Ecology & Evolution | Nov 10, 2026 | 12 min read | By Bryan White

CRISPR-Cas9 Depletion of Ubiquitous Ribosomal and Mitochondrial cDNA Halves Sequencing Depth Requirements in Human Intestinal Single-Cell Transcriptomics

Genomics * Biotech & Cellular Systems * CRISPR

Single-cell RNA sequencing (scRNA-seq) has transformed mucosal immunology and gastroenterology by resolving the transcriptional states of absorptive enterocytes, goblet cells, Paneth cells, enteroendocrine lineages, tuft cells, and lamina propria leukocytes across health and inflammatory bowel disease (IBD) [1, 2, 5] . Yet translating clinical single-cell atlases into large-scale longitudinal patient cohorts remains constrained by a formidable biochemical and economic bottleneck: in enzymatically dissociated human intestinal biopsies, between 50% and 85% of all sequenced reads map to ubiquitous, non-variable housekeeping species -- principally nuclear-encoded cytoplasmic ribosomal RNAs (18S, 5.8S, and 28S rRNA), mitochondrial ribosomal RNAs (12S MT-RNR1 and 16S MT-RNR2), and high-copy mitochondrial messenger RNAs released by dissociative cell lysis [1, 2, 3] .

In a methodological study conducted at Seoul National University Hospital and Seoul National University College of Medicine (published in Genomics & Informatics), Dong Jun Kim, Se Jong Koh, Hyung-Joo Kim, and colleagues demonstrate that integrating programmable CRISPR-Cas9 ribonucleoprotein (RNP) cleavage directly into post-amplification single-cell cDNA library construction selectively ablates these non-variable amplicons prior to Illumina bridge amplification [1, 2] . By physically severing adapter-flanked housekeeping cDNA molecules, the CRISPR-Cas9 depletion workflow recovers full transcriptional complexity, rare mucosal cell-type clusters, and differential gene-expression sensitivity at 50% of the conventional sequencing depth -- cutting flow-cell consumption and per-patient sequencing expenditure in half without distorting biological stoichiometry [1, 2] .

1. The Biophysical & Enzymatic Pathology of Dissociated Intestinal Mucosa

Why does human gastrointestinal tissue suffer from far more severe uninformative read saturation than peripheral blood mononuclear cells (PBMCs) or lymphoid organs? The answer lies in the physiological architecture and bioenergetic demands of the intestinal epithelium [1, 5] :

* Anchorage-Dependent Anoikis and Mechanical Shear: Intestinal epithelial cells (IECs) are polarized columnar cells anchored to a basal lamina via integrin-laminin focal adhesions and linked laterally by tight junctions (zonula occludens) and desmosomes . Stripping IECs from endoscopic pinch biopsies requires prolonged incubation with chelating agents (EDTA) and proteolytic enzymes (collagenase, dispase, or liberase) at 37 deg C, followed by mechanical trituration [1, 5] . Detachment from the extracellular matrix immediately triggers anoikis -- caspase-mediated programmed cell death -- and plasma-membrane permeabilization.

* Extreme Mitochondrial Bioenergetic Density: Absorptive enterocytes and crypt-base Paneth cells maintain some of the highest mitochondrial volume fractions in the human body to power ATP-dependent sodium - glucose cotransporters (SGLT1), basolateral Na + /K + -ATPase pumps, and constitutive antimicrobial peptide secretion. When stressed or partially lysed during dissociation, each compromised enterocyte releases thousands of copies of circular 16.5-kilobase mitochondrial genomes and polyadenylated mitochondrial transcripts (MT-CO1 , MT-CO 2 , MT-CO3 , MT-ND4 , MT-CYB , and MT-ATP6), alongside 12S and 16S mitochondrial rRNAs [1, 3] .

* Endogenous Pancreatic and Mucosal RNases: The intestinal lumen and brush-border glycocalyx harbor high concentrations of endogenous ribonucleases, bile salts, and mucin glycoproteins. Partial endonucleolytic nicking of cytoplasmic 28S and 18S rRNA generates internal poly(A)-like tracts or primeable 3' hydroxyl termini that undergo internal priming by oligo(dT) primers during droplet reverse transcription [1, 3, 4] .

