Meeting Minutes - Cut Tag Project Report

- **Date**: 2025-08-30

1. Nucleosome Architecture and Histone Organization

- **Key Points**
  - The nucleosome core particle consists of ~147 bp of DNA wrapped ~1.7 turns around a histone octamer.
  - **Histone spatial arrangement**:
    - Two copies each of H2A, H2B, H3, and H4 form the octamer.
    - The central scaffold is the (H3–H4)(_2) tetramer; two H2A–H2B dimers dock on either side to complete the octamer.
    - DNA engages the histone octamer via the acidic patch (notably on H2A–H2B) and multiple basic histone tails.
  - **Histone H1 (linker histone)**:
    - Binds near the nucleosome dyad and the entering/exiting linker DNA.
    - **Stabilizing roles**:
      - Enhances DNA wrapping and reduces nucleosome breathing.

2. CUT&Tag vs ChIP-seq: Methods and Practical Considerations

- **Method Overviews**
  - **ChIP-seq** (Chromatin Immunoprecipitation followed by sequencing):
    - Workflow: Crosslink (often formaldehyde) → chromatin fragmentation (sonication or MNase) → immunoprecipitation with antibody → reverse crosslinks → library prep → sequencing.
  - **CUT&Tag** (Cleavage Under Targets and Tagmentation):
    - Workflow: Lightly permeabilize nuclei/cells → primary antibody binds target → secondary antibody recruits Protein A/G–Tn5 transposase loaded with adapters → tethered Tn5 cleaves and tags DNA at antibody-bound sites → direct PCR amplification → sequencing.
- **Comparative Discussion**
  - **Resolution and signal-to-noise**:
    - **CUT&Tag**: High resolution and excellent signal-to-noise due to tethered cleavage and minimal background.
    - **ChIP-seq**: Broader fragments; background from nonspecific fragmentation and immunoprecipitation.
  - **DNA fragment characteristics**:
    - **CUT&Tag**: Generates DNA fragments of highly consistent and appropriate lengths, which significantly improves the efficiency and accuracy of *de novo* DNA sequencing.
    - **ChIP-seq**: Produces a broader range of fragment sizes due to nonspecific sonication, complicating subsequent sequencing and assembly.
  - **Workflow complexity and turnaround**:
    - **CUT&Tag**: Fewer steps; direct library generation via Tn5 reduces hands-on time.

        - **ChIP-seq**: More steps; longer turnaround.
    - **Compatibility**:
        - **CUT&Tag**: Works well for histone marks and many transcription factors; dependent on antibody quality and accessibility.

3. Nucleic Acid Chemistry, Directionality, and Nomenclature

- **DNA structure and numbering**
    - Backbone: Repeating deoxyribose and phosphate groups; bases attached at C1′.
    - Sugar numbering: C1′ (base), C2′ (H in DNA; OH in RNA), C3′ (OH), C4′, C5′ (phosphoryl linkage).
    - Strands are antiparallel with polarity (5′ \rightarrow 3′).
- **Phosphates on nucleotide building blocks**
    - Clarification from our discussion:
        - The phosphodiester backbone between nucleotides contains one phosphate per linkage.
        - The 5′ end of a nucleic acid chain typically bears a monophosphate (5′-P) or hydroxyl, depending on preparation.
        - The incoming building blocks for polymerization are nucleoside triphosphates:
            - For DNA synthesis: dNTPs (dATP, dCTP, dGTP, dTTP).
            - For RNA synthesis: NTPs (ATP, CTP, GTP, UTP).
        - These carry a triphosphate on the 5′ carbon; during incorporation, pyrophosphate (PPi) is released, leaving a single phosphate in the backbone.
        - Conclusion recorded: Each incorporated nucleotide contributes one phosphate in the backbone; the free nucleotide substrates are triphosphates prior to incorporation.
- **Abbreviations**
    - dNTP: Deoxyribonucleoside triphosphate (dATP, dCTP, dGTP, dTTP).
    - ATP/GTP/CTP/TTP: Ribonucleoside triphosphates for RNA (note: UTP replaces TTP in RNA; TTP is typically used in DNA contexts as dTTP).
- **Template vs non-template strands**
    - Template strand: The DNA strand read by polymerases to synthesize a complementary strand.
    - Non-template (coding) strand: Same sequence as the RNA transcript (except T ↔ U), not directly read during transcription.
- **Directional conventions in core processes**
    - DNA replication: DNA polymerases extend DNA only (5′ \rightarrow 3′); leading strand continuous, lagging strand discontinuous (Okazaki fragments).
    - Transcription: RNA polymerase synthesizes RNA (5′ \rightarrow 3′) while reading the DNA template (3′ \rightarrow 5′).
    - Translation: Ribosomes read mRNA (5′ \rightarrow 3′); polypeptide synthesized from N-terminus to C-terminus.

4. DNA Sequencing Quality Metrics and GC Content

- **Q-scores (Phred quality)**
    - Definition: $Q = -10 \log_{10}(P_{\text{error}})$.
    - Common thresholds:
        - Q20: $P_{\text{error}} = 1\%$ (accuracy ~99%).
        - Q30: $P_{\text{error}} = 0.1\%$ (accuracy ~99.9%).
    - Usage:
        - Reported as per-base scores and aggregate metrics (e.g., %≥Q30).
        - Higher %≥Q30 indicates better overall run quality; filters for downstream analyses.
- **GC content**
    - Meaning: Fraction of G and C bases in reads or genome regions.
    - Why we check it:
        - Detects library prep or sequencing biases (e.g., drop-outs at extreme GC).
        - Flags contamination or sample swaps if observed GC deviates from expected.
        - Informs normalization and peak-calling (AT/GC bias corrections) and guides PCR conditions.