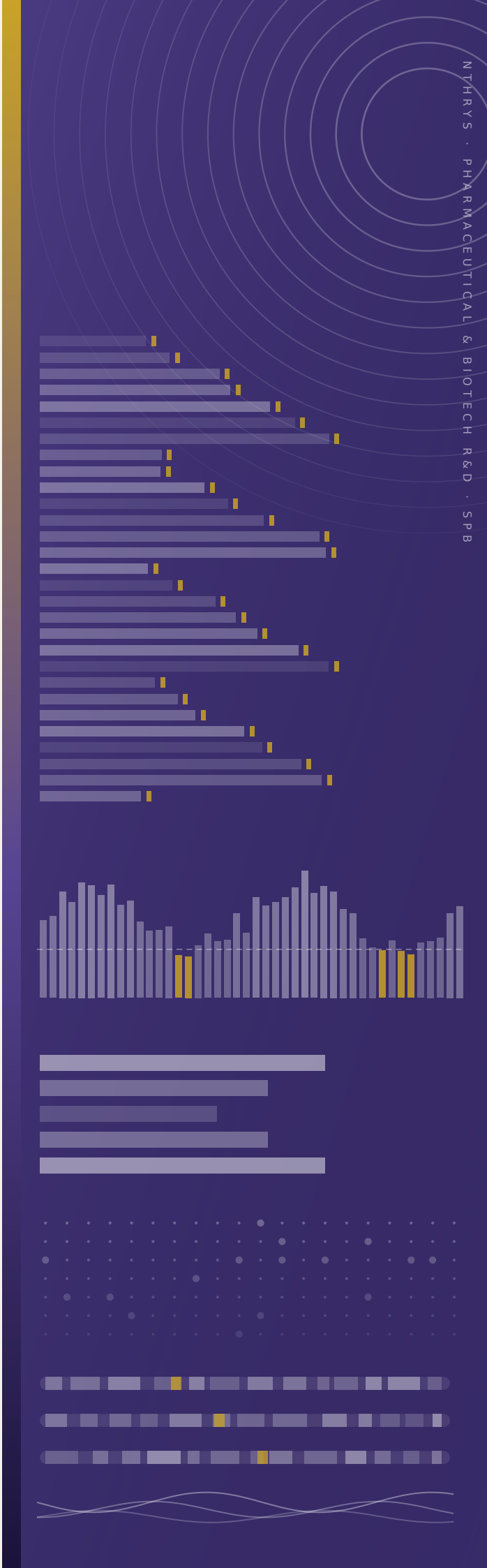
NTHRYS

OPC PVT LTD · HYDERABAD · INDIA

SECTOR CATALOGUE 04 · NGS DATA ANALYSIS

Pharmaceutical & Biotech *R&D*

Thirty-three industrial training programmes, one for each distinct type of NGS data analysis performed across pharmaceutical and biotechnology research and development, from target discovery through preclinical characterisation, clinical development and regulatory submission. Every programme answers the questions a laboratory asks before it commits budget — why the analysis is done, when it is done, what it detects, what it does not detect, what makes it hard, who does this job and what a trainee can do afterwards — and is followed by its complete ordered workflow, step by step.



33	668	3240	551
ANALYSIS TYPES	STEPS	HOURS	DAYS

CATALOGUE SPB · 2026 EDITION

INDUSTRIAL TRAINING · DISCOVERY & DEVELOPMENT PRACTICE

Pharmaceutical & Biotech R&D — *Analysis Types*

Thirty-three distinct types of NGS data analysis performed across drug discovery and development, each an independent analysis with its own input, its own pipeline and its own report set. Each is listed under its formal title with the industry name and the shorthand a laboratory uses in practice. Every analysis type occupies a two-page entry: the first page answers why, when, what it detects, what it does not detect, who does the job and what follows; the second sets out the complete ordered workflow.

CODE	ANALYSIS TYPE · INDUSTRY NAME & SHORT NAME	STEPS	HOURS	DAYS	FEE (INR)
SPB-01	**Target Discovery Transcriptomic Analysis** Transcriptomic target identification · Target discovery RNA-Seq	20	96	16	2,20,000
SPB-02	**Differential Expression & Pathway Analysis** Differential expression and enrichment analysis · DE analysis · pathway analysis	18	86	15	2,00,000
SPB-03	**Single-Cell Target Discovery Analysis** Single-cell transcriptomic target discovery · scRNA target discovery	23	112	19	2,55,000
SPB-04	**Pooled CRISPR Screen Analysis** Pooled genetic screen analysis · CRISPR screen analysis	22	106	18	2,45,000
SPB-05	**Perturb-Seq & Arrayed Screen Analysis** Single-cell perturbation screen analysis · Perturb-Seq analysis	24	116	20	2,65,000
SPB-06	**RNAi & shRNA Screen Analysis** RNA interference screen analysis · RNAi screen analysis	17	82	14	1,90,000
SPB-07	**Cell Line Genomic Authentication Analysis** Cell line identity and authentication analysis · Cell line authentication	14	66	11	1,50,000
SPB-08	**Cell Line Panel & Model Characterisation Analysis** Preclinical model genomic characterisation · Model characterisation	19	92	16	2,10,000
SPB-09	**Patient-Derived Xenograft Genomic Analysis** PDX model genomic analysis · PDX analysis	20	96	16	2,20,000
SPB-10	**Organoid & Primary Model Genomic Analysis** Organoid and primary culture genomic analysis · Organoid analysis	19	92	16	2,10,000
SPB-11	**Preclinical Pharmacogenomic & Toxicogenomic Analysis** Toxicogenomic and preclinical safety analysis · Toxicogenomics	18	88	15	2,05,000
SPB-12	**Drug Response & Sensitivity Prediction Analysis** Genomic drug response modelling · Sensitivity prediction	21	102	17	2,35,000
SPB-13	**Mechanism of Action Deconvolution Analysis** Compound mechanism deconvolution · MoA deconvolution	20	98	17	2,25,000
SPB-14	**Target Engagement & Pathway Modulation Analysis** Pharmacodynamic transcriptomic analysis · Target engagement analysis	18	88	15	2,05,000
SPB-15	**Biomarker Discovery Analysis** Genomic and transcriptomic biomarker discovery · Biomarker discovery	22	106	18	2,45,000
SPB-16	**Biomarker Validation & Assay Bridging Analysis** Biomarker validation and platform bridging · Biomarker validation	20	96	16	2,20,000
SPB-17	**Clinical Trial Genomic Stratification Analysis** Trial enrichment and stratification analysis · Trial stratification	21	102	17	2,35,000

Continued overleaf. Programme codes read SPB–01 and protocol references PRT/SPB/01.

Analysis Types — *18 to 33*

CODE	ANALYSIS TYPE · INDUSTRY NAME & SHORT NAME	STEPS	HOURS	DAYS	FEE (INR)
SPB-18	**Companion Diagnostic Development Analysis** Companion diagnostic analytical development · CDx development	24	118	20	2,70,000
SPB-19	**Pharmacogenomic Trial Analysis** Clinical trial pharmacogenomic analysis · Trial PGx analysis	19	92	16	2,10,000
SPB-20	**Resistance Mechanism Discovery Analysis** Acquired resistance mechanism discovery · Resistance discovery	21	102	17	2,35,000
SPB-21	**Immunogenicity & Anti-Drug Antibody Repertoire Analysis** Immunogenicity repertoire analysis · Immunogenicity analysis	18	88	15	2,05,000
SPB-22	**Antibody Discovery Repertoire Analysis** Antibody repertoire mining and discovery · Antibody discovery analysis	22	106	18	2,45,000
SPB-23	**CAR-T & Cell Therapy Product Characterisation Analysis** Cell therapy product genomic characterisation · Cell therapy characterisation	23	112	19	2,55,000
SPB-24	**Vector Integration Site Analysis** Integration site analysis for gene therapy · Integration site analysis	22	108	18	2,45,000
SPB-25	**On-Target Gene Editing Outcome Analysis** Editing outcome quantification · On-target editing analysis	20	98	17	2,25,000
SPB-26	**Off-Target Editing Assessment Analysis** Genome-wide off-target editing assessment · Off-target assessment	24	118	20	2,70,000
SPB-27	**mRNA & Nucleic Acid Therapeutic Sequence QC Analysis** Nucleic acid therapeutic sequence quality analysis · mRNA sequence QC	19	92	16	2,10,000
SPB-28	**Viral Vector Genomic Quality Analysis** Viral vector genomic characterisation and QC · Vector genomic QC	21	102	17	2,35,000
SPB-29	**Cell Bank & Adventitious Agent Screening Analysis** Sequencing-based adventitious agent testing · Adventitious agent screening	20	96	16	2,20,000
SPB-30	**Microbiome–Drug Interaction Analysis** Pharmacomicrobiomics analysis · Pharmacomicrobiomics	19	92	16	2,10,000
SPB-31	**Multi-Omic Integration Analysis** Integrated multi-omic analysis · Multi-omic integration	22	106	18	2,45,000
SPB-32	**Real-World & Cohort Genomic Evidence Analysis** Real-world genomic evidence analysis · RWE genomic analysis	20	98	17	2,25,000
SPB-33	**Regulatory Bioinformatics Submission Analysis** Bioinformatics regulatory submission preparation · Regulatory submission analysis	18	88	15	2,05,000
Complete Sector Catalogue 02 — all thirty-nine analysis types		**668**	**3240**	**551**	**74,45,000**

Each fee is derived independently from the step count, analytical complexity and reporting burden of that specific analysis type. There is no banded pricing. Days are calculated at six contact hours per working day and rounded up. Fees are per participant and exclusive of GST and the one-time registration charge. In-house delivery at the client laboratory and institutional rates are available on application.

SPB-01
SECTOR 01

Target Discovery Transcriptomic Analysis

(**Transcriptomic target identification** · Target discovery RNA-Seq)

WHY IT IS DONE

Before a programme exists there is a disease and a hypothesis. Transcriptomic comparison between diseased and healthy tissue is the most direct route from that hypothesis to a ranked list of genes worth pursuing as drug targets.

WHAT IT ANSWERS

Which genes and pathways are differentially regulated in this disease state, and which of them are plausible and tractable as therapeutic targets?

WHEN IT IS DONE

At the earliest stage of a discovery programme, in indication expansion for an existing asset, and in target triage where several hypotheses compete for resource.

WHAT TRIGGERS IT

A new discovery programme, an indication expansion decision, an asset due diligence exercise, or a target triage review.

HOW IT IS DONE

Study design and confounding are assessed before analysis, expression is quantified and normalised, and differential expression is modelled with covariates carried explicitly. Pathway and network analysis follows, cell composition is deconvolved to separate compositional from regulatory change, and candidates are ranked against druggability, expression specificity and existing evidence.

WHAT IT PRODUCES

A differential expression table with effect sizes and adjusted significance, pathway and network results, deconvolution-adjusted findings, a ranked target list with tractability annotation, and a reproducibility record.

WHAT IT DETECTS

Differentially expressed genes with effect size and significance, pathway and network level dysregulation, cell-type-attributable signal where deconvolution is applied, and candidate targets ranked against tractability criteria.

WHAT IT DOES NOT DETECT

Expression change does not establish causation, so a differentially expressed gene may be a consequence rather than a driver. Post-transcriptional and protein-level regulation is invisible, tissue heterogeneity confounds bulk comparison, and no transcriptomic analysis alone establishes that modulating a target changes disease.

WHAT MAKES IT HARD

Distinguishing a change in cell composition from a change in cell state is the recurring error, and bulk data cannot separate them without deconvolution. Batch and confounding structure in disease cohorts is frequently severe, and unadjusted analysis produces target lists that describe the sampling rather than the disease.

WHO DOES THIS JOB

Computational biologist in discovery research, mid to senior level, embedded in a target discovery or translational group.

WHAT YOU CAN DO AFTERWARDS

Design and analyse discovery transcriptomic studies, control confounding and composition, and produce ranked target lists with defensible evidence.

WHO HIRES FOR IT

Pharmaceutical discovery groups, biotech target discovery teams, academic drug discovery units, and contract research organisations supporting discovery.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**96 hours · 16 days**	Prior RNA-Seq analysis experience	**2,20,000**

SPB-01
WORKFLOW

Target Discovery Transcriptomic Analysis

(**Transcriptomic target identification** · Target discovery RNA-Seq)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Programme question definition**	The discovery question and success criteria defined before any analysis begins	Programme protocol	*Defined question*
02	**Study design and power assessment**	Sample size, group structure and confounding assessed against the question	Design review	*Design assessment*
03	**Metadata and covariate assembly**	Clinical, demographic and technical covariates assembled for every sample	Metadata compilation	*Covariate matrix*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified against acceptance criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
05	**Read preprocessing**	Adapters and low-quality tails removed with strand information preserved	fastp	*Trimmed FASTQ*
06	**Alignment or pseudo-alignment**	Reads aligned or quantified against a version-locked reference annotation	STAR, Salmon	*Quantification output*
07	**Library metric assessment**	Duplication, gene body coverage and strand specificity assessed per sample	RSeQC, Qualimap	*Library metrics*
08	**Expression matrix assembly**	Counts assembled with annotation version and gene identifier mapping recorded	tximport, featureCounts	*Expression matrix*
09	**Outlier sample detection**	Samples deviating from the cohort identified and exclusion decided	Clustering, correlation analysis	*Outlier decisions*
10	**Normalisation**	Counts normalised with library composition and depth differences handled	DESeq2, edgeR normalisation	*Normalised matrix*
11	**Batch structure assessment**	Technical batch structure tested for association with biological grouping	Variance partitioning	*Batch assessment*
12	**Cell composition deconvolution**	Composition estimated so compositional change is separated from state change	Deconvolution	*Composition estimates*
13	**Differential expression modelling**	Expression modelled with covariates and composition carried in the design	DESeq2, limma	*Differential results*
14	**Effect size shrinkage**	Effect sizes shrunk so low-expression genes do not dominate rankings	Shrinkage estimation	*Shrunk effect sizes*
15	**Multiplicity control**	Significance adjusted across all tests with the method recorded	FDR control	*Adjusted results*
16	**Pathway and gene set analysis**	Enrichment performed with gene universe and database versions recorded	clusterProfiler, GSEA	*Enrichment results*
17	**Network and regulatory analysis**	Co-expression and regulatory structure analysed to identify hub candidates	WGCNA, network inference	*Network results*
18	**Tractability annotation**	Candidates annotated for druggability, expression specificity and existing evidence	Tractability resources	*Tractability annotation*
19	**Candidate ranking**	Targets ranked combining statistical, biological and tractability evidence	Ranking framework	*Ranked target list*
20	**Reproducibility documentation**	Versions, parameters and decisions documented for independent re-execution	Documentation procedure	*Reproducibility record*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**96 h · 16 d**	**SPB-01**	**2,20,000**

SPB-02
SECTOR 01

Differential Expression & Pathway Analysis

(**Differential expression and enrichment analysis** · DE analysis · pathway analysis)

WHY IT IS DONE

It is the most frequently performed analysis in all of biology and the most frequently performed badly. Design flaws, unmodelled batch structure and multiple testing errors produce results that do not replicate, and replication failure is the dominant cost in preclinical research.

WHAT IT ANSWERS

Which genes change between these conditions, by how much, with what statistical confidence, and what biological processes do those changes implicate?

WHEN IT IS DONE

In every comparative transcriptomic experiment across discovery, mechanism, toxicology and translational work.

WHAT TRIGGERS IT

Completion of any comparative RNA sequencing experiment, or reanalysis of an existing dataset under a new question.

HOW IT IS DONE

Design and power are assessed before any modelling, counts are normalised with library composition handled, and the model is specified with covariates and batch structure included explicitly. Testing accounts for multiplicity, effect size shrinkage is applied, and enrichment is performed with the gene universe and database versions recorded.

WHAT IT PRODUCES

The specified model with covariates, normalised expression matrices, differential expression results with shrunk effect sizes and adjusted significance, enrichment results with database versions, and a full reproducibility record.

WHAT IT DETECTS

Differentially expressed genes with fold change and adjusted significance, enriched pathways and gene sets, direction of pathway regulation, and the effect of covariates on the result.

WHAT IT DOES NOT DETECT

It does not establish mechanism or direction of causation, and enrichment results reflect the annotation databases used rather than biology itself. Underpowered designs produce results that are statistically significant and biologically meaningless, and no analysis recovers power a design did not have.

WHAT MAKES IT HARD

The design determines the result and cannot be repaired analytically. Confounded batch structure produces confident false findings, the gene universe choice materially changes enrichment output, and shrinkage is routinely omitted so that low-expression genes dominate the top of the list.

WHO DOES THIS JOB

Computational biologist or bioinformatician across discovery, translational and toxicology functions, entry to mid level. The most universally required analytical skill in the sector.

WHAT YOU CAN DO AFTERWARDS

Specify and defend statistical models, control batch and confounding, apply appropriate multiplicity correction and shrinkage, and perform enrichment reproducibly.

WHO HIRES FOR IT

Every pharmaceutical and biotech research function, contract research organisations, academic research groups, and translational research units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**86 hours · 15 days**	Prior RNA-Seq analysis experience, basic statistics	**2,00,000**

SPB-02
WORKFLOW

Differential Expression & Pathway Analysis

(**Differential expression and enrichment analysis** · DE analysis · pathway analysis)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Experimental design review**	Group structure, replication and confounding reviewed before analysis	Design review	*Design assessment*
02	**Power assessment**	Statistical power assessed against the effect size the experiment can detect	Power calculation	*Power assessment*
03	**Metadata assembly**	Sample covariates and batch information assembled and verified	Metadata compilation	*Covariate matrix*
04	**Data quality assessment**	Sequence quality and library metrics assessed per sample	FastQC, MultiQC	*QC record*
05	**Preprocessing and quantification**	Reads preprocessed and expression quantified against a locked annotation	fastp, STAR or Salmon	*Quantification output*
06	**Expression matrix assembly**	Counts assembled with annotation version recorded	tximport, featureCounts	*Expression matrix*
07	**Low-expression filtering**	Genes below the detection threshold removed with the criterion recorded	Filtering	*Filtered matrix*
08	**Normalisation**	Library size and composition differences normalised	DESeq2, edgeR	*Normalised matrix*
09	**Exploratory structure analysis**	Sample relationships examined for expected and unexpected structure	PCA, clustering	*Structure assessment*
10	**Outlier identification**	Outlying samples identified with exclusion decisions documented	Outlier analysis	*Outlier decisions*
11	**Batch and covariate assessment**	Technical variables tested for association with the biological contrast	Association testing	*Confounding assessment*
12	**Model specification**	The statistical model specified explicitly with all covariates justified	Model specification	*Specified model*
13	**Dispersion estimation**	Gene-wise dispersion estimated with shrinkage across genes	Dispersion estimation	*Dispersion estimates*
14	**Differential testing**	Contrasts tested under the specified model	DESeq2, limma-voom	*Test results*
15	**Effect size shrinkage**	Fold change estimates shrunk for ranking stability	Shrinkage estimation	*Shrunk effect sizes*
16	**Multiplicity correction**	Significance adjusted across the tested gene set	FDR control	*Adjusted results*
17	**Enrichment analysis**	Pathway enrichment performed with universe and database versions recorded	clusterProfiler, GSEA	*Enrichment results*
18	**Reproducibility documentation**	Full parameter, version and decision record assembled	Documentation procedure	*Reproducibility record*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**86 h · 15 d**	**SPB-02**	**2,00,000**

SPB-03
SECTOR 01

Single-Cell Target Discovery Analysis

(**Single-cell transcriptomic target discovery** · scRNA target discovery)

WHY IT IS DONE

A target expressed on a disease-driving cell population and nowhere else is a far better asset than one expressed broadly. Single-cell analysis resolves which population expresses what, which is the difference between a tractable target and an unacceptable safety profile.

WHAT IT ANSWERS

Which cell populations drive this disease, what do they express uniquely, and which of those genes are viable targets with acceptable expression elsewhere?

WHEN IT IS DONE

In target discovery where cellular specificity matters, in immunology and oncology programmes particularly, and in safety assessment of a proposed target's expression across normal tissue.

WHAT TRIGGERS IT

A discovery programme requiring cell-type resolution, a safety question about target expression breadth, or an unexplained response heterogeneity in a bulk study.

HOW IT IS DONE

Dissociation protocol is recorded because it determines which cells are measured, then data are processed with ambient and doublet correction, integrated across samples, and clustered. Populations are annotated against references, disease-associated populations are identified by compositional and state analysis, and candidate targets are assessed for expression specificity across the full atlas.

WHAT IT PRODUCES

An annotated cell atlas with composition, disease-associated population identification, population-specific expression profiles, target candidate specificity assessment across all populations, and a dissociation bias caveat.

WHAT IT DETECTS

Cell populations with type and state annotation, population-specific and state-specific expression, disease-associated populations absent or expanded relative to control, and cell surface expression profiles for accessible target classes.

WHAT IT DOES NOT DETECT

Cells lost during dissociation are absent entirely, and dissociation bias systematically under-represents fragile populations. Protein-level expression is not measured, spatial context is destroyed, and low-abundance populations fall below detection at typical cell numbers.

WHAT MAKES IT HARD

Compositional change and state change are easily confused, and both look like differential expression in a naive comparison. Expression specificity claims require an adequate normal tissue reference, and asserting specificity without one is the mistake that surfaces later as toxicity.

WHO DOES THIS JOB

Computational biologist with single-cell competence in discovery, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Process and integrate single-cell discovery data, separate compositional from state change, and assess target expression specificity defensibly.

