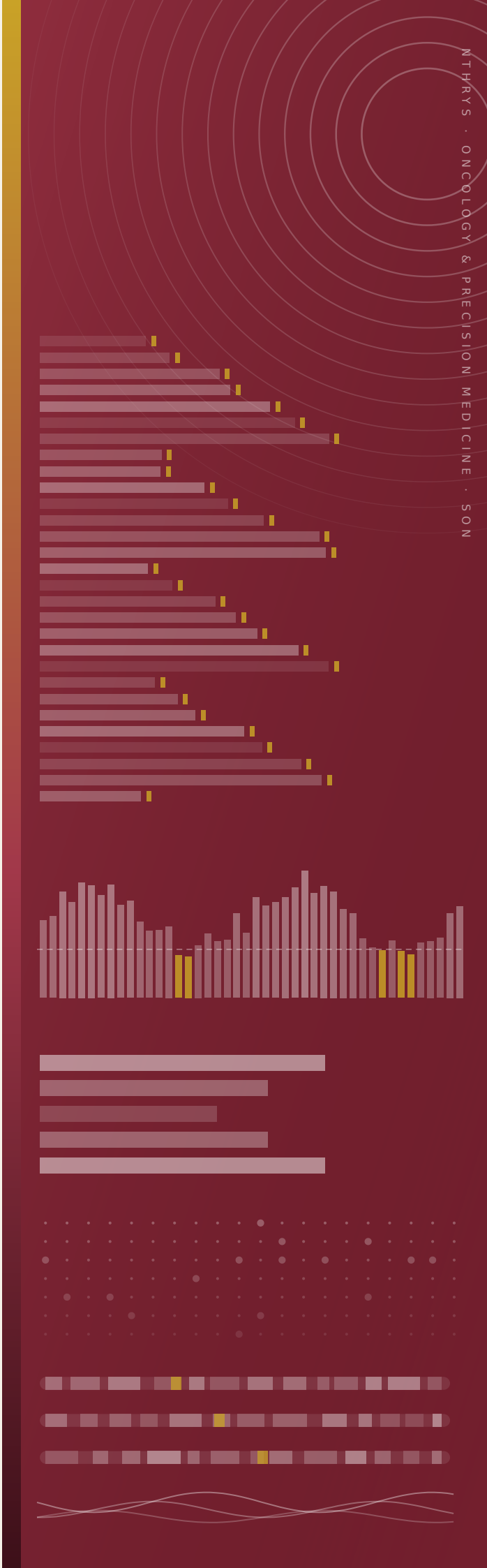
NTHRYS

OPC PVT LTD · HYDERABAD · INDIA

SECTOR CATALOGUE 02 · NGS DATA ANALYSIS

Oncology & *Precision Medicine*

Thirty-nine industrial training programmes, one for each distinct type of NGS data analysis performed in cancer centres, molecular pathology laboratories and precision oncology programmes. 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.



39	771	3750	636
ANALYSIS TYPES	STEPS	HOURS	DAYS

CATALOGUE SON · 2026 EDITION

INDUSTRIAL TRAINING · MOLECULAR ONCOLOGY PRACTICE

Oncology & Precision Medicine — *Analysis Types*

Thirty-nine distinct types of NGS data analysis performed in molecular oncology, 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)
SON-01	**Tumour–Normal Paired Somatic Analysis** Paired somatic variant analysis · T/N analysis · paired somatic	26	128	22	2,90,000
SON-02	**Tumour-Only Somatic Analysis** Tumour-only variant analysis · Tumour-only · unpaired somatic	24	118	20	2,70,000
SON-03	**Solid Tumour Panel Analysis** Targeted oncology panel analysis · Oncology panel · hotspot panel	20	96	16	2,20,000
SON-04	**Comprehensive Genomic Profiling** Comprehensive genomic profiling analysis · CGP	28	138	23	2,98,000
SON-05	**Somatic Whole Exome Analysis** Tumour whole exome analysis · Tumour WES	24	116	20	2,65,000
SON-06	**Somatic Whole Genome Analysis** Tumour whole genome analysis · Tumour WGS	27	132	22	2,95,000
SON-07	**Somatic Copy Number Analysis** Somatic CNA analysis · Somatic CNV · CNA analysis	20	98	17	2,25,000
SON-08	**Tumour Purity & Ploidy Analysis** Purity, ploidy and allele-specific copy number analysis · Purity–ploidy fitting	18	88	15	2,05,000
SON-09	**Somatic Structural Variant Analysis** Somatic structural rearrangement analysis · Somatic SV	22	108	18	2,45,000
SON-10	**Loss of Heterozygosity Analysis** Somatic LOH and allelic imbalance analysis · LOH · HLA LOH	16	76	13	1,75,000
SON-11	**Mutational Signature Analysis** Somatic mutational signature analysis · Signature analysis · SBS analysis	17	82	14	1,90,000
SON-12	**Tumour Mutational Burden Analysis** TMB determination · TMB	16	78	13	1,80,000
SON-13	**Microsatellite Instability Analysis** MSI status determination · MSI	15	72	12	1,65,000
SON-14	**Homologous Recombination Deficiency Analysis** HRD and genomic scar analysis · HRD · genomic scar	18	88	15	2,05,000
SON-15	**Focal Amplification & Extrachromosomal DNA Analysis** ecDNA and focal amplification analysis · ecDNA analysis	17	84	14	1,95,000
SON-16	**Subclonal Reconstruction & Tumour Evolution Analysis** Clonality and cancer cell fraction analysis · Clonality analysis · CCF analysis	20	98	17	2,25,000
SON-17	**Clonal Haematopoiesis Analysis** CHIP detection and interpretation · CHIP analysis	16	78	13	1,80,000
SON-18	**Circulating Tumour DNA Analysis** ctDNA and liquid biopsy analysis · ctDNA analysis · liquid biopsy	25	122	21	2,80,000
SON-19	**Tumour-Informed MRD Analysis** Bespoke panel minimal residual disease analysis · MRD · tumour-informed MRD	26	128	22	2,90,000
SON-20	**Tumour-Naive & Methylation MRD Analysis** Fixed-panel residual disease analysis · Fixed-panel MRD	22	106	18	2,45,000

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

Analysis Types — *21 to 39*

CODE	ANALYSIS TYPE · INDUSTRY NAME & SHORT NAME	STEPS	HOURS	DAYS	FEE (INR)
SON-21	**cfDNA Fragmentomics Analysis** Cell-free DNA fragmentation analysis · Fragmentomics	18	88	15	2,05,000
SON-22	**cfDNA Methylation & Tissue-of-Origin Analysis** Plasma methylation and tissue-of-origin analysis · MCED · TOO analysis	20	98	17	2,25,000
SON-23	**Resistance Mutation Monitoring Analysis** Serial resistance tracking analysis · Resistance monitoring	18	86	15	2,00,000
SON-24	**RNA Fusion Analysis** RNA-based gene fusion detection · Fusion analysis · RNA fusion panel	19	92	16	2,10,000
SON-25	**Tumour Transcriptome & Molecular Subtyping Analysis** Tumour RNA-Seq and expression subtyping · Tumour RNA-Seq · expression subtyping	20	96	16	2,20,000
SON-26	**Oncogenic Splice Variant Analysis** Splicing alteration and exon skipping analysis · Exon skipping analysis	16	78	13	1,80,000
SON-27	**Myeloid & Haematological Malignancy Panel Analysis** Myeloid and haematology panel analysis · Myeloid panel · haem panel	23	112	19	2,55,000
SON-28	**Lymphoid Clonality Analysis** IG and TR clonality assessment · Clonality assay · IG/TR clonality	17	82	14	1,90,000
SON-29	**Post-Transplant Chimerism Analysis** Donor chimerism monitoring · Chimerism monitoring	15	72	12	1,65,000
SON-30	**Hereditary Cancer Predisposition Analysis** Hereditary cancer panel analysis · Hereditary cancer panel	21	102	17	2,35,000
SON-31	**Germline–Somatic Integration Analysis** Paired germline and somatic reporting · Paired germline reporting	18	86	15	2,00,000
SON-32	**Neoantigen Prediction & Ranking Analysis** Neoantigen identification and prioritisation · Neoantigen analysis	21	102	17	2,35,000
SON-33	**Immune Repertoire & Immunotherapy Biomarker Analysis** TCR repertoire and immune biomarker analysis · TCR profiling · IO biomarkers	19	92	16	2,10,000
SON-34	**Tumour Methylation Classification Analysis** Methylation-based tumour classification · Methylation classifier · CNS classifier	18	88	15	2,05,000
SON-35	**Oncogenic Viral Integration Analysis** Viral integration and oncogenic virus analysis · HPV/HBV/EBV integration analysis	16	78	13	1,80,000
SON-36	**Tissue-of-Origin Analysis for Cancer of Unknown Primary** Cancer of unknown primary tissue-of-origin analysis · CUP analysis	17	82	14	1,90,000
SON-37	**Single-Cell Tumour Analysis** Single-cell tumour profiling analysis · scRNA tumour analysis	22	108	18	2,45,000
SON-38	**Spatial Tumour Profiling Analysis** Spatial transcriptomics analysis in oncology · Spatial oncology analysis	21	102	17	2,35,000
SON-39	**Oncology Pharmacogenomic Analysis** Chemotherapy pharmacogenomic analysis · DPYD, UGT1A1, TPMT analysis	15	72	12	1,65,000
Complete Sector Catalogue 02 — all thirty-nine analysis types		**771**	**3750**	**636**	**85,93,000**

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.

SON-01
SECTOR 01

Tumour–Normal Paired Somatic Analysis

(**Paired somatic variant analysis** · T/N analysis · paired somatic)

WHY IT IS DONE

Comparing tumour against the patient's own normal tissue is the only way to separate acquired mutations from inherited variation with confidence. Every other somatic approach is an approximation of this one, and it remains the reference standard against which tumour-only methods are judged.

WHAT IT ANSWERS

Which mutations were acquired by this tumour rather than inherited, at what allele fraction, and which of them are clinically actionable?

WHEN IT IS DONE

Wherever a matched normal sample is obtainable and the clinical or research question requires confident somatic calls, particularly in whole exome and whole genome oncology work and in trial enrolment.

WHAT TRIGGERS IT

An oncology referral with matched normal tissue available, a research protocol requiring confident somatic calls, or trial screening demanding a defined mutation profile.

HOW IT IS DONE

Tumour and normal are aligned and processed identically, sample pairing is verified from the data, and somatic calling proceeds with germline subtraction. Cross-contamination and read orientation artefacts are modelled explicitly, calls are filtered against a panel of normals and population frequencies, and surviving variants are annotated, tiered and reviewed before reporting.

WHAT IT PRODUCES

A filtered somatic variant call set in mutation annotation format, the pairing verification record, contamination and orientation bias models, tier-assigned actionable findings, and a signed molecular report.

WHAT IT DETECTS

Somatic single nucleotide variants and small insertions and deletions across the sequenced territory, with germline variation subtracted and allele fractions reported against tumour purity.

WHAT IT DOES NOT DETECT

Variants below the assay's limit of detection in low-purity samples, mutations confined to unsampled tumour regions, and anything outside the sequenced territory. Structural variants, copy number and fusions require separate analyses. Contamination of the normal by tumour, common in haematological malignancy and adjacent tissue, silently removes real somatic calls.

WHAT MAKES IT HARD

Tumour purity governs everything and is rarely as high as assumed, so the effective limit of detection differs case by case. Tumour-in-normal contamination is the failure that produces confident false negatives, and formalin fixation generates a characteristic artefact signature that must be modelled rather than filtered by threshold.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics or computational biologist in oncology, mid to senior level, working with a molecular pathologist who holds reporting responsibility.

WHAT YOU CAN DO AFTERWARDS

Run a complete paired somatic pipeline, verify sample pairing from data, model contamination and artefact explicitly, and produce a defensible actionable variant set with tier assignment.

WHO HIRES FOR IT

Cancer centres and comprehensive cancer hospitals, oncology-focused diagnostic laboratories, pharmaceutical companies running translational oncology programmes, and oncology CROs.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
26	**128 hours · 22 days**	Prior germline variant analysis experience	**2,90,000**

SON-01
WORKFLOW

Tumour–Normal Paired Somatic Analysis

(**Paired somatic variant analysis** · T/N analysis · paired somatic)

COMPLETE WORKFLOW — 26 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Case authorisation and consent verification**	Test scope, germline finding policy and specimen consent confirmed before analysis	LIMS record, consent form	*Analysis authorisation*
02	**Specimen adequacy and tumour content recording**	Pathologist tumour content estimate and specimen adequacy recorded as analytical inputs	Pathology report	*Adequacy record*
03	**Data receipt and integrity check**	Checksums and read counts verified for tumour and normal against run records	md5sum, seqkit stats	*Verified FASTQ pair*
04	**Sequence quality assessment**	Quality, adapter content and duplication measured for both samples against criteria	FastQC, MultiQC	*QC acceptance record*
05	**Read preprocessing**	Adapters and low-quality tails removed identically across tumour and normal	fastp	*Trimmed FASTQ*
06	**Reference and resource version lock**	Reference build and all filtering resources confirmed and checksummed	Reference registry	*Version lock record*
07	**Alignment**	Tumour and normal aligned with identical parameters and read groups assigned	bwa-mem2	*Aligned BAMs*
08	**Duplicate marking**	Duplicates flagged with flow-cell-appropriate optical distance in both samples	Picard MarkDuplicates	*Marked BAMs*
09	**Base quality recalibration**	Systematic base quality error corrected against version-matched known sites	GATK BQSR	*Analysis-ready BAMs*
10	**Sample pairing verification**	Tumour and normal confirmed to originate from the same individual before subtraction	somalier relate, CrosscheckFingerprints	*Pairing verification*
11	**Tumour-in-normal assessment**	Normal assessed for tumour contamination that would suppress somatic calls	Allele fraction review, contamination estimation	*TiN assessment*
12	**Coverage assessment**	Depth measured across the sequenced territory for both samples with gaps declared	mosdepth	*Coverage record*
13	**Contamination estimation**	Cross-individual contamination estimated in both samples	GetPileupSummaries, CalculateContamination	*Contamination model*
14	**Panel of normals application**	A process-matched panel of normals prepared for recurrent artefact filtering	Internal PoN	*Authorised panel of normals*
15	**Somatic variant calling**	Somatic variants called with the normal used to subtract germline variation	Mutect2	*Raw somatic VCF*
16	**Read orientation bias modelling**	Formalin and oxidation artefact modelled from read orientation patterns	LearnReadOrientationModel	*Orientation model*
17	**Somatic filtering**	Calls filtered using contamination, orientation and panel of normals evidence	FilterMutectCalls	*Filtered somatic VCF*
18	**Population frequency filtering**	Residual germline-like variation removed using population allele frequencies	gnomAD	*Population-filtered calls*
19	**Allele fraction and depth review**	Calls assessed against tumour purity for plausibility and detection limit	Allele fraction analysis	*Reviewed calls*
20	**Read-level review of reportable calls**	Every reportable variant inspected at read level in tumour and normal	IGV review	*Reviewed variant set*
21	**Annotation**	Consequence, transcript and hotspot annotation applied on canonical oncology transcripts	VEP, oncology transcript set	*Annotated variants*
22	**Oncogenicity assessment**	Variants assessed for oncogenic effect against curated somatic evidence	ClinGen somatic criteria, OncoKB	*Oncogenicity calls*
23	**Clinical tiering**	Findings assigned evidence tiers for therapeutic, prognostic and diagnostic significance	AMP/ASCO/CAP tiering	*Tiered findings*
24	**Therapy and trial matching**	Tiered findings matched against approved therapies and open trials	Curated therapy and trial resources	*Therapy matches*
25	**Molecular tumour board preparation**	Findings assembled with evidence for multidisciplinary review	Board preparation template	*Board pack*
26	**Report generation and sign-out**	Result, tiering, limitations and tumour content issued under dual signature	Validated molecular report template	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
26	**128 h · 22 d**	**SON-01**	**2,90,000**

SON-02
SECTOR 01

Tumour-Only Somatic Analysis

(**Tumour-only variant analysis** · Tumour-only · unpaired somatic)

WHY IT IS DONE

A matched normal is frequently unavailable — archival blocks, referred specimens, resource constraints — yet the clinical question remains. Tumour-only analysis substitutes statistical filtering for the missing normal, and is the operational reality for the majority of oncology panels performed worldwide.

WHAT IT ANSWERS

Which of the variants observed in this tumour are most likely somatic rather than inherited, and which are actionable for treatment selection?

WHEN IT IS DONE

Wherever no matched normal exists, which covers most routine solid tumour panel testing, most archival material and most referred specimens.

WHAT TRIGGERS IT

An oncology referral with no matched normal available, archival block testing, or a rapid panel pathway where obtaining a normal sample would delay the result.

HOW IT IS DONE

After alignment and processing, calling proceeds without a matched normal against a panel of normals built from the same assay. Population frequency filters remove germline-likely variation, allele fraction patterns and hotspot evidence are used to infer somatic origin, and germline-suspicious findings are flagged for referral rather than reported as somatic.

WHAT IT PRODUCES

A filtered candidate somatic call set, the panel of normals provenance record, a germline-suspicious variant flag list with referral recommendations, tier-assigned actionable findings, and a signed report stating the tumour-only limitation.

WHAT IT DETECTS

Candidate somatic variants after germline-likely variation has been filtered out statistically, with actionable findings prioritised for therapy matching.

WHAT IT DOES NOT DETECT

It cannot definitively distinguish somatic from germline, so rare germline variants are misclassified as somatic and genuine somatic variants at germline-like frequencies are removed. Private germline variation in under-represented ancestries is systematically over-called as somatic, and incidental germline cancer predisposition findings surface without a consent framework to handle them.

WHAT MAKES IT HARD

The whole method rests on filtering decisions that cannot be verified without the normal. Panel of normals quality is the single strongest determinant of accuracy, ancestry bias in population filters is a genuine equity problem, and the germline referral pathway must exist before the first incidental predisposition finding appears.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics, mid level. The most commonly performed somatic analysis in routine diagnostic practice.

WHAT YOU CAN DO AFTERWARDS

Build and govern a panel of normals, apply germline filtering defensibly, recognise ancestry bias in filtering, and operate a germline referral pathway for incidental findings.

WHO HIRES FOR IT

Oncology diagnostic laboratories, hospital molecular pathology services, private diagnostic chains offering cancer panels, and reference laboratories.

COMMERCIALS

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

SON-02
WORKFLOW

Tumour-Only Somatic Analysis

(**Tumour-only variant analysis** · Tumour-only · unpaired somatic)

COMPLETE WORKFLOW — 24 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Case authorisation and consent verification**	Scope confirmed including the policy on incidental germline predisposition findings	LIMS record, consent form	*Analysis authorisation*
02	**Specimen adequacy and tumour content recording**	Tumour content and specimen adequacy recorded as inputs to detection limit	Pathology report	*Adequacy record*
03	**Data receipt and quality assessment**	Integrity verified and sequence quality measured against acceptance criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
04	**Read preprocessing**	Adapters and low-quality tails removed with retention accounting	fastp	*Trimmed FASTQ*
05	**Alignment and duplicate marking**	Reads aligned and duplicates flagged with assay-appropriate handling	bwa-mem2, Picard	*Marked BAM*
06	**Base quality recalibration**	Base quality error corrected against version-matched known sites	GATK BQSR	*Analysis-ready BAM*
07	**Coverage certification**	Depth certified per target against the assay's validated threshold	mosdepth, bedtools coverage	*Coverage certificate*
08	**Panel of normals construction**	A panel built from process-matched normals to capture recurrent artefacts	Internal PoN pipeline	*Authorised panel of normals*
09	**Panel of normals provenance record**	Composition, size, chemistry and build date of the panel documented	PoN register	*Provenance record*
10	**Tumour-only variant calling**	Variants called without a matched normal using the panel of normals	Mutect2 tumour-only	*Raw variant set*
11	**Read orientation bias modelling**	Formalin and oxidation artefact modelled and applied as a filter	LearnReadOrientationModel	*Orientation model*
12	**Artefact and panel filtering**	Recurrent artefacts and panel-matching calls removed	FilterMutectCalls	*Artefact-filtered calls*
13	**Population frequency filtering**	Germline-likely variation removed using ancestry-appropriate frequency thresholds	gnomAD by ancestry	*Population-filtered calls*
14	**Germline-likely flagging**	Variants with germline-consistent allele fractions and frequencies flagged rather than reported	Allele fraction and frequency logic	*Germline-suspicious list*
15	**Hotspot and somatic evidence weighting**	Known somatic hotspots retained despite frequency filters that would otherwise remove them	Curated hotspot catalogues	*Hotspot-retained calls*
16	**Ancestry bias assessment**	Filtering reviewed for over-calling in ancestries under-represented in reference databases	Ancestry-stratified review	*Bias assessment*
17	**Read-level review of reportable calls**	Every reportable variant inspected at read level	IGV review	*Reviewed variant set*
18	**Annotation**	Consequence and hotspot annotation applied on canonical oncology transcripts	VEP, oncology transcript set	*Annotated variants*
19	**Oncogenicity assessment**	Variants assessed for oncogenic effect against curated somatic evidence	ClinGen somatic criteria, OncoKB	*Oncogenicity calls*
20	**Clinical tiering**	Findings assigned evidence tiers for clinical significance	AMP/ASCO/CAP tiering	*Tiered findings*
21	**Therapy and trial matching**	Findings matched against approved therapies and open trials	Curated therapy and trial resources	*Therapy matches*
22	**Germline referral pathway application**	Germline-suspicious findings routed for confirmatory germline testing under the agreed pathway	Referral protocol	*Referral record*
23	**Report generation**	Result assembled with the tumour-only limitation stated explicitly	Validated report template	*Draft report*
24	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
24	**118 h · 20 d**	**SON-02**	**2,70,000**

SON-03
SECTOR 01

Solid Tumour Panel Analysis

(**Targeted oncology panel analysis** · Oncology panel · hotspot panel)

WHY IT IS DONE

Most treatment decisions turn on a small number of genes with approved or trial-linked therapies. A focused panel answers those questions at high depth, on small and degraded specimens, within a turnaround that fits an oncology clinic.

WHAT IT ANSWERS

Are any of the actionable alterations in this panel present in this tumour, at what allele fraction, and what therapy or trial does each imply?

WHEN IT IS DONE

At diagnosis for treatment selection in solid tumours, at progression to reassess, and as the default first-line molecular test in most oncology services.

WHAT TRIGGERS IT

A new cancer diagnosis requiring molecular characterisation, disease progression on current therapy, or a trial screening requirement.

HOW IT IS DONE

The panel design version and its validated claim scope are confirmed, specimen adequacy and tumour content are recorded, and deep calling proceeds with formalin artefact modelling. Coverage is certified per amplicon or exon, low-fraction calls are reviewed at read level, and findings are tiered against therapy and trial evidence.

WHAT IT PRODUCES

A per-target coverage certificate, the annotated variant set with allele fractions, tumour content and adequacy record, tier-assigned actionable findings with therapy and trial matches, and a signed molecular report.

WHAT IT DETECTS

Single nucleotide variants, small insertions and deletions, hotspot alterations and panel-scoped copy number changes at the high depth a focused design permits.

WHAT IT DOES NOT DETECT

Any gene outside the panel, which is the principal risk when a rare but actionable alteration lies elsewhere. Fusions require RNA analysis in most designs, large structural rearrangements are missed, and panel-derived tumour mutational burden is an estimate calibrated to the design rather than a measurement.

WHAT MAKES IT HARD

Specimen quality dominates outcomes. Low tumour content, small biopsies and long-fixed blocks all degrade sensitivity in ways a coverage metric alone does not reveal, so tumour content must be recorded and the effective limit of detection stated per case rather than as an assay-wide claim.

WHO DOES THIS JOB

Clinical bioinformatician or molecular scientist in oncology, entry to mid level. The highest-volume analytical role in cancer diagnostics.

WHAT YOU CAN DO AFTERWARDS

Operate an oncology panel end to end, model formalin artefact, state case-specific detection limits against tumour content, and tier findings for therapy matching.

WHO HIRES FOR IT

Every hospital offering cancer treatment, oncology diagnostic chains, molecular pathology laboratories, and reference laboratories at scale.

COMMERCIALS

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

SON-03
WORKFLOW

Solid Tumour Panel Analysis

(**Targeted oncology panel analysis** · Oncology panel · hotspot panel)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Panel design version verification**	Design version, gene content and validated claim scope confirmed and recorded	Panel design register	*Authorised panel version*
02	**Specimen adequacy and tumour content recording**	Tumour content, necrosis and block age recorded as analytical inputs	Pathology report	*Adequacy record*
03	**Nucleic acid quality assessment**	Input quantity and fragmentation assessed as predictors of assay performance	Fragment analysis, quantification	*Input quality record*
04	**Data receipt and quality assessment**	Integrity verified and sequence quality measured against panel acceptance criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
05	**Chemistry-appropriate preprocessing**	Adapters removed and primers trimmed where the design is amplicon-based	fastp, primer trimming	*Preprocessed FASTQ*
06	**Alignment and duplicate handling**	Reads aligned with duplicate marking applied or suppressed according to chemistry	bwa-mem2, Picard	*Analysis-ready BAM*
07	**Per-target coverage certification**	Depth certified per amplicon or exon against the validated threshold	mosdepth, bedtools coverage	*Coverage certificate*
08	**Deep variant calling**	Variants called at panel depth with a low allele fraction floor	Validated somatic caller	*Panel variant set*
09	**Formalin artefact filtering**	Deamination artefacts identified and removed by orientation and context modelling	Orientation bias model	*Artefact-filtered calls*
10	**Panel of normals filtering**	Recurrent design-specific artefacts removed using a process-matched panel	Internal PoN	*Filtered calls*
11	**Population frequency filtering**	Germline-likely variation removed with hotspot retention	gnomAD, hotspot catalogues	*Population-filtered calls*
12	**Low-fraction call review**	Calls near the detection floor inspected at read level individually	IGV review	*Reviewed low-fraction calls*
13	**Panel copy number calling**	Copy number called across genes carrying a validated claim	CNVkit, panel-specific method	*Panel CNV calls*
14	**Detection limit statement**	Effective limit of detection derived from tumour content and achieved depth	Sensitivity calculation	*Case-specific LOD*
15	**Annotation**	Consequence and hotspot annotation applied on canonical oncology transcripts	VEP, oncology transcript set	*Annotated variants*
16	**Oncogenicity assessment**	Variants assessed for oncogenic effect against curated somatic evidence	ClinGen somatic criteria, OncoKB	*Oncogenicity calls*
17	**Clinical tiering**	Findings assigned evidence tiers for therapeutic and prognostic significance	AMP/ASCO/CAP tiering	*Tiered findings*
18	**Therapy and trial matching**	Findings matched against approved therapies and open trials	Curated therapy and trial resources	*Therapy matches*
19	**Report generation**	Result, coverage, tumour content and case-specific detection limit assembled	Validated report template	*Draft report*
20	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**96 h · 16 d**	**SON-03**	**2,20,000**

SON-04
SECTOR 01

Comprehensive Genomic Profiling

(**Comprehensive genomic profiling analysis** · CGP)

WHY IT IS DONE

A single assay reporting mutations, copy number, fusions and the immunotherapy biomarkers together removes the sequential testing that exhausts small biopsies and delays treatment. It is the format most molecular tumour boards now expect and the one most trial screening protocols specify.

