BioPhorum
ADVANCED THERAPY MEDICINAL PRODUCTS

Recommendations for the Validation of rAAV Identity by Next Generation Sequencing

Jarrod Dean, Associate Director, Genomic Medicine Unit
BioAnalytics, NGS CMC
Sanofi

On behalf of BioPhorum ATMP Next Generation Sequencing workstream

USP Biologics Stakeholder Forum
Innovative Analytical Approaches CGT
Rockville, MD; 22FEB2024

CONNECT
COLLABORATE
ACCELERATE™

Permission to use

The contents of this report may be used unaltered as long as the copyright is acknowledged appropriately with correct source citation, as follows ‘Entity, Author(s), Editor, Title, Location: Year’

Disclaimer

This document represents a consensus view, and as such it does not represent fully the internal policies of the contributing companies.

Neither BioPhorum nor any of the contributing companies accept any liability to any person arising from their use of this document.

The views and opinions contained herein are that of the individual authors and should not be attributed to the authors’ employers.

CONNECT COLLABORATE ACCELERATE
is a trademark of BioPhorum Operations Group.

Agenda

- Introduction to BioPhorum & ATMP Phorum
- Product Identity testing
- Recombinant Adeno Associated Virus (rAAV) Identity Testing
- Historical Methods for rAAV Genomic Identity Testing
- Next Generation Sequencing for rAAV Identity Testing

- BioPhorum White Paper Outline for NGS ID Testing of rAAV
  - Initial considerations to Establish Internal NGS Workflows
  - Sequencing Library Preparation
  - Method Development Considerations: Sample Preparation
  - Sequencing
  - Bioinformatics (Analysis)
  - Validation Plan Proposal
  - Validation Plan Proposal (Specificity)
  - Validation Plan Proposal (Instrument & Analysis)
  - Long-read Sequencing Technologies
  - Comparison of Sequencing Technologies for rAAVs and Use

- Conclusion

BioPhorum: a co-ordinated program of industry change

BioPhorum creates an environment where the global biopharmaceutical and device industry can collaborate and accelerate their rate of progress, for the benefit of all. We do this by:

Bringing leaders together to create future visions that focus the industry's energy on the key emerging opportunities

Mobilizing communities of the top experts around these opportunities, up and down the biopharma value chain

Creating partnerships that enable change and provide the quickest route to implementation and results

Replacing isolation with collaboration so that the industry shares, learns and builds the best solutions together

…making the journey better, faster and cheaper than it would be for individual companies to do it on their own.

There are currently 11 Phorums providing a wealth of opportunities for companies to align their interests with similarly committed organizations.

150+
member companies

120+
global programs for change

7500+
leaders and subject matter experts

150
published papers, presentations and resources in the last 12 months

ADVANCED THERAPY MEDICINAL PRODUCTS (ATMP)

Formed in 2018, the ATMP Phorum supports the quest for better and faster development of Cell, Gene (in vivo & ex vivo) and RNA therapies.

In this collaboration, we connect drug development and contract manufacturing organizations with the aim of ensuring harmonization and alignment around many issues. These include potency assays, phase-appropriate guidance for critical quality attributes, operator safety, regulatory guidelines, and ATMP specific validation issues, as well as working towards resolving the current challenges to commercializing ATMP products.

The business of cell, gene and RNA therapies is as diverse as the patients which it serves and for that there is no one size fits all solution.

ATMP members have identified eleven high level areas of interest, along with additional specific challenges to resolve. Addressing these challenges will help to avoid multiple industry solutions.

HIGH LEVEL WORKSTREAMS

- Cell therapy (+/- gene modified)
- *In vivo* gene therapy
- RNA
- Development
- Analytics & assay validation
- Commercial Readiness & CMC
- Regulatory
- EHS & Biosafety
- Raw Materials

There are now two types of workstreams:

- **High Level Workstreams** allowing opportunities for quick topic discussion across the industry
- **In-Depth Workstreams** which will focus on a specific focused topic working towards a defined deliverable.
  - Currently **14 in-depth workstreams**

23
Workstreams
9 High Level + 14 In-Depth

3
F2F / Virtual Meetings

66
External publications, presentations, webinars & podcasts

900+
Active participants

34
Member Companies

BioPhorum

Product Identity Testing

- Product lot testing ensures the safety and efficacy of a therapeutic active pharmaceutical ingredient
- Identity testing is specific to each product, confirming that the product is what is stated on the final container/package label
- Must be able to distinguish from other products being manufactured / processed in the same laboratory
- Release method (i.e., qualified / validated)
- Identity is established in multiple Code of Federal Regulations (CFR) Title 21 parts/subchapters

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER F - BIOLOGICS

PART 610 -- GENERAL BIOLOGICAL PRODUCTS STANDARDS

Subpart B - General Provisions

Sec. 610.14 Identity.

The contents of a final container of each filling of each lot shall be tested for identity after all labeling operations shall have been completed. The identity test shall be specific for each product in a manner that will adequately identify it as the product designated on final container and package labels and circulars, and distinguish it from any other product being processed in the same laboratory. Identity may be established either through the physical or chemical characteristics of the product, inspection by macroscopic or microscopic methods, specific cultural tests, or in vitro or in vivo immunological tests.

https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=610.14

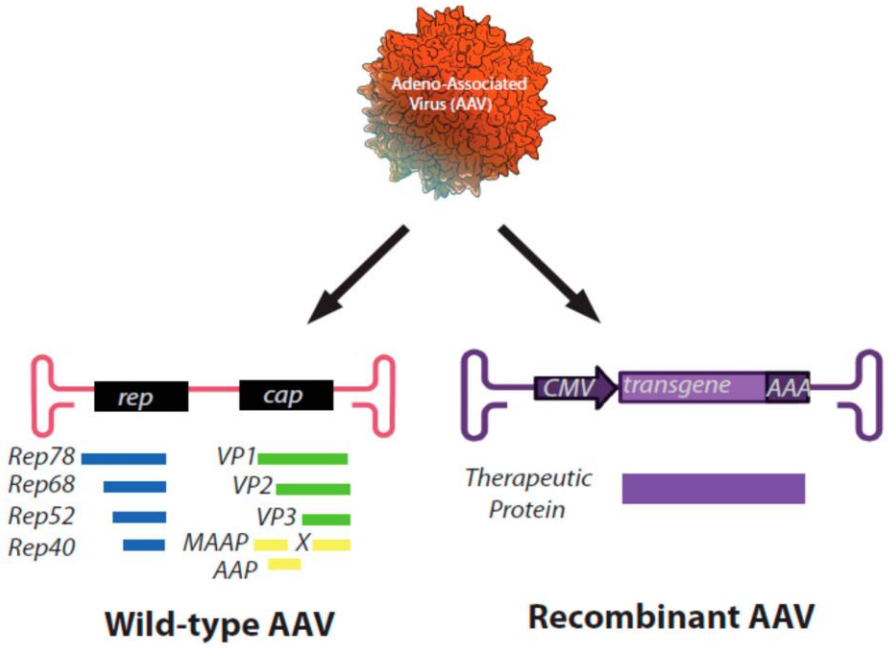
Title 21 / Chapter I / Subchapter D / Part 312

(a) ***Drug substance.*** A description of the drug substance, including its physical, chemical, or biological characteristics; the name and address of its manufacturer; the general method of preparation of the drug substance; the acceptable limits and analytical methods used to assure the identity, strength, quality, and purity of the drug substance; and information sufficient to support stability of the drug substance during the toxicological studies and the planned clinical studies. Reference to the current edition of the United States Pharmacopeia—National Formulary may satisfy relevant requirements in this paragraph.

