



FICHA TÉCNICA DEL CURSO

Control de Calidad y Preprocesamiento de Datos NGS

TC-BIO-01

Código:	FOR-EDU-005
Tipo de documento:	FORMATO
Fecha de emisión:	Enero 2026
Versión:	1.0
Elaboró:	Dirección General
Revisó:	Dirección General
Aprobó:	Dirección General

INFORMACIÓN DEL FORMATO

Este formato se utiliza para documentar la ficha técnica completa de cada curso o taller impartido por BioSeryl, especificando la estructura curricular, contenidos, metodología, evaluación y requisitos, conforme a NTC-ISO 21001:2025, Cláusula 8.1 (Planificación operacional) y Cláusula 8.5 (Prestación del servicio educativo). Una copia de esta ficha debe existir para cada curso activo.

Retención: Durante vigencia del curso + 2 años después de obsolescencia.

IDENTIFICACIÓN DEL PROGRAMA

Código del curso:	TC-BIO-01
Nombre completo:	Control de Calidad y Preprocesamiento de Datos NGS
Nombre corto (para marketing):	QC y Preprocesamiento NGS
Versión del currículo:	1.0
Fecha de creación:	Enero 2026
Última actualización:	Enero 2026
Próxima revisión:	Enero 2027
Estado:	<input type="checkbox"/> En desarrollo <input checked="" type="checkbox"/> Activo

CLASIFICACIÓN

Línea de producto:	<input checked="" type="checkbox"/> BIOINFORMATICA
Tipo de programa:	<input checked="" type="checkbox"/> Curso especializado (40–80h)
Nivel:	<input checked="" type="checkbox"/> Básico (sin prerrequisitos técnicos)
Modalidad:	<input checked="" type="checkbox"/> Sincrónico en vivo (videoconferencia)

DURACIÓN Y ESTRUCTURA

Duración total:	40 horas
Distribución:	Teoría: 40 % Práctica: 60 %
Número de módulos:	14
Número de sesiones:	14
Duración por sesión:	3 horas
Frecuencia sugerida:	2–3 veces/semana

PÚBLICO OBJETIVO

Perfil ideal:	Biólogos moleculares, microbiólogos, bioquímicos que generan o analizan datos NGS; estudiantes de maestría/doctorado con proyectos NGS; técnicos de laboratorios de secuenciación
Formación recomendada:	Biología molecular, bioinformática, ciencias biológicas (mínimo pregrado)
Prerrequisitos técnicos:	Ninguno (se enseña desde cero)
Prerrequisitos de software:	Ninguno (se usa entorno en la nube y Conda)

Cupo mínimo:	5 participantes
Cupo máximo:	25 participantes

DESCRIPCIÓN DEL PROGRAMA

Descripción:

Curso integral de control de calidad y preprocesamiento de datos de secuenciación de nueva generación (NGS), diseñado para investigadores que generan o analizan datos ómicos. Cubre el flujo completo desde la inspección de lecturas crudas (FastQC, MultiQC) hasta el filtrado, trimming y limpieza de datos para aplicaciones WGS, RNA-Seq, metagenómica y lecturas largas (PacBio, ONT). Incluye pipelines reproducibles con Docker/Singularity y estrategias de QC específicas para cada tipo de experimento NGS, con énfasis en la toma de decisiones informadas: distinguir problemas reales de advertencias cosméticas, optimizar parámetros según el contexto biológico, y documentar cada decisión para auditorías y reproducibilidad.

RESULTADOS DE APRENDIZAJE

Al finalizar este programa, el participante será capaz de:

No.	Resultado de aprendizaje
RA1	Interpretar reportes FastQC distinguiendo warnings cosméticos de problemas genuinos según contexto experimental.
RA2	Diseñar pipelines de preprocesamiento apropiados para diferentes tipos de experimentos NGS (WGS, RNA-Seq, Metagenómica).
RA3	Evaluar la calidad de lecturas largas PacBio y ONT con NanoPlot y PycoQC.
RA4	Implementar pipelines QC reproducibles usando contenedores Docker/Singularity.
RA5	Generar reportes MultiQC que documenten decisiones de QC para auditorías y reproducibilidad.