WHAT IT ANSWERS

What is the complete clinically relevant genomic profile of this tumour across every alteration class, and what does it imply for approved therapy, trial eligibility and prognosis?

WHEN IT IS DONE

At diagnosis in advanced or metastatic disease, in rare and difficult-to-treat cancers, at progression when the profile may have changed, and for trial screening.

WHAT TRIGGERS IT

Advanced or metastatic disease requiring full molecular characterisation, a rare tumour type, failure of first-line therapy, or trial eligibility assessment.

HOW IT IS DONE

Multiple analytical arms run against one library: small variant calling, copy number segmentation with purity correction, fusion detection, and biomarker computation. Each arm carries its own quality gates, results are integrated into a single profile, and every finding is tiered against therapy, trial and prognostic evidence before molecular tumour board review.

WHAT IT PRODUCES

An integrated genomic profile across all alteration classes, per-arm quality records, biomarker values with calibration statements, tiered therapy and trial matches, and a signed molecular tumour board report.

WHAT IT DETECTS

Single nucleotide variants, insertions and deletions, copy number amplification and loss, fusions, tumour mutational burden, microsatellite instability, and where designed for it homologous recombination deficiency and loss of heterozygosity.

WHAT IT DOES NOT DETECT

Alterations outside the design, expression-level changes without a genomic correlate, and epigenetic mechanisms of resistance. Immunotherapy biomarkers derived from a panel are calibrated estimates rather than direct measurements, and their concordance with exome-derived values must be stated rather than assumed.

WHAT MAKES IT HARD

Integrating four analytical arms with different failure modes into one report is the difficulty. Each arm can fail independently while the others succeed, so per-arm quality gating is essential, and panel-calibrated biomarkers must be reported with their calibration basis rather than as absolute values.

WHO DOES THIS JOB

Senior clinical bioinformatician in cancer genomics, working with molecular pathology and the tumour board. A lead-level role in most laboratories.

WHAT YOU CAN DO AFTERWARDS

Operate a multi-arm comprehensive profiling assay, gate each arm independently, calibrate and report panel-derived biomarkers correctly, and present at a molecular tumour board.

WHO HIRES FOR IT

Comprehensive cancer centres, commercial profiling providers, pharmaceutical companies running precision oncology programmes, and large diagnostic chains.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
28	**138 hours · 23 days**	Prior oncology panel and copy number analysis experience	**2,98,000**

SON-04
WORKFLOW

Comprehensive Genomic Profiling

(**Comprehensive genomic profiling analysis** · CGP)

COMPLETE WORKFLOW — 28 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Assay version and claim scope verification**	Design version and the validated claim for each alteration class confirmed	Assay register	*Authorised assay version*
02	**Specimen adequacy and tumour content recording**	Tumour content, necrosis and specimen size recorded as inputs to every arm	Pathology report	*Adequacy record*
03	**Nucleic acid quality assessment**	DNA and RNA quality and quantity assessed against arm-specific requirements	Fragment analysis, quantification	*Input quality record*
04	**Data receipt and quality assessment**	Integrity verified and quality measured for all libraries against criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
05	**Read preprocessing**	Adapters and low-quality tails removed across DNA and RNA libraries	fastp	*Trimmed FASTQ*
06	**Alignment and processing**	DNA aligned and processed; RNA aligned with a splice-aware method	bwa-mem2, STAR	*Analysis-ready BAMs*
07	**Coverage certification**	Depth certified per target with arm-specific thresholds applied	mosdepth, bedtools coverage	*Coverage certificate*
08	**Small variant calling**	Somatic variants called at depth with a defined allele fraction floor	Validated somatic caller	*Variant set*
09	**Artefact modelling and filtering**	Formalin, orientation and recurrent panel artefacts modelled and removed	Orientation model, internal PoN	*Filtered variant set*
10	**Copy number segmentation**	Coverage normalised and segmented against a process-matched reference	CNVkit, GATK	*Copy number segments*
11	**Purity-corrected copy state assignment**	Segments converted to copy states using fitted tumour purity	Purity fitting	*Gene-level copy states*
12	**Fusion detection**	Fusion transcripts detected from the RNA arm with multiple callers	Arriba, STAR-Fusion	*Fusion calls*
13	**Fusion artefact filtering**	Read-through transcripts and alignment artefacts removed from fusion calls	Filtering, review	*Filtered fusions*
14	**Microsatellite instability determination**	Instability assessed at panel-covered microsatellite loci	MSIsensor-pro	*MSI status*
15	**Tumour mutational burden calculation**	Burden computed over the callable denominator with counting rules applied	Burden pipeline	*TMB value*
16	**Panel calibration application**	Panel-derived biomarker values calibrated against exome-derived reference values	Calibration model	*Calibrated biomarkers*
17	**Loss of heterozygosity assessment**	Allelic imbalance assessed where the design supports it	Allelic analysis	*LOH findings*
18	**Per-arm quality gating**	Each analytical arm gated independently so one failure does not pass silently	Arm-specific QC criteria	*Per-arm quality record*
19	**Detection limit statement**	Effective limits stated per alteration class against tumour content	Sensitivity calculation	*Class-specific LOD*
20	**Annotation**	Consequence, hotspot and fusion annotation applied on canonical transcripts	VEP, fusion annotation	*Annotated findings*
21	**Oncogenicity assessment**	All alteration classes assessed for oncogenic effect against curated evidence	ClinGen somatic criteria, OncoKB	*Oncogenicity calls*
22	**Profile integration**	Findings across all arms integrated into a single coherent genomic profile	Integration procedure	*Integrated profile*
23	**Clinical tiering**	Every finding assigned an evidence tier for clinical significance	AMP/ASCO/CAP tiering	*Tiered findings*
24	**Therapy and trial matching**	Findings matched against approved therapies, off-label evidence and open trials	Curated therapy and trial resources	*Therapy matches*
25	**Resistance and prognostic annotation**	Known resistance alterations and prognostic markers annotated	Curated resistance evidence	*Resistance annotation*
26	**Molecular tumour board preparation**	Integrated profile assembled with evidence for multidisciplinary review	Board preparation template	*Board pack*
27	**Report generation**	Integrated report assembled with per-arm quality and calibration statements	Validated CGP report template	*Draft report*
28	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
28	**138 h · 23 d**	**SON-04**	**2,98,000**

SON-05
SECTOR 01

Somatic Whole Exome Analysis

(**Tumour whole exome analysis** · Tumour WES)

WHY IT IS DONE

Exome-wide somatic analysis captures the full coding mutational landscape rather than a preselected gene list, which is what mutational burden, signature analysis and neoantigen prediction all require. It is the standard research and trial format and increasingly a clinical one in difficult cases.

WHAT IT ANSWERS

What is the complete coding mutational landscape of this tumour, and what do its burden, signatures and neoantigen content imply?

WHEN IT IS DONE

In trial screening, translational research, immunotherapy candidate assessment, cancers of unknown primary, and clinical cases where panel testing has not explained the disease.

WHAT TRIGGERS IT

Trial protocol requirement, immunotherapy assessment needing exome-derived burden, an uninformative panel result, or a research protocol.

HOW IT IS DONE

Tumour and normal exomes are processed identically with pairing verified, somatic calling proceeds with contamination and orientation bias modelling, and coverage is certified across the capture target. The call set is filtered, annotated and used to derive burden and signatures, with the callable exome footprint declared as the denominator.

WHAT IT PRODUCES

An exome-wide somatic call set, capture and coverage certification, the callable footprint used as burden denominator, mutational burden and signature results, and a signed report.

WHAT IT DETECTS

Coding somatic variants exome-wide with allele fractions, directly measured tumour mutational burden, mutational signatures, and the mutation set required for neoantigen prediction.

WHAT IT DOES NOT DETECT

Non-coding drivers, structural rearrangements, most fusions and promoter mutations including TERT, which is a recurrent and clinically important omission. Copy number from exome data is coarser than from genome data, and low-purity samples lose sensitivity across the whole exome rather than in identifiable regions.

WHAT MAKES IT HARD

Tumour mutational burden is a rate, so the denominator matters as much as the numerator. Declaring the callable exome footprint honestly is what makes the value comparable across cases, and laboratories that quote burden without it are reporting an uninterpretable number.

WHO DOES THIS JOB

Clinical bioinformatician or computational biologist in cancer genomics, mid to senior level, frequently in a translational or trial support role.

WHAT YOU CAN DO AFTERWARDS

Run exome-scale somatic analysis, derive burden with a declared denominator, extract mutational signatures, and prepare the mutation set for downstream neoantigen work.

WHO HIRES FOR IT

Academic cancer centres, pharmaceutical translational oncology groups, oncology CROs, and trial screening laboratories.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
24	**116 hours · 20 days**	Prior paired somatic analysis experience	**2,65,000**

SON-05
WORKFLOW

Somatic Whole Exome Analysis

(**Tumour whole exome analysis** · Tumour WES)

COMPLETE WORKFLOW — 24 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Case authorisation and consent verification**	Scope confirmed including germline findings arising from the matched normal	LIMS record, consent form	*Analysis authorisation*
02	**Specimen adequacy and tumour content recording**	Tumour content recorded as the scaling factor for all downstream metrics	Pathology report	*Adequacy record*
03	**Data receipt and quality assessment**	Integrity verified and quality measured for tumour and normal exomes	md5sum, FastQC, MultiQC	*QC acceptance record*
04	**Read preprocessing**	Adapters and low-quality tails removed identically across both samples	fastp	*Trimmed FASTQ*
05	**Alignment and processing**	Both exomes aligned, duplicate marked and recalibrated identically	bwa-mem2, Picard, GATK BQSR	*Analysis-ready BAMs*
06	**Sample pairing verification**	Tumour and normal confirmed to share an identity before subtraction	somalier, CrosscheckFingerprints	*Pairing verification*
07	**Capture performance assessment**	On-target fraction and enrichment measured for both libraries	Picard CollectHsMetrics	*Capture efficiency record*
08	**Coverage and callable footprint definition**	Depth measured and the callable exome footprint defined for burden denominator use	mosdepth, CallableLoci	*Callable footprint*
09	**Tumour-in-normal assessment**	Normal assessed for tumour contamination that would suppress somatic calls	Allele fraction review	*TiN assessment*
10	**Contamination estimation**	Cross-individual contamination estimated in both samples	CalculateContamination	*Contamination model*
11	**Somatic variant calling**	Somatic variants called exome-wide with germline subtraction	Mutect2	*Raw somatic VCF*
12	**Read orientation bias modelling**	Formalin and oxidation artefact modelled from orientation patterns	LearnReadOrientationModel	*Orientation model*
13	**Somatic filtering**	Calls filtered using contamination, orientation and panel of normals evidence	FilterMutectCalls, internal PoN	*Filtered somatic VCF*
14	**Population frequency filtering**	Residual germline-like variation removed using population frequencies	gnomAD	*Population-filtered calls*
15	**Exome copy number calling**	Copy number called from exome coverage against a process-matched reference	CNVkit, GATK	*Exome CNV calls*
16	**Tumour mutational burden calculation**	Burden computed over the declared callable footprint with counting rules stated	Burden pipeline	*TMB value*
17	**Mutational signature analysis**	Signatures decomposed from the filtered call set where counts permit	SigProfiler, signature fitting	*Signature contributions*
18	**Neoantigen candidate preparation**	Mutation set prepared in the format required for downstream neoantigen analysis	Peptide generation	*Neoantigen input set*
19	**Annotation**	Consequence and hotspot annotation applied on canonical oncology transcripts	VEP	*Annotated variants*
20	**Oncogenicity assessment**	Variants assessed for oncogenic effect against curated somatic evidence	ClinGen somatic criteria, OncoKB	*Oncogenicity calls*
21	**Driver identification**	Drivers separated from passengers using recurrence and functional evidence	Driver evidence resources	*Driver call set*
22	**Clinical tiering**	Actionable findings assigned evidence tiers	AMP/ASCO/CAP tiering	*Tiered findings*
23	**Report generation**	Result, burden with denominator, signatures and limitations assembled	Validated report template	*Draft report*
24	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

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

SON-06
SECTOR 01

Somatic Whole Genome Analysis

(**Tumour whole genome analysis** · Tumour WGS)

WHY IT IS DONE

Only whole genome analysis sees the complete somatic landscape — non-coding drivers, structural rearrangements, complex events, mitochondrial changes and viral integration — in one assay. It is the reference method for cancer genome characterisation and is entering routine use in national programmes.

WHAT IT ANSWERS

What is the complete somatic alteration landscape of this tumour across every class and across the entire genome, including the non-coding and structural events other assays cannot see?

WHEN IT IS DONE

In national genomic medicine programmes, in research and trial contexts requiring full characterisation, in paediatric and rare cancers, and where panel and exome testing have failed to explain the tumour.

WHAT TRIGGERS IT

Enrolment in a whole genome programme, a rare or paediatric cancer, an uninformative exome, or a research protocol requiring complete characterisation.

HOW IT IS DONE

Tumour and normal genomes are processed identically with pairing verified, and parallel arms address small variants, structural variants, copy number with purity and ploidy fitting, signatures, and complex rearrangement classification. Driver identification proceeds against curated evidence, and findings are integrated for tumour board review.

WHAT IT PRODUCES

A genome-wide somatic call set across all alteration classes, allele-specific copy number with purity and ploidy, structural and complex rearrangement classification, full-resolution signature results, and a signed integrated report.

WHAT IT DETECTS

Somatic variants genome-wide including non-coding and promoter mutations, structural rearrangements with breakpoints, copy number with allele specificity, complex rearrangement patterns, mutational signatures at full resolution, mitochondrial changes and viral integration.

WHAT IT DOES NOT DETECT

Expression and epigenetic mechanisms, which require RNA and methylation assays. Interpretation of non-coding somatic variation remains extremely limited even where detection is reliable, and the sheer volume of passenger events makes driver identification the binding constraint rather than detection.

WHAT MAKES IT HARD

Detection is no longer the problem; distinguishing driver from passenger among many thousands of somatic events is. Structural variant false positives remain high, complex rearrangement patterns need explicit classification rather than enumeration, and non-coding findings are abundant and largely uninterpretable.

WHO DOES THIS JOB

Senior clinical bioinformatician or cancer genomics lead. Among the most technically demanding roles in the sector and correspondingly scarce.

WHAT YOU CAN DO AFTERWARDS

Operate whole genome somatic analysis end to end, fit purity and ploidy, classify complex rearrangements, extract signatures at full resolution, and separate drivers from passengers defensibly.

WHO HIRES FOR IT

National genomic medicine programmes, comprehensive cancer centres, cancer genome research institutes, and pharmaceutical discovery groups.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
27	**132 hours · 22 days**	Prior somatic exome and structural variant experience	**2,95,000**

SON-06
WORKFLOW

Somatic Whole Genome Analysis

(**Tumour whole genome analysis** · Tumour WGS)

COMPLETE WORKFLOW — 27 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Case authorisation and consent verification**	Scope confirmed including non-coding findings and germline findings from the normal	LIMS record, consent form	*Analysis authorisation*
02	**Specimen adequacy and tumour content recording**	Tumour content recorded as the scaling factor for purity fitting and all metrics	Pathology report	*Adequacy record*
03	**Data receipt and quality assessment**	Integrity verified and quality measured for both genomes	md5sum, FastQC, MultiQC	*QC acceptance record*
04	**Read preprocessing**	Adapters and low-quality tails removed identically across both genomes	fastp	*Trimmed FASTQ*
05	**Alignment and processing**	Both genomes aligned, duplicate marked and recalibrated identically	bwa-mem2, Picard, GATK BQSR	*Analysis-ready BAMs*
06	**Sample pairing verification**	Tumour and normal confirmed to share an identity before subtraction	somalier, CrosscheckFingerprints	*Pairing verification*
07	**Coverage and callable region analysis**	Genome-wide depth measured and the callable footprint declared	mosdepth, CallableLoci	*Callable region BED*
08	**Tumour-in-normal and contamination assessment**	Tumour contamination of the normal and cross-individual contamination both assessed	Allele fraction review, CalculateContamination	*Contamination models*
09	**Somatic small variant calling**	Somatic variants called genome-wide with germline subtraction	Mutect2	*Raw somatic VCF*
10	**Read orientation bias modelling and filtering**	Artefact modelled and applied alongside panel of normals filtering	LearnReadOrientationModel, internal PoN	*Filtered somatic VCF*
11	**Structural variant ensemble calling**	Structural rearrangements called by multiple callers against the matched normal	Manta, GRIDSS, Delly	*Somatic SV set*
12	**Structural variant merging and genotyping**	Calls merged and uniformly re-genotyped for comparability	SURVIVOR, SVTyper	*Merged SV set*
13	**Copy number segmentation**	Depth and allelic imbalance segmented jointly across the genome	Copy number pipeline	*Copy number segments*
14	**Purity and ploidy fitting**	Tumour purity and ploidy fitted with alternative solutions enumerated	Purity-ploidy fitting	*Purity-ploidy solution*
15	**Allele-specific copy number assignment**	Segments assigned allele-specific absolute copy states	Allele-specific assignment	*Allele-specific CN*
16	**Whole genome doubling assessment**	Genome doubling status determined from ploidy and allelic patterns	Ploidy analysis	*WGD status*
17	**Complex rearrangement classification**	Multi-breakpoint events classified as patterns rather than enumerated separately	Complex event classification	*Complex event calls*
18	**Mutational signature analysis**	Signatures decomposed at full genome resolution across substitution classes	SigProfiler	*Signature contributions*
19	**Non-coding and promoter variant assessment**	Recurrent promoter and regulatory somatic variants assessed against curated evidence	Regulatory annotation	*Non-coding findings*
20	**Viral integration screening**	Non-human reads screened for oncogenic viral sequence and integration junctions	Viral screening	*Viral findings*
21	**Mitochondrial somatic analysis**	Somatic mitochondrial variants called with heteroplasmy quantified	Mitochondrial calling	*mtDNA findings*
22	**Annotation**	All alteration classes annotated on canonical oncology transcripts	VEP, SV annotation	*Annotated findings*
23	**Driver identification**	Drivers separated from passengers across all alteration classes	Driver evidence resources	*Driver call set*
24	**Oncogenicity and tiering**	Findings assessed for oncogenic effect and assigned clinical evidence tiers	ClinGen somatic criteria, AMP/ASCO/CAP	*Tiered findings*
25	**Molecular tumour board preparation**	Integrated genome findings assembled for multidisciplinary review	Board preparation template	*Board pack*
26	**Report generation**	Integrated genome report assembled with limitations and callable footprint	Validated report template	*Draft report*
27	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
27	**132 h · 22 d**	**SON-06**	**2,95,000**

SON-07
SECTOR 01

Somatic Copy Number Analysis

(**Somatic CNA analysis** · Somatic CNV · CNA analysis)

WHY IT IS DONE

Amplifications and deletions drive a large share of cancers and determine eligibility for several approved therapies. Unlike germline copy number, somatic copy number must be interpreted against tumour purity and ploidy, which changes the analysis fundamentally.

WHAT IT ANSWERS

Which genomic regions are amplified or deleted in this tumour, at what absolute copy number once purity and ploidy are accounted for, and which are clinically actionable?

WHEN IT IS DONE

As a standard arm of every solid tumour panel, exome and genome analysis, and specifically where therapy depends on an amplification or a homozygous deletion.

WHAT TRIGGERS IT

Routine inclusion in oncology profiling, a therapy decision resting on gene amplification, or suspected homozygous deletion of a tumour suppressor.

HOW IT IS DONE

Coverage is normalised against a matched process-specific reference, allele fractions at heterozygous sites are extracted, and segmentation proceeds jointly on depth and allelic imbalance. Purity and ploidy are fitted so segments can be converted to absolute copy number, and gene-level states are assigned and classified.

WHAT IT PRODUCES

Segmented copy number with absolute copy states, the purity and ploidy fit with alternative solutions considered, gene-level calls for actionable targets, a detection limit statement against purity, and a signed report.

WHAT IT DETECTS

Focal and broad amplifications and deletions, homozygous deletions, absolute copy number after purity correction, and gene-level copy state for actionable targets.

WHAT IT DOES NOT DETECT

Balanced rearrangements, copy-neutral loss of heterozygosity without allelic analysis, and subclonal copy number changes present in a minority of cells. Low tumour purity compresses the observable copy number signal until real events become indistinguishable from noise.

WHAT MAKES IT HARD

Purity and ploidy fitting frequently admits more than one mathematically valid solution, and choosing the wrong one shifts every copy number call in the genome. The alternative solutions must be examined and the choice justified rather than accepting the top-scoring fit automatically.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics, mid level. A distinct competency from germline copy number and consistently under-supplied.

WHAT YOU CAN DO AFTERWARDS

Fit purity and ploidy defensibly, convert segments to absolute copy number, recognise and resolve ambiguous solutions, and report gene-level actionable copy states.

WHO HIRES FOR IT

Oncology diagnostic laboratories, cancer centres, comprehensive profiling providers, and pharmaceutical translational groups.

COMMERCIALS

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

SON-07
WORKFLOW

Somatic Copy Number Analysis

(**Somatic CNA analysis** · Somatic CNV · CNA analysis)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Specimen and tumour content recording**	Tumour content recorded because copy number interpretation is scaled by it	Pathology report	*Adequacy record*
02	**Reference set preparation**	A process-matched normal reference set prepared from the same assay and chemistry	Internal reference set	*Authorised reference set*
03	**Coverage extraction**	Read depth extracted across targets or bins identically for sample and reference	mosdepth, bedtools coverage	*Coverage matrix*
04	**GC and mappability correction**	Systematic coverage bias removed before segmentation	GC and mappability correction	*Corrected coverage*
05	**Reference normalisation**	Sample coverage normalised against the reference set with denoising	PCA denoising, reference normalisation	*Normalised coverage*
06	**Heterozygous site extraction**	B-allele frequencies extracted at heterozygous positions for allelic analysis	Allele fraction extraction	*BAF track*
07	**Joint segmentation**	Depth and allelic imbalance segmented together rather than separately	Segmentation algorithm	*Segments*
08	**Purity and ploidy fitting**	Tumour purity and ploidy fitted with alternative solutions enumerated and scored	Purity-ploidy fitting	*Fitted solution*
09	**Solution adjudication**	The selected fit checked against pathology content and known driver copy states	Orthogonal checks	*Adjudication record*
10	**Absolute copy state assignment**	Segments converted to absolute integer copy number using the fitted solution	Copy state assignment	*Absolute copy states*
11	**Focal and broad event classification**	Events classified as focal or broad with amplitude and boundaries recorded	Event classification	*Classified events*
12	**Homozygous deletion identification**	Complete losses identified and distinguished from low coverage artefacts	Depth review, read-level check	*Homozygous deletion calls*
13	**Gene-level copy state assignment**	Copy states assigned to actionable genes with boundaries checked against gene extents	Gene overlap analysis	*Gene-level calls*
14	**Detection limit statement**	Minimum detectable copy change stated against the fitted purity	Sensitivity calculation	*Purity-adjusted LOD*
15	**Population and artefact filtering**	Recurrent artefacts and germline copy number polymorphism removed	Internal artefact database	*Filtered call set*
16	**Oncogenicity assessment**	Amplifications and deletions assessed against curated oncogene and suppressor evidence	OncoKB, curated evidence	*Oncogenicity calls*
17	**Clinical tiering**	Actionable copy number findings assigned evidence tiers	AMP/ASCO/CAP tiering	*Tiered findings*
18	**Orthogonal confirmation**	High-impact copy number findings confirmed by an independent method where required	FISH, ddPCR	*Confirmation record*
19	**Report generation**	Copy states, purity fit and detection limit assembled into the report	Validated report template	*Draft report*
20	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**98 h · 17 d**	**SON-07**	**2,25,000**

SON-08
SECTOR 01

Tumour Purity & Ploidy Analysis

(**Purity, ploidy and allele-specific copy number analysis** · Purity–ploidy fitting)

WHY IT IS DONE

Almost every somatic metric depends on tumour purity: allele fractions, copy number states, clonality, mutational burden and limit of detection are all scaled by it. Estimating it wrongly produces a report in which every number is wrong in the same direction.

WHAT IT ANSWERS

What proportion of this specimen is tumour, what is the tumour's genome ploidy, and what allele-specific copy number state does each segment carry?

WHEN IT IS DONE

Within every somatic copy number and clonality analysis, and specifically where a pathologist's visual tumour content estimate conflicts with the molecular data.

WHAT TRIGGERS IT

Any somatic analysis requiring absolute copy number or clonality, a discrepancy between pathology and molecular estimates, or an ambiguous copy number profile.

HOW IT IS DONE

Depth ratios and B-allele frequencies are extracted across the genome and segmented jointly. A grid of purity and ploidy combinations is fitted, alternative solutions are enumerated and scored, and the selected fit is checked against pathology estimates, known driver copy states and mutation allele fractions before being accepted.

WHAT IT PRODUCES

A purity estimate with bounds, ploidy call including whole genome doubling status, allele-specific copy number per segment, the enumerated alternative fits with the justification for the one selected, and a concordance check against pathology.

WHAT IT DETECTS

Tumour purity with confidence bounds, genome ploidy including whole genome doubling, allele-specific copy number per segment, and copy-neutral loss of heterozygosity.

WHAT IT DOES NOT DETECT

It cannot resolve purity reliably in very low tumour content samples where the signal is too weak to fit, and it cannot distinguish subclonal populations from an intermediate purity in a single sample. Highly rearranged genomes admit multiple valid fits that the data alone cannot separate.

WHAT MAKES IT HARD

Multiple solutions are the defining difficulty, and they typically differ by a factor of two in ploidy. Automated selection of the top-scoring fit is wrong often enough to matter, so orthogonal checks against pathology and mutation allele fractions are the only reliable safeguard.

WHO DOES THIS JOB

Senior clinical bioinformatician or computational cancer biologist. A specialist skill that most oncology laboratories rely on one or two people for.

WHAT YOU CAN DO AFTERWARDS

Fit purity and ploidy from depth and allelic data, enumerate and adjudicate alternative solutions, detect whole genome doubling, and validate fits against orthogonal evidence.

