

PROTOCOL

Developing Novel Genomic method for risk stratification and measurable residual disease assessment in Multiple Myeloma

Protocol version number with date: Protocol Version 1.0 dated 23.07.2026

Principal Investigator and CoPrincipal Investigator Name, Designation and Contact Details:

Name	Designation	Contact Details
Principal Investigator		
Dr. Bhausaheb Bagal MD, DM	Professor and Medical Oncologist	Tata Memorial Centre, Mumbai. E-mail: bagalbp@gmail.com
Co-Principal Investigator		
Dr. Prashant Tembhare	Clinician Scientist and Professor Hematopathology,	Tata Memorial Centre, Mumbai. E-mail: docprt@gmail.com
Dr. Nikhil Patkar	Clinician Scientist	Tata Memorial Centre, Mumbai. E-mail: nvpatkar@gmail.com
Dr. Navin Khattry	Professor, Medical Oncology	Tata Memorial Centre, Mumbai. E-mail: nkhattry@gmail.com
Dr. Pooja Prasad	Associate Professor, Medical Oncology	Homi Bhabha Cancer Hospital & Research Centre, Punjab. E-mail: poojaprasad1_02@rediffmail.com
Dr. Sushant Vinarkar	Assistant Professor, Pathology	Homi Bhabha Cancer Hospital & Research Centre, Punjab. E-mail: drsushantvinarkar@gmail.com
Dr. Jeevan Kumar	Senior Consultant, Department of Clinical Hematology & BMT	Tata Medical Centre, Kolkata. E-mail: jeevan.kumar@tmckolkata.com
Dr. Mayur Parihar	Senior Consultant, Department Lab Haematology and Cytogenetics	Tata Medical Centre, Kolkata. E-mail: mayurparihar@gmail.com

Institutional Name and Address:

Tata Memorial Centre,

Tata Memorial Hospital Dr. E Borges Road, Parel, Mumbai, Maharashtra 400012 and The Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Sector 22, Utsav Chowk-CISF Road, Kharghar, Navi Mumbai, Maharashtra. 410210

Collaborating Centre Name and Address:

- Homi Bhaba Cancer Hospital and Research Centre , New Chandigarh, Punjab.
- Tata Medical Centre, Kolkata.

Co- Investigators:

- Dr Lingaraj Nayak
- Dr Alok Shetty
- Dr Hasmukh Jain
- Dr Manju Sengar
- Dr Sumeet Mirgh
- Dr Dhanlaxmi Shetty
- Dr Sweta Rajpal
- Dr Pratusha Gudapati
- Dr PGS Subramanian
- Mrs.Sadhana Kannan

Contents

1. Background	5
2. Problem statement	5
3. Rationale of study	6
4. Study Objectives.....	6
5. Type of study Design.....	6
6. Methodology	7
6.1. Methodology For Objective 1	7
6.1.1. Study Design	7
6.1.2. Adaptive WGS	7
6.1.3. Translocations/CAN/Mutations	7
6.1.4. Sequencing.....	8
6.1.5. Bioinformatics Analysis	8
6.2. Methodology For Objective 2.....	8
6.2.1. Baseline Tumor Genotyping and PV Identification	8
6.2.2. Phased Variant Construction.....	8
7. Sample Collection for MRD	8
7.1. Timepoints for MRD sampling.....	8
7.2. Peripheral Blood (ccfDNA).....	9
7.3. Bone Marrow MRD	9
8. Study Area or Settings	10
9. Study Subjects	10
9.1. Inclusion Criteria	10
9.2. Exclusion Criteria	10
9.3. Withdrawal Criteria.....	10
10. Interventions	11
11. Safety Reporting	11
12. Primary and secondary outcome measures.....	11
12.1. Objective 1:.....	11
12.2. Objective 2 :.....	11
13. Sample Size estimation and sampling strategy.....	11
14. Data Collection methods	12

15.	Data quality assurance methods.....	12
15.1.	Pre-analytical:	13
15.2.	Analytical:	13
15.3.	Post-analytical:.....	13
15.4.	Data management plan:	14
16.	Statistical Analysis Plan.....	14
16.1.	Objective 1.....	14
16.2.	Objective 2.....	15
17.	Ethical Considerations	15
18.	Informed Consent	15
19.	Risks and Burden.....	15
20.	Benefit-Risk Assessment.....	16
21.	Confidentiality	16
22.	Genomic Data Considerations	16
23.	Compensation	16
24.	Vulnerable Populations	16
25.	Data Protection	16
26.	Plans for Publication of Results.....	16
27.	References	18

