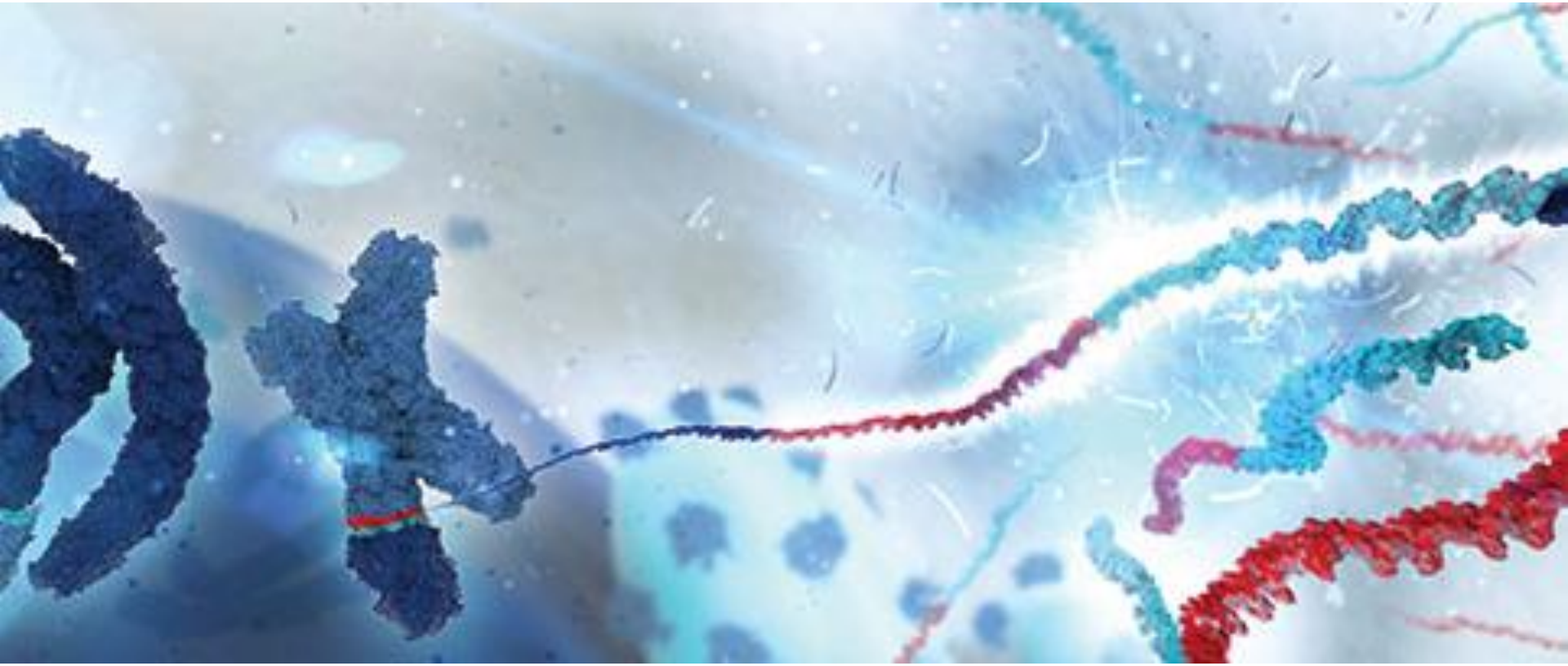
Analyzing Fusion Genes with Next-Generation Sequencing Technology



Raed Samara, Ph.D

Global Product Manager

Legal disclaimer

- QIAGEN products shown here are intended for molecular biology applications. These products are not intended for the diagnosis, prevention or treatment of a disease.
- For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.

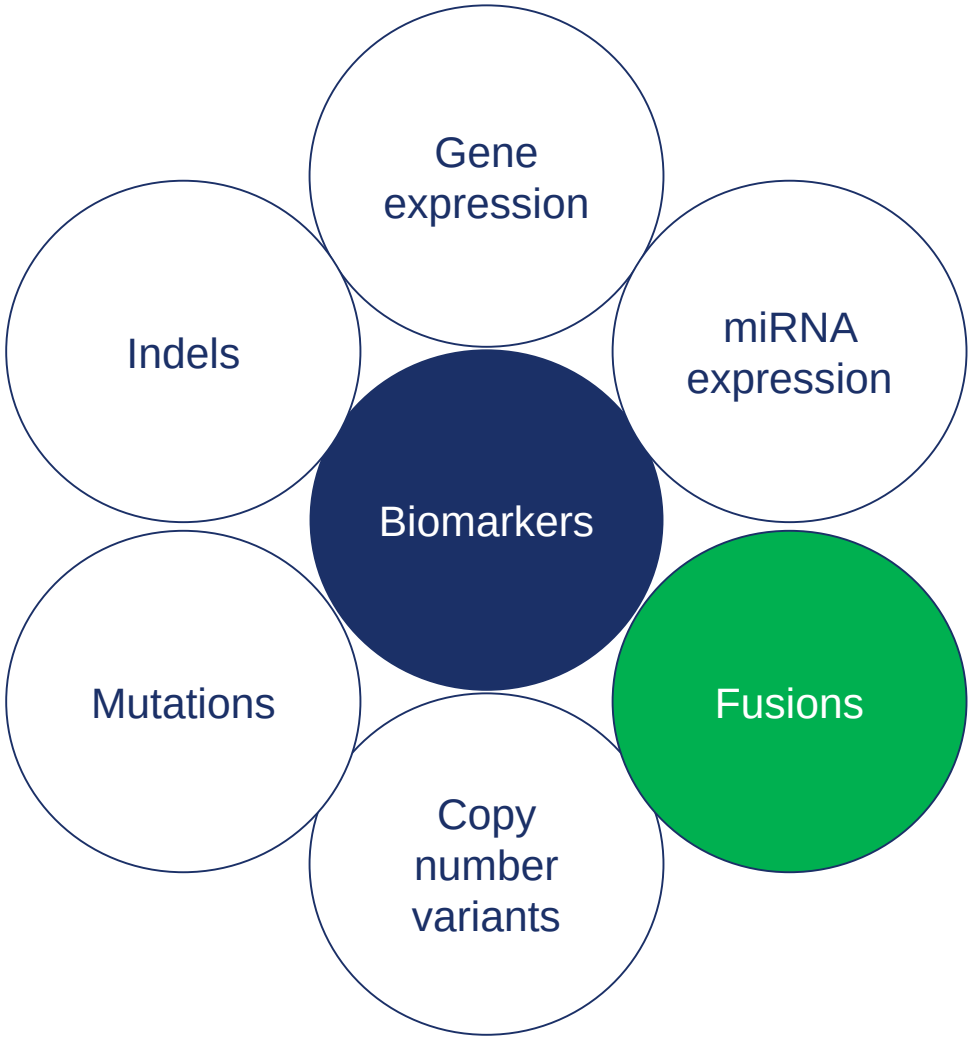
Agenda

1. Fusion genes: What they are and their historical perspective
2. Fusion gene detection: Current status
3. RNA sequencing vs. digital RNA sequencing
4. How to detect and accurately quantify novel fusion genes in your sample

Agenda

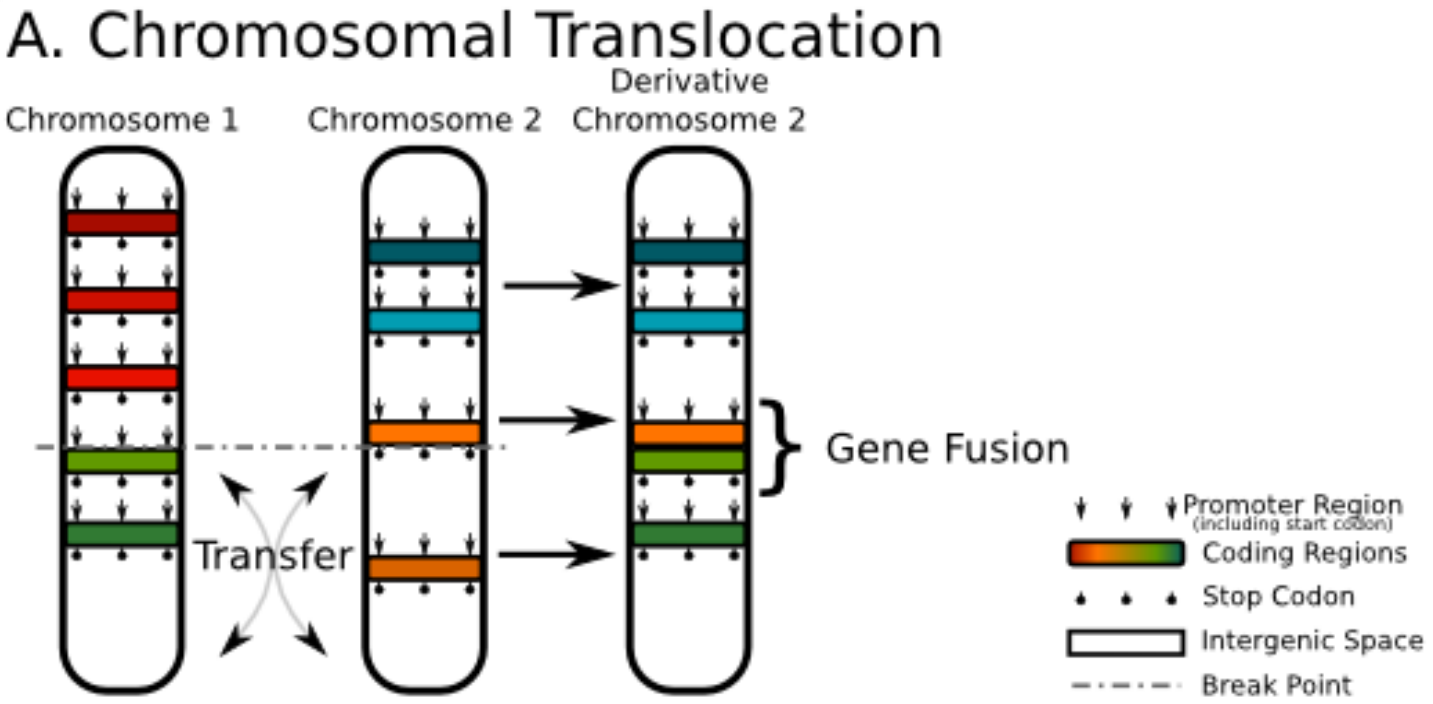
1. Fusion genes: What they are and their historical perspective
2. Fusion gene detection: Current status
3. RNA sequencing vs. digital RNA sequencing
4. How to detect and accurately quantify novel fusion genes in your sample

Gene fusions are one form of biomarkers



What is a gene fusion?

