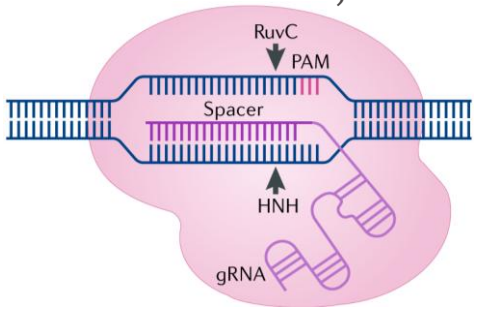
Bristol Myers Squibb®

Development of an NGS-Based Method for Characterization of sgRNA Sequence Impurities

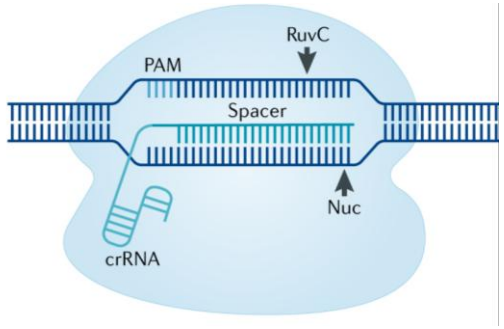
Athea Vichas, PhD
Senior Principal Scientist
June 10th, 2026

Impurities in the Manufacturing of sgRNA by Chemical Synthesis: Product and Process-Related

Cas9 sgRNA
100 nucleotides, ssRNA



Cas12a sgRNA
~66 nucleotides, ssDNA:ssRNA



- sgRNAs are generated by chemical synthesis
- Final sgRNA purity is determined by sequence, length, and coupling efficiency.
  - Purity and yield decrease with increased sequence length
  - Universal purity requirement may not be suitable
- FDA recommends each sgRNA be assessed for identity, functionality, purity, and safety.
- Process impurities: residual solvents and elements.
- Product impurities: sequence variants can result from incorrect base addition during chemical synthesis.

FDA CBER Webinar Q/A (2/29/2024) recommendations:

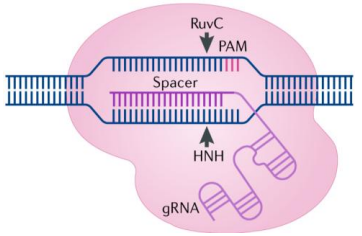
- Full length product ≥ 80%, and that sequence-based impurities of ≥ 1% be identified.
- Sequence variants be identified and quantified using an orthogonal NGS-based sequencing method.
- Risk Assessment on how sequence impurities may impact the safety and efficacy of drug product (e.g. off-target editing)

Considerations in Development of NGS-based sgRNA Sequence Impurity Assay

Library Preparation

Unbiased and representative of starting material

- Cas9 and Cas12a sgRNAs have different structures and compositions, leading to variable reaction efficiencies
- Avoid enrichment or depletion of any subset of sgRNA sequences
- Account for amplification bias



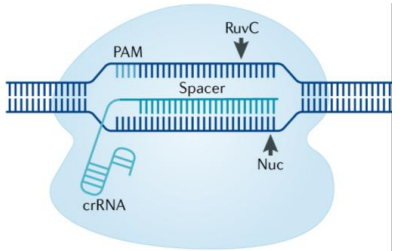
Cas9 sgRNA Structure

- 100-mer, all ssRNA
- High 2°structure at 3’ end
- Spacer region on 5’ end

Bioinformatic analysis

An analytical pipeline with the ability to process raw sequencing reads and report the frequency of all sgRNA sequence variants as a fraction of the total population.

- Account for false-positive PCR- and sequencing–induced errors
- Adequate sampling of sgRNA population
- Bulk-metric v. Single-molecule quantification