(b) ***Drug product.*** A list of all components, which may include reasonable alternatives for inactive compounds, used in the manufacture of the investigational drug product, including both those components intended to appear in the drug product and those which may not appear but which are used in the manufacturing process, and, where applicable, the quantitative composition of the investigational drug product, including any reasonable variations that may be expected during the investigational stage; the name and address of the drug product manufacturer; a brief general description of the manufacturing and packaging procedure as appropriate for the product; the acceptable limits and analytical methods used to assure the identity, strength, quality, and purity of the drug product; and information sufficient to assure the product's stability during the planned clinical studies. Reference to the current edition of the United States Pharmacopeia—National Formulary may satisfy certain requirements in this paragraph.

IND Applications

Recombinant Adeno-associated Virus Identity Testing



Modified from Bower et al. 2021

GUIDANCE DOCUMENT

Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs)

Guidance for Industry

JANUARY 2020

ii. Identity

We recommend that identity assays uniquely identify a product and distinguish it from other products in the same facility. This test is performed on the final labeled product to verify its contents (21 CFR 610.14). Sometimes, a single test is not sufficient to distinguish clearly among products, and therefore, it is good practice to use different types of test methods (e.g., vector genome restriction digest and protein capsid analysis).

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/chemistry-manufacturing-and-control-cmc-information-human-gene-therapy-investigational-new-drug

Identity is typically measured from the 5'-ITR through the 3'-ITR with emphasis on the therapeutic transgene of interest.

Table 5. Analytical Tests for Cell and Gene Therapy Biological Products

Test	Gene-Modified Cellular Gene Therapy Product	Gene Therapy Products	
		Viral	Nonviral
Identity of Biological Substance	• Surface marker determination • Species • Morphology • Bioassay • Biochemical Marker	—Restriction enzyme map —PCR —Immunoassay for expressed gene —Sequencing	—Restriction enzyme map —PCR —Immunoassay for expressed gene —Sequencing

United States Pharmacopeia (2023). *General Chapter, (1047) Gene Therapy Products*. USP-NF. Rockville, MD: United States Pharmacopeia.

Historical Methods for rAAV Genomic Identity Testing

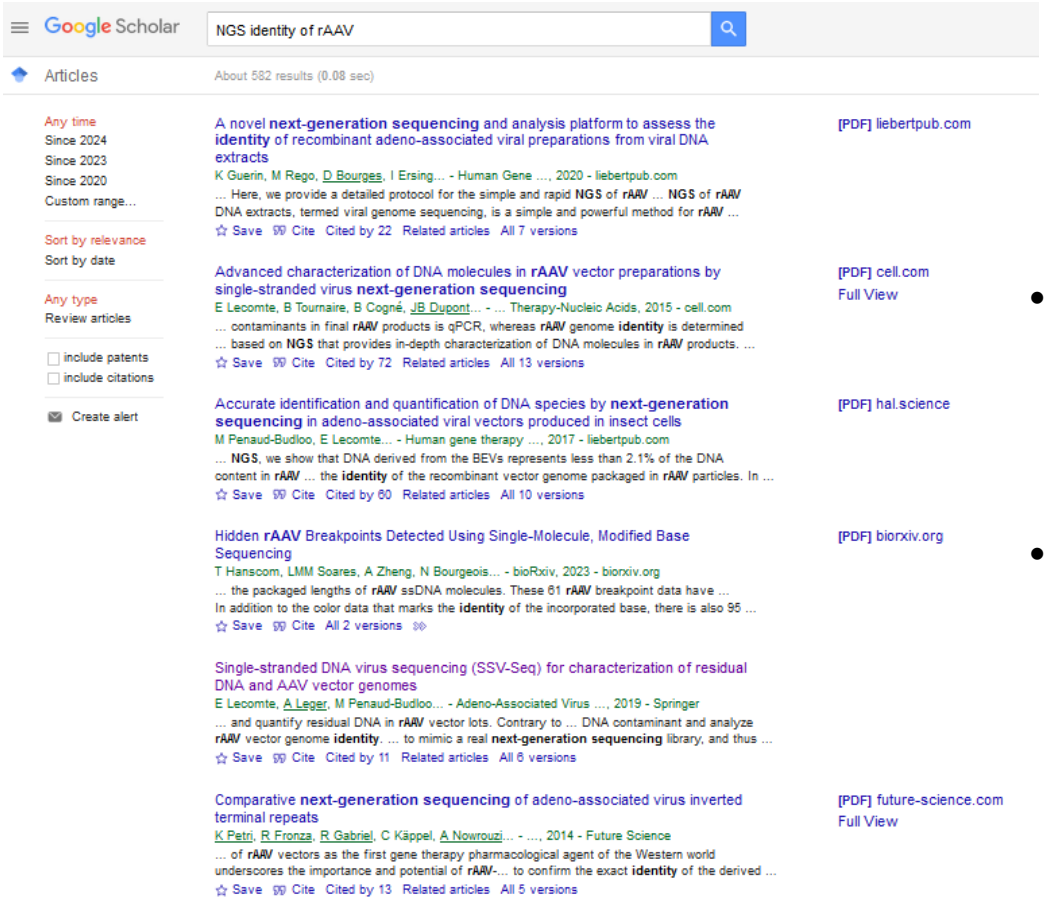
Method	Strategy	Pros	Cons
qPCR/dPCR	Product-specific amplification	- Well established methodologies - Minimum instrumentation with IQ/OQ - Use of product specific oligos - Multi-attribute with vector genome	- Might not span entire ITR-ITR sequence - Multiple assays required for genome integrity - PCR assay spanning complete ITR is difficult - Difficulty amplifying secondary structures
Restriction Enzyme Digest with PCR	Product-specific amplification with sequence specificity to restriction enzyme cut sites	- Target majority of transgene - Added sequence specificity with RE sites - Make unique fingerprint for each product - Compare pattern directly to reference	- Assay design more complex - Requires more instrumentation - Limited to number of samples per assay
Sanger Sequencing	Direct sequencing of full transgene	- Full ITR - ITR sequence coverage	- Low sequence depth (read accuracy) -Not enough reads for variant analyses -Difficulty sequencing through ITRs

- Non-inclusive list
- Shortcomings with above historical methods
  - Inability to provide an ITR – ITR full length sequence coverage readout (q-/dPCR, RE)
  - Depth of coverage insufficient for robust variant analyses (Sanger)

How can Next Generation Sequencing Improve rAAV Identity Confirmation?