A new gene formed by the chromosomal translocation of two parts of different genes

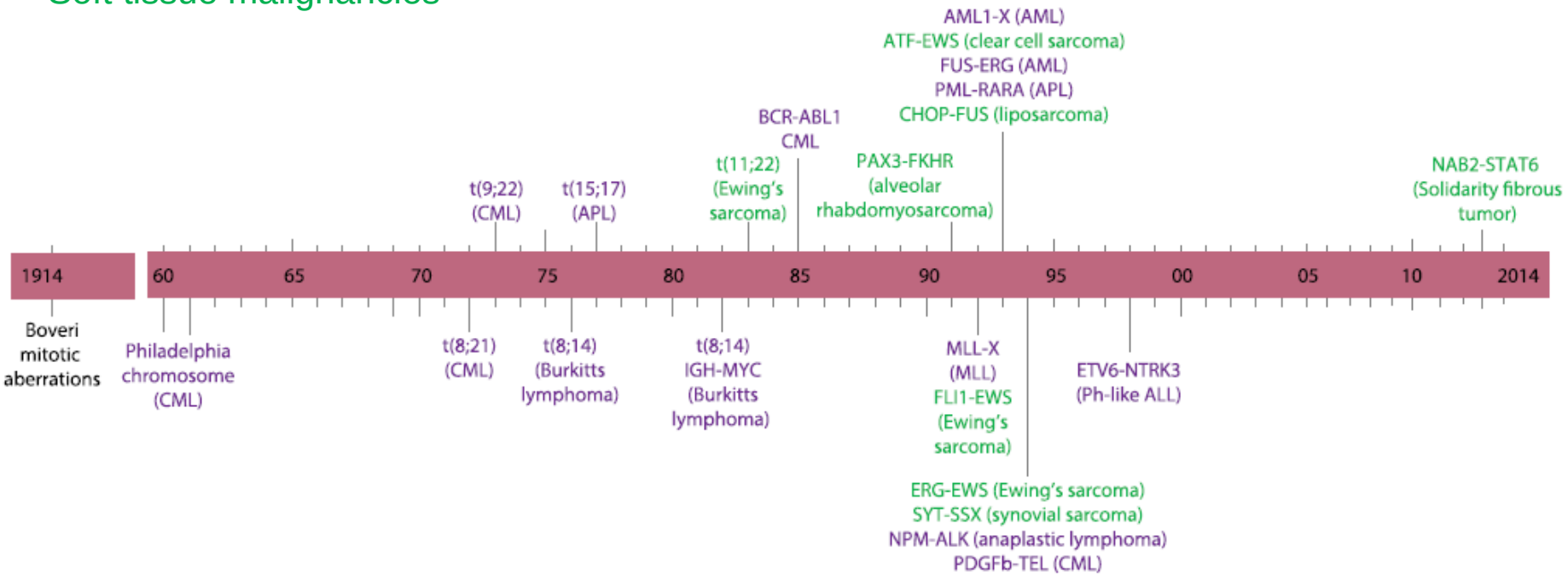


Source: Leonard, G. [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons
https://upload.wikimedia.org/wikipedia/commons/b/b0/Gene_Fusion_Types.png

Why are gene fusions important?

They are implicated in hematologic and soft tissue malignancies…

Timeline of gene fusion discoveries in:
Hematologic malignancies
Soft tissue malignancies

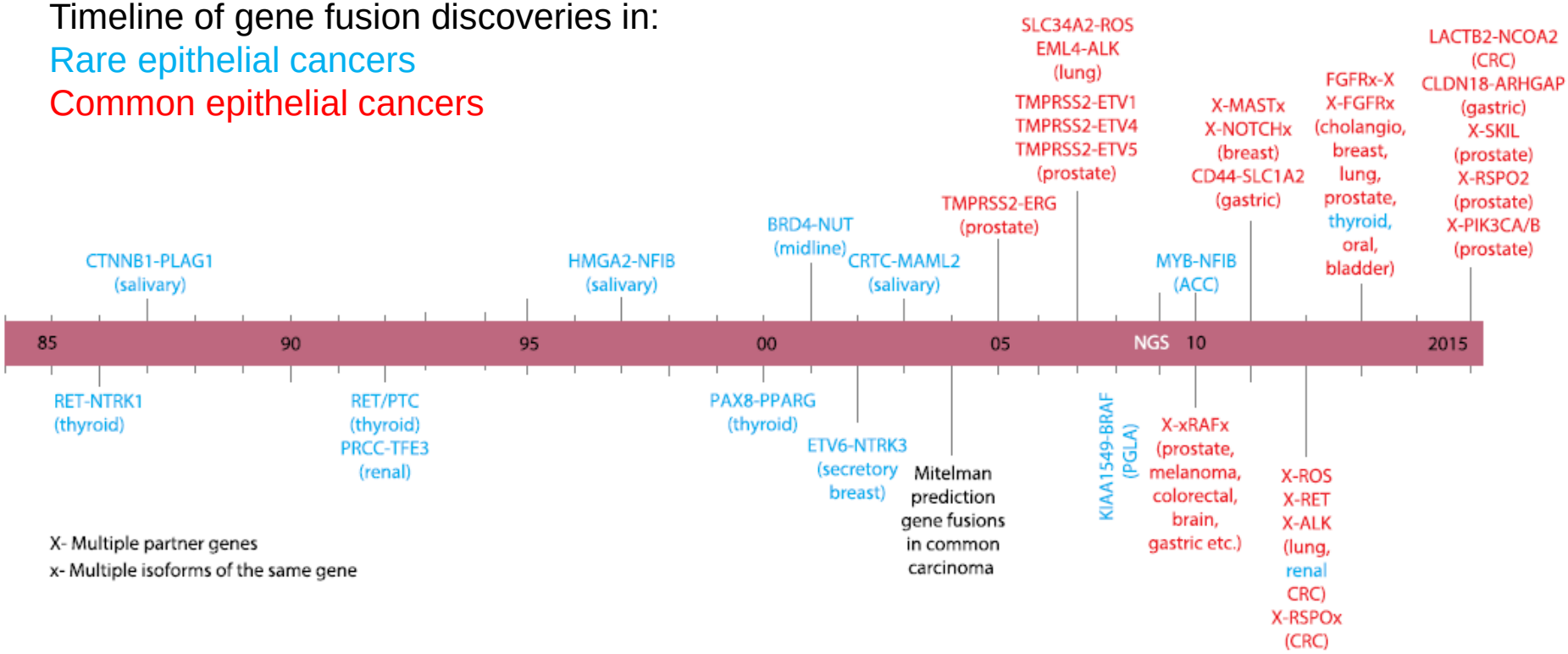


Source: Kumar-Sinha, C., Kalyana-Sundaram, S., Chinnaiyan, A.M. (2015) Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med. **7**, 129.

Why are gene fusions important?

And in rare and common epithelial cancer…

Timeline of gene fusion discoveries in:
Rare epithelial cancers
Common epithelial cancers

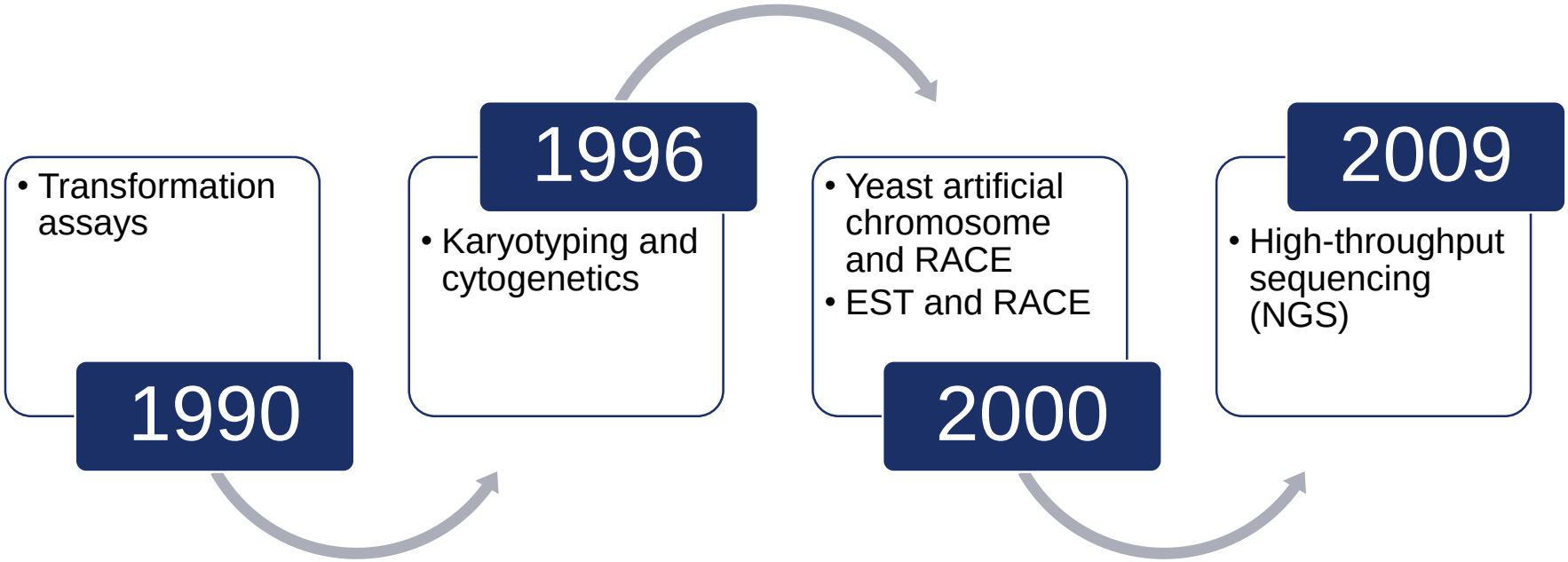


Source: Kumar-Sinha, C., Kalyana-Sundaram, S., Chinnaiyan, A.M. (2015) Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med. **7**, 129.