1. Background

Multiple myeloma (MM) is a malignant plasma cell disorder characterized by the clonal proliferation of abnormal plasma cells within the bone marrow. This leads to end-organ damage, classically manifested as anaemia, hypercalcemia, lytic bone lesions, and renal dysfunction (CRAB features). MM accounts for approximately 10–15% of all haematological malignancies and about 1–2% of all cancers worldwide. MM is a biologically complex disease and always preceded by premalignant conditions such as monoclonal gammopathy of undetermined significance (MGUS) and smouldering multiple myeloma (SMM). These precursor states represent a prolonged preclinical phase during which progressive genetic alterations accumulate and this increases further during course of treatment in clinical manifest phase. Thus, besides the presence of early “founder” genetic events, the disease demonstrates significant clonal heterogeneity and ongoing genomic evolution, contributing to its biological complexity. The clinical course of multiple myeloma is also highly heterogeneous. Some patients experience an indolent disease course with prolonged survival, while others exhibit aggressive disease characterized by early relapse and treatment resistance. Although therapeutic advances have significantly improved outcomes, multiple myeloma remains largely incurable, with most patients eventually relapsing and succumbing to the disease. The traditional staging and risk stratification systems such as the International Staging System (ISS) have primarily relied on measure of disease burden using biochemical parameters, including beta 2 macroglobulin and serum albumin. More recent refinements in risk stratification, including the Revised International Staging System (R-ISS) and updated classifications by the International Myeloma Working Group/International Myeloma Society (IMWG/IMS 2025), have incorporated key genomic abnormalities such as high-risk translocations (e.g., t(4;14), t(14;16)), TP53 alterations (deletion and mutation), and chromosomal abnormalities including gain(1q) and deletion(1p). These abnormalities are typically detected using fluorescence in situ hybridization (FISH). However, FISH has several inherent limitations. With the advent of novel therapeutic agents such as proteasome inhibitors, immunomodulatory drugs (IMiDs), monoclonal antibodies (e.g., anti-CD38), and more recently T-cell-engaging therapies including bispecific antibodies and CAR-T cells, outcomes in multiple myeloma have significantly improved. These advances have necessitated refinement of response assessment criteria to better capture deeper levels of disease control. The International Myeloma Working Group (IMWG) has accordingly updated response criteria, with increasing emphasis on achieving and accurately measuring deep responses, particularly at the level of measurable residual disease (MRD). Current response assessment incorporates, biochemical evaluation (serum/urine protein estimation, free light chains), bone marrow (BM)–based MRD assessment using highly sensitive techniques such as Multiparameter flow cytometry (next-generation flow, NGF), molecular methods such as next-generation sequencing (NGS) and imaging-based evaluation of extramedullary disease using functional modalities such as PET-CT and MRI MRD has emerged as a powerful prognostic marker like other hematological malignancies. MRD negativity is consistently associated with improved progression-free survival (PFS) and more importantly improved overall survival (OS). Besides this, the prognostic value of MRD is consistent across disease settings (newly diagnosed and relapsed/refractory myeloma) and is independent of treatment modality, retaining predictive value in both transplant-eligible and transplant-ineligible patients and applicable across therapeutic classes, including conventional therapies as well as novel immunotherapies and T cell redirecting therapies.

2. Problem statement

Despite significant advances in risk stratification, the current methods fail to accurately and comprehensively capture the genomic complexity of multiple myeloma resulting in suboptimal risk stratification. FISH, the current method for detection of prognostically relevant genomic alterations have several limitations including limited sensitivity, restriction to predefined genomic loci, inability to accurately quantify the fraction of tumour cells harboring specific abnormalities, and lack of resolution to detect sequence-level mutations such as TP53 point mutations. As a result, FISH based risk stratifications have several limitation in the current required stratification and most of the myeloma centers globally have shifted to sequencing based methods. Despite the well-established prognostic value of MRD, its accurate and standardized assessment remains a significant unmet challenge in routine

clinical practice. Currently, the most widely used method for MRD detection NGF, has a sensitivity is typically limited to 10^{-5} , and achieving higher sensitivity depends heavily on sample quality and cell yield and it requires significant technical expertise in multiparametric analysis. Importantly, the desirable threshold of 10^{-6} sensitivity is often difficult to achieve in real-world settings. Molecular MRD using NGS, although has ability to detect MRD to the threshold of 10^{-6} , however require baseline sample and is currently not available in India.

3. Rationale of study

Long-read sequencing using Oxford Nanopore Technology (ONT) particularly through adaptive whole genome sequencing (adaptive WGS) enables comprehensive, genome-wide profiling, including detection of structural rearrangements and phased haplotypes, offering a more complete characterization of tumour biology. In addition, ONT allows real-time sequencing with rapid turnaround times, flexible throughput, and reduced dependence on extensive laboratory infrastructure, making it potentially more cost-effective and scalable compared to conventional short length multi-step approaches. Thus adaptive WGS has potential to provide clinically deployable yet comprehensive and affordable risk stratification for newly diagnosed MM.

Phased variant-based sequencing (PhasED-Seq) represents a novel strategy for MRD detection by leveraging tumour-specific haplotypes, wherein multiple linked mutations on the same DNA molecule are tracked simultaneously. This approach is particularly well suited for B-cell malignancies, which are characterized by a high burden of somatic mutations and clonal diversity, enabling the identification of patient-specific phased variant signatures. Notably, PhasED-Seq has been successfully developed and clinically validated in diffuse large B-cell lymphoma (DLBCL), where it has demonstrated ultra-sensitive detection of circulating tumor DNA and strong prognostic utility. Therefore, developing and validating a nanopore-based PhasED sequencing assay for MRD detection has the potential to provide a higher sensitive technique (potentially to 10^{-7}) in BM and PB

4. Study Objectives

Objective 1 :Develop and evaluate assay for comprehensive genomic profiling and risk stratification of multiple myeloma beyond conventional cytogenetic approaches using adaptive sequencing by nanopore technology

