Auperon RNA-seq
sample prep protocol.

The bench-ready guide to preparing samples for Auperon engagements. Read prep, cell handling, vendor coordination, shipping, and quality standards — everything your team needs, in one document.

50–150M
READ DEPTH
paired-end reads / sample

Ribo-zero
LIBRARY PREP
preferred over poly-A

≥5 × 10^6
CELLS / SAMPLE
frozen pellet minimum

2–8 wks
ANALYSIS TURNAROUND
after we receive your data

Auperon by Van Heron Labs
Version 1.0 · July 2026
hello@auperon.io · auperon.io

§ 1

Overview & scope

This protocol covers sample preparation for Auperon engagements. Auperon is Van Heron Labs' cellular intelligence platform, productized — every engagement starts with RNA-seq data and produces multi-omic characterization, an interactive biological digital twin, and mechanism-backed recommendations.

Auperon does not run sequencing itself, coordinate vendors, or handle sample logistics. You send us data in one of three ways: (1) existing RNA-seq data you already have, (2) new data you generate with the vendor of your choice following the sample prep guidance in this document, or (3) public data (GEO, SRA, ENA, ArrayExpress). This document covers the sample prep guidance for option 2 — cell-type-specific handling, sample labeling, shipping best practices, and backup protocols for edge cases (low cell counts, protein-rich biofluids, BSL2+ material).

If your engagement needs vendor management, sequencing coordination, sample prep guidance beyond this document, or wet-lab work, that's **Van Heron Labs full-service** — the enterprise pathway that handles the pipeline end-to-end.

This document reflects Van Heron Labs' **preferred** RNA-seq setup — the configuration that produces the strongest Auperon output. That said, we can work with virtually any RNA-seq data. If your existing setup, vendor, or read structure deviates from what's below, send it anyway — most of the time we can accommodate. Flag deviations at scoping so we can advise.

Hand this document to your bench team. If any protocol here doesn't fit your setup, email **hello@auperon.io** before capture — we adapt.

> **Special case flags**
>
> **HPL, serum, lysates, extracts and other protein-rich biofluids:** perform RNA extraction on your side. See §6.
>
> **Low cell count (<1 × 10^6):** perform on-site RNA extraction and ship purified RNA. See §7.
>
> **BSL2+ material:** vendor written approval required in advance. See §10.

§ 2

Sequencing input requirements

Recommended defaults

Auperon's analysis maps the full transcriptional landscape at pathway, biomarker candidate, and existing-molecule modulator level. Depth and prep choice matter.

- **Read depth:** 50–150M paired-end reads per sample.
- **Library prep:** ribo-zero preferred over poly-A selection.
- **Format:** paired-end (over single-end) — better alignment accuracy and splice junction capture.

Depth by use case

- **Baseline characterization / discovery:** 50–75M paired-end reads sufficient.
- **Deep comparative analysis / differential expression:** 100–150M recommended.
- **Long-tail transcript detection / novel biomarker discovery:** 150M+ recommended.

Why ribo-zero over poly-A

Ribo-zero captures the full transcriptome — including mitochondrial reads. Poly-A selection captures only long-tailed mature mRNAs and drops the rest. Three specific reasons this matters for Auperon:

1. Ribo-zero preserves mitochondrial signal. Mitochondria are the metabolic hub of the cell — central to almost every question Auperon is asked to answer (productivity, viability, differentiation, potency, stress response). Mitochondrial mRNAs are polyadenylated, but their poly-A tails are much shorter than nuclear mRNAs (roughly 25-55 nt vs 150-250 nt). Poly-A selection systematically depletes short-tailed mRNAs — you lose the metabolic signal exactly where you most want it. Ribo-Zero Gold-family kits deplete both cytoplasmic AND mitochondrial rRNA, so mitochondrial mRNAs are preserved in the library.

2. Ribo-zero catches non-coding and regulatory signal. lncRNAs, snoRNAs, snRNAs, and partially degraded or bimorphic transcripts are recovered by rRNA depletion but lost with poly-A selection. Many carry regulatory information you may not want to throw away.

3. Poly-A yields shift with cell state. Poly-A tail length varies with cell stress, translational state, and growth phase — introducing a hidden covariate between samples. Ribo-zero doesn't have this problem.