Agenda

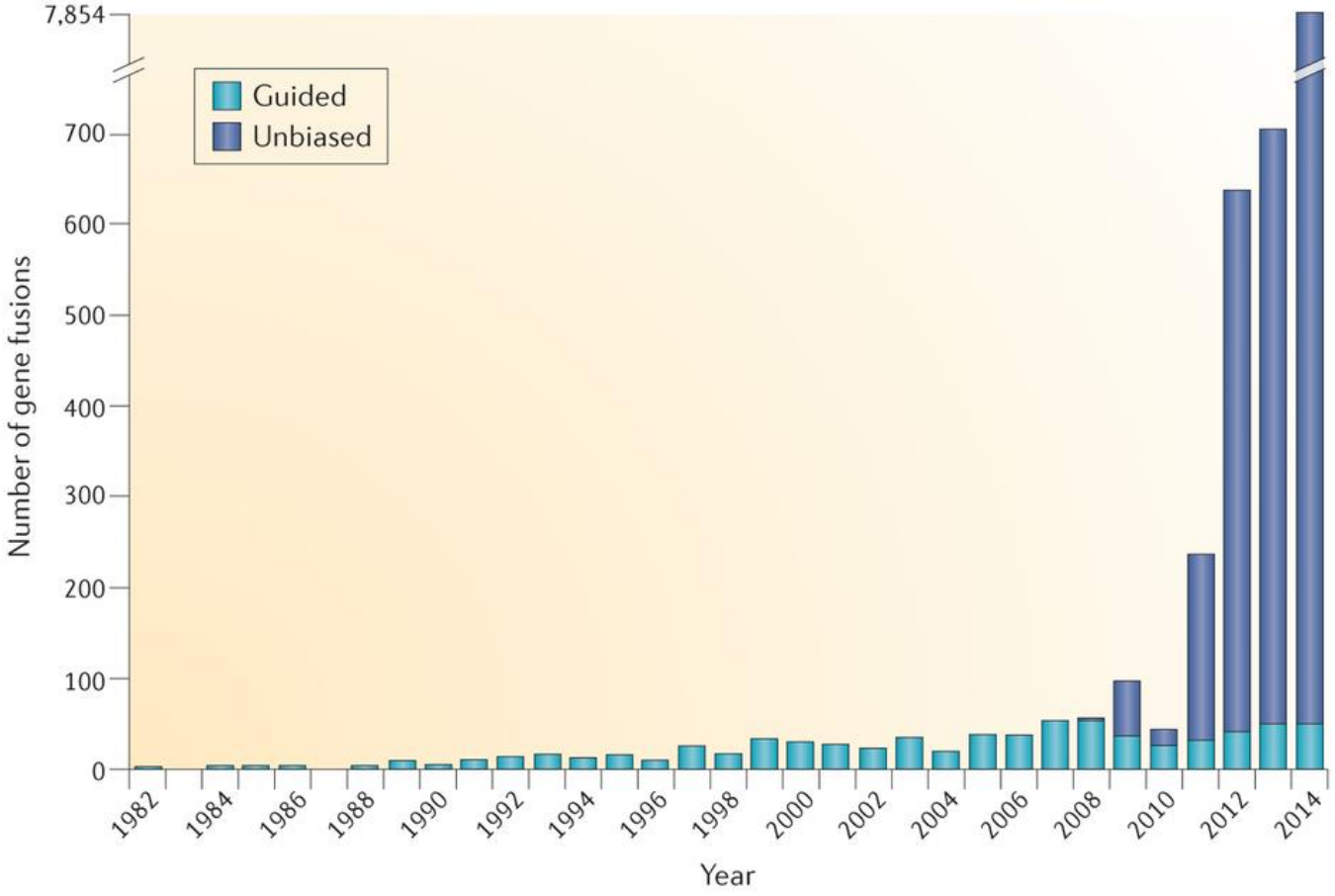
1. Fusion genes: What they are and their historical perspective
2. Fusion gene detection: Current status
3. RNA sequencing vs. digital RNA sequencing
4. How to detect and accurately quantify novel fusion genes in your sample

Evolution of technologies to discover and detect gene fusions



NGS has enabled the discovery of gene fusions

At an unprecedented rapid pace…



Source: Mertens, F., Johansson, B., Fioretos, T., Mitelman, F. (2015) The emerging complexity of gene fusions in cancer. Nat Rev Cancer **15**, 371–81.

Current technologies: Advantages and disadvantages

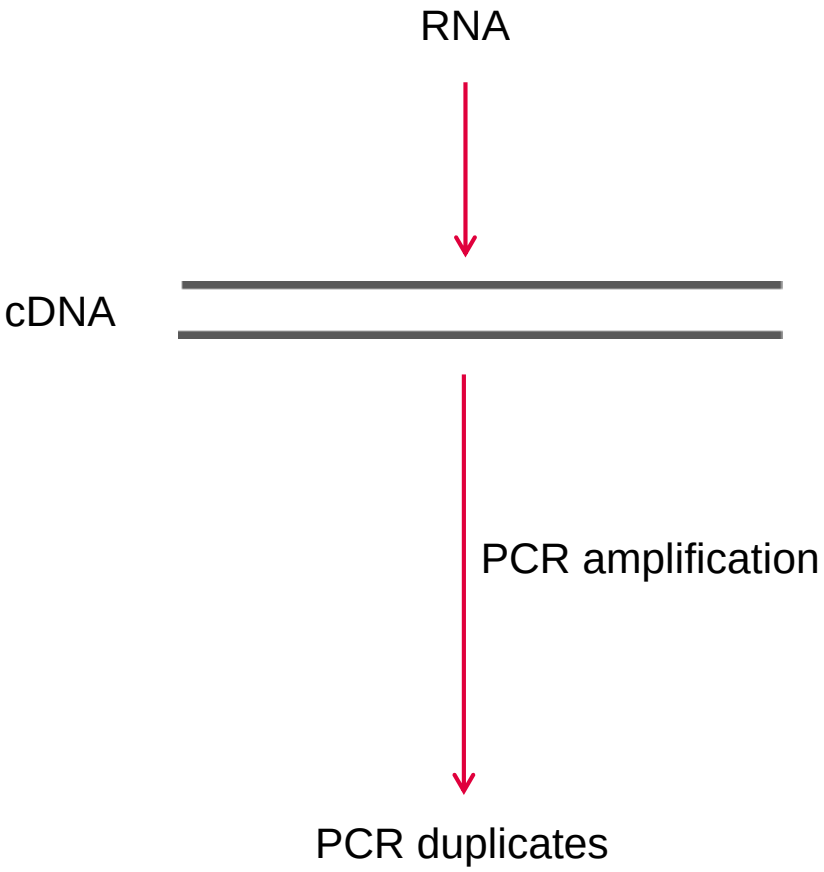
Current technology	Advantages	Disadvantages
PCR-based	Accuracy	Limited sample and assay throughput Requires a lot of RNA
Whole transcriptome sequencing (WTS)	Throughput power	Expensive Complex data
FISH (Fluorescence in situ hybridization)	Established method for routine testing	Laborious
Conventional Targeted RNA sequencing	Manageable data Relatively low per-sample cost	PCR duplicates, which limit the sensitivity of the panel and reduce the ability to confidently detect low-abundance fusions

Agenda

1. Fusion genes: What they are and their historical perspective
2. Fusion gene detection: Current status
3. RNA sequencing vs. digital RNA sequencing
4. How to detect and accurately quantify novel fusion genes in your sample

PCR amplification: The necessary evil

PCR amplification is required for target enrichment, but…



Disadvantages of PCR duplicates

PCR duplicates limit accurate quantification

Conventional targeted RNA sequencing

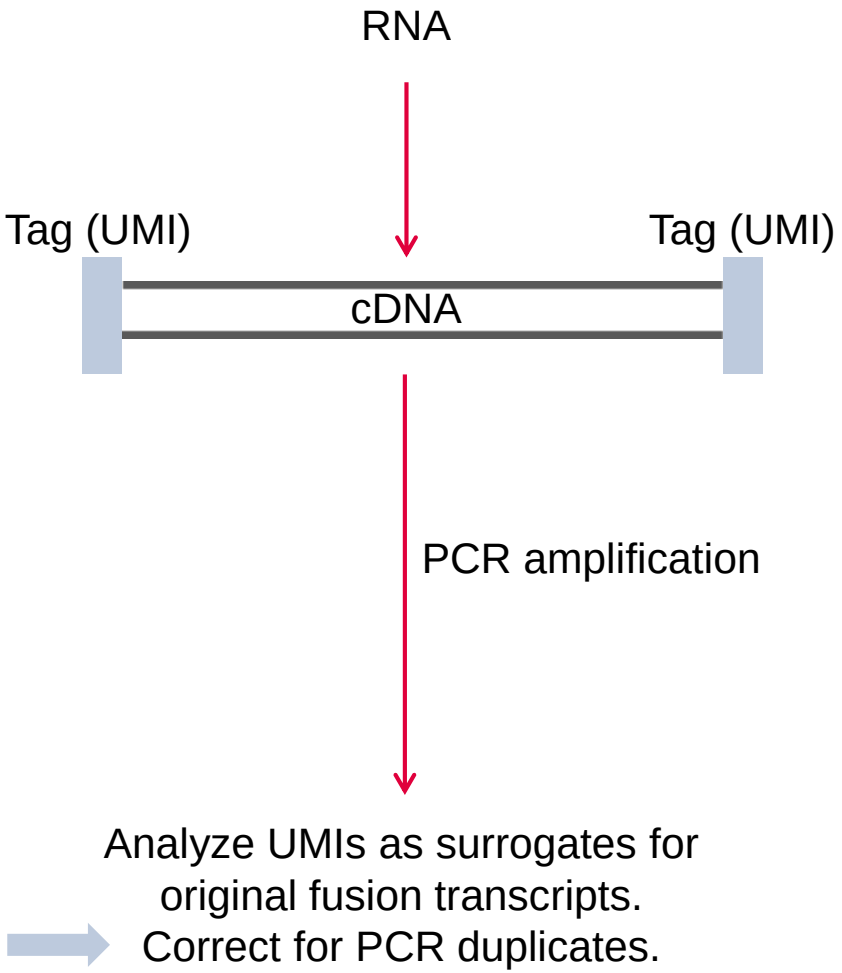
Five library fragments that look exactly the same.
Cannot tell whether they represent:
1. Five **unique** fusion transcripts, or
2. Five **copies** of the **same** fusion transcript **(PCR duplicates**)

Quantification based on non-unique reads does not reflect quantities of original fusion molecules

If you cannot eliminate PCR duplicates, correct for them

How to correct for PCR duplicates?

Capture and freeze original status of fusion transcripts by tagging them before any amplification



What is a Unique Molecular Index (UMI)?