Objective 2 :Develop phased sequencing-based assays for ultra-sensitive measurable residual disease detection in multiple myeloma through tracking of tumor-specific haplotypes and evaluate performance against conventional MRD methods

5. Type of study Design

This is a multicentre, prospective, non-interventional (observational), translational, diagnostic accuracy and assay-development/validation study. No experimental therapeutic intervention is administered and patients are not randomized; all samples analysed are obtained from routine diagnostic and disease-monitoring procedures. The study addresses two linked objectives conducted in parallel: (i) a cross-sectional method-comparison study evaluating adaptive whole-genome sequencing (WGS) against conventional FISH-based cytogenetic risk stratification at diagnosis in newly diagnosed multiple myeloma; and (ii) a paired, method-comparison (concordance) study evaluating PhasED-Seq-based measurable residual disease (MRD) detection in peripheral blood and bone marrow against standard next-generation flow (NGF) MRD, with longitudinal follow-up of participants for correlation with progression-free and overall survival.

6. Methodology

6.1. Methodology For Objective 1

6.1.1. Study Design

Development and validation of adaptive WGS for risk stratification Plasma cells from newly diagnosed MM will be enriched using CD138+ magnetic beads followed by adaptive WGS by long read sequencing using ONT as detailed below.

Sample collection and processing:

- 1) Bone marrow aspirates and peripheral blood samples will be collected in EDTA tubes.
- 2) Mononuclear cells will be isolated using density gradient centrifugation.
- 3) Plasma cells will be enriched using CD138-based magnetic bead selection.
- 4) High-molecular-weight genomic DNA will be extracted using Maxwell RSC genomic DNA kit (Promega, WI).
- 5) The library preparation workflow (Ligation Sequencing kit) includes DNA end repair, adapter ligation and post ligation cleanup of the library.

6.1.2. Adaptive WGS

Sequencing Platform Adaptive whole genome sequencing will be performed using nanopore-based long-read sequencing technology using Oxford Nanopore PromethION 2 Solo sequencer with R10.4.1 flow cells (FLO-PRO114M), enabling real-time data acquisition and adaptive sampling generating an expected yield of ~80–150 Gb per flow cell under standard conditions, with potential yields exceeding 150 Gb depending on DNA quality and run performance, with 1–3 samples per flow cell for whole genome sequencing (or up to 4–6 samples with adaptive sampling-guided multiplexing depending on coverage requirements), and an anticipated mean read length (N50) of 30–50 kb when high molecular weight DNA (>50 kb fragment size) is used Adaptive Sampling Strategy

Adaptive WGS will be implemented using real-time selective sequencing to enrich genomic regions of interest while retaining genome-wide coverage.

6.1.3. Translocations/CAN/Mutations

Regions containing MM translocation hotspots within the IGH locus will be designed and obtained from a literature review of known breakpoint regions. Shall covered common and rare MM translocations that are screened by FISH such as,

- c-MAF [t(14;16)]
- MAFB [t(14;20)]
- FGFR/MMSET [t(4;14)]
- CCND1/CCND2/CCND3 [t(11;14), t(12;14), t(6;14)]
- cMYC alterations
- Copy number alteration hotspots (e.g., 1q gain, 1p del, del17p, del13q)
- Genes implicated in myeloma pathogenesis (e.g., KRAS, NRAS, TP53, BRAF)
- Genome-Wide Copy Number Alterations

Based on SNPs in European and African Populations, with probes as equally distributed as close to 1MB

- Immunoglobulin loci (IGH, IGK, IGL)

6.1.4. Sequencing

DNA molecules will be computationally aligned in real time to the reference genome. Reads mapping to target regions will be preferentially retained. Off-target reads will be selectively rejected to improve sequencing efficiency and depth over regions of interest.

Sequencing Depth and Coverage

- Targeted regions: ≥ 50 – $100\times$ effective coverage
- Genome-wide background coverage: ~ 5 – $15\times$

Sequencing will be continued until predefined coverage thresholds are achieved.

6.1.5. Bioinformatics Analysis

Base calling will be performed using Dorado with the super-accuracy (SUP) model then Nano plot or FastQC will be used for raw read QC, followed by its alignment to GRCh37 human reference genome using Minimap2 and then detection of fusions using JAFFAL, LongGF and CtatLR. CNV calling using cuteSV or CNVkit and finally targeted SNV and insertion/deletion calling using deepvariant, ClairS and medaka.

6.2. Methodology For Objective 2

PhasEDSeq (Phased Variant–based) MRD Detection in Multiple Myeloma.

6.2.1. Baseline Tumor Genotyping and PV Identification

In diagnostic BM samples, CD138+ plasma cells will be isolated from bone marrow aspirates. Tumor DNA will undergo adaptive WGS using the Oxford Nanopore PromethION platform as described in Objective 1 for identification of somatic SNVs, indels, and structural variants (e.g., IGH translocations). The detection of PVs will be done using short read sequencing (Illumina/Aviti) using a capture-based panel.

6.2.2. Phased Variant Construction

We will be using the AULE framework for these purposes. Starting from raw FASTQs, the AULE pipeline proceeds through to minimal residual disease (MRD) quantification, employing UMI-based error modelling to detect single nucleotide variants leveraging phased variant tracking — as described by Kurtz et al. (2021) — to monitor tumour specific MRD status for individual patients.

