

Protocol

Molecular and Pharmacogenetic Marker Evaluation in Relation to the Toxicity and Clinical Response of Acute Lymphoblastic Leukemia Treatment in Indian Children (MPGx-INDALL): Protocol for a Prospective Observational Cohort Study

Swetambri Sharma^{1*}, MSc; Shuvadeep Ganguly^{2*}, DM; Kamali Murugadoss², MSc; Smita Kayal², DM; Swaminathan Keerthivasagam², DM; Jaikumar Ramamoorthy³, DM; Archana Singh⁴, PhD; Deepam Pushpam¹, DM; Jayanthi Mathaiyan⁵, MD; Yvonne Gloor⁶, PhD; Frederic Baleyrier⁶, MD, PhD; Marc Ansari^{6*}, MD; Sameer Bakhshi^{1*}, MD; Biswajit Dubashi^{2*}, DM; Chakradhara Rao S Uppugunduri^{2,6*}, PhD

¹Department of Medical Oncology, All India Institute of Medical Sciences, New Delhi, India

²Department of Medical Oncology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

³Department of Pediatrics, Jawaharlal Institute of Post Graduate Medical Education and Research, , Puducherry, India

⁴Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India

⁵Department of Pharmacology, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India

⁶Department of Pediatrics, Gynecology and Obstetrics, CANSEARCH Research Platform in Pediatric Oncology and Haematology, University of Geneva, Geneva, Switzerland

*these authors contributed equally

Corresponding Author:

Chakradhara Rao S Uppugunduri, PhD
Department of Medical Oncology
Jawaharlal Institute of Postgraduate Medical Education and Research
Dhanvantri Nagar
Puducherry 605006
India
Phone: 91 7675935408
Email: uscrao@jipmer.ac.in

Abstract

Background: Understanding interindividual variability in treatment response and toxicity is essential for optimizing outcomes in pediatric acute lymphoblastic leukemia (ALL). Molecular and pharmacogenetic markers hold promise in predicting treatment efficacy and adverse effects, particularly in genetically diverse populations. This protocol outlines the methodology for a prospective, nonrandomized observational cohort designed to evaluate molecular and pharmacogenetic factors associated with treatment response and toxicity in Indian children diagnosed with ALL.

Objective: The primary objective is to identify genetic markers associated with treatment-related toxicity and therapeutic response. Secondary objectives include evaluating associations between the occurrence of early toxicities and quality of life during active ALL treatment, specific pharmacogenetic variants, and survival outcomes along with generating data to support the future implementation of personalized treatment strategies in Indian children with ALL.

Methods: In this prospective, observational cohort, 556 children (≤ 18 years of age) with newly diagnosed ALL treated under the Indian Childhood Collaborative Leukemia–Acute Lymphoblastic Leukemia 2014 (ICiCLE-ALL-14) protocol at two Indian centers will be enrolled, aiming for a minimum of 500 evaluable children. Eligible participants will be enrolled prior to the initiation of chemotherapy and followed longitudinally throughout treatment. Clinical and laboratory data (demographics, nutritional assessment, quality of life, comorbidities, treatment regimen, toxicity graded by Common Terminology Criteria for Adverse Events v5.0, remission status, and survival) will be collected at predefined intervals up to day 100 of the maintenance phase. Germline and somatic DNA will be sampled at diagnosis and remission. The first phase will use whole-exome sequencing to discover candidate variants by implementing a candidate gene prioritization strategy. The second phase will genotype the top candidates in the full cohort using array technology. Associations with early treatment-related toxicities,

steroid response, and survival will be tested by multivariable regression and Cox models. A machine learning approach with pharmacogenetic predictors as classifiers will be implemented further with cross-validation and sensitivity analyses.

Results: Ethical committees approved the protocol version 1.0 in 2020: IEC-1167/06.11.2020 (All India Institute of Medical Sciences, New Delhi), JIP/IEC/2020/201 (Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry), and AO_2021-00048 (UNIGE, Geneva). Funding was received from Swiss National Science Foundation, Switzerland; Department of Biotechnology, India; and CANSEARCH Foundation, Switzerland. Recruitment began in December 2022 and is likely to conclude by 2027. A comprehensive analysis of the complete study cohort is anticipated to be completed by 2027.

Conclusions: The MPG_x-INDALL (Molecular and Pharmacogenetic Marker Evaluation in Relation to the Toxicity and Clinical Response of Acute Lymphoblastic Leukemia Treatment in Indian Children) study will generate actionable insights for individualized ALL therapy in India via systematically evaluating germline and somatic markers in a large ethnically distinct cohort.

Trial Registration: ClinicalTrials.gov NCT05512169; <https://clinicaltrials.gov/study/NCT05512169>

International Registered Report Identifier (IRRID): DERR1-10.2196/79865

JMIR Res Protoc 2026;15:e79865; doi: [10.2196/79865](https://doi.org/10.2196/79865)

Keywords: pharmacogenetics; germline variants; somatic variants; toxicity; quality of life; treatment-related toxicity; ICiCLE; acute lymphoblastic leukemia; pediatric ALL; sequencing

Introduction

Background

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy accounting for about 25% of all cancers in children aged 2-15 years [1]. Globally, the odds of surviving ALL have improved dramatically, with 5-year overall survival (OS) reaching up to 90% in high-income countries (HICs) [2,3]. The proportion of patient deaths due to treatment-related toxicity (TRT) during treatment remains significant, ranging from 2% to 24% [4]. There is clearly a survival gap between low-income countries and HICs. Over recent years, the treatment protocols in ALL treatment have been focused on minimizing TRT and long-term adverse effects while preserving favorable survival outcomes. The protocols developed by BFM (Berlin-Frankfurt-Münster), the Children's Oncology Group, United Kingdom ALL group, and the French protocol for the treatment of ALL in children and adolescents are widely used in HICs [5].

Most of the oncology centers in India recently implemented these protocols, or modified versions of them, for the treatment of childhood ALL, especially the more recently implemented Indian Childhood Collaborative Leukemia–Acute Lymphoblastic Leukemia 2014 (ICiCLE-ALL-14) protocol [6-11]. The ICiCLE-ALL-14 protocol study on the Indian pediatric population reported a 4-year OS rate at 74%, and the event-free survival (EFS) rate was 62% [12]. Risk stratification clearly marked differences in the outcomes as well, with 4-year EFS and OS at 76% (95% CI 72%-79%) and 88% (95% CI 85%-90%) in the standard risk group, 70% (95% CI 66%-74%) and 80% (95% CI 77%-83%) in the intermediate risk group, 61% (95% CI 51%-64%) and 73% (95% CI 70%-76%) in the high risk groups, and 69% (95% CI 62%-75%) and 77% (95% CI 70%-83%) in T-ALL cases [12]. The cumulative incidence of treatment deaths and relapses at centers varied from 2% (95% CI 1%-5%) to 13% (95% CI 10%-17%; $P \leq .001$) and 21% (95% CI 17%-26%)