WHO HIRES FOR IT

Cancer genomics laboratories, comprehensive profiling providers, pharmaceutical translational oncology, and cancer genome research programmes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**88 hours · 15 days**	Prior somatic copy number experience	**2,05,000**

SON-08
WORKFLOW

Tumour Purity & Ploidy Analysis

(**Purity, ploidy and allele-specific copy number analysis** · Purity–ploidy fitting)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Input data verification**	Coverage and allelic data confirmed adequate in extent and depth for fitting	Coverage and BAF review	*Input adequacy record*
02	**Pathology estimate capture**	The pathologist's tumour content estimate recorded as an independent comparator	Pathology report	*Pathology estimate*
03	**Coverage extraction and correction**	Depth extracted and corrected for GC and mappability bias genome-wide	mosdepth, bias correction	*Corrected coverage*
04	**Heterozygous site identification**	Germline heterozygous sites identified from the normal for allelic analysis	Germline genotyping	*Heterozygous site set*
05	**B-allele frequency extraction**	Allele fractions measured at heterozygous sites across the tumour	Allele fraction extraction	*BAF track*
06	**Joint segmentation**	Depth ratio and allelic imbalance segmented together across the genome	Segmentation algorithm	*Segments*
07	**Solution grid fitting**	A grid of purity and ploidy combinations fitted and scored against the data	Purity-ploidy grid search	*Candidate solutions*
08	**Alternative solution enumeration**	Competing solutions enumerated rather than discarding all but the top score	Solution enumeration	*Alternative solutions*
09	**Mutation allele fraction check**	Candidate solutions tested against clonal mutation allele fractions for consistency	Allele fraction consistency check	*Consistency assessment*
10	**Driver copy state check**	Solutions tested against expected copy states at known drivers for the tumour type	Driver state review	*Plausibility check*
11	**Pathology concordance check**	Fitted purity compared against the pathologist estimate with discrepancies examined	Concordance review	*Concordance record*
12	**Solution selection and justification**	The final solution selected with the reasoning for rejecting alternatives recorded	Adjudication procedure	*Selected solution*
13	**Allele-specific copy number assignment**	Major and minor allele copy numbers assigned to every segment	Allele-specific assignment	*Allele-specific CN*
14	**Whole genome doubling determination**	Genome doubling status determined from ploidy and allelic configuration	WGD analysis	*WGD status*
15	**Copy-neutral loss identification**	Regions with allelic loss at normal total copy number identified	LOH analysis	*Copy-neutral LOH regions*
16	**Confidence interval derivation**	Confidence bounds derived for the purity estimate from the fit landscape	Confidence estimation	*Purity bounds*
17	**Downstream scaling record**	Purity and ploidy recorded as the scaling inputs for all dependent analyses	Scaling record	*Downstream parameters*
18	**Reporting**	Purity, ploidy, allele-specific states and adjudication assembled for downstream use	Analysis report template	*Purity-ploidy report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**88 h · 15 d**	**SON-08**	**2,05,000**

SON-09
SECTOR 01

Somatic Structural Variant Analysis

(**Somatic structural rearrangement analysis** · Somatic SV)

WHY IT IS DONE

Structural rearrangements create fusion oncogenes, delete tumour suppressors, amplify drivers and disrupt regulatory architecture. Several are direct therapy targets, and in some cancers rearrangement rather than point mutation is the dominant mechanism.

WHAT IT ANSWERS

What somatic rearrangements does this tumour carry, where exactly are their breakpoints, which genes do they disrupt or fuse, and which are actionable?

WHEN IT IS DONE

In tumour whole genome analysis as standard, in sarcomas and haematological malignancies where rearrangement is the defining event, and where a fusion is suspected but not detected by RNA analysis.

WHAT TRIGGERS IT

Whole genome oncology analysis, a tumour type defined by characteristic rearrangements, or a suspected fusion requiring DNA-level breakpoint confirmation.

HOW IT IS DONE

Multiple callers with complementary strengths run against the tumour with the matched normal used to subtract germline rearrangements. Calls are merged, re-genotyped and filtered by class, breakpoints are reviewed at read level, junction sequence is assembled, and gene disruption and fusion consequences are assessed and classified.

WHAT IT PRODUCES

A somatic rearrangement call set with breakpoint coordinates, germline subtraction evidence, junction sequences, gene disruption and fusion assessment, complex event classification, and a signed report.

WHAT IT DETECTS

Somatic deletions, duplications, inversions, insertions and translocations with breakpoint coordinates, fusion-generating rearrangements, and complex multi-breakpoint events.

WHAT IT DOES NOT DETECT

Rearrangements with breakpoints in repetitive or segmentally duplicated sequence are unreliable, and short reads resolve balanced events far more poorly than unbalanced ones. Subclonal rearrangements below the detection threshold are missed, and low purity suppresses supporting read counts below callable levels.

WHAT MAKES IT HARD

False positives dominate unfiltered somatic structural calling, and the matched normal is the strongest available filter, so tumour-in-normal contamination degrades it badly. Complex events need classification as patterns rather than enumeration as dozens of separate breakpoints.

WHO DOES THIS JOB

Senior clinical bioinformatician in cancer genomics. One of the hardest positions in this sector to recruit for.

WHAT YOU CAN DO AFTERWARDS

Run somatic ensemble structural calling with germline subtraction, review breakpoints competently, assemble junctions, and classify complex rearrangement patterns.

WHO HIRES FOR IT

Cancer genome programmes, sarcoma and haematology specialist centres, comprehensive cancer centres, and cancer research institutes.

COMMERCIALS

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

SON-09
WORKFLOW

Somatic Structural Variant Analysis

(**Somatic structural rearrangement analysis** · Somatic SV)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Alignment verification**	Tumour and normal alignments verified for pairing, read groups and coverage adequacy	samtools stats	*Alignment verification*
02	**Sample pairing verification**	Tumour and normal confirmed to share an identity before germline subtraction	somalier, CrosscheckFingerprints	*Pairing verification*
03	**Tumour-in-normal assessment**	Normal assessed for tumour content that would remove true somatic rearrangements	Allele fraction review	*TiN assessment*
04	**Caller selection and version lock**	Callers selected for complementary class strengths with versions recorded	Validation record	*Caller configuration*
05	**Ensemble somatic calling**	Multiple callers run with the normal used to subtract germline rearrangements	Manta, GRIDSS, Delly	*Per-caller SV sets*
06	**Call merging**	Calls merged with defined breakpoint tolerance and caller agreement recorded	SURVIVOR, Jasmine	*Merged SV set*
07	**Uniform re-genotyping**	All merged calls re-genotyped uniformly for comparability across callers	SVTyper, graph genotyping	*Genotyped SV set*
08	**Class-specific filtering**	Each rearrangement class filtered against its own quality criteria	Per-class thresholds	*Filtered SV set*
09	**Repeat and blacklist annotation**	Calls in repeats, segmental duplications and blacklisted regions flagged	RepeatMasker, blacklist	*Context-annotated calls*
10	**Panel of normals filtering**	Recurrent process artefacts removed using a structural panel of normals	Internal SV PoN	*Artefact-filtered calls*
11	**Copy number corroboration**	Depth evidence assessed at breakpoints to separate balanced from unbalanced events	Depth analysis	*Balance assessment*
12	**Population frequency filtering**	Common germline structural polymorphism removed	gnomAD-SV, internal database	*Rare SV set*
13	**Breakpoint read-level review**	Both breakpoints of every reportable call inspected at read level	IGV review	*Reviewed calls*
14	**Junction assembly**	Local assembly performed to reconstruct the sequence across each junction	Local assembly	*Junction sequences*
15	**Gene disruption assessment**	Genes interrupted at each breakpoint identified with transcript consequence	Gene and transcript overlap	*Disruption assessment*
16	**Fusion consequence assessment**	Predicted fusions assessed for reading frame and oncogenic plausibility	Frame and orientation analysis	*Fusion predictions*
17	**Regulatory disruption assessment**	Enhancer hijacking and domain disruption assessed for position effect	Regulatory and TAD annotation	*Regulatory assessment*
18	**Complex event classification**	Multi-breakpoint events classified as recognisable patterns	Complex event classification	*Classified complex events*
19	**Per-class sensitivity determination**	Sensitivity established separately for each rearrangement class	Reference material benchmarking	*Per-class sensitivity*
20	**Oncogenicity and tiering**	Findings assessed for oncogenic effect and assigned clinical evidence tiers	OncoKB, AMP/ASCO/CAP	*Tiered findings*
21	**Orthogonal confirmation**	Reportable rearrangements confirmed by a class-appropriate independent method	Breakpoint PCR, FISH, RNA confirmation	*Confirmation record*
22	**Report generation and sign-out**	Rearrangements, breakpoints and sensitivity issued under dual signature	Validated report template	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**108 h · 18 d**	**SON-09**	**2,45,000**

SON-10
SECTOR 01

Loss of Heterozygosity Analysis

(**Somatic LOH and allelic imbalance analysis** · LOH · HLA LOH)

WHY IT IS DONE

Loss of the second allele is how most tumour suppressors are inactivated, and it is invisible to copy number analysis when the loss is copy-neutral. Loss of heterozygosity at the HLA locus additionally explains immunotherapy resistance, which makes it directly clinically relevant.

WHAT IT ANSWERS

Which genomic regions have lost heterozygosity in this tumour, is the loss copy-neutral or accompanied by deletion, and does it affect a tumour suppressor or the HLA locus?

WHEN IT IS DONE

Alongside somatic copy number analysis as standard, in homologous recombination deficiency assessment, in hereditary cancer second-hit confirmation, and in immunotherapy resistance investigation.

WHAT TRIGGERS IT

Homologous recombination deficiency scoring, a germline tumour suppressor variant requiring second-hit evidence, immunotherapy resistance investigation, or routine inclusion in profiling.

HOW IT IS DONE

Heterozygous sites are identified in the normal, B-allele frequencies are measured in the tumour, and allelic imbalance is segmented across the genome. Segments are classified as copy-neutral or deletion-associated using depth, tumour suppressor second hits are identified, and HLA loss is assessed with a locus-specific method.

WHAT IT PRODUCES

Genome-wide loss of heterozygosity segments with copy-neutral status distinguished, informative site density per region, tumour suppressor second-hit findings, HLA loss assessment, and the genome-wide loss fraction for scoring use.

WHAT IT DETECTS

Regional and genome-wide loss of heterozygosity, copy-neutral loss, allelic imbalance short of complete loss, second-hit events at tumour suppressor loci, and loss of heterozygosity at the HLA locus.

WHAT IT DOES NOT DETECT

Regions with too few informative heterozygous sites cannot be assessed, which particularly affects small genes and homozygous stretches in consanguineous individuals. Low purity compresses allelic imbalance below detectability, and without a matched normal, germline homozygosity is indistinguishable from somatic loss.

WHAT MAKES IT HARD

Informative site density limits what can be assessed and varies enormously across the genome, so declaring uninformative regions is essential. Copy-neutral loss is invisible without allelic analysis, and the HLA locus needs a specialist method because standard alignment fails there.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics, mid level, usually alongside copy number and homologous recombination work.

WHAT YOU CAN DO AFTERWARDS

Detect and classify loss of heterozygosity including copy-neutral events, assess informativeness honestly, confirm tumour suppressor second hits, and assess HLA loss.

WHO HIRES FOR IT

Cancer genomics laboratories, immunotherapy biomarker programmes, hereditary cancer services, and pharmaceutical translational oncology.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
16	**76 hours · 13 days**	Prior allele-specific copy number experience	**1,75,000**

SON-10
WORKFLOW

Loss of Heterozygosity Analysis

(**Somatic LOH and allelic imbalance analysis** · LOH · HLA LOH)

COMPLETE WORKFLOW — 16 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Germline heterozygous site identification**	Heterozygous positions identified in the matched normal as the informative site set	Germline genotyping	*Heterozygous site set*
02	**Informative site density mapping**	Density of informative sites mapped across the genome to define assessable regions	Density analysis	*Informativeness map*
03	**B-allele frequency extraction**	Allele fractions measured at informative sites across the tumour	Allele fraction extraction	*BAF track*
04	**Coverage extraction and correction**	Depth extracted and corrected for GC and mappability bias	mosdepth, bias correction	*Corrected coverage*
05	**Purity and ploidy input**	Fitted purity and ploidy applied as the scaling inputs for allelic interpretation	Purity-ploidy solution	*Scaling parameters*
06	**Allelic imbalance segmentation**	Allelic imbalance segmented across the genome jointly with depth	Segmentation algorithm	*Imbalance segments*
07	**Loss of heterozygosity classification**	Segments classified as retaining or having lost heterozygosity	Classification logic	*LOH segments*
08	**Copy-neutral loss identification**	Losses occurring at normal total copy number identified separately	Depth and allelic comparison	*Copy-neutral LOH*
09	**Partial imbalance assessment**	Allelic imbalance short of complete loss quantified and reported separately	Imbalance quantification	*Partial imbalance findings*
10	**Uninformative region declaration**	Regions with insufficient informative sites declared rather than reported as retained	Informativeness thresholds	*Uninformative region list*
11	**Genome-wide loss fraction calculation**	The proportion of the genome showing loss computed for scoring use	Fraction calculation	*Genome-wide LOH fraction*
12	**Tumour suppressor second-hit assessment**	Germline tumour suppressor variants tested for somatic second-hit loss	Gene-level LOH review	*Second-hit findings*
13	**HLA locus loss assessment**	Loss of heterozygosity at the HLA locus assessed with a locus-specific method	LOHHLA-class method	*HLA LOH result*
14	**Purity-adjusted confidence assessment**	Reliability of calls assessed against fitted purity with low-confidence regions flagged	Confidence assessment	*Confidence record*
15	**Clinical interpretation**	Findings interpreted for immunotherapy resistance, scarring scores and predisposition	Clinical interpretation	*Interpretation record*
16	**Reporting**	LOH segments, fraction, second hits and HLA status assembled for downstream use	Report template	*LOH report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
16	**76 h · 13 d**	**SON-10**	**1,75,000**

SON-11
SECTOR 01

Mutational Signature Analysis

(**Somatic mutational signature analysis** · Signature analysis · SBS analysis)

WHY IT IS DONE

The pattern of mutations records the processes that produced them — ultraviolet damage, tobacco, defective mismatch repair, homologous recombination deficiency, prior chemotherapy. Signatures therefore reveal mechanism and treatment sensitivity that no individual mutation does.

WHAT IT ANSWERS

Which mutational processes have been active in this tumour, in what proportion, and do any of them imply a therapeutic vulnerability or an environmental exposure?

WHEN IT IS DONE

In exome and genome somatic analysis where mutation counts are sufficient, in homologous recombination deficiency assessment, in cancers of unknown aetiology, and in treatment-related malignancy investigation.

WHAT TRIGGERS IT

Sufficient somatic mutation count from exome or genome analysis, suspected repair deficiency, unexplained tumour aetiology, or suspected therapy-related cancer.

HOW IT IS DONE

The somatic call set is filtered to high-confidence variants and formalin and orientation artefacts are removed. Mutations are decomposed into trinucleotide context, contributions are fitted against the reference signature catalogue with confidence assessment, and results are interpreted only where mutation count supports them.

WHAT IT PRODUCES

Signature contributions with confidence bounds, the mutation count and its adequacy for the analysis, artefact removal evidence, and interpretation of therapeutically or aetiologically relevant signatures.

WHAT IT DETECTS

Single base substitution, doublet and insertion-deletion signatures with proportional contributions, repair deficiency signatures, and exposure-associated patterns.

WHAT IT DOES NOT DETECT

Signature extraction is unreliable below a few dozen mutations, which excludes most panel data and many low-burden tumours. Signatures are correlative rather than causal, several are cosmetically similar and confusable, and formalin artefact produces a pattern that mimics a real signature if not removed first.

WHAT MAKES IT HARD

Artefact is the trap. Formalin damage produces a characteristic pattern that decomposition will happily assign to a real signature, so artefact removal must precede fitting. Low mutation counts produce unstable fits that look definitive, which makes stating count adequacy essential.

WHO DOES THIS JOB

Computational cancer biologist or clinical bioinformatician, mid to senior level, in translational and research-facing settings.

WHAT YOU CAN DO AFTERWARDS

Extract and fit mutational signatures correctly, remove artefact before decomposition, judge count adequacy, and interpret repair deficiency and exposure signatures.

WHO HIRES FOR IT

Cancer research institutes, pharmaceutical translational oncology, academic cancer centres, and comprehensive profiling providers.

COMMERCIALS

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

SON-11
WORKFLOW

Mutational Signature Analysis

(**Somatic mutational signature analysis** · Signature analysis · SBS analysis)

COMPLETE WORKFLOW — 17 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Mutation count adequacy assessment**	Total somatic mutation count assessed against the minimum required for stable fitting	Count assessment	*Adequacy record*
02	**High-confidence call set preparation**	The call set restricted to high-confidence somatic variants before decomposition	Filtering procedure	*High-confidence call set*
03	**Formalin artefact removal**	Deamination artefact identified by orientation and context and removed	Orientation bias model	*Artefact-removed set*
04	**Oxidation artefact removal**	Oxidative damage artefact identified and removed before decomposition	Artefact modelling	*Cleaned call set*
05	**Germline contamination check**	Residual germline variation checked for, since it distorts context distributions	Frequency review	*Contamination check*
06	**Trinucleotide context assignment**	Each substitution assigned its trinucleotide context on the pyrimidine strand	Context assignment	*Context matrix*
07	**Indel and doublet matrix construction**	Insertion-deletion and doublet substitution matrices constructed separately	Matrix construction	*ID and DBS matrices*
08	**Reference catalogue selection**	The reference signature catalogue and its version selected and recorded	Signature catalogue register	*Catalogue version*
09	**Signature fitting**	Reference signatures fitted to the observed matrices with contributions estimated	Signature fitting	*Signature contributions*
10	**Confidence and stability assessment**	Fit stability assessed by resampling with confidence bounds derived	Bootstrap analysis	*Confidence bounds*
11	**Cosine similarity assessment**	Reconstruction quality assessed against the observed profile	Similarity calculation	*Reconstruction quality*
12	**Confusable signature review**	Signatures with similar profiles reviewed against clinical plausibility	Discrimination review	*Discrimination record*
13	**Repair deficiency signature interpretation**	Mismatch repair and homologous recombination signatures interpreted for therapy relevance	Clinical interpretation	*Deficiency interpretation*
14	**Exposure signature interpretation**	Environmental and therapy-related signatures interpreted for aetiology	Clinical interpretation	*Exposure interpretation*
15	**Therapy-related malignancy assessment**	Prior treatment signatures assessed where a secondary malignancy is suspected	Treatment history review	*Therapy-related assessment*
16	**Integration with genomic findings**	Signature results integrated with repair gene mutations and instability scores	Integration procedure	*Integrated interpretation*
17	**Reporting**	Contributions, confidence, count adequacy and interpretation assembled	Report template	*Signature report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
17	**82 h · 14 d**	**SON-11**	**1,90,000**

SON-12
SECTOR 01

Tumour Mutational Burden Analysis

(**TMB determination** · TMB)

WHY IT IS DONE

Tumour mutational burden is an approved companion biomarker for immune checkpoint therapy across tumour types. It is also one of the least standardised measurements in oncology, so how it is computed determines whether a patient is offered a therapy.

WHAT IT ANSWERS

How many somatic mutations per megabase does this tumour carry, over what defined denominator, and does that value cross the threshold for immunotherapy eligibility?

WHEN IT IS DONE

In comprehensive profiling and exome analysis for immunotherapy decisions, in trial screening with a burden eligibility criterion, and wherever checkpoint inhibitor therapy is being considered.

WHAT TRIGGERS IT

Consideration of checkpoint inhibitor therapy, trial eligibility screening, or routine inclusion within comprehensive profiling.

HOW IT IS DONE

The somatic call set is filtered under explicit counting rules covering variant class, allele fraction floor, germline filtering and driver inclusion. The callable denominator is computed from actual coverage rather than assumed design size, burden is calculated, and panel-to-exome calibration is applied and stated where relevant.

WHAT IT PRODUCES

The burden value with the counting rules and denominator both stated, the callable footprint used, the calibration basis for panel-derived values, and threshold assessment against the applicable eligibility criterion.

WHAT IT DETECTS

Somatic mutation count over a defined callable denominator, expressed as mutations per megabase, with the counting rules and threshold applied stated explicitly.

WHAT IT DOES NOT DETECT

It is a count, not a mechanism, and it does not identify which mutations generate immunogenic neoantigens. Panel-derived values are estimates calibrated to exome, and they diverge substantially at the clinically decisive threshold. Tumour-only burden is inflated by residual germline variation, particularly in under-represented ancestries.

WHAT MAKES IT HARD

Every element of the calculation is a choice — which variants count, what allele fraction floor applies, whether drivers are included, what the denominator is — and different choices shift values across the decision threshold. Two laboratories can report different burdens for the same tumour and both be defensible.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics, mid level, working with molecular pathology on threshold policy.

WHAT YOU CAN DO AFTERWARDS

Compute burden under explicit and documented rules, derive a genuine callable denominator, apply panel calibration correctly, and defend a value at the decision threshold.

WHO HIRES FOR IT

Oncology diagnostic laboratories, comprehensive profiling providers, immunotherapy trial screening laboratories, and pharmaceutical companies.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
16	**78 hours · 13 days**	Prior somatic variant calling experience	**1,80,000**

SON-12
WORKFLOW

Tumour Mutational Burden Analysis

(**TMB determination** · TMB)

COMPLETE WORKFLOW — 16 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Counting rule specification**	The variant classes, allele fraction floor and inclusion rules defined and recorded	Assay specification	*Counting rules*
02	**Threshold and guideline recording**	The eligibility threshold and its source guideline recorded with version	Guideline register	*Threshold record*
03	**Somatic call set preparation**	The somatic call set prepared under the specified counting rules	Filtering procedure	*Counting call set*
04	**Germline filtering verification**	Germline removal verified, since residual germline inflates the value	Frequency and PoN review	*Germline filtering record*
05	**Artefact removal verification**	Formalin and orientation artefacts confirmed removed before counting	Artefact model review	*Artefact removal record*
06	**Driver inclusion decision**	Whether driver mutations are counted decided and applied consistently	Documented rule	*Driver rule application*
07	**Allele fraction floor application**	The specified allele fraction floor applied uniformly across the call set	Filtering	*Floor-applied set*
08	**Callable denominator computation**	The denominator computed from achieved coverage rather than nominal design size	Coverage-based calculation	*Callable denominator*
09	**Burden calculation**	Mutations per megabase calculated over the computed callable denominator	Burden calculation	*Raw burden value*
10	**Panel-to-exome calibration**	Panel-derived values calibrated against exome reference where applicable	Calibration model	*Calibrated burden*
11	**Confidence interval derivation**	Statistical uncertainty derived from mutation count and denominator size	Confidence estimation	*Burden bounds*
12	**Tumour content adjustment assessment**	Effect of tumour content on burden assessed and low-content cases flagged	Content review	*Content adjustment record*
13	**Threshold assessment**	The value assessed against the applicable threshold with proximity noted	Threshold comparison	*Eligibility assessment*
14	**Concordance context statement**	Expected divergence from other assays at the decision threshold stated	Concordance data	*Concordance statement*
15	**Interpretation**	Burden interpreted alongside microsatellite status and immune context	Integrated interpretation	*Interpretation record*
16	**Reporting**	Value, rules, denominator, calibration and threshold assessment assembled	Report template	*TMB report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
16	**78 h · 13 d**	**SON-12**	**1,80,000**

SON-13
SECTOR 01

Microsatellite Instability Analysis

(**MSI status determination** · MSI)

WHY IT IS DONE

Microsatellite instability identifies mismatch repair deficient tumours, which respond to checkpoint immunotherapy across cancer types and may indicate Lynch syndrome in the patient's family. It is a small analysis with disproportionate clinical consequence.

WHAT IT ANSWERS

Is this tumour microsatellite unstable, at what fraction of assessed loci, and does that indicate immunotherapy eligibility or possible Lynch syndrome?

WHEN IT IS DONE

Universally in colorectal and endometrial cancer, in any tumour being considered for checkpoint therapy, and as a standard arm of comprehensive profiling.

WHAT TRIGGERS IT

Colorectal or endometrial cancer diagnosis, consideration of immunotherapy, a family history suggesting Lynch syndrome, or routine profiling inclusion.

HOW IT IS DONE

Microsatellite loci are genotyped from the sequencing data with locus-specific models, either against the matched normal or against a baseline built from stable samples. The unstable locus fraction is computed, status is called against a validated threshold, and low tumour content cases are flagged as potentially uninformative.

WHAT IT PRODUCES

Per-locus instability calls, the unstable fraction, the overall status against the validated threshold, an informativeness assessment against tumour content and coverage, and Lynch syndrome referral recommendation where indicated.

WHAT IT DETECTS

Instability at assessed microsatellite loci with the unstable fraction reported, an overall status call against the validated threshold, and concordance context with immunohistochemistry where available.

WHAT IT DOES NOT DETECT

Not all mismatch repair deficient tumours are unstable at the loci assessed, so a stable result does not exclude deficiency, and immunohistochemistry remains complementary rather than redundant. It also does not distinguish somatic from germline causation, which requires separate testing.

WHAT MAKES IT HARD

Low tumour content produces false stable results, which is the clinically dangerous direction of error since it withholds an effective therapy. The number of adequately covered informative loci must be reported, and cases falling below the informative minimum should be declared uninformative rather than stable.

WHO DOES THIS JOB

Clinical bioinformatician or molecular scientist in oncology, entry to mid level.

WHAT YOU CAN DO AFTERWARDS

Determine microsatellite status from sequencing data, judge informativeness against tumour content, avoid false stable calls, and trigger appropriate Lynch syndrome referral.