Next Generation Sequencing for rAAV Identity Testing

- Massively parallel sequencing technology from an unbiased, holistic perspective
- Drastically increase the per base coverage of the sample, providing greater coverage depth
  - Increase confidence of variant calling



- Growing interest in industry of using NGS as an analytical tool
  - Identity and other parameters for product characterization
- Approaches are being provided by CROs/CDMOs
  - Offering validated solutions for identity testing and confirmation

BioPhorum Whitepaper for NGS Identity Testing of rAAV

Objectives

- Build on recent industry interest and promote NGS as an analytical tool for rAAV testing
- Define NGS rAAV identity testing workflow using a modular approach via short-read sequencing
    - Sample preparation
    - Library preparation
    - Sequencing
    - BioInformatics / Analysis
- Considerations for method development and control strategies for each workflow module
    - Identify target system suitability and sample acceptance criteria to ensure quality results
- Considerations for method validation
- High-level overview of other technologies and how they can be used for rAAV characterization
- Whitepaper is currently in draft and includes input from 29 SMEs across 21 member companies

Initial Considerations for Internal NGS Workflows

- Segregation of Pre and Post PCR Activities
  - Avoid contamination during workflow
    - No template clean area (reagent preparation)
    - Pre-PCR clean area (sample preparation)
    - Post-PCR area (PCR and sequencing workflows)

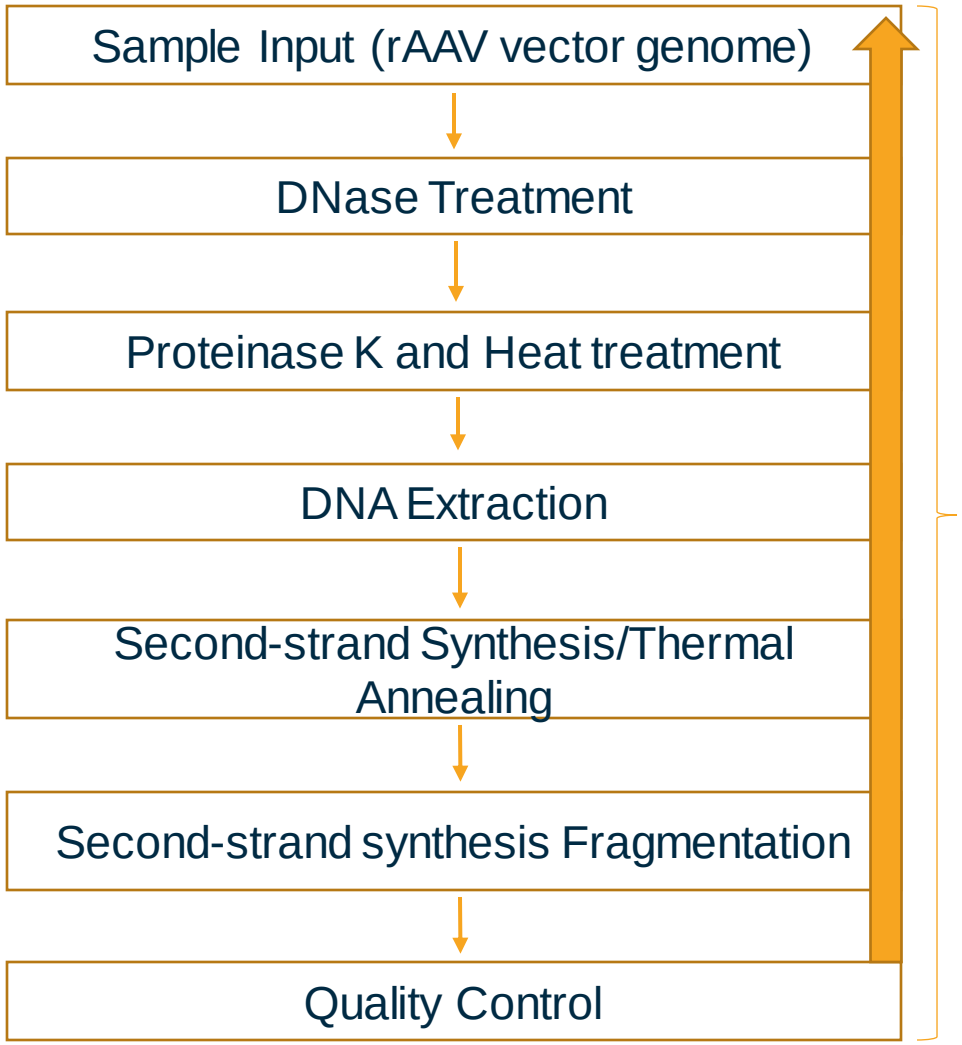
  Engineering controls (i.e., dedicated BSC) in lab if space is limited

- Lab Space and Utilities
  - Instrument location (i.e., vibration considerations, direct sunlight)
  - Lab temperature and humidity considerations
  - Uninterrupted power supply (UPS)
  - High speed network for large data management
  - Other considerations depending on sequencer (i.e., Nitrogen line)

- Instrument and Analysis Validation and GMP Compliance for Data Integrity and Storage
  - Adherence to cGMP guidance (i.e., 21 CFR Part 11 and Annex 11) for all workflow instruments
  - Validation plan for analysis application and environment
    - Requires input from matrixed team (i.e., Digital, IT and local business quality representatives)
  - Data storage considerations
  - Production environment support with established change control process