7. Sample Collection for MRD

7.1. Timepoints for MRD sampling

MRD will be evaluated in PB and BM post consolidation in NDMM patients. This will be post-consolidation therapy, such as autologous stem cell transplantation in transplant-eligible patients or systemic therapy regimens (e.g., VRD, Dara-VRD) in transplant-ineligible patients after approximately 8–12 cycles of therapy. We will include patients who have achieved VGPR or better and are undergoing BM MRD as part of routine clinical care. At the

same time an extra BM and peripheral blood sample will be collected for PhasEdSeq MRD evaluation. Since this is to be correlated with BM NGF MRD, we will have to collect at an identical timepoint of BM MRD (+/- 3 days).

7.2. Peripheral Blood (ccfDNA)

Approximately 30 mL of peripheral blood will be collected in cfDNA stabilization tubes (e.g., Streck tubes) for ctDNA-based MRD assessment. This volume is necessary to obtain sufficient circulating tumor DNA (ctDNA) molecules for ultra-sensitive MRD detection. The analytical sensitivity of ctDNA MRD assays is directly dependent on the number of genome equivalents sampled. Increasing plasma volume increases the number of available genome equivalents and improves the achievable limit of detection. This approach is consistent with published literature on ultra-sensitive ctDNA-based MRD detection (PMID: 37349125).

Plasma will be separated within 72 hours by double centrifugation. cfDNA will be extracted using QIAamp Circulating Nucleic Acid Kit (Qiagen) or equivalent validated kits, with input cfDNA quantified using Qubit High Sensitivity assays. Typical cfDNA yield is expected to range between 10–50 ng per 10 mL blood. Given the fragmented nature of ccfDNA, sequencing will prioritize maximizing molecule recovery rather than read length.

7.3. Bone Marrow MRD

Bone marrow aspirates will be collected at predefined timepoints. Both bulk mononuclear cells and CD138-enriched plasma cells (where feasible) will be analyzed.

Library Preparation and Sequencing

Sequencing will be performed on the Illumina Sequencers, targeting Plasma cfDNA with maximal achievable depth (target ≥ 30 –60 Gb per sample depending on yield) and Bone marrow DNA: ≥ 30 Gb per sample.

Bioinformatics Pipeline: integrates UMI sequences from separate FASTQ files, performs BWA-MEM alignment with base quality recalibration following GATK Best Practices, and marks duplicates using UMI information (RX tags) to produce analysis-ready BAMs. These BAMs are then passed through a UMI collapse module that generates consensus reads from UMI families, applying configurable family-size and quality thresholds to correct sequencing errors and substantially reduce artifacts. The core analytical step is the phased variants module, which performs patient-specific variant discovery following the Kurtz et al. 2021 methodology, filters candidates against a panel of normals (PON) and gnomAD population frequencies, and quantifies MRD with statistical confidence scoring. In parallel, the plasma scoring module quantifies tumour-informed variants in plasma cfDNA using both raw and UMI-collapsed BAMs with population frequency filtering. Finally, the output tables module consolidates results across all modules into standardised reports with QC metrics.

MRD Quantification:

MRD burden will be expressed as the fraction of PV-positive reads divided by total informative reads, where informative reads are defined as high-quality reads spanning at least one PV locus.

Limit of Detection (LoD):

A patient-specific limit of detection (LoD95) will be calculated based on the number of informative reads and number of trackable PVs and LoD will be reported individually using a binomial model.

Definition of MRD Positivity:

- **MRD-positive:** Detection of ≥ 1 high-confidence PV-supported read above background threshold
- **MRD-negative:** No detectable PV signal within assay sensitivity limits
- **MRD-unevaluable:** Insufficient informative reads (minimum threshold: $\geq 50,000$ informative reads)

8. Study Area or Settings

This is a multicentric, investigator-initiated collaborative study coordinated by Tata Memorial Centre (TMC), Mumbai. The collaborating sites include Tata Medical Center (Kolkata) and Homi Bhabha Cancer Hospital & Research Centre (HBCHRC), Mullanpur. Eligible patients will be recruited at all participating centers during routine clinical care. Clinical samples (bone marrow aspirates and peripheral blood) will be collected during routine diagnostic evaluation and disease monitoring at each participating site. Initial sample processing, including plasma cell enrichment, DNA extraction, and standard-of-care investigations such as FISH and next-generation flow (NGF) MRD, will be performed at the respective participating centers according to standardized study protocols. Advanced genomic analyses, including adaptive whole-genome sequencing, PhasED-Seq assay development, centralized bioinformatics analyses, data integration, and overall study coordination will be performed at the Hematopathology Laboratory and Bioinformatics Facility of Tata Memorial Centre, Mumbai, which will serve as the coordinating center. Advanced wet laboratory work, genomic sequencing, and computational analyses will be carried out at the Hematopathology Laboratory and Bioinformatics Facility of Tata Memorial Centre, Mumbai.

