# nature portfolio

Corresponding author(s):	Sigurdur Yngvi Kristinsson
Last updated by author(s	): Dec 2, 2022

### **Reporting Summary**

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

<u> </u>				
<b>S</b> 1	121	tic:	ħΠ	$\sim$

For	all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Confirmed
	$oxed{\boxtimes}$ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	🔀 A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	The statistical test(s) used AND whether they are one- or two-sided  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
	A description of all covariates tested
$\boxtimes$	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
	For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
$\boxtimes$	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
$\boxtimes$	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
$\boxtimes$	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
	Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.
So <sup>.</sup>	ftware and code
Poli	cy information about <u>availability of computer code</u>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

#### Data

Data collection

Data analysis

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability

REDCap 12.0.29

R (version 3.6.3)

- For clinical datasets or third party data, please ensure that the statement adheres to our policy

EuroFlow™ developed and validated a Multiple Myeloma Minimal Residual Disease (MM-MRD) database containing representative flow cytometry data sets from normal healthy bone marrow samples processed in different standardized centers. The database (available through Infinicyt™) when used with files which follow standardized operating procedures allows for an automated analysis of the complete bone marrow sample (https://www.cytognos.com/euroflow-databases/resources/multiple-myeloma/). Due to Icelandic law on ethics in research, data privacy regulations, and per informed consent for participants in this study, the

patient-level data used for this study cannot be shared. We encourage researchers or parties interested in collaboration for non-commercial use to apply to the corresponding author (sigyngvi@hi.is). The request will be reviewed by the iStopMM team to verify whether the data sharing is within the restrictions of the study's ethical approvals.

#### Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender

Analysis on sex differences were performed and reported.

Population characteristics

The median age of all screened participants was 61 years (range 41-103), 34.619 (46%) were male and 40.803 (54%) were female. A total of 193 individuals were diagnosed with SMM, of which 116 (60%) were male, 77 (40%) were female and the median age was 70 years (range 44-92).

Recruitment

Recruitment is described in "Rögnvaldsson S et al. Iceland screens, treats, or prevents multiple myeloma (iStopMM): a population-based screening study for monoclonal gammopathy of undetermined significance and randomized controlled trial of follow-up strategies. Blood Cancer J. 2021 May 17;11(5):94. doi: 10.1038/s41408-021-00480-w. PMID: 34001889; PMCID: PMC8128921." which is cited in the manuscript. All individuals born before in 1975 and earlier residing in Iceland September 9th, 2016 (148,704 individuals) were invited to participate. A letter containing a detailed information brochure and consent form was mailed to them and an extensive campaign on social and conventional media was launched introducing the study to the Icelandic public. This campaign was followed by phone calls to those who had not yet signed up for the study. The only exclusion criterion was previously known lymphoproliferative disease other than MGUS (including previous MM and SMM). A total of 80,759 participants (54.3% of the underlying Icelandic population) gave informed consent for participating in the screening and during the sampling phase a total of 75,422 blood samples were collected.

Persons from the general population self selected into the study cohort and participants may have been healthier, more health-aware, and have higher socioeconomic status than the general population. Participation was lower in the youngest and oldest age groups, higher among females, and slightly higher in rural areas. However, screening was free and available at all healthcare institutions in Iceland and the prevalence numbers are adjusted for age and sex. Furthermore, more than half of the national population were screened as part of this study and we believe that the study population is representative of the underlying general population.

Ethics oversight

The iStopMM study protocol was approved by the Icelandic National Bioethics Committee (Number 16-022, date: 2016-04-26) and the Icelandic Data Protection Agency

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one belo	w that is the best fit for your research	. If you are not sure, read the appropriate sections before making your selection.
X Life sciences	Rehavioural & social sciences	Fcological evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

### Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The study on prevalence of SMM is descriptive and sample size was therefore not determined. In the iStopMM study all inhabitants in Iceland over the age of 40 were invited to participate.

Data exclusions

Our SMM cohort diagnosed through the iStopMM study and the only exclusion criterion in the iStopMM study was previously known lymphoproliferative disease other than MGUS or SMM.

Replication

Our findings can not be replicated in any existing cohort that we know of, to replicate the findings another population would have to be screened with blood samples and bone marrow testing.

Randomization

Randomization was not used for this study. We report results for all patients diagnosed with SMM in the iStopMM study and estimate prevalence of SMM using arm 3 of the iStopMM study. The randomization for the iStopMM RCT is described in Rögnvaldsson S et al. Iceland screens, treats, or prevents multiple myeloma (iStopMM): a population-based screening study for monoclonal gammopathy of undetermined significance and randomized controlled trial of follow-up strategies. Blood Cancer J. 2021 May 17;11(5):94. doi: 10.1038/s41408-021-00480-w. PMID: 34001889; PMCID: PMC8128921." which is cited in the manuscript. Participants with an M protein or pathological FLC results were randomized into three study arms in a dynamic, non-predetermined manner and to avoid skewed distribution of high-risk MGUS and light chain MGUS, randomization is carried out by blocks of having an M protein >1.5g/dL and having light chain MGUS.