WHO HIRES FOR IT

Pharmaceutical discovery groups, immunology and oncology biotech, single-cell technology providers, and translational research units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
23	**112 hours · 19 days**	Prior single-cell RNA-Seq analysis experience	**2,55,000**

SPB-03
WORKFLOW

Single-Cell Target Discovery Analysis

(**Single-cell transcriptomic target discovery** · scRNA target discovery)

COMPLETE WORKFLOW — 23 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Programme question definition**	The target discovery question and specificity requirements defined	Programme protocol	*Defined question*
02	**Dissociation protocol recording**	Protocol recorded because it determines which populations survive to measurement	Protocol record	*Dissociation record*
03	**Viability and yield assessment**	Cell viability and recovery assessed before library preparation	Viability assessment	*Viability record*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
05	**Alignment and counting**	Reads aligned and per-cell count matrices generated	CellRanger, STARsolo	*Count matrices*
06	**Empty droplet removal**	True cells distinguished from ambient-only barcodes	EmptyDrops	*Cell-called matrix*
07	**Ambient correction**	Ambient transcript contamination estimated and corrected	SoupX, CellBender	*Corrected matrix*
08	**Doublet detection**	Multiplet barcodes identified and removed	scDblFinder	*Filtered matrix*
09	**Quality filtering**	Cells filtered on transcript count, gene count and mitochondrial fraction	QC filtering	*Quality-filtered matrix*
10	**Normalisation and feature selection**	Counts normalised and variable features selected	Normalisation, HVG selection	*Normalised matrix*
11	**Dimensionality reduction**	Principal components computed and neighbour graph constructed	PCA, neighbour graph	*Reduced representation*
12	**Sample integration**	Samples integrated with batch effect controlled and integration benchmarked	Harmony, scVI, kBET	*Integrated embedding*
13	**Clustering**	Cells clustered at a resolution justified against marker separation	Leiden clustering	*Cell clusters*
14	**Reference-based annotation**	Clusters annotated against reference atlases with confidence recorded	Reference mapping	*Annotated populations*
15	**Marker-based annotation refinement**	Annotations refined against canonical marker evidence	Marker analysis	*Refined annotations*
16	**Compositional analysis**	Population proportions compared between conditions with compositional methods	scCODA, propeller	*Composition results*
17	**State versus composition separation**	Expression change separated into compositional and state components	Pseudobulk and state analysis	*Separated effects*
18	**Population-specific expression profiling**	Expression profiled per population to identify specific markers	Marker detection	*Population profiles*
19	**Disease-associated population identification**	Populations expanded or altered in disease identified and characterised	Comparative analysis	*Disease-associated populations*
20	**Surface expression assessment**	Candidate targets assessed for surface accessibility where relevant	Surfaceome annotation	*Surface candidates*
21	**Normal tissue specificity assessment**	Candidate expression assessed across normal tissue reference atlases	Reference atlas comparison	*Specificity assessment*
22	**Candidate ranking**	Targets ranked on specificity, expression level and tractability	Ranking framework	*Ranked candidates*
23	**Dissociation bias caveat documentation**	Populations likely lost to dissociation documented as a stated limitation	Limitation template	*Bias caveat*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
23	**112 h · 19 d**	**SPB-03**	**2,55,000**

SPB-04
SECTOR 01

Pooled CRISPR Screen Analysis

(**Pooled genetic screen analysis** · CRISPR screen analysis)

WHY IT IS DONE

Screens interrogate thousands of genes for a functional phenotype in a single experiment, which is the fastest route from a phenotype to the genes that cause it. They are the workhorse of functional genomics in drug discovery.

WHAT IT ANSWERS

Which genes, when perturbed, produce the screened phenotype, at what effect size, and with what confidence given the screen's design and depth?

WHEN IT IS DONE

In target identification, in synthetic lethality and drug sensitisation screens, in resistance mechanism discovery, and in essentiality profiling across model panels.

WHAT TRIGGERS IT

A functional genomics programme, a resistance mechanism question, a synthetic lethality hypothesis, or drug mechanism investigation.

HOW IT IS DONE

Library composition and representation are assessed at every timepoint since representation loss invalidates the screen, guides are counted and normalised, and gene-level effects are modelled from guide-level data with variable guide efficiency accounted for. Control gene behaviour is used to calibrate noise, and hits are ranked with reagent-level consistency required.

WHAT IT PRODUCES

Library representation metrics across timepoints, guide-level counts and normalisation, gene-level effect sizes with confidence, control calibration evidence, ranked hit lists, and guide-consistency assessment per hit.

WHAT IT DETECTS

Guide-level and gene-level enrichment or depletion, effect sizes with statistical confidence, essential gene signal as a positive control, and hits ranked against screen-specific noise.

WHAT IT DOES NOT DETECT

It reports the consequence of perturbation, not the mechanism behind it. Genes not represented in the library are invisible, guides with poor efficiency produce false negatives indistinguishable from true ones, and off-target guide activity produces false positives that only reagent-level analysis reveals.

WHAT MAKES IT HARD

Representation is everything and is lost silently through bottlenecks in culture, so it must be measured rather than assumed. A gene-level hit driven by a single guide is usually an off-target artefact, and requiring multi-guide consistency is the main defence against a hit list full of them.

WHO DOES THIS JOB

Computational biologist in functional genomics, mid to senior level, working closely with screening laboratories.

WHAT YOU CAN DO AFTERWARDS

Assess library representation, model gene-level effects from guide data, calibrate against controls, and produce hit lists with reagent-level support.

WHO HIRES FOR IT

Pharmaceutical functional genomics groups, biotech target discovery teams, screening platform companies, and academic functional genomics units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
22	**106 hours · 18 days**	Prior counting-based sequencing analysis experience	**2,45,000**

SPB-04
WORKFLOW

Pooled CRISPR Screen Analysis

(**Pooled genetic screen analysis** · CRISPR screen analysis)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Screen design review**	Library, coverage, timepoints and replication reviewed against the question	Design review	*Design assessment*
02	**Library composition verification**	Library guide composition verified against the intended design	Library sequencing	*Composition record*
03	**Representation assessment at plasmid stage**	Guide representation measured in the plasmid pool as the baseline	Counting analysis	*Baseline representation*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified per timepoint and replicate	md5sum, FastQC	*QC acceptance record*
05	**Guide counting**	Guide sequences extracted and counted per sample	Counting pipeline	*Guide count matrix*
06	**Representation assessment per timepoint**	Representation measured at every timepoint to detect bottlenecks	Representation analysis	*Representation metrics*
07	**Bottleneck detection**	Coverage loss through culture detected and its impact assessed	Bottleneck analysis	*Bottleneck assessment*
08	**Count normalisation**	Counts normalised for sequencing depth and library composition	Normalisation	*Normalised counts*
09	**Replicate correlation assessment**	Replicate agreement assessed as a screen quality indicator	Correlation analysis	*Replicate metrics*
10	**Control guide behaviour assessment**	Non-targeting and essential gene controls assessed to calibrate the screen	Control analysis	*Control calibration*
11	**Screen quality scoring**	Overall screen quality scored against control separation metrics	Quality scoring	*Screen quality score*
12	**Guide-level effect estimation**	Log fold changes estimated per guide between conditions	Effect estimation	*Guide effects*
13	**Guide efficiency weighting**	Guides weighted by predicted or observed efficiency where data support it	Efficiency modelling	*Weighted effects*
14	**Gene-level effect modelling**	Gene effects modelled from guide-level data with variance accounted for	MAGeCK, drugZ	*Gene-level effects*
15	**Statistical significance estimation**	Significance estimated against the screen's own null distribution	Statistical modelling	*Significance estimates*
16	**Multiplicity control**	Significance adjusted across all genes tested	FDR control	*Adjusted results*
17	**Guide consistency assessment**	Hits assessed for support across multiple independent guides	Consistency analysis	*Consistency assessment*
18	**Single-guide artefact filtering**	Hits driven by a single guide flagged as probable off-target artefacts	Artefact filtering	*Filtered hit list*
19	**Copy number bias assessment**	Amplified region bias assessed since it produces false essentiality signal	Copy number correction	*Bias-corrected results*
20	**Hit ranking**	Hits ranked combining effect size, significance and guide consistency	Ranking framework	*Ranked hit list*
21	**Comparison against reference screens**	Hits compared against public essentiality and screen datasets for context	Reference comparison	*Contextual comparison*
22	**Validation prioritisation**	Hits prioritised for orthogonal validation with rationale recorded	Prioritisation criteria	*Validation list*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**106 h · 18 d**	**SPB-04**	**2,45,000**

SPB-05
SECTOR 01

Perturb-Seq & Arrayed Screen Analysis

(**Single-cell perturbation screen analysis** · Perturb-Seq analysis)

WHY IT IS DONE

A pooled screen tells you a perturbation had an effect; a single-cell perturbation screen tells you what that effect was at the level of the whole transcriptome. That converts a hit list into mechanistic hypotheses without a separate follow-up experiment for each gene.

WHAT IT ANSWERS

What transcriptional consequence does each perturbation produce, which perturbations converge on the same programme, and what does that reveal about pathway structure?

WHEN IT IS DONE

In mechanism-of-action work following a pooled screen, in pathway mapping, in regulatory network inference, and in target deconvolution for a compound with unknown mechanism.

WHAT TRIGGERS IT

A pooled screen requiring mechanistic follow-up, a pathway mapping objective, or a compound with an unresolved mechanism.

HOW IT IS DONE

Guide assignment per cell is established with confidence and multiplets excluded, then perturbation efficiency is estimated so unperturbed cells can be down-weighted. Signatures are computed against matched controls, weak effects are pooled across cells, and perturbations are clustered by signature similarity to infer pathway structure.

WHAT IT PRODUCES

Per-cell guide assignments with confidence, perturbation efficiency estimates, per-perturbation transcriptional signatures with statistical support, signature clustering revealing pathway convergence, and inferred regulatory relationships.

WHAT IT DETECTS

Perturbation-specific transcriptional signatures, convergence between perturbations indicating shared pathway membership, dose and efficiency effects at single-cell resolution, and inferred regulatory relationships.

WHAT IT DOES NOT DETECT

Incomplete perturbation efficiency means many cells assigned a guide are not actually perturbed, and this dilutes every signature unless modelled. Protein-level and post-transcriptional consequences are invisible, and cells with lethal perturbations are absent from the data by the time it is collected.

WHAT MAKES IT HARD

Guide assignment ambiguity and incomplete perturbation efficiency both dilute signal, and treating every guide-assigned cell as perturbed is the standard error that flattens real effects. Statistical power is limited by cells per perturbation, so design decisions bind the analysis far more than method choice.

WHO DOES THIS JOB

Senior computational biologist with single-cell and functional genomics competence. A specialist and scarce combination.

WHAT YOU CAN DO AFTERWARDS

Assign guides with confidence, model perturbation efficiency, extract signatures with adequate power, and infer pathway structure from signature convergence.

WHO HIRES FOR IT

Pharmaceutical functional genomics groups, single-cell focused biotech, screening platform companies, and academic systems biology units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
24	**116 hours · 20 days**	Prior single-cell and screen analysis experience	**2,65,000**

SPB-05
WORKFLOW

Perturb-Seq & Arrayed Screen Analysis

(**Single-cell perturbation screen analysis** · Perturb-Seq analysis)

COMPLETE WORKFLOW — 24 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Screen design review**	Perturbation set, cell numbers per perturbation and controls reviewed	Design review	*Design assessment*
02	**Power assessment per perturbation**	Cells per perturbation assessed against detectable effect size	Power calculation	*Power assessment*
03	**Data receipt and quality assessment**	Sequence integrity and quality verified for transcriptome and guide libraries	md5sum, FastQC	*QC acceptance record*
04	**Transcriptome counting**	Per-cell transcriptome counts generated	CellRanger, STARsolo	*Transcriptome matrix*
05	**Guide capture counting**	Guide identity reads counted per cell barcode	Guide counting	*Guide count matrix*
06	**Empty droplet and ambient correction**	Empty droplets removed and ambient contamination corrected	EmptyDrops, CellBender	*Corrected matrix*
07	**Doublet detection**	Transcriptomic and guide-based multiplet detection applied	scDblFinder, guide multiplicity	*Filtered matrix*
08	**Guide assignment**	Guides assigned to cells with a confidence threshold rather than by maximum count	Assignment model	*Guide assignments*
09	**Multiple guide cell handling**	Cells receiving multiple guides identified and handled under a documented rule	Assignment policy	*Multi-guide decisions*
10	**Assignment confidence recording**	Assignment confidence recorded per cell for downstream weighting	Confidence estimation	*Assignment confidence*
11	**Quality filtering**	Cells filtered on transcriptome and guide capture quality	QC filtering	*Quality-filtered cells*
12	**Normalisation**	Counts normalised with technical covariates identified	Normalisation	*Normalised matrix*
13	**Perturbation efficiency estimation**	Efficiency estimated so unperturbed guide-assigned cells can be down-weighted	Efficiency estimation	*Efficiency estimates*
14	**Control cell definition**	Non-targeting control cells defined as the comparison baseline	Control definition	*Control set*
15	**Per-perturbation signature computation**	Transcriptional signature computed per perturbation against controls	Signature computation	*Perturbation signatures*
16	**Weak effect pooling**	Effects pooled across cells to recover signal below single-cell detection	Pooling methods	*Pooled effects*
17	**Statistical significance estimation**	Significance estimated with cell number and efficiency accounted for	Statistical modelling	*Significance estimates*
18	**Multiplicity control**	Significance adjusted across perturbations and genes	FDR control	*Adjusted results*
19	**Signature clustering**	Perturbations clustered by signature similarity to reveal pathway convergence	Clustering	*Perturbation clusters*
20	**Pathway structure inference**	Pathway relationships inferred from convergence and signature hierarchy	Inference methods	*Pathway structure*
21	**Regulatory relationship inference**	Regulatory relationships inferred between perturbed genes and responders	Network inference	*Regulatory relationships*
22	**Lethal perturbation absence assessment**	Perturbations depleted from the population identified as probable lethals	Depletion analysis	*Lethality assessment*
23	**Cross-perturbation validation**	Findings checked for consistency across independent guides targeting the same gene	Consistency analysis	*Validation assessment*
24	**Reproducibility documentation**	Assignments, thresholds, versions and decisions documented	Documentation procedure	*Reproducibility record*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
24	**116 h · 20 d**	**SPB-05**	**2,65,000**

SPB-06
SECTOR 01

RNAi & shRNA Screen Analysis

(**RNA interference screen analysis** · RNAi screen analysis)

WHY IT IS DONE

Large historical screening datasets were generated by RNA interference, and reanalysing them against modern methods extracts value from investments already made. Knockdown also models partial inhibition better than knockout, which is closer to what a drug actually does.

WHAT IT ANSWERS

Which genes produce the screened phenotype on knockdown, and how much of the observed signal is seed-based off-target effect rather than genuine target silencing?

WHEN IT IS DONE

In reanalysis of historical screening data, where partial knockdown better models pharmacological inhibition, and in target validation alongside knockout evidence.

WHAT TRIGGERS IT

Reanalysis of legacy screening data, a target validation requirement using an orthogonal modality, or a partial-inhibition hypothesis.

HOW IT IS DONE

Reagent representation is assessed, counts are normalised, and seed sequence effects are modelled explicitly since they dominate uncorrected RNA interference data. Gene-level effects are estimated after seed correction, and results are compared against knockout data for the same genes where available.

WHAT IT PRODUCES

Representation metrics, seed effect model and correction evidence, gene-level effects after correction, reagent-level consistency assessment, and knockout concordance comparison.

WHAT IT DETECTS

Gene-level knockdown phenotypes with effect sizes, seed sequence driven off-target signal, and concordance or discordance with knockout evidence for the same genes.

WHAT IT DOES NOT DETECT

Knockdown efficiency is not measured by the screen itself, so a negative result may reflect poor silencing rather than gene dispensability. Off-target seed effects are pervasive and cannot be fully removed, only estimated and corrected.

WHAT MAKES IT HARD

Seed-based off-target effect is the defining problem and is large enough to generate entire false hit lists. Correcting for it requires explicit modelling rather than filtering, and legacy analyses that predate that correction should be regarded as unreliable rather than merely dated.

WHO DOES THIS JOB

Computational biologist in functional genomics, mid level, frequently working on legacy data reanalysis.

WHAT YOU CAN DO AFTERWARDS

Model and correct seed-based off-target effects, estimate gene-level knockdown phenotypes, and reconcile knockdown against knockout evidence.

WHO HIRES FOR IT

Pharmaceutical functional genomics groups, biotech target validation teams, and academic groups with legacy screening data.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
17	**82 hours · 14 days**	Prior screen analysis experience	**1,90,000**

SPB-06
WORKFLOW

RNAi & shRNA Screen Analysis

(**RNA interference screen analysis** · RNAi screen analysis)

COMPLETE WORKFLOW — 17 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Screen and library review**	Library design, reagent chemistry and screen structure reviewed	Design review	*Design assessment*
02	**Representation assessment**	Reagent representation assessed at baseline and each timepoint	Representation analysis	*Representation metrics*
03	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC	*QC acceptance record*
04	**Reagent counting**	Reagent sequences extracted and counted per sample	Counting pipeline	*Count matrix*
05	**Normalisation**	Counts normalised for depth and composition differences	Normalisation	*Normalised counts*
06	**Replicate correlation assessment**	Replicate agreement assessed as a quality indicator	Correlation analysis	*Replicate metrics*
07	**Control behaviour assessment**	Control reagent behaviour assessed to calibrate screen performance	Control analysis	*Control calibration*
08	**Seed sequence extraction**	Seed sequences extracted from every reagent for off-target modelling	Sequence analysis	*Seed assignments*
09	**Seed effect modelling**	Systematic seed-driven effects modelled across the whole library	Seed effect model	*Seed model*
10	**Seed effect correction**	Reagent effects corrected for the modelled seed contribution	Correction procedure	*Corrected effects*
11	**Reagent-level effect estimation**	Effects estimated per reagent after seed correction	Effect estimation	*Reagent effects*
12	**Gene-level effect modelling**	Gene effects modelled from corrected reagent-level data	Gene-level modelling	*Gene effects*
13	**Statistical significance estimation**	Significance estimated with multiplicity controlled	Statistical modelling, FDR	*Adjusted results*
14	**Reagent consistency assessment**	Hits assessed for support across independent reagents	Consistency analysis	*Consistency assessment*
15	**Knockdown efficiency caveat documentation**	Absence of efficiency measurement documented as a limitation on negatives	Limitation template	*Efficiency caveat*
16	**Knockout concordance comparison**	Results compared against knockout data for the same genes where available	Cross-modality comparison	*Concordance analysis*
17	**Hit ranking and prioritisation**	Hits ranked and prioritised with seed correction evidence attached	Ranking framework	*Ranked hit list*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
17	**82 h · 14 d**	**SPB-06**	**1,90,000**

SPB-07
SECTOR 01

Cell Line Genomic Authentication Analysis

(**Cell line identity and authentication analysis** · Cell line authentication)

WHY IT IS DONE

A substantial proportion of cell lines in circulation are misidentified or cross-contaminated, and experiments run on the wrong line produce results that cannot replicate. Authentication is the cheapest available protection against an entire programme built on a false premise.

WHAT IT ANSWERS

Is this cell line what it is labelled as, is it contaminated with another line, and does it match the reference profile for that identity?

WHEN IT IS DONE

On receipt of any line, at defined passage intervals, before any pivotal experiment, and at publication or regulatory submission.

WHAT TRIGGERS IT

Receipt of a new line, a scheduled authentication interval, an unexpected experimental result, or a submission requirement.

HOW IT IS DONE

An identity profile is generated by the chosen marker method, compared against reference databases with match thresholds applied, and assessed for the mixed signal that indicates cross-contamination. Species is confirmed, contamination proportion is estimated where present, and the profile is archived for future comparison.

WHAT IT PRODUCES

The identity profile, reference database match with percentage and threshold, cross-contamination assessment with proportion, species confirmation, contamination screening results, and an archived profile record.

WHAT IT DETECTS

Short tandem repeat or single nucleotide polymorphism identity profiles, match against reference databases, cross-contamination from another line with the proportion estimated, species identity, and mycoplasma or other contamination where screened.

WHAT IT DOES NOT DETECT

It confirms identity, not fitness for purpose — an authenticated line may still have drifted genomically from the reference through passage. It also cannot detect contamination by a line with an identical or absent reference profile.

WHAT MAKES IT HARD

Partial contamination is the difficult case, because a line that is ninety percent correct still matches its reference while producing unreliable biology. Reference database coverage is incomplete, so an unmatched profile means unknown rather than wrong.

WHO DOES THIS JOB

Bioinformatician or research scientist with analytical responsibility, entry to mid level, frequently within a quality function.

WHAT YOU CAN DO AFTERWARDS

Generate and interpret authentication profiles, detect partial cross-contamination, and maintain an authentication record system.

WHO HIRES FOR IT

Pharmaceutical and biotech research operations, cell banking services, contract research organisations, and academic core facilities.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
14	**66 hours · 11 days**	Prior genotype analysis experience	**1,50,000**

SPB-07
WORKFLOW

Cell Line Genomic Authentication Analysis

(**Cell line identity and authentication analysis** · Cell line authentication)

COMPLETE WORKFLOW — 14 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Line identity and provenance recording**	Claimed identity, source and passage history recorded on receipt	Inventory record	*Provenance record*
02	**Marker method selection**	The identity marker method selected and its reference database identified	Method selection	*Selected method*
03	**Sample preparation verification**	Sample confirmed to derive from the culture being authenticated	Chain of custody	*Custody record*
04	**Identity profile generation**	The identity profile generated across the marker panel	Genotyping	*Identity profile*
05	**Profile quality assessment**	Profile completeness and marker call quality assessed	Quality assessment	*Profile quality record*
06	**Reference database matching**	The profile matched against reference databases with version recorded	Database matching	*Match results*
07	**Match threshold application**	Match percentage assessed against the applicable identity threshold	Threshold criteria	*Match determination*
08	**Mixed signal assessment**	Additional alleles indicating cross-contamination identified	Mixture analysis	*Mixture assessment*
09	**Contamination proportion estimation**	Where mixed, the proportion of the contaminating line estimated	Proportion estimation	*Contamination estimate*
10	**Species confirmation**	Species identity confirmed independently of the line identity match	Species analysis	*Species confirmation*
11	**Microbial contamination screening**	Mycoplasma and other contamination screened where in scope	Contamination screening	*Contamination results*
12	**Unmatched profile handling**	Profiles without a database match reported as unmatched rather than incorrect	Reporting policy	*Unmatched determination*
13	**Profile archiving**	The profile archived in queryable form for future comparison	Archive procedure	*Archived profile*
14	**Authentication certificate issue**	Authentication outcome documented with method, database and thresholds	Certificate template	*Authentication certificate*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
14	**66 h · 11 d**	**SPB-07**	**1,50,000**

SPB-08
SECTOR 01

Cell Line Panel & Model Characterisation Analysis

(**Preclinical model genomic characterisation** · Model characterisation)

WHY IT IS DONE

Choosing preclinical models on the basis of their genomic features rather than convenience determines whether an efficacy result translates. Characterising a panel also enables genotype-response association, which is how a biomarker hypothesis is generated before any patient is dosed.