* Ambient RNA Droplet Encapsulation: In microfluidic water-in-oil droplet platforms (such as 10x Genomics Chromium), cell-free cytoplasmic and mitochondrial transcripts floating in the suspension buffer are co-encapsulated with intact viable cells and barcoded gel beads [1, 6] . Every droplet consequently receives a heavy background "soup" of ubiquitous transcripts that compete directly with low-copy transcription factors, cytokines, and G-protein-coupled receptors for reverse transcriptase enzyme kinetics and PCR amplification cycles [1, 6] .

2. Thermodynamic & Kinetic Limitations of Standard Poly(A) Selection and Computational Filtering

In bulk eukaryotic RNA-seq, cytoplasmic rRNA is routinely removed prior to reverse transcription using either oligo(dT) magnetic bead capture or RNase H-mediated enzymatic digestion (where single-stranded DNA probes hybridize to rRNA, trigger RNase H cleavage of the RNA strand, and are subsequently degraded by DNase I) [3, 4] . However, droplet-based single-cell workflows cannot easily execute multi-step magnetic bead washes or sequential RNase H / DNase I buffer exchanges inside picoliter-volume emulsion droplets without disrupting reverse transcription fidelity or causing catastrophic loss of picogram-scale mRNA inputs [3, 4] .

Conversely, downstream bioinformatic tools -- such as SoupX , CellBender , and standard mitochondrial read-percentage thresholds in Seurat or Scanpy -- operate strictly post hoc [1, 6] . While an algorithm can subtract estimated ambient counts or discard droplets exceeding 20%-50% mitochondrial reads, those discarded reads have already consumed physical clusters on the sequencing flow cell [1, 2] . When 70% of an Illumina NovaSeq flow cell is occupied by MT-RNR2 , MT-CO1 , and 18S rRNA amplicons, rare regulatory transcripts in crypt stem cells (LGR5 , ASCL2) or lamina propria innate lymphoid cells drop below the Poisson sampling threshold (producing technical "dropouts") unless the investigator doubles or triples total sequencing depth [1, 2, 4] .

3. Molecular Mechanism of Post-Library CRISPR-Cas9 Depletion (DASH / CRISPRclean Architecture)

To eliminate uninformative sequences before flow-cell loading without altering droplet microfluidics, the workflow evaluated by Kim et al. builds upon the Depletion of Abundant Sequences by Hybridization (DASH) principle originally formulated by Gu, DeRisi, and colleagues [1, 3] and refined for commercial single-cell library pipelines (such as Jumpcode Genomics' CRISPRclean single-cell boost protocol) [1, 2, 4] .

Rather than targeting fragile native RNA inside intact cells, the enzymatic intervention takes place at the amplified double-stranded cDNA (dscDNA) or final adapter-ligated sequencing library stage [1, 3, 4] :

* Guide RNA (sgRNA) Tiling Array Design: A high-density pool of hundreds of synthetic single-guide RNAs (sgRNAs) is computationally tiled across the non-variable target sequences -- specifically nuclear 18S, 5.8S, and 28S rDNA, mitochondrial 12S and 16S rDNA, polyubiquitin housekeeping loci, and hyper-abundant mitochondrial protein-coding genes -- ensuring dense coverage adjacent to canonical Streptococcus pyogenes Cas9 (SpCas9) 5'-NGG-3' protospacer adjacent motifs (PAMs) while bioinformatically excluding off-target homology to variable human protein-coding exons [1, 3, 4] .

* Ribonucleoprotein (RNP) Complex Assembly & R-Loop Formation: Recombinant SpCas9 endonuclease is pre-complexed in vitro with the multiplexed sgRNA pool in a stoichiometric excess relative to target cDNA molecules. When incubated with the barcoded single-cell cDNA library at 37 deg C, SpCas9 scans double-stranded cDNA for PAM motifs, unwinds the target duplex to form a directional RNA-DNA heteroduplex (R-loop), and positions the target and non-target DNA strands into its HNH and RuvC catalytic domains, respectively [3, 4] .