to 45% (95% CI 39%-51%), respectively. About 39% of the relapsed cases were early relapses, and steroid poor response indicated worse EFS in univariate analyses having 53% of EFS in poor responders [12]. These results indicate addressing the issue of improving survival rates, but when it comes to the prevention of toxicities, we have no prospectively collected data from this geographical region. Further, delayed presentation at diagnosis, treatment abandonment [13-15], treatment-associated infections, adverse effects [16, 17], higher steroid nonresponse rates (up to 18%-40%) [6,13], and higher incidence of relapse simultaneously account for the death of children with ALL undergoing treatment [14,18]. ALL is a genetically heterogeneous disease. Specific genetic variations have been identified as biomarkers to predict the toxicity and efficacy of drug therapies used to manage ALL [19], for example, mutations in the IL7R signaling components *JAK1* and *KRAS* with steroid resistance [20]. However, baseline prognostic stratification using genetic screening and risk-adapted protocols is yet to be uniformly accessible across low-resource settings and even in high-resource settings. The use of pharmacogenetic markers to guide the dosing of chemotherapeutic drugs is also nonexistent in low or low-middle income settings. Geographic and ethnic variations in childhood cancers [21] and variations in supportive care therapy further increase the need for population-specific, preemptive markers for optimized treatment management.

Indian children may harbor distinct genetic mutations [20] and risk profiles not well captured in Western studies. As Gogoi et al [12] emphasize, scaling up genetic profiling is both significant and transformative. Incorporating genetic testing into the management of childhood ALL—especially in India's diverse and resource-limited setting—might help in the reduction of treatment-related mortality. Building on these insights, we would like to use this opportunity to combine this current standardized treatment protocol and the unique ethnic population of India, which also has high rates of inbreeding practices and consanguinity in certain parts of the country [22]. This prospective observational cohort

study aims to evaluate the association of genetic variants (both germline and somatic) with TRTs, steroid response, and clinical outcomes among patients with childhood ALL being treated with the ICiCLe protocol.

Objectives

The primary objectives of the study are as follows:

1. To study the associations of static germline genetic variants with early chemotherapy-related toxicities (treatment-related toxicities) in children with ALL undergoing the ICiCLe treatment protocol.
2. To investigate the somatic and germline genetic markers associated with the efficacy and toxicity of glucocorticoid response, respectively.
3. To biobank biological samples and clinical data for future association analyses to develop biomarkers predicting the efficacy and toxicity of the treatment protocol.

The secondary objectives of the study are as follows:

1. To study the impact of the occurrence of early toxicities on quality of life (QoL) during active ALL treatment (as estimated by PedsQL Cancer module tool).
2. To evaluate the association of genetic variants (somatic and germline) with OS, nonrelapse mortality, relapse-free survival (RFS), and EFS.

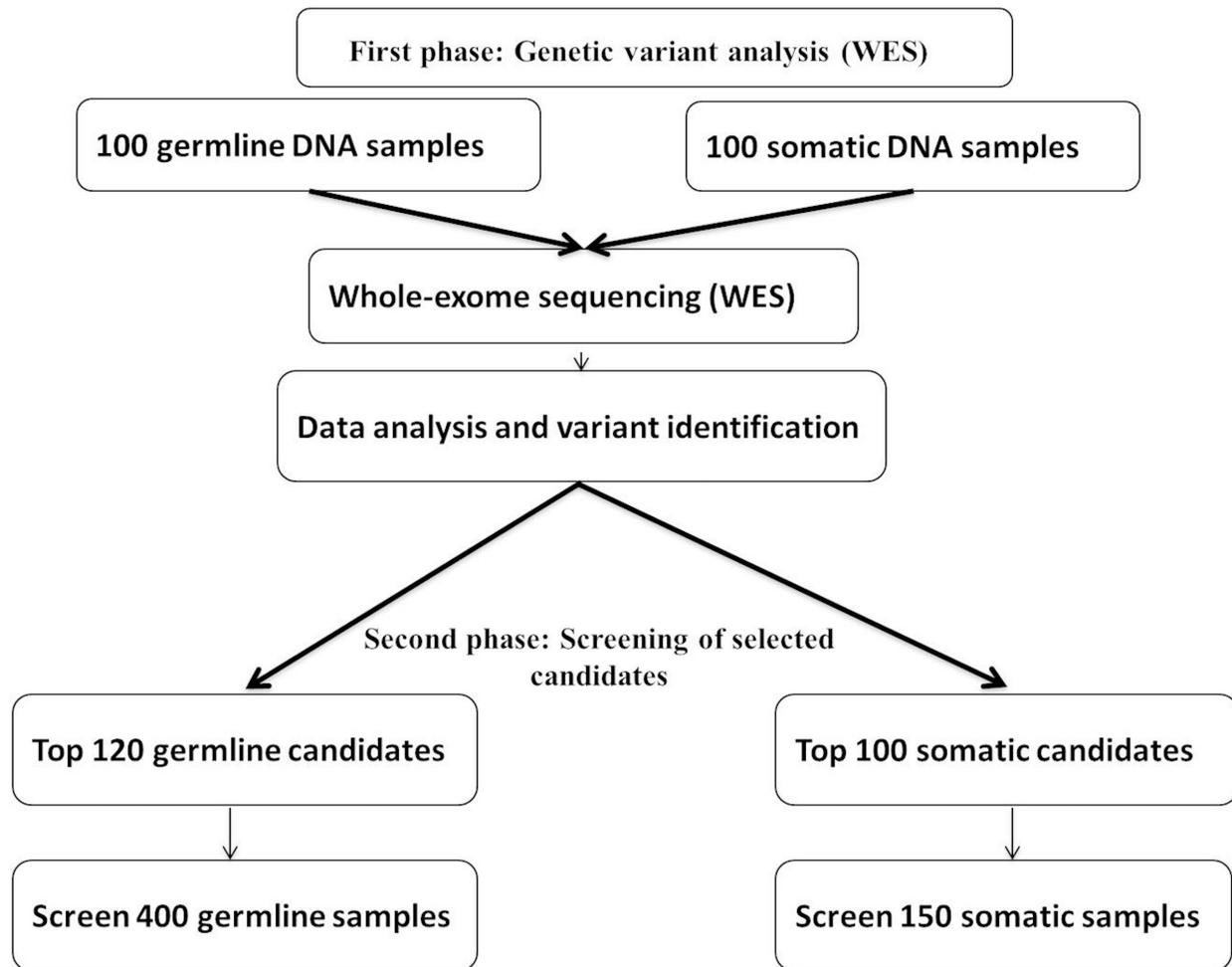
Methods

Study Design

The proposed study is a prospective observational multicentric cohort study, where patient recruitment and clinical

follow-up are being done at Regional Cancer Centre in Jawaharlal Institute of Post-Graduate Medical Education and Research, Puducherry in south India and at Dr. B.R.A. Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi in north India. The study is being carried out in 2 phases. The first phase includes the identification of germline genetic variants associated with TRTs and somatic mutations with that of steroid response using whole-exome sequencing (WES; n=100) for selecting candidate gene variants. WES combined with array technologies can shed light on many important coding region and regulatory region variants of interest. Through this exercise, we may also be able to identify any novel variants at a reasonable frequency in this ethnically distinct population. Further, selected variants from the first phase analysis will be genotyped in the rest of the cohort (n=400). Final association analyses will be performed using the entire cohort. First phase analyses will be conducted at the CANSEARCH Research platform of pediatric oncology and hematology, University of Geneva, Switzerland. Second phase analyses will be carried out using array technology at the two centers in India in collaboration with the University of Geneva, Switzerland. The study flow, including various phases, is shown in [Figure 1](#).