WHAT IT ANSWERS

What genomic and transcriptomic features characterise each model in this panel, and which models carry the alterations relevant to the programme's hypothesis?

WHEN IT IS DONE

When assembling a model panel for efficacy studies, in biomarker hypothesis generation from panel screening data, and in selecting models for mechanism studies.

WHAT TRIGGERS IT

Assembly of a preclinical model panel, a biomarker hypothesis requiring genotype-response association, or model selection for an efficacy study.

HOW IT IS DONE

Each model is characterised across mutation, copy number, fusion and expression, with authentication confirmed first. Alterations are annotated against pathway and functional evidence, models are profiled against the diversity of the clinical population, and the panel is assessed for coverage of the features the programme depends on.

WHAT IT PRODUCES

Per-model genomic and transcriptomic characterisation, pathway alteration status per model, panel diversity assessment against clinical population, model selection recommendations, and a reference profile archive.

WHAT IT DETECTS

Mutations, copy number alterations, fusions and expression profiles per model, pathway alteration status, and the panel's coverage of the genomic diversity relevant to the indication.

WHAT IT DOES NOT DETECT

Genomic characterisation does not predict whether a model recapitulates disease biology, and long-passaged lines diverge substantially from the tumours they derived from. Microenvironment and immune context are absent entirely from cell line models.

WHAT MAKES IT HARD

Panel composition bias is the persistent problem, since widely used lines over-represent certain genotypes and under-represent others, so a panel result may not generalise. Comparing models to clinical population diversity is the step that reveals this and is frequently skipped.

WHO DOES THIS JOB

Computational biologist in preclinical or translational research, mid level.

WHAT YOU CAN DO AFTERWARDS

Characterise model panels across alteration classes, assess panel diversity against clinical populations, and support evidence-based model selection.

WHO HIRES FOR IT

Pharmaceutical preclinical research, oncology biotech, contract research organisations, and model system providers.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
19	**92 hours · 16 days**	Prior somatic and copy number analysis experience	**2,10,000**

SPB-08
WORKFLOW

Cell Line Panel & Model Characterisation Analysis

(**Preclinical model genomic characterisation** · Model characterisation)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Panel scope definition**	The models in the panel and the characterisation scope defined	Programme protocol	*Panel definition*
02	**Authentication verification**	Every model authenticated before characterisation proceeds	Authentication analysis	*Authentication clearance*
03	**Data receipt and quality assessment**	Sequence integrity and quality verified for every model and assay	md5sum, FastQC	*QC acceptance record*
04	**Alignment and processing**	Reads aligned and processed identically across all models	bwa-mem2, STAR, Picard	*Analysis-ready data*
05	**Coverage assessment**	Coverage assessed per model against characterisation requirements	mosdepth	*Coverage record*
06	**Variant calling**	Sequence variants called with a consistent pipeline across the panel	Validated caller	*Variant call sets*
07	**Germline and artefact filtering**	Germline variation and recurrent artefacts removed from somatic call sets	Population and PoN filtering	*Filtered variants*
08	**Copy number analysis**	Copy number called with purity and ploidy addressed where relevant	Copy number analysis	*Copy number profiles*
09	**Fusion detection**	Fusions detected from RNA where transcriptomic data are available	Fusion callers	*Fusion calls*
10	**Expression quantification**	Expression quantified and normalised across the panel	Salmon, DESeq2 normalisation	*Expression matrix*
11	**Pathway alteration status assignment**	Pathway-level alteration status assigned per model	Pathway annotation	*Pathway status*
12	**Driver annotation**	Alterations annotated against driver and functional evidence resources	OncoKB, curated evidence	*Driver annotation*
13	**Clinical population comparison**	Panel genomic diversity compared against the clinical population distribution	Reference cohort comparison	*Diversity comparison*
14	**Coverage gap identification**	Genotypes under-represented in the panel relative to patients identified	Gap analysis	*Coverage gap record*
15	**Model selection criteria application**	Models selected against the programme's specific feature requirements	Selection criteria	*Model selection*
16	**Passage and drift documentation**	Passage number and known drift documented per model	Provenance record	*Drift documentation*
17	**Biological fidelity caveat documentation**	Limits of genomic characterisation as a fidelity measure documented	Limitation template	*Fidelity caveat*
18	**Reference profile archiving**	Profiles archived for future comparison and drift monitoring	Archive procedure	*Archived profiles*
19	**Panel characterisation reporting**	Per-model profiles, diversity assessment and selections assembled	Report template	*Panel report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SPB-08**	**2,10,000**

SPB-09
SECTOR 01

Patient-Derived Xenograft Genomic Analysis

(**PDX model genomic analysis** · PDX analysis)

WHY IT IS DONE

Xenograft models retain patient tumour heterogeneity better than cell lines, which makes them the closest preclinical proxy to clinical response. Analysing them requires separating human tumour signal from mouse stroma, which is a distinct technical problem with real consequences if handled naively.

WHAT IT ANSWERS

What is the genomic profile of the human tumour in this xenograft, once mouse-derived sequence has been removed, and how faithfully does it match the originating patient tumour?

WHEN IT IS DONE

In characterising xenograft models before efficacy studies, in tracking genomic drift across passages, and in confirming fidelity to the originating patient tumour.

WHAT TRIGGERS IT

Establishment or acquisition of a xenograft model, a passage milestone requiring drift assessment, or an efficacy study requiring model characterisation.

HOW IT IS DONE

Reads are aligned against combined human and mouse references so ambiguous reads can be assigned or discarded, mouse content is quantified, and human-assigned reads proceed to variant, copy number and expression analysis. Profiles are compared against the originating patient sample and across passages to quantify drift.

WHAT IT PRODUCES

Mouse content quantification, disambiguated human tumour profile across alteration classes, fidelity comparison against the originating tumour, passage drift analysis, and lymphoproliferative transformation screening.

WHAT IT DETECTS

Human tumour mutations, copy number and expression after mouse read removal, mouse stromal content proportion, genomic drift across passages, and fidelity to the originating patient sample.

WHAT IT DOES NOT DETECT

The human immune compartment is absent, so immuno-oncology mechanisms cannot be modelled. Lymphoproliferative transformation from human immune cells in the graft can occur and produces a sample that is no longer the tumour, which must be actively screened for.

WHAT MAKES IT HARD

Mouse reads mapping to conserved human sequence generate convincing false variants if not disambiguated, and simple filtering is insufficient. High stromal content compresses tumour signal, and drift across passages means a model characterised once may no longer be that model several passages later.

WHO DOES THIS JOB

Computational biologist in preclinical oncology, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Disambiguate human and mouse sequence rigorously, characterise xenograft tumours, quantify fidelity and passage drift, and screen for transformation.

WHO HIRES FOR IT

Pharmaceutical preclinical oncology, xenograft model providers, contract research organisations, and translational oncology programmes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**96 hours · 16 days**	Prior somatic analysis and species-aware alignment experience	**2,20,000**

SPB-09
WORKFLOW

Patient-Derived Xenograft Genomic Analysis

(**PDX model genomic analysis** · PDX analysis)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Model provenance recording**	Originating patient tumour, establishment date and passage recorded	Model record	*Provenance record*
02	**Reference preparation**	Combined human and mouse reference prepared and version locked	Reference preparation	*Combined reference*
03	**Data receipt and quality assessment**	Sequence integrity and quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
04	**Read preprocessing**	Adapters and low-quality tails removed	fastp	*Trimmed FASTQ*
05	**Combined reference alignment**	Reads aligned against human and mouse together so ambiguity is explicit	bwa-mem2, combined reference	*Aligned BAM*
06	**Species disambiguation**	Reads assigned to human or mouse with ambiguous reads discarded	Disambiguation method	*Disambiguated reads*
07	**Mouse content quantification**	Mouse-derived read proportion quantified as a sample quality attribute	Content quantification	*Mouse content*
08	**Human tumour content assessment**	Effective human tumour content assessed after stromal contribution	Content assessment	*Tumour content*
09	**Duplicate marking and recalibration**	Human reads duplicate marked and base qualities recalibrated	Picard, GATK BQSR	*Analysis-ready BAM*
10	**Coverage assessment**	Coverage assessed on human-assigned reads against requirements	mosdepth	*Coverage record*
11	**Variant calling**	Somatic variants called on the disambiguated human data	Validated caller	*Variant call set*
12	**Artefact filtering**	Residual mouse-derived artefacts and recurrent errors filtered	Artefact filtering	*Filtered variants*
13	**Copy number analysis**	Copy number called with stromal dilution accounted for	Copy number analysis	*Copy number profile*
14	**Expression profiling**	Expression quantified on human-assigned reads where RNA data exist	Salmon, STAR	*Expression profile*
15	**Originating tumour comparison**	Model profile compared against the originating patient tumour where available	Concordance analysis	*Fidelity assessment*
16	**Passage drift analysis**	Profiles compared across passages to quantify genomic drift	Drift analysis	*Drift assessment*
17	**Clonal selection assessment**	Subclonal loss and selection across passage assessed	Clonal analysis	*Selection assessment*
18	**Lymphoproliferative transformation screening**	Human immune-derived transformation screened for by marker and profile analysis	Transformation screening	*Transformation assessment*
19	**Model suitability determination**	Suitability for the intended study determined against all findings	Suitability criteria	*Suitability determination*
20	**Characterisation reporting**	Profile, fidelity, drift and suitability assembled	Report template	*PDX report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**96 h · 16 d**	**SPB-09**	**2,20,000**

SPB-10
SECTOR 01

Organoid & Primary Model Genomic Analysis

(**Organoid and primary culture genomic analysis** · Organoid analysis)

WHY IT IS DONE

Organoids retain the genomic and phenotypic features of the tissue they derive from far better than immortalised lines, which makes them the preferred model for translational work. They also drift, become clonally selected and can be overgrown by normal tissue, all of which must be monitored genomically.

WHAT IT ANSWERS

Does this organoid or primary culture still represent the tissue it derived from, genomically and clonally, and is it suitable for the intended experiment?

WHEN IT IS DONE

At establishment, at defined passage intervals, before pivotal experiments, and when comparing results across organoid lines in a biobank.

WHAT TRIGGERS IT

Establishment of a new organoid line, a passage milestone, an unexpected phenotypic change, or biobank quality review.

HOW IT IS DONE

Identity is authenticated against the originating tissue, then genomic profile is characterised with low-input methods appropriate to limited material. Clonal composition is compared across passages to detect selection, normal overgrowth is screened for by driver alteration allele fraction, and cross-contamination is assessed.

WHAT IT PRODUCES

Identity authentication against originating tissue, genomic characterisation, clonal composition across passages, normal overgrowth assessment, cross-contamination screening, and a suitability determination.

WHAT IT DETECTS

Genomic profile of the culture, fidelity to the originating tissue, clonal selection and subclonal loss across passage, normal tissue overgrowth in tumour organoids, and cross-contamination between lines.

WHAT IT DOES NOT DETECT

Genomic fidelity does not guarantee phenotypic fidelity, and epigenetic drift occurs without genomic change. Microenvironment, immune and vascular components are absent, which bounds what any organoid result can support.

WHAT MAKES IT HARD

Normal cell overgrowth in tumour organoids is the characteristic failure and looks like a clean result rather than a failed one, since the culture simply stops carrying tumour alterations. Detecting it requires tracking driver allele fractions rather than presence or absence.

WHO DOES THIS JOB

Computational biologist in translational or preclinical research, mid level, working with organoid biobank operations.

WHAT YOU CAN DO AFTERWARDS

Authenticate and characterise organoid models, track clonal selection across passage, detect normal overgrowth, and make suitability determinations.

WHO HIRES FOR IT

Pharmaceutical translational research, organoid biobanks, biotech using primary models, and academic translational units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
19	**92 hours · 16 days**	Prior somatic and low-input analysis experience	**2,10,000**

SPB-10
WORKFLOW

Organoid & Primary Model Genomic Analysis

(**Organoid and primary culture genomic analysis** · Organoid analysis)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Model provenance recording**	Originating tissue, establishment date and passage recorded	Model record	*Provenance record*
02	**Originating tissue reference verification**	A reference profile from the originating tissue confirmed available	Reference verification	*Reference availability*
03	**Identity authentication**	Culture identity authenticated against the originating tissue	Authentication analysis	*Identity clearance*
04	**Input quantity assessment**	Available material assessed against method requirements	Quantification	*Input record*
05	**Low-input method selection**	Analytical methods selected appropriate to limited material	Method selection	*Selected methods*
06	**Data receipt and quality assessment**	Sequence integrity and quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
07	**Alignment and processing**	Reads aligned and processed with low-input appropriate settings	bwa-mem2, Picard	*Analysis-ready BAM*
08	**Coverage assessment**	Coverage and uniformity assessed given limited input	mosdepth	*Coverage record*
09	**Variant calling**	Variants called with parameters appropriate to the input and coverage	Validated caller	*Variant call set*
10	**Copy number analysis**	Copy number called with limited-input noise accounted for	Copy number analysis	*Copy number profile*
11	**Originating tissue concordance**	Profile compared against the originating tissue for fidelity	Concordance analysis	*Fidelity assessment*
12	**Driver allele fraction tracking**	Driver alteration allele fractions tracked as the overgrowth indicator	Fraction analysis	*Driver fractions*
13	**Normal overgrowth assessment**	Loss of tumour alterations assessed as evidence of normal cell overgrowth	Overgrowth analysis	*Overgrowth assessment*
14	**Clonal composition analysis**	Clonal composition characterised and compared across passages	Clonal analysis	*Clonal composition*
15	**Passage selection assessment**	Subclonal loss and selection across passage quantified	Selection analysis	*Selection assessment*
16	**Cross-contamination screening**	Contamination from other lines in the biobank screened for	Contamination screening	*Contamination assessment*
17	**Phenotypic fidelity caveat documentation**	Limits of genomic fidelity as evidence of phenotypic fidelity documented	Limitation template	*Fidelity caveat*
18	**Suitability determination**	Suitability for the intended experiment determined and documented	Suitability criteria	*Suitability determination*
19	**Characterisation reporting**	Identity, fidelity, clonality and suitability assembled	Report template	*Organoid report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SPB-10**	**2,10,000**

SPB-11
SECTOR 01

Preclinical Pharmacogenomic & Toxicogenomic Analysis

(**Toxicogenomic and preclinical safety analysis** · Toxicogenomics)

WHY IT IS DONE

Toxicity is the leading cause of candidate attrition, and finding it late is the most expensive failure in drug development. Transcriptomic signatures in preclinical species detect mechanisms of injury before histopathology shows them, and connect a finding to a pathway rather than an observation.

WHAT IT ANSWERS

Does this compound produce transcriptional evidence of organ injury or a hazard mechanism in preclinical species, at what dose, and what pathway does it implicate?

WHEN IT IS DONE

In dose-range finding and repeat-dose toxicology studies, in mechanism investigation following a histopathological finding, and in candidate selection between structurally related compounds.

WHAT TRIGGERS IT

A regulatory toxicology study, an unexplained histopathological finding, a candidate selection decision, or a safety signal requiring mechanistic explanation.

HOW IT IS DONE

Study design, dose groups and timepoints are assessed, expression is quantified per organ and dose-response modelled rather than treated as pairwise comparison. Signatures are scored against established injury reference sets, pathway mechanisms are identified, and transcriptomic points of departure are derived for comparison against apical endpoints.

WHAT IT PRODUCES

Dose-response models per gene and pathway, injury signature scores against reference sets, mechanistic pathway findings, transcriptomic point of departure estimates, species comparison, and correlation against histopathology.

WHAT IT DETECTS

Dose-responsive transcriptional changes in target organs, established injury signatures, pathway-level hazard mechanisms, species differences in response, and the point of departure from transcriptomic dose-response modelling.

WHAT IT DOES NOT DETECT

A transcriptional change is not itself an adverse effect, and adaptive responses are frequently misread as toxicity. It does not predict human toxicity directly, since species differences in metabolism and receptor biology are substantial and this analysis measures response rather than translatability.

WHAT MAKES IT HARD

Separating adaptive response from adverse effect is a judgement that transcriptomics alone cannot settle, so correlation against apical endpoints is essential. Dose-response modelling rather than pairwise testing is what makes the analysis regulatory-relevant, and it is frequently done as the latter.

WHO DOES THIS JOB

Computational toxicologist or bioinformatician in preclinical safety, mid to senior level, working with toxicologic pathology.

WHAT YOU CAN DO AFTERWARDS

Model transcriptomic dose-response, score injury signatures, derive points of departure, and correlate molecular with apical findings.

WHO HIRES FOR IT

Pharmaceutical preclinical safety groups, contract research organisations running toxicology, agrochemical safety programmes, and regulatory science units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**88 hours · 15 days**	Prior differential expression analysis experience	**2,05,000**

SPB-11
WORKFLOW

Preclinical Pharmacogenomic & Toxicogenomic Analysis

(**Toxicogenomic and preclinical safety analysis** · Toxicogenomics)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Study design review**	Dose groups, timepoints, species and organs reviewed against the question	Design review	*Design assessment*
02	**Regulatory context confirmation**	Applicable guideline expectations for the study confirmed	Guideline review	*Regulatory context*
03	**Sample and organ inventory**	Samples inventoried by organ, dose group, timepoint and animal	Inventory compilation	*Sample inventory*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC, MultiQC	*QC acceptance record*
05	**Preprocessing and quantification**	Reads preprocessed and expression quantified per organ	fastp, STAR or Salmon	*Expression matrices*
06	**Species reference verification**	Species-appropriate reference and annotation versions confirmed	Reference registry	*Reference record*
07	**Outlier and quality filtering**	Outlying samples identified and exclusion decisions documented	Outlier analysis	*Outlier decisions*
08	**Normalisation**	Counts normalised within organ and study	Normalisation	*Normalised matrices*
09	**Dose-response modelling**	Expression modelled as a function of dose rather than pairwise contrasts	Dose-response modelling	*Dose-response models*
10	**Benchmark dose estimation**	Gene and pathway level benchmark doses estimated	BMD modelling	*Benchmark dose estimates*
11	**Transcriptomic point of departure derivation**	A transcriptomic point of departure derived per organ	POD derivation	*Point of departure*
12	**Injury signature scoring**	Established injury signatures scored against the observed profiles	Signature scoring	*Injury signature scores*
13	**Pathway mechanism identification**	Hazard mechanisms identified at pathway level	Pathway analysis	*Mechanism findings*
14	**Adaptive versus adverse assessment**	Responses assessed as adaptive or adverse against apical evidence	Interpretation framework	*Adaptivity assessment*
15	**Histopathology correlation**	Molecular findings correlated against histopathological observations	Correlation analysis	*Correlation record*
16	**Species comparison**	Responses compared across species where multi-species data exist	Cross-species analysis	*Species comparison*
17	**Human translatability caveat documentation**	Limits of preclinical response as a human predictor documented	Limitation template	*Translatability caveat*
18	**Regulatory reporting**	Dose-response, points of departure and mechanisms assembled to guideline expectations	Report template	*Toxicogenomics report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**88 h · 15 d**	**SPB-11**	**2,05,000**

SPB-12
SECTOR 01

Drug Response & Sensitivity Prediction Analysis

(**Genomic drug response modelling** · Sensitivity prediction)

WHY IT IS DONE

Predicting which tumours or models respond to a compound from their molecular features is what converts a broad indication into a targeted one. It is the analytical foundation of every biomarker-defined development strategy.

WHAT IT ANSWERS

Which molecular features predict response to this compound, how well do they predict, and does the model hold in data it was not trained on?

WHEN IT IS DONE

After panel screening data exist, in biomarker hypothesis generation before clinical development, and in retrospective analysis of response heterogeneity.

WHAT TRIGGERS IT

Completion of a model panel screen, unexplained response heterogeneity, or a biomarker strategy decision before clinical development.

HOW IT IS DONE

Response data are curated with the response metric defined explicitly, features are assembled and confounding by lineage assessed, and models are trained with cross-validation structured so that related samples do not span folds. Performance is evaluated on genuinely held-out data, feature importance is extracted, and transfer to independent or clinical datasets is tested where possible.

WHAT IT PRODUCES

The defined response metric and curated response matrix, model performance with validation structure stated, feature importance with interpretation, lineage confounding assessment, and independent dataset transfer results.

WHAT IT DETECTS

Molecular features associated with sensitivity or resistance, predictive model performance with validation, feature importance and interpretability, and the subset of models or tumours predicted to respond.

WHAT IT DOES NOT DETECT

Association is not causation, and a predictive feature may be a marker of a lineage that happens to be sensitive rather than a mechanism. Models trained on cell line panels frequently fail to transfer to patient data, and performance on held-out cell lines overstates clinical usefulness substantially.

WHAT MAKES IT HARD

Data leakage through related samples spanning cross-validation folds inflates performance systematically, and lineage confounding produces predictors that merely identify a tissue type. Transfer from preclinical to clinical data is the real test and is usually far worse than internal validation suggests.

WHO DOES THIS JOB

Computational biologist or data scientist in translational research, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Build and validate response prediction models without leakage, assess lineage confounding, and test transfer to independent data honestly.

