems. As an additional safety, failed DNA repair leads to the death of the mutated cell. In very rare cases, DNA repair fails, and the cell survives. If the mutation leads to altered function of a protein involved in growth control, the cell starts to multiply uncontrollably, and a tumor emerges.

The Role of Fusion Transcripts

Besides single nucleotide and copy number variants, many types of cancer are caused by gene fusions. Fusion events bring together parts of two different genes, often changing the activity of one of the partners. For example, a protein that is only active in the presence of an activating ligand in healthy cells may be permanently switched on after either losing the regulatory domain controlled by this ligand or by fusion to a gene that supports constitutive activation. Such "always-on" proteins then activate downstream pathways that regulate cell growth, survival, proliferation, and apoptosis, which in turn leads to carcinogenesis[1]. Examples are the tyrosine kinases *ALK*, *ROS1*, *RET*, *FGFR1/2/3*, *BRAF*, *EGFR*, and *NTRK1*.

Kinases are excellent therapy targets since the fusion is usually the tumor driver and thus present in every cancer cell. Drugs have been approved for all solid tumors with *NTRK1/2/3* fusion, for lung cancers harboring *ALK* or *ROS1* fusion, in cholangiocellular carcinoma and bladder cancer harboring *FGFR* alterations, and in all solid tumors with RET fusion. Other approved and new drugs are in clinical trials for tumors harboring gene fusions, e.g., Dabrafenib + Trametinib + HCQ in brain tumors with *BRAF* fusion, Belvarafenib for all solid tumors with *BRAF* fusion, or Zenocutuzumab for certain tumor entities with *NRG1* fusion. Beyond therapy, fusions in solid tumors have also gained importance in diagnosis and tumor classification, especially in sarcomas[2,3,4,5].

Fusion Detection from RNA

Fusion transcripts result from structural variants on the DNA level such as translocations, inversions, or deletions. Their breakpoints are mostly situated in very large intronic regions, with multiple possible breakpoint locations resulting in the same fusion transcript on the RNA level. Intronic sequences are especially prone to translocations if the intronic sequence is non-unique in the genome. This makes both the design of DNA enrichment panels and the detection of translocations from these regions challenging. In addition, for clinically relevant splice variants such as the *MET* exon 14 skipping variant, the underlying variants on the DNA level remain mostly undetected.

Conversely, not all fusion genes detected at the DNA level form a relevant fusion gene at the RNA and protein levels. This is especially the case in chromosomally highly unstable tumors, where there is an accumulation of putative fusion genes at the DNA level. Most of them are not transcribed or break the reading frame, so no meaningful fusion protein is formed.

Analysis on the RNA level thus is both more sensitive and allows for much better determination of the biological consequence of fusions. Using targeted enrichment, a focused analysis is possible. By including probes that span fusion breakpoint sequences, panel enrichment can significantly increase detection sensitivity with respect to total RNA sequencing. High sensitivity is crucial when working with clinical FFPE specimens that combine degraded material quality with potentially low tumor content.