Poly-A prep still works — the analysis just weights and interprets it differently. **Flag the format at kickoff so we adjust parameters.**

Why not 3'-biased approaches

3'-biased RNA-seq methods sequence only ~90-400 nt near the 3' end of each transcript, using poly(dT) priming to anchor at the poly-A tail. They are dramatically cheaper (roughly 10M reads per sample vs 30-50M+ for full-length) and dramatically faster (turnaround measured in days rather than weeks). For simple differential expression on highly-expressed, well-annotated genes, they work.

They are **not suitable for Auperon** for five reasons:

- **No isoform or splicing information.** Alternative splicing, isoform switches, and 3' UTR variants (which regulate stability and translation) are invisible with 3'-only reads. Auperon's modulator identification and biomarker discovery can depend on this detail.
- **No 5' UTR or coding-sequence signal.** Translation regulation happens at the 5' end. 3'-biased methods don't see it.
- **Same mitochondrial problem as poly-A selection.** 3'-biased methods rely on poly(dT) priming to reach the poly-A tail. Mitochondrial short-tail mRNAs bind poorly and are systematically underrepresented — the same metabolic-signal loss.
- **Non-coding RNAs are mostly missed.** Most lncRNAs and snoRNAs lack a long poly-A tail; 3'-biased methods can't see them. Some folks want this information for future internal work.
- **Depends on 3' UTR annotation quality.** For proprietary strains, engineered lines, or less-characterized organisms, 3' UTR annotations are often incomplete. Reads either map ambiguously or get discarded.

When 3'-biased methods are the right choice: high-throughput screening across dozens or hundreds of samples where you already know which genes matter, on well-annotated protein-coding transcripts. Great for quick pilots and screens; not for mechanism-backed discovery.

If you already have 3'-biased data, send it anyway — we can often find useful signal. But for a new engagement, we strongly recommend full-length, ribo-zero, paired-end sequencing at the depths described above.

Other formats we handle

Auperon supports formats beyond the recommended bulk ribo-zero prep. Note the format in your project description at kickoff:

- **3′ poly-A RNA-seq** — analysis weights coding transcripts; loses some non-coding signal.
- **Oxford Nanopore / long-read RNA-seq** — excellent for isoform-level analysis, splice detection.
- **Single-cell RNA-seq (+ or - 3′ biased)** — supported; scoped differently vs bulk (see below)
- **Existing published datasets** — supported for discovery-mode and feasibility work.

A note on single-cell projects

Virtually all droplet-based single-cell RNA-seq (scRNA-seq) datasets — and single-nucleus RNA-seq (snRNA-seq) datasets even more so — contain some level of **ambient RNA contamination**. Ambient RNA is the pool of mRNA molecules released into the cell suspension (typically from stressed or apoptotic cells) that gets co-encapsulated with intact cells during droplet formation.

In practice, this shows up as **marker genes and receptors from non-target cell types appearing highly expressed in target-cell populations** — reads that must have originated from contaminating cells in the suspension, not from target-cell biology and thus confound results.

This is a ubiquitous, well-characterized phenomenon in the single-cell literature, and it's mitigable — through dissociation protocol optimization, dead-cell removal, and computational correction downstream. But it's frequently under-appreciated by teams designing single-cell experiments.

Flag single-cell or single-nucleus workflows at scoping so we can advise on both experimental design and analysis approach.

§ 3

Sample capture — timing by study goal

The right time to capture your cells depends on what you're trying to optimize. Growth phase and physiological state at capture determine what biology the analysis will see. Below are the capture windows we recommend by study goal, with the reasoning behind each — and, equally important, what your cells should **not** be doing at the moment of capture.

For proliferation studies / general culture optimization

Capture cells that are **actively proliferating, healthy, and unstressed**. In practice:

- **Post-passage recovery complete** — typically 48-72 hours after passage or thaw. Immediately post-passage cells are transcriptionally noisy from mechanical/enzymatic stress; you're capturing recovery signal, not steady-state biology.
- **Pre-confluence** — typically 60-80% confluence for adherent cells; below stationary phase for suspension. Confluence triggers contact inhibition, growth arrest, and dramatic transcriptional shifts.

- **Actively dividing** — verify log-phase growth, not lag or plateau.
- **Alive** — dead or dying cells contribute noise from stress-response gene signatures and ambient RNA that mask baseline biology.

Why: proliferation biomarkers and modulator recommendations are only meaningful when cells are in a proliferative state. Capturing at the wrong phase gives you information about the wrong biology.