Tag (barcode) to identify unique RNA molecules

TATCGTACAGAT

(12 nucleotides long)

Incorporate this random barcode (signature) into the original RNA molecules (before amplification) to preserve their uniqueness

Achieve accurate quantification with UMIs

Count and analyze single original molecules (not total reads) = digital sequencing

Conventional targeted RNA sequencing

Five library fragments that look exactly the same.
Cannot tell whether they represent:
1. Five **unique** fusion transcripts, or
2. Five **copies** of the **same** fusion transcript **(PCR duplicates**)

UMIs before any amplification

Digital sequencing with UMIs

Five **unique** fusion transcripts since 5 molecular barcodes are detected

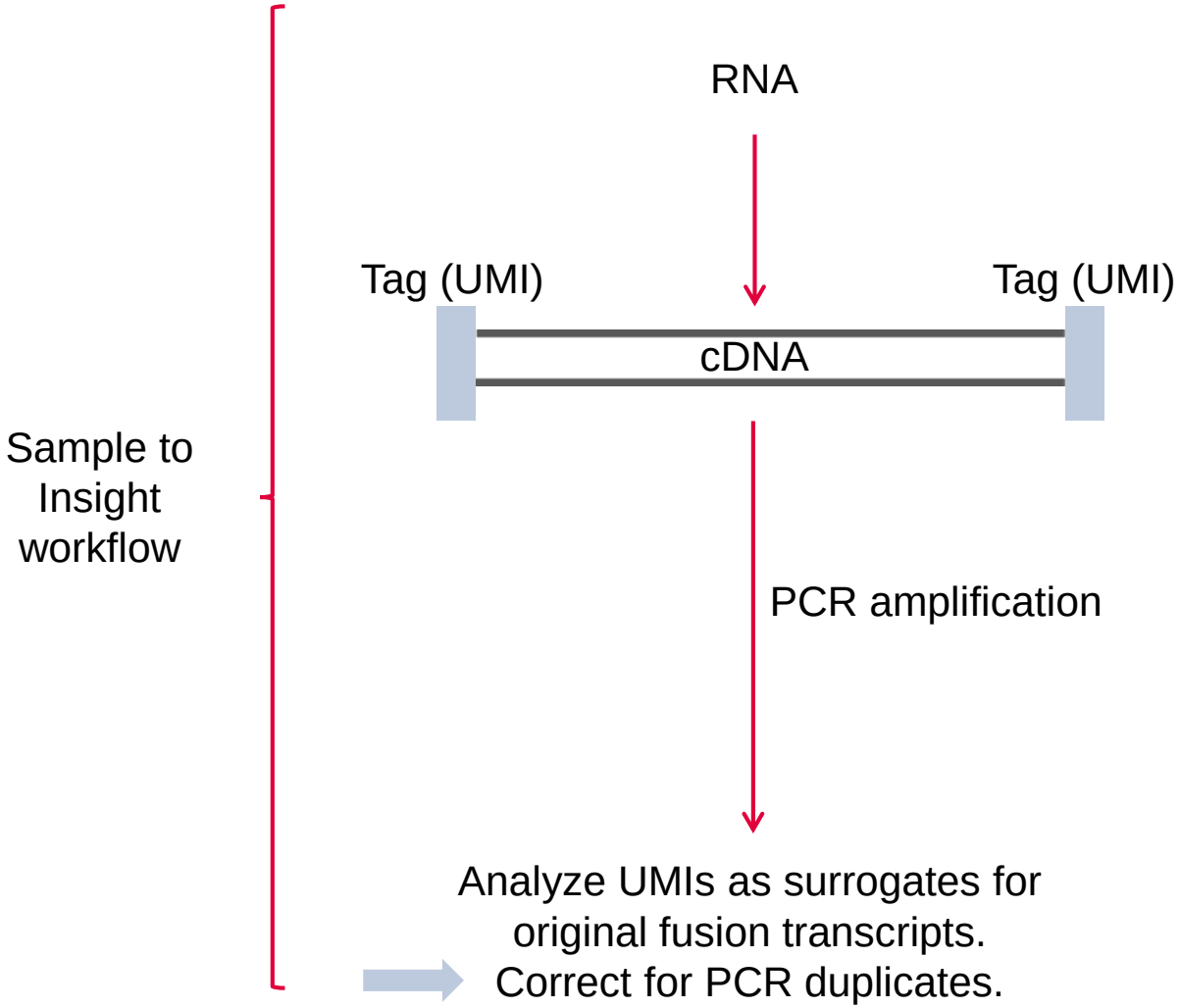
Five **copies** of the **same** fusion transcript (**PCR duplicates**) since 1 molecular barcode is detected

Agenda

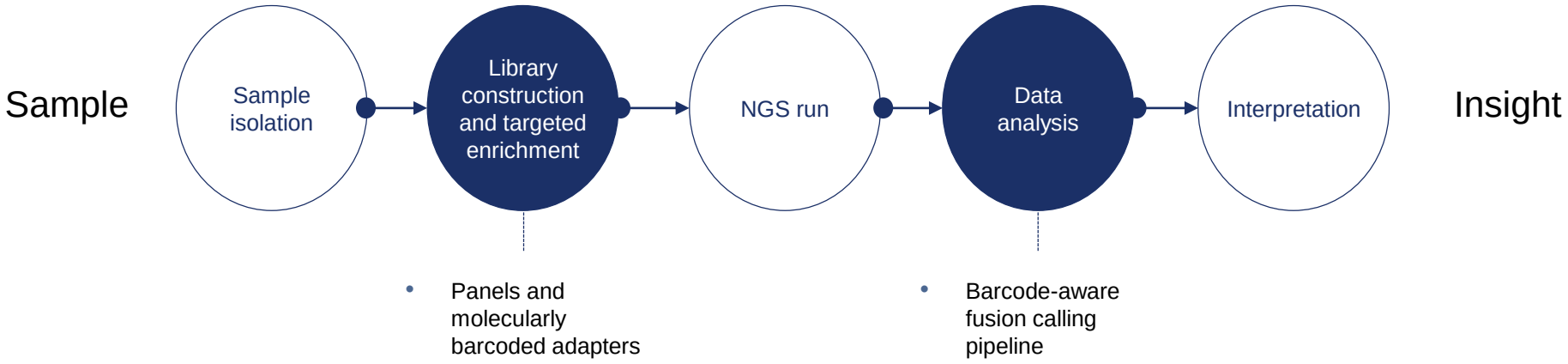
1. Fusion genes: What they are and their historical perspective
2. Fusion gene detection: Current status
3. RNA sequencing vs. digital RNA sequencing
4. How to detect and accurately quantify novel fusion genes in your sample

QIAGEN's solution

QIAseq targeted RNAscan workflow



Panels, molecularly-barcoded (UMI) adapters and data analysis algorithms

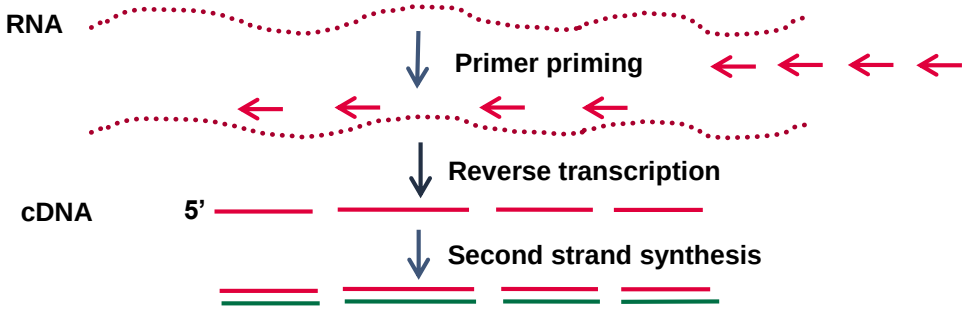


QIAGEN's solution: QIAseq targeted RNAscan Panels

Current technology	Disadvantages	How QIAseq targeted RNAscan panels address disadvantages of current technologies
PCR-based	Limited sample and assay throughput Requires a lot of RNA	Profile hundreds of gene fusions in up to 384 samples simultaneously Requires 10 ng RNA
Whole transcriptome sequencing (WTS)	Expensive Complex data	Cost-effective Easy data analysis
FISH (Fluorescence in situ hybridization)	Laborious	Streamlined, automation-friendly protocol
Conventional Targeted RNA sequencing	PCR duplicates, which reduce the ability to confidently detect low-abundance fusions	Digital sequencing removes PCR duplicates to increase sensitivity of panel, and detect low-abundance fusions