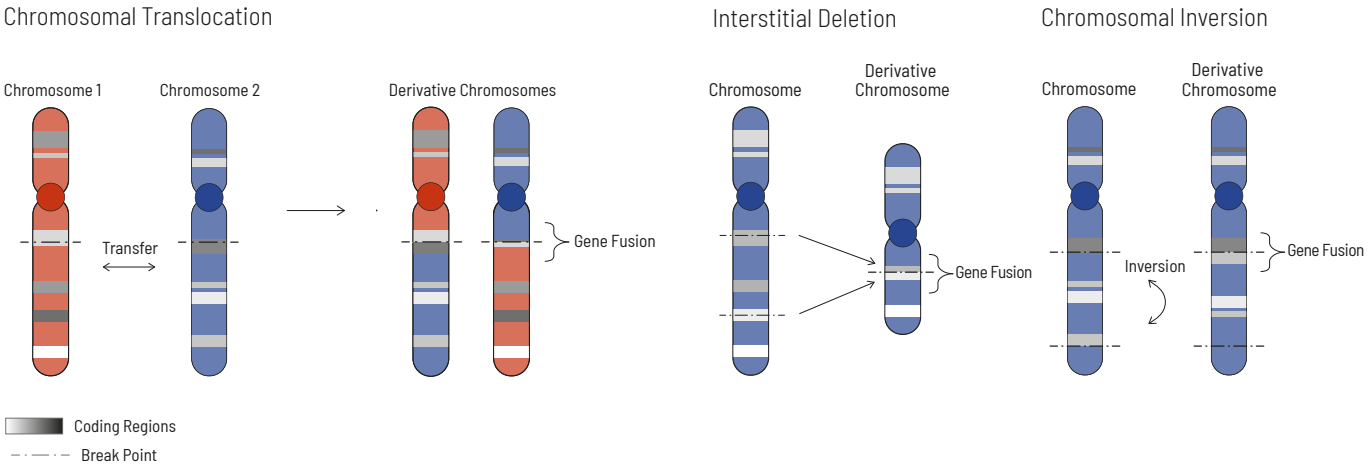

Figure 1: *Gene fusions result from changes to the chromosomes that join the genomic regions of two genes which were previously at separate locations. Different mechanisms can lead to such breakpoints.* ***Chromosomal Translocation:*** *Two chromosomes exchange their arms in a translocation, a partial gene from chromosome 1 is joined to a partial gene from chromosome 2;* ***Interstitial Deletion:*** *The loss of a chromosomal fragment joins two genes previously separated by a large distance;* ***Chromosomal Inversion:*** *The direction of a chromosomal fragment is changed in an inversion.*

Designing a Therapy-Relevant Fusion Detection Panel

Our aim in developing the fusion detection panel was to cover the relevant fusions known to be important for targeted therapeutic drugs, fusions implicated in resistance to treatment, as well as fusions that are relevant for tumor diagnosis or subclassification. In addition, we include important transcript variants such as *EGFR*vIII and *MET* exon 14 skipping. We designed probes spanning known recurrent fusion breakpoints. In addition, all coding exons of relevant genes are targeted to also detect fusions where only one partner is known as well as rare breakpoints for known gene pairs.

We decided to use hybridization enrichment for our panel design for two main reasons. Firstly, we can capture fusions where we only know one fusion gene, by designing probes to capture that gene's exonic sequence. For fusions with an unknown partner, the molecules thus captured will extend into that partner's exonic sequence. Secondly, hybridization enrichment avoids some issues of amplicon-based strategies, particularly allelic dropout which can happen when primer binding sites have patient-specific mutations and lack of amplification in highly fragmented samples. To design the hybridization panel, we partnered with Twist Bioscience. The Twist Target Enrichment workflow is currently one of the best available, giving excellent coverage uniformity and high capture efficiency.

In addition, we have an automated library preparation pipeline set up for high-throughput processing of samples, allowing us to process the fusion panel alongside CeGaT's ExomeXtra® and other enrichment panels with no additional process complexity. This means that the same process is used for patients' DNA samples (e.g. exomes) and their RNA samples for fusion detection, and the resulting sequencing-ready libraries are sequenced alongside one another on the same flowcells. This streamlined process removes the need to collect a minimum number of RNA samples until a separate library preparation or sequencing run is feasible. Finally, the technology is agnostic to the downstream sequencing instrument, which differentiates it from other approaches that are limited to a specific vendor's sequencing platform or instrument.

Experimental Workflow

The fusion detection workflow starts with RNA extracted from fresh-frozen or FFPE tumor material. Following reverse transcription, 10 ng of cDNA from each sample is used as input to the Twist Target Enrichment workflow using the fusion detection panel. The sequencing-ready libraries are sequenced on the Illumina NovaSeq system in paired-end mode (PE100), to 5 Gbp (50 million reads) per sample.

The resulting data is analyzed using a pipeline built around the STAR aligner[6], STAR-Fusion[7], and custom components. The final outputs are fusion calls and FFPM (fusion fragments per million) values, a normalized abundance score describing the strength of the observed event.

An example implementation of the software pipeline is available as docker image on our website at **www.cegat.com/fusions**.

Figure 2: *STAR-Fusion visualization of the ETV6-NTRK3 fusion detected in reference material. Top left, transcript variants of ETV6; top right, transcript variants of NTRK3; blue arc, split read ("junction read") evidence; purple arcs, split pair ("spanning fragment") evidence.*

Validation and Results

To validate the panel design, we ordered publicly available, certified reference materials: SeraCare Fusion RNA mix v4, covering 16 fusions and two transcript variants, and SeraCare FFPE *NTRK* Fusion RNA, covering 15 fusions (including 4 alternative events fusing *ETV6* and *NTRK3*). Both standards were sequenced in triplicates. We also included five clinical samples, extracting RNA from FFPE material.

All expected fusions were detected in all replicates with high FFPM values, indicating 100% sensitivity of the fusion panel on this standard, despite very low RNA quality (RIN=1). A few additional (putative false-positive) fusion calls were detected, but these were at low levels close to the detection limit and had a large distance to the scores of the expected fusions.