Figure 1. Study flow.



Patient Recruitment

Consecutive patients of pediatric ALL presenting to any of the treatment centers and being treated with the ICiCLE-ALL-14 protocol will be screened for the study.

The inclusion criteria of the study include patients (1) that are older than 1 year and younger than 18 years at enrollment; (2) that are previously untreated except for patients who received treatment up to 7 days of only steroid with documented baseline steroid dose details; (3) with an ALL diagnosis confirmed by morphology and immunophenotyping by flow cytometry; (4) with self-declared Indian origins; (5) who fulfill ICiCLE treatment protocol inclusion criteria and receive treatment as per the protocol; and (6) who provide written informed consent to participate in the study, signed by the participant, parent, or guardian. The exclusion criteria of the study include (1) previously treated patients and (2) patients with Down syndrome.

Biological Sampling

The schedule of biological sampling planned for this study is mentioned in [Multimedia Appendix 1](#). In brief, saliva or buccal swabs at the time of diagnosis or saliva or buccal swabs or blood at complete remission will be collected for germline DNA. For somatic DNA extraction, bone marrow samples will be collected at the time of diagnosis and at

relapse (if any). In children with an inadequate amount of bone marrow, a peripheral blood sample will be considered for somatic DNA extraction, provided the percentage of blasts in peripheral blood is 80% or above. All saliva or buccal swabs and white blood cell pellets obtained from blood or bone marrow samples will be stored at 4 and -80°C , respectively, until extraction. Plasma samples (4 mL whole-blood samples) will also be collected at the time of diagnosis (for drug level measurements), at remission, and at relapse (if any). Plasma separation will be performed immediately and stored at -80°C . All biological sample aliquots remaining after addressing the study objectives will be stored in biobanks at respective centers for future investigations. Plasma or whole-blood samples will be collected at predefined time points. The details of the timelines for sample collection to estimate drug levels and biobanking are given in [Multimedia Appendix 2](#).

DNA Preparation Protocol, Samples Storage, and Shipment

DNA extraction will be performed in batches using QIAamp DNA Mini Kit. DNA quality and quantity will be determined using Nanodrop 260/280 ratio and Qubit [1] (Qubit 4 Fluorometer by Thermo Fisher Scientific). Integrity will be measured using TapeStation, with genomic quality number values greater than 7.5 being used for phase 1 analyses. An

aliquot of 1 μg DNA will be shared with the Geneva team for phase 1 sequencing analyses and storage in the BaHOP (Biobank of the Pediatric Hematology and Oncology Unit, IRB number: PB_2017-00533). The remaining aliquots will be stored in their respective biobanks in properly labeled vials at -80°C temperature. DNA QC thresholds were as follows: minimum A260/280 ratio of 1.8–2.0, Qubit-measured concentration 20 $\text{ng}/\mu\text{L}$ or higher, and TapeStation genomic quality number less than 7.5 for phase 1 sequencing. At least 1 μg of germline DNA and 1 μg of somatic DNA per sample are required for shipment, with 2 or more additional aliquots stored at -80°C at the center's biobank for future investigations. DNA samples will be shipped on dry ice, with a maximum acceptable transit time of 48 hours and temperature logs maintained. For plasma sample separation, whole blood must be centrifuged within 30 minutes of collection, and separated plasma shall be aliquoted into pre-labeled cryovials and immediately stored at -80°C until shipment on dry ice. All the harmonized laboratory and sampling procedures have been finalized and are provided in the laboratory manual ([Multimedia Appendix 2](#)) to support consistent implementation across centers.

Clinical Data Collection, Patient Follow-Up, and Recording of Outcomes

The baseline sociodemographic details, including socioeconomic status by the modified Kuppaswamy socioeconomic scale and clinical and laboratory details (immunophenotype, karyotyping, and molecular tests) during disease presentation, will be collected [23]. All patients will undergo risk stratification and receive treatment as per ICiCLE-ALL-14 protocol and use uniform case record forms implemented in REDCap (Research Electronic Data Capture). The data will be anonymized by each center before the sharing of the clinical data and genetic data with other parties and in the public domain.

Patients will be followed up at each outpatient visit and any unscheduled visit due to toxicity. At each visit, all the available details of drug dose, modifications, or omissions due to toxicities will be recorded. This will aid in assessing the intensities of the treatment in relation to the incidence of toxicities and other clinical outcomes. All drug-related toxicities as per common terminology criteria for adverse events (version 5.0) [24] occurring from the first day of induction until day 100 of maintenance therapy will be recorded. For objective 1 (primary end point analyses), a predefined set of clinically significant early toxicities will be evaluated, including grade 3 or higher hematological, hepatotoxicity, pancreatitis, neurotoxicity, and serious infections. These events are explicitly mapped to the case record forms to ensure harmonized data capture across study sites. Post day 100 of maintenance, the patient will continue to be monitored for long-term outcomes including relapses (if any), need for hematopoietic stem cell transplantation, and mortality.

The data on steroid response at the end of 7 days (1 week of steroid prophylaxis), bone marrow morphological remission status, and minimal residual disease using flow cytometry

status at the end of induction or consolidation (if applicable) will be systematically recorded and analyzed in relation to somatic genetic variants.

Other clinical outcomes to be collected are nonrelapse mortality (time from study enrollment to death due to cause other than relapse of disease), incidence of relapse of disease (ie, the duration between the day of complete remission and the day of occurrence of relapse), RFS (time from study enrollment to relapse for patients attaining complete remission), OS (the duration from enrollment in the study to death due to any cause), and EFS (from treatment initiation to the first induction failure, nonresponse, or the progression of the disease, death from any cause).

Additionally, QoL will be assessed by the interview of children and their parents or caregivers using the age-specific PedsQL Cancer Module v3.0 questionnaire [25]. QoL will be assessed longitudinally initially at baseline, then at end of induction, consolidation, interim maintenance, and at day 100 of maintenance phases of the ICiCLE-ALL-14 protocol. We have obtained a license to use the questionnaires (work order 312005), and translated versions are validated as per the guidelines of the Mapi research trust and are made available at the Mapi research trust. Nutritional status will be evaluated by anthropometric assessment at baseline and serially along with the recording of the implementation of any nutritional recommendations. The case record form template used for clinical data collection and XML files of the REDCap database template are available as [Multimedia Appendices 3](#) and [4](#), respectively.