For titer / production optimization

Capture at **two timepoints — early AND late in the production window**:

- **Early production** — as production is ramping up. Reveals what's driving production onset (metabolic switching, secretory machinery activation, unfolded-protein-response engagement, etc.).
- **Late production** — at or near peak production, before senescence or death signals dominate. Reveals what sustains production and what's beginning to shut it down.

Why: production physiology changes dramatically across a production run. Single-timepoint sampling gives you only part of the story; two-timepoint sampling captures the full arc. Media and process optimization built on both timepoints supports both the ramp AND the sustain — where single-timepoint work often optimizes one at the expense of the other.

For viability / stress response studies

- **Match sampling exactly across conditions** — same nominal timepoint, same growth phase, same media age.
- If comparing stressed vs unstressed cells, capture at the same phase in each condition — do not compare early-log stressed cells to mid-log unstressed cells.
- For acute stress studies, define stress duration and capture at the same duration in every replicate.

Why: mismatched growth phase across conditions conflates stress signal with growth-phase signal. The analysis can't tell them apart, and neither can you.

For anti-differentiation / maintenance of stemness

- Capture cells **while they are still in the target state** (pre-differentiation).
- Verify target-state markers at capture — e.g., stemness markers for iPSCs / MSCs, naïveté markers for T cells.
- Multiple timepoints across the maintenance window if you're studying maintenance duration.

For serum elimination / chemically-defined transition

- Capture cells **adapted to the target media condition** — not during acute transition, unless the transition itself is the study subject.
- Extended culture (multiple passages) in target media before capture is strongly recommended.
- If comparing serum-containing vs serum-free, ensure equivalent adaptation time in both conditions.

For potency studies

- Capture at the physiological state where potency is measured.
- For activation-dependent potency (T cells, CAR-T), capture at both pre-activation and post-activation states.
- Match activation conditions across replicates precisely.

Universal rules for comparative studies

Same conditions throughout:

Matched growth phase across all conditions being compared.

Same media age across all samples (media aged differently has meaningfully different composition).

Same handling protocol — temperature, timing, air exposure, cell density at capture.

If any condition can't hit these exactly, flag it at scoping — we can run covariate corrections in some cases, but not all.

§ 4

Sample preparation — frozen cell pellets

Auperon's standard input is **frozen cell pellets**. The vendor performs RNA extraction — no on-site lysis, no RNAlater, no purified RNA required (except for specific cases in §5 and §6).

Who handles RNA extraction?

We generally recommend letting your sequencing vendor perform the RNA extraction. Ship frozen cell pellets; let the vendor extract on their end. This is the default in most Auperon engagements.

Why:

Eliminates a variation source — one protocol, one lab, one operator across all your samples meaningfully reduces batch effects vs multiple in-house extractions.

Standardized with downstream library prep — vendor extraction is validated against their library prep workflow, so you get more consistent library yields.

Cost — RNA extraction kits (Trizol, RNeasy, MagMax, etc.) run $200-500+ per kit; bundling extraction into the sequencing quote is usually cheaper than doing it yourself.

Simpler for your team — one less bench step, and frozen cell pellets ship more stably than purified RNA in most cases.

Upstream QC — if extraction fails, the vendor knows first and can flag it before library prep.

When to extract in-house instead:

Serum, plasma, HPL, or other protein-rich biofluids — matrix effects require specialized protocols. See §6.

Low cell counts (<1 × 10^6) — vendor extraction has material loss; concentrate purified RNA on your side first. See §7.

BSL2+ material without vendor biosafety approval. See §10.

Regulatory / GMP / chain-of-custody constraints — extract in-house under your quality system if required.

If in doubt, email **hello@auperon.io** before capture — we advise based on your specific system.

Target input

Target: 5 × 10^6 cells per sample. Highly recommended: 5 × 10^7 cells where platform yield allows. Pooling is generally fine if needed.

General workflow

1. Collect cell samples into a labeled RNase-free centrifuge tube; remove culture medium.
2. Rinse cell pellet with **cold PBS** — **not RNAlater** (high density makes RNAlater difficult to remove and can interfere with extraction).
3. Remove supernatant by centrifuge (~200–300 ×g, ~5 min for standard mammalian cells; adjust for cell type).
4. Snap-freeze the dry pellet in liquid nitrogen — **use LN_2 vapor or a LN_2-chilled block**, not direct plunge (see §9 for tube safety).
5. Transfer to –80 °C freezer for storage until shipment.