WHO HIRES FOR IT

Pharmaceutical translational and precision medicine groups, oncology biotech, computational drug discovery companies, and academic pharmacogenomics units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
21	**102 hours · 17 days**	Prior genomic analysis and machine learning basics	**2,35,000**

SPB-12
WORKFLOW

Drug Response & Sensitivity Prediction Analysis

(**Genomic drug response modelling** · Sensitivity prediction)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Response metric definition**	The response endpoint defined explicitly since different metrics give different answers	Metric specification	*Defined response metric*
02	**Response data curation**	Response values curated across models with assay and batch recorded	Data curation	*Curated response matrix*
03	**Response data quality assessment**	Response reproducibility assessed where replicate measurements exist	Reproducibility analysis	*Response quality record*
04	**Feature assembly**	Genomic, transcriptomic and lineage features assembled per model	Feature assembly	*Feature matrix*
05	**Feature quality filtering**	Features with insufficient variance or excessive missingness removed	Filtering	*Filtered features*
06	**Lineage confounding assessment**	Lineage tested as a confounder of the apparent molecular association	Confounding analysis	*Lineage assessment*
07	**Sample relatedness assessment**	Related or derivative models identified to prevent leakage across folds	Relatedness analysis	*Relatedness record*
08	**Cross-validation structure design**	Fold structure designed so related samples never span train and test	CV design	*Validation structure*
09	**Baseline model establishment**	A simple baseline model established as the comparison for any complex model	Baseline modelling	*Baseline performance*
10	**Model training**	Predictive models trained within the designed validation structure	Model training	*Trained models*
11	**Hyperparameter selection**	Hyperparameters selected within training folds only, never on test data	Nested validation	*Selected hyperparameters*
12	**Performance evaluation**	Performance evaluated on genuinely held-out data with appropriate metrics	Evaluation	*Performance estimates*
13	**Confidence interval derivation**	Performance uncertainty derived by resampling	Resampling	*Performance intervals*
14	**Feature importance extraction**	Feature contributions extracted with stability across folds assessed	Importance analysis	*Feature importance*
15	**Interpretability assessment**	Important features assessed for biological plausibility and mechanism	Interpretation review	*Interpretability record*
16	**Lineage-controlled re-evaluation**	Performance re-evaluated with lineage controlled to test genuine molecular signal	Controlled evaluation	*Lineage-controlled performance*
17	**Independent dataset transfer**	Model applied to an independent dataset without retraining	Transfer evaluation	*Transfer performance*
18	**Clinical data transfer assessment**	Transfer to patient data assessed where such data exist	Clinical transfer evaluation	*Clinical transfer performance*
19	**Failure mode analysis**	Models examined for systematic failure in identifiable subgroups	Subgroup analysis	*Failure mode record*
20	**Deployment limitation documentation**	Conditions under which the model should not be applied documented	Limitation template	*Deployment limitations*
21	**Reporting**	Metric definition, validation structure, performance and limitations assembled	Report template	*Prediction report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SPB-12**	**2,35,000**

SPB-13
SECTOR 01

Mechanism of Action Deconvolution Analysis

(**Compound mechanism deconvolution** · MoA deconvolution)

WHY IT IS DONE

Compounds emerging from phenotypic screens work without anyone knowing why, and regulators, partners and internal decision-makers all require a mechanism. Transcriptional and genetic signatures connect an unknown compound to known biology faster than target-by-target investigation.

WHAT IT ANSWERS

What pathway or target does this compound act through, inferred from the transcriptional and genetic consequences of treating cells with it?

WHEN IT IS DONE

After a phenotypic screen produces an active compound, when an established compound shows unexpected activity, and in competitor compound characterisation.

WHAT TRIGGERS IT

An active compound from phenotypic screening with unknown mechanism, unexpected activity of a known compound, or characterisation of a competitor asset.

HOW IT IS DONE

Compounds are profiled transcriptionally across doses and timepoints with concentration-response structure preserved, signatures are computed against vehicle controls, and similarity is measured against reference compound and genetic perturbation libraries. Candidate mechanisms are ranked by concordance and tested against orthogonal evidence before any conclusion.

WHAT IT PRODUCES

Dose and time resolved transcriptional signatures, similarity scores against reference compound and perturbation libraries, ranked candidate mechanisms with concordance evidence, polypharmacology assessment, and orthogonal validation recommendations.

WHAT IT DETECTS

Compound-induced transcriptional signatures, similarity to reference compound and genetic perturbation signatures, implicated pathways, and candidate targets ranked by signature concordance.

WHAT IT DOES NOT DETECT

Signature similarity implies shared downstream consequence, not shared molecular target, so two compounds with matching signatures may act at different points in one pathway. Off-target polypharmacology produces composite signatures that resolve to no single mechanism.

WHAT MAKES IT HARD

Cytotoxicity dominates transcriptional response at high concentration and washes out mechanism-specific signal, so dose selection determines whether anything interpretable emerges. Reference library composition bounds what can be discovered, and a compound acting through unrepresented biology returns no match rather than a novel mechanism.

WHO DOES THIS JOB

Computational biologist in chemical biology or discovery, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Generate interpretable compound signatures across dose, match against reference libraries, and rank mechanistic hypotheses with orthogonal validation planned.

WHO HIRES FOR IT

Pharmaceutical chemical biology groups, phenotypic screening biotech, compound profiling platform companies, and academic chemical biology units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**98 hours · 17 days**	Prior transcriptomic and screen analysis experience	**2,25,000**

SPB-13
WORKFLOW

Mechanism of Action Deconvolution Analysis

(**Compound mechanism deconvolution** · MoA deconvolution)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Compound and question definition**	The compound, its activity and the mechanistic question defined	Programme protocol	*Defined question*
02	**Concentration range design**	Concentrations selected to produce mechanism signal below cytotoxic collapse	Dose design	*Concentration set*
03	**Cytotoxicity assessment**	Cytotoxicity measured across the concentration range to bound interpretation	Viability assay data	*Cytotoxicity profile*
04	**Timepoint selection**	Timepoints selected to capture primary rather than only secondary response	Timepoint design	*Timepoint set*
05	**Cell model selection**	Cell models selected for relevance and reference library compatibility	Model selection	*Selected models*
06	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC	*QC acceptance record*
07	**Expression quantification**	Expression quantified against a locked reference annotation	Salmon, STAR	*Expression matrix*
08	**Vehicle control normalisation**	Expression normalised against matched vehicle controls per plate and batch	Control normalisation	*Normalised expression*
09	**Plate and batch correction**	Plate position and batch effects corrected before signature computation	Batch correction	*Corrected expression*
10	**Signature computation**	Transcriptional signatures computed per compound, dose and timepoint	Signature computation	*Compound signatures*
11	**Signature strength assessment**	Signature magnitude assessed to identify uninformative low-activity conditions	Strength assessment	*Signature strength*
12	**Cytotoxic signature exclusion**	Conditions dominated by cytotoxic response excluded from mechanism inference	Exclusion criteria	*Excluded conditions*
13	**Reference library matching**	Signatures matched against reference compound signature libraries	Similarity analysis	*Compound similarity*
14	**Genetic perturbation library matching**	Signatures matched against genetic perturbation reference signatures	Similarity analysis	*Perturbation similarity*
15	**Similarity significance assessment**	Match significance assessed against the reference library null distribution	Statistical assessment	*Match significance*
16	**Pathway enrichment analysis**	Signature genes analysed for pathway enrichment to support mechanism hypotheses	Enrichment analysis	*Pathway results*
17	**Polypharmacology assessment**	Composite signatures assessed for evidence of multiple mechanisms	Composite analysis	*Polypharmacology assessment*
18	**Candidate mechanism ranking**	Mechanisms ranked on combined signature, perturbation and pathway evidence	Ranking framework	*Ranked mechanisms*
19	**Reference library coverage caveat**	Limits imposed by reference library composition documented	Limitation template	*Coverage caveat*
20	**Orthogonal validation planning**	Validation experiments specified to test the leading hypotheses	Validation planning	*Validation plan*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**98 h · 17 d**	**SPB-13**	**2,25,000**

SPB-14
SECTOR 01

Target Engagement & Pathway Modulation Analysis

(**Pharmacodynamic transcriptomic analysis** · Target engagement analysis)

WHY IT IS DONE

A compound that reaches its target and modulates the intended pathway may still fail, but one that does neither will certainly fail. Demonstrating engagement and downstream modulation is what justifies dose selection and continued investment, and regulators expect it.

WHAT IT ANSWERS

Did this compound engage its target and modulate the intended pathway at the doses tested, and what is the exposure-response relationship?

WHEN IT IS DONE

In preclinical pharmacodynamic studies, in early clinical dose selection, in proof-of-mechanism studies, and in comparing candidate molecules on pathway modulation.

WHAT TRIGGERS IT

A pharmacodynamic study, a dose selection decision, a proof-of-mechanism requirement, or candidate comparison.

HOW IT IS DONE

A pathway-specific signature is defined and validated in advance rather than derived post hoc, samples are profiled across doses and timepoints, and modulation is scored against that pre-specified signature. Exposure data are integrated to model exposure-response, duration is characterised, and specificity is assessed against off-target pathway signatures.

WHAT IT PRODUCES

The pre-specified pathway signature with validation, modulation scores across dose and time, exposure-response models, duration of effect characterisation, specificity assessment, and dose recommendation evidence.

WHAT IT DETECTS

Pathway-specific transcriptional modulation, dose and exposure response relationships, the magnitude and duration of modulation, and differences between candidate compounds on the same pathway.

WHAT IT DOES NOT DETECT

Transcriptional modulation is a downstream proxy for engagement, not engagement itself, and a compound may engage without producing transcriptional change. It also cannot distinguish on-target from off-target routes to the same downstream signature.

WHAT MAKES IT HARD

Defining the pharmacodynamic signature before the study is what makes the result credible, and deriving it afterwards from the same data is circular. Tissue accessibility limits where modulation can be measured, and surrogate tissue modulation may not reflect the target organ.

WHO DOES THIS JOB

Computational biologist in translational pharmacology, mid level, working with clinical pharmacology and pharmacometrics.

WHAT YOU CAN DO AFTERWARDS

Pre-specify and validate pharmacodynamic signatures, model exposure-response, characterise duration, and generate dose selection evidence.

WHO HIRES FOR IT

Pharmaceutical translational medicine and clinical pharmacology groups, biotech development teams, and contract research organisations.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**88 hours · 15 days**	Prior differential expression analysis experience	**2,05,000**

SPB-14
WORKFLOW

Target Engagement & Pathway Modulation Analysis

(**Pharmacodynamic transcriptomic analysis** · Target engagement analysis)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Pathway signature pre-specification**	The pharmacodynamic signature defined and locked before study samples are analysed	Signature specification	*Pre-specified signature*
02	**Signature validation**	The signature validated in an independent system before use as a readout	Validation study	*Signature validation record*
03	**Study design review**	Doses, timepoints, tissue and sampling reviewed against the readout	Design review	*Design assessment*
04	**Tissue accessibility assessment**	Whether the sampled tissue reflects the target organ assessed explicitly	Tissue assessment	*Accessibility assessment*
05	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC	*QC acceptance record*
06	**Expression quantification**	Expression quantified against a locked reference annotation	Salmon, STAR	*Expression matrix*
07	**Baseline normalisation**	Post-dose samples normalised against matched pre-dose or vehicle baselines	Baseline normalisation	*Normalised expression*
08	**Signature scoring**	The pre-specified signature scored per sample	Signature scoring	*Signature scores*
09	**Modulation quantification**	Magnitude of pathway modulation quantified relative to baseline	Quantification	*Modulation magnitude*
10	**Dose-response modelling**	Modulation modelled against administered dose	Dose-response modelling	*Dose-response model*
11	**Exposure-response modelling**	Modulation modelled against measured exposure rather than dose alone	Exposure-response modelling	*Exposure-response model*
12	**Duration characterisation**	Duration of modulation characterised across timepoints	Time-course analysis	*Duration profile*
13	**Inter-individual variability assessment**	Variability in modulation across subjects quantified	Variability analysis	*Variability assessment*
14	**Off-target pathway assessment**	Modulation of unintended pathways assessed as a specificity measure	Specificity analysis	*Specificity assessment*
15	**Engagement proxy limitation documentation**	The distinction between transcriptional modulation and direct engagement documented	Limitation template	*Proxy caveat*
16	**Target coverage calculation**	Proportion of the dosing interval with adequate modulation calculated	Coverage calculation	*Target coverage*
17	**Dose recommendation derivation**	Dose recommendations derived from exposure-response and duration evidence	Derivation framework	*Dose recommendation*
18	**Reporting**	Signature, modulation, exposure-response and dose evidence assembled	Report template	*Pharmacodynamic report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**88 h · 15 d**	**SPB-14**	**2,05,000**

SPB-15
SECTOR 01

Biomarker Discovery Analysis

(**Genomic and transcriptomic biomarker discovery** · Biomarker discovery)

WHY IT IS DONE

A biomarker that identifies responders converts a failed trial into a successful one in a defined population. Discovery is where that possibility is created, and where most biomarker programmes fail through overfitting long before they reach validation.

WHAT IT ANSWERS

Which molecular features distinguish responders from non-responders, or predict the outcome of interest, with performance that might survive independent validation?

WHEN IT IS DONE

In retrospective analysis of clinical or preclinical response data, in exploratory trial arms, and in indication expansion where a responsive subgroup is suspected.

WHAT TRIGGERS IT

Availability of response-linked molecular data, unexplained response heterogeneity in a trial, or an indication expansion hypothesis.

HOW IT IS DONE

The clinical question is specified as predictive or prognostic and the design assessed for whether it can answer that, then features are assembled with confounding examined. Candidate biomarkers are identified with multiplicity controlled, stability is assessed by resampling, performance is estimated with realistic optimism correction, and a validation plan is specified before any claim is made.

WHAT IT PRODUCES

The specified biomarker question and design adequacy assessment, candidate features with multiplicity-controlled significance, stability analysis across resampling, optimism-corrected performance estimates, and a pre-specified validation plan.

WHAT IT DETECTS

Candidate predictive and prognostic features, multivariate signatures with performance estimates, feature stability across resampling, and the effect size a biomarker-defined population would show.

WHAT IT DOES NOT DETECT

It cannot distinguish predictive from prognostic without an appropriate comparator arm, which is the most consequential confusion in the field. Discovery-stage performance always overstates validation performance, and no discovery analysis establishes clinical utility.

WHAT MAKES IT HARD

Predictive and prognostic biomarkers require different designs and are constantly conflated, so a marker of good outcome is presented as a marker of drug benefit. Overfitting is universal at discovery stage and only honest optimism correction and independent validation reveal it.

WHO DOES THIS JOB

Computational biologist or biostatistician in translational medicine, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Specify biomarker questions correctly, control multiplicity and overfitting, assess feature stability, and produce credible performance estimates with validation plans.

WHO HIRES FOR IT

Pharmaceutical translational medicine, precision medicine groups, diagnostic partners, and biotech development teams.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
22	**106 hours · 18 days**	Prior omics analysis and statistics experience	**2,45,000**

SPB-15
WORKFLOW

Biomarker Discovery Analysis

(**Genomic and transcriptomic biomarker discovery** · Biomarker discovery)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Biomarker question specification**	The question specified as predictive or prognostic with the distinction explicit	Question specification	*Specified question*
02	**Design adequacy assessment**	The study design assessed for whether it can answer the specified question	Design review	*Design adequacy*
03	**Endpoint definition**	The clinical endpoint defined precisely with censoring and timing rules	Endpoint specification	*Defined endpoint*
04	**Cohort assembly and characterisation**	The analysis cohort assembled with attrition and representativeness documented	Cohort assembly	*Cohort definition*
05	**Confounding assessment**	Clinical and technical confounders identified and their structure examined	Confounding analysis	*Confounding assessment*
06	**Feature assembly**	Candidate molecular features assembled with quality and missingness assessed	Feature assembly	*Feature matrix*
07	**Batch and technical correction**	Technical variation corrected before any association testing	Batch correction	*Corrected features*
08	**Missing data assessment**	Missing data mechanisms examined and handling strategy documented	Missingness analysis	*Missing data strategy*
09	**Univariate association testing**	Individual features tested against the endpoint with confounders modelled	Association testing	*Univariate results*
10	**Multiplicity control**	Significance adjusted across all features tested	FDR control	*Adjusted results*
11	**Multivariate signature derivation**	Multivariate signatures derived within a defined validation structure	Model derivation	*Candidate signatures*
12	**Resampling stability assessment**	Feature selection stability assessed across resampled datasets	Stability analysis	*Stability assessment*
13	**Optimism correction**	Performance estimates corrected for optimism from selection on the same data	Optimism correction	*Corrected performance*
14	**Predictive versus prognostic testing**	Interaction with treatment tested where the design permits	Interaction testing	*Predictive assessment*
15	**Effect size estimation in biomarker population**	Effect size estimated for the biomarker-defined population	Effect estimation	*Population effect estimate*
16	**Prevalence estimation**	Biomarker prevalence estimated in the intended clinical population	Prevalence analysis	*Prevalence estimate*
17	**Clinical utility distinction documentation**	The distinction between validity and utility documented explicitly	Limitation template	*Utility caveat*
18	**Assay feasibility assessment**	Feasibility of measuring the biomarker in a clinical assay assessed	Feasibility review	*Feasibility assessment*
19	**Validation plan specification**	The independent validation plan specified before any claim is made	Validation planning	*Pre-specified validation plan*
20	**Sample size planning for validation**	Validation sample size derived from the discovery effect estimate	Power calculation	*Validation sample size*
21	**Reporting**	Question, design, candidates, stability, performance and validation plan assembled	Report template	*Discovery report*
22	**Reproducibility documentation**	Versions, parameters and all analytical decisions documented	Documentation procedure	*Reproducibility record*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**106 h · 18 d**	**SPB-15**	**2,45,000**

SPB-16
SECTOR 01

Biomarker Validation & Assay Bridging Analysis

(**Biomarker validation and platform bridging** · Biomarker validation)

WHY IT IS DONE

A biomarker discovered on a research platform must work on the clinical assay that will eventually be deployed, and those two rarely agree perfectly. Bridging between them is where most biomarker programmes fail after discovery has succeeded.

WHAT IT ANSWERS

Does this biomarker replicate in independent data, and does the clinical assay reproduce the research platform's classification for the same samples?

WHEN IT IS DONE

After discovery produces a candidate, before any prospective clinical use, and whenever the measurement platform changes during development.

WHAT TRIGGERS IT

A validated discovery candidate progressing toward clinical use, a platform change during development, or a companion diagnostic partnership.

HOW IT IS DONE

An independent validation cohort is assessed for adequacy and the pre-specified analysis executed without modification, then the same samples are measured on both platforms and concordance assessed at both continuous and classification level. Thresholds are re-derived for the clinical platform rather than transferred, and discordance cases are examined individually.

WHAT IT PRODUCES

Independent validation results against the pre-specified analysis, platform concordance at continuous and classification level, re-derived clinical thresholds, discordance case analysis, and a validity assessment against intended use.

WHAT IT DETECTS

Replication of the biomarker association in independent data, concordance between research and clinical platforms, threshold transferability, and the classification discordance rate between platforms.

WHAT IT DOES NOT DETECT

Validation in one population does not establish generalisability to another, and platform concordance at the population level can coexist with substantial individual-level discordance. It does not establish clinical utility, only analytical and clinical validity.

WHAT MAKES IT HARD

Threshold transfer between platforms is where programmes break, because a cut-point optimal on one platform misclassifies on another even at high correlation. Individual-level discordance near the threshold is what matters clinically and is invisible in a correlation coefficient.

WHO DOES THIS JOB

Computational biologist or biostatistician in biomarker development, mid to senior level, working with diagnostic partners.

WHAT YOU CAN DO AFTERWARDS

Execute pre-specified validation, assess platform concordance at classification level, re-derive thresholds correctly, and examine discordance cases.

WHO HIRES FOR IT

Pharmaceutical biomarker and companion diagnostic groups, diagnostic companies, contract research organisations, and regulatory affairs functions.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**96 hours · 16 days**	Prior biomarker analysis experience	**2,20,000**

SPB-16
WORKFLOW

Biomarker Validation & Assay Bridging Analysis

(**Biomarker validation and platform bridging** · Biomarker validation)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Pre-specified analysis lock**	The validation analysis locked in writing before validation data are examined	Analysis specification	*Locked analysis plan*
02	**Validation cohort adequacy assessment**	Cohort size, composition and representativeness assessed against the plan	Cohort review	*Adequacy assessment*
03	**Cohort independence verification**	Independence from the discovery cohort verified at sample level	Independence check	*Independence record*
04	**Data receipt and quality assessment**	Validation data quality verified against the same criteria as discovery	QC procedures	*QC acceptance record*
05	**Pre-specified analysis execution**	The locked analysis executed without modification	Execution procedure	*Validation results*
06	**Replication assessment**	Replication assessed against the pre-specified success criteria	Criteria comparison	*Replication determination*
07	**Effect size comparison**	Validation effect size compared against the discovery estimate	Comparison analysis	*Effect comparison*
08	**Bridging sample set assembly**	A sample set spanning the clinical range assembled for platform bridging	Sample selection	*Bridging sample set*
09	**Dual platform measurement**	The same samples measured on research and clinical platforms	Measurement	*Paired measurements*
10	**Continuous concordance assessment**	Agreement assessed on the continuous measurement scale	Concordance analysis	*Continuous concordance*
11	**Systematic bias assessment**	Systematic offset between platforms quantified across the range	Bias analysis	*Bias assessment*
12	**Threshold re-derivation**	The clinical platform threshold derived independently rather than transferred	Threshold derivation	*Clinical threshold*
13	**Classification concordance assessment**	Agreement assessed at the classification level using the derived threshold	Classification analysis	*Classification concordance*
14	**Threshold region analysis**	Concordance assessed specifically among samples near the threshold	Region analysis	*Threshold region concordance*
15	**Discordance case examination**	Individual discordant cases examined for systematic causes	Case review	*Discordance analysis*
16	**Reproducibility assessment**	Within and between run reproducibility assessed on the clinical platform	Reproducibility study	*Reproducibility metrics*
17	**Generalisability assessment**	Applicability to populations beyond the validation cohort assessed	Generalisability review	*Generalisability assessment*
18	**Analytical validity determination**	Analytical validity determined against the intended use	Validity criteria	*Validity determination*
19	**Utility distinction documentation**	The distinction between validity and demonstrated clinical utility documented	Limitation template	*Utility caveat*
20	**Reporting**	Validation, concordance, thresholds and discordance assembled	Report template	*Validation report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**96 h · 16 d**	**SPB-16**	**2,20,000**

SPB-17
SECTOR 01

Clinical Trial Genomic Stratification Analysis

(**Trial enrichment and stratification analysis** · Trial stratification)

WHY IT IS DONE

Enrolling only patients whose tumours or genomes carry the relevant feature raises the observable effect size, reduces sample size and shortens trials. Stratification design determines whether a genuinely effective drug shows an effect at all.