Schedule of Patient Participation in the Study

The patient follow-up data will be collected for the study objectives from the day of induction treatment until day 100 in the maintenance phase of the ICiCLE treatment protocol for ALL. Each patient will be scheduled for data collection and laboratory tests as a part of routine clinical care for 37 to 42 weeks from the day of the initiation of the treatment to evaluate the primary objectives and one of the secondary objectives related to QoL. However, the survival outcomes would require having follow-up data up to 1 year post treatment. The details of the planned procedures for patients and timelines are shown in [Multimedia Appendix 2](#).

Detection and Analysis of Genetic Variants

First Phase Genetic Variant Analysis

Whole-Exome Sequencing

WES will be performed in 100 germline DNA samples and 100 DNA samples from leukemic cells (same patients). WES will use the TWIST Biosciences TWIST Comprehensive Exome workflow according to manufacturers' protocols: (1) library preparation; (2) sequencing using an Illumina NovaSeq6000 system with a mean coverage of 70 \times for germline DNA samples and 150 \times for somatic DNA samples; and (3) raw data integration, storage, and analyses. The

sequencing will be done at the Health 2030 Genome Center, Campus Biotech, Geneva, Switzerland, and the sequences will be stored on a secured server of the University of Geneva until analyses and at the corresponding centers until the end of the study and archival.

Genetic Data Analysis—First Phase

For genetic data analysis, whole exomes and flanking regions ($n=100$, paired somatic and germline DNA) captured will be mapped to the GRCh38 reference and the variants called using the DRAGEN v.4.0.3 analysis platform. Plink 1.9 will be used to perform the quality control of the resulting sequences including call rate, Hardy-Weinberg equilibrium, sex mismatch, relatedness, heterogeneity, and ethnicity. Quality control investigations on cryptic relatedness and concordance between the reported and genetic sex will be included for patient selection. Principal component analyses will be an integral part of the analyses due to the possibility of unique clusters. Thus, the population structure is also accounted for while performing association analyses. When it comes to pharmacogenetic association (objective 1), frequencies might be impacted by the population structure, but for somatic genetic association, we expect no impact of population structures.

We will perform association analysis via (1) a candidate gene approach with filtered variants or mutations to the selected genes using candidate prioritization strategies from a systematic review (PROSPERO CRD420251112557 and CRD42021229748) and candidates associated with phenotypic measures used for toxicity definitions or steroid sensitivity from public datasets, for example, UK Biobank (application 91515), and (2) a hypothesis-free, exome-wide, and association analysis. The predicted effects of missense variants on protein function will be assessed in silico using SIFT and PolyPhen2 incorporated into variant effect predictor tools [26]. Variant filtering will be performed based on the 1000 Genomes, UK Biobank, IndiGen variation [22], and the NHLBI GO Exome Sequencing projects [27]. Fisher exact test (allelic association) and the Cochran-Armitage test for trends will be implemented in PLINK [28] to search for associations between clinical outcomes and genetic variants. Association analysis for quantitative and binary data will be analyzed using general linearized models in PLINK, with a P value significance threshold of .05 after adjustments for multiple testing that will be performed using the Benjamini-Hochberg false discovery rate method [29]. Candidate prioritization strategies for phase 1 analyses will be communicated in a separate report. Phase 1 analyses will thus enable the creation of a leukemia sequencing database of a single ethnicity, along with its clinical and follow-up data, all obtained from a single study. This will serve as a valuable resource for researchers investigating questions related to ALL treatment in children.

Second Phase Genotyping for Selected Candidates From Phase 1 Analysis

Somatic short variants will be called using GATK Mutect2 with matched tumor (bone marrow) and normal (buccal

swab) samples. To control for germline mutations, we will provide a panel of normals generated from the buccal swab samples (GATK CreateSomaticPanelOfNormals) and use the gnomAD af-only resource (af-only-gnomad.vcf.gz; reference set of germline mutations) as our germline resource. Cross-sample contamination will be estimated (GATK GetPileupSummaries+CalculateContamination) and used during filtering. The resulting calls will be filtered and annotated for downstream analyses (GATK FilterMutectCalls) “purity thresholds.” Tumor purity and copy number will be inferred using PureCN, as well as computing cancer cell fraction for each somatic variant. Samples with estimated tumor purity p less than 20% will be excluded from primary discovery analyses (but retained in QC tables). This threshold may be refined after inspecting the purity distribution. The rationale behind choosing this threshold is as follows: for a heterozygous clonal mutation ($c=1$) in a diploid region, the expected variant allele frequency (VAF) of a clonal mutation would be $p/2$ (eg, when $p=1.0$, $\text{VAF}=0.50$, when $p=0.20$, $\text{VAF}=0.10$). A mutation present in only 10% of the tumor cells (subclonal, cancer cell fraction=0.1), and tumor purity $p=0.20$, would correspond to a VAF of 0.01, which corresponds to 1.5 alternate reads at 150 \times . This is below a reliable detection level that can be confounded with sequencing noise. Additionally, to reduce false positives from sequencing noise, we will require a minimum of 5 alternate reads for somatic calls to pass filters for downstream analyses.

Phase 2 analyses will include screening 400 germline DNA samples for approximate top candidates identified in phase 1’s germline sequencing association analysis with TRT. Somatic DNA analysis includes the screening of 150 samples from the top 100 candidates from phase 1 analysis. The final TRT association analyses will be conducted for 500 patients for the top candidates and for steroid response (steroid or prednisone response rate on day 8; good responders: peripheral blood blast count $<1000/\mu\text{L}$; poor responders: peripheral blast count $\geq 1000/\mu\text{L}$) [7] in 250 patients before meta-analysis with phase 1 participants. Phase 2 genotyping will be realized by genotyping arrays. We further plan to develop a project for cost-effective screening for implementation in a clinic, for example, custom open array.

Withdrawal and Discontinuation

The expected dropout rate is low for the entire period of the study. However, it is expected that approximately 5% of the recruited patients may discontinue or drop the study, and this will be managed by recruiting new patients into the study, if the data are not available for use to evaluate primary objectives.