QIAseq Targeted RNAscan Panel: Workflow

As low as 10 ng

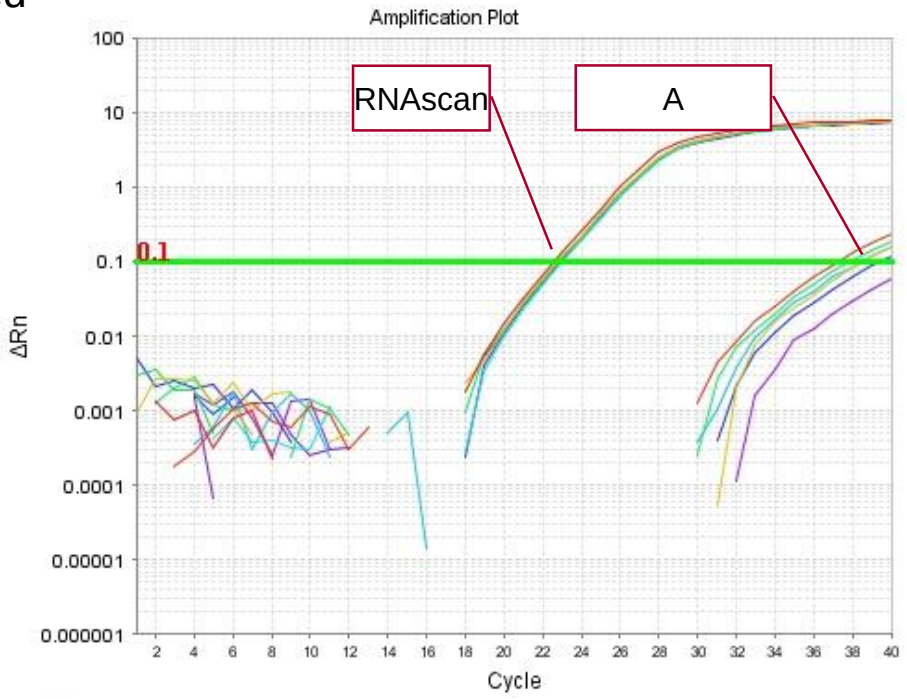


QIAseq RNAscan has higher RNA conversion rate

RNA conversion step is crucial

More RNA converted ⟶ More fusions detected

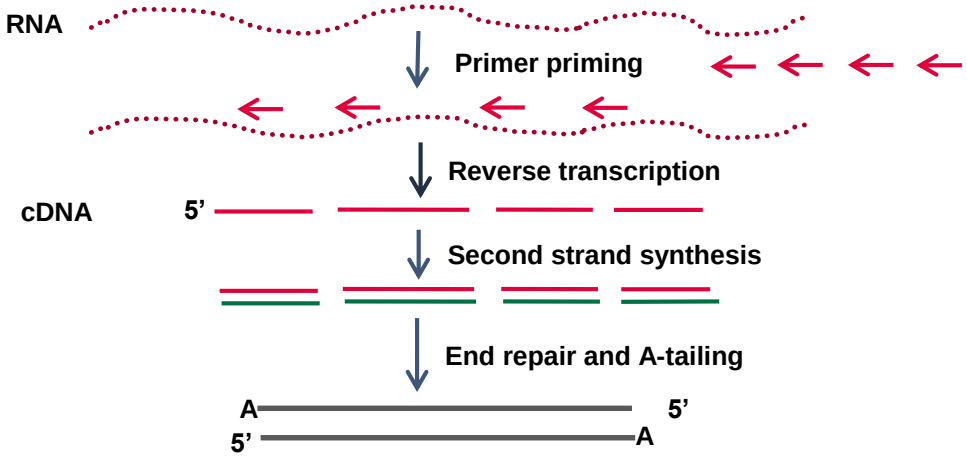
- Test kits:
  - QIAseq Targeted RNAscan panel kit
  - **Competitor A** panel kit
- Sample: For each reaction,10 ng universal RNA spiked with 1 pg of an artificial RNA molecule.
- Follow each kit's standard workflow to generate 2nd strand product.
- Detect converted spike-in RNA with qPCR
- Result showed RNAscan has superior RNA conversion efficiency.
- Higher conversion efficiency means higher sensitivity, especially with low input.



	RNAscan	A
C_T (average+/-SD)	22.8 +/– 0.05	38.5 +/– 0.53

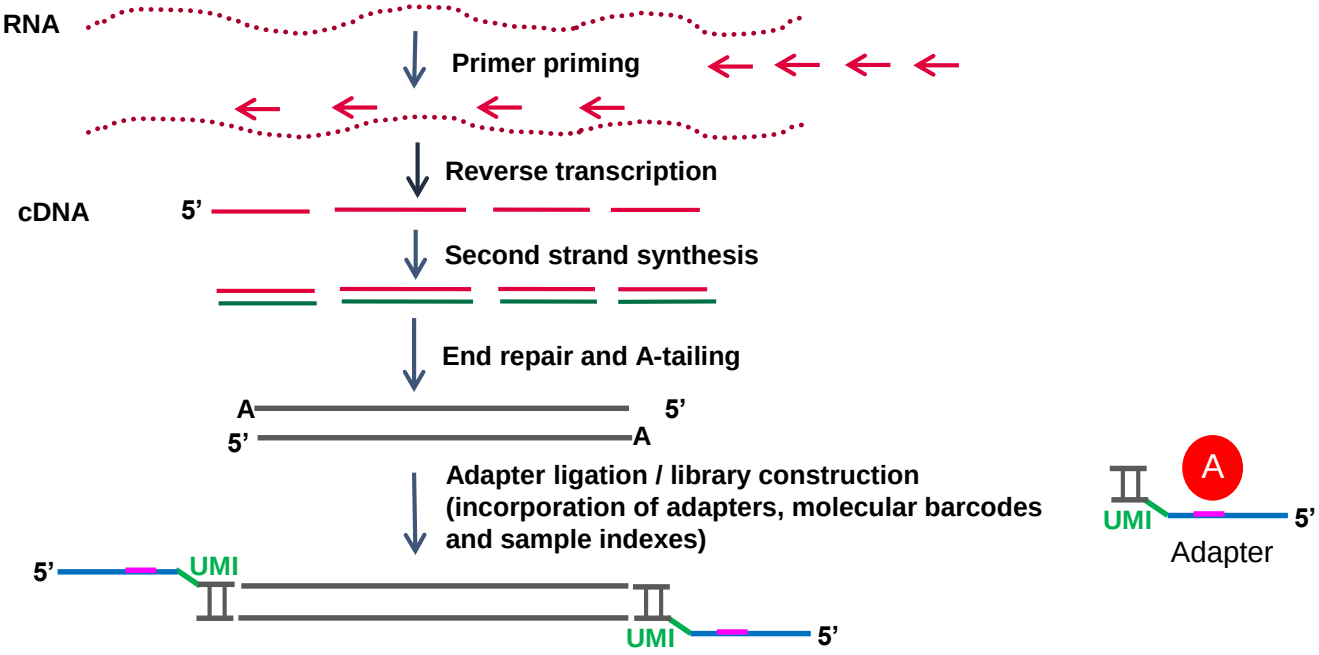
QIAseq Targeted RNAscan Panel: Workflow

As low as 10 ng



QIAseq Targeted RNAscan Panel: Workflow

As low as 10 ng



Achieve accurate quantification with UMIs

Count and analyze single original molecules (not total reads) = digital sequencing

Conventional targeted RNA sequencing

Five library fragments that look exactly the same.
Cannot tell whether they represent:
1. Five **unique** fusion transcripts, or
2. Five **copies** of the **same** fusion transcript (PCR duplicates)

Molecular barcodes
before any amplification

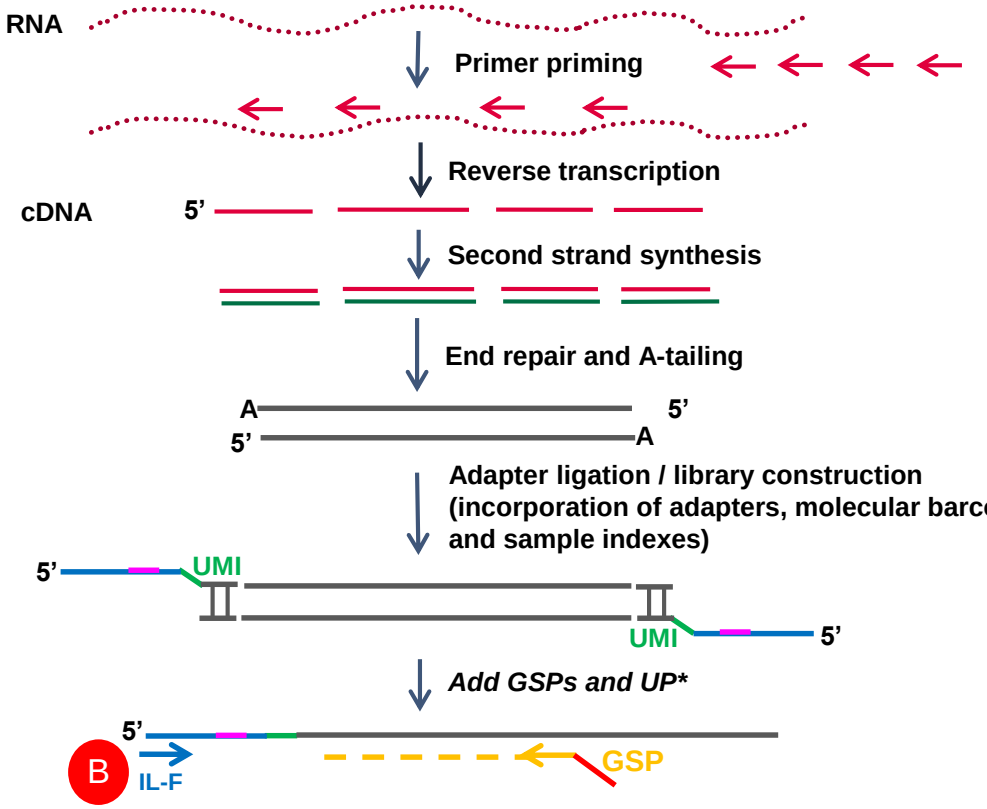
Digital sequencing with UMIs

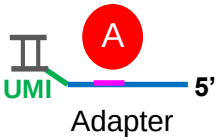
Five **unique** fusion transcripts
since 5 molecular barcodes are detected

Five **copies** of the **same** fusion transcript (PCR duplicates)
since 1 molecular barcode is detected

QIAseq Targeted RNAscan Panel: Workflow

As low as 10 ng

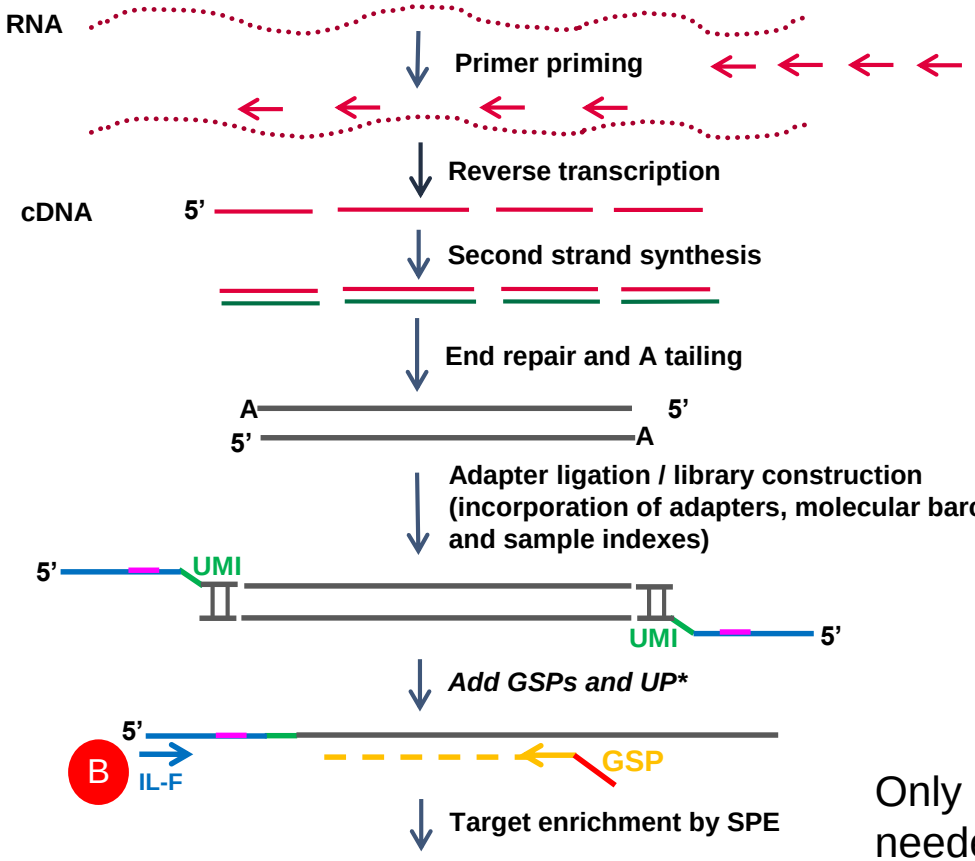


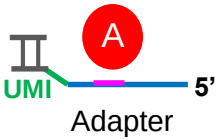


Up to 1000 primers
in a single pool

QIAseq Targeted RNAscan Panel: Workflow (L)

