

# Ultra-sensitive variant calling and transcript quantification using Unique Molecular Identifiers

## Archive

[August 2020](#)

---

[July 2020](#)

---

[June 2020](#)

---

[May 2020](#)

---

[March 2020](#)

---

[February 2020](#)

---

[January 2020](#)

---

[November 2019](#)

---

[March 2018](#)

---

[January 2018](#)

---

[December 2017](#)

---

[November 2017](#)

---

[October 2017](#)

---

[September 2017](#)

---

PANDURANG KOLEKAR JANUARY 23, 2018  
AVADIS NGS, DNA-SEQ, NEXT GENERATION  
SEQUENCING, NGS, RNA-SEQ, STRAND  
NGS ASHG, TRANSCRIPT QUANTIFICATION, UMI,  
UNIQUE MOLECULAR IDENTIFIER, VARIANT  
CALLING

## Background

Next Generation Sequencing (NGS) technologies have remarkably revolutionized the medical and genomics research. The incremental cost reductions and size of the throughput at molecular resolution helped penetration and acceptance of

the NGS methodologies into worldwide labs and clinics. The third generation wave of NGS technologies are knocking the doors to provide impetus to the dream of preventive, predictive, personalized, and precision (P4) medicine initiative.

At the core of NGS technologies lies the fine tuned, optimised, sensitive molecular biology and chemistry protocols, which helps to accurately snapshot the response of cells at molecular resolution under varying genotypic conditions and environmental impacts. To enable the understanding of genotype-phenotype relationships, accurate quantification of sequenced reads plays a key role before arriving at conclusions and deriving actionable insights from the NGS data.

## Challenges and opportunities

An accurate quantification of NGS data requires discriminating PCR duplicate reads from identical molecules that are of unique origin.

Computationally, PCR duplicates are identified as sequence reads that align to the same genomic coordinates using reference genome guided

August 2017

---

July 2017

---

April 2017

---

February 2017

---

January 2017

---

October 2016

---

September 2016

---

August 2016

---

July 2016

---

June 2016

---

April 2016

---

February 2016

---

September 2015

---

August 2015

---

June 2015

---

January 2015

---

October 2014

---

alignment. However, identical molecules can be independently generated during library preparation and can have unique cellular origins. Thus, false identification of these molecules as PCR duplicates can lead to erroneous analysis and interpretation of NGS data.

On the other hand, it is unclear how much noise or bias PCR amplification introduces and its effect on accuracy of quantification. Generally RNA-Seq methods work with small starting amounts of RNA that require PCR amplification to generate sequenceable sized libraries. To assess the effects of amplification, reads that originated from the same RNA molecule (PCR-duplicates) need to be identified.

In case of variant calling methods, sequencing errors makes it further difficult to distinguish actual variant calls from the sequencing artifacts. This problem is more prominent during the detection of low frequency somatic variations, which are expected to be called in liquid biopsy samples, wherein the proportion of circulating tumor DNA (ctDNA) is very low as compared to normal cell free DNA (cfDNA). Accurate detection of such variants holds key in early stage cancer

---

September 2014

---

August 2014

---

February 2014

---

January 2014

---

December 2013

---

August 2013

---

April 2013

---

June 2012

---

April 2012

---

March 2012

---

February 2012

---

January 2012

---

December 2011

---

November 2011

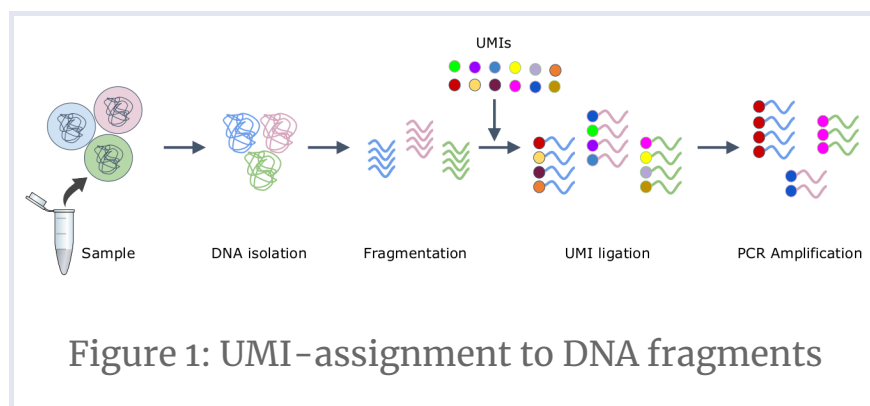
diagnosis and monitoring.