WHO HIRES FOR IT

Oncology diagnostic laboratories, colorectal and gynaecological cancer services, molecular pathology departments, and hereditary cancer services.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
15	**72 hours · 12 days**	Prior somatic analysis experience	**1,65,000**

SON-13
WORKFLOW

Microsatellite Instability Analysis

(**MSI status determination** · MSI)

COMPLETE WORKFLOW — 15 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Locus panel authorisation**	The microsatellite loci assessed and their validated performance confirmed	Locus panel register	*Authorised locus panel*
02	**Tumour content recording**	Tumour content recorded because low content produces false stable results	Pathology report	*Content record*
03	**Coverage assessment at loci**	Depth confirmed adequate at each microsatellite locus for genotyping	mosdepth, per-locus coverage	*Coverage record*
04	**Baseline selection**	Matched normal or a stable-sample baseline selected and recorded	Baseline procedure	*Selected baseline*
05	**Locus genotyping**	Repeat length distributions genotyped at each locus in the tumour	MSIsensor-pro, locus genotyping	*Locus genotypes*
06	**Distribution comparison**	Tumour length distributions compared against the baseline per locus	Distribution comparison	*Comparison results*
07	**Per-locus instability calling**	Each locus called stable or unstable against validated criteria	Instability criteria	*Per-locus calls*
08	**Informative locus counting**	Loci with adequate coverage and informativeness counted	Informativeness assessment	*Informative locus count*
09	**Unstable fraction calculation**	The proportion of informative loci showing instability calculated	Fraction calculation	*Unstable fraction*
10	**Status determination**	Overall status called against the validated threshold	Threshold comparison	*MSI status*
11	**Uninformative case flagging**	Cases with too few informative loci declared uninformative rather than stable	Minimum locus criteria	*Informativeness decision*
12	**Low content false-stable assessment**	Cases with low tumour content assessed for false stable risk	Content and fraction review	*False-stable assessment*
13	**Immunohistochemistry correlation**	Result correlated against mismatch repair immunohistochemistry where available	Concordance review	*Correlation record*
14	**Lynch syndrome referral assessment**	Germline referral considered where instability suggests inherited mismatch repair deficiency	Referral criteria	*Referral recommendation*
15	**Reporting**	Status, unstable fraction, informativeness and referral recommendation assembled	Report template	*MSI report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
15	**72 h · 12 d**	**SON-13**	**1,65,000**

SON-14
SECTOR 01

Homologous Recombination Deficiency Analysis

(**HRD and genomic scar analysis** · HRD · genomic scar)

WHY IT IS DONE

Tumours unable to repair double-strand breaks by homologous recombination are sensitive to PARP inhibition, and that sensitivity extends beyond BRCA-mutated tumours. Genomic scarring detects the functional deficiency itself rather than only its commonest genetic cause.

WHAT IT ANSWERS

Does this tumour show the genomic scarring characteristic of homologous recombination deficiency, and does it therefore qualify for PARP inhibitor therapy?

WHEN IT IS DONE

In ovarian, breast, prostate and pancreatic cancer where PARP inhibition is a treatment option, and wherever a BRCA-negative tumour may still be repair deficient.

WHAT TRIGGERS IT

Consideration of PARP inhibitor therapy, a tumour type with established repair deficiency prevalence, or a negative BRCA result in a clinically suspicious case.

HOW IT IS DONE

Allele-specific copy number is derived with purity and ploidy fitted, and the three scar components are computed across the genome. The composite score is calculated against the validated threshold, deficiency signatures are assessed where mutation counts allow, and the result is integrated with BRCA and related gene status.

WHAT IT PRODUCES

The three component scar scores, the composite instability score against the validated threshold, purity-adjusted confidence, signature corroboration where available, and integrated repair gene status.

WHAT IT DETECTS

Loss of heterozygosity burden, telomeric allelic imbalance and large-scale state transitions, the composite genomic instability score, and deficiency signatures where mutation counts permit.

WHAT IT DOES NOT DETECT

It measures historical scarring rather than current repair capacity, so a tumour that has restored function through reversion still scores as deficient. It gives no information about the specific gene responsible, and scores in low-purity samples are unreliable in the direction of falsely negative.

WHAT MAKES IT HARD

The score depends entirely on allele-specific copy number, which depends on purity and ploidy fitting, so an error two steps upstream propagates silently into a therapy decision. Reversion mutations restoring function are invisible to scarring and must be sought separately in resistant disease.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics, mid to senior level, working with molecular pathology on threshold governance.

WHAT YOU CAN DO AFTERWARDS

Compute genomic scar components and composite scores, judge reliability against purity, corroborate with signatures, and integrate with repair gene mutation status.

WHO HIRES FOR IT

Gynaecological and breast oncology services, comprehensive profiling providers, pharmaceutical companies developing PARP inhibitors, and oncology diagnostic laboratories.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**88 hours · 15 days**	Prior allele-specific copy number experience	**2,05,000**

SON-14
WORKFLOW

Homologous Recombination Deficiency Analysis

(**HRD and genomic scar analysis** · HRD · genomic scar)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Scoring method authorisation**	The scar scoring method, components and validated threshold confirmed	Assay register	*Authorised scoring method*
02	**Tumour content recording**	Tumour content recorded because low purity biases scores toward negative	Pathology report	*Content record*
03	**Coverage extraction and correction**	Depth extracted genome-wide and corrected for GC and mappability bias	mosdepth, bias correction	*Corrected coverage*
04	**Heterozygous site extraction**	Germline heterozygous sites identified and allele fractions measured in tumour	Germline genotyping, BAF extraction	*BAF track*
05	**Joint segmentation**	Depth and allelic imbalance segmented together across the genome	Segmentation algorithm	*Segments*
06	**Purity and ploidy fitting**	Purity and ploidy fitted with alternative solutions enumerated	Purity-ploidy fitting	*Fitted solution*
07	**Allele-specific copy number assignment**	Major and minor allele copy numbers assigned per segment	Allele-specific assignment	*Allele-specific CN*
08	**Loss of heterozygosity component**	Regions of loss above the size threshold identified and totalled	LOH component calculation	*LOH score*
09	**Telomeric allelic imbalance component**	Imbalanced segments extending to a telomere counted	TAI component calculation	*TAI score*
10	**Large-scale state transition component**	Transitions between copy states above the size threshold counted	LST component calculation	*LST score*
11	**Composite score calculation**	Component scores combined into the composite instability score	Composite calculation	*Composite score*
12	**Threshold assessment**	Composite score assessed against the validated eligibility threshold	Threshold comparison	*Eligibility assessment*
13	**Purity-adjusted confidence assessment**	Score reliability assessed against fitted purity with low-confidence cases flagged	Confidence assessment	*Confidence record*
14	**Signature corroboration**	Homologous recombination deficiency signatures assessed where counts permit	Signature fitting	*Signature corroboration*
15	**Repair gene status integration**	BRCA and related gene mutation and methylation status integrated	Gene status review	*Integrated repair status*
16	**Reversion mutation screening**	Restorative reversion mutations screened for in previously treated disease	Reversion analysis	*Reversion findings*
17	**Clinical interpretation**	Deficiency status interpreted for PARP inhibitor eligibility	Clinical interpretation	*Interpretation record*
18	**Reporting**	Components, composite score, confidence and gene status assembled	Report template	*HRD report*

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

SON-15
SECTOR 01

Focal Amplification & Extrachromosomal DNA Analysis

(**ecDNA and focal amplification analysis** · ecDNA analysis)

WHY IT IS DONE

Oncogenes carried on extrachromosomal circular DNA reach copy numbers chromosomal amplification cannot, segregate unequally at division, and drive rapid resistance evolution. Distinguishing this mechanism from chromosomal amplification changes prognosis and treatment expectation.

WHAT IT ANSWERS

Is this amplification extrachromosomal or chromosomal, what is its structure and copy number, and which oncogenes does it carry?

WHEN IT IS DONE

Where very high focal copy number is observed, in aggressive tumours with rapid resistance, in glioblastoma and neuroblastoma particularly, and in resistance investigation.

WHAT TRIGGERS IT

Extreme focal amplification detected during copy number analysis, unexpectedly rapid therapy resistance, or a tumour type with known extrachromosomal prevalence.

HOW IT IS DONE

High copy number seed regions are identified from copy number analysis, structural variant breakpoints within and around them are extracted, and amplicon structure is reconstructed by graph analysis. The reconstruction is classified as circular, breakage-fusion-bridge or linear, and constituent oncogenes and regulatory elements are annotated.

WHAT IT PRODUCES

Amplicon structural reconstructions, mechanism classification with confidence, copy number per amplicon segment, the oncogene and enhancer content, and recommendations for cytogenetic confirmation.

WHAT IT DETECTS

Focal amplifications with copy number, circular against linear amplicon structure, amplicon architecture and constituent segments, and the oncogenes carried within.

WHAT IT DOES NOT DETECT

Short-read data infers rather than observes circularity, so classification is probabilistic and low-copy circular elements may be missed entirely. Cell-to-cell heterogeneity in copy number is invisible in bulk sequencing, and confirming circularity definitively needs cytogenetic or long-read methods.

WHAT MAKES IT HARD

Reconstruction is an inference from breakpoint graphs rather than a direct observation, so confidence must be stated. Distinguishing extrachromosomal from breakage-fusion-bridge amplification is genuinely difficult and the two carry different biological implications.

WHO DOES THIS JOB

Senior clinical bioinformatician or computational cancer biologist. A specialist and rapidly growing niche.

WHAT YOU CAN DO AFTERWARDS

Reconstruct amplicon architecture from copy number and breakpoint data, classify amplification mechanism, and state confidence in circularity inference.

WHO HIRES FOR IT

Neuro-oncology and paediatric oncology centres, cancer research institutes, pharmaceutical discovery groups, and comprehensive cancer centres.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
17	**84 hours · 14 days**	Prior somatic copy number and structural variant experience	**1,95,000**

SON-15
WORKFLOW

Focal Amplification & Extrachromosomal DNA Analysis

(**ecDNA and focal amplification analysis** · ecDNA analysis)

COMPLETE WORKFLOW — 17 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Seed region identification**	Regions of extreme focal copy number identified as amplicon seeds	Copy number analysis	*Seed regions*
02	**Copy number amplitude quantification**	Absolute copy number at each seed quantified against fitted purity	Copy state assignment	*Amplitude record*
03	**Breakpoint extraction**	Structural variant breakpoints within and bounding each seed extracted	SV calling	*Breakpoint set*
04	**Breakpoint quality review**	Breakpoints supporting reconstruction inspected at read level	IGV review	*Reviewed breakpoints*
05	**Amplicon graph construction**	Copy number segments and breakpoints assembled into a connectivity graph	Graph construction	*Amplicon graph*
06	**Cycle detection**	Circular paths through the graph detected as candidate extrachromosomal structures	Graph cycle analysis	*Candidate cycles*
07	**Structure reconstruction**	Amplicon architecture reconstructed from graph paths and copy number	AmpliconArchitect-class reconstruction	*Reconstructed amplicons*
08	**Mechanism classification**	Structures classified as circular, breakage-fusion-bridge or linear amplification	Classification criteria	*Mechanism classification*
09	**Confidence assessment**	Confidence in circularity inference assessed from supporting evidence	Confidence criteria	*Confidence record*
10	**Constituent segment annotation**	Segments composing each amplicon annotated with genomic origin	Segment annotation	*Annotated segments*
11	**Oncogene content annotation**	Oncogenes carried within each amplicon identified with copy number	Gene overlap analysis	*Oncogene content*
12	**Enhancer and regulatory content assessment**	Regulatory elements co-amplified with oncogenes identified	Regulatory annotation	*Regulatory content*
13	**Heterogeneity limitation statement**	The invisibility of cell-to-cell copy number variation in bulk data stated	Limitation template	*Limitation statement*
14	**Prognostic interpretation**	Findings interpreted for aggressiveness and resistance evolution implications	Clinical interpretation	*Prognostic interpretation*
15	**Cytogenetic confirmation recommendation**	Confirmation by metaphase or long-read methods recommended where decisive	Confirmation criteria	*Confirmation recommendation*
16	**Actionability assessment**	Amplified oncogenes assessed against available targeted options	OncoKB, therapy resources	*Actionability assessment*
17	**Reporting**	Reconstructions, mechanism, oncogene content and confidence assembled	Report template	*Amplicon report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
17	**84 h · 14 d**	**SON-15**	**1,95,000**

SON-16
SECTOR 01

Subclonal Reconstruction & Tumour Evolution Analysis

(**Clonality and cancer cell fraction analysis** · Clonality analysis · CCF analysis)

WHY IT IS DONE

A tumour is a population, not a single genome. Whether a mutation is present in every cancer cell or only a minority determines whether targeting it will work, and subclonal architecture predicts which resistance mechanisms are already latent before treatment begins.

WHAT IT ANSWERS

Which mutations are clonal and which are subclonal, how many distinct populations does this tumour contain, and in what order did they arise?

WHEN IT IS DONE

In research and translational studies of resistance and evolution, in multi-region and longitudinal sampling, and where treatment failure needs a mechanistic explanation.

WHAT TRIGGERS IT

A research or translational protocol on tumour evolution, multi-region or serial sampling, or investigation of unexpected treatment resistance.

HOW IT IS DONE

Purity and allele-specific copy number are established first, then mutation allele fractions are converted to cancer cell fractions with copy state and multiplicity accounted for. Fractions are clustered to infer populations, phylogenetic relationships are reconstructed, and driver mutations are placed on the resulting tree.

WHAT IT PRODUCES

Cancer cell fraction per mutation with bounds, clonal and subclonal classification, the inferred population structure, a phylogenetic reconstruction with drivers placed, and a detection floor statement.

WHAT IT DETECTS

Cancer cell fraction per mutation, clonal against subclonal classification, the number and size of distinct subclonal populations, and the inferred order in which they arose.

WHAT IT DOES NOT DETECT

Bulk sequencing infers populations from allele fraction clustering rather than observing cells, so distinct subclones with similar fractions merge into one cluster. Spatially separated subclones absent from the sampled region are invisible, and small subclones fall below the detection floor entirely.

WHAT MAKES IT HARD

Everything depends on purity and copy number being right, since cancer cell fraction is computed through them. Cluster number is a modelling choice rather than an observation, and over-fitting produces subclones that do not exist while under-fitting merges ones that do.

WHO DOES THIS JOB

Computational cancer biologist or senior clinical bioinformatician, predominantly in research and translational settings.

WHAT YOU CAN DO AFTERWARDS

Convert allele fractions to cancer cell fractions correctly, infer population structure defensibly, reconstruct evolutionary relationships, and state the detection floor honestly.

WHO HIRES FOR IT

Cancer research institutes, academic cancer centres, pharmaceutical translational oncology, and resistance-focused research programmes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
20	**98 hours · 17 days**	Prior purity, ploidy and somatic calling experience	**2,25,000**

SON-16
WORKFLOW

Subclonal Reconstruction & Tumour Evolution Analysis

(**Clonality and cancer cell fraction analysis** · Clonality analysis · CCF analysis)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Input adequacy assessment**	Mutation count and coverage depth assessed as adequate for clustering	Adequacy review	*Input adequacy record*
02	**Purity and ploidy input**	Fitted purity and ploidy applied as the scaling inputs for fraction calculation	Purity-ploidy solution	*Scaling parameters*
03	**Allele-specific copy number input**	Allele-specific copy states applied per mutation position	Allele-specific CN	*Copy state per mutation*
04	**High-confidence mutation selection**	Mutations restricted to high-confidence calls with adequate depth	Filtering procedure	*Selected mutation set*
05	**Multiplicity estimation**	The number of mutant copies per cell estimated for each mutation	Multiplicity estimation	*Multiplicity assignments*
06	**Cancer cell fraction calculation**	Allele fractions converted to cancer cell fractions using purity, copy state and multiplicity	CCF calculation	*Cancer cell fractions*
07	**Confidence interval derivation**	Confidence bounds derived per mutation from read depth and copy state	Confidence estimation	*CCF bounds*
08	**Detection floor determination**	The minimum detectable cancer cell fraction derived from depth and purity	Sensitivity calculation	*Detection floor*
09	**Clustering**	Cancer cell fractions clustered to infer distinct subclonal populations	Clustering algorithm	*Population clusters*
10	**Cluster number selection**	The number of populations selected with over-fitting and under-fitting assessed	Model selection	*Selected cluster count*
11	**Clonal and subclonal classification**	Mutations classified as clonal or subclonal against the cluster structure	Classification logic	*Clonality assignments*
12	**Population size estimation**	The proportion of cancer cells in each population estimated	Population estimation	*Population sizes*
13	**Phylogenetic reconstruction**	Ancestral relationships between populations inferred from fraction structure	Phylogenetic inference	*Population tree*
14	**Driver placement**	Driver mutations placed on the reconstructed tree to infer event ordering	Driver mapping	*Ordered driver events*
15	**Multi-region integration**	Regions or timepoints integrated where multiple samples are available	Multi-sample integration	*Integrated structure*
16	**Sampling limitation statement**	The invisibility of unsampled spatial populations stated explicitly	Limitation template	*Limitation statement*
17	**Resistance potential assessment**	Subclonal populations carrying known resistance alterations identified	Resistance annotation	*Resistance assessment*
18	**Therapy implication interpretation**	Clonality interpreted for the likely durability of targeted therapy	Clinical interpretation	*Therapy implications*
19	**Visualisation preparation**	Fraction distributions and tree structures prepared for interpretation	Visualisation	*Figures*
20	**Reporting**	Fractions, populations, tree and detection floor assembled	Report template	*Clonality report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**98 h · 17 d**	**SON-16**	**2,25,000**

SON-17
SECTOR 01

Clonal Haematopoiesis Analysis

(**CHIP detection and interpretation** · CHIP analysis)

WHY IT IS DONE

Blood-derived clonal mutations mimic tumour mutations in liquid biopsy and appear as somatic findings in matched normals, corrupting both. They also carry independent risk of haematological malignancy and cardiovascular disease, so detecting them matters in their own right.

WHAT IT ANSWERS

Are the low-fraction variants observed in this blood sample derived from clonal haematopoiesis rather than from a solid tumour, and do they carry independent clinical significance?

WHEN IT IS DONE

In every circulating tumour DNA analysis to avoid misattribution, when interpreting a matched normal that shows somatic-looking variants, and in dedicated risk assessment.

WHAT TRIGGERS IT

Low-fraction variants in plasma or in a matched normal, an older patient undergoing liquid biopsy, or a dedicated haematological risk assessment.

HOW IT IS DONE

Deep sequencing of the white cell fraction is analysed in parallel with plasma, low-fraction calling is applied against an empirical error model, and variants are assessed against the characteristic gene set and variant spectrum. Concordance between compartments assigns origin, and clone size is estimated and tracked.

WHAT IT PRODUCES

Variants with allele fractions in both compartments, origin assignment with the concordance evidence, clone size estimates, an exclusion list for liquid biopsy interpretation, and haematological risk assessment where indicated.

WHAT IT DETECTS

Low-fraction variants in genes characteristic of clonal haematopoiesis, their allele fractions, and concordance between plasma and white cell fractions that identifies blood origin.

WHAT IT DOES NOT DETECT

It cannot definitively assign origin from plasma alone, since a genuine tumour mutation may occur in the same genes. Without a paired white cell sample the distinction is probabilistic, and small clones fall below detection while remaining biologically real.

WHAT MAKES IT HARD

The whole analysis lives at low allele fraction where error dominates, so an empirical error model is essential. Without a paired white cell sample origin assignment is inference rather than determination, and this limitation is routinely glossed over in liquid biopsy reporting.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics or haematology, mid level, usually alongside liquid biopsy work.

WHAT YOU CAN DO AFTERWARDS

Detect low-fraction clonal variants against an empirical error floor, assign origin using paired compartment evidence, and apply exclusions correctly in liquid biopsy interpretation.

WHO HIRES FOR IT

Liquid biopsy providers, haematology services, cancer centres running plasma programmes, and cardiovascular risk research groups.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
16	**78 hours · 13 days**	Prior low-fraction variant calling experience	**1,80,000**

SON-17
WORKFLOW

Clonal Haematopoiesis Analysis

(**CHIP detection and interpretation** · CHIP analysis)

COMPLETE WORKFLOW — 16 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Gene set specification**	The clonal haematopoiesis gene set and variant spectrum specified and recorded	Gene set register	*Authorised gene set*
02	**Paired sample availability assessment**	Availability of a white cell fraction assessed, since origin assignment depends on it	Sample inventory	*Configuration record*
03	**White cell fraction sequencing**	The cellular fraction sequenced deeply in parallel with plasma where available	Deep sequencing	*White cell data*
04	**Coverage adequacy assessment**	Depth confirmed adequate at gene set positions for low-fraction calling	mosdepth	*Coverage record*
05	**Error model construction**	Background error rates measured empirically at known-negative positions	Error profiling	*Empirical error model*
06	**Low-fraction variant calling**	Variants called on statistical significance against the error model	Low-fraction caller	*Candidate variants*
07	**Artefact filtering**	Strand and position bias artefacts removed before interpretation	Bias metrics	*Filtered candidates*
08	**Read-level review**	Surviving candidates inspected at read level individually	IGV review	*Reviewed candidates*
09	**Gene and spectrum assessment**	Variants assessed against the characteristic gene set and mutation spectrum	Spectrum comparison	*Spectrum assessment*
10	**Compartment concordance analysis**	Allele fractions compared between plasma and white cell fractions	Concordance analysis	*Concordance results*
11	**Origin assignment**	Variants assigned haematopoietic or tumour origin with the evidence recorded	Assignment logic	*Origin assignments*
12	**Clone size estimation**	Clone size estimated from allele fraction with confidence bounds	Fraction estimation	*Clone size estimates*
13	**Liquid biopsy exclusion list generation**	Haematopoietic variants compiled for exclusion from tumour interpretation	Exclusion list generation	*Exclusion list*
14	**Serial trajectory analysis**	Clone size tracked across timepoints where serial samples exist	Trajectory analysis	*Trajectory results*
15	**Haematological risk assessment**	Findings assessed against criteria for haematological malignancy risk	Risk criteria	*Risk assessment*
16	**Reporting**	Variants, origin assignment, clone sizes and exclusions assembled	Report template	*CHIP report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
16	**78 h · 13 d**	**SON-17**	**1,80,000**

SON-18
SECTOR 01

Circulating Tumour DNA Analysis

(**ctDNA and liquid biopsy analysis** · ctDNA analysis · liquid biopsy)

WHY IT IS DONE

Plasma testing gives a molecular result without a biopsy, samples the whole tumour burden rather than one site, and can be repeated as often as clinically useful. It is now standard where tissue is unavailable and increasingly used alongside tissue rather than instead of it.

WHAT IT ANSWERS

What tumour-derived alterations are detectable in this plasma sample, at what allele fraction, and what is the tumour fraction of the circulating DNA?

WHEN IT IS DONE

When tissue is unobtainable or exhausted, at progression to identify resistance mechanisms, for serial monitoring during treatment, and where a rapid molecular result is needed.

WHAT TRIGGERS IT

Insufficient or unobtainable tissue, disease progression requiring resistance assessment, serial monitoring on therapy, or an urgent treatment decision.

HOW IT IS DONE

Molecular identifiers are extracted and consensus reads constructed to suppress error, an empirical background model is built, and low-fraction calling proceeds against it. Clonal haematopoiesis variants are excluded using the white cell fraction, tumour fraction is estimated, and the achieved limit of detection is stated per target region.

WHAT IT PRODUCES

Consensus-based variant calls with allele fractions, the background error model, the clonal haematopoiesis exclusion record, tumour fraction estimate, a per-region limit of detection statement, and a signed report.

WHAT IT DETECTS

Somatic variants at very low allele fraction using molecular consensus, copy number changes at sufficient tumour fraction, selected fusions and resistance mutations, and estimated tumour fraction.

WHAT IT DOES NOT DETECT

A negative plasma result does not exclude disease, because shedding varies enormously by tumour type, site and burden — low-shedding tumours and early-stage disease frequently yield nothing. Clonal haematopoiesis variants masquerade as tumour signal, and copy number and structural analysis require far higher tumour fraction than point mutation detection.

WHAT MAKES IT HARD

Sensitivity is the entire assay, and it depends on input material, molecular recovery, consensus depth and the error floor together. The pre-analytical variables dominate — collection tube, time to processing, plasma volume — so an analysis that ignores them reports a sensitivity the sample never supported.

WHO DOES THIS JOB

Senior clinical bioinformatician in liquid biopsy or cancer genomics. A specialist and heavily recruited skill set.

WHAT YOU CAN DO AFTERWARDS

Build molecular consensus, construct empirical error models, exclude clonal haematopoiesis correctly, estimate tumour fraction, and state achieved sensitivity per region.

WHO HIRES FOR IT

Liquid biopsy companies, comprehensive cancer centres, pharmaceutical monitoring programmes, and diagnostic chains offering plasma testing.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
25	**122 hours · 21 days**	Prior somatic analysis and molecular consensus experience	**2,80,000**

SON-18
WORKFLOW

Circulating Tumour DNA Analysis

(**ctDNA and liquid biopsy analysis** · ctDNA analysis · liquid biopsy)

COMPLETE WORKFLOW — 25 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Pre-analytical record capture**	Collection tube, time to processing and plasma volume recorded as analytical inputs	Pre-analytical log	*Pre-analytical record*
02	**Input quantity assessment**	Extracted cell-free DNA quantity measured and converted to genome equivalents	Quantification	*Input quantity record*
03	**Fragment profile assessment**	Fragment size profile assessed for genomic DNA contamination from cell lysis	Fragment analysis	*Fragment profile*
04	**Library and molecular identifier verification**	Molecular identifier structure verified as parseable before processing	Structure verification	*UMI structure record*
05	**Data receipt and quality assessment**	Integrity verified and sequence quality measured against acceptance criteria	md5sum, FastQC	*QC acceptance record*
06	**Molecular identifier extraction**	Identifiers extracted to read tags with spacers removed and pairing preserved	fgbio, UMI-tools	*Tagged reads*
07	**Alignment retaining tags**	Reads aligned with molecular tags preserved through alignment	bwa-mem2, ZipperBams	*Tagged BAM*
08	**Molecular family grouping**	Reads grouped into families by identifier and position with edits allowed	GroupReadsByUmi	*Grouped BAM*
09	**Family size distribution analysis**	Family sizes and singleton fraction analysed against expected molecular recovery	Distribution analysis	*Family metrics*
10	**Consensus read construction**	Single-strand and duplex consensus reads constructed from families	CallDuplexConsensusReads	*Consensus reads*
11	**Consensus filtering**	Consensus reads filtered on family size, quality and error rate	FilterConsensusReads	*Filtered consensus BAM*
12	**Overlap clipping**	Overlapping mate regions clipped to prevent double counting of molecules	ClipBam	*Clipped BAM*
13	**Molecular recovery assessment**	Unique molecules recovered compared against input genome equivalents	Recovery calculation	*Recovery record*
14	**Background error model construction**	Per-context error rates measured on consensus reads at negative positions	Error profiling	*Empirical error model*
15	**Consensus depth assessment**	Consensus and duplex depth measured per target region	Depth analysis	*Consensus depth record*
16	**Low-fraction variant calling**	Variants called on significance against the error model rather than by threshold	Low-fraction caller	*Candidate variants*
17	**Artefact and bias filtering**	Strand, position and context artefacts removed from candidate calls	Bias metrics	*Filtered candidates*
18	**Read-level review**	Every reportable call inspected at read level in consensus data	IGV review	*Reviewed calls*
19	**Clonal haematopoiesis exclusion**	Haematopoietic variants excluded using the white cell fraction comparison	Exclusion list application	*Tumour-attributed calls*
20	**Tumour fraction estimation**	Circulating tumour fraction estimated from variant fractions and copy number	Tumour fraction estimation	*Tumour fraction*
21	**Copy number assessment**	Copy number assessed where tumour fraction supports it, otherwise declared uninformative	Plasma copy number analysis	*Copy number findings*
22	**Per-region detection limit statement**	Achieved limit of detection stated per target region from consensus depth	Sensitivity calculation	*Per-region LOD*
23	**Annotation and oncogenicity assessment**	Variants annotated and assessed for oncogenic effect against curated evidence	VEP, OncoKB	*Oncogenicity calls*
24	**Clinical tiering and therapy matching**	Findings tiered and matched against approved therapies and trials	AMP/ASCO/CAP, therapy resources	*Tiered findings*
25	**Report generation and sign-out**	Findings, tumour fraction, sensitivity and limitations issued under dual signature	Validated report template	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
25	**122 h · 21 d**	**SON-18**	**2,80,000**

SON-19
SECTOR 01

Tumour-Informed MRD Analysis

(**Bespoke panel minimal residual disease analysis** · MRD · tumour-informed MRD)

WHY IT IS DONE

Detecting residual disease after curative treatment identifies recurrence months before imaging and selects patients who need further therapy from those who do not. Tracking mutations already known to be in the patient's tumour pushes sensitivity below one tumour molecule in ten thousand.