Statistics and Analyses Methodology

SPSS (version 25; IBM Corp) or R statistical software (packages: survival, cmprsk, glmnet, and mbmdr [30]) will be used for statistical analysis. The appropriate genetic models (dominant, recessive, or additive) for specific genotypes or haplotypes will subsequently be derived based on the results of those analyses. Nonparametric or chi-square tests will be used to test differences in sex, age, and other

categorical variables between genotype groups. The allele genotype frequencies, Hardy-Weinberg equilibriums, and haplotypes will be analyzed using Plink v1.9, the Haploview software (Broad Institute), and PHASE [31]. The frequency of all incidents and toxic events will be compared between genotype or haplotype or demographic- and disease characteristic-based groups using chi-square tests. The cumulative incidences of clinical outcomes, such as TRTs, will be estimated in relation to genotype or haplotype groups using cumulative incidence curves (cmprsk package in R). Survival curves for OS, nonrelapse mortality, RFS, and EFS will be calculated using Kaplan-Meier estimations, and log-rank tests will be used to compare the differences between genotype groups in univariate analysis. Univariate Cox regression analysis will be used to estimate hazard ratios with 95% CI; multivariate Cox regression analysis will then be used to estimate the impact of genotypes or haplotypes on clinical outcomes in the presence of other covariates. The influence of multiple genes on clinical parameters will be assessed in gene-gene interaction epistasis models incorporating various genetic and clinical confounders using multivariate regression analysis and multifactor-dimensionality reduction analysis [32]. All primary genetic association models will be adjusted for a predefined set of covariates but not limited to age, sex, immunophenotype, baseline white blood cell count, measurable residual disease, risk category, nutritional status category, socioeconomic status (modified Kuppuswamy), and treatment center.

We will use appropriate statistical tests for less frequent outcomes, for example, Least Absolute Shrinkage and Selection Operator-penalized regression [33], thus avoiding the overfitting of the data and biasing results due to the collinearity of the factors included. Confounders, such as socioeconomic and nutritional status, will also be included in the analyses. Reduced QoL due to the occurrence of toxicity will also be evaluated using logistic regression (after defining the impaired scores for both physical and emotional components). PedsQL scores are collected longitudinally across treatment phases; hence, we will attempt mixed-effects modeling to appropriately account for within-patient correlation and temporal trends. The objective of the final joint analyses is to determine the sensitivity and specificity of gene variants combined with other patient-specific characteristics to predict the clinical outcomes. This will be analyzed by decision curve analysis using all the categorical variables (some continuous variables would be converted into categorical variables using the X-tile program from Rimm lab, Yale School of Medicine), including genotypes along with reclassification improvement and integrated discrimination improvement in R using the “rms” package. The concordance index would be used to test a new model in comparison to the previous model. The performance of this strategy would be compared with that of receiver operating characteristic curve analysis. We will be applying suitable models only after the exploration of the data based on the collinearity matrix of the variables. However, we hereby propose possibilities for the analyses a priori. Appropriate statistical tools will be implemented upon

discussions with peers to derive meaningful conclusions from the data collected in secondary analyses.

Additionally, for multiplicity control, the 2-stage discovery-validation design incorporates explicit control for multiple comparisons. In phase 1 (WES discovery), variant associations will undergo Benjamini-Hochberg false discovery rate (FDR) control (target FDR=0.05) with the reporting of both q -values and unadjusted p -values. Only the top ≈ 120 predefined candidates meeting prespecified criteria (biological relevance, effect size, and FDR ranking) will advance to phase 2 validation. In phase 2, confirmation analyses will apply a Bonferroni-adjusted significance threshold ($\alpha = .05/\text{number of candidate variants}$). Joint (meta-analytic) models combining phase 1 and phase 2 will apply the same validation-level α threshold. This strategy balances discovery sensitivity with stringent clinical-grade validation.

Power Considerations for Primary Objectives

For germline pharmacogenetic variant associations, the sample size ($n=500$) was calculated using a power calculator for genome-wide association studies using a 2-stage design [34]. In this 2-stage design, 20% of the patients are genotyped for all the candidate markers, and then the remaining 80% are genotyped for the top 120 candidates (0.02% of markers analyzed in phase 1). This design has sufficient power for the phase 1 analyses: replicating and joint analyses. Criteria used for estimation are 5% to 15% minor allele frequency; dominant mode of inheritance; α value of .00001 (0.05/5000; 2-sided); power ($1 - \beta$) of 80%; early toxicities prevalence of 40%; and relative risk of 3.5 to estimate for the gene variant or variants. We used a stringent α value of .00001 (instead of .0004), that is, adequate for joint analysis (.05/120 = .0004) as required when developing models for clinical use. We acknowledge that this approach may lack smaller but clinically important associations in resource-limited settings. This design reduces the total amount of genotyping required by 80% (and thus associated costs), and there is a 90.1% probability that the associated markers will be followed in stage 2 analysis. Considering 5% of patients with loss to follow-up and an additional 5% of samples being not analyzable due to laboratory issues, a total of 556 patients need to be recruited to get 500 evaluable patients for the primary objective 1. For secondary outcomes (OS, EFS, QoL, and steroid response), the power considerations are detailed in [Multimedia Appendix 5](#).

Handling of Missing Data

This study expects to have minimal missing data, as a dry run for collecting data will be implemented, and documented issues will be fixed before initiating the recruitment of the patients. Moreover, the study includes participants who are under routine clinical care settings (in-patient settings and out-patient settings) for about 2.5 years, and our primary objective's data collection is limited to a maximum of 40 weeks; hence, we foresee dropouts (5%) from the study. Most importantly, the social worker coordinating the ALL

treatment at both centers will ensure that patients receive this standard treatment. For secondary survival outcomes, we will collect the data retrospectively through forms that are currently filled out by the participating centers. The few dropouts that may occur shall be replaced by the recruitment of new participants. Patients whose germline samples could not be collected or analyzed will be excluded from the study and will not be analyzed. Furthermore, patients requesting the withdrawal of their informed consent will not be analyzed by the end of the study.

Ethical Considerations

The ethical approval to conduct the study was obtained from the Institute Ethics Committee, All India Institute of Medical Sciences, New Delhi, India (IEC-1167/06.11.2020); Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry (JIP/IEC/2020/201); and a positive stance from the ethics committee of Canton of Geneva (AO_2021-00048). All study procedures will be conducted in accordance with the principles of the Helsinki Declaration and in accordance with national and local regulations. All consecutive eligible patients will be recruited in the study after informed consent from a legal guardian and patient assent (if applicable). The participant or their legal representatives may withdraw consent to participate in the study at any time without justification, and without affecting the quality of routine standard medical care. For any withdrawal of consent, the data collected prior to that point will be used unless the participant or his or her legal representative does not wish to do so in writing.

Results

As this is a protocol, no clinical or genetic results are presented here. This section summarizes the current study status. Ethical approval has been obtained from all participating centers, and recruitment began in December 2022. A total of 346 participants were recruited for the study, and recruitment is ongoing. Phase I exome sequencing for both germline and somatic DNA has been completed. As per the most recent internal review, recruitment is projected to conclude by 2027. Data cleaning, harmonization, and preliminary genetic analyses for phase I are planned to begin in early 2026, with full cohort analyses expected in 2027.