* Site-Specific Double-Strand Break (DSB) Excision: The HNH and RuvC nuclease domains catalyze concerted hydrolysis of the phosphodiester backbone 3 base pairs upstream of the PAM, cleaving the full-length barcoded cDNA molecule into fragmented sub-segments [3] . Because multiple sgRNAs tile each target transcript at 50-150 bp intervals, ubiquitous ribosomal and mitochondrial cDNAs are shredded into short internal fragments [3, 4] .

* Adapter Uncoupling & Exponential PCR Exclusion: On an Illumina sequencing instrument, only intact library molecules carrying both the P5 (cell barcode + unique molecular identifier [UMI] + Read 1 primer site) and P7 (sample index + Read 2 primer site) terminal adapters can undergo solid-phase bridge amplification on the flow-cell lawn [1, 3] . Once SpCas9 severs the internal cDNA insert between P5 and P7, the cleaved halves cannot serve as templates during subsequent library indexing PCR or flow-cell cluster generation, and short magnetic-bead size selection (e.g., AMPure XP / SPRIselect beads) physically removes the digested fragments [1, 3, 4] .

4. Comparative Parameter Budget: Standard vs. CRISPR-Depleted Intestinal scRNA-seq

By physically purging high-copy non-variable cDNAs prior to sequencing, the CRISPR-Cas9 depletion protocol shifts the entire sequencing read budget toward

informative, variable mRNA transcripts [1, 2, 4] . The table below synthesizes the quantitative performance metrics across standard versus CRISPR-Cas9-depleted intestinal single-cell workflows across full (100%) and halved (50%) sequencing depths [1, 2, 3, 4] .

| Analytical / Economic Parameter | | Standard Uncleaned scRNA-seq (100% Depth) | | CRISPR-Cas9 Depleted scRNA-seq (100% Depth) | | CRISPR-Cas9 Depleted scRNA-seq (50% Half-Depth) | | Mechanistic & Clinical Interpretation |

| Raw Sequencing Depth per Cell | | ~50,000-60,000 read pairs / cell | | ~50,000-60,000 read pairs / cell | | ~25,000-30,000 read pairs / cell | | Halves NovaSeq flow-cell cluster allocation per patient biopsy [1, 2] . |

| Ribosomal & Unwanted Ubiquitous Read Fraction | | 50%-78% of total aligned reads | | <5%-12% of total aligned reads | | <5%-12% of total aligned reads | | Multi-tiled SpCas9 RNP cleavage eliminates rRNA and mtRNA dominance [1, 3, 4] . |

| Informative Exonic Protein-Coding Reads | | 22%-45% of total reads | | 75%-88% of total reads | | 75%-88% of total reads | | Redirects sequencing bandwidth to variable nuclear mRNAs [1, 2, 4] . |

| Median Unique Genes Detected per Cell | | Baseline reference (1.0x) | | 1.4x-1.8x increase vs. baseline | | 1.05x-1.25x (Exceeds 100%-depth control) | | Half-depth CRISPR libraries recover more informative genes than full-depth untreated libraries [1, 2] . |

| Cell-Type Proportion Concordance (Pearson r) | | Reference standard (r = 1.00) | | r > 0.98 across epithelial & immune clusters | | r > 0.97 across all lineages | | Zero distortion of enterocyte, goblet, Paneth, T-cell, or macrophage ratios [1, 2] . |

| Net Sequencing Cost per Sample | | 100% baseline NGS cost | | ~110% (100% NGS + ~10% RNP reagent overhead) | | ~55%-60% (~40%-50% net savings) | | Modest CRISPR reagent cost is outweighed by 50% reduction in sequencing run cost [1, 2] . |

5. Empirical Performance at Half Sequencing Depth: Saturation Kinetics and Rare Lineage Resolution

To rigorously test whether CRISPR-Cas9 depletion allows investigators to halve sequencing depth without losing biological signal, the Seoul National University Hospital investigators performed paired library preparation on human intestinal mucosal samples -- splitting cDNA pools into untreated control libraries and CRISPR-Cas9-depleted libraries, and systematically downsampling sequencing reads to 50% depth [1, 2] .