As low as 10 ng





Up to 1000 primers in a single pool

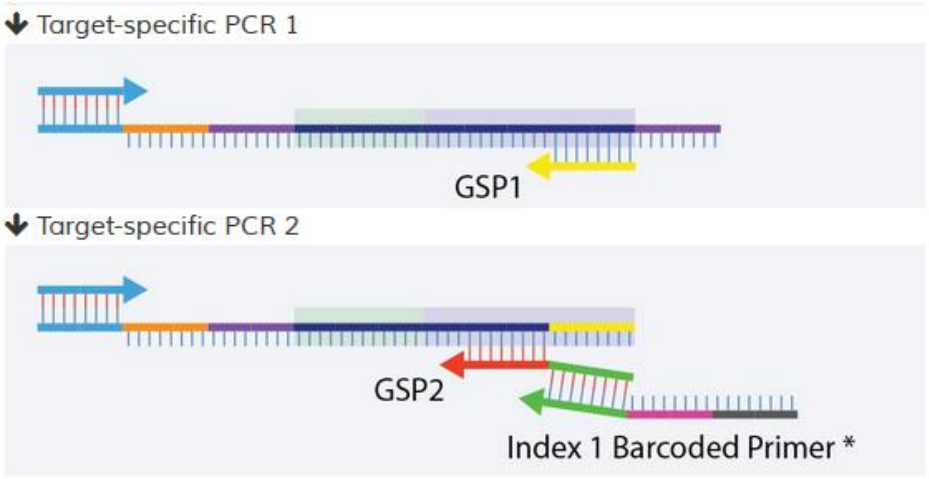
Only one primer is needed

Advantage of Single Primer Extension (SPE)

SPE

No need for long fragments to give space for two primers apart from the fusion junction, so shorter fragments are better tolerated. This approach will help deal with FFPE sample challenges where many RNA fragments may not have enough length to cover both GS1 and GS2 sequence in competitor A panels. Competitor A panels might require longer fragments, which will reduce the efficiency of PCR and the crossing of spanning reads across fusion junctions.

Nested PCR



One gene/target-specific primer	Two gene/target-specific primers

- Flexibility in primer design
- Generate relatively small library fragments to maintain compatibility with fragmented RNA (FFPE samples)

Design based on junction

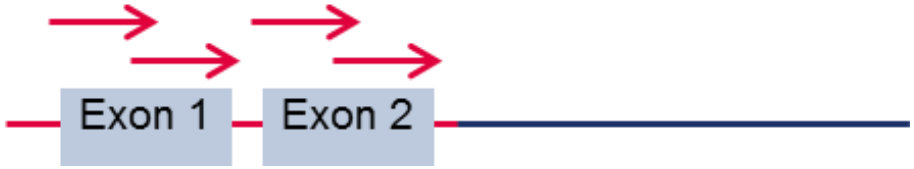
Primers target both ends of a fusion

fusion pair	5' chr	start	end	strand	3' chr	start	end	strand	ENSTs and exons
ETV6-NTRK3	12	11853560	11853561	+	15	87940752	87940753	–	ETV6 (ENST00000396373) exon 4 to NTRK3 (ENST00000394480) exon 15

Primer design approaches

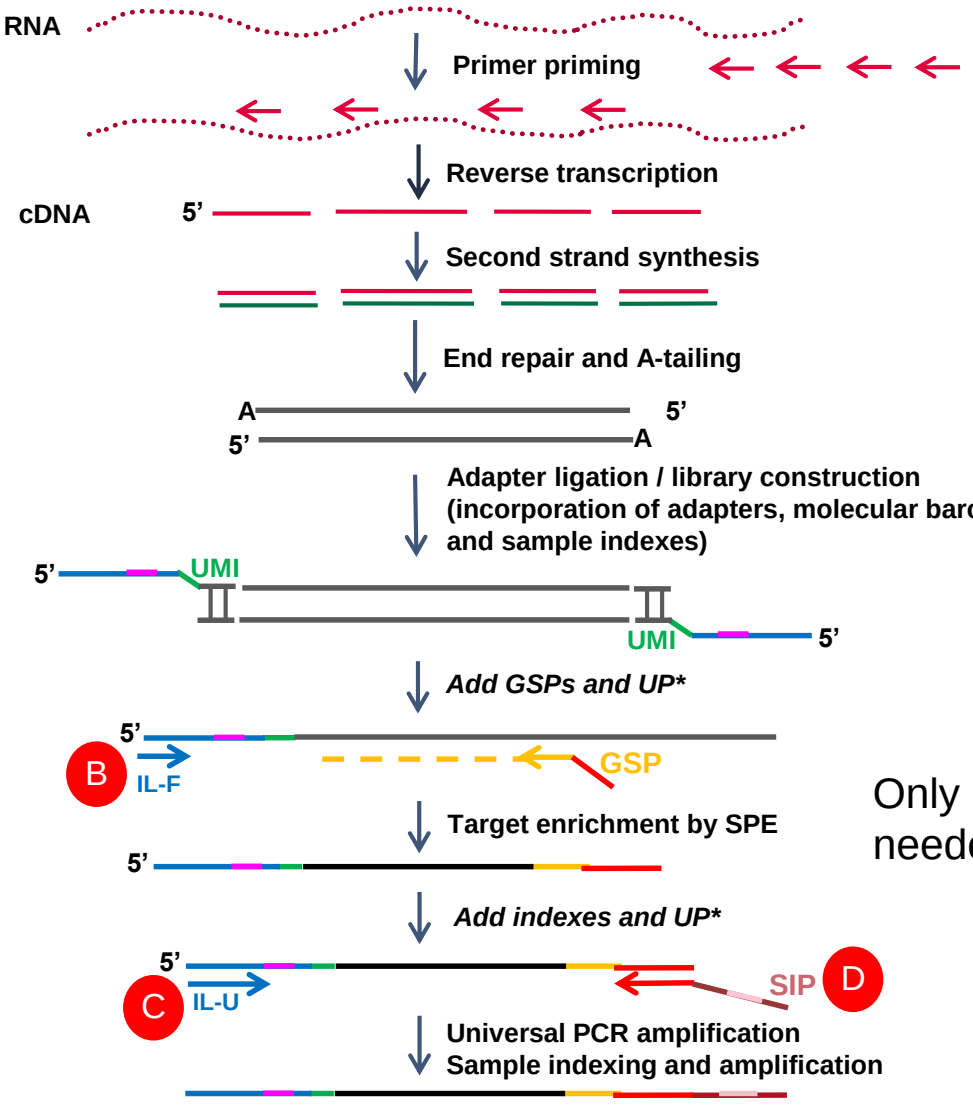
Exon- or gene-based design for unknown partners

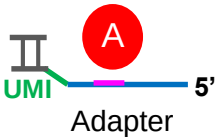
Primers target one end of a fusion



QIAseq Targeted RNAscan Panel: Workflow (I)

As low as 10 ng



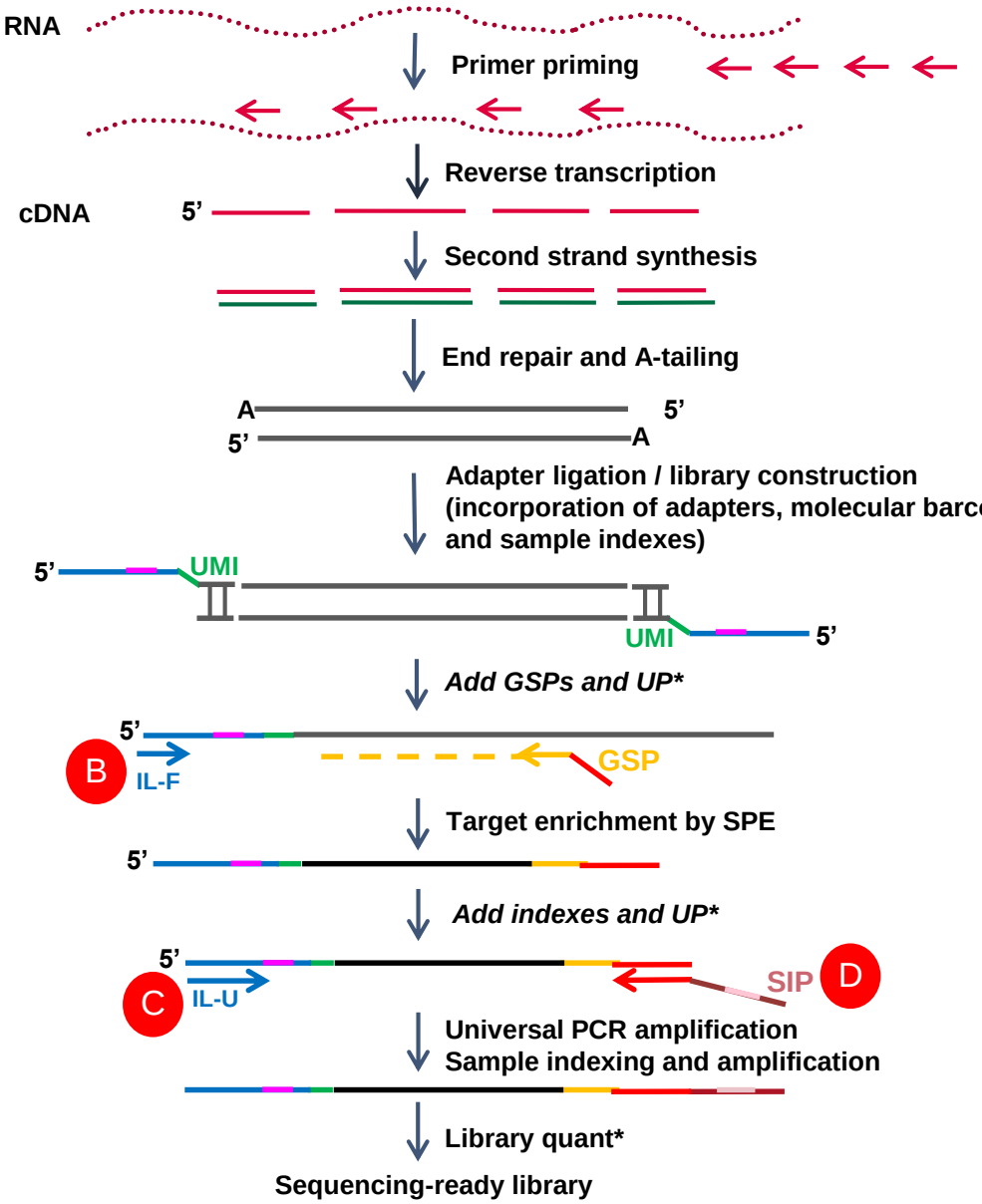


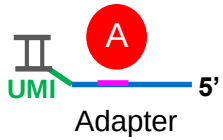
- Up to 1000 primers in a single pool
- Dual sample indexing system; up to 384 samples