WHAT IT ANSWERS

Which patients should be enrolled or stratified on genomic criteria, what proportion of the screened population will qualify, and how does that affect the trial's operational feasibility?

WHEN IT IS DONE

In trial design before protocol finalisation, during enrolment for screening performance monitoring, and in retrospective subgroup analysis of completed trials.

WHAT TRIGGERS IT

Design of a biomarker-defined trial, screening performance concerns during enrolment, or retrospective subgroup investigation.

HOW IT IS DONE

The stratification hypothesis is specified with the interaction test defined in advance, screening feasibility is modelled from prevalence data, and assay failure rates are incorporated into enrolment projections. During conduct, screening performance is monitored, and analysis follows the pre-specified plan with multiplicity across subgroups controlled.

WHAT IT PRODUCES

Prevalence and screening feasibility models, enrolment projections incorporating assay failure, screening performance monitoring, pre-specified subgroup analysis with interaction tests, and multiplicity-controlled results.

WHAT IT DETECTS

Eligible patient proportion from screening data, biomarker prevalence in the enrolled population, screening assay performance and failure rate, and subgroup effect estimates with appropriate multiplicity handling.

WHAT IT DOES NOT DETECT

It cannot establish that the biomarker-defined subgroup benefits more than others without a comparator design that permits interaction testing. Retrospective subgroup findings are hypothesis-generating regardless of statistical significance, and presenting them otherwise is the most common analytical error in trial reporting.

WHAT MAKES IT HARD

Screening failure rate is routinely underestimated at design stage and then dominates enrolment timelines, because samples fail quality control at rates the feasibility model never included. Post hoc subgroup analysis is the standard route to a false conclusion, and pre-specification is the only protection.

WHO DOES THIS JOB

Computational biologist or biostatistician in clinical development, mid to senior level, working with trial operations.

WHAT YOU CAN DO AFTERWARDS

Model screening feasibility realistically, monitor enrolment performance, and analyse biomarker-defined subgroups with correct interaction and multiplicity handling.

WHO HIRES FOR IT

Pharmaceutical clinical development, precision medicine groups, clinical research organisations, and trial biomarker operations.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
21	**102 hours · 17 days**	Prior biomarker and clinical data analysis experience	**2,35,000**

SPB-17
WORKFLOW

Clinical Trial Genomic Stratification Analysis

(**Trial enrichment and stratification analysis** · Trial stratification)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Stratification hypothesis specification**	The biomarker hypothesis and its role in the trial specified before design lock	Protocol specification	*Specified hypothesis*
02	**Interaction test pre-specification**	The interaction test establishing predictive benefit specified in advance	Statistical plan	*Pre-specified interaction test*
03	**Biomarker prevalence estimation**	Prevalence estimated in the intended enrolment population	Prevalence data	*Prevalence estimate*
04	**Screening assay performance assessment**	Assay success rate, turnaround and failure modes assessed	Assay performance data	*Assay performance record*
05	**Sample failure rate modelling**	Specimen and assay failure incorporated into enrolment projections	Failure modelling	*Failure rate model*
06	**Screening feasibility modelling**	Patients screened per patient enrolled modelled from prevalence and failure	Feasibility modelling	*Feasibility projection*
07	**Site capability assessment**	Site sample handling capability assessed against assay requirements	Site assessment	*Site capability record*
08	**Sample size derivation**	Sample size derived for the biomarker-defined population and interaction test	Power calculation	*Sample size*
09	**Stratification factor definition**	Stratification factors defined for randomisation	Randomisation design	*Stratification factors*
10	**Screening data monitoring**	Screening performance monitored against projection during enrolment	Monitoring procedure	*Screening monitoring*
11	**Assay failure investigation**	Failures investigated by cause and site during conduct	Failure analysis	*Failure investigation*
12	**Enrolled population characterisation**	Biomarker distribution in the enrolled population characterised against expectation	Population analysis	*Population characterisation*
13	**Selection bias assessment**	Differences between screened and enrolled populations assessed	Bias analysis	*Selection bias assessment*
14	**Pre-specified primary analysis**	The primary biomarker analysis executed as pre-specified	Analysis execution	*Primary results*
15	**Interaction testing**	Treatment by biomarker interaction tested as pre-specified	Interaction analysis	*Interaction results*
16	**Subgroup multiplicity control**	Multiplicity across all subgroups examined controlled and documented	Multiplicity control	*Adjusted subgroup results*
17	**Exploratory subgroup labelling**	Exploratory analyses labelled as hypothesis-generating irrespective of significance	Labelling policy	*Exploratory labelling*
18	**Consistency assessment**	Consistency across regions, sites and strata assessed	Consistency analysis	*Consistency assessment*
19	**Assay-related missingness assessment**	Patients excluded by assay failure assessed for systematic difference	Missingness analysis	*Missingness assessment*
20	**Regulatory presentation preparation**	Results prepared to regulatory expectations for biomarker-defined claims	Presentation preparation	*Regulatory package*
21	**Reporting**	Feasibility, screening performance, primary and subgroup results assembled	Report template	*Stratification report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SPB-17**	**2,35,000**

SPB-18
SECTOR 01

Companion Diagnostic Development Analysis

(**Companion diagnostic analytical development** · CDx development)

WHY IT IS DONE

A drug approved for a biomarker-defined population requires an approved test to identify that population, and the two are reviewed together. The analytical development of that test is a regulated activity with requirements far beyond research biomarker work.

WHAT IT ANSWERS

Does the companion assay meet the analytical performance requirements for its intended use, and is its concordance with the clinical trial assay sufficient to support the drug label?

WHEN IT IS DONE

In parallel with clinical development from the point a biomarker strategy is adopted, and in bridging studies when the assay changes during development.

WHAT TRIGGERS IT

Adoption of a biomarker-defined development strategy, a diagnostic partnership, or an assay change requiring a bridging study.

HOW IT IS DONE

Intended use is defined precisely because it determines every requirement that follows, then analytical validation is executed covering sensitivity, specificity, precision, reproducibility and interference. Concordance against the clinical trial assay is established on a defined sample set, performance near the threshold is characterised specifically, and the evidence package is assembled to regulatory expectations.

WHAT IT PRODUCES

The intended use statement, analytical validation results across all required parameters, clinical trial assay concordance with threshold-region analysis, failure mode characterisation, and a regulatory-ready evidence package.

WHAT IT DETECTS

Analytical sensitivity, specificity, precision and reproducibility of the assay, concordance against the clinical trial assay, performance at the clinically relevant threshold, and failure modes across specimen types.

WHAT IT DOES NOT DETECT

Analytical performance does not establish clinical utility, which comes from the trial. It also cannot anticipate every real-world specimen condition, and post-market performance frequently differs from validation study performance.

WHAT MAKES IT HARD

Performance near the decision threshold is what matters and is where every assay is weakest, so characterising it specifically rather than reporting overall concordance is essential. Bridging studies when the assay changes mid-development are the commonest cause of programme delay.

WHO DOES THIS JOB

Computational biologist or assay development scientist in diagnostics, senior level, working with regulatory affairs.

WHAT YOU CAN DO AFTERWARDS

Define intended use, execute analytical validation to regulatory standard, establish concordance with threshold-region focus, and assemble submission evidence.

WHO HIRES FOR IT

Diagnostic companies, pharmaceutical companion diagnostic groups, regulatory consultancies, and contract diagnostic developers.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
24	**118 hours · 20 days**	Prior biomarker validation and assay analysis experience	**2,70,000**

SPB-18
WORKFLOW

Companion Diagnostic Development Analysis

(**Companion diagnostic analytical development** · CDx development)

COMPLETE WORKFLOW — 24 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Intended use definition**	The intended use defined precisely since it determines every requirement	Intended use specification	*Intended use statement*
02	**Regulatory pathway determination**	The applicable regulatory pathway and its evidence expectations determined	Regulatory review	*Pathway determination*
03	**Specimen type scope definition**	Specimen types within intended use defined with handling requirements	Scope specification	*Specimen scope*
04	**Analytical performance requirement derivation**	Required sensitivity, specificity and precision derived from intended use	Requirement derivation	*Performance requirements*
05	**Reference material selection**	Reference and control materials selected for validation studies	Material selection	*Reference materials*
06	**Analytical sensitivity study**	Limit of detection established across specimen types and variant classes	Sensitivity study	*Sensitivity results*
07	**Analytical specificity study**	Specificity established including interference and cross-reactivity	Specificity study	*Specificity results*
08	**Precision study**	Within-run, between-run and between-site precision established	Precision study	*Precision results*
09	**Reproducibility study**	Reproducibility established across operators, instruments and sites	Reproducibility study	*Reproducibility results*
10	**Threshold region characterisation**	Performance characterised specifically among samples near the decision threshold	Threshold study	*Threshold performance*
11	**Specimen quality impact study**	Performance assessed across the range of real-world specimen quality	Specimen study	*Quality impact results*
12	**Failure mode characterisation**	Assay failure modes and their rates characterised by cause	Failure analysis	*Failure mode record*
13	**Clinical trial assay comparison design**	The concordance study against the trial assay designed with sample selection justified	Study design	*Concordance study design*
14	**Concordance sample set assembly**	Samples spanning the clinical range and threshold region assembled	Sample selection	*Concordance sample set*
15	**Concordance measurement**	Both assays run on the same samples under controlled conditions	Measurement	*Paired results*
16	**Positive and negative agreement calculation**	Agreement calculated with confidence intervals at the decision threshold	Agreement analysis	*Agreement metrics*
17	**Discordance investigation**	Every discordant sample investigated to establish cause	Case investigation	*Discordance analysis*
18	**Bridging analysis**	Bridging established where the trial and commercial assays differ	Bridging analysis	*Bridging results*
19	**Software and pipeline validation**	The analytical pipeline validated with version control and change control	Pipeline validation	*Software validation record*
20	**Traceability documentation**	Traceability from specimen to reported result documented end to end	Traceability documentation	*Traceability record*
21	**Risk analysis**	Risks to correct result identified with mitigations documented	Risk analysis	*Risk file*
22	**Post-market surveillance planning**	Post-market performance monitoring planned before launch	Surveillance planning	*Surveillance plan*
23	**Evidence package assembly**	All studies assembled into a submission-ready evidence package	Package assembly	*Evidence package*
24	**Regulatory review preparation**	Package reviewed against submission expectations before filing	Review procedure	*Submission-ready package*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
24	**118 h · 20 d**	**SPB-18**	**2,70,000**

SPB-19
SECTOR 01

Pharmacogenomic Trial Analysis

(**Clinical trial pharmacogenomic analysis** · Trial PGx analysis)

WHY IT IS DONE

Variability in drug exposure and adverse events across a trial population frequently has a genetic explanation, and finding it can rescue a dose, refine a label or explain an outlier safety signal that would otherwise stop a programme.

WHAT IT ANSWERS

Do genetic variants explain the variation in exposure, response or adverse events observed in this trial population?

WHEN IT IS DONE

In exploratory pharmacogenomic arms of clinical trials, in investigating unexplained exposure variability or adverse events, and in label development for a marketed product.

WHAT TRIGGERS IT

An unexplained adverse event cluster, exposure variability exceeding expectations, an exploratory pharmacogenomic protocol, or a label development requirement.

HOW IT IS DONE

Consent scope for genetic analysis is confirmed, pharmacogene diplotypes are called and translated to phenotype, and association is tested against pharmacokinetic and safety endpoints with population structure controlled. Power is assessed honestly, exposure-response is modelled by genotype group, and findings are placed against existing guideline evidence.

WHAT IT PRODUCES

Diplotype and phenotype assignments across the population, association results with power assessment, genotype-stratified exposure-response models, adverse event association analysis, and label and guideline implications.

WHAT IT DETECTS

Associations between pharmacogene variants and pharmacokinetic parameters, exposure-response relationships stratified by genotype, adverse event associations, and diplotype-phenotype relationships in the trial population.

WHAT IT DOES NOT DETECT

Trials are rarely powered for genetic association, so absence of association usually means insufficient power rather than absence of effect. Rare variants are unmeasurable at trial scale, and non-genetic sources of variability frequently exceed the genetic contribution.

WHAT MAKES IT HARD

Power is almost always the binding constraint and is frequently not stated, so null results are misread as evidence of no effect. Ancestry structure in a multinational trial confounds association unless controlled, and pharmacogene diplotype calling from trial-grade data is often incomplete.

WHO DOES THIS JOB

Computational biologist or pharmacogenomics scientist in clinical development, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Call pharmacogene diplotypes at trial scale, test association with appropriate power and structure control, and interpret findings for labelling.

WHO HIRES FOR IT

Pharmaceutical clinical pharmacology and development groups, contract research organisations, regulatory science functions, and pharmacogenomics consultancies.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
19	**92 hours · 16 days**	Prior pharmacogenomic and clinical data analysis experience	**2,10,000**

SPB-19
WORKFLOW

Pharmacogenomic Trial Analysis

(**Clinical trial pharmacogenomic analysis** · Trial PGx analysis)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Consent scope verification**	Genetic analysis consent verified for every subject before analysis	Consent review	*Analysis authorisation*
02	**Analysis plan specification**	The pharmacogenomic analysis plan specified with endpoints defined	Statistical plan	*Specified plan*
03	**Gene and variant scope definition**	Pharmacogenes and variants in scope defined with definition versions locked	Scope specification	*Gene scope*
04	**Data receipt and quality assessment**	Genotype data quality assessed with call rate and concordance checks	QC procedures	*QC acceptance record*
05	**Sample identity verification**	Genetic samples verified against clinical records for correct attribution	Identity verification	*Identity clearance*
06	**Diplotype calling**	Pharmacogene diplotypes called against the versioned definition tables	PharmCAT	*Diplotype calls*
07	**Structural and repeat allele analysis**	Structural and repeat alleles resolved with dedicated methods	Specialist analysis	*Structural findings*
08	**Call completeness assessment**	Diplotype call rate assessed with uncallable subjects identified	Completeness analysis	*Completeness record*
09	**Phenotype translation**	Diplotypes translated to metabolic phenotype using current activity scores	Activity score tables	*Metabolic phenotypes*
10	**Ancestry inference**	Genetic ancestry inferred for population structure control	PCA projection	*Ancestry assignment*
11	**Population structure control**	Ancestry included in association models to control confounding	Structure control	*Controlled models*
12	**Power assessment**	Statistical power assessed given phenotype frequencies and cohort size	Power calculation	*Power assessment*
13	**Pharmacokinetic association testing**	Associations tested between phenotype groups and exposure parameters	Association testing	*PK association results*
14	**Exposure-response modelling by genotype**	Exposure-response modelled separately by metabolic phenotype group	Modelling	*Genotype-stratified models*
15	**Adverse event association testing**	Associations tested between phenotype and adverse event occurrence	Association testing	*Safety association results*
16	**Multiplicity control**	Significance adjusted across genes and endpoints tested	Multiplicity control	*Adjusted results*
17	**Guideline evidence comparison**	Findings compared against existing prescribing guideline evidence	Guideline comparison	*Evidence comparison*
18	**Labelling implication assessment**	Implications for product labelling assessed with regulatory input	Labelling assessment	*Labelling implications*
19	**Reporting**	Diplotypes, associations, power and labelling implications assembled	Report template	*PGx trial report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SPB-19**	**2,10,000**

SPB-20
SECTOR 01

Resistance Mechanism Discovery Analysis

(**Acquired resistance mechanism discovery** · Resistance discovery)

WHY IT IS DONE

Every targeted therapy eventually fails, and the mechanism determines whether a next-generation compound is possible. Identifying resistance early shapes the follow-on molecule and the combination strategy before the first one loses its market.

WHAT IT ANSWERS

How does resistance to this compound arise, through which molecular mechanisms, and does a tractable intervention exist for each?

WHEN IT IS DONE

In preclinical resistance modelling before clinical data exist, in analysing paired pre-treatment and progression clinical samples, and in designing next-generation or combination strategies.

WHAT TRIGGERS IT

Emergence of clinical resistance, a preclinical resistance modelling programme, or a next-generation compound design decision.

HOW IT IS DONE

Resistant models or paired clinical samples are compared against their sensitive counterparts across mutation, copy number, fusion and expression, with convergence across independent lineages required before a mechanism is claimed. Bypass signalling is assessed at pathway level, transformation is screened for, and each mechanism is assessed for tractability.

WHAT IT PRODUCES

Mechanism findings per model or patient across alteration classes, convergence analysis across independent instances, bypass pathway assessment, transformation screening, and tractability assessment per mechanism.

WHAT IT DETECTS

On-target resistance mutations, bypass pathway activation, target amplification or loss, phenotypic and lineage transformation signatures, and convergence of mechanisms across independent resistant models or patients.

WHAT IT DOES NOT DETECT

Non-genetic and epigenetic resistance requires expression and chromatin analysis rather than sequence alone, and drug tolerance persister states leave no genomic mark. Mechanisms in unsampled sites are invisible, and a single resistant model may reveal a mechanism that does not generalise.

WHAT MAKES IT HARD

A mechanism found in one resistant clone frequently does not generalise, so convergence across independent instances is what distinguishes a real mechanism from a passenger event. Non-genetic resistance is common and entirely missed by sequence-focused analysis.

WHO DOES THIS JOB

Computational biologist in translational oncology or discovery, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Compare resistant against sensitive states across alteration classes, require convergence before claiming mechanism, and assess tractability.

WHO HIRES FOR IT

Pharmaceutical translational oncology, targeted therapy biotech, academic resistance research programmes, and combination strategy groups.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
21	**102 hours · 17 days**	Prior somatic and screen analysis experience	**2,35,000**

SPB-20
WORKFLOW

Resistance Mechanism Discovery Analysis

(**Acquired resistance mechanism discovery** · Resistance discovery)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Resistance question definition**	The resistance question and available material defined	Programme protocol	*Defined question*
02	**Sample pairing verification**	Pre-treatment and resistant samples verified as originating from the same source	Identity verification	*Pairing verification*
03	**Resistance model provenance recording**	Derivation route for resistant models recorded including selection conditions	Model record	*Provenance record*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified for all samples	md5sum, FastQC	*QC acceptance record*
05	**Alignment and processing**	All samples aligned and processed identically	bwa-mem2, Picard, GATK BQSR	*Analysis-ready BAMs*
06	**Baseline profile establishment**	The sensitive or pre-treatment profile established as the comparator	Baseline analysis	*Baseline profile*
07	**Variant calling**	Variants called in resistant samples with the baseline used for comparison	Validated caller	*Variant call sets*
08	**Emergent variant identification**	Variants absent at baseline and present at resistance identified	Comparative analysis	*Emergent variants*
09	**Detection limit reconciliation**	Baseline detection limits assessed so emergence is distinguished from prior invisibility	Sensitivity comparison	*Emergence assessment*
10	**On-target resistance assessment**	Mutations in the drug target assessed for known or plausible resistance mechanism	Target analysis	*On-target findings*
11	**Copy number analysis**	Target amplification and loss assessed as resistance mechanisms	Copy number analysis	*Copy number findings*
12	**Fusion and structural analysis**	Structural mechanisms of resistance assessed where data permit	Structural analysis	*Structural findings*
13	**Expression profiling**	Expression changes assessed for bypass pathway activation	Expression analysis	*Expression findings*
14	**Bypass pathway assessment**	Pathway-level activation assessed as a non-genetic resistance route	Pathway analysis	*Bypass findings*
15	**Lineage transformation screening**	Phenotypic and lineage transformation screened for in expression data	Transformation analysis	*Transformation assessment*
16	**Convergence analysis**	Mechanisms assessed for recurrence across independent resistant instances	Convergence analysis	*Convergence assessment*
17	**Passenger discrimination**	Non-recurrent findings assessed as probable passengers rather than mechanisms	Discrimination criteria	*Passenger assessment*
18	**Functional validation planning**	Validation experiments specified to test candidate mechanisms causally	Validation planning	*Validation plan*
19	**Tractability assessment**	Each mechanism assessed for whether an intervention exists or is feasible	Tractability review	*Tractability assessment*
20	**Combination strategy derivation**	Combination or next-generation strategies derived from the mechanism set	Strategy derivation	*Strategy recommendations*
21	**Reporting**	Mechanisms, convergence, tractability and strategy assembled	Report template	*Resistance report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SPB-20**	**2,35,000**

SPB-21
SECTOR 01

Immunogenicity & Anti-Drug Antibody Repertoire Analysis

(**Immunogenicity repertoire analysis** · Immunogenicity analysis)

WHY IT IS DONE

Biologic therapeutics provoke immune responses that neutralise the drug, alter its exposure and occasionally cause harm. Characterising the responding repertoire explains why some patients respond immunologically and others do not, which conventional antibody assays cannot.

WHAT IT ANSWERS

What immune repertoire response has this biologic provoked, which clones expanded, and do the responding patients share features that predict immunogenicity?

WHEN IT IS DONE

In clinical development of biologics where anti-drug antibodies are detected, in comparing immunogenicity across formulations, and in investigating loss of response.

WHAT TRIGGERS IT

Detection of anti-drug antibodies in a trial, unexplained loss of response, a formulation comparison, or a biosimilar comparability programme.

HOW IT IS DONE

Baseline and post-exposure repertoires are sequenced with depth normalisation planned in advance, clonotypes are assembled with error correction, and expansion is identified against the baseline repertoire. HLA is typed and tested for association with immunogenicity status, and repertoire features are compared between responders and non-responders.