CONTENIDO MODULAR

Módulo	Título	Horas	Temas principales	RA
1	Fundamentos de NGS y control de calidad	3	Plataformas Illumina/PacBio/ONT, estándares de calidad, fuentes de error en secuenciación	RA1
2	FastQC: interpretación de reportes	3	Módulos FastQC, métricas por base/secuencia, adaptadores, duplicados, sobre-representación	RA1
3	MultiQC: integración de reportes	3	Agregación multi-muestra, reportes personalizados, alertas automatizadas	RA1
4	Trimming y filtrado con Fastp	3	Fastp, recorte adaptativo, filtrado por calidad, deduplicación	RA2
5	Trimming con Trimmomatic y Cutadapt	3	Trimmomatic (PE/SE), Cutadapt, parámetros óptimos por plataforma	RA2
6	Preprocesamiento para WGS	3	Flujo WGS, control de cobertura, estimación de profundidad, detección de contaminación	RA2, RA5
7	Preprocesamiento para RNA-Seq	3	rRNA removal, strandedness, UMI handling, gene body coverage	RA2, RA5
8	Preprocesamiento para metagenómica	3	Host removal, adapter trimming para metagenómica, calidad de reads largos	RA2, RA5
9	QC de lecturas largas PacBio	3	NanoPlot, PacBio HiFi QVs, read length distribution, concordancia	RA3
10	QC de lecturas ONT	3	PycoQC, NanoComp, calidad por barcode, estimación de N50	RA3
11	Pipelines QC con Conda	3	Entornos Conda, instalación reproducible, versionado de herramientas QC	RA4
12	Contenerización con Docker/Singularity	3	Imágenes Docker para QC, Dockerfiles, Singularity para HPC	RA4
13	Reportes MultiQC para auditoría	3	Personalización MultiQC, exportación para publicación, documentación de decisiones de QC	RA5
14	Proyecto integrador de QC	4	Pipeline QC completo multi-muestra con reporte auditado	RA5
Total		40		

METODOLOGÍA

Enfoque pedagógico

- | |
|--|
| <input checked="" type="checkbox"/> Aprendizaje activo con ejercicios prácticos
<input checked="" type="checkbox"/> Estudio de casos reales |
|--|

Descripción metodológica

Aprendizaje progresivo con enfoque problema-solución: cada sesión inicia con Análisis de dataset real problemático (baja calidad, contaminación, sesgos), identifica métricas anómalas, evalúa herramientas disponibles, diseña un flujo de corrección, ejecuta el Análisis y valida mejoras con métricas post-QC. Los participantes aplican procedimientos guiados, comparan estrategias alternativas y documentan decisiones metodológicas. Evaluación Diagnóstica (FOR-EDU-006): Cuestionario inicial aplicado en la primera sesión para identificar conocimientos previos y ajustar profundidad de contenidos. Evaluación Formativa: Retroalimentación continua durante cada sesión mediante revisión de ejercicios prácticos, monitoreo de comprensión y resolución de dudas en tiempo real. Evaluación Sumativa (FOR-EDU-007): Proyecto final integrador con calificación mínima de 4.0/5.0 para aprobación.

EVALUACIÓN

Instrumentos de evaluación

Instrumento	Peso (%)	Descripción
Ejercicios durante sesiones	40 %	Ejercicios prácticos por módulo
Proyecto final/integrador	50 %	Pipeline/proyecto integrador con reporte
Participación activa	10 %	Asistencia y participación en sesiones
Total	100 %	

Criterios de aprobación

Modalidad de aprobación:	<input checked="" type="checkbox"/> Ambos
Criterio:	Calificación mínima: 4.0/5.0 Asistencia mínima: 80 %
Certificado otorgado:	<input checked="" type="checkbox"/> Digital

MATERIALES Y RECURSOS

Tipo	Descripción
Datasets de práctica	Datasets de práctica (SRA NCBI, ENA/EBI)
Scripts de código	Scripts documentados (Bash) para cada sesión

Tipo	Descripción
Lecturas complementarias	Guía de instalación de herramientas (Conda environment YAML)
Lecturas complementarias	Manuales de referencia de herramientas

BIBLIOGRAFÍA DE REFERENCIA

No.	Referencia bibliográfica
1	Hamid D. Ismail (2023). <i>Bioinformatics: A Practical Guide to Next Generation Sequencing Data Analysis</i> . CRC Press. Capítulos 1, 4, 5: QC, FASTX-Toolkit, QualiMap, RSeQC. ISBN: 978-1-032-40900-9
2	Wang, X. (2023). <i>Next-Generation Sequencing Data Analysis</i> . CRC Press. Capítulos 1, 4.2, 8, 11: QC, ISO, validación clínica. ISBN: 978-0-367-34989-9
3	Datta, S. & Guha, S. (2023). <i>Statistical Analysis of Microbiome Data</i> . Springer. Capítulo 15: ITS analysis, ITSx, UNITE, decontam. ISBN: 978-1-0716-5008-0
4	FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
5	Trimmomatic: http://www.usadellab.org/cms/?page=trimmomatic
6	cutadapt: https://cutadapt.readthedocs.io/
7	fastp: https://github.com/OpenGene/fastp
8	BBTools (BBDuk, BBSplit): https://jgi.doe.gov/data-and-tools/bbtools/
9	MultiQC: https://multiqc.info/
10	NanoPlot/PycoQC: https://github.com/wdecoster/NanoPlot

INSTRUCTOR(ES) ASIGNADO(S)

Nombre	FOR-EDU-004

CONTENIDO PROGRAMÁTICO
Módulo 1 — Fundamentos de NGS y control de calidad (3 horas)

Unidad	Contenido
1.1	Plataformas de secuenciación Illumina: química, flujo de trabajo y formatos de datos
1.2	Plataformas PacBio y ONT: fundamentos de secuenciación de lecturas largas
1.3	Estándares de calidad en NGS: métricas de calidad Q-score, profundidad y cobertura
1.4	Fuentes de error en secuenciación: sesgos de amplificación, errores de secuenciación y contaminación