WHAT IT ANSWERS

Is tumour-derived DNA still detectable in this patient's plasma after treatment, at what level, and does the trajectory across timepoints indicate recurrence?

WHEN IT IS DONE

After curative-intent surgery or chemoradiotherapy, at defined surveillance intervals thereafter, and to guide decisions on adjuvant therapy escalation or de-escalation.

WHAT TRIGGERS IT

Completion of curative-intent treatment, a scheduled surveillance timepoint, or an adjuvant therapy decision point.

HOW IT IS DONE

The primary tumour is sequenced and clonal trunk variants are selected against strict criteria, avoiding clonal haematopoiesis genes and germline-suspicious sites. A bespoke panel is designed and validated, plasma is sequenced very deeply with molecular consensus, and detection is called against a statistical threshold derived from the assay's error model.

WHAT IT PRODUCES

The tracked variant panel with selection rationale, per-timepoint detection status with quantified levels, the statistical detection framework, serial trajectory analysis, and a clinical report with recurrence implications.

WHAT IT DETECTS

Tumour-derived molecules at extremely low fraction using patient-specific tracked variants, quantified residual disease level, and trajectory across serial timepoints.

WHAT IT DOES NOT DETECT

It only tracks variants present in the sequenced tumour region, so a recurrence driven by a subclone absent from the original specimen is missed. Low-shedding disease and certain anatomical sites yield no plasma signal, and a negative result reduces but never eliminates recurrence probability.

WHAT MAKES IT HARD

Variant selection determines everything downstream. Choosing subclonal variants produces false negatives at recurrence, and including clonal haematopoiesis or germline sites produces false positives. Detection at this depth is a statistical call, not an observation, so the threshold must be derived rather than asserted.

WHO DOES THIS JOB

Senior clinical bioinformatician in minimal residual disease or liquid biopsy. Among the most sought-after skill sets in commercial oncology diagnostics.

WHAT YOU CAN DO AFTERWARDS

Select trackable variants defensibly, design and validate a bespoke panel, apply statistical detection frameworks, and interpret serial trajectories.

WHO HIRES FOR IT

Minimal residual disease companies, comprehensive cancer centres, pharmaceutical adjuvant trial programmes, and surveillance service providers.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
26	**128 hours · 22 days**	Prior ctDNA and somatic analysis experience	**2,90,000**

SON-19
WORKFLOW

Tumour-Informed MRD Analysis

(**Bespoke panel minimal residual disease analysis** · MRD · tumour-informed MRD)

COMPLETE WORKFLOW — 26 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Programme enrolment and timepoint scheduling**	Enrolment confirmed with the surveillance timepoint schedule recorded	Programme protocol	*Enrolment record*
02	**Tumour specimen adequacy assessment**	Primary tumour specimen assessed as adequate for variant discovery	Pathology review	*Adequacy record*
03	**Tumour sequencing**	The primary tumour sequenced to identify candidate trackable variants	Tumour sequencing	*Tumour variant set*
04	**Germline sequencing**	A germline sample sequenced to exclude inherited variants from tracking	Germline sequencing	*Germline variant set*
05	**Somatic variant identification**	Somatic variants identified by germline subtraction with high confidence required	Somatic calling	*Somatic call set*
06	**Clonality assessment for selection**	Cancer cell fractions computed so only clonal trunk variants are selected	CCF calculation	*Clonality assignments*
07	**Clonal haematopoiesis gene exclusion**	Variants in haematopoietic clonal genes excluded from candidate tracking	Gene exclusion	*Filtered candidates*
08	**Germline-suspicious exclusion**	Variants with germline-consistent characteristics excluded from tracking	Exclusion logic	*Cleared candidates*
09	**Trackability assessment**	Candidates assessed for assay design feasibility and sequence context	Design feasibility review	*Trackable candidates*
10	**Variant selection and rationale recording**	The final tracked variant set selected with selection rationale documented	Selection procedure	*Tracked variant panel*
11	**Bespoke panel design**	A patient-specific panel designed against the selected variants	Panel design	*Bespoke panel design*
12	**Panel validation**	The bespoke panel validated for coverage, uniformity and detection performance	Validation procedure	*Panel validation record*
13	**Pre-analytical record capture**	Collection and processing variables recorded for every plasma timepoint	Pre-analytical log	*Pre-analytical records*
14	**Plasma input quantification**	Cell-free DNA input measured and converted to genome equivalents per timepoint	Quantification	*Input records*
15	**Deep plasma sequencing**	Plasma sequenced to very high depth across the bespoke panel	Deep sequencing	*Plasma data*
16	**Molecular consensus construction**	Consensus reads constructed to suppress sequencing error	fgbio consensus	*Consensus BAM*
17	**Background error model construction**	Error rates measured at tracked positions in control material	Error profiling	*Empirical error model*
18	**Per-variant evidence quantification**	Supporting molecule counts quantified at every tracked variant position	Molecule counting	*Evidence counts*
19	**Statistical detection framework application**	Detection called against a statistical threshold derived from the error model	Statistical framework	*Detection call*
20	**Detection threshold derivation record**	The derivation of the detection threshold recorded rather than asserted	Derivation documentation	*Threshold derivation*
21	**Quantification of residual disease**	Disease level quantified in tumour molecules per millilitre or equivalent units	Quantification	*Residual disease level*
22	**Per-timepoint sensitivity statement**	Achieved sensitivity stated per timepoint against molecules interrogated	Sensitivity calculation	*Timepoint sensitivity*
23	**Serial trajectory analysis**	Detection status and level analysed across timepoints against treatment	Trajectory analysis	*Trajectory results*
24	**Recurrence interpretation**	Trajectory interpreted for recurrence probability and lead time	Clinical interpretation	*Recurrence interpretation*
25	**Report generation**	Detection status, level, sensitivity and trajectory assembled per timepoint	Validated report template	*Draft report*
26	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed clinical report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
26	**128 h · 22 d**	**SON-19**	**2,90,000**

SON-20
SECTOR 01

Tumour-Naive & Methylation MRD Analysis

(**Fixed-panel residual disease analysis** · Fixed-panel MRD)

WHY IT IS DONE

Bespoke panel design needs tumour tissue and several weeks of turnaround, neither of which is always available. A fixed panel combining recurrent mutations with tumour-associated methylation gives residual disease detection without any tumour sequencing, and works from the first timepoint.

WHAT IT ANSWERS

Is residual or recurrent disease detectable using a fixed assay, without reference to this patient's own tumour sequence?

WHEN IT IS DONE

Where tumour tissue is unavailable or inadequate for bespoke design, where immediate monitoring is required, and in surveillance programmes covering many patients uniformly.

WHAT TRIGGERS IT

Unavailable or inadequate tumour tissue, a requirement for immediate monitoring, or a programme needing one uniform assay across a cohort.

HOW IT IS DONE

Plasma is processed for both mutation and methylation analysis from the same input, mutation calling proceeds with molecular consensus against an error model, methylation signal is scored against tumour-associated regions, and the two analytes are combined into a single classifier with a validated threshold.

WHAT IT PRODUCES

Mutation and methylation results separately and combined, the classifier score against the validated threshold, sensitivity statement relative to tumour-informed methods, and serial trajectory across timepoints.

WHAT IT DETECTS

Recurrent mutations from the fixed panel, tumour-associated methylation signal, combined multi-analyte detection status, and quantified disease level.

WHAT IT DOES NOT DETECT

Sensitivity is materially lower than tumour-informed tracking because the panel is not personalised. Tumours lacking any tracked mutation or methylation feature are undetectable, methylation signal is tissue-dependent and confounded by inflammation, and clonal haematopoiesis again mimics tumour signal.

WHAT MAKES IT HARD

Combining two analytes with different error structures into one defensible call is the analytical problem, and the classifier threshold must be validated on real longitudinal outcomes rather than tuned on spiked material. Sensitivity relative to tumour-informed assays must be stated honestly rather than implied as equivalent.

WHO DOES THIS JOB

Senior clinical bioinformatician with methylation and classification competence, in liquid biopsy or surveillance settings.

WHAT YOU CAN DO AFTERWARDS

Run multi-analyte residual disease detection, combine mutation and methylation evidence into a validated classifier, and state sensitivity relative to tumour-informed alternatives.

WHO HIRES FOR IT

Liquid biopsy and early detection companies, cancer surveillance programmes, comprehensive cancer centres, and diagnostic chains.

COMMERCIALS

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

SON-20
WORKFLOW

Tumour-Naive & Methylation MRD Analysis

(**Fixed-panel residual disease analysis** · Fixed-panel MRD)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Assay scope confirmation**	The fixed mutation panel and methylation region set confirmed and recorded	Assay register	*Authorised assay version*
02	**Pre-analytical record capture**	Collection and processing variables recorded as they affect both analytes	Pre-analytical log	*Pre-analytical record*
03	**Plasma input quantification**	Cell-free DNA input measured and split appropriately between analytes	Quantification	*Input record*
04	**Dual library preparation verification**	Mutation and methylation libraries confirmed prepared from the same input	Library records	*Preparation record*
05	**Data receipt and quality assessment**	Integrity and sequence quality verified for both library types	md5sum, FastQC	*QC acceptance record*
06	**Molecular consensus construction**	Consensus reads constructed for the mutation arm to suppress error	fgbio consensus	*Consensus BAM*
07	**Mutation arm error modelling**	Background error rates measured empirically at panel positions	Error profiling	*Mutation error model*
08	**Mutation calling**	Variants called at panel positions against the error model	Low-fraction caller	*Mutation calls*
09	**Clonal haematopoiesis exclusion**	Haematopoietic variants excluded using the white cell comparison	Exclusion list application	*Tumour-attributed mutations*
10	**Methylation conversion quality assessment**	Conversion efficiency assessed for the methylation arm	Conversion controls	*Conversion QC*
11	**Methylation region scoring**	Tumour-associated methylation regions scored against the control distribution	Region scoring	*Methylation scores*
12	**Cell composition control**	Blood cell composition effects controlled before methylation interpretation	Deconvolution correction	*Corrected methylation*
13	**Methylation error and background modelling**	Background methylation signal modelled from control cohorts	Background modelling	*Methylation background model*
14	**Analyte combination**	Mutation and methylation evidence combined into a single classifier score	Combined classifier	*Classifier score*
15	**Threshold application**	The classifier score assessed against the outcome-validated threshold	Threshold comparison	*Detection call*
16	**Classifier validation record**	The outcome-based validation of the threshold recorded rather than assumed	Validation documentation	*Validation record*
17	**Quantification**	Residual disease level quantified from the combined evidence	Quantification	*Disease level*
18	**Relative sensitivity statement**	Sensitivity stated relative to tumour-informed tracking rather than implied equivalent	Comparison data	*Relative sensitivity*
19	**Serial trajectory analysis**	Detection status and level analysed across timepoints	Trajectory analysis	*Trajectory results*
20	**Clinical interpretation**	Results interpreted for recurrence risk with limitations stated	Clinical interpretation	*Interpretation record*
21	**Report generation**	Combined results, threshold and relative sensitivity assembled	Validated report template	*Draft report*
22	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed clinical report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**106 h · 18 d**	**SON-20**	**2,45,000**

SON-21
SECTOR 01

cfDNA Fragmentomics Analysis

(**Cell-free DNA fragmentation analysis** · Fragmentomics)

WHY IT IS DONE

Tumour-derived fragments differ from normal ones in length, end position and genomic distribution because they reflect the chromatin of the cell they came from. This carries tumour signal independent of any mutation, and works at fractions too low for variant calling.

WHAT IT ANSWERS

Does the fragmentation profile of this plasma sample carry tumour signal, and what does the pattern imply about tumour presence, burden or tissue origin?

WHEN IT IS DONE

In early detection where mutation-based signal is too sparse, as a complement to mutation calling in low tumour fraction samples, and in tissue-of-origin inference.

WHAT TRIGGERS IT

An early detection or screening protocol, a mutation-negative plasma sample where suspicion remains, or a tissue-of-origin question.

HOW IT IS DONE

Pre-analytical variables are recorded because they confound the entire analysis, fragment lengths and end motifs are extracted genome-wide, and coverage periodicity is analysed for nucleosome positioning. Profiles are compared against a matched control cohort and combined into a classifier with validated thresholds.

WHAT IT PRODUCES

Fragment size and end motif profiles, nucleosome positioning analysis, genome-wide fragmentation profiles against controls, tissue-of-origin inference, and classifier scores with the pre-analytical record.

WHAT IT DETECTS

Fragment size distribution shifts, fragment end motif preferences, nucleosome positioning inferred from coverage periodicity, genome-wide fragmentation profiles, and tissue-of-origin signal derived from them.

WHAT IT DOES NOT DETECT

It gives no information about specific mutations or actionable targets. Signal is highly sensitive to pre-analytical handling, and processing delay produces fragmentation changes indistinguishable from tumour signal. Inflammation and other pathology shift profiles in similar directions.

WHAT MAKES IT HARD

The signal is real but fragile. Time from collection to plasma separation alters fragmentation as much as tumour presence does, so the pre-analytical record is not administrative overhead but an analytical prerequisite, and a control cohort processed identically is mandatory.

WHO DOES THIS JOB

Computational biologist or senior clinical bioinformatician in liquid biopsy and early detection. A newer specialism with few experienced practitioners.

WHAT YOU CAN DO AFTERWARDS

Extract and interpret fragmentation features, control for pre-analytical confounding, build classifiers against matched cohorts, and infer tissue of origin.

WHO HIRES FOR IT

Early cancer detection companies, liquid biopsy providers, screening programme research groups, and academic cancer centres.

COMMERCIALS

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

SON-21
WORKFLOW

cfDNA Fragmentomics Analysis

(**Cell-free DNA fragmentation analysis** · Fragmentomics)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Pre-analytical record capture**	Tube type, time to separation and centrifugation recorded as primary confounders	Pre-analytical log	*Pre-analytical record*
02	**Control cohort matching**	Control samples processed identically selected for comparison	Cohort selection	*Matched controls*
03	**Input quantification**	Cell-free DNA input measured and genomic contamination assessed	Quantification, fragment analysis	*Input record*
04	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
05	**Alignment**	Reads aligned with settings preserving accurate fragment endpoints	bwa-mem2	*Aligned BAM*
06	**Duplicate handling**	Duplicates handled without collapsing distinct fragments sharing endpoints	Duplicate policy	*Processed BAM*
07	**Fragment length extraction**	Fragment lengths extracted genome-wide from properly paired reads	Fragment extraction	*Fragment length data*
08	**Size distribution profiling**	Length distributions profiled and compared against matched controls	Distribution analysis	*Size profile*
09	**Short fragment ratio calculation**	The ratio of short to long fragments computed across genomic windows	Ratio calculation	*Fragment ratio profile*
10	**End motif extraction**	Sequence motifs at fragment ends extracted and frequencies computed	Motif extraction	*End motif profile*
11	**Nucleosome positioning inference**	Coverage periodicity analysed to infer nucleosome positioning	Periodicity analysis	*Nucleosome profile*
12	**Window protection scoring**	Protection scores computed at transcription start sites and regulatory regions	Protection scoring	*Protection profile*
13	**Genome-wide profile construction**	Fragmentation features assembled into genome-wide profiles	Profile construction	*Fragmentation profiles*
14	**Batch and technical correction**	Technical variation between processing batches corrected before classification	Batch correction	*Corrected profiles*
15	**Classifier application**	Profiles classified against reference cohorts with validated thresholds	Classifier	*Classifier score*
16	**Tissue-of-origin inference**	Tissue signal inferred from protection and positioning patterns	Inference model	*Tissue inference*
17	**Confounding assessment**	Inflammation, pregnancy and other confounders assessed against the result	Clinical correlation	*Confounding assessment*
18	**Reporting**	Profiles, classifier score, tissue inference and pre-analytical record assembled	Report template	*Fragmentomics report*

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

SON-22
SECTOR 01

cfDNA Methylation & Tissue-of-Origin Analysis

(**Plasma methylation and tissue-of-origin analysis** · MCED · TOO analysis)

WHY IT IS DONE

Methylation patterns are tissue-specific and are established early in tumour development, so plasma methylation detects cancer at lower burden than mutation and simultaneously indicates which organ it came from. This combination is what makes multi-cancer early detection possible at all.

WHAT IT ANSWERS

Does this plasma sample carry a cancer-associated methylation signal, and if so what tissue did the tumour-derived DNA originate from?

WHEN IT IS DONE

In multi-cancer early detection screening, in cancers of unknown primary, in low tumour fraction samples where mutation analysis fails, and in residual disease monitoring.

WHAT TRIGGERS IT

A screening or early detection protocol, cancer of unknown primary, a mutation-negative plasma sample with clinical suspicion, or residual disease surveillance.

HOW IT IS DONE

Plasma methylation is measured by conversion or enrichment, quality controlled for conversion efficiency and cell composition, and normalised with age and composition corrected. Cancer signal is classified against reference cohorts and tissue-of-origin predicted with ranked probabilities and calibrated confidence.

WHAT IT PRODUCES

Cancer signal detection with classifier score, tissue-of-origin ranking with probabilities, methylated tumour fraction, quality control covering conversion and composition, and calibrated confidence statements.

WHAT IT DETECTS

Cancer-associated methylation signal at low tumour fraction, tissue-of-origin prediction with ranked probability, and quantified methylated tumour fraction.

WHAT IT DOES NOT DETECT

It identifies neither specific mutations nor actionable targets. Tissue-of-origin prediction degrades sharply at low signal and between tissues with similar methylation, inflammation produces confounding signal, and age and cell composition shift baseline methylation substantially.

WHAT MAKES IT HARD

It is a classification problem resting entirely on reference cohort quality and covariate control. Age and blood composition shift methylation enough to misclassify healthy samples, and tissue-of-origin confidence must be calibrated to real outcomes rather than reported as a bare top-ranked prediction.

WHO DOES THIS JOB

Senior clinical bioinformatician or computational biologist with methylation and classification competence, in early detection settings.

WHAT YOU CAN DO AFTERWARDS

Run plasma methylation classification with covariate control, predict tissue of origin with calibrated confidence, and quantify methylated tumour fraction.

WHO HIRES FOR IT

Multi-cancer early detection companies, screening programmes, liquid biopsy providers, and cancer of unknown primary services.

COMMERCIALS

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

SON-22
WORKFLOW

cfDNA Methylation & Tissue-of-Origin Analysis

(**Plasma methylation and tissue-of-origin analysis** · MCED · TOO analysis)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Assay scope confirmation**	The methylation region set and classification target list confirmed	Assay register	*Authorised assay version*
02	**Pre-analytical record capture**	Collection and processing variables recorded as methylation confounders	Pre-analytical log	*Pre-analytical record*
03	**Input quantification**	Cell-free DNA input measured and converted to genome equivalents	Quantification	*Input record*
04	**Conversion or enrichment quality assessment**	Conversion efficiency or enrichment performance assessed against controls	Conversion controls	*Conversion QC*
05	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
06	**Methylation-aware alignment**	Reads aligned with a method appropriate to the conversion chemistry	Bismark, BWA-meth	*Aligned BAM*
07	**Duplicate handling**	Duplicates handled appropriately for low-input cell-free material	Duplicate policy	*Processed BAM*
08	**Methylation calling**	Per-site methylation levels called with coverage thresholds applied	Methylation extraction	*Methylation calls*
09	**Coverage adequacy assessment**	Coverage assessed at classification regions with inadequate regions declared	Coverage analysis	*Coverage record*
10	**Cell composition estimation**	Blood cell composition estimated as a major confounder of plasma methylation	Deconvolution	*Composition estimates*
11	**Age and covariate correction**	Age and composition effects corrected before classification	Covariate correction	*Corrected methylation*
12	**Batch effect control**	Processing batch effects assessed and corrected against reference cohorts	Batch correction	*Corrected data*
13	**Reference cohort matching**	Age and composition matched references selected for comparison	Cohort selection	*Matched references*
14	**Cancer signal classification**	Profiles classified for cancer-associated signal against validated thresholds	Classifier	*Cancer signal score*
15	**Tissue-of-origin prediction**	Tissue of origin predicted with ranked probabilities across candidate tissues	Tissue classifier	*Ranked predictions*
16	**Confidence calibration**	Prediction confidence calibrated against real outcome data rather than model score	Calibration model	*Calibrated confidence*
17	**Methylated tumour fraction estimation**	The tumour-derived methylated fraction estimated with bounds	Fraction estimation	*Tumour fraction*
18	**Confounding assessment**	Inflammation and non-malignant pathology assessed as alternative explanations	Clinical correlation	*Confounding assessment*
19	**Clinical interpretation**	Results interpreted for the referral question with limitations stated	Clinical interpretation	*Interpretation record*
20	**Reporting**	Signal score, tissue ranking, calibrated confidence and quality assembled	Report template	*Methylation report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**98 h · 17 d**	**SON-22**	**2,25,000**

SON-23
SECTOR 01

Resistance Mutation Monitoring Analysis

(**Serial resistance tracking analysis** · Resistance monitoring)

WHY IT IS DONE

Targeted therapies fail through identifiable genetic mechanisms, and those mechanisms often appear in plasma before radiological progression. Detecting the specific resistance alteration determines whether a next-line targeted option exists or whether the class is exhausted.

WHAT IT ANSWERS

Has a resistance mechanism emerged in this patient on targeted therapy, which one, at what level, and does a next-line targeted option address it?

WHEN IT IS DONE

At intervals during targeted therapy, at the first sign of biochemical or radiological progression, and before selecting a next-line agent.

WHAT TRIGGERS IT

Scheduled monitoring on targeted therapy, rising tumour markers, early radiological progression, or a next-line treatment decision.

HOW IT IS DONE

A baseline profile is established for comparison, plasma is sequenced deeply with molecular consensus at each timepoint, and known resistance loci are interrogated at high sensitivity. Emergent alterations are identified against baseline, clonal haematopoiesis is excluded, levels are quantified and trajectories analysed against therapy timing.

WHAT IT PRODUCES

Per-timepoint resistance findings with allele fractions, the emergence analysis against baseline, trajectory plots aligned to treatment, next-line therapy implications, and a serial monitoring report.

WHAT IT DETECTS

Known resistance mutations at low allele fraction, gene amplification-mediated resistance, emergent alterations absent at baseline, and the trajectory of each across timepoints.

WHAT IT DOES NOT DETECT

Non-genetic resistance including phenotypic transformation, drug efflux and pathway rewiring is entirely invisible, and these account for a substantial share of treatment failure. Resistance confined to a non-shedding site is missed, and low tumour fraction suppresses emerging subclones below detection.

WHAT MAKES IT HARD

Distinguishing a genuinely emergent subclone from a variant that was always present below the previous timepoint's detection limit requires the detection limit at each timepoint to be known and stated. Without it, emergence is an artefact of varying sensitivity rather than a biological observation.

WHO DOES THIS JOB

Clinical bioinformatician in liquid biopsy or cancer genomics, mid to senior level, working closely with treating oncologists.

WHAT YOU CAN DO AFTERWARDS

Run serial plasma monitoring, distinguish true emergence from sensitivity variation, quantify resistance trajectories, and link findings to next-line options.

WHO HIRES FOR IT

Cancer centres running targeted therapy programmes, liquid biopsy providers, pharmaceutical resistance research groups, and oncology diagnostic laboratories.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**86 hours · 15 days**	Prior ctDNA and somatic analysis experience	**2,00,000**

SON-23
WORKFLOW

Resistance Mutation Monitoring Analysis

(**Serial resistance tracking analysis** · Resistance monitoring)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Baseline profile establishment**	The pre-treatment molecular profile established as the emergence comparator	Baseline analysis	*Baseline profile*
02	**Resistance locus specification**	Known resistance loci for the therapy in use specified and recorded	Resistance locus register	*Locus specification*
03	**Timepoint scheduling**	Monitoring timepoints scheduled against treatment and clinical milestones	Monitoring protocol	*Timepoint schedule*
04	**Pre-analytical record capture**	Collection and processing variables recorded at every timepoint	Pre-analytical log	*Pre-analytical records*
05	**Plasma input quantification**	Cell-free DNA input measured per timepoint as a sensitivity determinant	Quantification	*Input records*
06	**Deep plasma sequencing**	Plasma sequenced deeply across resistance loci at each timepoint	Deep sequencing	*Plasma data*
07	**Molecular consensus construction**	Consensus reads constructed to suppress sequencing error	fgbio consensus	*Consensus BAM*
08	**Error model construction**	Background error rates measured at monitored positions	Error profiling	*Empirical error model*
09	**Resistance locus interrogation**	Known resistance positions interrogated at maximum sensitivity	Targeted calling	*Resistance calls*
10	**Amplification-mediated resistance assessment**	Copy number gain at resistance-associated genes assessed where fraction permits	Plasma copy number analysis	*Amplification findings*
11	**Clonal haematopoiesis exclusion**	Haematopoietic variants excluded from resistance interpretation	Exclusion list application	*Tumour-attributed calls*
12	**Per-timepoint detection limit derivation**	The detection limit derived at each timepoint from input and consensus depth	Sensitivity calculation	*Timepoint LOD*
13	**Emergence analysis**	New findings distinguished from previously undetectable ones using timepoint limits	Emergence logic	*Emergence assessment*
14	**Level quantification**	Resistance variant levels quantified per timepoint with bounds	Quantification	*Level estimates*
15	**Trajectory analysis**	Levels analysed across timepoints aligned to treatment and imaging	Trajectory analysis	*Trajectory results*
16	**Mechanism interpretation**	Findings interpreted as on-target, bypass or transformation-associated where identifiable	Mechanism interpretation	*Mechanism assessment*
17	**Next-line option assessment**	Findings matched against next-line targeted options addressing the mechanism	Therapy resources	*Next-line options*
18	**Reporting**	Findings, emergence assessment, trajectory and options assembled per timepoint	Report template	*Monitoring report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**86 h · 15 d**	**SON-23**	**2,00,000**

SON-24
SECTOR 01

RNA Fusion Analysis

(**RNA-based gene fusion detection** · Fusion analysis · RNA fusion panel)

WHY IT IS DONE

Fusion oncogenes are among the most actionable alterations in oncology, with approved therapies across several fusion families and tumour types. RNA detects the expressed fusion transcript directly, which DNA analysis can only infer from a breakpoint.