Only one primer is needed

QIAseq Targeted RNAscan Panel: Workflow (I)

8–9 hours

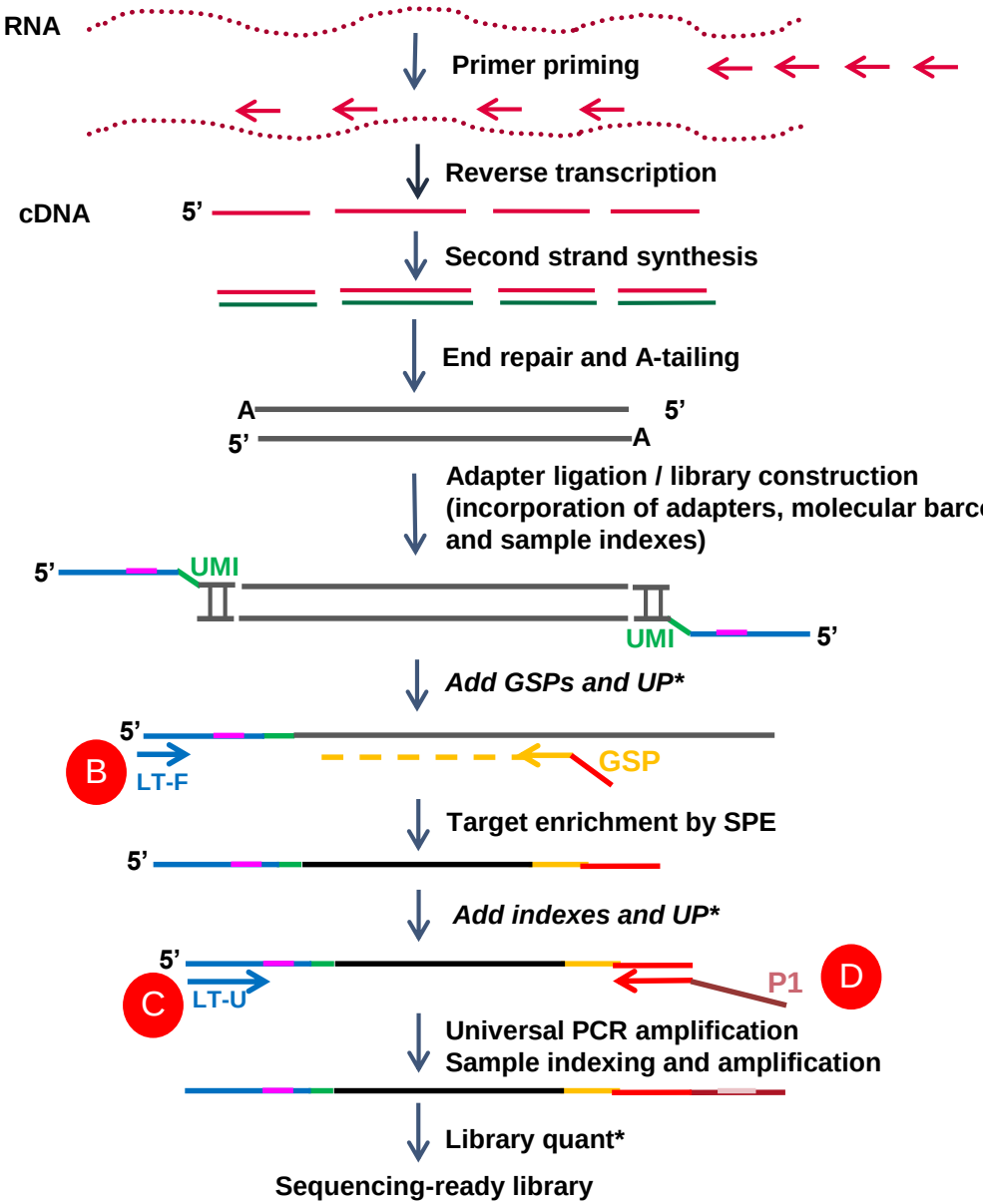


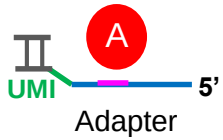


MB: Molecular barcode
GSP: Gene specific primer
IL-F: Forward primer
IL-U: Universal primer
SIP: Sample index primer
* Preceded by bead cleanup

QIAseq Targeted RNAscan Panel: Workflow (L)

8–9 hours





MB: Molecular barcode
GSP: Gene specific primer
LT-F: Forward primer
LT-U: Universal primer
P1: P1 adapter sequence
* Preceded by bead cleanup

Features and benefits of QIAseq targeted RNAscan panels

Feature	Benefit
Low RNA input (as low as 10 ng of unenriched RNA)	Preserve sample
Unique RNA conversion chemistry	Convert more RNA to cDNA to detect more fusions
Unique molecular indices (UMIs)	Enables single molecule counting, removal of PCR duplicates and error correction; these features correct for amplification bias, distinguish clonal replicates and deliver absolute measurements of fusion transcripts
High primer multiplexing capability (up to 1000 primers)	Detects a large number of fusions
Single pool of primers	Easier sample handling
SPE-based target enrichment	Flexibility in primer design Generate relatively small library fragments to maintain compatibility with fragmented RNA (FFPE samples)
High sample multiplexing (up to 384 samples)	Increased sample throughput to decrease sequencing costs
No prior knowledge of breakpoint needed	Discover novel partners
Automation-friendly workflow	Streamline operations for high throughput
Suite of complementary data analysis tools	Save resources
Very cost-effective compared to competition	Save $$$

Simplified logistics

Two boxes are needed to support the workflow

Panel box (kit)

- Reagents for cDNA synthesis
- Fusion-specific primers for enrichment
- All required buffers and enzymes
- Magnetic beads

Index box (kit)

- Molecularly-barcoded library adapters, primers to prepare libraries that are sample indexed and sequencing platform-specific

Product offerings

At a glance

Panels

- RNAscan
  - Cataloged
  - Custom
  - Extended

Indices

- Illumina
  - 12-index
  - 96-index (4 sets, for up to 384-plex)
- Ion Torrent
  - 12-index
  - 96-index

QIAseq Targeted RNAscan Panels

List of panels

Panel	Variant (cat) number	Number of primers	Types of coverage
Hematology panel	FHS-001Z	156	Breakpoint
Solid tumor panel	FHS-002Z	101	Breakpoint
Lung cancer panel	FHS-003Z	137	Breakpoint
Oncology panel	FHS-3001Z	950	Breakpoint

Primers target both ends of a fusion

QIAseq indexes

For QIAseq targeted RNAscan panels

Cat ID	Name	Sequencing platform	Number of samples per sequencing run that can be multiplexed	Number of samples each kit (SAP ID) can process
333714	QIAseq 12-Index I (48)	Illumina	12 samples per sequencing run	48 samples
333727	QIAseq 96-Index I Set A (384)	Illumina	96 samples per sequencing run	384 samples
333737	QIAseq 96-Index I Set B (384)	Illumina	96 samples per sequencing run (192 if used with Set A)	384 samples
333747	QIAseq 96-Index I Set C (384)	Illumina	96 samples per sequencing run (288 if used with Sets A, B)	384 samples
333757	QIAseq 96-Index I Set D (384)	Illumina	96 samples per sequencing run (384 if used with Sets A, B, C)	384 samples
333764	QIAseq 12-Index L (48)	Ion Torrent	12 samples per sequencing run	48 samples
333777	QIAseq 96-Index L (384)	Ion Torrent	96 samples per sequencing run	384 samples

QIAseq Targeted RNAscan: Complete fusion call

Application: Proof of sensitivity and specificity of the QIAseq Targeted RNAscan panels

- Engineered positive control for 12 popular fusions as testing material
- QIAseq Targeted RNAscan panel was designed based on fusion‘s breakpoint and sequence used from both genes
- Control FFPE sample was used with 15 ng as starting material; same amount of normal universal RNA was used as negative control
- Library was run with sequencer MiSeq
- Fusion was detected with current QIAseq Targeted RNAscan data analysis pipeline

Fusion Partners included

	RNA Fusion	Primary Cancer Tissue
1	EML4-ALK	Lung
2	NPM1-ALK	Lymphoid
3	KIF5B-RET	Lung
4	NCOA4-RET	Thyroid
5	CD74-ROS1	Lung
6	SLC34A2-ROS1	Lung, Stomach
7	TPM3-NTRK1	Lung, Large Intestine
8	TFG-NTRK1	Thyroid (rare)
9	FGFR3-BAIAP2L1	Urinary tract (rare)
10	FGFR3-TACC3	Urinary tract, CNS
11	PAX8-PPARG	Thyroid
12	ETV6-NTRK3	Kidney, Breast, Soft Tissue

RNAscan: High sensitivity and specificity

100% sensitivity (12/12 positive), 0% positive

Fusion in sample	UMIs for detected crossing-junction reads
EML4-ALK	18
NPM1-ALK	5
KIF5B-RET	25
NCOA4-RET	46
CD74-ROS1	11
SLC34A2-ROS1	8
TPM3-NTRK1	46
TFG-NTRK1	5
FGFR3-BAIAP2L1	37
FGFR3-TACC3	60
PAX8-PPARG	28
ETV6-NTRK3	30

All fusions in control were detected with at least 5 MTs

RNAscan: Report of fusion with details

Example of positive fusion detected: In positive controls, FGFR3-TACC3 fusion is one the 12 positive fusions

FGFR3--TACC3

Summary

Role	Gene(s)	Donor			Acceptor			Links	Crossing		Spanning	
		Chr	Position	Strand	Chr	Position	Strand		Reads	Tags	Reads	Tags
Donor	FGFR3	chr4	1,806,935	+	chr4	1,807,115	+	1	32	**15**	0	**0**
Acceptor	TACC3	chr4	1,739,701	+	chr4	1,739,701	+	2	720	**459**	0	**0**
Fusion	FGFR3--TACC3	chr4	1,806,935	+	chr4	1,739,701	+	1	99	**60**	0	**0**

QIAseq targeted RNAscan panels: Unique value propositions (UVPs)

Confidently detect low-abundance fusions

Molecular barcodes

- Correct for PCR duplicates
- Unmatched sensitivity

Flexibility to detect known and novel fusions

Single primer extension and powerful pipelines

- Only one of the two targets need to be known/targeted
- Targets are defined based on breakpoint, exons or genes

QIAseq targeted RNAscan panels: Unique value propositions (UVPs)

Completely integrated and optimized workflow: From Sample to Insight

Digital sequencing with molecular barcodes

- Remove PCR duplicates and errors for accurate quantification

Barcode-aware, comprehensive data analysis pipelines

- Confidently call low-abundance fusion transcripts and identify novel variants

Single primer extension approach

- No need for nested PCR primer design; ability to detect more fusions

Throughput flexibility

- Molecularly profile hundreds of fusion transcripts in hundreds of samples, simultaneously

Compatibility with fragmented or preserved RNA

- Sequence RNA from FFPE samples or liquid biopsy samples

Fast custom design and manufacturing turnaround time

- Receive custom panel or extend an existing panel in 2–3 weeks.