Discussion

Principal Findings

To date, there have been no comprehensive studies for the treatment of childhood ALL across India's unique ethnic population in a standardized treatment setting with predefined methodology. We would like to use this opportunity to initiate this comprehensive project to serve as a benchmark pharmacogenetic study and to provide valuable resources (biological material and clinical data) for future investigations aimed at personalizing ALL treatment in Indian children. The novelty of the study lies in the fact that the 2-stage design with candidate prioritization strategies is implemented

in a prospective cohort that exclusively covers the ICiCLE protocol that is uniformly used to treat patients with pediatric ALL across the Indian population. Moreover, harmonized treatment protocols in India have increasingly demonstrated benefits to survival progressively. This study design will enable us to identify both common and specific variants as markers of TRTs and clinical outcomes of ALL treatment in Indian children.

The study's results will be useful in the future to apply therapies to children and adolescents from India and elsewhere being treated for ALL. The concept of avoiding lost-to-follow-up patients by preventing the development of toxicities through personalized dosing is innovative. Thus, identifying such patients earlier using genetic markers would enable the physician to plan a proper management of their therapies well in advance, which is advantageous in limited-resource settings. Glucocorticoids are the principal components of pediatric ALL therapy protocols, including ICiCLE. It has been established through previous studies that there is interindividual variability in the treatment response to glucocorticoids as well as in susceptibilities to their toxicity [19,20]. Steroid nonresponders are usually at a higher risk of treatment failure, but response assessment is done after 8 days of treatment in the current treatment regimen. Therefore, with the early identification of responders, this study will help in reducing deaths related to steroid nonresponders. As the treatment protocol involves several chemotherapeutic drugs, there is an inadvertent need to identify genetic markers involved in drug metabolism, transport, and pharmacodynamics. So far, studies have been focused on selecting single genetic markers, but there has been emerging evidence suggesting associations between single-nucleotide polymorphisms and the development of toxicity or resistance or relapse to ALL therapies. Various single-nucleotide polymorphisms (target genes or metabolic enzymes) have been identified that are associated with increased susceptibility to drugs used in treatments, such as thiopurine methyltransferase, nudix hydrolase 15 for 6-mercaptopurine, methylenetetrahydrofolate reductase for methotrexate, and asparagine synthetase for L-asparaginase. This study will assess the relevance of these findings in addition to shedding light on novel variants, leading to a more comprehensive genetic profile for each patient that may aid in developing precision medicine strategies tailored specifically for the Indian population.

Therefore, it is our expectation that this study will provide direction on the utility of using genetic biomarkers for patient stratification. If the proposed genetic stratification criteria could be clinically justified, it would be possible to involve preemptive genetic testing in routine clinical practice. As the study spans 2 centers and 2 laboratory environments, we prospectively established harmonized standard operating procedures and predetermined covariate sets. Batch-effect control measures, including technical replicates and platform-specific QC pipelines, have been incorporated to minimize systematic bias. The study will include whole-exome discovery with downstream genotyping; integration of clinical, QoL, and biological data; and biobanking for future

work. This study design is not devoid of any limitations; for example, the design includes the evaluation of the most frequent variations and lacks power to quantify the effect of rare variants. The 2-phase design of the study optimizes the costs of genotyping but compromises study power to quantify the effect of low-frequency variations that may have impact together in a gene-gene interaction model. Moreover, the sample size of the study is calculated considering objective 1, that is, germline genetic predisposition. Within this study, we do not plan to collect late toxicities due to follow-up limited to day 100 of maintenance for primary objectives and reduced power for detecting modest genetic effects. However, patients will be followed up to obtain the retrospective data collection on toxicities and other clinical outcomes to assess long-term clinical outcomes in relation to germline and somatic genetic variants. Another major limitation includes the inclusion of patients from 2 geographical locations ie, North and South India, that are distinct in terms of ethnic origins. However, there is no clear distinction that was observed in terms of pharmacogene variant distribution between these 2 distinct groups [14]. This limitation, indeed, is an advantage to identify the variants that may have clinical utility for both populations, so that countrywide policies can be made efficient for translating and implementing the findings

in clinics. Another limitation is the long recruitment time that may have an impact on the institute's policies and practices in treating the disease using the ICiCLE protocol. This can be resolved by considering treatment year as a confounder in the association analyses. This protocol study is reported in accordance with the SPIROS (Standardized Protocol Items Recommendations for Observational Studies) and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational studies and the TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence) guideline for studies involving the development of machine learning-based prediction models.

Conclusion

To conclude, this study will prospectively evaluate the association of germline genetic variants with TRT and clinical outcomes in patients with pediatric ALL of Indian origin. The study will also explore somatic variant associations with clinical outcomes in the background of socioeconomic and nutritional status. In addition, the impact of toxicity occurrence on QoL will also be assessed as well, along with storing biological material and clinical data for future studies.

Acknowledgments

We acknowledge the support received from the clinical research center, University Hospitals of Geneva, especially for building the REDCap database. We also acknowledge Resonance INC (Dr Scott Howard) for monitoring the study during the initial phase. We thank Dr. Aurore Britan-Wood for critical feedback and support on the database and the management of the study. We thank Mme Zoe Giorgetta for her administrative support. Assistance by Mme Valerie Steck admin and Mme Thanh Mai is also acknowledged. We thank Florence Schmidt from the CANSEARCH foundation for their unending support. We thank the entire research team members (past and present) of Jawaharlal Institute of Postgraduate Medical Education and Research (Dr Charanraj Goud, Dr Yadav Nisha, Dr Govindan Krishnamoorthy, and Ms Sindhuja Jayamurthi) and team members from AIIMS (Ms Farhana Jahana, Ms Himanshi Sharma, Ms Mamta, Ms Muskan David, and Mr Chakshu Sharma) for their support during various stages of the study. We thank the NOCI India team (Dr Prasanth Ganesan) for their support in monitoring the study. We thank Dr Youssef Daali, Dr Caroline Samer, Prof Jules Desmeules, and the late Prof Hans Peter Wagner for supporting the idea of a pharmacogenomic study in acute lymphoblastic leukemia. Continued mentoring support from Prof Maja Krajinovic is highly appreciated.

Funding

The study is supported by the Swiss National Science Foundation (Switzerland; grant IZLIZ3_200243), the Department of Biotechnology (India; grant IC-12044(12)/3/2022-ICD-DBT), and the CANSEARCH Foundation (Switzerland).

Authors' Contributions

MA, SB, BD, and CRSU conceptualized the study. CRSU developed the methodological framework. SS, SG, MA, AS, SB, BD, and CRSU contributed to the design and optimization of experimental protocols. Original manuscript, tables, and figures were prepared by SS, SG, and CRSU. Reviewing and editing the manuscript were performed by SS, SG, KM, SB, BD, YG, and CRSU. Project supervision, recruitment, and administration are overseen by SG, S Kayal, S Keerthivasagam, AS, JM, JR, YG, FB, MA, SB, DP, BD, and CRSU. Funding was secured by SB, BD, and CRSU. All authors have read and approved the final manuscript. SG and SS contributed equally and held joint first authorship. CRSU, SB, BD, and MA held joint last authorship.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of biological sampling.