WHAT IT PRODUCES

Baseline and post-exposure repertoire profiles, expanded clone identification with dynamics, HLA association analysis, responder against non-responder repertoire comparison, and immunogenicity risk feature assessment.

WHAT IT DETECTS

B-cell and T-cell repertoire changes following exposure, clonal expansion associated with the response, HLA associations with immunogenicity, and repertoire differences between responding and non-responding patients.

WHAT IT DOES NOT DETECT

It does not measure antibody binding or neutralisation, which require functional assays. Repertoire from blood may not reflect the tissue compartment where the response is generated, and it cannot establish that an expanded clone is drug-specific without additional experimental work.

WHAT MAKES IT HARD

Repertoire metrics are exquisitely sensitive to sampling depth and input, so cross-timepoint comparison requires normalisation designed in from the start. Demonstrating that an expanded clone is drug-specific rather than coincidental needs experimental confirmation that repertoire sequencing alone cannot provide.

WHO DOES THIS JOB

Computational immunologist or bioinformatician in biologics development, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Analyse repertoire responses to biologics with depth normalisation, identify expansion, and test HLA and repertoire association with immunogenicity.

WHO HIRES FOR IT

Biologics development groups, biosimilar developers, immunogenicity assessment functions, and contract research organisations supporting biologics.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**88 hours · 15 days**	Prior immune repertoire analysis experience	**2,05,000**

SPB-21
WORKFLOW

Immunogenicity & Anti-Drug Antibody Repertoire Analysis

(**Immunogenicity repertoire analysis** · Immunogenicity analysis)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Study design review**	Sampling timepoints, compartments and depth reviewed against the question	Design review	*Design assessment*
02	**Depth normalisation planning**	Sequencing depth planned so cross-timepoint comparison is valid	Normalisation planning	*Depth plan*
03	**Baseline sample verification**	Pre-exposure baseline samples verified available for every subject compared	Sample inventory	*Baseline verification*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC	*QC acceptance record*
05	**Repertoire extraction**	Receptor sequences extracted from targeted or bulk data	MiXCR, IgBLAST	*Extracted repertoire*
06	**Error correction**	Sequencing error corrected without destroying genuine clonal variation	Error correction	*Corrected sequences*
07	**Clonotype assembly**	Clonotypes assembled with consistent definition criteria	Clonotype assembly	*Clonotype tables*
08	**Repertoire quality assessment**	Recovery, depth and error metrics assessed against acceptance criteria	Repertoire QC	*QC record*
09	**Depth normalisation**	Repertoires normalised to comparable depth before any comparison	Normalisation	*Normalised repertoires*
10	**Diversity metric computation**	Diversity indices computed on normalised repertoires	Diversity calculation	*Diversity metrics*
11	**Baseline comparison**	Post-exposure repertoires compared against matched baselines	Comparative analysis	*Change assessment*
12	**Clonal expansion identification**	Expanded clones identified against baseline with statistical support	Expansion analysis	*Expanded clones*
13	**Expansion dynamics analysis**	Clone trajectories analysed across timepoints	Dynamics analysis	*Clonal dynamics*
14	**HLA typing**	Subject HLA alleles typed for immunogenicity association testing	HLA typing	*HLA assignments*
15	**HLA association testing**	HLA alleles tested for association with immunogenicity status	Association testing	*HLA associations*
16	**Responder comparison**	Repertoire features compared between responders and non-responders	Comparative analysis	*Responder comparison*
17	**Specificity limitation documentation**	The inability to establish drug specificity from repertoire alone documented	Limitation template	*Specificity caveat*
18	**Reporting**	Repertoire change, expansion, HLA association and comparison assembled	Report template	*Immunogenicity report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**88 h · 15 d**	**SPB-21**	**2,05,000**

SPB-22
SECTOR 01

Antibody Discovery Repertoire Analysis

(**Antibody repertoire mining and discovery** · Antibody discovery analysis)

WHY IT IS DONE

Sequencing the antibody repertoire of an immunised animal or a convalescent donor and mining it computationally identifies candidate binders far faster than screening hybridomas one at a time. It has become the dominant route into antibody discovery programmes.

WHAT IT ANSWERS

Which antibody sequences in this repertoire are the strongest candidates for the target, based on clonal expansion, maturation and lineage structure?

WHEN IT IS DONE

After immunisation or donor collection in an antibody discovery campaign, in mining historical repertoire datasets, and in optimising a lead antibody by lineage analysis.

WHAT TRIGGERS IT

Completion of an immunisation campaign, availability of convalescent or patient donor material, or a lead optimisation objective.

HOW IT IS DONE

Sequences are assembled with error correction appropriate to the very high similarity between clonal variants, gene segments assigned and junctions characterised, and clonal lineages reconstructed. Expansion and maturation are quantified, chain pairing is established where single-cell data permit, and candidates are ranked on lineage and maturation evidence with developability liabilities screened.

WHAT IT PRODUCES

The assembled repertoire with gene assignment, reconstructed clonal lineages with maturation trajectories, expansion quantification, paired chain assignments where available, ranked candidate sequences, and developability liability screening.

WHAT IT DETECTS

Antibody heavy and light chain sequences with gene segment assignment, clonal lineages and their expansion, somatic hypermutation load and maturation trajectory, paired chain information where the method provides it, and candidate sequences ranked by lineage evidence.

WHAT IT DOES NOT DETECT

Sequence does not establish binding, affinity or developability, all of which require expression and testing. Repertoire from blood may not represent the tissue where the relevant response resides, and unpaired sequencing loses the chain pairing that determines specificity.

WHAT MAKES IT HARD

Distinguishing genuine clonal variants from sequencing error is the core technical problem, because real variants differ by as little as one base and error correction that is too aggressive destroys the maturation signal being sought. Unpaired data cannot recover chain pairing, which limits everything downstream.

WHO DOES THIS JOB

Computational biologist in antibody discovery, mid to senior level, working with protein engineering.

WHAT YOU CAN DO AFTERWARDS

Assemble repertoires with appropriate error correction, reconstruct clonal lineages, establish chain pairing, and rank candidates on lineage evidence.

WHO HIRES FOR IT

Antibody discovery biotech, pharmaceutical biologics groups, antibody platform companies, and contract discovery organisations.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
22	**106 hours · 18 days**	Prior immune repertoire analysis experience	**2,45,000**

SPB-22
WORKFLOW

Antibody Discovery Repertoire Analysis

(**Antibody repertoire mining and discovery** · Antibody discovery analysis)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Campaign design review**	Immunisation or donor strategy and sampling reviewed against discovery goals	Design review	*Design assessment*
02	**Sample and compartment recording**	Tissue and compartment sampled recorded as it bounds the repertoire recovered	Sample record	*Compartment record*
03	**Chain pairing capability determination**	Whether the method provides chain pairing determined and recorded	Method review	*Pairing capability*
04	**Data receipt and quality assessment**	Sequence integrity and quality verified per library	md5sum, FastQC	*QC acceptance record*
05	**Read merging and preprocessing**	Paired reads merged to recover full variable region sequence	Read merging	*Merged reads*
06	**Unique molecular identifier processing**	Molecular identifiers used to build consensus where the method provides them	UMI consensus	*Consensus sequences*
07	**Error correction calibration**	Error correction calibrated so genuine clonal variants survive	Calibration procedure	*Calibrated correction*
08	**Sequence assembly**	Heavy and light chain sequences assembled with quality thresholds	Assembly	*Assembled sequences*
09	**Gene segment assignment**	Variable, diversity and joining segments assigned against reference databases	IgBLAST, MiXCR	*Segment assignments*
10	**Junction characterisation**	Junction regions characterised including insertions and deletions	Junction analysis	*Junction characterisation*
11	**Productivity filtering**	Non-productive rearrangements identified and separated	Productivity analysis	*Productive sequences*
12	**Clonal lineage reconstruction**	Sequences grouped into clonal lineages by shared rearrangement	Lineage clustering	*Clonal lineages*
13	**Somatic hypermutation quantification**	Mutation load quantified per sequence and per lineage	Mutation analysis	*Hypermutation profiles*
14	**Maturation trajectory reconstruction**	Lineage trees built to reconstruct affinity maturation trajectories	Lineage tree construction	*Maturation trees*
15	**Clonal expansion quantification**	Lineage abundance quantified as evidence of antigen-driven selection	Abundance analysis	*Expansion metrics*
16	**Selection pressure analysis**	Evidence of selection assessed from mutation patterns in framework and binding regions	Selection analysis	*Selection assessment*
17	**Chain pairing assignment**	Heavy and light chains paired where single-cell data permit	Pairing analysis	*Paired sequences*
18	**Public and convergent sequence identification**	Sequences shared across donors identified as convergent responses	Convergence analysis	*Convergent sequences*
19	**Germline reversion assessment**	Germline-reverted forms assessed to understand the maturation pathway	Reversion analysis	*Reversion assessment*
20	**Developability liability screening**	Sequences screened for aggregation, glycosylation and stability liabilities	Liability screening	*Liability assessment*
21	**Candidate ranking**	Candidates ranked on expansion, maturation, convergence and developability	Ranking framework	*Ranked candidates*
22	**Expression and testing prioritisation**	Candidates prioritised for expression and binding characterisation	Prioritisation criteria	*Testing list*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**106 h · 18 d**	**SPB-22**	**2,45,000**

SPB-23
SECTOR 01

CAR-T & Cell Therapy Product Characterisation Analysis

(**Cell therapy product genomic characterisation** · Cell therapy characterisation)

WHY IT IS DONE

An engineered cell product is a living drug whose composition varies between batches and patients. Characterising what is actually in the infusion bag, and what happens to it afterwards, is both a release requirement and the route to understanding why some patients respond.

WHAT IT ANSWERS

What is the composition, transgene status and clonal structure of this cell therapy product, and how does it persist and evolve after infusion?

WHEN IT IS DONE

In product release characterisation, in comparability between manufacturing processes, and in correlative analysis of persistence and response after infusion.

WHAT TRIGGERS IT

A manufacturing batch requiring characterisation, a process comparability exercise, or a correlative study on persistence and response.

HOW IT IS DONE

Transgene integration and copy number are quantified, transduced proportion determined, and single-cell profiling used to establish composition and differentiation state. Repertoire is analysed for clonal diversity, and post-infusion samples are compared against the product to characterise persistence, expansion and phenotype change.

WHAT IT PRODUCES

Transgene copy number and transduced proportion, product composition and differentiation state profile, repertoire diversity and clonal structure, post-infusion persistence and phenotype evolution, and batch comparability assessment.

WHAT IT DETECTS

Transgene presence and copy number per cell, transduced cell proportion, cell type and differentiation state composition, T-cell receptor repertoire and clonal structure, and post-infusion persistence and phenotype evolution.

WHAT IT DOES NOT DETECT

It does not measure cytotoxic function, which requires functional assays. Cells in tissue compartments are not sampled by blood analysis, and the infusion product composition does not predict in vivo behaviour reliably.

WHAT MAKES IT HARD

Product composition varies enormously between patients because the starting material does, so comparability must be assessed against that baseline variation rather than against a fixed specification. Distinguishing product-derived from endogenous cells post-infusion requires the transgene or clonal markers rather than phenotype.

WHO DOES THIS JOB

Computational biologist in cell therapy development, senior level, working with manufacturing and translational science.

WHAT YOU CAN DO AFTERWARDS

Characterise engineered cell products across transgene, composition and repertoire, and analyse post-infusion persistence and evolution.

WHO HIRES FOR IT

Cell therapy developers, contract manufacturing organisations, academic cell therapy programmes, and pharmaceutical cell therapy units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
23	**112 hours · 19 days**	Prior single-cell and repertoire analysis experience	**2,55,000**

SPB-23
WORKFLOW

CAR-T & Cell Therapy Product Characterisation Analysis

(**Cell therapy product genomic characterisation** · Cell therapy characterisation)

COMPLETE WORKFLOW — 23 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Product and process recording**	Manufacturing process, batch identity and starting material recorded	Batch record	*Product record*
02	**Starting material characterisation**	Apheresis starting material characterised as the baseline for comparability	Baseline analysis	*Starting material profile*
03	**Data receipt and quality assessment**	Sequence integrity and quality verified per assay and sample	md5sum, FastQC	*QC acceptance record*
04	**Transgene sequence verification**	The transgene sequence verified against the intended construct	Sequence verification	*Construct verification*
05	**Transgene copy number quantification**	Vector copy number per cell quantified against release specification	Copy number quantification	*Copy number result*
06	**Transduced proportion determination**	The proportion of transduced cells determined	Transduction analysis	*Transduced proportion*
07	**Single-cell data processing**	Per-cell transcriptome data processed with ambient and doublet correction	CellRanger, CellBender, scDblFinder	*Processed matrix*
08	**Transgene detection per cell**	Transgene expression detected per cell to link identity to phenotype	Transgene quantification	*Per-cell transgene status*
09	**Quality filtering**	Cells filtered on transcript count and quality metrics	QC filtering	*Filtered matrix*
10	**Normalisation and integration**	Data normalised and integrated across batches where multiple are compared	Normalisation, Harmony	*Integrated matrix*
11	**Cell type composition analysis**	Product composition determined by cell type and lineage	Annotation, composition analysis	*Composition profile*
12	**Differentiation state characterisation**	Memory and effector differentiation states characterised	State analysis	*Differentiation profile*
13	**Exhaustion phenotype assessment**	Exhaustion-associated programmes assessed in the product	Programme scoring	*Exhaustion assessment*
14	**Repertoire extraction**	T-cell receptor repertoire extracted from the product	Repertoire extraction	*Product repertoire*
15	**Clonal diversity quantification**	Repertoire diversity quantified as a product attribute	Diversity analysis	*Diversity metrics*
16	**Starting material comparability**	Product composition compared against its own starting material	Comparability analysis	*Comparability assessment*
17	**Batch comparability assessment**	Batches compared against baseline variation rather than a fixed specification	Comparability analysis	*Batch comparability*
18	**Post-infusion sample processing**	Post-infusion samples processed with the same pipeline	Processing pipeline	*Post-infusion data*
19	**Product-derived cell identification**	Infused cells distinguished from endogenous by transgene and clonal markers	Identification logic	*Product-derived cells*
20	**Persistence quantification**	Product-derived cell abundance quantified across timepoints	Quantification	*Persistence profile*
21	**Phenotype evolution analysis**	Phenotype and state change after infusion characterised	Evolution analysis	*Evolution profile*
22	**Clonal dynamics analysis**	Repertoire dynamics after infusion analysed for expansion and contraction	Dynamics analysis	*Clonal dynamics*
23	**Reporting**	Product characterisation, comparability and persistence assembled	Report template	*Product report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
23	**112 h · 19 d**	**SPB-23**	**2,55,000**

SPB-24

SECTOR 01

Vector Integration Site Analysis

(**Integration site analysis for gene therapy** · Integration site analysis)

WHY IT IS DONE

Integrating vectors insert semi-randomly, and insertion near a proto-oncogene can drive clonal expansion and malignancy. Integration site analysis is the safety assessment that follows every integrating gene therapy patient, and it is a regulatory requirement rather than an option.

WHAT IT ANSWERS

Where has the vector integrated, how many distinct clones are present, and is any clone expanding in a pattern that suggests insertional oncogenesis?

WHEN IT IS DONE

In preclinical vector safety assessment, in release testing of transduced products, and in long-term follow-up of gene therapy patients at defined intervals.

WHAT TRIGGERS IT

Preclinical vector characterisation, product release testing, or a scheduled patient follow-up timepoint.

HOW IT IS DONE

Vector-genome junctions are recovered by a method that preserves abundance information, junctions are mapped to the genome with repetitive-region ambiguity handled, and unique integration sites are defined with duplicate collapse. Clonal abundance is quantified, diversity indices computed, and clonal dynamics tracked across timepoints with enrichment near oncogenes tested statistically.

WHAT IT PRODUCES

Integration site coordinates with gene annotation, clonal abundance per site, diversity metrics across the population, longitudinal clonal dynamics, oncogene proximity enrichment analysis, and a safety assessment against expansion criteria.

WHAT IT DETECTS

Integration site coordinates with nearby genes, clonal abundance per site, clonal diversity of the transduced population, expansion of individual clones across timepoints, and enrichment near cancer-associated genes.

WHAT IT DOES NOT DETECT

It does not establish that an expanding clone is malignant, only that expansion is occurring. Sites in repetitive sequence are unmappable, sampling limits detection of small clones, and blood sampling does not represent tissue-resident populations.

WHAT MAKES IT HARD

Quantifying clonal abundance rather than merely listing sites is what makes the analysis a safety assessment, and it requires method choices that preserve abundance through library preparation. Distinguishing genuine clonal expansion from sampling variation across timepoints needs statistical treatment rather than visual comparison.

WHO DOES THIS JOB

Computational biologist in gene therapy development, senior level, working with regulatory and clinical safety.

WHAT YOU CAN DO AFTERWARDS

Recover and map integration sites with abundance preserved, quantify clonal dynamics longitudinally, and produce regulatory-grade safety assessments.

WHO HIRES FOR IT

Gene therapy developers, vector manufacturers, regulatory safety functions, and academic gene therapy programmes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
22	**108 hours · 18 days**	Prior structural variant and alignment analysis experience	**2,45,000**

SPB-24
WORKFLOW

Vector Integration Site Analysis

(**Integration site analysis for gene therapy** · Integration site analysis)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Study scope and regulatory context definition**	Whether the analysis is preclinical, release or follow-up defined with expectations	Protocol review	*Scope definition*
02	**Sample and timepoint recording**	Sample source, timepoint and cell fraction recorded	Sample record	*Sample record*
03	**Vector construct sequence registration**	The vector sequence registered as the junction detection reference	Sequence registry	*Registered construct*
04	**Input material quantification**	Input DNA quantified as it bounds the clones recoverable	Quantification	*Input record*
05	**Junction recovery method verification**	The recovery method confirmed to preserve clonal abundance information	Method verification	*Method record*
06	**Data receipt and quality assessment**	Sequence integrity and quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
07	**Vector sequence trimming**	Vector-derived sequence trimmed to leave genomic flanking sequence	Trimming	*Trimmed reads*
08	**Genomic alignment**	Flanking sequence aligned to the host genome	bwa-mem2	*Aligned reads*
09	**Repetitive region ambiguity handling**	Sites in repetitive sequence handled explicitly rather than discarded silently	Ambiguity handling	*Ambiguity record*
10	**Integration site definition**	Unique integration sites defined with position tolerance applied	Site definition	*Integration sites*
11	**Duplicate collapse**	Amplification duplicates collapsed so abundance reflects clones not reads	Duplicate collapse	*Collapsed sites*
12	**Clonal abundance quantification**	Abundance quantified per integration site as a clone proxy	Abundance quantification	*Clonal abundances*
13	**Sampling depth assessment**	Clones recoverable given input and depth assessed as a detection bound	Depth assessment	*Detection bound*
14	**Diversity metric computation**	Clonal diversity and evenness computed across the population	Diversity calculation	*Diversity metrics*
15	**Gene annotation of sites**	Integration sites annotated against nearby genes and regulatory features	Annotation	*Site annotation*
16	**Cancer gene proximity analysis**	Enrichment near cancer-associated genes tested statistically	Enrichment testing	*Proximity analysis*
17	**Genomic feature distribution analysis**	Site distribution across genomic features compared against expectation	Distribution analysis	*Distribution profile*
18	**Longitudinal clone tracking**	Individual clones tracked across timepoints	Tracking analysis	*Clone trajectories*
19	**Expansion significance testing**	Clonal expansion tested statistically against sampling variation	Statistical testing	*Expansion assessment*
20	**Dominant clone assessment**	Clones exceeding dominance thresholds identified and characterised	Dominance criteria	*Dominant clone findings*
21	**Safety assessment against criteria**	Findings assessed against the programme's defined safety criteria	Safety criteria	*Safety assessment*
22	**Regulatory reporting**	Sites, abundance, dynamics and safety assessment assembled to expectations	Report template	*Integration site report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**108 h · 18 d**	**SPB-24**	**2,45,000**

SPB-25
SECTOR 01

On-Target Gene Editing Outcome Analysis

(**Editing outcome quantification** · On-target editing analysis)

WHY IT IS DONE

Editing efficiency is the primary readout of any genome editing programme, and the spectrum of outcomes at the target site determines whether the intended biological effect is achieved. Reporting a single efficiency figure conceals outcomes that matter clinically.

WHAT IT ANSWERS

What proportion of alleles were edited at the target site, what outcomes were produced, and how many are the intended edit rather than an unintended one?

WHEN IT IS DONE

In editing optimisation, in preclinical characterisation of a therapeutic edit, in product release testing, and in comparing guide and delivery conditions.

WHAT TRIGGERS IT

An editing experiment requiring quantification, guide optimisation, product release testing, or a delivery method comparison.

HOW IT IS DONE

Target amplification is designed to avoid allele dropout, reads are aligned with an indel-aware method against the reference and expected edited sequences, and outcomes are quantified allele by allele. Large deletion detection is addressed by design or explicitly declared as outside scope, and background from unedited controls is subtracted.

WHAT IT PRODUCES

Allele-level outcome distribution, editing efficiency with the outcome spectrum broken out, precise edit frequency where applicable, large deletion assessment or explicit scope limitation, and control background subtraction evidence.

WHAT IT DETECTS

Insertion and deletion spectra at the target site, precise edit frequency where homology-directed repair is intended, allele-level outcome distribution, and large deletions or rearrangements at the target where the method detects them.

WHAT IT DOES NOT DETECT

Standard amplicon methods miss large deletions and rearrangements that remove primer binding sites entirely, which systematically underestimates disruption. Chromosomal-scale consequences such as translocations and chromosome loss require different methods, and protein-level consequence is not measured.

WHAT MAKES IT HARD

Amplicon-based quantification is blind to the outcomes that remove the amplicon, so efficiency is systematically underestimated and the most disruptive outcomes are the ones missed. Distinguishing a genuine precise edit from unedited reference sequence requires the assay to resolve them unambiguously.