9. Study Subjects

9.1. Inclusion Criteria

1. Patients with newly diagnosed multiple myeloma (NDMM) diagnosed according to the International Myeloma Working Group (IMWG) diagnostic criteria.
2. Age ≥ 18 years, irrespective of sex.
3. Willing and able to provide written informed consent.
4. Availability of an adequate bone marrow aspirate sample for genomic analysis.
5. For the MRD objective, patients achieving VGPR or better according to the IMWG response criteria, who have completed consolidation therapy (ASCT for transplant-eligible patients or approximately 8–12 cycles of standard systemic therapy for transplant-ineligible patients) and are undergoing routine bone marrow MRD assessment as part of standard clinical care.

9.2. Exclusion Criteria

1. Patients who do not provide written informed consent.
2. Patients with inadequate or poor-quality bone marrow samples for genomic analysis.
3. For the MRD objective, patients who have not achieved **VGPR or better**, are not undergoing routine bone marrow MRD assessment, or have inadequate bone marrow and/or peripheral blood samples for PhasED-Seq analysis.

9.3. Withdrawal Criteria

Participants are free to withdraw consent and discontinue participation in the research component of the study at any time, without giving reasons and without any effect on their ongoing standard clinical care. Upon withdrawal, no further research samples will be collected from the participant. Data and sample analyses already completed prior to withdrawal will be retained and may be used in the final analysis, in a de-identified form, unless the participant specifically requests destruction of stored samples, which will be honoured to the extent technically feasible. Participants may also be withdrawn by the investigator if they become ineligible (e.g., inadequate sample quality) or if continued participation is no longer in their interest. All withdrawals and the reasons for them will be documented.

10. Interventions

This study is non-interventional. No therapeutic intervention will be administered as part of this study, and all patients will receive treatment, monitoring, and response assessment according to standard institutional protocols and IMWG recommendations for multiple myeloma. Participation in the study will not influence treatment decisions or clinical management.

The study procedures involve additional genomic analyses performed on bone marrow samples obtained during routine diagnostic evaluation or disease monitoring and no additional bone marrow will be performed. Similarly, MRD assessment will be done for patients undergoing routine MRD assessment and no additional procedure is required.

Comparison : Comparison of adaptive WGS versus conventional FISH based risk stratification and NGF MRD versus PhasEd-seq MRD in PB and BM.

11. Safety Reporting

This is a non-interventional, observational translational study; no investigational drug, device, or therapeutic procedure is administered, and all treatment decisions remain per standard institutional and IMWG-recommended practice. Consequently, formal adverse event (AE) / serious adverse event (SAE) reporting as defined for interventional trials is not applicable.

12. Primary and secondary outcome measures

12.1. Objective 1:

Primary objective: Comparison of adaptive WGS versus conventional FISH based risk stratification

12.2. Objective 2 :

Primary outcomes: Comparison of PhasEd-Seq MRD in PB and BM versus NGF MRD Secondary objectives: PFS and OS of MRD negative versus positive cohort.

13. Sample Size estimation and sampling strategy

This study is a prospective observational translational genomics study aimed at assay development and comparative evaluation and there is no formal hypothesis testing. Therefore, the sample size is based on feasibility and expected patient accrual at the study center, while ensuring adequate representation of genomic risk groups.

Tata Memorial Centre evaluates approximately 120–150 new cases of multiple myeloma annually. Similarly, TMC Kolkata and HBCHRC Mullanpur evaluates 100 new cases each. Over the proposed study duration, we anticipate enrolling approximately 450 patients with newly diagnosed multiple myeloma for genomic profiling under Objective 1.

This sample size is expected to provide adequate representation of commonly observed cytogenetic risk groups, including high-risk abnormalities such as del(17p), t(4;14), t(14;16), and gain/amplification of 1q, which together occur in approximately 25–30% of patients. This will allow meaningful comparison between conventional FISH-based risk stratification and long-read genomic profiling.

For Objective 2, a subset of patients achieving deep responses (CR or VGPR) and undergoing routine MRD assessment will be included. Based on institutional experience, approximately 65-75% of treated patients reach

MRD timepoints, and therefore we anticipate that 300 patients will be evaluable for MRD analysis using phased sequencing approaches.

Among patients undergoing MRD assessment, institutional experience suggests MRD positivity rates of approximately 60–65% in patients treated without autologous stem cell transplantation (ASCT) and 35–40% in patients undergoing ASCT-based therapy. Based on these expected rates, the anticipated MRD cohort will include an adequate number of both MRD-positive and MRD-negative cases to enable meaningful comparison between phased sequencing-based MRD detection and next-generation flow cytometry (NGF). This sample size is therefore considered sufficient to evaluate concordance and assess the ability of phased sequencing to detect residual disease relative to standard MRD methodologies.

14. Data Collection methods

Clinical data will be collected prospectively from institutional electronic medical record systems. Baseline variables will include age, sex, disease stage (ISS/R-ISS), laboratory parameters (hemoglobin, serum calcium, creatinine, β 2-microglobulin, LDH), serum protein electrophoresis, immunofixation, and serum free light chain measurements. Treatment details and response assessments will be recorded according to International Myeloma Working Group (IMWG) response criteria (2016).

Bone marrow aspirate samples obtained during routine diagnostic evaluation and MRD assessment will be used for laboratory measurements. Plasma cell enrichment will be performed using CD138 magnetic bead-based cell separation systems using Miltenyi Biotec AutoMACS. DNA and RNA will be extracted using automated extraction systems such as (QIAasympphony, Qiagen).