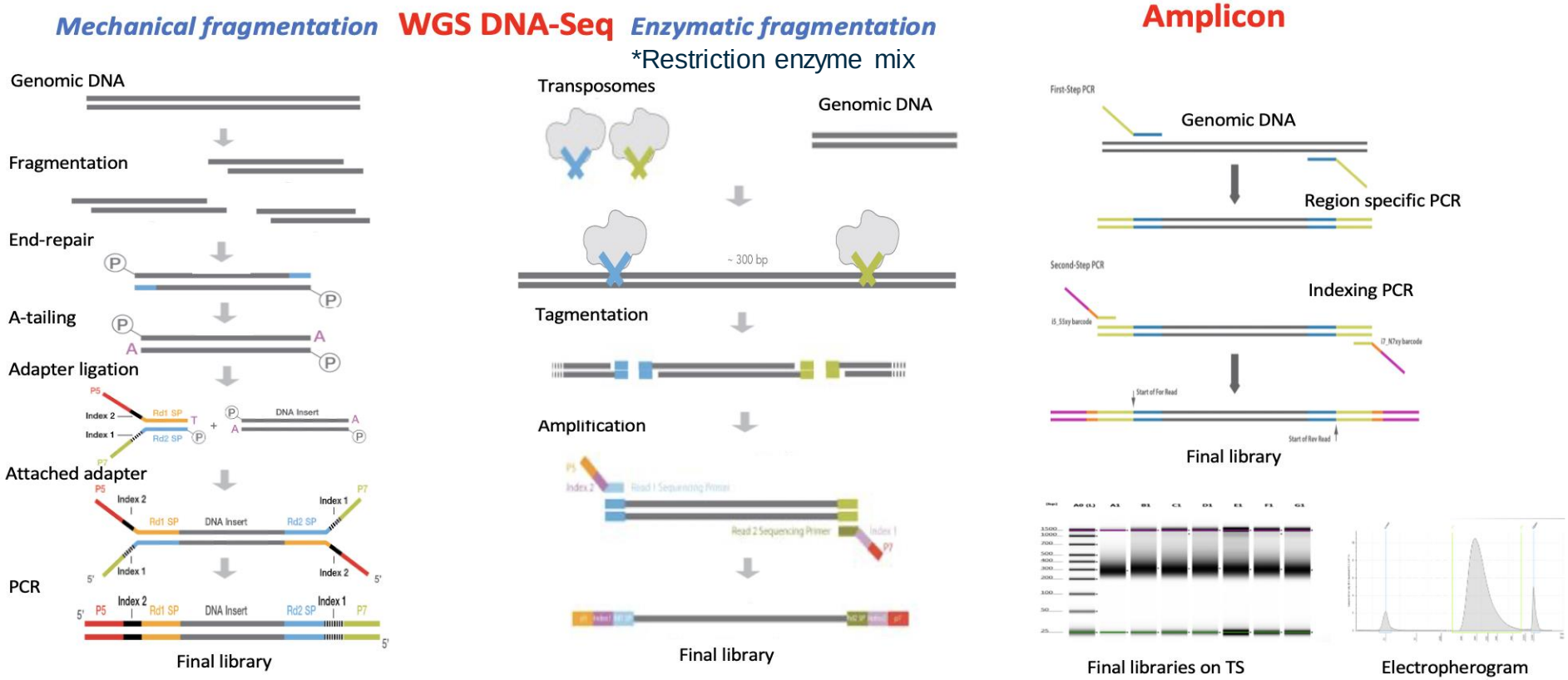
Method Development Considerations: Sample Preparation

Example Short-read SBS workflow



- Each step could be considered a critical assay attribute
  - Internal method development should be considered
  - Define robust range to ensure optimal method performance
- Establish control to monitor assay performance
  - Controls may not transfer over to final version of method
    - Spiked DNA control to show DNase treatment efficiency
- Suggested quality metrics for performance monitoring
  - Final extracted DNA
    - Yield and purity
  - DNA sizing (post fragmentation)
    - Fall within established size range
- Potential opportunity for universal reference standard
  - Should be run during each analysis as a quality control

Sequencing Library Preparation



Modified from: https://know ledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000006561

- Many commercial library kits available for sequencing library generation
  - Recommended to perform due diligence and compare kits for performance check
- Recommended quality metrics
  - DNA library size distribution, purity, and concentration
- “One size fits all” method approach may not be appropriate when considering unique transgene configurations

Sequencing

- Sequencing run quality is directly related to sequencing library quality
  - Robust wet-lab procedure is required for high quality results
- Required results will inform sequencing platform and associated kit
  - Depth of coverage and/or number of samples, per run

Illumina MiSeq

Maximum Output	8.5 Gb (500-cycle MiSeq Reagent Kit v2), 5.1 Gb (300-cycle MiSeq Reagent Kit v2), 1.2 Gb (300-cycle MiSeq Reagent Micro Kit v2), 0.850 Gb (50-cycle MiSeq Reagent Kit v2), 0.5 Gb (500-cycle MiSeq Reagent Nano Kit v2), 0.3 Gb (300-cycle MiSeq Reagent Nano Kit v2)
Maximum Reads per Run	Up to 15 million
System Compatibility	MiSeq, MiSeq, MiSeq FGx in Research Mode, MiSeqDx in Research Mode
Reagent Type	Cluster Generation, Paired-End Sequencing, Sequencing by Synthesis
Nucleic Acid Type	RNA, DNA
Technology	Sequencing

Reagent Kit	No. of Reads	Kit Size (cycles)	Output (max.)	2 × 75 Output	2 × 300 Output
MiSeq Reagent Kit v3	25 M	150, 600	15 Gb	3.8 Gb	15 Gb

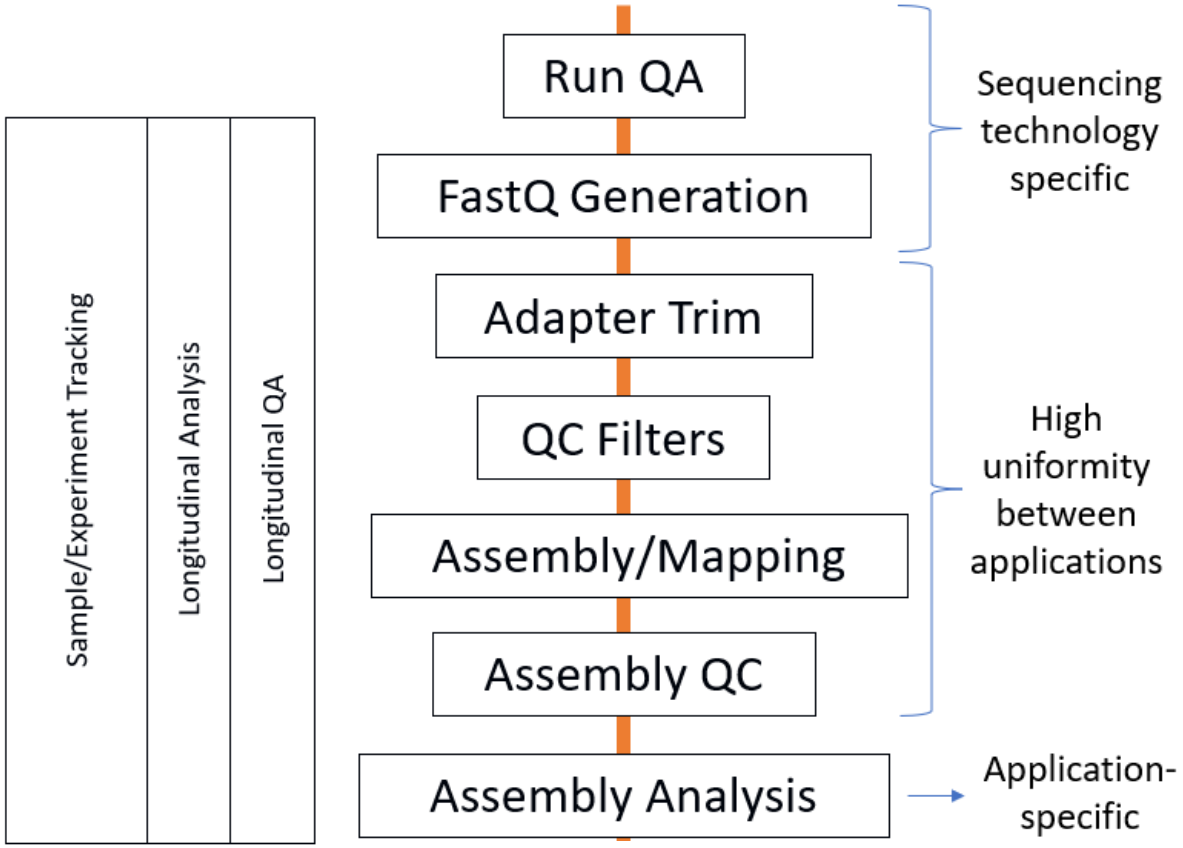
Gb = gigabases, M = millions

https://emea.illumina.com

- Sequencing quality considerations
  - Cluster density range
  - Phred score (% Q30)
  - Reads passing filter
  - Number of reads per sample