[\[DOCX File \(Microsoft Word File\), 19 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Laboratory manual.

[[PDF File \(Adobe File\)](#), 2069 KB-[Multimedia Appendix 2](#)]

Multimedia Appendix 3

MPX-IND ALL Case Record Form Version 1.3.

[[PDF File \(Adobe File\)](#), 831 KB-[Multimedia Appendix 3](#)]

Multimedia Appendix 4

Electronic case record form template (REDCap).

[[RAR File \(RAR archive File\)](#), 71 KB-[Multimedia Appendix 4](#)]

Multimedia Appendix 5

Objectives 2 and 3.

[[DOCX File \(Microsoft Word File\)](#), 15 KB-[Multimedia Appendix 5](#)]

References

1. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(2):83-103. [doi: [10.3322/caac.21219](#)] [Medline: [24488779](#)]
2. Hunger SP, Lu X, Devidas M, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group. *J Clin Oncol*. May 10, 2012;30(14):1663-1669. [doi: [10.1200/JCO.2011.37.8018](#)] [Medline: [22412151](#)]
3. Gaynon PS. Childhood acute lymphoblastic leukaemia and relapse. *Br J Haematol*. Dec 2005;131(5):579-587. [doi: [10.1111/j.1365-2141.2005.05773.x](#)] [Medline: [16351633](#)]
4. Dharia NV, Kugener G, Guenther LM, et al. A first-generation pediatric cancer dependency map. *Nat Genet*. Apr 2021;53(4):529-538. [doi: [10.1038/s41588-021-00819-w](#)] [Medline: [33753930](#)]
5. Reismüller B, Steiner M, Pichler H, et al. High hyperdiploid acute lymphoblastic leukemia (ALL)—a 25-year population-based survey of the Austrian ALL-BFM (Berlin-Frankfurt-Münster) Study Group. *Pediatr Blood Cancer*. Jun 2017;64(6):e26327. [doi: [10.1002/pbc.26327](#)] [Medline: [27804199](#)]
6. Roy P, Narula G, Arora B, et al. Implementation of risk adapted therapeutic strategy for childhood acute lymphoblastic leukaemia—interim report of the pilot InPOG-ALL-15-01 study. *Pediatric Hematology Oncology Journal*. 2018;3(3):S19. [doi: [10.1016/j.phoj.2018.11.052](#)]
7. Das N, Banavali S, Bakhshi S, et al. Protocol for ICiCle-ALL-14 (InPOG-ALL-15-01): a prospective, risk stratified, randomised, multicentre, open label, controlled therapeutic trial for newly diagnosed childhood acute lymphoblastic leukaemia in India. *Trials*. Jan 31, 2022;23(1):102. [doi: [10.1186/s13063-022-06033-1](#)] [Medline: [35101099](#)]
8. Pui CH, Yang JJ, Hunger SP, et al. Childhood acute lymphoblastic leukemia: progress through collaboration. *J Clin Oncol*. Sep 20, 2015;33(27):2938-2948. [doi: [10.1200/JCO.2014.59.1636](#)]
9. Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. *The Lancet*. Mar 2008;371(9617):1030-1043. [doi: [10.1016/S0140-6736\(08\)60457-2](#)]
10. Karrman K, Forestier E, Heyman M, et al. Clinical and cytogenetic features of a population-based consecutive series of 285 pediatric T-cell acute lymphoblastic leukemias: rare T-cell receptor gene rearrangements are associated with poor outcome. *Genes Chromosomes Cancer*. Sep 2009;48(9):795-805. [doi: [10.1002/gcc.20684](#)] [Medline: [19530250](#)]
11. Gupta SK, Singh M, Chandrashekar PH, et al. Clinical and prognostic impact of copy number alterations and associated risk profiles in a cohort of pediatric B-cell precursor acute lymphoblastic leukemia cases treated under ICiCle protocol. *Hemasphere*. Oct 2022;6(10):e782. [doi: [10.1097/HS9.0000000000000782](#)] [Medline: [36204689](#)]
12. Gogoi MP, Das P, Das N, et al. Risk stratified treatment for childhood acute lymphoblastic leukaemia: a multicentre observational study from India. *Lancet Reg Health Southeast Asia*. Jun 2025;37:100593. [doi: [10.1016/j.lansea.2025.100593](#)] [Medline: [40487914](#)]
13. Pandey A, Ahlawat S, Singh A, Singh S, Murari K, Aryan R. Outcomes and impact of minimal residual disease (MRD) in pediatric, adolescent and young adults (AYA) with acute lymphoblastic leukemia treated with modified MCP 841 protocol. *Cancer Res Stat Treat*. 2020;3(2):183-191. [doi: [10.4103/CRST.CRST_85_20](#)]
14. Kodidela S, Dorababu P, Thakkar DN, et al. Association of NUDT15*3 and FPGS 2572C>T variants with the risk of early hematologic toxicity during 6-MP and low-dose methotrexate-based maintenance therapy in Indian patients with acute lymphoblastic leukemia. *Genes (Basel)*. May 28, 2020;11(6):594. [doi: [10.3390/genes11060594](#)] [Medline: [32481505](#)]
15. Ganesan P, Jain H, Bagal B, et al. Outcomes in adolescent and young adult (AYA) acute lymphoblastic leukemia (ALL): a report from the Indian acute leukemia research database (INWARD) of the hematology cancer consortium (HCC). *Blood*. Nov 13, 2019;134(Supplement_1):1306. [doi: [10.1182/blood-2019-127091](#)]