Expected Fusion	Replicate 1	Replicate 2	Replicate 3
TPM3-NTRK1	88	78	64
LMNA-NTRK1	88	102	80
IRF2BP2-NTRK1	125	132	113
SQSTM1-NTRK1	71	63	51
TFG-NTRK1	67	50	46
AFAP1-NTRK2	93	73	98
NACC2-NTRK2	37	42	27
QKI-NTRK2	33	34	24
TRIM24-NTRK2	38	36	41
PAN3-NTRK2	47	38	35
ETV6-NTRK3 (E5N14) chr12:12022903-chr15:88576276	165	171	137
ETV6-NTRK3 (E5N15) chr12:12022903-chr15:88483984	145	160	126
ETV6-NTRK3 (E4N14) chr12:12006495-chr15:88576276	149	168	137
ETV6-NTRK3 (E4N15) chr12:12006495- chr15:88483984	159	164	142
BTBD1-NTRK3	88	95	78
Additional calls, max FFPM	2,5	3,8	1,8
Additional calls, FFPM >= 1	4	5	6
Additional calls, FFPM < 1	13	17	18

Table 1: *Fusion panel validation results from the SeraCare FFPE NTRK Fusion RNA standard, sequenced in triplicates. Table shows FFPM (fusion fragments per million) values. All expected fusions are found with high support. Additional (putative false-positive) calls are present but with very low support.*

Clinical Benefit

Fusion detection is an important part of tumor diagnostics and therapy. Depending on the tumor entity, fusions may be the main causative alteration. In prostate cancers analyzed at CeGaT, we find 61% of cases with *TMPRSS2-ERG* fusions, and about 28% of glioblastomas show the *EGFR*vIII transcript variant. In a retrospective analysis of 1,369 clinical cases covering 89 different tumor entities, clinically relevant fusions were detected in 123 (9%) of the cases.

To investigate the benefit of using a targeted enrichment approach, we compared the results from the RNA panel to fusion transcript detection from total RNA sequencing in a cohort of 91 patients.

Of those, 27 patients (30%) had clinically relevant fusions detected by our fusion panel. The enrichment approach has much higher sensitivity than total RNA sequencing: A third of these events was missed in the matched total RNA data. For 59 of these patients, we also had exome sequencing data available, including specific enrichment of intronic translocation regions. Comparing translocation calls from these with the fusion calls from the RNA enrichment panel, we find that more than half of the events (56%) were not detectable from the DNA data.

Conclusion

At CeGaT, we want to offer comprehensive and affordable tumor diagnostics to our clients. Sensitive detection of fusion transcripts is highly relevant in tumor diagnostics, with implications for therapy and prognosis. Targeted enrichment of RNA is a comprehensive and sensitive method for this task, and the Twist Target Enrichment technology is a robust workflow that allows us to integrate RNA fusion detection into automated laboratory processes and sequence samples alongside exomes without the need for a different library preparation process or separate sequencing runs. Together with our automated data analysis, we can accurately analyze patient samples to detect more than 80 therapeutically actionable fusion variants in a high-throughput manner, allowing us to report clinically actionable results and support treatment decisions.

Literature

1. Taniue and Akimitsu, Fusion Genes and RNAs in Cancer Development. *Noncoding RNA 2021*. PMID 33557176.
2. Powers, The ever-changing world of gene fusions in cancer: a secondary gene fusion and progression. Oncogene 2019. PMID 31616060.
3. Huang *et al.*, Establishing an RNA fusions panel in soft tissue sarcoma with clinical validation. *Scientific Reports 2023*. PMID: 36928336.
4. Perry *et al.*, Biology and Therapy of Dominant Fusion Oncoproteins Involving Transcription Factor and Chromatin Regulators in Sarcoma. *Annual Review of Cancer Biology, 2019*. doi: 10.1146/annurev-cancerbio-030518-055710.
5. Mertens *et al.*, Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. *Genes Chromosomes Cancer, 2015*. PMID 26684580.
6. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics 2013*. PMID 23104886.
7. Haas *et al.*, Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology, 2019*. PMID 31639029.

About Us

CeGaT is a global provider of genetic analyses for a wide range of medical, research, and pharmaceutical applications.

Founded in 2009 in Tübingen, Germany, the company combines state-of-the-art sequencing technology with medical expertise – with the aim of identifying the genetic causes of diseases and supporting patient care. For researchers and pharmaceutical companies, CeGaT offers a broad portfolio of sequencing services and tumor analyses. CeGaT generates the data basis for clinical studies and medical innovations and drives science forward with its own insights.

The owner-managed company stands for independence, comprehensive personal customer service, and outstanding quality. CeGaT's laboratory is accredited according to CAP/CLIA, DIN EN ISO 15189, and DIN EN ISO/IEC 17025 and thus meets the highest international standards. To obtain first-class results, all processes are carried out in-house under scientific supervision.

CeGaT GmbH
Genetic Diagnostics
Paul-Ehrlich-Str. 23
72076 Tübingen
Germany

Phone: +49 7071 56544-220
Fax: +49 7071 56544-56
Email: diagnostic-support@cegat.com
Web: www.cegat.com/diagnostics

DAkkS Deutsche Akkreditierungsstelle D-ML-13206-01-00

Accredited by DAkkS according to DIN EN ISO 15189:2014

CAP ACCREDITED COLLEGE of AMERICAN PATHOLOGISTS

CLIA CERTIFIED ID: 99D2130225

2023/08