Módulo 2 — FastQC: interpretación de reportes (3 horas)

Unidad	Contenido
2.1	Módulos FastQC: Per base sequence quality, per sequence quality scores
2.2	Detección de adaptadores, duplicados y secuencias sobre-representadas
2.3	Interpretación de warnings: distinción entre problemas cosméticos y genuinos
2.4	Reportes combinados y exportación de resultados

Módulo 3 — MultiQC: integración de reportes (3 horas)

Unidad	Contenido
3.1	Agregación multi-muestra de reportes FastQC
3.2	Personalización de reportes MultiQC con módulos adicionales
3.3	Alertas automatizadas y generación de resúmenes
3.4	Exportación de reportes para documentación y auditoría

Módulo 4 — Trimming y filtrado con Fastp (3 horas)

Unidad	Contenido
4.1	Fastp: recorte adaptativo y filtrado por calidad
4.2	Deduplicación y corrección de errores con Fastp
4.3	Parámetros óptimos según plataforma y tipo de experimento



Unidad	Contenido
4.4	Reportes integrados de Fastp

Módulo 5 — Trimming con Trimmomatic y Cutadapt (3 horas)

Unidad	Contenido
5.1	Trimmomatic: modos paired-end y single-end
5.2	Cutadapt: recorte de adaptadores y primers
5.3	Parámetros óptimos por plataforma de secuenciación
5.4	Evaluación de calidad post-trimming

Módulo 6 — Preprocesamiento para WGS (3 horas)

Unidad	Contenido
6.1	Flujo completo de preprocesamiento para genomas completos
6.2	Control de cobertura y estimación de profundidad
6.3	Detección de contaminación en muestras WGS
6.4	Documentación de decisiones para reproducibilidad

Módulo 7 — Preprocesamiento para RNA-Seq (3 horas)

Unidad	Contenido
7.1	rRNA removal y enriquecimiento de mRNA
7.2	Detección de strandedness en librerías RNA-Seq
7.3	Manejo de UMI (Unique Molecular Identifiers)
7.4	Gene body coverage y evaluación de calidad

Módulo 8 — Preprocesamiento para metagenómica (3 horas)

Unidad	Contenido
8.1	Host removal: filtrado de genoma huésped
8.2	Adapter trimming específico para metagenómica
8.3	Calidad de reads largos en metagenómica
8.4	Estrategias de QC según tipo de muestra ambiental

Módulo 9 — QC de lecturas largas PacBio (3 horas)



Unidad	Contenido
9.1	NanoPlot: visualización de calidad de lecturas largas
9.2	PacBio HiFi QVs y métricas de precisión
9.3	Read length distribution y concordancia
9.4	Comparación entre corridas y condiciones

Módulo 10 — QC de lecturas ONT (3 horas)

Unidad	Contenido
10.1	PycoQC: control de calidad para datos Oxford Nanopore
10.2	NanoComp: comparación entre corridas ONT
10.3	Calidad por barcode en corridas multiplexadas
10.4	Estimación de N50 y métricas de ensamblaje

Módulo 11 — Pipelines QC con Conda (3 horas)

Unidad	Contenido
11.1	Creación de entornos Conda reproducibles
11.2	Instalación y versionado de herramientas QC
11.3	Automatización de pipelines con Conda
11.4	Exportación de entornos para reproducción

Módulo 12 — Contenerización con Docker/Singularity (3 horas)

Unidad	Contenido
12.1	Imágenes Docker para pipelines de QC
12.2	Creación de Dockerfiles con herramientas preinstaladas
12.3	Singularity para entornos HPC
12.4	Estrategias de versionado y distribución de imágenes

Módulo 13 — Reportes MultiQC para auditoría (3 horas)

Unidad	Contenido
13.1	Personalización avanzada de MultiQC
13.2	Exportación de reportes para publicación y auditoría
13.3	Documentación de decisiones de QC en el reporte
13.4	Automatización de reportes multi-muestra

Módulo 14 — Proyecto integrador de QC (4 horas)

Unidad	Contenido
14.1	Diseño de pipeline QC completo multi-muestra
14.2	Ejecución integrada de herramientas de QC
14.3	Generación de reporte auditado con MultiQC
14.4	Presentación y defensa de resultados

CONTROL DE VERSIONES DEL CURRÍCULO

Versión	Fecha	Cambios realizados	Autor
1.0	Enero 2026	Versión inicial	Dirección General

APROBACIÓN

Diseñó: Nombre: _____ Cargo: Director Departamento de Bioinformática Fecha: Enero 2026	Revisó: Nombre: _____ Cargo: Dirección General Fecha: Enero 2026	Aprobó: Nombre: _____ Cargo: Dirección General Fecha: Enero 2026
--	--	--



CONTROL DE CAMBIOS

Versión	Fecha	Descripción del Cambio	Autor
1.0	Enero 2026	Emisión inicial del formato	Dirección General