WHO DOES THIS JOB

Computational biologist in genome editing, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Quantify editing outcomes allele by allele, recognise and address amplicon blind spots, and report efficiency with the full outcome spectrum.

WHO HIRES FOR IT

Genome editing biotech, gene therapy developers, editing platform companies, and academic editing laboratories.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**98 hours · 17 days**	Prior amplicon and indel analysis experience	**2,25,000**

SPB-25
WORKFLOW

On-Target Gene Editing Outcome Analysis

(**Editing outcome quantification** · On-target editing analysis)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Editing design recording**	Guide sequence, nuclease, repair template and delivery recorded	Design record	*Design record*
02	**Amplicon design verification**	Amplicon designed to avoid primer sites affected by expected edits	Design verification	*Amplicon design*
03	**Allele dropout risk assessment**	Risk of dropout from variants or deletions under primers assessed	Risk assessment	*Dropout risk record*
04	**Control sample inclusion**	Unedited controls processed identically to establish background	Control preparation	*Control samples*
05	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC	*QC acceptance record*
06	**Primer trimming**	Primer sequences removed before outcome analysis	Trimming	*Trimmed reads*
07	**Reference sequence preparation**	Reference and intended edited sequences prepared for alignment	Sequence preparation	*Reference set*
08	**Indel-aware alignment**	Reads aligned with settings that resolve insertions and deletions accurately	Indel-aware aligner	*Aligned reads*
09	**Allele-level outcome classification**	Each read classified by its specific outcome at the target site	Outcome classification	*Outcome classifications*
10	**Indel spectrum quantification**	The distribution of insertion and deletion sizes quantified	Spectrum analysis	*Indel spectrum*
11	**Precise edit quantification**	Reads matching the intended edit quantified where a template was used	Precise edit analysis	*Precise edit frequency*
12	**Imperfect repair quantification**	Outcomes resembling but not matching the intended edit quantified separately	Outcome analysis	*Imperfect repair frequency*
13	**Control background subtraction**	Background outcomes from unedited controls subtracted	Background subtraction	*Corrected frequencies*
14	**Editing efficiency calculation**	Overall editing efficiency calculated with the outcome spectrum stated	Efficiency calculation	*Editing efficiency*
15	**Large deletion assessment**	Large deletions removing primer sites assessed or declared outside scope	Long-range analysis	*Large deletion assessment*
16	**Structural consequence assessment**	Chromosomal-scale consequences assessed where the method permits	Structural analysis	*Structural assessment*
17	**Amplicon blind spot documentation**	Outcomes invisible to the assay documented as a stated limitation	Limitation template	*Blind spot statement*
18	**Reproducibility assessment**	Outcome distributions compared across replicates	Reproducibility analysis	*Reproducibility metrics*
19	**Condition comparison**	Editing outcomes compared across guides or delivery conditions	Comparative analysis	*Condition comparison*
20	**Reporting**	Outcome spectrum, efficiency, limitations and comparisons assembled	Report template	*Editing outcome report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**98 h · 17 d**	**SPB-25**	**2,25,000**

SPB-26
SECTOR 01

Off-Target Editing Assessment Analysis

(**Genome-wide off-target editing assessment** · Off-target assessment)

WHY IT IS DONE

An editing therapeutic that modifies unintended sites carries genotoxic risk, and demonstrating that it does not is a regulatory requirement for clinical entry. It is the most methodologically contested area in editing development and the one regulators scrutinise most closely.

WHAT IT ANSWERS

Has this editing reagent modified any site other than the intended one, at what frequency, and is the evidence sufficient to support clinical use?

WHEN IT IS DONE

In preclinical characterisation before clinical entry, in guide selection between candidates, and in regulatory submission support.

WHAT TRIGGERS IT

Progression of an editing candidate toward clinical use, guide selection, or a regulatory submission requirement.

HOW IT IS DONE

Candidate sites are nominated by orthogonal methods combining computational prediction with empirical genome-wide detection, since neither alone is accepted. Nominated sites are verified by targeted deep sequencing at sensitivity sufficient for the risk threshold, structural consequences are assessed separately, and the relevant cell type is used because chromatin context governs accessibility.

WHAT IT PRODUCES

Candidate site nominations from each orthogonal method, verified editing frequency per site with detection limits stated, structural consequence assessment, cancer gene proximity analysis, and a regulatory evidence package with method limitations declared.

WHAT IT DETECTS

Candidate off-target sites from both computational prediction and empirical genome-wide methods, editing frequency at each candidate site verified by targeted deep sequencing, structural consequences including translocations, and sites in cancer-associated genes.

WHAT IT DOES NOT DETECT

No method detects all off-target editing, and each has characteristic blind spots, which is why orthogonal approaches are required rather than preferred. Very low frequency events fall below detection, and cell-type-specific chromatin context means results in one cell type do not transfer to another.

WHAT MAKES IT HARD

Sensitivity is the whole argument, and stating the detection limit at each verified site is what makes a negative meaningful. Relying on computational prediction alone is no longer acceptable, and results generated in a convenient cell type do not support claims about the therapeutic cell type.

WHO DOES THIS JOB

Senior computational biologist in genome editing safety, working with regulatory affairs. A specialist and heavily scrutinised role.

WHAT YOU CAN DO AFTERWARDS

Nominate off-target candidates by orthogonal methods, verify at declared sensitivity, assess structural consequences, and assemble regulatory evidence.

WHO HIRES FOR IT

Genome editing therapeutic developers, gene therapy companies, regulatory science functions, and editing safety consultancies.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
24	**118 hours · 20 days**	Prior editing and low-frequency variant analysis experience	**2,70,000**

SPB-26
WORKFLOW

Off-Target Editing Assessment Analysis

(**Genome-wide off-target editing assessment** · Off-target assessment)

COMPLETE WORKFLOW — 24 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Reagent and cell type definition**	The editing reagent and the therapeutically relevant cell type defined	Programme protocol	*Scope definition*
02	**Regulatory expectation review**	Applicable regulatory expectations for off-target evidence reviewed	Regulatory review	*Expectation record*
03	**Risk threshold definition**	The editing frequency threshold of concern defined to set required sensitivity	Risk specification	*Sensitivity requirement*
04	**Computational prediction**	Candidate sites predicted computationally with mismatch and bulge tolerance stated	Prediction tools	*Predicted candidates*
05	**Empirical nomination method selection**	A genome-wide empirical nomination method selected as an orthogonal approach	Method selection	*Selected method*
06	**Empirical nomination execution**	Genome-wide empirical nomination executed in a relevant system	Empirical method	*Empirically nominated sites*
07	**Method concordance assessment**	Overlap and divergence between nomination methods assessed	Concordance analysis	*Nomination concordance*
08	**Candidate site consolidation**	Candidates from all methods consolidated into a single verification list	Consolidation	*Verification list*
09	**Cell type relevance verification**	Verification performed in the therapeutically relevant cell type	Cell type selection	*Cell type record*
10	**Verification amplicon design**	Amplicons designed for every candidate site with dropout risk assessed	Design procedure	*Verification amplicons*
11	**Control sample preparation**	Unedited controls prepared for background estimation at every site	Control preparation	*Control samples*
12	**Targeted deep sequencing**	Every candidate site sequenced to the depth the risk threshold requires	Deep sequencing	*Verification data*
13	**Per-site depth verification**	Achieved depth verified per site against the required sensitivity	Depth analysis	*Depth record*
14	**Background error modelling**	Site-specific background error modelled from control samples	Error modelling	*Background models*
15	**Editing frequency quantification**	Editing frequency quantified per site against site-specific background	Frequency analysis	*Site frequencies*
16	**Per-site detection limit statement**	The detection limit stated for every site so negatives are interpretable	Limit calculation	*Per-site limits*
17	**Statistical significance assessment**	Frequencies assessed for significance against background	Statistical testing	*Significance results*
18	**Translocation assessment**	Translocations between on-target and off-target sites assessed	Structural analysis	*Translocation findings*
19	**Chromosomal abnormality assessment**	Larger chromosomal consequences assessed by an appropriate method	Chromosomal analysis	*Chromosomal findings*
20	**Cancer gene proximity analysis**	Confirmed off-target sites assessed for proximity to cancer-associated genes	Annotation analysis	*Proximity assessment*
21	**Donor variability assessment**	Off-target profiles compared across donors where genotype varies	Cross-donor analysis	*Donor variability*
22	**Method limitation documentation**	Blind spots of every method used documented explicitly	Limitation template	*Method limitations*
23	**Risk assessment**	Overall genotoxic risk assessed against the defined threshold	Risk assessment	*Risk determination*
24	**Regulatory package assembly**	Nomination, verification, limits and risk assembled into a submission package	Package assembly	*Regulatory package*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
24	**118 h · 20 d**	**SPB-26**	**2,70,000**

SPB-27
SECTOR 01

mRNA & Nucleic Acid Therapeutic Sequence QC Analysis

(**Nucleic acid therapeutic sequence quality analysis** · mRNA sequence QC)

WHY IT IS DONE

For a nucleic acid therapeutic the sequence is the product, so confirming identity and characterising impurities is a direct quality attribute rather than a research question. Truncated, fragmented and mis-incorporated species affect both potency and safety.

WHAT IT ANSWERS

Does this nucleic acid product match its intended sequence, what impurity species are present, and are the critical quality attributes within specification?

WHEN IT IS DONE

In batch release testing, in process development and comparability, in stability studies, and in investigating a potency deviation.

WHAT TRIGGERS IT

A manufacturing batch requiring release, a process change requiring comparability, a stability timepoint, or a potency investigation.

HOW IT IS DONE

Sequencing chemistry is selected to resolve full-length species rather than fragmenting them, reads are aligned to the intended construct, and identity is confirmed across the full length. Truncation and fragmentation are quantified from read length distributions, variants are called against sequencing error background, tail length is characterised, and residual template is quantified.

WHAT IT PRODUCES

Full-length identity confirmation, impurity species quantification by type, sequence variant analysis against error background, tail length distribution, residual template quantification, and specification comparison.

WHAT IT DETECTS

Full-length sequence identity against the intended construct, truncated and fragmented species with their proportions, sequence variants and mis-incorporations, poly-adenosine tail length distribution, and residual template or plasmid sequence.

WHAT IT DOES NOT DETECT

It does not measure secondary structure, modified nucleoside incorporation directly, lipid formulation attributes or biological potency, all of which require separate methods. Sequencing chemistry itself introduces artefacts that mimic product impurities and must be controlled.

WHAT MAKES IT HARD

Separating product impurity from sequencing artefact requires a well-characterised control and an explicit error model, because both produce the same signatures. Full-length characterisation needs read lengths that span the construct, and fragmenting approaches destroy exactly the information being sought.

WHO DOES THIS JOB

Computational biologist in analytical development or quality control, mid to senior level, working with analytical chemistry.

WHAT YOU CAN DO AFTERWARDS

Confirm nucleic acid product identity, quantify impurity species against error background, characterise tail length, and compare against specification.

WHO HIRES FOR IT

mRNA and nucleic acid therapeutic developers, contract manufacturers, analytical development functions, and quality control laboratories.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
19	**92 hours · 16 days**	Prior sequencing analysis and long-read basics	**2,10,000**

SPB-27
WORKFLOW

mRNA & Nucleic Acid Therapeutic Sequence QC Analysis

(**Nucleic acid therapeutic sequence quality analysis** · mRNA sequence QC)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Construct sequence registration**	The intended product sequence registered as the analytical reference	Sequence registry	*Registered construct*
02	**Critical quality attribute definition**	Sequence-related quality attributes and their specifications defined	Specification review	*Defined attributes*
03	**Sequencing method selection**	A method capable of resolving full-length species selected	Method selection	*Selected method*
04	**Control material preparation**	Reference material processed identically to establish method background	Control preparation	*Control data*
05	**Data receipt and quality assessment**	Sequence integrity and run quality verified against acceptance criteria	md5sum, run QC	*QC acceptance record*
06	**Read preprocessing**	Reads preprocessed with settings preserving length information	Preprocessing	*Processed reads*
07	**Alignment to intended construct**	Reads aligned to the registered construct sequence	Alignment	*Aligned reads*
08	**Full-length identity confirmation**	Identity confirmed across the entire construct length	Coverage and identity analysis	*Identity confirmation*
09	**Read length distribution analysis**	Length distribution analysed to quantify truncated and fragmented species	Length analysis	*Length distribution*
10	**Truncation quantification**	Truncated species quantified with their termination positions characterised	Truncation analysis	*Truncation profile*
11	**Fragmentation quantification**	Fragmented species quantified as a proportion of total	Fragmentation analysis	*Fragmentation profile*
12	**Sequence variant analysis**	Variants called against the construct with error background subtracted	Variant analysis	*Variant profile*
13	**Method error background characterisation**	Sequencing error background characterised from control material	Background analysis	*Error background*
14	**Poly-adenosine tail characterisation**	Tail length distribution characterised where applicable to the product	Tail analysis	*Tail profile*
15	**Residual template quantification**	Residual DNA template or plasmid sequence quantified	Template analysis	*Residual template*
16	**Impurity species classification**	All impurity species classified by type and quantified	Classification	*Impurity profile*
17	**Specification comparison**	Every measured attribute compared against its specification	Specification comparison	*Compliance assessment*
18	**Batch comparability assessment**	Results compared against prior batches for process consistency	Comparability analysis	*Comparability assessment*
19	**Quality control reporting**	Identity, impurity profile and specification compliance assembled	Report template	*Sequence QC report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SPB-27**	**2,10,000**

SPB-28
SECTOR 01

Viral Vector Genomic Quality Analysis

(**Viral vector genomic characterisation and QC** · Vector genomic QC)

WHY IT IS DONE

Viral vector preparations contain the intended genome alongside truncated forms, reverse packaged sequence, host and helper DNA and recombinants. Characterising this mixture is a release requirement and a direct determinant of both potency and safety.

WHAT IT ANSWERS

What genomic species are present in this vector preparation, in what proportions, and do the impurities fall within specification?

WHEN IT IS DONE

In vector batch release, in process development and comparability, in investigating potency variability, and in supporting regulatory submissions.

WHAT TRIGGERS IT

A vector batch requiring release, a process change, a potency deviation investigation, or a submission requirement.

HOW IT IS DONE

Nuclease treatment removes unencapsidated DNA before extraction so that packaged content is what is measured, then sequencing chemistry appropriate to full-length characterisation is applied. Reads are classified against vector, host and helper references, genome integrity is assessed across the full construct, terminal repeat regions are examined specifically, and impurity proportions are quantified against specification.

WHAT IT PRODUCES

Packaged content classification across vector, host and helper origin, full-length genome integrity assessment, terminal repeat region analysis, truncated and recombinant species quantification, and specification comparison.

WHAT IT DETECTS

Full-length vector genome integrity, truncated and incomplete genome species with proportions, reverse packaged and inverted terminal repeat aberrations, host cell and helper plasmid DNA packaging, and recombinant species.

WHAT IT DOES NOT DETECT

It does not measure infectious titre, transduction efficiency or empty-to-full capsid ratio, which require separate analytical methods. Sequencing of packaged material cannot distinguish encapsidated from surface-associated DNA without appropriate sample treatment.

WHAT MAKES IT HARD

Terminal repeat regions are structurally difficult and are exactly where aberrations that affect potency occur, so a method that cannot resolve them is not characterising the product. Failing to remove unencapsidated DNA before analysis produces impurity figures that describe the buffer rather than the vector.

WHO DOES THIS JOB

Computational biologist in vector analytical development, senior level, working with process development and quality control.

WHAT YOU CAN DO AFTERWARDS

Characterise packaged vector content, resolve terminal repeat regions, quantify impurity species, and support release and comparability decisions.

WHO HIRES FOR IT

Gene therapy developers, viral vector contract manufacturers, analytical development functions, and quality control laboratories.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
21	**102 hours · 17 days**	Prior sequencing analysis and assembly experience	**2,35,000**

SPB-28
WORKFLOW

Viral Vector Genomic Quality Analysis

(**Viral vector genomic characterisation and QC** · Vector genomic QC)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Vector construct registration**	Vector genome sequence and terminal repeat structure registered	Sequence registry	*Registered construct*
02	**Reference set assembly**	Vector, host cell and helper plasmid references assembled for classification	Reference assembly	*Reference set*
03	**Critical quality attribute definition**	Genomic quality attributes and specifications defined	Specification review	*Defined attributes*
04	**Nuclease treatment verification**	Unencapsidated DNA removal verified before extraction	Treatment verification	*Treatment record*
05	**Input quantification**	Extracted vector genome quantity measured	Quantification	*Input record*
06	**Sequencing method selection**	A method capable of resolving full-length genomes and terminal repeats selected	Method selection	*Selected method*
07	**Data receipt and quality assessment**	Sequence integrity and run quality verified against criteria	md5sum, run QC	*QC acceptance record*
08	**Read classification**	Reads classified against vector, host and helper references	Classification	*Classified reads*
09	**Packaging composition quantification**	Proportions of vector, host and helper sequence quantified	Composition analysis	*Packaging composition*
10	**Full-length genome assessment**	Genome integrity assessed across the entire construct	Coverage analysis	*Integrity assessment*
11	**Coverage uniformity analysis**	Coverage uniformity analysed to detect systematic packaging defects	Uniformity analysis	*Uniformity profile*
12	**Terminal repeat region analysis**	Terminal repeat structure and integrity analysed specifically	Specialised analysis	*Terminal repeat assessment*
13	**Truncated species quantification**	Incomplete genome species quantified with truncation positions mapped	Truncation analysis	*Truncation profile*
14	**Reverse packaged species assessment**	Reverse packaged sequence identified and quantified	Orientation analysis	*Reverse packaging findings*
15	**Recombinant species detection**	Recombinant and rearranged genome species detected	Recombination analysis	*Recombinant findings*
16	**Host cell DNA characterisation**	Packaged host sequence characterised by genomic origin and size	Host analysis	*Host DNA profile*
17	**Helper sequence quantification**	Helper plasmid sequence packaging quantified	Helper analysis	*Helper profile*
18	**Sequence variant analysis**	Variants within the vector genome called against error background	Variant analysis	*Variant profile*
19	**Specification comparison**	Measured attributes compared against release specifications	Specification comparison	*Compliance assessment*
20	**Batch comparability assessment**	Results compared against prior batches for process consistency	Comparability analysis	*Comparability assessment*
21	**Quality control reporting**	Composition, integrity, impurities and compliance assembled	Report template	*Vector QC report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SPB-28**	**2,35,000**

SPB-29
SECTOR 01

Cell Bank & Adventitious Agent Screening Analysis

(**Sequencing-based adventitious agent testing** · Adventitious agent screening)

WHY IT IS DONE

Biological manufacturing requires assurance that cell banks and raw materials are free of adventitious agents, and sequencing detects agents that culture and antibody methods cannot. Regulators now accept and increasingly expect sequencing-based testing alongside conventional assays.

WHAT IT ANSWERS

Does this cell bank or material contain any viral, bacterial, fungal or mycoplasmal agent, and does the evidence support release?

WHEN IT IS DONE

In master and working cell bank characterisation, in raw material and feedstock testing, in end-of-production cell testing, and in investigating a contamination event.

WHAT TRIGGERS IT

Cell bank establishment or qualification, raw material qualification, end-of-production testing, or a suspected contamination event.

HOW IT IS DONE

Host sequence is depleted or computationally subtracted, non-host reads are classified against comprehensive reference databases, and hits are assessed against negative and reagent controls processed identically. Detected agents are confirmed by targeted analysis, spike-in controls establish sensitivity, and unassigned sequence is assembled and characterised.

WHAT IT PRODUCES

Host depletion and classification results, agent detections with species assignment and abundance, control comparison establishing reagent background, spike-in derived sensitivity, unassigned sequence characterisation, and a release recommendation.

WHAT IT DETECTS

Viral, bacterial, fungal and mycoplasmal sequence at high sensitivity, endogenous retroviral element expression, integrated viral sequence, and species-level identification of any agent detected.

WHAT IT DOES NOT DETECT

It detects nucleic acid, not viable or infectious agent, so a positive requires follow-up to establish significance. Agents absent from reference databases cannot be identified, only flagged as unassigned, and sensitivity depends on the agent's abundance relative to host background.

WHAT MAKES IT HARD

Reagent and laboratory background contaminants are ubiquitous and will be detected, so controls processed identically are what separate a finding from background. Sensitivity must be demonstrated by spike-in rather than asserted, and unassigned sequence needs characterisation rather than dismissal.

WHO DOES THIS JOB

Computational biologist in biosafety or quality control, mid to senior level, working with virology and quality assurance.

WHAT YOU CAN DO AFTERWARDS

Run sequencing-based adventitious agent screening with proper controls, demonstrate sensitivity by spike-in, and characterise unassigned sequence.