Percent PhiX control spike consideration, depending on complexity of library, to monitor error rate

Bioinformatics (Analysis)



Multiple Tooling Options

Step	Tool	Language	Note
Adapter Trimming	Trimmomatic	Java	Widely used
	bbDuk	Java	K-mer based, part of bbTools
	Cutadapt	Python	Simple
	Trim Galore	Perl and Python	Wrapper for Cutadapt
Assembly	SPAdes	C++	Small genomes, single-cell and metagenomics
	Velvet	C	One of the first de- novo assemblers
	ABySS	C++	Large genomes
Mapping	BBMap	Java	Accurate, fast, high precision
	BWA	C	Widely used for large genomes
	Novoalign	C++	Commercial, accurate for a cost
	Bowtie2	C++	Balances speed and accuracy

- Requires coding environment to support multiple scripting languages
  - Multiple programs should be evaluated during method development as they could impact results
  - Code development should be conducted using semantic version control to track and comment changes
- Final outputs could include alignment mapping (depth of reads), consensus sequence, and variant detection
- Quality criteria should be built around percent of reads aligning to transgene and depth coverage

Validation Plan Proposal

TABLE

TYPE OF ANALYTICAL PROCEDURE; CHARACTERISTICS	IDENTIFICATION	TESTING FOR IMPURITIES		ASSAY; dissolution (measurement only); content/potency
		Quantitation	Limit	
Accuracy	-	+	-	+
Precision				
Repeatability	-	+	-	+
Interm. Precision	-	+[1]	-	+[1]
Specificity[2]	+	+	+	+
Detection Limit	-	-[3]	+	-
Quantitation Limit	-	+	-	-
Linearity	-	+	-	+
Range	-	+	-	+

NOTE: - signifies that this characteristic is not normally evaluated; + signifies that this characteristic is normally evaluated

[1] in cases where reproducibility (see Glossary) has been performed, intermediate precision is not needed

[2] lack of specificity of one analytical procedure could be compensated by other supporting analytical procedure(s)

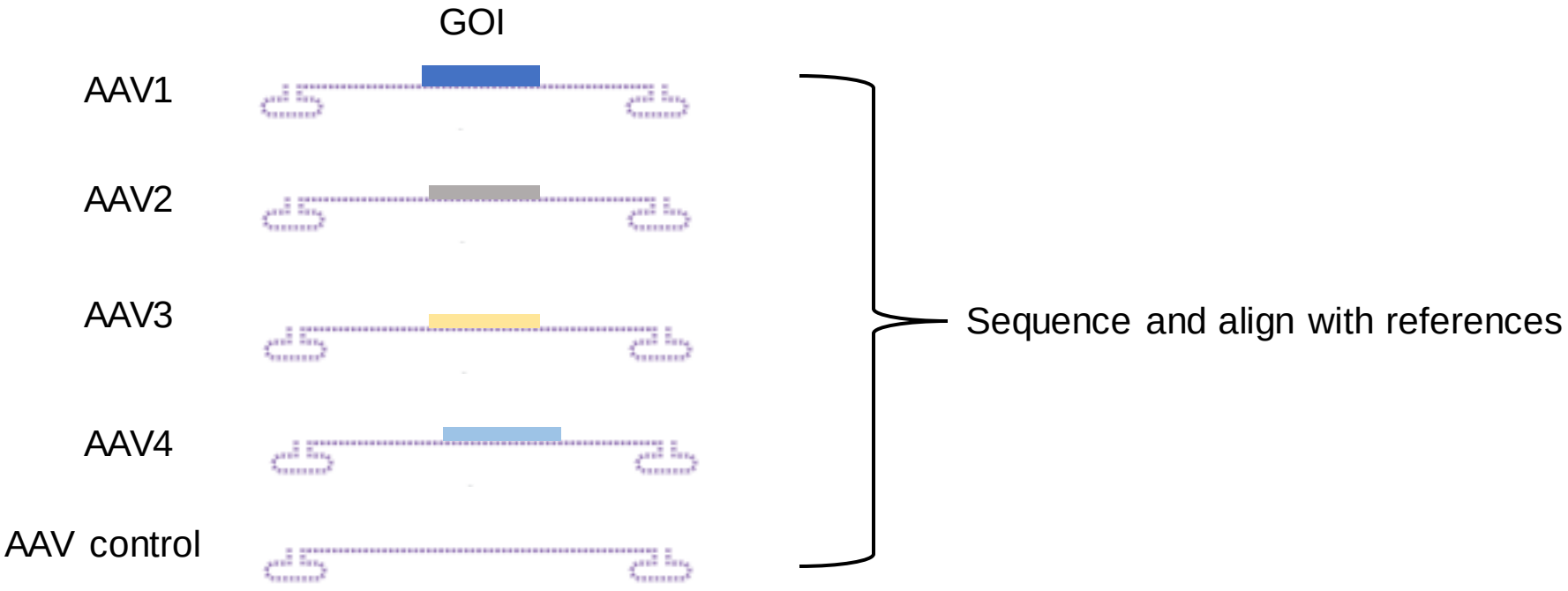
[3] may be needed in some cases

ICH Q2(R2) Guidelines

- NGS is a complicated assay
- Will detect all sequences present in a sample
  - Common elements between rAAV vectors?
- Reads are fragmented for short-read workflow
  - Not continuous ITR – ITR sequence detection
- Best way to prove identity confirmation of a specific product?