# Unique Molecular Identifiers

In order to overcome these challenges, protocols offering assignment of Unique Molecular Identifiers (UMIs) to DNA molecules while preparing NGS sequencing libraries are gaining wide attention (Figure 1). Unique Molecular Identifiers (UMIs) are short random nucleotide sequences (4 – 20 bases) and are increasingly being used in high-throughput next generation sequencing (NGS) experiments to distinguish individual DNA/cDNA molecules <sup>1, 2</sup>.

Following terms are synonymous with UMIs:

Unique Molecular Tags (UMTs)  
Random Molecular Tags (RMTs)  
Molecular Barcode



*Note: Click on image to enlarge.*

The UMI-tagged NGS data allow users to 1) accurately quantify the expression levels of genes in different cells using single cell RNA-Seq experiments and 2) detect low frequency variants with better sensitivity and specificity using UMI based DNA-Seq experiments (Figure 2 and 3). These applications of UMI are of particular interest in areas like liquid biopsy based cancer diagnosis and monitoring using cfDNA<sup>3</sup>, differential expression of transcriptome at cellular levels instead of a tissue to study cell-to-cell heterogeneity<sup>4</sup> etc.

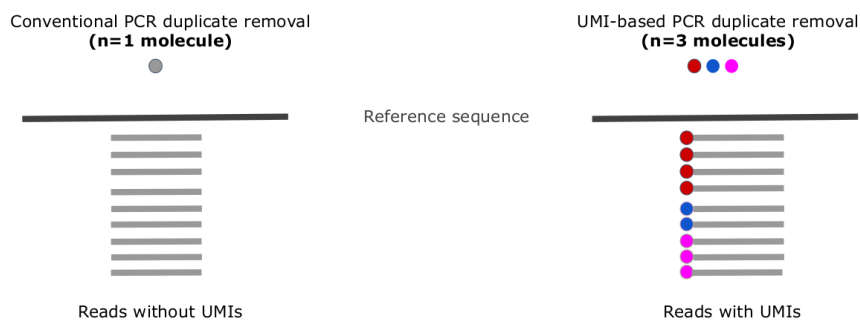
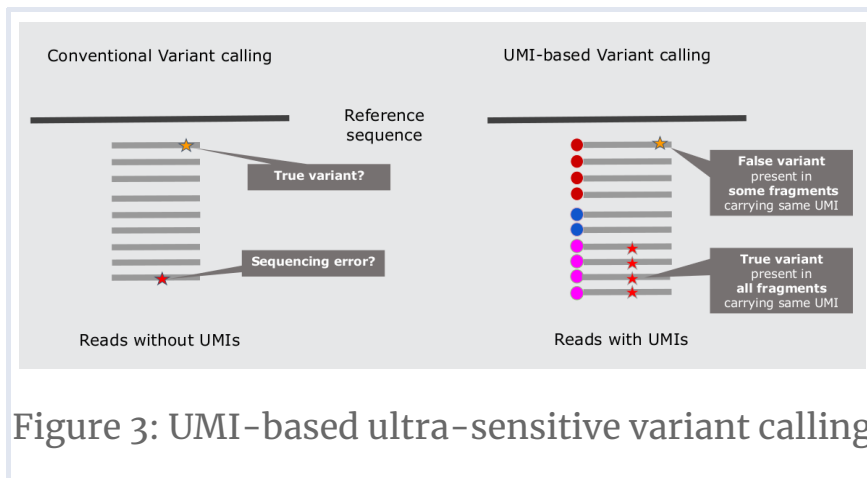


Figure 2: UMI-based PCR duplicate removal and accurate read quantification

*Note: Click on image to enlarge.*



*Note: Click on image to enlarge.*

However, there is hardly any Bioinformatics software which provides “end-to-end solution” supporting UMI-aware custom data import, QC metrics, consensus alignment, quantification, variant calling, and a genome browser to explore and analyse UMI-tagged NGS data.

## Bioinformatics Solution

Strand NGS v3.1 which was released during ASHG 2017 <sup>5, 6</sup> supports the necessary features to explore, analyze and visualize the UMI-tagged NGS data allowing researchers to harness the potential of big data to gain deeper insights.

In case of open source pipelines, users have to

make use of multiple third party software and scripts to perform alignment, filtering, and post-processing of BAM files. And for variant inspection users have to use another software to visualize and verify the variant reads. Most of the open source software also require users to have basic understanding and knowledge of scripting and linux operating systems to run command line tools.

Strand NGS, on the other hand provides user-friendly end-to-end solutions for all the necessary steps for data import, pre-processing, alignment, filtering, variant calling, and genome browser based visualizations and variant verification (Figure 4).