WHO HIRES FOR IT

Biologics manufacturers, cell and gene therapy companies, contract testing laboratories, and biosafety testing providers.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**96 hours · 16 days**	Prior metagenomic analysis experience	**2,20,000**

SPB-29
WORKFLOW

Cell Bank & Adventitious Agent Screening Analysis

(**Sequencing-based adventitious agent testing** · Adventitious agent screening)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Testing scope definition**	Material type, agents in scope and regulatory expectations defined	Protocol review	*Scope definition*
02	**Reference database assembly**	Comprehensive reference databases assembled with versions recorded	Database assembly	*Reference databases*
03	**Control design**	Negative, reagent and spike-in controls designed into the run	Control design	*Control plan*
04	**Spike-in control preparation**	Spike-in organisms prepared at defined levels to establish sensitivity	Spike preparation	*Spike controls*
05	**Sample and control processing**	Samples and controls extracted and prepared identically	Processing	*Prepared libraries*
06	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample and control	md5sum, FastQC	*QC acceptance record*
07	**Host sequence depletion**	Host reads removed computationally or by wet-lab depletion	Host depletion	*Non-host reads*
08	**Depletion efficiency assessment**	Host removal efficiency quantified as it determines effective sensitivity	Efficiency analysis	*Depletion metrics*
09	**Taxonomic classification**	Non-host reads classified against the reference databases	Kraken2, classification	*Classification results*
10	**Reagent background subtraction**	Taxa present in reagent controls treated as background rather than findings	Background analysis	*Background-corrected results*
11	**Negative control comparison**	Sample results compared against negative controls processed in parallel	Control comparison	*Control comparison record*
12	**Spike-in recovery assessment**	Spike-in organisms quantified to establish achieved sensitivity	Recovery analysis	*Sensitivity determination*
13	**Detection threshold application**	Detections assessed against the abundance threshold for reporting	Threshold criteria	*Threshold decisions*
14	**Targeted confirmation**	Detected agents confirmed by targeted analysis or assembly	Confirmation analysis	*Confirmation results*
15	**Endogenous retroviral assessment**	Endogenous retroviral element expression assessed and distinguished from exogenous	Element analysis	*Endogenous assessment*
16	**Integrated viral sequence assessment**	Integrated viral sequence in the host genome identified and characterised	Integration analysis	*Integration findings*
17	**Unassigned sequence assembly**	Unclassified sequence assembled and characterised rather than dismissed	Assembly, characterisation	*Unassigned characterisation*
18	**Viability limitation documentation**	The distinction between nucleic acid detection and viable agent documented	Limitation template	*Viability caveat*
19	**Risk assessment**	Detections assessed for significance against the material's intended use	Risk assessment	*Risk determination*
20	**Release recommendation reporting**	Findings, sensitivity, controls and recommendation assembled	Report template	*Adventitious agent report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**96 h · 16 d**	**SPB-29**	**2,20,000**

SPB-30
SECTOR 01

Microbiome–Drug Interaction Analysis

(**Pharmacomicrobiomics analysis** · Pharmacomicrobiomics)

WHY IT IS DONE

Gut bacteria metabolise drugs, modify their bioavailability and influence immune-mediated response, which explains a meaningful share of the variability that pharmacokinetics alone cannot. It has become a routine correlative arm in immuno-oncology and metabolic programmes.

WHAT IT ANSWERS

Does the microbiome composition or functional capacity of these subjects relate to drug exposure, response or toxicity?

WHEN IT IS DONE

In correlative arms of clinical trials, in preclinical models investigating variability, and in investigating unexplained exposure or response heterogeneity.

WHAT TRIGGERS IT

A correlative trial arm, unexplained response or exposure variability, or a hypothesis about microbial drug metabolism.

HOW IT IS DONE

Study design and confounding are assessed with diet, antibiotics and geography treated as primary confounders, then composition is profiled taxonomically and functionally with compositional data methods used throughout. Associations are tested with confounders modelled, longitudinal change is assessed, and drug-metabolising capacity is examined specifically.

WHAT IT PRODUCES

Taxonomic and functional profiles with diversity metrics, confounder-adjusted association results, longitudinal compositional change, drug-metabolising gene capacity analysis, and interpretation bounded by causality limitations.

WHAT IT DETECTS

Taxonomic composition and diversity, functional gene content including drug-metabolising enzyme families, associations between composition and drug outcome, and compositional change over the course of treatment.

WHAT IT DOES NOT DETECT

It cannot establish causation from association, and microbiome composition is confounded by diet, geography, medication and disease itself to a degree that defeats naive comparison. Functional capacity from gene content does not establish that the function is active in vivo.

WHAT MAKES IT HARD

Microbiome data are compositional, so standard statistical methods produce spurious associations unless compositional approaches are used. Antibiotic exposure is the strongest determinant of composition and frequently correlates with disease severity, which confounds almost every association unless explicitly modelled.

WHO DOES THIS JOB

Computational biologist with microbiome competence in translational research, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Analyse microbiome data with compositional methods, control the dominant confounders, and interpret drug-microbiome associations with appropriate causal caution.

WHO HIRES FOR IT

Pharmaceutical translational medicine, microbiome therapeutic companies, immuno-oncology programmes, and academic pharmacomicrobiomics groups.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
19	**92 hours · 16 days**	Prior metagenomic analysis experience	**2,10,000**

SPB-30
WORKFLOW

Microbiome–Drug Interaction Analysis

(**Pharmacomicrobiomics analysis** · Pharmacomicrobiomics)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Study design and confounder specification**	Design reviewed with diet, antibiotics and geography specified as primary confounders	Design review	*Design assessment*
02	**Sample collection protocol recording**	Collection, storage and transport protocol recorded as composition determinants	Protocol record	*Collection record*
03	**Metadata assembly**	Diet, medication, antibiotic exposure and clinical variables assembled	Metadata compilation	*Covariate matrix*
04	**Control sample inclusion**	Negative and reagent controls processed alongside samples	Control processing	*Control data*
05	**Data receipt and quality assessment**	Sequence integrity and quality verified per sample	md5sum, FastQC	*QC acceptance record*
06	**Read preprocessing**	Adapters and low-quality sequence removed with retention accounting	fastp	*Trimmed reads*
07	**Host read removal**	Host-derived reads removed before microbial analysis	Host depletion	*Non-host reads*
08	**Taxonomic profiling**	Community composition profiled to species level where data permit	MetaPhlAn, Kraken2	*Taxonomic profiles*
09	**Functional profiling**	Functional gene content profiled including drug-metabolising families	HUMAnN, functional profiling	*Functional profiles*
10	**Contamination assessment**	Reagent contaminants identified against controls and removed	Decontamination	*Cleaned profiles*
11	**Compositional transformation**	Data transformed with compositional methods before any statistical analysis	Compositional transform	*Transformed data*
12	**Diversity computation**	Alpha and beta diversity computed on transformed data	Diversity analysis	*Diversity metrics*
13	**Confounder association assessment**	Diet, antibiotics and geography tested for association with composition	Association testing	*Confounder assessment*
14	**Confounder-adjusted association testing**	Drug outcome associations tested with confounders modelled explicitly	Mixed models	*Adjusted associations*
15	**Multiplicity control**	Significance adjusted across taxa and functions tested	FDR control	*Adjusted results*
16	**Longitudinal change analysis**	Compositional change across treatment analysed within subjects	Longitudinal modelling	*Change analysis*
17	**Drug-metabolising capacity analysis**	Genes encoding relevant metabolising enzymes analysed specifically	Targeted functional analysis	*Metabolising capacity*
18	**Causality limitation documentation**	The inability to infer causation from association documented explicitly	Limitation template	*Causality caveat*
19	**Reporting**	Profiles, adjusted associations, longitudinal change and caveats assembled	Report template	*Pharmacomicrobiomics report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SPB-30**	**2,10,000**

SPB-31
SECTOR 01

Multi-Omic Integration Analysis

(**Integrated multi-omic analysis** · Multi-omic integration)

WHY IT IS DONE

Genomic, transcriptomic, proteomic and clinical data each describe part of a system, and the mechanism usually sits in the relationship between them rather than in any one layer. Integration is where a programme moves from correlation within a layer to a mechanistic account across layers.

WHAT IT ANSWERS

What structure emerges when these data layers are analysed jointly, and what mechanistic or predictive insight does that yield beyond any single layer?

WHEN IT IS DONE

In mechanism studies where several data types exist, in biomarker development combining modalities, and in cohort studies with multi-platform profiling.

WHAT TRIGGERS IT

Availability of multiple aligned data layers on the same samples, a mechanism question spanning layers, or a multi-modal biomarker objective.

HOW IT IS DONE

Each layer is quality controlled and batch corrected independently before any integration, sample alignment across layers is verified, and layers are scaled appropriately given their different distributions. Integration is performed with the method matched to the question, shared and specific variation are separated, and integrated findings are validated against single-layer results.

WHAT IT PRODUCES

Per-layer quality and batch correction records, sample alignment verification, shared and layer-specific variation decomposition, integrated groupings with driving features, and comparison against single-layer analyses.

WHAT IT DETECTS

Shared and layer-specific variation structure, associations between layers, integrated sample groupings, features driving cross-layer structure, and predictive performance of combined against single-layer models.

WHAT IT DOES NOT DETECT

Integration does not create information that is not present, and combining underpowered layers produces a confidently wrong integrated result. Batch structure differs between layers and integrating without correcting each independently propagates artefact into the joint model.

WHAT MAKES IT HARD

Layer-specific batch structure is the dominant failure, because integration methods will happily find shared structure that is entirely technical. Sample misalignment between layers is common in real projects and produces integrated results that are meaningless rather than merely weak.

WHO DOES THIS JOB

Senior computational biologist or data scientist in translational research, with statistical depth across modalities.

WHAT YOU CAN DO AFTERWARDS

Quality control and correct each layer independently, verify sample alignment, integrate appropriately, and validate integrated findings against single-layer results.

WHO HIRES FOR IT

Pharmaceutical translational and systems biology groups, precision medicine programmes, biotech with multi-modal platforms, and academic systems units.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
22	**106 hours · 18 days**	Prior multi-platform omics analysis experience	**2,45,000**

SPB-31
WORKFLOW

Multi-Omic Integration Analysis

(**Integrated multi-omic analysis** · Multi-omic integration)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Integration question specification**	The question requiring integration specified, since it determines the method	Question specification	*Specified question*
02	**Layer inventory**	Available data layers inventoried with platform and processing recorded	Inventory	*Layer inventory*
03	**Sample alignment verification**	Sample correspondence across layers verified at identity level	Identity verification	*Alignment verification*
04	**Per-layer quality assessment**	Each layer assessed independently against its own quality criteria	Layer QC	*Per-layer QC records*
05	**Per-layer batch assessment**	Batch structure assessed separately within each layer	Batch analysis	*Per-layer batch records*
06	**Per-layer batch correction**	Batch effects corrected independently within each layer before integration	Batch correction	*Corrected layers*
07	**Missing data assessment**	Missingness patterns assessed per layer and across layers	Missingness analysis	*Missingness record*
08	**Feature filtering**	Uninformative and unreliable features removed within each layer	Filtering	*Filtered layers*
09	**Scaling and transformation**	Layers transformed and scaled to comparable distributions	Transformation	*Scaled layers*
10	**Layer-specific structure characterisation**	Dominant variation within each layer characterised before integration	Per-layer analysis	*Layer structure*
11	**Integration method selection**	The integration method selected against the specified question and layer types	Method selection	*Selected method*
12	**Integration execution**	Layers integrated with parameters recorded	Integration method	*Integrated representation*
13	**Shared and specific variance decomposition**	Variation decomposed into shared and layer-specific components	Variance decomposition	*Variance decomposition*
14	**Technical structure assessment**	Integrated structure tested for technical rather than biological drivers	Technical assessment	*Technical structure record*
15	**Integrated grouping identification**	Sample groupings emerging from integration identified and characterised	Clustering	*Integrated groupings*
16	**Driving feature extraction**	Features driving integrated structure extracted per layer	Feature analysis	*Driving features*
17	**Cross-layer association analysis**	Associations between layers characterised at feature level	Association analysis	*Cross-layer associations*
18	**Single-layer comparison**	Integrated findings compared against single-layer analyses	Comparative analysis	*Comparison record*
19	**Added value assessment**	Whether integration added information beyond single layers assessed explicitly	Value assessment	*Added value determination*
20	**Predictive performance comparison**	Combined and single-layer predictive performance compared where applicable	Performance comparison	*Performance comparison*
21	**Stability assessment**	Integrated structure assessed for stability across resampling	Stability analysis	*Stability record*
22	**Reporting**	Integration, decomposition, findings and added value assembled	Report template	*Integration report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**106 h · 18 d**	**SPB-31**	**2,45,000**

SPB-32
SECTOR 01

Real-World & Cohort Genomic Evidence Analysis

(**Real-world genomic evidence analysis** · RWE genomic analysis)

WHY IT IS DONE

Regulators and payers increasingly accept evidence from routine clinical data linked to genomic profiles, which is faster and cheaper than a trial and covers populations trials exclude. Generating it credibly requires methods that trials do not need, because the data were never designed for analysis.

WHAT IT ANSWERS

What does linked real-world genomic and clinical data show about treatment patterns, outcomes and biomarker-outcome relationships in routine practice?

WHEN IT IS DONE

In post-approval evidence generation, in payer and health technology assessment support, in rare subgroup analysis where trials are infeasible, and in external control arm construction.

WHAT TRIGGERS IT

A post-approval evidence requirement, a payer evidence request, a rare subgroup question, or an external control arm need.

HOW IT IS DONE

Data provenance and completeness are assessed, cohorts are defined with explicit inclusion criteria and index dates, and confounding by indication is addressed through design rather than adjustment alone. Missing data mechanisms are examined, outcome definitions are validated, sensitivity analyses probe key assumptions, and comparability against trial populations is assessed explicitly.

WHAT IT PRODUCES

Data provenance and completeness assessment, cohort definitions with attrition accounting, confounding control design, missing data mechanism analysis, outcome results with sensitivity analyses, and trial population comparability assessment.

WHAT IT DETECTS

Biomarker prevalence in routine populations, treatment patterns by genomic subgroup, outcome associations with confounding addressed, and comparability between real-world and trial populations.

WHAT IT DOES NOT DETECT

It cannot establish causal treatment effect without design features that address confounding by indication, which is severe in routine data. Missing data are informative rather than random, testing is not performed at random, and unmeasured confounding cannot be excluded by any statistical method.

WHAT MAKES IT HARD

Confounding by indication is the central problem and cannot be solved by adjusting for observed variables, since the reason a patient received a treatment is usually the thing that predicts their outcome. Selective testing means genomically profiled patients differ systematically from those who were not.

WHO DOES THIS JOB

Computational epidemiologist or biostatistician with genomic competence, senior level, in evidence generation functions.

WHAT YOU CAN DO AFTERWARDS

Construct cohorts from routine data, address confounding by design, analyse missing data mechanisms, and assess comparability against trial populations.

WHO HIRES FOR IT

Pharmaceutical real-world evidence groups, health economics and outcomes research functions, payers, and evidence generation consultancies.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**98 hours · 17 days**	Prior cohort and clinical data analysis experience	**2,25,000**

SPB-32
WORKFLOW

Real-World & Cohort Genomic Evidence Analysis

(**Real-world genomic evidence analysis** · RWE genomic analysis)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Research question and estimand specification**	The question and the target estimand specified before data are examined	Protocol specification	*Specified estimand*
02	**Data source assessment**	Provenance, coverage and known limitations of each data source assessed	Source review	*Source assessment*
03	**Linkage verification**	Linkage between genomic and clinical records verified at patient level	Linkage verification	*Linkage record*
04	**Cohort definition**	Inclusion, exclusion and index date defined explicitly and applied	Cohort definition	*Defined cohort*
05	**Attrition documentation**	Patient attrition through each criterion documented	Attrition analysis	*Attrition record*
06	**Selective testing assessment**	Differences between tested and untested patients assessed as a selection bias	Selection analysis	*Selection assessment*
07	**Exposure definition**	Treatment exposure defined with timing and duration rules explicit	Exposure definition	*Defined exposure*
08	**Outcome definition and validation**	Outcomes defined and validated against source records where possible	Outcome validation	*Validated outcomes*
09	**Covariate assembly**	Confounders assembled from the pre-index period	Covariate assembly	*Covariate matrix*
10	**Missing data mechanism analysis**	Missingness examined as informative rather than assumed random	Missingness analysis	*Missingness assessment*
11	**Confounding by indication assessment**	The reasons treatment was given assessed as the dominant confounder	Confounding analysis	*Indication confounding assessment*
12	**Design-based confounding control**	Confounding addressed through design such as matching or new-user design	Design methods	*Controlled design*
13	**Balance assessment**	Covariate balance after design control assessed and reported	Balance diagnostics	*Balance assessment*
14	**Genomic subgroup definition**	Biomarker-defined subgroups defined with assay heterogeneity accounted for	Subgroup definition	*Defined subgroups*
15	**Outcome analysis**	Outcomes analysed within the controlled design	Outcome modelling	*Outcome results*
16	**Sensitivity analysis**	Key assumptions probed through pre-specified sensitivity analyses	Sensitivity analysis	*Sensitivity results*
17	**Unmeasured confounding assessment**	Robustness to unmeasured confounding quantified	Quantitative bias analysis	*Bias assessment*
18	**Trial population comparability**	The real-world population compared against trial populations	Comparability analysis	*Comparability assessment*
19	**Causal limitation documentation**	Limits on causal interpretation documented explicitly	Limitation template	*Causal caveat*
20	**Reporting**	Cohort, design, results, sensitivity and comparability assembled	Report template	*Evidence report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**98 h · 17 d**	**SPB-32**	**2,25,000**

SPB-33
SECTOR 01

Regulatory Bioinformatics Submission Analysis

(**Bioinformatics regulatory submission preparation** · Regulatory submission analysis)

WHY IT IS DONE

Genomic evidence supporting a submission must be reproducible, traceable and defensible to a reviewer who was not present when it was generated. Programmes are delayed far more often by inadequate documentation of correct analysis than by incorrect analysis.

WHAT IT ANSWERS

Is the genomic evidence in this submission reproducible, traceable from raw data to conclusion, and documented to the standard a regulatory reviewer requires?

WHEN IT IS DONE

In preparing regulatory submissions containing genomic evidence, in responding to regulatory questions, and in preparing for inspection of computational systems.

WHAT TRIGGERS IT

A submission containing genomic evidence, a regulatory question on analytical methods, or an inspection readiness exercise.

HOW IT IS DONE

The analytical chain is reconstructed from raw data to reported result with every transformation identified, environments and versions are captured, and independent re-execution is attempted to confirm reproducibility. Analytical decisions are documented with their rationale, data provenance is traced, and the evidence package is assembled against the applicable regulatory expectations.

WHAT IT PRODUCES

Reconstructed analytical chain with all transformations, environment and version capture, independent re-execution results, documented decision rationale, data provenance trace, and a submission-ready evidence package.

WHAT IT DETECTS

Gaps in reproducibility, traceability breaks between raw data and reported results, undocumented analytical decisions, version control failures, and inconsistencies between reported and reproducible results.

WHAT IT DOES NOT DETECT

It does not assess whether the science is correct, only whether it is documented and reproducible. A perfectly documented flawed analysis passes this assessment, and scientific review is a separate exercise.

WHAT MAKES IT HARD

Retrospective reconstruction is far harder than contemporaneous documentation, and the analytical decisions that most need justification are the ones nobody recorded at the time. Re-execution frequently fails to reproduce reported numbers exactly, and explaining small discrepancies is more difficult than preventing them.

WHO DOES THIS JOB

Bioinformatician or computational scientist in regulatory affairs or quality, senior level. A role that is expanding rapidly and is difficult to recruit for.

WHAT YOU CAN DO AFTERWARDS

Reconstruct and document analytical chains, capture environments, demonstrate reproducibility, and assemble regulatory-grade evidence packages.

WHO HIRES FOR IT

Pharmaceutical regulatory affairs, diagnostic regulatory functions, gene and cell therapy developers, and regulatory consultancies.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**88 hours · 15 days**	Prior pipeline and validation experience	**2,05,000**

SPB-33
WORKFLOW

Regulatory Bioinformatics Submission Analysis

(**Bioinformatics regulatory submission preparation** · Regulatory submission analysis)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Submission scope definition**	The genomic evidence within the submission scope identified	Submission review	*Scope definition*
02	**Regulatory expectation review**	Applicable expectations for computational evidence reviewed	Guidance review	*Expectation record*
03	**Analytical chain reconstruction**	Every step from raw data to reported result identified and ordered	Chain reconstruction	*Analytical chain*
04	**Raw data provenance tracing**	Raw data traced to source with custody and integrity established	Provenance tracing	*Provenance record*
05	**Software inventory**	Every tool, version and dependency in the chain inventoried	Inventory compilation	*Software inventory*
06	**Environment capture**	Computational environments captured in reproducible form	Environment capture	*Environment record*
07	**Parameter documentation**	Every non-default parameter documented with its rationale	Parameter documentation	*Parameter record*
08	**Analytical decision documentation**	Decisions such as filtering and exclusion documented with rationale	Decision documentation	*Decision record*
09	**Reference and resource version capture**	Reference genomes, annotations and databases captured with versions	Version capture	*Resource record*
10	**Independent re-execution**	The chain re-executed independently to test reproducibility	Re-execution	*Re-execution results*
11	**Result concordance assessment**	Re-executed results compared against reported results	Concordance analysis	*Concordance assessment*
12	**Discrepancy investigation**	Any discrepancy investigated and explained rather than accepted	Investigation procedure	*Discrepancy record*
13	**Pipeline validation evidence assembly**	Validation evidence for the analytical pipeline assembled	Validation records	*Validation evidence*
14	**Change control evidence assembly**	Version changes during the programme documented with impact assessed	Change control records	*Change control evidence*
15	**Traceability matrix construction**	A matrix linking every reported result to its source data and code built	Traceability construction	*Traceability matrix*
16	**Data integrity assessment**	Data integrity controls assessed against expectations	Integrity assessment	*Integrity record*
17	**Gap remediation**	Identified documentation gaps remediated before submission	Remediation procedure	*Remediation record*
18	**Submission package assembly**	All evidence assembled into a reviewer-ready package	Package assembly	*Submission package*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**88 h · 15 d**	**SPB-33**	**2,05,000**

Build and staff a *discovery informatics function*

Each of the thirty-three programmes in this catalogue trains one complete type of NGS data analysis as a pharmaceutical or biotechnology research function performs it — not a tool, not a step, but the whole analysis from the data that arrives to the report that is signed. Trainees run the pipeline on live clinical data, break it, diagnose the failure, and produce the deliverable report set a laboratory would place in a case file and defend to an assessor. Programmes are named the way laboratories name their own work, so a trainee returns to the bench able to run the analysis on the requisition rather than a generic version of it. Individual programmes may be taken alone, combined into a track, or delivered in-house at the client laboratory against that laboratory's own assays and gene lists.

CATALOGUE

SPB · Sector 04 of 16

SCOPE

33 analysis types · 668 steps

DURATION

3240 contact hours · 551 working days

COMPLETE CATALOGUE FEE

INR 48,39,000 (list 74,45,000)

15%	22%	28%	35%
ANY THREE PROGRAMMES	CORE TRACK	PROFESSIONAL TRACK	COMPLETE CATALOGUE