WHAT IT ANSWERS

Does this tumour express an oncogenic fusion transcript, which partners are involved, is the fusion in frame, and is it targetable?

WHEN IT IS DONE

In lung, thyroid, sarcoma and paediatric tumours as standard, in any tumour where a targetable fusion family is plausible, and where DNA analysis has found a rearrangement needing expression confirmation.

WHAT TRIGGERS IT

A tumour type with characteristic fusions, consideration of a fusion-targeted therapy, or a DNA rearrangement requiring transcript-level confirmation.

HOW IT IS DONE

RNA quality and library metrics are assessed before interpretation, reads are aligned with a fusion-aware method, and multiple callers are run for complementary sensitivity. Calls are merged, artefacts and read-through transcripts are filtered, breakpoints and reading frame are assessed, and expression imbalance is examined for unpartnered driver expression.

WHAT IT PRODUCES

Fusion calls with partners, breakpoints and supporting reads, reading frame assessment, the RNA quality record bounding the negative, expression imbalance findings, and tiered actionable fusion results.

WHAT IT DETECTS

Expressed fusion transcripts with partner genes and breakpoints, reading frame status, novel partners of known oncogenic drivers, and expression imbalance suggesting an unidentified fusion.

WHAT IT DOES NOT DETECT

Fusions not expressed in the sampled material, and any fusion in a sample whose RNA has degraded — which is common in formalin-fixed tissue and is the dominant failure mode. Enrichment-based panels detect only designed partners, and promoter-swap events without a chimeric transcript are missed.

WHAT MAKES IT HARD

A negative fusion result is only meaningful if RNA quality supported detection, so the quality metrics are part of the result rather than preliminary housekeeping. Read-through transcripts and alignment artefacts generate convincing false fusions that require filtering and review.

WHO DOES THIS JOB

Clinical bioinformatician with RNA competence in cancer genomics, mid level.

WHAT YOU CAN DO AFTERWARDS

Run fusion detection with multiple callers, filter artefacts and read-through events, bound negatives by RNA quality, and assess reading frame and actionability.

WHO HIRES FOR IT

Oncology diagnostic laboratories, lung and sarcoma specialist services, paediatric oncology, and comprehensive profiling providers.

COMMERCIALS

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

SON-24
WORKFLOW

RNA Fusion Analysis

(**RNA-based gene fusion detection** · Fusion analysis · RNA fusion panel)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Assay design verification**	The fusion detection design and its partner coverage confirmed and recorded	Design register	*Authorised design version*
02	**Specimen and RNA quality assessment**	RNA integrity and fragmentation assessed as the bound on any negative result	RNA quality metrics	*RNA quality record*
03	**Input quantity assessment**	RNA input quantity measured against the assay's validated minimum	Quantification	*Input record*
04	**Data receipt and quality assessment**	Integrity and sequence 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	**Library metric assessment**	Duplication, complexity and on-target metrics assessed for the RNA library	RSeQC, Qualimap	*Library metrics*
07	**Fusion-aware alignment**	Reads aligned with a method retaining chimeric alignments	STAR chimeric mode	*Aligned BAM*
08	**Multi-caller fusion detection**	Fusions detected using multiple callers for complementary sensitivity	Arriba, STAR-Fusion, FusionCatcher	*Per-caller fusion sets*
09	**Call merging and agreement recording**	Calls merged with caller agreement recorded as a confidence signal	Merging procedure	*Merged fusion set*
10	**Read-through and artefact filtering**	Adjacent gene read-through and alignment artefacts removed	Filtering criteria	*Filtered fusions*
11	**Supporting read review**	Split and spanning read evidence reviewed for every reportable fusion	IGV, evidence review	*Reviewed fusions*
12	**Breakpoint and exon assessment**	Breakpoints assigned to exon boundaries with partner orientation determined	Breakpoint annotation	*Breakpoint assignments*
13	**Reading frame assessment**	Predicted transcripts assessed for reading frame and functional domain retention	Frame analysis	*Frame assessment*
14	**Expression imbalance analysis**	Driver gene expression imbalance assessed to detect unpartnered fusions	Imbalance analysis	*Imbalance findings*
15	**Known fusion catalogue matching**	Fusions matched against catalogued oncogenic fusion events	Fusion catalogues	*Catalogue matches*
16	**Negative bound statement**	Any negative result bounded explicitly by the measured RNA quality	Limitation template	*Bounded negative statement*
17	**Oncogenicity and tiering**	Fusions assessed for oncogenic effect and assigned clinical evidence tiers	OncoKB, AMP/ASCO/CAP	*Tiered findings*
18	**Orthogonal confirmation**	Reportable fusions confirmed by an independent method where required	RT-PCR, FISH	*Confirmation record*
19	**Report generation and sign-out**	Fusions, evidence, frame and RNA quality bound issued under dual signature	Validated report template	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SON-24**	**2,10,000**

SON-25
SECTOR 01

Tumour Transcriptome & Molecular Subtyping Analysis

(**Tumour RNA-Seq and expression subtyping** · Tumour RNA-Seq · expression subtyping)

WHY IT IS DONE

Expression subtypes predict prognosis and treatment response in ways mutation profiles do not, and several established subtyping schemes now guide clinical management. Transcriptome analysis also reveals pathway activation and immune context that genomic analysis cannot.

WHAT IT ANSWERS

What molecular subtype does this tumour belong to, which pathways are activated, and what is its immune microenvironment composition?

WHEN IT IS DONE

Where a validated subtyping scheme informs management, in immunotherapy candidate assessment, in cancers of unknown primary, and in translational research.

WHAT TRIGGERS IT

A tumour type with a clinically used subtyping scheme, immunotherapy assessment, unknown primary investigation, or a research protocol.

HOW IT IS DONE

RNA quality and tumour content are assessed, expression is quantified and normalised against an appropriate reference, and batch effects are controlled. Subtype is assigned using the validated classifier with confidence reported, pathway activity is scored, and immune composition is estimated by deconvolution against reference profiles.

WHAT IT PRODUCES

Normalised expression matrices, subtype assignment with confidence, pathway activation scores, immune deconvolution estimates, signature scores, and a report bounded by RNA quality and tumour content.

WHAT IT DETECTS

Gene expression profiles, assigned molecular subtype with confidence, pathway activation scores, immune cell composition by deconvolution, and immunotherapy-associated expression signatures.

WHAT IT DOES NOT DETECT

It cannot identify the genomic alteration driving an expression change, and subtype assignment degrades badly with degraded RNA or low tumour content. Bulk expression averages tumour and stroma together, so deconvolution estimates are model-dependent rather than measured.

WHAT MAKES IT HARD

Classifiers are trained on specific platforms and preparation methods, so applying them to different data without recalibration produces confident wrong subtypes. Tumour content confounds every expression measurement, and deconvolution is an estimate whose model dependence should be stated.

WHO DOES THIS JOB

Clinical bioinformatician or computational biologist with transcriptomics competence, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Quantify and normalise tumour expression, apply subtype classifiers with correct calibration, score pathway activity, and estimate immune composition honestly.

WHO HIRES FOR IT

Academic cancer centres, pharmaceutical translational oncology, comprehensive profiling providers, and precision oncology programmes.

COMMERCIALS

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

SON-25
WORKFLOW

Tumour Transcriptome & Molecular Subtyping Analysis

(**Tumour RNA-Seq and expression subtyping** · Tumour RNA-Seq · expression subtyping)

COMPLETE WORKFLOW — 20 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Classifier and platform verification**	The subtyping classifier and the platform it was trained on confirmed	Classifier register	*Authorised classifier version*
02	**Specimen and tumour content recording**	Tumour content recorded because expression is averaged across all cells present	Pathology report	*Content record*
03	**RNA quality assessment**	RNA integrity assessed as a determinant of classification reliability	RNA quality metrics	*RNA quality record*
04	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
05	**Read preprocessing**	Adapters and low-quality tails removed with strand preserved	fastp	*Trimmed FASTQ*
06	**Splice-aware alignment**	Reads aligned with an annotation-guided splice-aware method	STAR	*Aligned BAM*
07	**Library metric assessment**	Duplication, gene body coverage and strand specificity assessed	RSeQC, Qualimap	*Library metrics*
08	**Expression quantification**	Gene and transcript expression quantified against the reference annotation	featureCounts, Salmon	*Expression matrices*
09	**Normalisation**	Expression normalised using a method compatible with the classifier	Normalisation procedure	*Normalised expression*
10	**Batch effect assessment**	Batch structure assessed and corrected before classification	Batch correction	*Corrected expression*
11	**Classifier calibration check**	Compatibility between the data and the classifier's training platform verified	Calibration check	*Calibration record*
12	**Subtype assignment**	Molecular subtype assigned with confidence scores across candidate classes	Subtype classifier	*Subtype assignment*
13	**Low-confidence declaration**	Assignments below the confidence threshold declared unclassifiable	Threshold criteria	*Confidence decision*
14	**Pathway activity scoring**	Pathway activation scored using validated signature sets	Pathway scoring	*Pathway scores*
15	**Immune deconvolution**	Immune and stromal composition estimated against reference profiles	Deconvolution	*Composition estimates*
16	**Deconvolution model recording**	The reference profile set used for deconvolution recorded as a dependency	Model documentation	*Model record*
17	**Immunotherapy signature scoring**	Immunotherapy-associated expression signatures scored	Signature scoring	*Signature scores*
18	**Tumour content confounding assessment**	Expression results assessed against tumour content as a confounder	Confounding review	*Confounding assessment*
19	**Clinical interpretation**	Subtype and pathway findings interpreted for prognosis and treatment	Clinical interpretation	*Interpretation record*
20	**Reporting**	Subtype, confidence, pathway and immune results assembled with quality bounds	Report template	*Transcriptome report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
20	**96 h · 16 d**	**SON-25**	**2,20,000**

SON-26
SECTOR 01

Oncogenic Splice Variant Analysis

(**Splicing alteration and exon skipping analysis** · Exon skipping analysis)

WHY IT IS DONE

Some of the most actionable alterations in oncology are splicing events rather than protein-coding mutations, and several are directly targetable. DNA analysis predicts splice impact; RNA analysis demonstrates it, which is what a therapy decision requires.

WHAT IT ANSWERS

Does this tumour carry an oncogenic splicing alteration, is it expressed, and is the resulting transcript targetable?

WHEN IT IS DONE

Where a targetable splicing alteration is plausible for the tumour type, where DNA analysis found a variant with predicted splice impact, and in splicing factor mutated malignancies.

WHAT TRIGGERS IT

A tumour type with known targetable splicing events, a DNA variant predicted to affect splicing, or a splicing factor mutation requiring functional characterisation.

HOW IT IS DONE

RNA quality is assessed and splice-aware alignment performed, then splice junctions are quantified and compared against a matched control cohort. Aberrant events are characterised by type and supporting reads, linked to any candidate DNA variant, and assessed for reading frame and therapeutic relevance.

WHAT IT PRODUCES

Aberrant splicing events with type and supporting reads, junction-level quantification against controls, linkage to candidate DNA variants, reading frame assessment, and tiered actionable findings.

WHAT IT DETECTS

Exon skipping events, intron retention, cryptic splice site activation, aberrant isoform expression, and the functional consequence of splice-region DNA variants.

WHAT IT DOES NOT DETECT

Events in transcripts not expressed in the sampled material, and anything lost to RNA degradation. Nonsense-mediated decay removes some aberrant transcripts before they can be observed, and distinguishing a pathogenic splicing change from normal isoform variation requires a control cohort that many laboratories lack.

WHAT MAKES IT HARD

Normal isoform diversity is extensive, so aberrance is only definable relative to a matched control cohort, and building that cohort is the main barrier. Nonsense-mediated decay silently removes the very evidence being sought unless it is considered explicitly.

WHO DOES THIS JOB

Clinical bioinformatician with transcriptomics competence in cancer genomics, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Detect and characterise aberrant splicing against a control cohort, link events to DNA variants, account for transcript decay, and assess therapeutic relevance.

WHO HIRES FOR IT

Oncology diagnostic laboratories with RNA capability, lung and haematological cancer services, pharmaceutical translational groups, and academic cancer centres.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
16	**78 hours · 13 days**	Prior RNA-Seq and variant analysis experience	**1,80,000**

SON-26
WORKFLOW

Oncogenic Splice Variant Analysis

(**Splicing alteration and exon skipping analysis** · Exon skipping analysis)

COMPLETE WORKFLOW — 16 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Control cohort verification**	A matched control cohort confirmed available for defining aberrance	Cohort register	*Control cohort record*
02	**RNA quality assessment**	RNA integrity assessed as the bound on splicing analysis reliability	RNA quality metrics	*RNA quality record*
03	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
04	**Read preprocessing**	Adapters and low-quality tails removed with strand preserved	fastp	*Trimmed FASTQ*
05	**Splice-aware alignment**	Reads aligned with a method reporting novel and annotated junctions	STAR two-pass	*Aligned BAM*
06	**Junction extraction**	Splice junctions extracted with supporting read counts	Junction extraction	*Junction table*
07	**Junction normalisation**	Junction usage normalised against local expression for comparability	Normalisation	*Normalised junctions*
08	**Control comparison**	Junction usage compared against the matched control cohort distribution	Statistical comparison	*Comparison results*
09	**Aberrant event detection**	Statistically aberrant splicing events identified against controls	FRASER, rMATS	*Aberrant events*
10	**Event characterisation**	Events characterised as skipping, retention or cryptic activation with read support	Event annotation	*Characterised events*
11	**Nonsense-mediated decay assessment**	Decay considered as a mechanism removing evidence of aberrant transcripts	NMD analysis	*Decay assessment*
12	**DNA variant linkage**	Aberrant events linked to candidate splice-region variants from DNA analysis	Variant-event correlation	*Linkage record*
13	**Reading frame and domain assessment**	Predicted transcripts assessed for frame and functional domain consequence	Frame analysis	*Frame assessment*
14	**Known oncogenic event matching**	Events matched against catalogued targetable splicing alterations	Curated catalogues	*Catalogue matches*
15	**Oncogenicity and tiering**	Findings assessed for oncogenic effect and assigned evidence tiers	OncoKB, AMP/ASCO/CAP	*Tiered findings*
16	**Reporting**	Events, evidence, linkage and RNA quality bound assembled	Report template	*Splicing report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
16	**78 h · 13 d**	**SON-26**	**1,80,000**

SON-27
SECTOR 01

Myeloid & Haematological Malignancy Panel Analysis

(**Myeloid and haematology panel analysis** · Myeloid panel · haem panel)

WHY IT IS DONE

Haematological malignancies are classified, risk-stratified and treated on molecular findings, and current classification systems make specific mutations diagnostic in their own right. The analysis differs from solid tumour work because the sample is blood or marrow, purity is high, and low-fraction subclones carry direct prognostic weight.

WHAT IT ANSWERS

What mutations does this patient's myeloid or lymphoid malignancy carry, at what allele fractions, and what do they imply for classification, risk group and treatment?

WHEN IT IS DONE

At diagnosis of acute leukaemia, myelodysplastic syndrome, myeloproliferative neoplasm or lymphoma, at relapse, and at defined points during treatment for subclone tracking.

WHAT TRIGGERS IT

A new haematological malignancy diagnosis, relapse or transformation, treatment response assessment, or transplant decision-making.

HOW IT IS DONE

Blood or marrow is sequenced deeply with panel-specific handling of difficult regions including internal tandem duplications. Low-fraction calling proceeds against an empirical error model, variants are annotated against haematology-specific evidence, classification and risk criteria are applied, and clonal architecture is assessed for subclone tracking.

WHAT IT PRODUCES

The variant set with allele fractions and clonal structure, internal tandem duplication findings with ratios, classification and risk stratification against current criteria, subclone tracking baseline, and a signed report.

WHAT IT DETECTS

Single nucleotide variants and insertions and deletions including internal tandem duplications, low-fraction subclones, panel-scoped copy number changes, and the mutation set required for current classification and risk stratification.

WHAT IT DOES NOT DETECT

Fusions and translocations central to haematological classification usually require RNA or cytogenetic methods, and large structural events are missed by panel designs. Distinguishing a somatic mutation from clonal haematopoiesis in an older patient is genuinely difficult without a germline comparator.

WHAT MAKES IT HARD

Internal tandem duplications and other length-variable events defeat standard callers and need dedicated handling. Allele fraction carries prognostic meaning here rather than being incidental, so accurate quantification matters, and germline predisposition mutations surface frequently and need a referral pathway.

WHO DOES THIS JOB

Clinical bioinformatician in haematology genomics, mid to senior level, working closely with haematopathology.

WHAT YOU CAN DO AFTERWARDS

Operate a myeloid and lymphoid panel including length-variable events, quantify allele fractions defensibly, apply current classification and risk criteria, and establish subclone tracking.

WHO HIRES FOR IT

Haematology and bone marrow transplant centres, haematological malignancy reference laboratories, hospital molecular pathology, and diagnostic chains.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
23	**112 hours · 19 days**	Prior somatic panel and low-fraction calling experience	**2,55,000**

SON-27
WORKFLOW

Myeloid & Haematological Malignancy Panel Analysis

(**Myeloid and haematology panel analysis** · Myeloid panel · haem panel)

COMPLETE WORKFLOW — 23 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Panel version and claim verification**	Panel version, gene content and validated claim scope confirmed	Panel register	*Authorised panel version*
02	**Specimen type and blast content recording**	Blood or marrow origin and blast percentage recorded as analytical inputs	Haematopathology report	*Specimen record*
03	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
04	**Read preprocessing**	Adapters and low-quality tails removed with retention accounting	fastp	*Trimmed FASTQ*
05	**Alignment and duplicate handling**	Reads aligned with duplicate handling appropriate to the chemistry	bwa-mem2, Picard	*Analysis-ready BAM*
06	**Per-target coverage certification**	Depth certified per target against the validated threshold	mosdepth, bedtools coverage	*Coverage certificate*
07	**Background error model construction**	Error rates measured empirically to support low-fraction calling	Error profiling	*Empirical error model*
08	**Deep variant calling**	Variants called with a low allele fraction floor supported by panel depth	Validated caller	*Variant set*
09	**Internal tandem duplication detection**	Length-variable duplications detected with a method built for them	ITD-specific caller	*ITD findings*
10	**Allelic ratio quantification**	Duplication allelic ratios quantified as they carry prognostic weight	Ratio calculation	*Allelic ratios*
11	**Insertion and deletion review**	Larger insertions and deletions reviewed at read level for accuracy	IGV review	*Reviewed indels*
12	**Panel copy number calling**	Copy number called across genes carrying a validated claim	Panel CNV method	*Copy number findings*
13	**Artefact and recurrent error filtering**	Design-specific recurrent artefacts removed using an internal database	Internal artefact database	*Filtered call set*
14	**Clonal haematopoiesis discrimination**	Variants assessed for clonal haematopoiesis against malignancy-associated patterns	Spectrum and fraction review	*CHIP discrimination*
15	**Germline predisposition flagging**	Variants suggesting inherited predisposition flagged for referral	Predisposition gene review	*Predisposition flags*
16	**Annotation**	Variants annotated on haematology-appropriate canonical transcripts	VEP, haematology transcripts	*Annotated variants*
17	**Oncogenicity assessment**	Variants assessed for oncogenic effect against haematology-specific evidence	ClinGen somatic criteria, curated evidence	*Oncogenicity calls*
18	**Classification criteria application**	Current disease classification criteria applied to the molecular findings	Current classification criteria	*Classification assignment*
19	**Risk stratification**	Risk group assigned using the applicable stratification system	Risk stratification criteria	*Risk group*
20	**Clonal architecture assessment**	Allele fractions assessed to describe clonal and subclonal structure	Fraction analysis	*Clonal architecture*
21	**Subclone tracking baseline**	A baseline established for tracking specific clones at later timepoints	Baseline procedure	*Tracking baseline*
22	**Report generation**	Findings, classification, risk group and architecture assembled	Validated report template	*Draft report*
23	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed molecular report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
23	**112 h · 19 d**	**SON-27**	**2,55,000**

SON-28
SECTOR 01

Lymphoid Clonality Analysis

(**IG and TR clonality assessment** · Clonality assay · IG/TR clonality)

WHY IT IS DONE

Distinguishing a malignant lymphoid clone from a reactive expansion is the central diagnostic question in lymphoma and lymphoproliferative disease. Sequencing-based clonality has largely replaced fragment analysis because it identifies the clonal sequence itself, which can then be tracked for residual disease.

WHAT IT ANSWERS

Is there a dominant lymphoid clone in this specimen, what is its rearranged sequence, and can it be distinguished from a reactive polyclonal background?

WHEN IT IS DONE

At diagnosis of suspected lymphoma or lymphoproliferative disease, to establish a trackable clonal sequence for later monitoring, and in staging of involved tissue.

WHAT TRIGGERS IT

Suspected lymphoid malignancy on morphology, an equivocal immunophenotype, or the need to establish a clonal marker before treatment.

HOW IT IS DONE

Rearranged loci are amplified or captured with designs covering the required gene segments, reads are processed with careful primer handling, and rearrangements are assembled and assigned to gene segments. Clonal dominance is assessed against the repertoire background using validated criteria, and clonal sequences are recorded for tracking.

WHAT IT PRODUCES

The assembled repertoire with frequency distribution, dominant clone identification with sequence, clonality assessment against validated criteria, the trackable clonal sequence record, and a signed report.

WHAT IT DETECTS

Immunoglobulin and T-cell receptor rearrangements, clonal dominance against the polyclonal background, the specific clonal sequence for downstream tracking, and multiple clones where present.

WHAT IT DOES NOT DETECT

Clones using rearrangements outside the primer or capture design are missed, and somatic hypermutation in the primer binding region causes dropout in germinal centre derived lymphomas — a well-recognised false negative. Clonality is not synonymous with malignancy, since reactive conditions can produce restricted repertoires.

WHAT MAKES IT HARD

Somatic hypermutation causing primer dropout is the characteristic false negative and must be addressed by design or by acknowledging it in the report. Judging dominance requires a sufficient background repertoire, so low-input and heavily infiltrated samples give unreliable clonality assessment.

WHO DOES THIS JOB

Clinical bioinformatician in haematology or immunogenetics, mid level.

WHAT YOU CAN DO AFTERWARDS

Assemble and analyse lymphoid repertoires, assess clonal dominance against validated criteria, recognise hypermutation-driven dropout, and record trackable clonal sequences.

WHO HIRES FOR IT

Haematopathology laboratories, lymphoma specialist centres, immunogenetics services, and haematology reference laboratories.

COMMERCIALS

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

SON-28
WORKFLOW

Lymphoid Clonality Analysis

(**IG and TR clonality assessment** · Clonality assay · IG/TR clonality)

COMPLETE WORKFLOW — 17 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Assay design verification**	The gene segments and rearrangement types covered by the design confirmed	Design register	*Authorised design version*
02	**Specimen adequacy assessment**	Specimen assessed for lymphoid cellularity and input adequacy	Pathology review	*Adequacy record*
03	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
04	**Primer and adapter handling**	Primers trimmed with care taken not to distort rearrangement sequence	Primer trimming	*Trimmed reads*
05	**Read pairing and merging**	Paired reads merged where overlap permits to recover full rearrangements	Read merging	*Merged reads*
06	**Rearrangement assembly**	Rearranged sequences assembled from reads with quality thresholds applied	Assembly procedure	*Assembled rearrangements*
07	**Gene segment assignment**	Variable, diversity and joining segments assigned against reference databases	IgBLAST, MiXCR	*Segment assignments*
08	**Junction characterisation**	Junction regions characterised including insertions and deletions at joins	Junction analysis	*Junction characterisation*
09	**Clonotype definition**	Clonotypes defined by consistent criteria and frequencies computed	Clonotype definition	*Clonotype table*
10	**Repertoire background assessment**	The polyclonal background assessed as the comparator for dominance	Repertoire analysis	*Background assessment*
11	**Somatic hypermutation assessment**	Hypermutation load assessed, including its effect on primer binding	Mutation analysis	*Hypermutation assessment*
12	**Primer dropout risk assessment**	Risk of hypermutation-driven dropout assessed and stated for the specimen	Dropout risk analysis	*Dropout risk statement*
13	**Clonal dominance evaluation**	Dominance evaluated against validated criteria rather than by inspection	Dominance criteria	*Dominance assessment*
14	**Multiple clone assessment**	Additional clones assessed and reported where present	Multi-clone analysis	*Multiple clone findings*
15	**Trackable sequence recording**	The clonal sequence recorded in a form usable for later residual disease tracking	Sequence recording	*Trackable sequence*
16	**Clinical interpretation**	Clonality interpreted in the context of morphology and immunophenotype	Clinical correlation	*Interpretation record*
17	**Report generation and sign-out**	Clonality assessment, sequences and dropout risk issued under dual signature	Validated report template	*Signed report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
17	**82 h · 14 d**	**SON-28**	**1,90,000**

SON-29
SECTOR 01

Post-Transplant Chimerism Analysis

(**Donor chimerism monitoring** · Chimerism monitoring)

WHY IT IS DONE

After allogeneic transplant, the proportion of donor to recipient cells indicates engraftment success and predicts relapse. Sequencing-based chimerism reaches sensitivity an order of magnitude beyond fragment analysis, which brings forward the point at which intervention is possible.

WHAT IT ANSWERS

What proportion of this patient's haematopoietic cells are donor-derived, in which lineages, and how is that proportion changing across timepoints?

WHEN IT IS DONE

At defined intervals after allogeneic transplant, when relapse is suspected, before and after donor lymphocyte infusion, and in graft failure investigation.

WHAT TRIGGERS IT

A scheduled post-transplant monitoring timepoint, suspected relapse or graft failure, or assessment before a donor lymphocyte infusion.

HOW IT IS DONE

Informative markers differing between donor and recipient are identified from pre-transplant samples, sensitivity is calculated from the number available, and post-transplant samples are sequenced deeply at those markers. Proportions are quantified with confidence bounds, lineage-specific fractions are assessed where sorted, and trajectory is analysed.

WHAT IT PRODUCES

The informative marker set with derived sensitivity, donor and recipient proportions with bounds per timepoint, lineage-specific results where applicable, trajectory analysis, and a clinical report with intervention implications.

WHAT IT DETECTS

Donor and recipient proportions at high sensitivity, lineage-specific chimerism in sorted fractions, and trajectory across serial timepoints.

WHAT IT DOES NOT DETECT

It reports cellular origin, not malignancy, so recipient cells may be normal or malignant and chimerism alone does not distinguish them. Sensitivity depends on the number of informative markers between donor and recipient, which is reduced in related and HLA-matched donors.