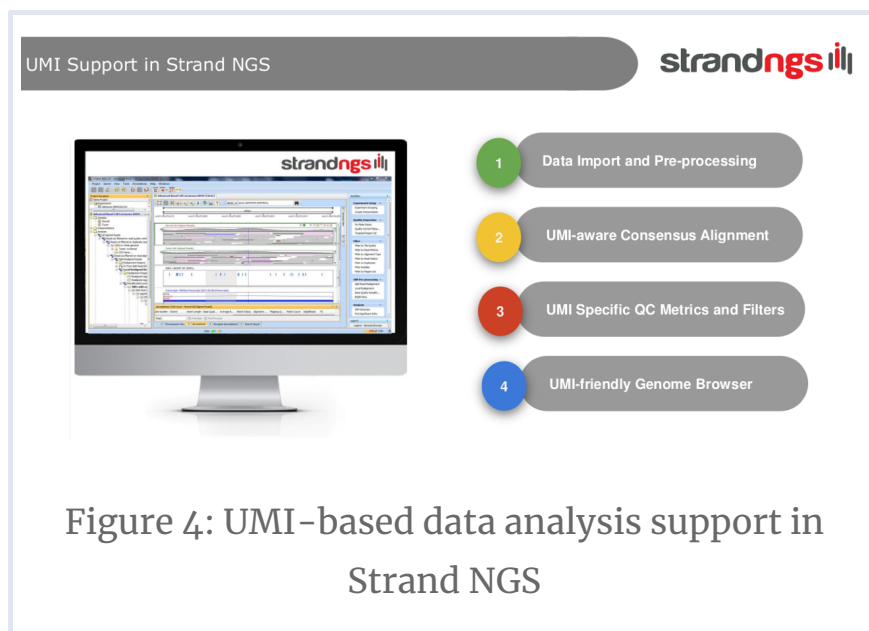


Figure 4: UMI-based data analysis support in Strand NGS

*Note: Click on image to enlarge.*

More wonderful features are in pipeline to empower researchers in the journey from reads to discoveries.

## Webinar and resources

Recently, we delivered a webinar on “Unique Molecular Identifier-powered Ultra-sensitive Variant Calling using Strand NGS”. The recording for the webinar is available at <http://www.strand-ngs.com/learn/webinar-recordings>. Please feel free to share it with your friends, colleagues and connections, who might be interested in this topic and recent trends in NGS.

Visit the Strand NGS website at <http://www.strand-ngs.com/> to learn more about NGS data analyses with Strand NGS through recorded webinars, tutorials and reference manuals.

## References

1. Kivioja et al. Nature Methods 9, 72–74 (2012) doi:[10.1038/nmeth.1778](https://doi.org/10.1038/nmeth.1778)
2. Islam et al. Nature Methods 11, 163–166 (2014) doi:[10.1038/nmeth.2772](https://doi.org/10.1038/nmeth.2772)



3. Phallen et al. Science Translational Medicine 9(403): eaan2415 (2017)  
doi:[10.1126/scitranslmed.aan2415](https://doi.org/10.1126/scitranslmed.aan2415)
4. Ofengeim et al. Trends in Molecular Medicine, 23(6), 563–576, (2017)  
doi:[10.1016/j.molmed.2017.04.006](https://doi.org/10.1016/j.molmed.2017.04.006)
5. ASHG 2017  
[<http://www.ashg.org/2017meeting/> ]
6. Strand Life Sciences Announces the Release of Strand NGS v3.1 at ASHG 2017  
[<https://www.prnewswire.com/news-releases/strand-life-sciences-announces-the-release-of-strand-ngs-v31-at-ashg-2017-300538443.html>]

## Contact author:

Dr. Pandurang Kolekar

Bioinformatics Engineer

Strand Life Sciences Pvt. Ltd., Bengaluru, INDIA

Send Mail to [pandurang \[at\] strandls.com](mailto:pandurang@strandls.com)



TAGGED ASHG, TRANSCRIPT QUANTIFICATION,  
UMI, UNIQUE MOLECULAR IDENTIFIER, VARIANT  
CALLING

**PREVIOUS POST** [Webinar on Unravelling complex mutational events in clinical cases using the power of NGS data analysis on 31 Jan](#)

**NEXT POST** [Webinar on Diagnosis of Rare Diseases using NGS Based Multi-gene Testing- Case studies 18 Apr](#)

## Pages

[Review of selected publications citing Strand NGS](#)

---

[Strand NGS v2.7](#)

---

[Calling narrow and broad peaks from ChIP-Seq data in Strand NGS](#)

[Subscribe](#)

2017 © Strand Genomics, Inc. All rights reserved