16. Faruqui N, Joshi R, Lowe J, et al. Exploring barriers to accessing care for childhood cancers In India: a qualitative study (Inpog-Acc-16-03). *Pediatr Hematol Oncol J*. 2018;3(3):S1-S2. [doi: [10.1016/j.phoj.2018.11.004](https://doi.org/10.1016/j.phoj.2018.11.004)]
17. Ahuja S, Sharma J, Gupta S, et al. Patient tracking during treatment of children with cancer in India—an exploratory study. *Cancer Rep*. Jun 2022;5(6):e1359. URL: <https://onlinelibrary.wiley.com/toc/25738348/5/6> [doi: [10.1002/cnr2.1359](https://doi.org/10.1002/cnr2.1359)]
18. Arya LS, Kotikanyadanam SP, Bhargava M, et al. Pattern of relapse in childhood ALL: challenges and lessons from a uniform treatment protocol. *J Pediatr Hematol Oncol*. Jul 2010;32(5):370-375. [doi: [10.1097/MPH.0b013e3181d7ae0d](https://doi.org/10.1097/MPH.0b013e3181d7ae0d)] [Medline: [20463606](https://pubmed.ncbi.nlm.nih.gov/20463606/)]
19. Khera S, Trehan A, Bhatia P, Singh M, Bansal D, Varma N. Prevalence of TPMT, ITPA and NUDT 15 genetic polymorphisms and their relation to 6MP toxicity in north Indian children with acute lymphoblastic leukemia. *Cancer Chemother Pharmacol*. Feb 2019;83(2):341-348. [doi: [10.1007/s00280-018-3732-3](https://doi.org/10.1007/s00280-018-3732-3)] [Medline: [30474703](https://pubmed.ncbi.nlm.nih.gov/30474703/)]
20. Li Y, Buijs-Gladdines JGCAM, Canté-Barrett K, et al. IL-7 receptor mutations and steroid resistance in pediatric T cell acute lymphoblastic leukemia: a genome sequencing study. *PLoS Med*. Dec 2016;13(12):e1002200. [doi: [10.1371/journal.pmed.1002200](https://doi.org/10.1371/journal.pmed.1002200)] [Medline: [27997540](https://pubmed.ncbi.nlm.nih.gov/27997540/)]
21. Stiller CA, Parkin DM. Geographic and ethnic variations in the incidence of childhood cancer. *Br Med Bull*. Oct 1996;52(4):682-703. [doi: [10.1093/oxfordjournals.bmb.a011577](https://doi.org/10.1093/oxfordjournals.bmb.a011577)] [Medline: [9039726](https://pubmed.ncbi.nlm.nih.gov/9039726/)]
22. Sahana S, Bhojar RC, Sivadas A, et al. Pharmacogenomic landscape of Indian population using whole genomes. *Clin Transl Sci*. Apr 2022;15(4):866-877. [doi: [10.1111/cts.13153](https://doi.org/10.1111/cts.13153)] [Medline: [35338580](https://pubmed.ncbi.nlm.nih.gov/35338580/)]
23. Saleem SM, Jan SS. Modified Kuppuswamy socioeconomic scale updated for the year 2021. *Indian J Forensic Community Med*. Apr 28, 2021;8(1):1-3. [doi: [10.18231/j.ijfcm.2021.001](https://doi.org/10.18231/j.ijfcm.2021.001)]
24. Common Terminology Criteria for Adverse Events (CTCAE): Version 5.0. US Department of Health and Human Services; Nov 27, 2017. URL: <https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events/ctcae-v5-5x7.pdf> [Accessed 2026-02-04]
25. Varni JW, Burwinkle TM, Katz ER, Meeske K, Dickinson P. The PedsQL in pediatric cancer: reliability and validity of the pediatric quality of life inventory generic core scales, multidimensional fatigue scale, and cancer module. *Cancer*. Apr 1, 2002;94(7):2090-2106. [doi: [10.1002/cncr.10428](https://doi.org/10.1002/cncr.10428)] [Medline: [11932914](https://pubmed.ncbi.nlm.nih.gov/11932914/)]
26. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. *Nature*. Sep 24, 2009;461(7263):489-494. [doi: [10.1038/nature08365](https://doi.org/10.1038/nature08365)] [Medline: [19779445](https://pubmed.ncbi.nlm.nih.gov/19779445/)]
27. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature New Biol*. Oct 28, 2010;467(7319):1061-1073. [doi: [10.1038/nature09534](https://doi.org/10.1038/nature09534)]
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. [doi: [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8)] [Medline: [25722852](https://pubmed.ncbi.nlm.nih.gov/25722852/)]
29. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J R Stat Soc B*. Feb 1, 2004;66(1):187-205. [doi: [10.1111/j.1467-9868.2004.00439.x](https://doi.org/10.1111/j.1467-9868.2004.00439.x)]
30. R Project. URL: <https://www.r-project.org/> [Accessed 2026-02-09]
31. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. Apr 2001;68(4):978-989. [doi: [10.1086/319501](https://doi.org/10.1086/319501)] [Medline: [11254454](https://pubmed.ncbi.nlm.nih.gov/11254454/)]
32. Hsieh AR, Hsiao CL, Chang SW, Wang HM, Fann CSJ. On the use of multifactor dimensionality reduction (MDR) and classification and regression tree (CART) to identify haplotype-haplotype interactions in genetic studies. *Genomics*. Feb 2011;97(2):77-85. [doi: [10.1016/j.ygeno.2010.11.003](https://doi.org/10.1016/j.ygeno.2010.11.003)] [Medline: [21111805](https://pubmed.ncbi.nlm.nih.gov/21111805/)]
33. Li Z, Sillanpää MJ. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor Appl Genet*. Aug 2012;125(3):419-435. [doi: [10.1007/s00122-012-1892-9](https://doi.org/10.1007/s00122-012-1892-9)] [Medline: [22622521](https://pubmed.ncbi.nlm.nih.gov/22622521/)]
34. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. Feb 2006;38(2):209-213. [doi: [10.1038/ng1706](https://doi.org/10.1038/ng1706)] [Medline: [16415888](https://pubmed.ncbi.nlm.nih.gov/16415888/)]

Abbreviations

ALL: acute lymphoblastic leukemia

EFS: event-free survival

FDR: false discovery rate

HIC: high-income country

ICiLe-ALL-14: Indian Childhood Collaborative Leukemia—Acute Lymphoblastic Leukemia 2014

MPGx-INDALL: Molecular and Pharmacogenetic Marker Evaluation in Relation to the Toxicity and Clinical Response of Acute Lymphoblastic Leukemia Treatment in Indian Children

OS: overall survival

PIROS: Standardized Protocol Items Recommendations for Observational Studies

QoL: quality of life

REDCap: Research Electronic Data Capture

RFS: relapse-free survival

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence

TRT: treatment-related toxicity

VAF: variant allele frequency

WES: whole-exome sequencing

Edited by Javad Sarvestan; peer-reviewed by Mehul Jani, Xun Ding; submitted 23.Aug.2025; accepted 24.Dec.2025; published 17.Mar.2026

Please cite as:

Sharma S, Ganguly S, Murugadoss K, Kayal S, Keerthivasagam S, Ramamoorthy J, Singh A, Pushpam D, Mathaiyan J, Gloor Y, Baleyrier F, Ansari M, Bakhshi S, Dubashi B, Uppugunduri CRS

Molecular and Pharmacogenetic Marker Evaluation in Relation to the Toxicity and Clinical Response of Acute Lymphoblastic Leukemia Treatment in Indian Children (MPGx-INDALL): Protocol for a Prospective Observational Cohort Study

JMIR Res Protoc 2026;15:e79865

URL: <https://www.researchprotocols.org/2026/1/e79865>

doi: [10.2196/79865](https://doi.org/10.2196/79865)

© Swetambri Sharma, Shuvadeep Ganguly, Kamali Murugadoss, Smita Kayal, Swaminathan Keerthivasagam, Jaikumar Ramamoorthy, Archana Singh, Deepam Pushpam, Jayanthi Mathaiyan, Yvonne Gloor, Frederic Baleyrier, Marc Ansari, Sameer Bakhshi, Biswajit Dubashi, Chakradhara Rao S Uppugunduri. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 17.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.