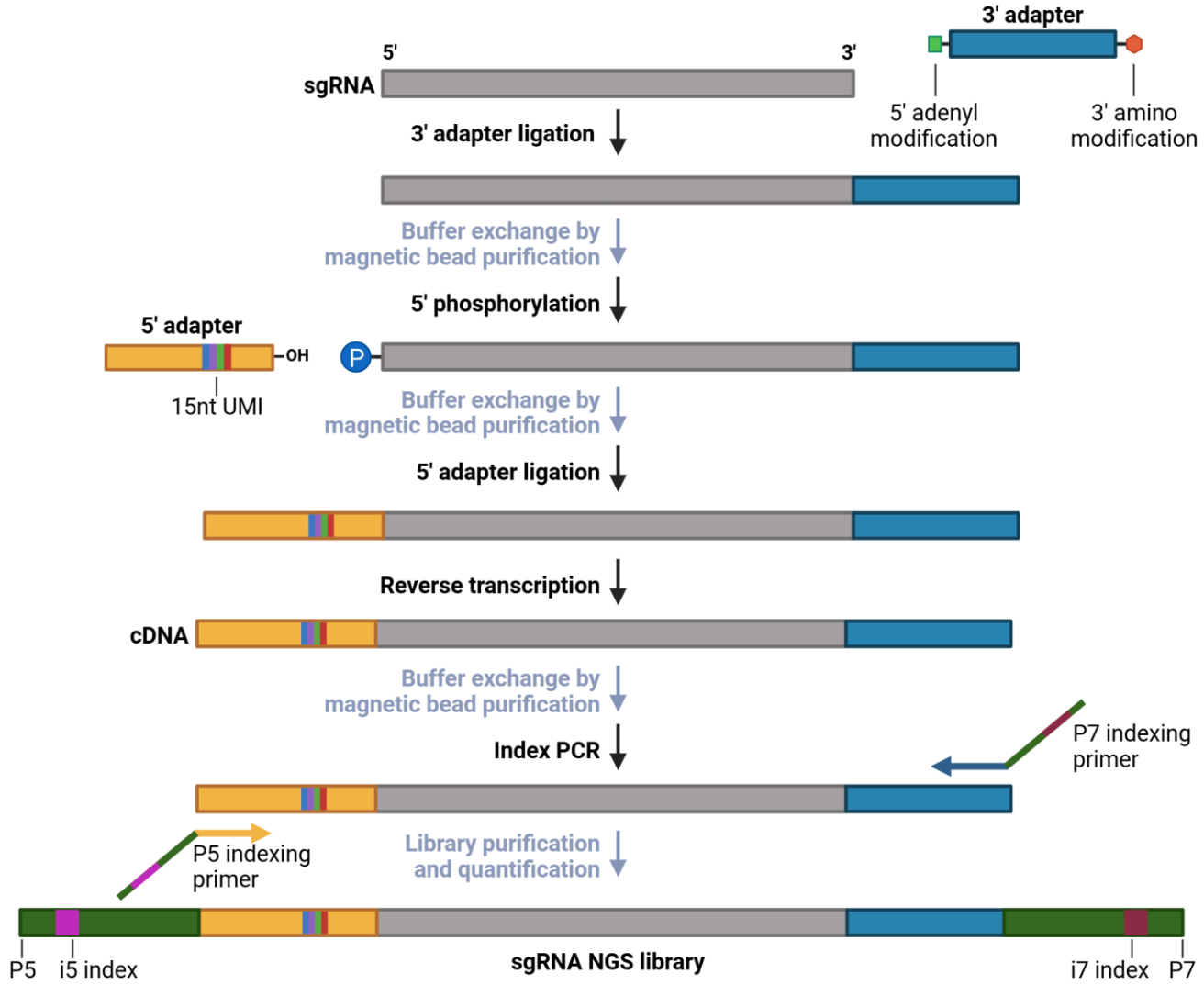


Cas12a sgRNA Structure

- ~66-mer, ssDNA:ssRNA
- High 2°structure at 5’ end
- Spacer region on 3’ end

UMI-Integrated Library Prep for Signal-Molecule Resolution

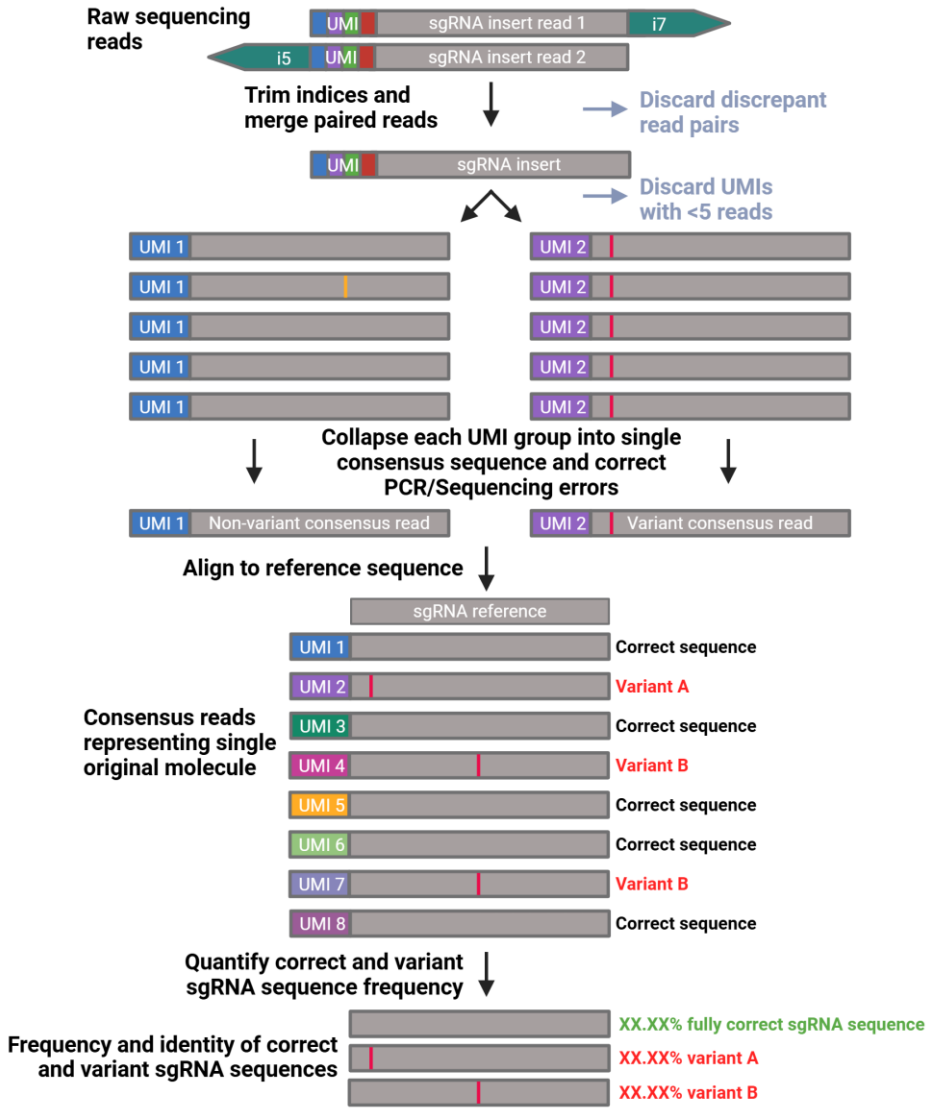
- Developed a next-generation sequencing method to assay millions of individual sgRNA molecules at single-nucleotide resolution.
- The workflow converts sgRNAs into Illumina-compatible libraries using universal adapter ligation, reverse transcription, and PCR for adapter and barcode addition.
- The process ensures unbiased representation by preventing enrichment or depletion of any sgRNA subset.
- Unique molecular identifiers (UMIs) individually label each sgRNA, improving quantification and minimizing false-positive sgRNA variant detection.