In single-cell transcriptomics, library complexity follows a Michaelis-Menten-like saturation curve (often modeled via the Lander-Waterman equation or non-parametric Good-Turing frequency estimators), where the number of newly discovered unique molecular identifiers (UMIs) per additional million reads asymptotically approaches the total number of captured reverse-transcribed molecules in the cDNA pool [3, 4] . In untreated intestinal libraries, because ubiquitous ribosomal and mitochondrial amplicons outnumber rare mRNAs by two to three orders of magnitude, early sequencing reads are overwhelmingly consumed by re-sequencing duplicate housekeeping UMIs.

Consequently, rare mRNAs only begin to register on the flow cell at high read depths (50,000-100,000 reads/cell) [1, 4] .

When CRISPR-Cas9 cleaves the top 100-200 hyper-abundant non-variable species prior to indexing PCR, the effective molar fraction of informative nuclear mRNAs increases by two- to threefold [1, 3, 4] . As a result, the saturation curve rises much more steeply at low read depths. At just 50% sequencing depth , the CRISPR-depleted intestinal libraries yielded higher valid exonic mapping rates, superior UMI complexity per cell, and clearer separation of principal component (PCA) and Uniform Manifold Approximation and Projection (UMAP) clusters than the untreated control libraries sequenced at 100% depth [1, 2] .

Crucially, the authors verified that enzymatic depletion introduced zero off-target cleavage bias across variable protein-coding genes [1, 2] :

* Lineage Stoichiometry Preservation: Quantitative deconvolution confirmed that the relative proportions of absorptive enterocytes, MUC2 + goblet cells, LGR5 + / OLFM4 + crypt stem and transit-amplifying cells, CD3D + intraepithelial and lamina propria T lymphocytes, CD79A + plasma/B cells, and myeloid macrophages remained invariant between untreated full-depth and CRISPR-treated half-depth datasets [1, 2] .

* Low-Abundance Transcription Factor Recovery: By stripping away the masking veil of rRNA and ambient mtRNA, half-depth CRISPR libraries enhanced the detection sensitivity of low-copy mucosal cytokines, chemokines, and lineage-defining transcription factors essential for dissecting Crohn's disease and ulcerative colitis pathogenesis [1, 5] .

6. Preserving QC Metrics: Navigating the Mitochondrial Apoptosis Paradox

One classical objection to enzymatically depleting mitochondrial transcripts in scRNA-seq is that bioinformaticians routinely use the percentage of mitochondrial reads (percent.mt) as a quality-control proxy to filter out broken, apoptotic, or empty droplets [6] . When a cell's plasma membrane ruptures during dissociation, small cytoplasmic mRNAs leak out into the suspension buffer faster than large, membrane-enclosed mitochondria; thus, a high ratio of mitochondrial-to-nuclear RNA typically flags a dying cell [1, 6] .

How does targeted CRISPR-Cas9 depletion avoid blinding downstream QC algorithms to genuinely damaged droplets? First, the sgRNA panel specifically targets the non-variable, hyper-abundant 12S and 16S mitochondrial ribosomal RNAs (MT-RNR1 and MT-RNR2) and uninformative nuclear rRNAs/pseudogenes that dominate ambient noise, while either sparing a calibrated subset of canonical mitochondrial protein-coding mRNAs or preserving nuclear-to-cytoplasmic splice-junction metrics (such as intronic vs. exonic read ratios and total nuclear UMI complexity) that robustly distinguish intact viable cells from lysed cellular debris [1, 4, 6] . Second, removing ambient MT-RNR1/2 prevents viable, metabolically active intestinal epithelial cells -- which naturally possess high baseline mitochondrial content -- from being falsely discarded by arbitrary percent.mt cutoffs [1, 5] .