WHAT MAKES IT HARD

Sensitivity is determined by informative marker count, which varies per donor-recipient pair, so a per-case sensitivity statement is required rather than an assay-wide claim. Related donors reduce informativeness substantially and the report must reflect that.

WHO DOES THIS JOB

Clinical bioinformatician in haematology or transplant genomics, mid level.

WHAT YOU CAN DO AFTERWARDS

Select informative markers and derive per-case sensitivity, quantify chimerism with bounds, analyse lineage-specific fractions, and interpret trajectories for intervention.

WHO HIRES FOR IT

Bone marrow transplant centres, haematology reference laboratories, transplant immunology services, and hospital molecular pathology.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
15	**72 hours · 12 days**	Prior genotype analysis experience	**1,65,000**

SON-29
WORKFLOW

Post-Transplant Chimerism Analysis

(**Donor chimerism monitoring** · Chimerism monitoring)

COMPLETE WORKFLOW — 15 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Pre-transplant sample verification**	Donor and recipient pre-transplant samples confirmed available and correctly attributed	Sample inventory, identity check	*Sample verification*
02	**Informative marker identification**	Markers differing between donor and recipient identified genome-wide	Marker analysis	*Informative marker set*
03	**Marker count and sensitivity derivation**	Assay sensitivity derived from the number of informative markers available	Sensitivity calculation	*Per-case sensitivity*
04	**Relatedness assessment**	Donor-recipient relatedness assessed as it determines marker availability	Relatedness analysis	*Relatedness record*
05	**Timepoint scheduling**	Monitoring timepoints scheduled against the transplant protocol	Monitoring protocol	*Timepoint schedule*
06	**Cell fraction preparation verification**	Sorted lineage fractions confirmed where lineage-specific chimerism is required	Sorting records	*Fraction verification*
07	**Deep marker sequencing**	Post-transplant samples sequenced deeply at the informative marker set	Deep sequencing	*Sequencing data*
08	**Coverage adequacy assessment**	Depth confirmed adequate at markers for the sensitivity being claimed	mosdepth	*Coverage record*
09	**Allele counting**	Donor and recipient allele counts quantified at every informative marker	Allele counting	*Allele counts*
10	**Error and background modelling**	Background error at marker positions modelled for low-level detection	Error profiling	*Background model*
11	**Chimerism quantification**	Donor and recipient proportions computed across markers with bounds	Quantification	*Chimerism proportions*
12	**Outlier marker exclusion**	Markers behaving inconsistently excluded with the reason recorded	Outlier analysis	*Exclusion record*
13	**Lineage-specific analysis**	Chimerism computed separately for each sorted lineage fraction	Per-fraction quantification	*Lineage results*
14	**Trajectory analysis**	Proportions analysed across timepoints for engraftment and relapse signal	Trajectory analysis	*Trajectory results*
15	**Reporting**	Proportions, sensitivity, lineage results and trajectory assembled	Report template	*Chimerism report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
15	**72 h · 12 d**	**SON-29**	**1,65,000**

SON-30
SECTOR 01

Hereditary Cancer Predisposition Analysis

(**Hereditary cancer panel analysis** · Hereditary cancer panel)

WHY IT IS DONE

Identifying an inherited predisposition changes surveillance, surgical decisions and therapy for the patient, and identifies at-risk relatives who can be tested and managed before they develop cancer. It is the analysis with the widest reach beyond the individual patient.

WHAT IT ANSWERS

Does this patient carry a germline pathogenic variant predisposing to cancer, and what does it imply for their management and for their family?

WHEN IT IS DONE

At diagnosis where family history, age of onset or tumour phenotype suggests predisposition, following an incidental germline-suspicious finding in tumour testing, and in cascade testing of relatives.

WHAT TRIGGERS IT

A suggestive family history, early onset disease, a characteristic tumour phenotype, a germline-suspicious somatic finding, or a known familial variant.

HOW IT IS DONE

Germline material is sequenced with per-gene coverage certified, sequence and copy number analysis run together, and pseudogene-bearing genes routed to specialist methods. Variants are classified under gene-specific criteria, second-hit evidence is assessed against tumour data where available, and management and cascade implications are set out.

WHAT IT PRODUCES

Per-gene coverage certification, classified germline variants, copy number findings, second-hit evidence where paired tumour data exist, management recommendations, and cascade testing guidance for relatives.

WHAT IT DETECTS

Germline sequence variants and copy number changes across predisposition genes, founder variants, and second-hit evidence in the tumour where paired data are available.

WHAT IT DOES NOT DETECT

Genes outside the panel, deep intronic and regulatory variants, and mosaic predisposition present in some tissues but not blood. Pseudogene-bearing predisposition genes need dedicated methods, and moderate-penetrance findings often lack management guidance even when detected.

WHAT MAKES IT HARD

Copy number changes are a substantial share of yield in these genes and are still frequently omitted. Pseudogene interference in several core predisposition genes produces both false positives and missed variants unless a dedicated method is used, and moderate-penetrance findings require careful communication.

WHO DOES THIS JOB

Clinical bioinformatician or variant scientist in cancer genetics, mid level, working with genetic counselling.

WHAT YOU CAN DO AFTERWARDS

Run hereditary cancer panels including copy number, manage pseudogene-bearing loci, classify under gene-specific criteria, and support cascade testing pathways.

WHO HIRES FOR IT

Cancer genetics services, hereditary cancer clinics, oncology diagnostic laboratories, and diagnostic chains offering predisposition testing.

COMMERCIALS

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

SON-30
WORKFLOW

Hereditary Cancer Predisposition Analysis

(**Hereditary cancer panel analysis** · Hereditary cancer panel)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Panel version authorisation**	Gene content, version and gene-disease validity tiers confirmed and recorded	Panel register	*Authorised panel version*
02	**Consent and scope confirmation**	Scope, secondary findings election and cascade implications confirmed	Consent record	*Analysis authorisation*
03	**Family history capture**	Family history captured in structured form for interpretation and cascade planning	Pedigree, structured capture	*Family history record*
04	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC, MultiQC	*QC acceptance record*
05	**Read preprocessing**	Adapters and low-quality tails removed with retention accounting	fastp	*Trimmed FASTQ*
06	**Alignment and processing**	Reads aligned, duplicate marked and recalibrated within panel intervals	bwa-mem2, Picard, GATK BQSR	*Analysis-ready BAM*
07	**Per-gene coverage certification**	Depth certified per gene and exon with shortfalls named individually	mosdepth, bedtools coverage	*Coverage certificate*
08	**Identity verification**	Fingerprint captured to confirm attribution before interpretation	somalier extract	*Identity clearance*
09	**Germline variant calling**	Variants called across the predisposition gene set with validated parameters	Validated caller	*Germline VCF*
10	**Copy number calling**	Exon-level copy number called against a matched reference pool	gCNV, DECoN	*Copy number calls*
11	**Pseudogene-aware locus handling**	Genes with pseudogene interference routed to dedicated methods	Specialist assay, long-read	*Specialist locus results*
12	**Founder variant screening**	Population-specific founder variants interrogated at required sensitivity	Founder variant panel	*Founder findings*
13	**Annotation and population filtering**	Variants annotated with gene-specific frequency thresholds applied	VEP, gnomAD, ClinVar	*Annotated variants*
14	**Gene-specific classification**	Variants classified under gene-specific specifications rather than generic criteria	ACMG-AMP with gene specifications	*Classified variants*
15	**Second-hit assessment**	Tumour data assessed for somatic second hits where paired data exist	LOH and somatic review	*Second-hit findings*
16	**Penetrance and management mapping**	Findings mapped to penetrance estimates and management guidance	Management guidelines	*Management recommendations*
17	**Moderate-penetrance handling**	Moderate-penetrance findings handled under the documented reporting policy	Reporting policy	*Policy application record*
18	**Cascade testing guidance**	Guidance prepared for testing at-risk relatives with the specific variant	Cascade protocol	*Cascade guidance*
19	**Orthogonal confirmation**	Reportable findings confirmed by an independent method	Sanger, MLPA, ddPCR	*Confirmation record*
20	**Report generation**	Findings, coverage, management and cascade guidance assembled	Validated report template	*Draft report*
21	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed clinical report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SON-30**	**2,35,000**

SON-31
SECTOR 01

Germline–Somatic Integration Analysis

(**Paired germline and somatic reporting** · Paired germline reporting)

WHY IT IS DONE

When tumour and normal are both sequenced, germline findings arise whether or not they were sought. Handling them deliberately turns an incidental risk into clinical value — confirming second hits, identifying therapy-relevant germline variants, and detecting predisposition that changes family management.

WHAT IT ANSWERS

What germline findings arise from this paired analysis, how do they interact with the somatic profile, and what do they mean for the patient and their relatives?

WHEN IT IS DONE

In every paired tumour-normal analysis, in tumour-only analysis where germline-suspicious variants surface, and wherever therapy eligibility depends on germline as well as somatic status.

WHAT TRIGGERS IT

A paired tumour-normal analysis, a germline-suspicious finding in tumour-only testing, or a therapy decision requiring germline status.

HOW IT IS DONE

Germline analysis is performed on the normal with consent scope confirmed, findings are classified under germline criteria, and second-hit evidence is assessed by combining the germline variant with somatic copy number and loss of heterozygosity. Therapy-relevant germline status is determined and reporting follows the agreed disclosure framework.

WHAT IT PRODUCES

Classified germline findings, biallelic inactivation evidence integrating germline and somatic data, therapy-relevant germline status, the consent and disclosure record, and cascade testing recommendations.

WHAT IT DETECTS

Germline pathogenic variants in predisposition genes, biallelic inactivation through germline variant plus somatic second hit, germline variants determining therapy eligibility, and the germline contribution to observed mutational signatures.

WHAT IT DOES NOT DETECT

Germline findings outside the analysed gene set, and mosaic predisposition absent from the sequenced normal. Where the normal is blood in a haematological malignancy, germline and somatic cannot be separated without a non-haematopoietic comparator such as skin.

WHAT MAKES IT HARD

Consent is the crux, because paired sequencing generates germline findings whether or not the patient agreed to receive them. The disclosure framework must exist before the first finding, and in haematological malignancy the blood normal is not a germline sample at all.

WHO DOES THIS JOB

Clinical bioinformatician spanning germline and somatic work, mid to senior level, with genetic counselling involvement.

WHAT YOU CAN DO AFTERWARDS

Integrate germline and somatic evidence, establish biallelic inactivation, determine therapy-relevant germline status, and operate a disclosure framework.

WHO HIRES FOR IT

Comprehensive cancer centres, cancer genetics services, oncology diagnostic laboratories, and precision oncology programmes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
18	**86 hours · 15 days**	Prior germline and somatic analysis experience	**2,00,000**

SON-31
WORKFLOW

Germline–Somatic Integration Analysis

(**Paired germline and somatic reporting** · Paired germline reporting)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Consent scope verification**	Germline reporting scope and secondary findings election confirmed before analysis	Consent record	*Analysis authorisation*
02	**Normal sample suitability assessment**	The normal confirmed suitable as a germline comparator for the malignancy type	Sample review	*Suitability record*
03	**Disclosure framework confirmation**	The framework for returning germline findings confirmed before any are generated	Laboratory policy	*Framework confirmation*
04	**Germline gene scope definition**	The germline genes to be analysed defined and recorded	Gene scope register	*Authorised germline scope*
05	**Germline coverage certification**	Depth certified across the germline gene scope in the normal sample	mosdepth, bedtools coverage	*Coverage certificate*
06	**Germline variant calling**	Variants called in the normal using germline-appropriate parameters	Validated germline caller	*Germline VCF*
07	**Germline copy number calling**	Copy number called in the normal across predisposition genes	gCNV, DECoN	*Germline CNV calls*
08	**Annotation and population filtering**	Germline variants annotated with appropriate frequency thresholds	VEP, gnomAD, ClinVar	*Annotated germline variants*
09	**Germline classification**	Variants classified under germline criteria with gene-specific specifications	ACMG-AMP with gene specifications	*Classified germline variants*
10	**Somatic second-hit assessment**	Somatic loss of heterozygosity and mutation assessed at germline variant loci	LOH and somatic review	*Second-hit evidence*
11	**Biallelic inactivation determination**	Biallelic inactivation established by combining germline and somatic evidence	Integration logic	*Biallelic findings*
12	**Therapy-relevant germline determination**	Germline status determined where it governs therapy eligibility	Therapy criteria	*Therapy-relevant status*
13	**Signature contribution assessment**	Germline contribution to observed mutational signatures assessed	Signature integration	*Signature interpretation*
14	**Somatic-germline reconciliation**	Somatic call set reviewed for variants that are in fact germline	Cross-comparison	*Reconciliation record*
15	**Incidental finding handling**	Incidental predisposition findings handled under the disclosure framework	Disclosure framework	*Disclosure record*
16	**Cascade implication assessment**	Family implications assessed with cascade testing guidance prepared	Cascade protocol	*Cascade guidance*
17	**Report generation**	Integrated germline and somatic findings assembled with disclosure record	Validated report template	*Draft report*
18	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed clinical report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**86 h · 15 d**	**SON-31**	**2,00,000**

SON-32
SECTOR 01

Neoantigen Prediction & Ranking Analysis

(**Neoantigen identification and prioritisation** · Neoantigen analysis)

WHY IT IS DONE

Personalised cancer vaccines and engineered cell therapies both depend on identifying which of a tumour's mutations produce peptides the patient's immune system can recognise. This analysis is the computational core of individualised immunotherapy.

WHAT IT ANSWERS

Which mutations in this tumour generate peptides that bind this patient's HLA alleles, are actually expressed, and are most likely to be immunogenic?

WHEN IT IS DONE

In personalised cancer vaccine development, in adoptive cell therapy target selection, and in translational research on immunotherapy response.

WHAT TRIGGERS IT

Enrolment in a personalised vaccine or cell therapy programme, or a research protocol on immunogenicity.

HOW IT IS DONE

Somatic mutations are called from exome or genome data and the patient's HLA alleles typed to high resolution. Mutant peptide sequences are generated including frameshift and fusion-derived neo-open reading frames, binding is predicted against the patient's specific alleles, expression support is required from RNA, and candidates are ranked on combined evidence.

WHAT IT PRODUCES

The candidate peptide set with mutation origin, patient-specific HLA typing, binding predictions with scores, expression support per candidate, antigen presentation machinery status, and a ranked prioritised list.

WHAT IT DETECTS

Mutation-derived candidate peptides, predicted binding affinity to the patient's specific HLA alleles, expression support for each candidate, and a ranked prioritisation for therapeutic use.

WHAT IT DOES NOT DETECT

Prediction is not demonstration — binding affinity correlates only moderately with genuine immunogenicity, and validated experimental confirmation remains necessary. Antigen processing and presentation upstream of binding are poorly modelled, and tumours that have lost HLA or presentation machinery will not present anything regardless of prediction.

WHAT MAKES IT HARD

Prediction quality is bounded by HLA typing accuracy and by expression evidence, and candidates without expression support are frequently carried forward regardless. Frameshift and fusion-derived neo-open reading frames are the most promising sources and the most often omitted because they need dedicated handling.

WHO DOES THIS JOB

Computational immunologist or senior bioinformatician in immuno-oncology. A specialist role concentrated in biotech and translational settings.

WHAT YOU CAN DO AFTERWARDS

Generate candidate neoantigens including frameshift-derived peptides, predict binding against patient-specific HLA, require expression support, and rank candidates defensibly.

WHO HIRES FOR IT

Personalised cancer vaccine companies, cell therapy developers, immuno-oncology biotech, and academic immunotherapy programmes.

COMMERCIALS

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

SON-32
WORKFLOW

Neoantigen Prediction & Ranking Analysis

(**Neoantigen identification and prioritisation** · Neoantigen analysis)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Programme scope confirmation**	The therapeutic programme and its candidate requirements confirmed	Programme protocol	*Scope confirmation*
02	**Somatic variant input preparation**	High-confidence somatic variants prepared as the candidate mutation source	Somatic call set	*Mutation input set*
03	**Variant class inclusion decision**	Missense, frameshift, fusion and splice-derived sources included by documented rule	Inclusion rules	*Class inclusion record*
04	**HLA typing to high resolution**	Patient HLA class I and class II alleles typed to the resolution binding requires	HLA typing method	*HLA allele set*
05	**HLA typing confirmation**	Typing confirmed by an orthogonal method given its impact on prediction	Orthogonal typing	*Typing confirmation*
06	**Antigen presentation machinery assessment**	Presentation pathway genes assessed for loss or mutation	Gene status review	*Presentation status*
07	**HLA loss of heterozygosity assessment**	Somatic loss at the HLA locus assessed as it removes presentation capacity	HLA LOH analysis	*HLA LOH result*
08	**Mutant peptide generation**	Mutant peptide sequences generated across the required length range	Peptide generation	*Candidate peptides*
09	**Frameshift neo-open reading frame generation**	Frameshift-derived novel reading frames translated as candidate sources	Neo-ORF generation	*Frameshift candidates*
10	**Fusion-derived peptide generation**	Peptides spanning fusion junctions generated where fusions are present	Fusion peptide generation	*Fusion candidates*
11	**Binding affinity prediction**	Binding predicted against the patient's specific alleles rather than common ones	Binding prediction	*Binding predictions*
12	**Presentation likelihood prediction**	Processing and presentation likelihood predicted beyond binding affinity alone	Presentation prediction	*Presentation scores*
13	**Expression support requirement**	Candidates required to have expression support from RNA data	Expression integration	*Expression-supported candidates*
14	**Allele fraction and clonality weighting**	Candidates weighted by mutation clonality so subclonal targets are deprioritised	Clonality integration	*Clonality-weighted candidates*
15	**Self-similarity filtering**	Candidates too similar to self peptides filtered to reduce tolerance risk	Similarity analysis	*Filtered candidates*
16	**Manufacturability assessment**	Candidates assessed for synthesis and formulation feasibility	Manufacturability criteria	*Feasible candidates*
17	**Ranking**	Candidates ranked on combined binding, presentation, expression and clonality evidence	Ranking model	*Ranked candidate list*
18	**Diversity and coverage assessment**	The selected set assessed for allele coverage and target diversity	Coverage analysis	*Set composition assessment*
19	**Validation recommendation**	Experimental validation recommended for the prioritised candidates	Validation criteria	*Validation recommendation*
20	**Limitation statement**	The gap between predicted binding and demonstrated immunogenicity stated	Limitation template	*Limitation statement*
21	**Reporting**	Ranked candidates with all supporting evidence assembled for the programme	Report template	*Neoantigen report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SON-32**	**2,35,000**

SON-33
SECTOR 01

Immune Repertoire & Immunotherapy Biomarker Analysis

(**TCR repertoire and immune biomarker analysis** · TCR profiling · IO biomarkers)

WHY IT IS DONE

Whether a patient responds to checkpoint immunotherapy depends on the immune context of the tumour as much as on its mutations. Repertoire diversity, clonal expansion and immune infiltration together predict response better than any single genomic marker.

WHAT IT ANSWERS

What is the immune repertoire and infiltration status of this tumour, and do the biomarkers indicate likely benefit from checkpoint immunotherapy?

WHEN IT IS DONE

Before checkpoint immunotherapy, during treatment to assess immune response, in trial biomarker programmes, and in resistance investigation.

WHAT TRIGGERS IT

Consideration of checkpoint therapy, on-treatment response assessment, trial biomarker requirements, or immunotherapy resistance investigation.

HOW IT IS DONE

Repertoire is extracted from targeted or bulk RNA data and clonotypes assembled with rigorous quality control. Diversity and clonality metrics are computed against matched controls, immune infiltration is estimated by deconvolution, signature scores are calculated, and antigen presentation integrity is assessed from the genomic data.

WHAT IT PRODUCES

Repertoire metrics with clonotype tables, expanded clone identification and dynamics, immune deconvolution estimates, immune signature scores, presentation machinery status, and an integrated biomarker summary.

WHAT IT DETECTS

T-cell receptor repertoire diversity and clonality, expanded clones and their dynamics, immune cell infiltration by deconvolution, immune-related expression signatures, and antigen presentation machinery integrity.

WHAT IT DOES NOT DETECT

It cannot determine what the expanded clones are specific for without additional experimental work. Repertoire from bulk tissue reflects only the sampled region, blood repertoire correlates imperfectly with tumour, and none of these biomarkers is individually decisive for response prediction.

WHAT MAKES IT HARD

Repertoire metrics are extremely sensitive to sequencing depth and input material, so comparisons across samples require normalisation that is frequently omitted. No single biomarker predicts response, so integration is the actual skill and it must avoid implying more certainty than the evidence supports.

WHO DOES THIS JOB

Computational immunologist or clinical bioinformatician in immuno-oncology, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Extract and normalise immune repertoires, compute diversity and clonality correctly, estimate infiltration, and integrate multiple biomarkers without overstating predictive power.

WHO HIRES FOR IT

Immuno-oncology biotech, pharmaceutical biomarker programmes, academic cancer immunology groups, and comprehensive cancer centres.

COMMERCIALS

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

SON-33
WORKFLOW

Immune Repertoire & Immunotherapy Biomarker Analysis

(**TCR repertoire and immune biomarker analysis** · TCR profiling · IO biomarkers)

COMPLETE WORKFLOW — 19 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Analysis scope definition**	Repertoire, infiltration and signature components in scope defined and recorded	Assay register	*Scope definition*
02	**Specimen and input assessment**	Specimen type and input adequacy assessed for repertoire recovery	Sample review	*Input record*
03	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
04	**Sequencing depth normalisation planning**	Depth differences across samples addressed before any repertoire comparison	Normalisation planning	*Normalisation plan*
05	**Repertoire extraction**	Receptor sequences extracted from targeted or bulk transcriptome data	MiXCR, TRUST4	*Extracted repertoire*
06	**Clonotype assembly**	Clonotypes assembled with error correction and quality thresholds applied	Clonotype assembly	*Clonotype table*
07	**Repertoire quality control**	Recovery, coverage and error rates assessed against acceptance criteria	Repertoire QC	*Repertoire QC record*
08	**Diversity metric computation**	Diversity indices computed with depth normalisation applied	Diversity calculation	*Diversity metrics*
09	**Clonality metric computation**	Clonality and evenness computed against normalised repertoires	Clonality calculation	*Clonality metrics*
10	**Expanded clone identification**	Dominant and expanded clones identified against the repertoire background	Expansion analysis	*Expanded clones*
11	**Serial dynamics analysis**	Clone dynamics analysed across timepoints where serial samples exist	Dynamics analysis	*Clonal dynamics*
12	**Immune deconvolution**	Immune cell composition estimated from expression against reference profiles	Deconvolution	*Composition estimates*
13	**Immune signature scoring**	Interferon, cytotoxicity and exclusion signatures scored	Signature scoring	*Signature scores*
14	**Antigen presentation assessment**	Presentation machinery integrity assessed from genomic and expression data	Gene status review	*Presentation status*
15	**Checkpoint expression assessment**	Checkpoint molecule expression quantified where in scope	Expression quantification	*Checkpoint expression*
16	**Tumour-immune spatial context flagging**	Absence of spatial context flagged as a limitation of bulk analysis	Limitation template	*Spatial limitation statement*
17	**Biomarker integration**	Components integrated into a summary without overstating predictive certainty	Integration procedure	*Integrated summary*
18	**Clinical interpretation**	Integrated biomarkers interpreted for likely immunotherapy benefit	Clinical interpretation	*Interpretation record*
19	**Reporting**	Repertoire metrics, composition, signatures and integrated summary assembled	Report template	*Immune biomarker report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
19	**92 h · 16 d**	**SON-33**	**2,10,000**

SON-34
SECTOR 01

Tumour Methylation Classification Analysis

(**Methylation-based tumour classification** · Methylation classifier · CNS classifier)

WHY IT IS DONE

Methylation profiles classify tumours more reproducibly than morphology in several settings, most decisively in central nervous system tumours where classification now formally incorporates methylation class. It resolves diagnoses that histology leaves ambiguous.

WHAT IT ANSWERS

Which methylation class does this tumour belong to, with what confidence, and does that classification change the diagnosis reached by histology?

WHEN IT IS DONE

In central nervous system tumours as standard, in sarcomas and paediatric tumours with overlapping morphology, in cancers of unknown primary, and where histological diagnosis is equivocal.

WHAT TRIGGERS IT

A central nervous system tumour requiring classification, morphologically ambiguous histology, an unknown primary, or a rare tumour with overlapping differential.

HOW IT IS DONE

Methylation is measured on the validated platform, quality controlled and normalised with tumour content assessed. The profile is classified against the reference cohort with calibrated confidence, copy number is derived from the same data as corroboration, and the result is correlated against histology with discrepancies investigated.

WHAT IT PRODUCES

Class assignment with calibrated confidence score, the copy number profile derived from methylation data, tumour content assessment, histological correlation, and a report stating when confidence is insufficient to classify.

WHAT IT DETECTS

Methylation class assignment with calibrated confidence score, copy number profile derived from the same data, and characteristic alterations such as amplifications and deletions relevant to the class.

WHAT IT DOES NOT DETECT

Only classes present in the reference cohort can be assigned, so rare and newly described entities return low confidence rather than a diagnosis. Low tumour content, necrosis and inflammatory infiltrate all degrade classification, and a low score is uninformative rather than negative.

WHAT MAKES IT HARD

Confidence calibration is what makes the result usable, and a low score must be reported as uninformative rather than as a weak diagnosis. Tumour content below the validated minimum produces confident misclassification, so assessing it is a prerequisite rather than a formality.

WHO DOES THIS JOB

Clinical bioinformatician with methylation and classification competence, mid to senior level, in neuropathology or specialist tumour settings.

WHAT YOU CAN DO AFTERWARDS

Run methylation classification with proper quality and content control, interpret calibrated confidence correctly, derive corroborating copy number, and reconcile with histology.

WHO HIRES FOR IT

Neuropathology and neuro-oncology centres, paediatric oncology, sarcoma specialist services, and reference laboratories offering classification.