Data analysis with barcode-aware fusion caller – overview

- Barcode-aware fusion caller has been developed
- Caller is available on the cloud
- In conjunction with molecular barcodes incorporated in the workflow, the caller can confidently call low-abundance fusions (down to 1% expression level)
- Fusion caller will do the following:
  - Mapping
  - Alignment
  - Molecular barcode counting
  - Fusion calling
  - Fusion annotation – based on internal curated database

Data analysis with barcode-aware fusion caller – overview

Inputs

FASTQ or BAM files are uploaded into cloud-based data analysis portal

The following inputs are needed (by user):

- Set up file
- Panel used
- File lanes
  - 1-lane (MiSeq/HiSeq/NextSeq concatenated)
  - 4-lane (NextSeq individual lane files)

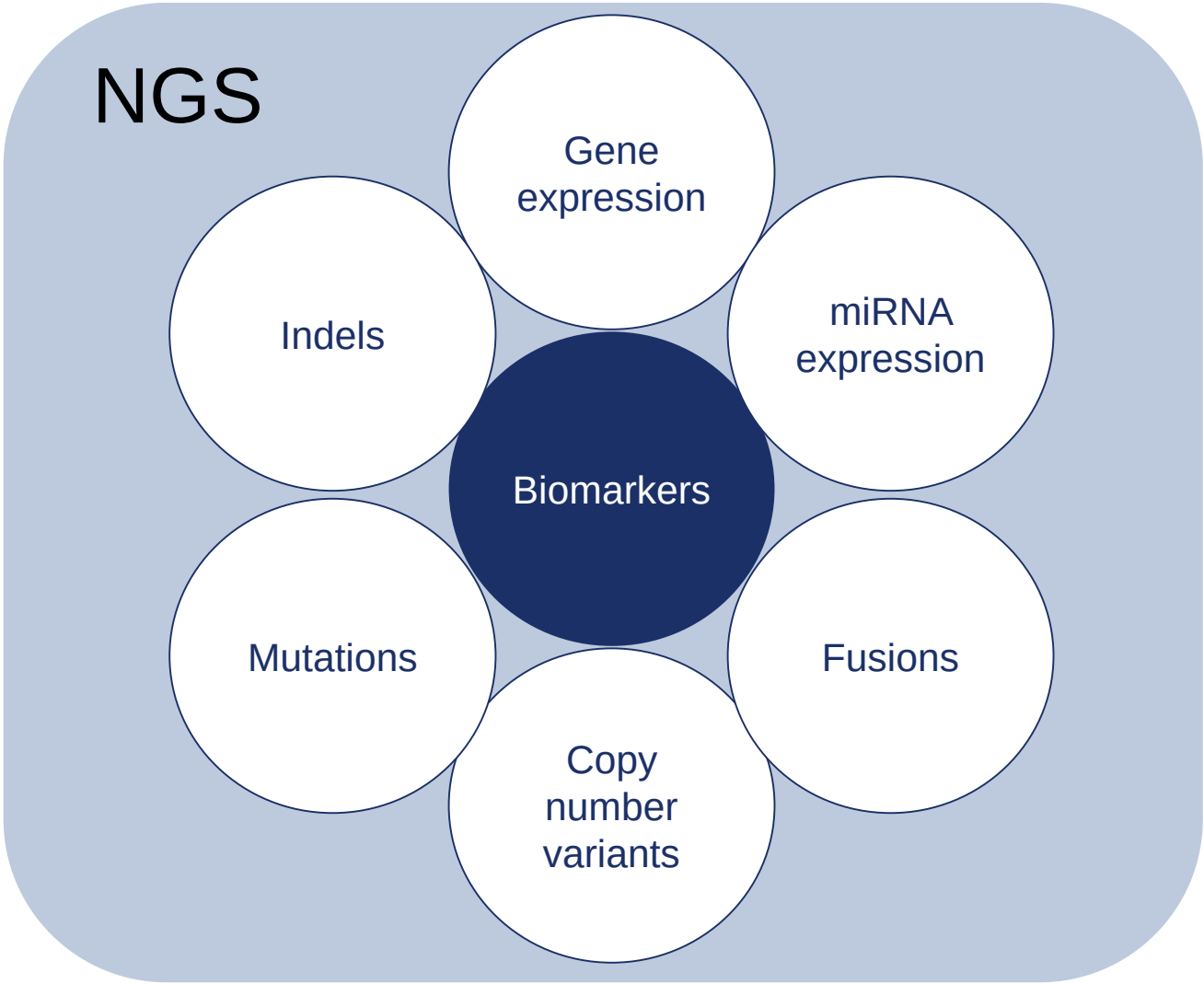
Data analysis with barcode-aware fusion caller – overview

Outputs

Summary file

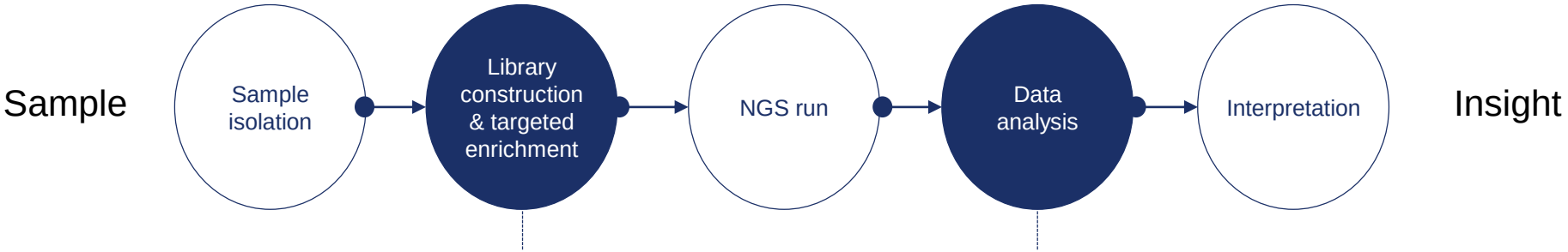
- Stats
  - Specificity
  - Uniformity
  - Molecular barcode counts
- Fusions
  - Curated/novel/no reference available
  - Statistical confidence
  - Graphical representations

NGS can be used to detect all biomarkers



Plug & Play: Many flavors, same Sample to Insight workflow

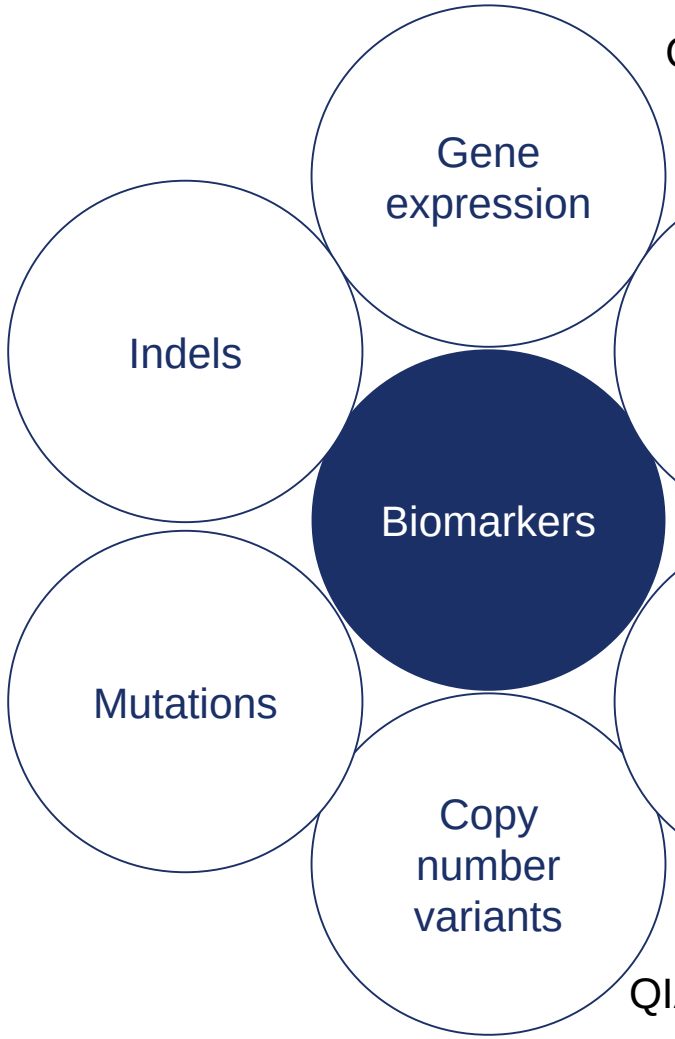
Different panels can be plugged into the same targeted NGS workflow



- DNA panels for variant analysis
- RNA panels for differential gene expression
- RNA panels for fusion profiling
- miRNome panel for miRNA expression

- Mutations, indels, copy number variants
- Gene expression levels
- Fusions
- miRNA levels

QIAseq solutions to detect all biomarkers using NGS



QIAseq targeted RNA panels

QIAseq targeted DNA panels

QIAseq miRNA sequencing system

QIAseq targeted DNA panels

QIAseq targeted RNAscan panels

QIAseq targeted DNA panels

Thank you for attending!

Questions?

Contact QIAGEN

Call: 1-800-426-8157

Email:

techservice-na@QAGEN.com

BRCsupport@QIAGEN.com

QIAseq.NGS@QIAGEN.com

QIAWebinars@QIAGEN.com