7. Implications for Clinical Multi-Omics, Spatial Transcriptomics, and Atlas-Scale Gastroenterology

The demonstration that CRISPR-Cas9 non-variable RNA deletion halves required sequencing depth in human intestinal biopsies carries immediate operational implications across translational genomics [1, 2, 5] :

* Doubling Clinical Cohort Power Under Fixed Grant Budgets: In translational IBD trials -- such as profiling mucosal biopsies from patients before and after anti-TNF (infliximab/adalimumab), anti-integrin (vedolizumab), or IL-23p19 (mirikizumab/risankizumab) biologics -- statistical power is dictated by the number of biological replicates (patients) rather than over-sequencing a handful of biopsies [5] . Cutting per-sample NovaSeq sequencing costs by 50% allows clinical consortia to double patient sample sizes within the same sequencing budget [1, 2] .

* Compatibility with Spatial Transcriptomics and FFPE Archival Tissue: Beyond dissociated droplet scRNA-seq, high-resolution spatial transcriptomics platforms (such as 10x Visium HD, Slide-seq, and Stereo-seq) and formalin-fixed paraffin-embedded (FFPE) clinical pathology blocks suffer from even more severe rRNA fragmentation and background capture. Post-library CRISPR-Cas9 RNP depletion plugs seamlessly into spatial cDNA workflows without requiring any modifications to slide capture chemistry [3, 4] .

* Generalizability Across High-Metabolism and Nuclease-Rich Organs: The biophysical challenges conquered in the intestine -- high mitochondrial density, enzymatic dissociation stress, and endogenous RNase activity -- are shared by pancreatic acinar tissue, renal proximal tubules, hepatic parenchymal biopsies, and necrotic solid tumor microenvironments [1, 3, 4] . Deploying multiplexed Cas9 RNP depletion as a standard pre-sequencing conditioning step establishes a scalable, cost-effective blueprint for the next generation of clinical single-cell medicine [1, 2] .

References & Primary Sources

* [1] Kim, D. J., Joh, C. S. Y., Jeong, S. Y., Kim, Y. J., Koh, S. J., & Kim, H. J. (2025). Non-variable RNA deletion using the CRISPR-Cas9 technique demonstrated improved outcomes in human intestine single-cell RNA sequencing data, even at half sequencing depths. Genomics & Informatics, 23, 43. -- https://doi.org/10.1186/s44342-025-00043-6

* [2] Bioengineer.org (2026). CRISPR Cleaves Ubiquitous RNA to Halve Intestinal Single-Cell Sequencing Costs. -- https://bioengineer.org/crispr-cleaves-ubiquitous-rna-to-halve-intestinal-single-cell-sequencing-costs/

* [3] Gu, W., Crawford, E. D., O'Donovan, B. D., Wilson, M. R., Chow, E. D., Retallack, H., & DeRisi, J. L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biology, 17, 41. -- https://doi.org/10.1186/s13059-016-0904-5

* [4] Prezza, G., Heckel, T., Dietrich, S., Homberger, C., Westermann, A. J., & Vogel, J. (2020). Improved bacterial RNA-seq by Cas9-based depletion of ribosomal

RNA reads. RNA, 26(8), 1069-1078. -- https://doi.org/10.1261/rna.075945.120

* [5] Elmentaite, R., Kumasaka, N., Roberts, K., Fleming, A., Dann, E., King, H. W., Kleshchevnikov, V., Dabrowska, M., Pritchard, S., Bolt, L., Wandrer, F., Humphreys, F., Mamanova, L., Huang, N.,ried, F., et al. (2021). Cells of the human intestinal tract mapped across space and time. Nature, 597(7875), 250-255. -- https://doi.org/10.1038/s41586-021-03852-1

* [6] Young, M. D., & Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience, 9(12), giaa151. -- https://doi.org/10.1093/gigascience/giaa151