Validation Plan Proposal (Specificity)

- Four (or more) different rAAV vectors and established control sequenced in same run
- The reference database will be a combination of the AAV transgene sequences
- Specificity
  - The reads from each rAAV will be aligned with the dataset
  - Therapeutic gene of interest (GOI) detected in corresponding reference only



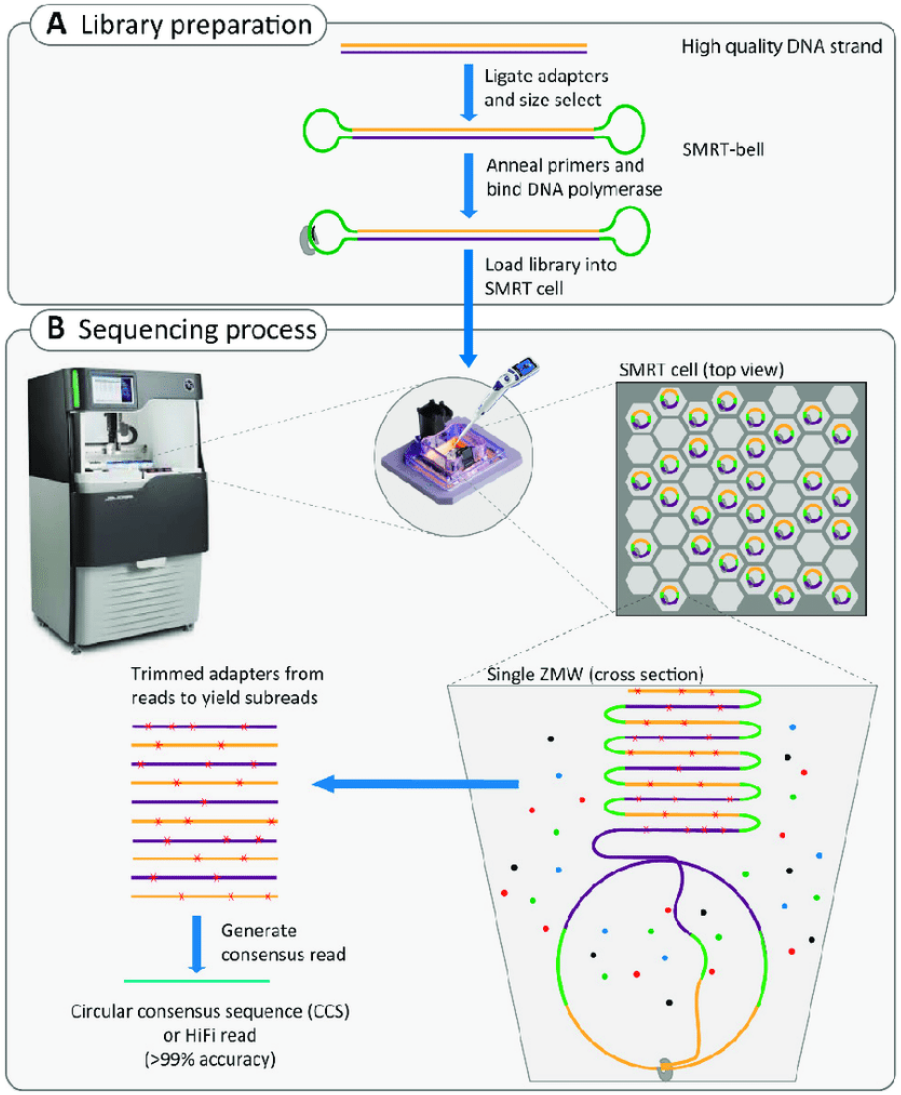
- Must also include ITR – ITR confirmation of each AAV product in separate analyses
- Consider running formulation buffer / NC preparation to control reagent suitability

Validation Plan Proposal (Instrument and Analysis)

- Instrumentation
  - Validation handled by internal company standards (i.e., URS, SOP, etc.)
  - Review of vendor IQ/OQ documentation and generation of supplemental protocol (if necessary)
- Analysis Pipeline and Environment
  - Conforms to regulatory standards (i.e., 21CFR Part 11, Annex 11, etc.)
    - Guidance for Industry and FDA Staff: General Principles of Software Validation
  - Internal risk assessment based upon function and non-functional requirements
  - Internal code review of pipeline
    - Use of version-controlled processes (i.e., GIT)
  - Inclusion of audit trail for application / environment performance confirmation
  - Documented test cases in UAT environment
  - Wet-lab validation work helps inform suitability of dry-lab validation
- Robust validation package should include full description and functionality of the analysis application, environment architecture, UAT and wet-lab results

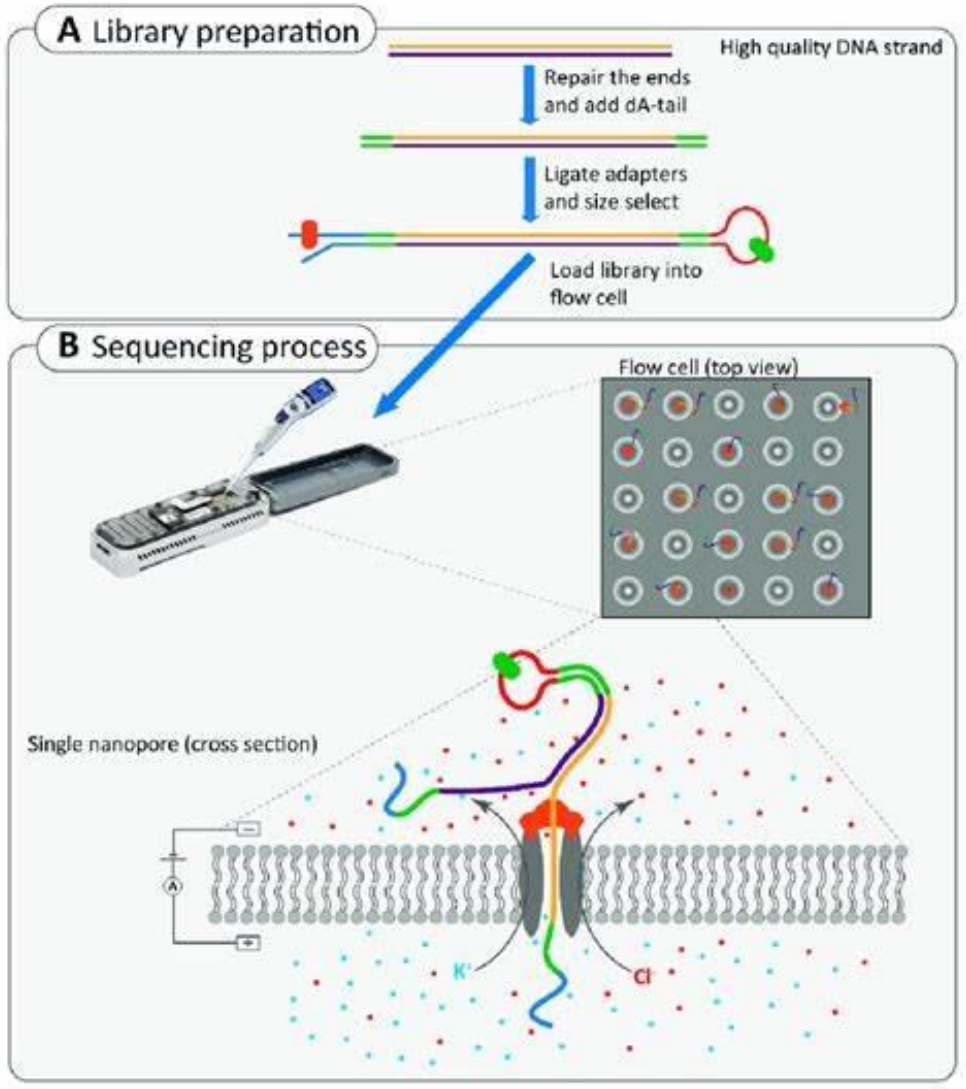
Long-Read Sequencing Technologies

Single-Molecule Real-Time (SMRT): PacBio



de Bruijn, S.E. *et al*. 2021. https://doi.org/10.3390/ijms22062943

Nanopore Sequencing: Oxford Nanopore Technologies



de Bruijn, S.E. *et al*. 2021. https://doi.org/10.3390/ijms22062943

Comparison of Sequencing Technologies for rAAV

Table 5. Summary of short-read and long-read NGS technologies.