Conventional cytogenetic risk assessment will be performed using fluorescence in situ hybridization (FISH) assays with probes targeting clinically relevant abnormalities including del(17p), t(4;14), t(14;16), t(14;20), gain/amplification of 1q21, and del(1p). Measurable residual disease (MRD) assessment as part of routine clinical care will be performed using next-generation flow cytometry (NGF) on multiparameter flow cytometers as per standardised institutional protocol.

Genomic measurements will be obtained using long-read sequencing platforms from Oxford Nanopore Technologies, using MinION sequencers. Sequencing output will include long-read genomic data enabling measurement of structural variants, copy number alterations, and somatic mutations. For MRD assessment using phased sequencing, tumor-specific haplotypes identified at diagnosis will be measured in follow-up samples using the same sequencing platform.

All clinical and genomic measurements will be captured and stored in institutional research databases and computational infrastructure.

15. Data quality assurance methods

Several measures will be implemented to ensure the accuracy, reliability, and completeness of clinical and laboratory data collected in this study.

Clinical data will be recorded using standardized case record forms (electro and verified against institutional electronic medical records. Data entry will be performed by trained research personnel and periodically reviewed by the study investigators to ensure completeness and consistency. Key clinical variables, response assessments, and laboratory parameters will be recorded according to International Myeloma Working Group (IMWG) diagnostic and response criteria.

Laboratory procedures will be conducted using standardized operating protocols. Plasma cell enrichment, nucleic acid extraction, and conventional cytogenetic testing will follow validated institutional protocols. Fluorescence in situ hybridization (FISH) and next-generation flow cytometry (NGF) assays used for clinical risk stratification and MRD assessment are routinely performed in the institution and undergo regular internal quality control and external proficiency testing. In fact the standardization of NGF given the required sensitivity, has been validated and published from our institute by our group.

For genomic sequencing, quality control will be implemented at multiple stages including nucleic acid quantification and integrity assessment, library preparation controls, and sequencing run quality metrics. Long-read sequencing data generated on Oxford Nanopore platforms will undergo standard quality filtering and read quality assessment before downstream analysis.

Bioinformatics pipelines will incorporate automated quality checks for read depth, alignment quality, and variant calling confidence. Reproducible computational workflows will be implemented to ensure consistency of analysis across samples.

Periodic data audits and cross-verification between clinical records, laboratory reports, and sequencing datasets will be performed to maintain overall data integrity throughout the study.

We have built a multi-center reproducibility and quality assessment for this proposal. All the centers perform gold-standard FISH and NGF MRD and institutional FISH/NGF assays are already undergoing external QC check.

For Adaptive sequencing and PhsEdSeq MRD we plan to standardise the work flow to ensure reproducibility as follows.

15.1. Pre-analytical:

Common SOPs for CD138+ enrichment, DNA extraction, library prep, and adaptive sampling; the target BED, reference genome, and Dorado/MinKNOW versions distributed as locked, checksummed files so every site verifies identical inputs; common run-level QC gates. Similarly, we will use standard set of probes for FISH analysis as orthogonal gold standard assays.

15.2. Analytical:

A single shared, version-controlled, containerized pipeline (<https://github.com/patkarlab/mm-awgs-nextflow>), with every step in pinned environments and each campaign run from one tagged release, so results trace to a frozen software state; offered to partner sites via the coordinating center's cloud facility. The PhasED-Seq/AULE pipeline is similarly pinned with shared panel-of-normals and common LoD/MRD positivity definitions.

15.3. Post-analytical:

Harmonized annotation, classification, and reporting. A shared characterized control runs periodically at each site, and inter-laboratory agreement will be tested by wet sample exchange (same DNA sequenced at two sites) and in-silico exchange (same data analysed at multiple sites), reported as Cohen's kappa against defined thresholds with discordances adjudicated against the gold standard. Partner-site results will have to be reproduced at the coordinating center before pooling.

The coordinating center will maintain a version registry for SOPs, BEDs, and pipeline releases; any change is re-validated before network-wide adoption.

All material is exchanged de-identified under appropriate MTAs and local ethics approval.

15.4. Data management plan:

All study data will be recorded using standardized case record forms (electronic) and entered into a secure electronic database maintained on institutional servers at Tata Memorial Centre. Each participant will be assigned a unique study identification number, and identifying information will be stored separately from research datasets to maintain confidentiality.

Clinical data, laboratory findings, and sequencing outputs will be linked through this study identification number to allow integrated analysis while ensuring that personal identifiers are not used in research datasets. Access to the database will be restricted to authorized study investigators and research personnel, with password-protected access controls.

Sequencing data generated during the study will be stored on the institution's secure high-compute facility with routine data backup systems in place. File organization and metadata documentation will follow standardized naming conventions to ensure traceability of samples and datasets. We will store raw fastqs as well as bam files. Once published the bam files and metadata will be publicly made available.

Regular checks will be performed to ensure completeness and consistency of entered data. Any discrepancies identified during review will be verified against source documents and corrected accordingly.

All study data will be stored in accordance with institutional data protection policies and applicable ethical guidelines. Only de-identified data will be used for research analysis and reporting. Data will be retained securely for the duration required by institutional and regulatory guidelines.