No lysis buffer or RNAlater needed. The vendor performs RNA extraction from the frozen pellets on arrival.

> **Backup samples**
>
> Pull duplicate pellets where possible. Keep two vials per condition as backups in –80 °C storage — these can travel with the primary set as replicates or stay behind as insurance.
>
> Indicate backups clearly in your label scheme (e.g., **_BKP** suffix) and list them in the manifest.

§ 5

Cell-type-specific handling notes

Same general workflow, with cell-type nuances. If your system falls outside these, email us at scoping — we'll adjust.

CHO (mAb production, biologics)

Standard mammalian centrifugation (~200–300 ×g, 5 min). For high-density production runs (>30 × 10^6 cells/mL), well under 1 mL of culture is needed per sample. **Pull by biological state, not calendar day** (e.g., mid-exponential vs late-stationary) so capture is consistent across process variations.

MSCs and adherent stem cells

5 × 10^6 cell minimum; >5 × 10^7 strongly preferred for platform depth. If your yield is genuinely low (e.g., microcarrier culture, early timepoints), fall back to on-site RNA extraction — see §7.

iPSCs / ESCs

Passage timing matters. Harvest at consistent confluence (typically 70–85%) across replicates to avoid confounding your biological signal with growth-phase artifacts.

T cells (CD4/CD8, TIL, Treg)

Capture by activation state, not calendar day. Pre-activation, peak-activation, and exhaustion each have distinct transcriptional profiles — decide which contrast you want to compare at scoping and pull consistently across replicates.

Microbial (E. coli, yeast)

See accompanying protocol we've provided.

§ 6

Special case — human platelet lysate & biofluids

For HPL and other protein-rich biofluid samples, **perform RNA extraction on your side prior to shipment.** Frozen pellets don't work for these matrices — the extraction chemistry differs from standard cell culture.

Recommended kits

- **Qiagen miRNeasy Serum/Plasma Advanced Kit** — recommended. qiagen.com › miRNeasy Serum/Plasma Advanced Kit
- **Zymo Research Direct-zol RNA Microprep Kit** — TRIzol-based workflows. zymoresearch.com › direct-zol

Both approaches are commonly used for low-input or inhibitor-rich biological fluids and generally improve RNA recovery and purity in these matrices.

Ship purified RNA on dry ice. Include RNA concentration and RIN (if measured) in your manifest.

§ 7

Backup — low cell count fallback

If a sample falls below 1 × 10^6 cells (common in microcarrier culture, early timepoints, or genuinely rare cell populations), **perform RNA extraction on-site and send purified total RNA instead** of a frozen pellet.

Recommended kit

Qiagen RNeasy Plus Mini Kit — includes genomic DNA removal and generally provides high-quality RNA suitable for transcriptomic applications. qiagen.com › RNeasy Plus Mini Kit

Flag the format change to Van Heron Labs and to the vendor before shipping so the intake and library prep pipeline is adjusted accordingly.

§ 8

Sample labeling & naming

Most sequencing vendors have naming constraints, and the rules below are broadly compatible with common vendors including Novogene, Genewiz/Azenta, and most academic cores. Sample names on tubes **must match exactly** with your vendor's Sample Information Form (SIF) or equivalent intake document.

Naming best practices

- Begin with a letter of the alphabet.
- Contain a maximum of **12 alphanumeric characters or underscores** — this is the tightest common constraint; if your vendor allows more, you can use more.
- Contain no hyphens, spaces, or special characters.
- Match exactly between tube labels and the intake document you submit to your vendor.

Example scheme

Encode process, condition, and replicate. Example: `P1_A_C_E_1` = Process 1, Cell line A, Control, Early timepoint, Bio rep 1. Choose a scheme that captures your experimental design in **≤12 characters** — we can help design it at kickoff.

Physical labeling

Label tubes on the side, not cap-only. Write name with cryo-marker and cover with clear tape. Cap-top labels alone rub off in transit or LN_2.

> **Vendor liability on label mismatches**
>
> Most sequencing vendors are not liable for samples shipped with labels inconsistent with the intake form. If names don't match, the vendor may attempt to process based on best-guess matching — or reject the sample. Get the labels right the first time.