Bristol Myers Squibb®

Bioinformatic Workflow for sgRNA Sequence Impurity Profiling

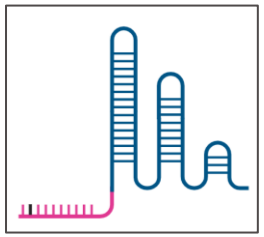
- Developed and optimized a bioinformatic pipeline to quantify sgRNA variant frequencies from raw sequencing reads.
- Extracts UMIs to group and collapse reads into consensus sequences, correcting PCR and sequencing errors.
- Enables accurate single-molecule quantification by representing each original sgRNA molecule with one consensus sequence.
- Aligns consensus sequences to the sgRNA reference, assigning descriptive ID tags to each correct or variant sequence.
- Reports the frequency and type (mismatch, insertion, deletion) of sgRNA variants by position, as a fraction of the total population.



Bristol Myers Squibb®

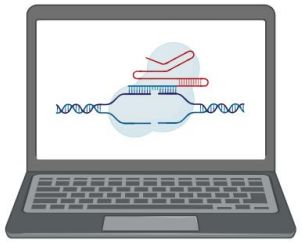
Assessing the functional impact of sgRNA sequence variants

Variant ID (type, position, length, nt change)	Sequence	UMI Count	Length (nt)	Frequency (%)	Category
Intended gRNA sequence	GAGAATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT	45882	100	78.69	WT
MM,93,G>A	GAGAATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCAGTGCTTTT	605	100	1.04	Cas-binding region
DEL,100,1	GAGAATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT	480	99	0.82	Cas-binding region
DEL,65,36	GAGAATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT	452	64	0.78	Cas-binding region
DEL,74,1	GAGAATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAGTGGCACCGAGTCGGTGCTTTT	397	99	0.68	Cas-binding region
MM,98,T>C	GAGAATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTT	362	100	0.62	Cas-binding region
DEL,1,4	ATCAAAATCGGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT	199	96	0.34	DNA targeting region
DEL,14,1	GAGAATCAAAATCGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT	152	99	0.26	DNA targeting region
DEL,1,13	GGTGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT	126	87	0.22	DNA targeting region



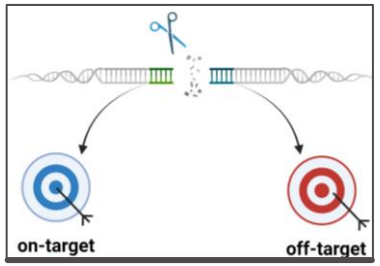
Location & Type of variant

- DNA Targeting region: safety impacting
- Cas-binding region: efficacy impacting
- Large truncation, indel, mismatch



Impact prediction tools

- In silico analysis of potential off-target sites
- Literature, structure/function studies



Safety Risk Assessment

- Is site already part of analytical control strategy?
- What is known about novel off-target?
- What is the frequency of sgRNA variant?

Thank you

athea.vichas@bms.com
www.linkedin.com/in/athea-vichas/

Bristol Myers Squibb®