16. Statistical Analysis Plan

16.1. Objective 1

Baseline demographic and clinical characteristics will be summarized using descriptive statistics. Continuous variables will be presented as mean $\pm$ standard deviation or median (interquartile range), while categorical variables will be expressed as frequencies and percentages. Genomic abnormalities identified by adaptive whole genome sequencing (WGS) will be compared with conventional fluorescence in situ hybridization (FISH). Agreement between the two methods for overall risk classification will be assessed using Cohen's kappa coefficient with corresponding 95% confidence intervals. Kappa values will be interpreted as poor (<0.20), fair ($0.21-0.40$), moderate ($0.41-0.60$), good ($0.61-0.80$), and excellent (>0.80) agreement. Difference in paired detection of individual genomic abnormalities will be evaluated using McNemar's test, and the additional genomic abnormalities identified by adaptive WGS beyond conventional FISH will be reported descriptively.

Progression-free survival (PFS) and overall survival (OS) will be estimated using the Kaplan–Meier method and compared using the log-rank test according to genomic risk groups defined by conventional FISH and adaptive WGS. Univariable Cox proportional hazards regression analysis will be performed to evaluate the association between genomic risk classification and survival outcomes. Multivariable Cox regression models will subsequently be constructed after adjustment for clinically relevant covariates to determine the independent prognostic value of adaptive WGS-based risk stratification. All statistical analyses will be two-sided, with a p value <0.05 considered statistically significant. All statistical analyses will be performed using R statistical software version 4.5.3.

16.2. Objective 2

Baseline demographic, clinical, and treatment characteristics will be summarized using descriptive statistics. Continuous variables will be presented as mean $\pm$ standard deviation or median (interquartile range), while categorical variables will be expressed as frequencies and percentages. Agreement between PhasED sequencing and next-generation flow cytometry (NGF) for MRD assessment will be evaluated using Cohen's kappa coefficient with corresponding 95% confidence intervals. Using NGF as the comparator, the sensitivity, specificity, of PhasED sequencing will be estimated. Differences in paired MRD detection rates will be compared using McNemar's test, and quantitative MRD levels, where available, will be assessed using Spearman's rank correlation coefficient. The frequency and proportion of patients achieving a very good partial response or better ($\geq$ VGPR) and measurable residual disease (MRD) negativity will be summarized and compared between study groups using the Chi-square test or Fisher's exact test, as appropriate.

Progression-free survival (PFS) and overall survival (OS) will be estimated using the Kaplan–Meier method and compared using the log-rank test. Univariate Cox proportional hazards regression analysis will be performed to evaluate factors associated with PFS and OS. Variables demonstrating statistical significance in univariate analysis will subsequently be entered into multivariable Cox regression models to identify independent prognostic factors. All statistical analyses will be two-sided, with a p value <0.05 considered statistically significant. All statistical analyses will be performed using R statistical software version 4.5.3.

17. Ethical Considerations

Written informed consent will be obtained from all participants before enrollment. Study samples will be collected as part of routine clinical care, with no additional visits or invasive procedures required solely for research. The study protocol will be approved by the Institutional Ethics Committees (IECs) of Tata Memorial Centre and collaborating sites before initiation and registered with the Clinical Trials Registry–India (CTRI) prior to enrollment of the first participant. The study will be conducted in accordance with the Declaration of Helsinki, ICMR National Ethical Guidelines (2017), and ICH-GCP. Participant confidentiality will be maintained through de-identification of data, with secure storage and access restricted to authorized study personnel only.

18. Informed Consent

Written informed consent will be obtained from every participant by a trained study investigator/co-investigator prior to any study-related sample collection or data use. The Patient Information Sheet and Consent Form will be provided in English and the local vernacular language, and adequate time will be given for the participant to ask questions before signing. Participation is entirely voluntary; participants may withdraw at any stage as described under Withdrawal Criteria.

19. Risks and Burden

As this is a non-interventional, observational translational study, no additional invasive procedures are performed solely for research purposes beyond an additional peripheral blood draw of approximately 30 mL for ctDNA-based and 4 ml bone marrow sample for PhasEdMRD assessment. This will be collected at time of routine blood draw and marrow and hence no other risk is anticipated.

20. Benefit-Risk Assessment

Participants are not expected to receive any direct clinical benefit, as research findings will not guide treatment. This observational study involves no experimental intervention; the only additional procedures are collection of an extra peripheral blood sample (~30 mL) and bone marrow sample (~4 mL) during routine sampling, with minimal risk of temporary discomfort or bruising. Risks to genomic data confidentiality will be minimized through de-identification and secure data handling. Given the minimal additional risk and the potential to improve future genomic risk stratification and MRD assessment in multiple myeloma, the overall risk-benefit profile is considered favourable.

21. Confidentiality

Each participant will be assigned a unique de-identified study ID. Identifying information will be stored separately from clinical and genomic research data, with restricted, password-protected access limited to authorized study personnel. Samples shared between collaborating sites for sequencing/QC purposes will be transferred in de-identified form under an appropriate Material Transfer Agreement (MTA) and with local ethics approval.