	Short-Read NGS	Long-Read NGS
Method Advantages	• High number of reads and low error rates allows for deep coverage of sequences and a lower threshold of detection for minor variants and impurities present • Greater accuracy for analysis of low-representative molecules • Large numbers of samples can be multiplexed via barcoding, reducing the per-sample cost significantly • Minimal length-bias results in a greater potential for accurate quantification • Ability to analyze methylation	• Long read length allows for single-read coverage across the entire packaged genome length (~5kb) reducing the complexity of bioinformatic analysis, and allowing for the identification of DNA truncation sites, chimeras, and other rearrangements • Ability to generate large number of reads allowing for deep coverage and identification of minor sequence variants • Ability to analyze methylation
Method Disadvantages	• Short reads do not allow for coverage of entire genome length (<800bp per read) from a single read and instead relies on post-processing bioinformatics for assembly • Poor sequencing efficiency in areas of high secondary structure (e.g., ITRs) • Computationally challenging to assemble and analyze • Inability to directionally phase repeat sequences (e.g., ITRs)	• Lower throughput and higher error rate • Read bias towards shorter reads reduces quantification accuracy for reads of varying lengths
Recommended Use Cases	• Allows for the identification and accurate quantification of single nucleotide polymorphisms (SNPs) and short indels • Quantification of non-payload DNA sequences (e.g., plasmid, host-cell, contaminant sequences)	• Identification of truncation hotspots, chimeras, overpackaging, and large rearrangements • Identification of non-payload DNA sequences (e.g., plasmid, host-cell, and other impurity sequences)

Dark Horse Consulting Group White Paper "Beyond Empty and Full: Understanding Heterogeneity in rAAV Products and Impurities"
https://www.darkhorseconsultinggroup.com/post/heterogeneity-in-raav-vector-genomes-etc

Short-read

Genomic Identity
Variant Detection
Contamination Detection
RNA-seq
% Indel Assessments
Adventitious Agent Testing
Viral Safety Testing
Integration Analyses

Long-read

Genomic Identity
Genomic Integrity
Partial/Truncation Analyses
Chimeric Sequence Identification
Contamination Detection
Contamination Sizing
Residual DNA Risk Assessments
Structural Variant Detection
Investigation Support
% Indel Assessments
Integration Analyses

- Combination of short- and long-read sequencing may be required for full analysis parameter understanding
- Robust method development must be conducted to understand impact of method and sequencing induced artifacts

Conclusions

- NGS can provide in-depth rAAV product understanding
  - Multiple sequencing technologies can provide enhanced insights to rAAV genomic diversity
- BioPhorum ATMP Next Generation Sequencing workstream is drafting a White Paper on the use of NGS for rAAV identity testing in a GMP setting
- There is expanding industry interest in establishing NGS as an analytical tool for rAAV
  - GMP services are being offered by external service providers
  - Updates to Regulatory Guidance are including the use of NGS [i.e., ICHQ5A(R2)]
- Current environment is supportive of utilizing NGS for rAAV release testing
  - Alignment with Regulatory Agencies through transparent communication on filings
    - Receive feedback from robust validation strategies and packages
  - Potential to establish universal controls to support method development and validation

Acknowledgements

AbbVie
Reshma Sundar

Bayer
Wenjie Yao
Jayashree Ray

Charles River Laboratories, Inc.
Georges Aboujaoudé

Eurofins BioPharma Product Testing
Weihong Wang

FUJIFILM Diosynth Biotechnologies
Andrew Hillhouse

Merck & Co., Inc., Rahway, NJ, USA
Meiping Chang

Passage Bio
Yuanyuan Gao

PolyPlus Transfection
Sylvain Julien

Regeneron
Aishwarya Shevade
David Van Houte

REGENXBIO Inc.
Bella Bidzhieva

Roche
Meenakshi Doma
Johannes Pschirer
Mehdi Sarmady

Sangamo Therapeutics
Rohit Rohra

Sanofi
Wei Zhang

SK pharmteco Cell & Gene US
Alexei Saveliev
Amanda Oran

UCB
Ishtiaq Khaliq

Ultragenyx Pharmaceutical
Wei Zhang

uniQure biopharma B.V.
Francesco Catania

BioPhorum
Niamh O'Kane

BioPhorum

For further information about BioPhorum and the ATMP Phorum, please contact Steven Wall (ATMP Phorum Director)