§ 9

Tubes, containers & LN_2 safety

Tube specifications

- DNase-/RNase-free 1.5 or 2.0 mL microcentrifuge tubes.
- Many vendors specify **unthreaded tubes** (snap-cap only) — confirm with your vendor before using screw-cap tubes.
- **Do not use 96-well plates** or non-standard tube formats — risk processing delays and vendor handling fees.
- Seal each tube with Parafilm before packaging.

LN_2 safety

Standard snap-cap microtubes can crack or pop when plunged into liquid nitrogen. This isn't a hypothetical — it happens routinely on high-volume capture days.

Two safe options:

- Use RNase-free **screw-cap tubes** (after confirming acceptable with your vendor).
- Flash-freeze in **LN_2 vapor** or on a **LN_2-chilled block**, then transfer to –80 °C.

Either way — keep caps fully secured.

§ 10

Shipping — packaging, courier, timing

General best practices for shipping frozen samples to any sequencing vendor. Always confirm your specific vendor's requirements, deadlines, and biosafety policies before shipment — the guidance below reflects what's common across major vendors.

Courier & timing

- **Ship frozen on dry ice, overnight.**
- FedEx overnight is a common choice. International express is also widely accepted, where applicable.
- Book shipment at least **48 hours in advance** via your vendor's shipping booking form or portal.
- **Ship Monday–Wednesday** to avoid weekend transit.
- Confirm holiday closures and current lead times with your vendor before booking.

Packaging

- **Dry ice is regulated** — outer box must carry the approved sticker (from your dry-ice supplier).
- Seal each tube with Parafilm.
- Place tubes **upright** in a container (50 mL tube or cryobox) with interior racks or supports — **never loose in bags**.
- Use cotton or absorbent paper to prevent tubes from being jostled by dry ice.
- Cover the polystyrene dry-ice box with an outer cardboard box.

Documents to include inside the box

- Copy of the vendor's Sample Information Form (SIF) or equivalent intake document.
- The Waybill.
- A printed sample manifest matching the tube labels.
- Also place a copy of the manifest inside the outer cardboard box (in case the inner is damaged).

Biosafety

> **BSL2+ material requires prior written approval**
>
> **Most sequencing vendors do not accept Biosafety Level 2+ material without prior written approval.** This includes blood-borne pathogens, purified viral genomes, and other BSL2+ agents.

> If your samples fall into this category, secure vendor approval before shipment. Any BSL2+ sample submitted without prior approval may be immediately destroyed or returned by the vendor.

§ 11

Quality standards

Auperon's output is only as good as the data it's built on. Every dataset goes through internal QC before we begin the full analysis pipeline.

Standard incoming-data checks

- Read quality distribution (base-level quality across read length).
- Alignment rate against the reference genome for your organism.
- Read count per sample vs the depth target for your engagement tier.
- RNA integrity indicators from the sequencing run (RIN where available).
- Library complexity and duplication rates.
- Comparability across samples in the experimental design.

If a specific sample fails, we flag it early. You may still get meaningful analysis on the remaining samples, or we may recommend re-sequencing that specific sample.

Refund policy for failed sequencing

> **If sequencing data doesn't meet our internal QC**
>
> If sequencing data you provide **fails our internal QC — meaning it's uninterpretable at the level required to produce mechanism-backed output — we refund the engagement in full** and provide our recommendations on experimental design for a future run.
>
> This protects both of us: you're not paying for output we can't confidently deliver, and we're not producing recommendations from data we don't trust.

§ 12

Contact & support

Questions about this protocol, edge cases in your setup, or anything else technical — we're a real team, and we reply within one business day.

Getting help

- **General questions:** hello@auperon.io
- **Technical scoping:** hello@auperon.io — subject line "Technical scoping"
- **Start an engagement:** auperon.io/start
- **Enterprise partnerships (VHL full-service):** vanheronlabs.com

Related documents

- **Auperon pricing & tiers:** auperon.io/pricing
- **FAQ:** auperon.io/faq
- **Case studies:** auperon.io/case-studies

Version note

This is **Auperon RNA-seq Sample Prep Protocol v1.0**, dated July 2026. Van Heron Labs updates this document as our platform and vendor relationships evolve. Always work from the version linked on auperon.io/science, and email us if anything in your printed copy contradicts current guidance online. Always default to Vendor and Kit instructions where applicable. This document is designed to serve only as a best practices guide and general reference.