COMMERCIALS

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

SON-34
WORKFLOW

Tumour Methylation Classification Analysis

(**Methylation-based tumour classification** · Methylation classifier · CNS classifier)

COMPLETE WORKFLOW — 18 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Reference cohort and classifier verification**	The classifier version and its reference cohort composition confirmed	Classifier register	*Authorised classifier version*
02	**Specimen and tumour content assessment**	Tumour content, necrosis and inflammatory infiltrate assessed as confounders	Pathology review	*Content record*
03	**Methylation data generation**	Genome-wide methylation measured on the validated platform	Methylation platform	*Raw methylation data*
04	**Signal and control quality assessment**	Detection rates and control probe performance assessed against criteria	Platform QC	*Methylation QC record*
05	**Normalisation**	Data normalised using the method the classifier was trained with	Normalisation procedure	*Normalised data*
06	**Batch effect assessment**	Batch structure assessed and corrected before classification	Batch correction	*Corrected data*
07	**Content adequacy check**	Tumour content confirmed above the classifier's validated minimum	Content criteria	*Adequacy decision*
08	**Classification**	The profile classified against the reference cohort across candidate classes	Classification model	*Class scores*
09	**Confidence calibration**	Scores calibrated to give interpretable confidence rather than raw model output	Calibration model	*Calibrated confidence*
10	**Threshold application**	Confidence assessed against validated thresholds for a reportable assignment	Threshold criteria	*Assignment decision*
11	**Uninformative declaration**	Low-confidence results declared uninformative rather than reported as weak calls	Declaration criteria	*Informativeness decision*
12	**Copy number derivation**	Copy number profile derived from the same methylation data as corroboration	Copy number derivation	*Copy number profile*
13	**Characteristic alteration assessment**	Class-characteristic amplifications and deletions assessed for corroboration	Alteration review	*Corroboration record*
14	**Histological correlation**	The molecular class correlated against the histological diagnosis	Correlation review	*Correlation record*
15	**Discrepancy investigation**	Discordance between methylation class and histology investigated and documented	Discrepancy procedure	*Discrepancy record*
16	**Clinical interpretation**	Class assignment interpreted for diagnosis, prognosis and management	Clinical interpretation	*Interpretation record*
17	**Report generation**	Class, calibrated confidence, copy number and correlation assembled	Validated report template	*Draft report*
18	**Clinical sign-out**	Report reviewed and authorised under dual signature	Sign-out procedure	*Signed clinical report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
18	**88 h · 15 d**	**SON-34**	**2,05,000**

SON-35
SECTOR 01

Oncogenic Viral Integration Analysis

(**Viral integration and oncogenic virus analysis** · HPV/HBV/EBV integration analysis)

WHY IT IS DONE

Several cancers are driven by integrated viral genomes, and where the virus inserts determines which host genes are disrupted or activated. Viral status also carries prognostic weight and, in head and neck and cervical cancer, directly influences treatment intensity.

WHAT IT ANSWERS

Is an oncogenic virus present in this tumour, is it integrated into the host genome, where, and which host genes does the integration affect?

WHEN IT IS DONE

In cervical, head and neck, hepatocellular and nasopharyngeal cancers, in lymphoproliferative disease associated with viral infection, and in unknown primary investigation.

WHAT TRIGGERS IT

A tumour type with established viral aetiology, a prognostically relevant viral status question, or unknown primary investigation.

HOW IT IS DONE

Reads not mapping to the human reference are screened against viral references, viral load and genotype are determined, and chimeric human-viral read pairs are identified. Integration sites are localised and refined, host gene disruption and activation are assessed, and integration is distinguished from episomal presence by junction evidence.

WHAT IT PRODUCES

Viral presence, load and genotype, integration status with junction evidence, integration site coordinates, affected host gene assessment, and prognostic interpretation.

WHAT IT DETECTS

Viral sequence presence and load, integration status distinguished from episomal presence, integration site coordinates, host genes disrupted or activated, and viral genotype.

WHAT IT DOES NOT DETECT

Integration into repetitive host sequence is unreliable, and low viral copy number falls below detection. It cannot determine whether the virus is causal or a bystander, and episomal virus without integration may still be biologically relevant while producing no integration signal.

WHAT MAKES IT HARD

Distinguishing genuine integration from contamination or from episomal virus rests on chimeric junction evidence, which is sparse and easily produced by alignment artefact. Junction-level review is therefore necessary rather than optional.

WHO DOES THIS JOB

Clinical bioinformatician in cancer genomics or infectious disease genomics, mid level.

WHAT YOU CAN DO AFTERWARDS

Detect and genotype oncogenic viruses, establish integration through junction evidence, localise integration sites, and assess host gene consequences.

WHO HIRES FOR IT

Head and neck and gynaecological oncology services, hepatology and liver cancer centres, virology-linked cancer laboratories, and research institutes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
16	**78 hours · 13 days**	Prior structural variant analysis experience	**1,80,000**

SON-35
WORKFLOW

Oncogenic Viral Integration Analysis

(**Viral integration and oncogenic virus analysis** · HPV/HBV/EBV integration analysis)

COMPLETE WORKFLOW — 16 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Viral reference selection**	The oncogenic virus references and genotypes in scope selected and recorded	Reference register	*Viral reference set*
02	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
03	**Host alignment**	Reads aligned to the human reference with unmapped reads retained	bwa-mem2	*Aligned BAM with unmapped*
04	**Non-host read extraction**	Reads failing to align to the host genome extracted for viral screening	Unmapped extraction	*Non-host read set*
05	**Viral screening**	Non-host reads screened against the viral reference set	Viral alignment, classification	*Viral hits*
06	**Contamination discrimination**	Laboratory and index contamination distinguished from genuine viral presence	Control comparison	*Contamination assessment*
07	**Viral load quantification**	Viral read fraction and estimated copy number quantified	Load calculation	*Viral load*
08	**Viral genotyping**	Viral genotype and subtype determined where clinically relevant	Genotyping analysis	*Genotype assignment*
09	**Chimeric read identification**	Read pairs spanning host and viral sequence identified	Chimeric analysis	*Chimeric read set*
10	**Junction assembly**	Host-viral junction sequences assembled from spanning reads	Local assembly	*Junction sequences*
11	**Junction read-level review**	Every candidate junction inspected at read level to exclude artefact	IGV review	*Reviewed junctions*
12	**Integration site localisation**	Integration sites localised to host genomic coordinates	Coordinate determination	*Integration sites*
13	**Episomal versus integrated discrimination**	Episomal presence distinguished from integration using junction evidence	Discrimination logic	*Integration status*
14	**Host gene impact assessment**	Host genes disrupted or activated by integration identified	Gene overlap analysis	*Host gene impact*
15	**Prognostic interpretation**	Viral status and integration interpreted for prognosis and management	Clinical interpretation	*Interpretation record*
16	**Reporting**	Viral status, load, genotype, integration and host impact assembled	Report template	*Viral integration report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
16	**78 h · 13 d**	**SON-35**	**1,80,000**

SON-36
SECTOR 01

Tissue-of-Origin Analysis for Cancer of Unknown Primary

(**Cancer of unknown primary tissue-of-origin analysis** · CUP analysis)

WHY IT IS DONE

Cancers presenting without an identifiable primary site are treated empirically and do poorly. Molecular tissue-of-origin prediction converts an untreatable-by-protocol diagnosis into one with a site-directed treatment pathway.

WHAT IT ANSWERS

Which tissue did this metastatic tumour originate from, with what confidence, and does the molecular profile also indicate actionable targets?

WHEN IT IS DONE

In metastatic disease where imaging, histology and immunohistochemistry have failed to identify a primary site, and where site-directed therapy would change management.

WHAT TRIGGERS IT

A confirmed cancer of unknown primary after standard workup, or a metastatic presentation with an ambiguous immunohistochemical profile.

HOW IT IS DONE

Expression, methylation or both are profiled and quality controlled with tumour content assessed. Classification against reference cohorts produces ranked predictions with calibrated confidence, mutational and copy number patterns are examined for corroboration, and the result is reconciled with clinical and immunohistochemical findings.

WHAT IT PRODUCES

Ranked tissue-of-origin predictions with calibrated probabilities, corroborating genomic patterns, actionable alteration findings, clinical and immunohistochemical reconciliation, and a report stating when confidence is insufficient.

WHAT IT DETECTS

Tissue-of-origin prediction with ranked probability from expression, methylation or both, supporting mutational and copy number patterns characteristic of the predicted site, and actionable alterations irrespective of origin.

WHAT IT DOES NOT DETECT

Origins absent from the reference cohort cannot be predicted, and poorly differentiated tumours frequently lose the tissue-specific signal that classification depends on. Prediction is probabilistic, and low tumour content or necrotic material degrades it substantially.

WHAT MAKES IT HARD

Confidence calibration decides clinical usefulness, since a confidently wrong prediction directs treatment down the wrong pathway. Corroborating the prediction against mutational patterns characteristic of the predicted site is the strongest available safeguard and is often skipped.

WHO DOES THIS JOB

Clinical bioinformatician with classification competence in cancer genomics, mid to senior level.

WHAT YOU CAN DO AFTERWARDS

Run tissue-of-origin classification with calibrated confidence, corroborate against genomic patterns, and integrate with clinical and pathological findings.

WHO HIRES FOR IT

Comprehensive cancer centres, unknown primary specialist services, molecular profiling providers, and academic oncology programmes.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
17	**82 hours · 14 days**	Prior expression or methylation classification experience	**1,90,000**

SON-36
WORKFLOW

Tissue-of-Origin Analysis for Cancer of Unknown Primary

(**Cancer of unknown primary tissue-of-origin analysis** · CUP analysis)

COMPLETE WORKFLOW — 17 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Referral confirmation**	Confirmation that standard workup has failed to identify a primary site	Clinical review	*Referral confirmation*
02	**Immunohistochemistry review**	Existing immunohistochemical results reviewed as the comparator	Pathology report	*IHC record*
03	**Specimen and tumour content assessment**	Tumour content and differentiation assessed as classification confounders	Pathology review	*Content record*
04	**Assay selection**	Expression, methylation or both selected according to specimen and reference availability	Assay selection criteria	*Selected assay*
05	**Data generation and quality assessment**	Data generated and quality assessed against platform criteria	Platform QC	*Quality record*
06	**Normalisation**	Data normalised using the method the classifier was trained with	Normalisation procedure	*Normalised data*
07	**Batch and technical correction**	Technical variation corrected before classification	Batch correction	*Corrected data*
08	**Reference cohort verification**	The reference cohort's tissue coverage confirmed to include plausible origins	Cohort review	*Cohort verification*
09	**Classification**	The profile classified against reference tissues with ranked probabilities	Classification model	*Ranked predictions*
10	**Confidence calibration**	Prediction confidence calibrated against outcome data	Calibration model	*Calibrated confidence*
11	**Differentiation confounding assessment**	Poor differentiation assessed as a cause of degraded tissue signal	Confounding review	*Differentiation assessment*
12	**Genomic corroboration**	Mutational and copy number patterns assessed against the predicted origin	Pattern comparison	*Corroboration record*
13	**Immunohistochemical reconciliation**	The molecular prediction reconciled against existing immunohistochemistry	Reconciliation review	*Reconciliation record*
14	**Actionable alteration assessment**	Actionable alterations identified irrespective of the origin prediction	OncoKB, therapy resources	*Actionable findings*
15	**Uninformative declaration**	Low-confidence predictions declared uninformative rather than reported as weak	Declaration criteria	*Informativeness decision*
16	**Clinical interpretation**	Prediction interpreted for site-directed treatment selection	Clinical interpretation	*Interpretation record*
17	**Report generation and sign-out**	Ranked prediction, confidence, corroboration and actionability issued under dual signature	Validated report template	*Signed clinical report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
17	**82 h · 14 d**	**SON-36**	**1,90,000**

SON-37
SECTOR 01

Single-Cell Tumour Analysis

(**Single-cell tumour profiling analysis** · scRNA tumour analysis)

WHY IT IS DONE

Bulk sequencing averages tumour, stroma and immune cells into a single profile that describes none of them. Single-cell analysis resolves the populations separately, which is how resistance subpopulations, immune states and tumour-microenvironment interactions become visible.

WHAT IT ANSWERS

What cell populations make up this tumour, which are malignant, what states are the immune and stromal cells in, and which subpopulations carry resistance potential?

WHEN IT IS DONE

In translational research on resistance and microenvironment, in immunotherapy mechanism studies, and increasingly in trial correlative programmes.

WHAT TRIGGERS IT

A translational research protocol, an immunotherapy mechanism study, or a trial correlative analysis requiring cellular resolution.

HOW IT IS DONE

Raw data are processed with ambient contamination and doublet removal, quality filtering and normalisation. Populations are clustered and annotated against reference profiles, malignant cells are separated by inferred copy number, tumour states and immune phenotypes are characterised, and interactions are inferred from ligand-receptor expression.

WHAT IT PRODUCES

The annotated cell atlas with population proportions, malignant cell identification with inferred copy number, tumour state and programme characterisation, immune and stromal phenotyping, and interaction analysis.

WHAT IT DETECTS

Cell populations with type assignment, malignant against non-malignant discrimination through inferred copy number, tumour cell states and programmes, immune and stromal composition and state, and cell-cell interaction inference.

WHAT IT DOES NOT DETECT

Cells lost during dissociation are absent from the data entirely, and dissociation bias systematically under-represents fragile and adherent populations. Spatial relationships are destroyed, and low-abundance populations fall below detection at typical cell numbers.

WHAT MAKES IT HARD

Separating malignant from normal cells is the crux and depends on inferred copy number, which is noisy in low-aneuploidy tumours. Dissociation bias is invisible in the data itself, so composition estimates must be interpreted against what the protocol is known to lose.

WHO DOES THIS JOB

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

WHAT YOU CAN DO AFTERWARDS

Process and quality control single-cell tumour data, separate malignant from normal cells, characterise states and immune phenotypes, and interpret composition against dissociation bias.

WHO HIRES FOR IT

Cancer research institutes, pharmaceutical translational oncology, academic cancer centres, and single-cell technology providers.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
22	**108 hours · 18 days**	Prior single-cell RNA-Seq analysis experience	**2,45,000**

SON-37
WORKFLOW

Single-Cell Tumour Analysis

(**Single-cell tumour profiling analysis** · scRNA tumour analysis)

COMPLETE WORKFLOW — 22 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Protocol and dissociation record capture**	Dissociation protocol recorded because it determines which cells survive to be measured	Protocol record	*Dissociation record*
02	**Specimen viability assessment**	Cell viability and yield assessed before library preparation	Viability assessment	*Viability record*
03	**Data receipt and quality assessment**	Integrity and sequence quality verified against acceptance criteria	md5sum, FastQC	*QC acceptance record*
04	**Alignment and counting**	Reads aligned and per-cell count matrices generated	CellRanger, STARsolo	*Count matrix*
05	**Empty droplet removal**	True cells distinguished from ambient-only barcodes	EmptyDrops	*Cell-called matrix*
06	**Ambient contamination correction**	Ambient transcript contamination estimated and corrected	SoupX, CellBender	*Corrected matrix*
07	**Doublet detection**	Multiplet barcodes detected and removed before clustering	scDblFinder	*Doublet-filtered matrix*
08	**Quality filtering**	Cells filtered on gene count, transcript count and mitochondrial fraction	QC filtering	*Filtered matrix*
09	**Normalisation and feature selection**	Counts normalised and variable features selected for dimensionality reduction	Normalisation, HVG selection	*Normalised matrix*
10	**Dimensionality reduction**	Principal components computed and a neighbour graph constructed	PCA, neighbour graph	*Reduced representation*
11	**Batch integration**	Samples integrated where multiple specimens are analysed together	Harmony, scVI	*Integrated embedding*
12	**Clustering**	Cells clustered into populations at a justified resolution	Leiden clustering	*Cell clusters*
13	**Cell type annotation**	Clusters annotated against reference profiles and marker evidence	Reference mapping, markers	*Annotated populations*
14	**Copy number inference**	Copy number inferred per cell to distinguish malignant from normal cells	inferCNV, CopyKAT	*Inferred copy number*
15	**Malignant cell identification**	Malignant cells identified from inferred copy number with confidence stated	Malignancy classification	*Malignant cell assignment*
16	**Tumour state characterisation**	Tumour cell programmes and states characterised within the malignant compartment	Programme analysis	*Tumour states*
17	**Immune phenotyping**	Immune populations phenotyped including exhaustion and activation states	Immune annotation	*Immune phenotypes*
18	**Stromal characterisation**	Stromal and endothelial populations characterised	Stromal annotation	*Stromal phenotypes*
19	**Composition quantification**	Population proportions quantified with dissociation bias stated as a caveat	Composition analysis	*Composition estimates*
20	**Cell-cell interaction inference**	Ligand-receptor interactions inferred between annotated populations	Interaction inference	*Interaction analysis*
21	**Resistance subpopulation assessment**	Populations carrying resistance-associated programmes identified	Programme scoring	*Resistance assessment*
22	**Reporting**	Atlas, malignant identification, states and interactions assembled with caveats	Report template	*Single-cell report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
22	**108 h · 18 d**	**SON-37**	**2,45,000**

SON-38
SECTOR 01

Spatial Tumour Profiling Analysis

(**Spatial transcriptomics analysis in oncology** · Spatial oncology analysis)

WHY IT IS DONE

Where cells sit relative to each other determines tumour behaviour — whether immune cells reach the tumour core, how the invasive margin is organised, how niches sustain resistance. Spatial profiling retains the geometry that dissociation destroys.

WHAT IT ANSWERS

How are cell populations arranged within this tumour, what distinguishes the tumour core from its margin, and which spatial niches relate to treatment response?

WHEN IT IS DONE

In immunotherapy mechanism research where immune exclusion is the question, in studies of the invasive margin and microenvironment, and in trial correlative work.

WHAT TRIGGERS IT

A research protocol on immune exclusion or microenvironment organisation, or a correlative study requiring spatial context.

HOW IT IS DONE

Spatial data are quality controlled with tissue registration confirmed, expression is normalised accounting for spatial technical gradients, and cell composition per location is estimated by deconvolution against reference profiles. Spatial domains are detected, spatially variable genes identified, and neighbourhood and interaction analysis performed with histology integrated.

WHAT IT PRODUCES

Spatially resolved expression with quality metrics, per-location composition estimates, spatial domain and niche assignments, spatially variable gene results, neighbourhood interaction analysis, and histology-integrated visualisations.

WHAT IT DETECTS

Spatially resolved expression, cell type composition per location by deconvolution, spatial domains and niches, spatially variable genes, immune exclusion patterns, and neighbourhood interactions.

WHAT IT DOES NOT DETECT

Resolution depends on the platform, and spot-based methods do not resolve single cells so composition is inferred rather than observed. Only the sampled section is represented, and gene panels in imaging-based platforms restrict analysis to designed targets.

WHAT MAKES IT HARD

Deconvolution is model-dependent and the reference profile choice materially changes composition estimates, so the reference must be stated. Spatial technical gradients across a section mimic biological gradients and must be controlled before any spatial conclusion is drawn.

WHO DOES THIS JOB

Computational biologist with spatial and single-cell competence, senior level. One of the newest and scarcest specialisms in this sector.

WHAT YOU CAN DO AFTERWARDS

Process spatial tumour data with technical gradient control, deconvolve composition with stated references, detect domains and niches, and integrate with histology.

WHO HIRES FOR IT

Cancer research institutes, pharmaceutical translational oncology, spatial technology providers, and academic immuno-oncology programmes.

COMMERCIALS

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

SON-38
WORKFLOW

Spatial Tumour Profiling Analysis

(**Spatial transcriptomics analysis in oncology** · Spatial oncology analysis)

COMPLETE WORKFLOW — 21 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Platform and resolution recording**	Platform and its spatial resolution recorded as the bound on all conclusions	Platform record	*Resolution record*
02	**Tissue section and registration verification**	Section quality confirmed and image registration to the expression grid verified	Registration check	*Registration record*
03	**Data receipt and quality assessment**	Integrity and platform quality metrics verified against criteria	Platform QC	*QC acceptance record*
04	**Segmentation or spot calling**	Cells segmented or spots called according to platform type	Segmentation, spot calling	*Spatial units*
05	**Per-unit quality filtering**	Spatial units filtered on transcript count and area with thresholds recorded	QC filtering	*Filtered units*
06	**Spatial technical gradient assessment**	Systematic gradients across the section assessed as confounders	Gradient analysis	*Gradient assessment*
07	**Normalisation**	Expression normalised accounting for spatial technical variation	Spatial normalisation	*Normalised expression*
08	**Reference profile selection**	The reference profile set for deconvolution selected and recorded	Reference selection	*Selected reference*
09	**Deconvolution**	Cell type composition estimated per spatial unit against the reference	cell2location, RCTD	*Composition estimates*
10	**Deconvolution sensitivity assessment**	Composition estimates assessed for dependence on the reference chosen	Sensitivity analysis	*Model dependence record*
11	**Spatial domain detection**	Spatially coherent domains detected across the section	BayesSpace, SpaGCN	*Spatial domains*
12	**Domain annotation**	Domains annotated as tumour core, margin, stroma or immune niche	Domain annotation	*Annotated domains*
13	**Spatially variable gene detection**	Genes with spatially structured expression identified statistically	SpatialDE, SPARK-X	*Spatially variable genes*
14	**Invasive margin analysis**	The tumour-stroma interface characterised as a distinct spatial region	Margin analysis	*Margin characterisation*
15	**Immune exclusion assessment**	Immune cell distribution relative to tumour assessed for exclusion patterns	Distribution analysis	*Exclusion assessment*
16	**Neighbourhood analysis**	Cell type co-occurrence quantified within spatial neighbourhoods	Neighbourhood analysis	*Neighbourhood results*
17	**Spatial interaction inference**	Ligand-receptor interactions inferred using spatial proximity constraints	Spatial interaction inference	*Interaction results*
18	**Histology integration**	Results integrated with the registered histological image	Image integration	*Integrated visualisations*
19	**Sampling limitation statement**	The single-section nature of the analysis stated as a limitation	Limitation template	*Limitation statement*
20	**Visualisation preparation**	Spatial maps prepared for interpretation and presentation	Visualisation	*Spatial figures*
21	**Reporting**	Domains, composition, exclusion and interactions assembled with caveats	Report template	*Spatial report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
21	**102 h · 17 d**	**SON-38**	**2,35,000**

SON-39
SECTOR 01

Oncology Pharmacogenomic Analysis

(**Chemotherapy pharmacogenomic analysis** · DPYD, UGT1A1, TPMT analysis)

WHY IT IS DONE

A small number of germline variants predict severe and occasionally fatal toxicity from widely used chemotherapy agents. Testing before the first dose prevents harm that is entirely predictable, and several health systems now mandate it.

WHAT IT ANSWERS

Does this patient carry variants predicting toxicity from the chemotherapy planned, and what dose adjustment does the applicable guideline recommend?

WHEN IT IS DONE

Before first administration of fluoropyrimidines, irinotecan or thiopurines, and following severe unexpected toxicity on any of them.

WHAT TRIGGERS IT

A planned prescription of an agent with pharmacogenomic guidance, or investigation of severe unexpected chemotherapy toxicity.

HOW IT IS DONE

Coverage is confirmed at all allele-defining positions, variants are called including repeat and structural alleles, and diplotypes are assigned against a versioned allele definition table. Phenotype is translated using the current activity score system and mapped to current dosing guidance with the guideline version recorded.

WHAT IT PRODUCES

Diplotype calls per gene with the allele definition version, metabolic phenotype translation, guideline-mapped dose recommendations with guideline version, ambiguity statement, and a durable record for the patient file.

WHAT IT DETECTS

Star-allele diplotypes at the relevant pharmacogenes, structural and repeat variants including promoter repeats, translated metabolic phenotype, and guideline-mapped dose recommendations.

WHAT IT DOES NOT DETECT

Only catalogued alleles are detected, so rare variants of unknown function default to normal and give false reassurance. It explains only the genetic component of toxicity, not organ function, interactions or adherence, and phenotype conversion by co-administered drugs is outside its scope.

WHAT MAKES IT HARD

Promoter repeat alleles and structural variation at these loci are missed by standard variant calling and need dedicated handling. Turnaround is the operational constraint, since a result arriving after the first dose has no preventive value.

WHO DOES THIS JOB

Clinical bioinformatician or pharmacogenomics scientist, mid level, working with oncology pharmacy.

WHAT YOU CAN DO AFTERWARDS

Call pharmacogene diplotypes including repeat and structural alleles, translate to phenotype correctly, and apply current dosing guidance within a clinically useful turnaround.

WHO HIRES FOR IT

Oncology pharmacy services, cancer centres, pharmacogenomics laboratories, and diagnostic chains offering pre-treatment testing.

COMMERCIALS

STEPS IN WORKFLOW	TRAINING DURATION	PREREQUISITE	TRAINING FEE (INR)
15	**72 hours · 12 days**	Prior variant analysis experience	**1,65,000**

SON-39
WORKFLOW

Oncology Pharmacogenomic Analysis

(**Chemotherapy pharmacogenomic analysis** · DPYD, UGT1A1, TPMT analysis)

COMPLETE WORKFLOW — 15 ORDERED STEPS

#	STEP	WHAT HAPPENS	TOOL / METHOD	OUTPUT
01	**Agent scope confirmation**	The chemotherapy agents in scope and their associated genes confirmed	Assay register	*Agent scope*
02	**Allele definition version lock**	Star-allele definition tables confirmed and versions recorded	Allele definition register	*Definition version*
03	**Turnaround requirement recording**	The clinical deadline recorded, since a late result has no preventive value	Clinical protocol	*Turnaround requirement*
04	**Coverage verification at defining positions**	Depth confirmed adequate at every allele-defining position	mosdepth	*Coverage certificate*
05	**Variant calling**	Variants called at defining positions including small insertions and deletions	Validated caller	*Variant set*
06	**Repeat allele analysis**	Promoter and intronic repeat alleles genotyped with a dedicated method	Repeat genotyping	*Repeat allele calls*
07	**Structural variant assessment**	Gene deletions, duplications and hybrids assessed at complex loci	Copy number analysis	*Structural findings*
08	**Phasing**	Variants phased into haplotypes where read or population data permit	Phasing method	*Phased haplotypes*
09	**Diplotype assignment**	Diplotypes assigned against the versioned allele definition table	PharmCAT	*Diplotype calls*
10	**Ambiguity documentation**	Unresolvable diplotypes and uncatalogued alleles documented explicitly	Ambiguity register	*Ambiguity statement*
11	**Phenotype translation**	Diplotypes translated to metabolic phenotype using current activity scores	Activity score tables	*Metabolic phenotype*
12	**Guideline mapping**	Phenotypes mapped to current dose recommendations for each agent	CPIC, DPWG guidelines	*Dose recommendations*
13	**Guideline version recording**	The guideline version applied recorded as recommendations are revised	Guideline register	*Guideline version*
14	**Durable record preparation**	Results formatted for permanent storage and reuse across future prescriptions	Structured record format	*Durable record*
15	**Report generation and sign-out**	Diplotypes, phenotype and dose recommendations issued under dual signature	Validated report template	*Signed report*

STEPS IN WORKFLOW	TRAINING DURATION	PROGRAMME CODE	TRAINING FEE (INR)
15	**72 h · 12 d**	**SON-39**	**1,65,000**

Build and staff a *molecular oncology service*

Each of the thirty-nine programmes in this catalogue trains one complete type of NGS data analysis as a molecular oncology laboratory 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

SON · Sector 02 of 16

SCOPE

39 analysis types · 771 steps

DURATION

3750 contact hours · 636 working days

COMPLETE CATALOGUE FEE

INR 55,85,000 (list 85,93,000)

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