22. Genomic Data Considerations

Given the nature of genomic sequencing, the possibility of incidental findings (e.g., unrelated germline variants) exists. The study is not designed to actively screen for or return incidental findings, and no individual genomic results will be returned to participants or treating physicians outside the research context. Sequencing data (BAM files) and associated metadata will be deposited in public repositories only after complete de-identification and only following publication, in accordance with journal and funder data-sharing requirements. Residual DNA/samples may be stored for future ethics-approved research; this will be explicitly consented for as a separate biobanking clause within the informed consent form, including the option for the participant to opt out of future-use storage while still participating in the primary study.

23. Compensation

Since this is a non-interventional study using samples collected as part of routine care, no financial compensation is planned for participation. There will be no additional cost to the participant for the extra blood draw or research-related genomic testing.

24. Vulnerable Populations

No vulnerable populations (minors, cognitively impaired individuals, prisoners, or economically/socially disadvantaged groups requiring special safeguards) are specifically targeted; all participants are adults capable of independent informed consent.

25. Data Protection

All data handling will comply with applicable institutional data protection policies and the Digital Personal Data Protection Act, 2023 (India), in addition to ICMR data-sharing guidelines. Only de-identified data will be used for research analysis and reporting.

26. Plans for Publication of Results

The results of this study will be disseminated through peer-reviewed scientific publications and presentations at national and international scientific/medical conferences, irrespective of whether the findings are positive, negative,

or inconclusive. Authorship will be determined in accordance with ICMJE authorship criteria. All publications and presentations will report only de-identified, aggregate data, and no individual participant will be identifiable in any publication, presentation, or report.

Project Implementation Plan along with milestone chart:

Study objectives	Year 1	Year 2	Year 3	Year 4
------------------	--------	--------	--------	--------

Objective 1: Development and validation of adaptive WGS for risk stratification	1. Logistics setup, SOP development, ethics approvals. 2. Sample collection NDMM patient 3. Optimization of CD138 enrichment, DNA and Bioinformatics pipeline development (SV, CNV, SNV) 4. Development of adaptive WGS workflow and establishment of aWGS at other centres.	1. Sample collection NDMM patient 2. Exchange of characterised sample to ensure reproducibility of adaptive WGS workflow 3. aWGS of recruited MM samples	1. Sample collection NDMM patient 2. Continuation of data analysis and reporting. 3. aWGS of recruited MM samples	1. Data analysis and reporting, and publication
Objective 2 :PhasED-Seq (Phased Variant-based) MRD Detection in Multiple Myeloma	1. Logistics setup, SOP development, and ethics approvals. 2. Baseline tumor profiling and phased variant (PV) identification 3. Development of PhasED-Seq pipeline (AULE + UMI). 4. Standardization of cfDNA and BM MRD workflows.	1. Baseline tumor profiling and phased variant (PV) 2. Development of PhasED-Seq pipeline (AULE + UMI). 3. Standardization of cfDNA and BM MRD workflows. 4. Initiation of data analysis and reporting.	1. MRD sequencing and analysis (PhasED-Seq). 2. Sensitivity, limit of detection (LoD), and concordance analysis. 3. Comparison of flow MRD with PV-MRD	1. Final analysis, reporting, and publication.

27. References

1. Rajkumar SV. Multiple myeloma: 2022 update on diagnosis, risk stratification, and management. Am J Hematol. 2022;97(8):1086-1107.
2. 2. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for multiple myeloma: a report from International Myeloma Working Group. J Clin Oncol. 2015;33(26):2863-2869
3. 3. Walker BA, Boyle EM, Wardell CP, et al. Mutational spectrum, copy number changes, and outcome: results from a sequencing study of patients with multiple myeloma. J Clin Oncol. 2015;33(33):3911-3920
4. 4. Kumar S, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol. 2016;17(8):e328-e346

5. 5 . Munshi NC, Avet-Loiseau H, Rawstron AC, et al. Association of minimal residual disease with superior survival outcomes in patients with multiple myeloma: a meta-analysis. *JAMA Oncol.* 2017;3(1):28-35.
6. 6. Kurtz DM, Scherer F, Jin MC, et al. Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nat Biotechnol.* 2021;39(12):1539-1547.
7. 7. Gargis AS, Kalman L, Bick DP, et al. Long-read sequencing technologies in clinical genomics: applications in hematologic malignancies. *J Mol Diagn.* 2022;24(3):213-226
8. 8. Patkar N, Kakirde C, Shaikh AF, et al. Clinical impact of panel-based error-corrected next-generation sequencing versus flow cytometry to detect measurable residual disease in acute myeloid leukemia. *Leukemia.* 2021;35(5):1392-1404.
9. 9. Bagal B, Tembhare PR, Nayak L et al. Minimal residual disease in newly diagnosed multiple myeloma patients not planned for upfront autologous stem cell transplant: patient disposition and MRD outcomes from a prospective clinical trial. *Clin Lymphoma Myeloma Leuk.* 2025;25(Suppl): S261. doi:10.1016/S2152-2650(25)03839-X
10. 10. Tembhare PR, Rawstron AC, Abeykoon JP, et al. Next-generation flow cytometry for measurable residual disease assessment in plasma cell myeloma: principles and clinical applications. *Front Oncol.* 2021;11:634839. doi:10.3389/fonc.2021.634839