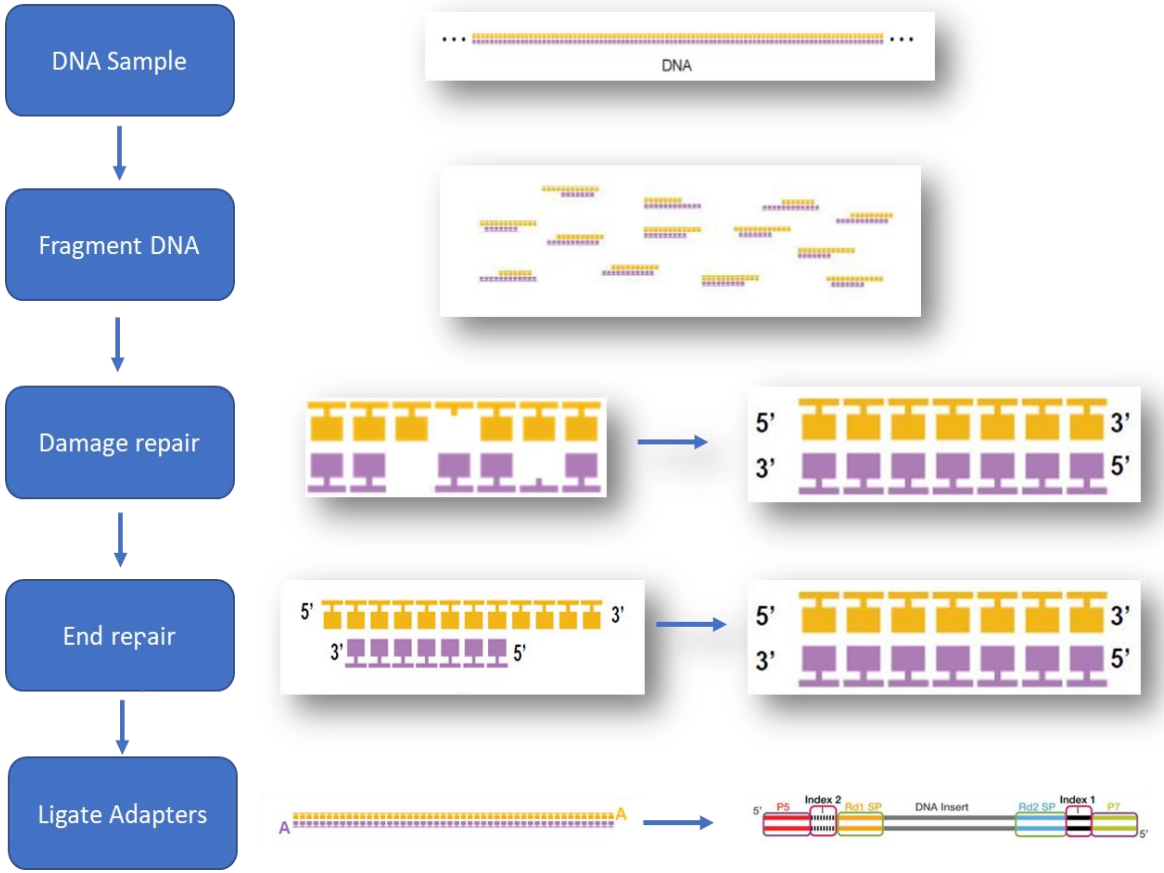
Steven.wall@biophorum.com

CONNECT
COLLABORATE
ACCELERATE™

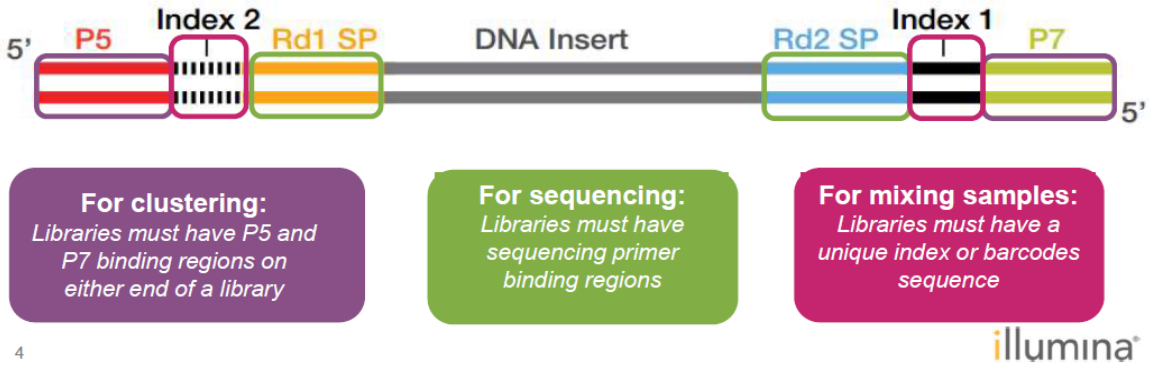
Extra Slides

Short-read Sequencing Overview

Sequencing by Synthesis Approach (Illumina)



Sample and Library Preparation

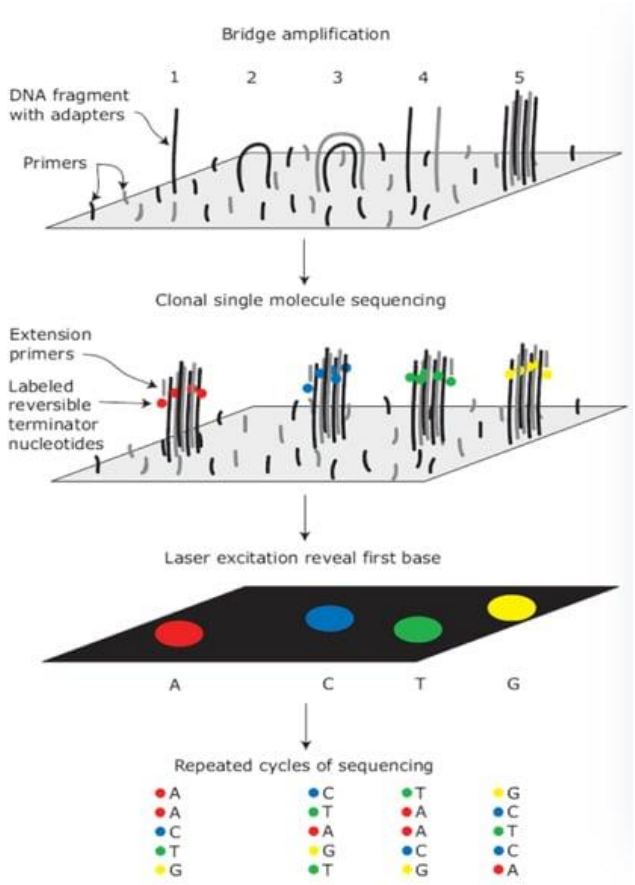


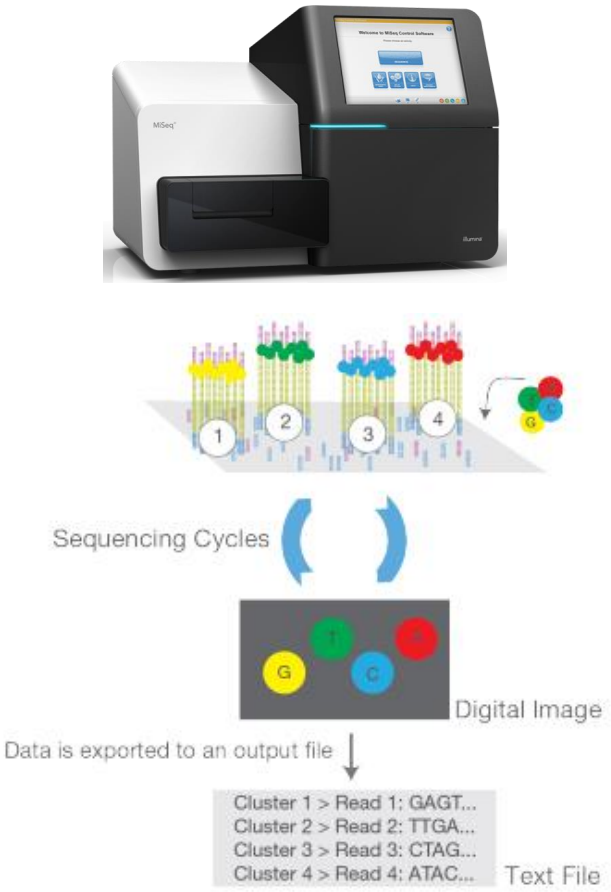
Final Sequencing Library Configuration

Modified from: https://www.futurelearn.com/info/courses/a-practical-guide-for-sars-cov-2-whole-genome-sequencing/0/steps/338197

Short-read Sequencing Overview

Sequencing by Synthesis Approach (Illumina)





https://www.sf2h.net/k-stock/data/uploads/2017/10/congres_SF2H_2018-06-06comp_corr.pdf

Modified from: https://1010genome.com/illumina-sequencing-explained/

Sequencing

Data Analysis

- Demultiplexing
  - Sample specific indexes
- .Fastq file generation (per sample)
  - Contains all reads
- Used for subsequent analyses
  - Reference alignment
  - Visualization