Blinding

For the iStopMM study, blinding is not possible for the clinicians because the RCT randomizes into two different follow-up strategies or no follow-up, but the statistical analysis in different outcomes between the groups is blinded. In this study on SMM prevalence we only report results for arm 2 and 3 because no patients have been diagnosed with SMM from arm 1.

### Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample

State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.

Sampling strategy

Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.

Data collection

Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.

Timing

Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Non-participation

State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

### Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample

Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

 $\label{procedure} \textit{Describe the data collection procedure, including who recorded the data and how.}$ 

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work?

#### Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access & import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

### Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a Involved in the study	
	X Antibodies	ChIP-seq	
$\boxtimes$	Eukaryotic cell lines	Flow cytometry	
$\boxtimes$	Palaeontology and archaeology	MRI-based neuroimaging	
$\boxtimes$	Animals and other organisms		
$\boxtimes$	Dual use research of concern		

#### **Antibodies**

Antibodies used

The Multiple Myeloma Minimal Residual Disease Panel (MM-MRD) kit was used for cell staining along with two EuroFlow recommended drop-in antibodies (CD27-BV510, clone O323, manufacturer BioLegend, catalogue number 302836, and CD138-BV421, clone MI15, manufacturer BD Biosciences, catalogue number 562935). The kit (manufacturer Cytognos, catalogue number CYT-MM-MRD) is comprised of a pre-mixed 6-color 8-antibody combination (CD19-PE-Cy7, clone SA287, CD38-FITC, multi epitope, CD45-PerCp-Cy5.5, clone EO1, CD56-PE, clone C5.9, CD81-APC-C750, clone M38, CD117-APC, clone 104D2, CylgKappa-APC, polyclonal, CylgLambda-APC-C750, polyclonal) for staining in two separate tubes. Lyophilized multicolor antibody vials in the MM-MRD kit were reconstituted in volume according to manufacturer's description. All antibodies were used without pre-dilution for staining of a target number of 5 million nucleated cells in 300 μL of staining buffer per tube following bulk red blood cell lysis of whole bone marrow samples.

Validation

The antibodies used for flow cytometry analysis in this study are part of a standardized and commercially available EuroFlow developed MM-MRD method.

Flores-Montero, J. et al. Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia 31, 2094–2103 (2017). https://doi.org/10.1038/leu.2017.29

### Eukaryotic cell lines

Cell line source(s)

Policy information about cell lines and Sex and Gender in Research

oney information about <u>centifies and sex and defider in Nescaren</u>

State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.

Authentication Describe the aut

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

#### Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

#### Animals and other research organisms

Policy information about <u>studies involving animals</u>; <u>ARRIVE guidelines</u> recommended for reporting animal research, and <u>Sex and Gender in Research</u>

Laboratory animals

For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Reporting on sex

Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

#### Clinical data

Policy information about <u>clinical studies</u>

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

This manuscript does not report results from a clinical trial. The iStopMM trial: ClinicalTrials.gov identifier: NCT03815279.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

A pilot recruitment phase for the iStopMM study was started in Akranes in Western Iceland on September 15th, 2016 and the wholenation recruitment phase commenced on November 15th, 2016, and continued until February 20th, 2018. Blood samples were collected for screening until December 31st 2020. First bone marrow samples were collected until December 31st 2021.

Outcomes

The primary outcome was to describe the prevalence of SMM in the general population and was estimated from arm 3 of the istopMM study. The estimate for each age group and corresponding standard errors, were used to report prevalence with 95 % confidence interval for each age group (Table 4).

#### Dual use research of concern

Policy information about <u>dual use research of concern</u>

#### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No Yes		
Public health		
National security	National security	
Crops and/or livest	ock	
Ecosystems		
Any other significan	nt area	
Experiments of concer	'n	
Does the work involve an	y of these experiments of concern:	
No Yes		
Demonstrate how	to render a vaccine ineffective	
	to therapeutically useful antibiotics or antiviral agents	
	nce of a pathogen or render a nonpathogen virulent	
	ibility of a pathogen	
Alter the host rang	e of a pathogen	
	diagnostic/detection modalities	
	nization of a biological agent or toxin	
	Illy harmful combination of experiments and agents	
ChIP-seq		
	v and final processed data have been deposited in a public database such as <u>GEO</u> .	
Confirm that you have	e deposited or provided access to graph files (e.g. BED files) for the called peaks.	
Data access links May remain private before public	For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.	
Files in database submissi	ion Provide a list of all files available in the database submission.	
Genome browser session (e.g. <u>UCSC</u> )	Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.	
Methodology		
Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.	
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.	
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.	
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.	
	(week)	
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.	

repository, provide accession details.

### Flow Cytometry

Confirm that:			
The axis labels state the marke	er and fluorochrome used (e.g. CD4-FITC).		
The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).			
All plots are contour plots with outliers or pseudocolor plots.			
A numerical value for number	of cells or percentage (with statistics) is provided.		
Methodology			
	EDTA-anti-coagulated bone marrow aspirate samples were processed and stained using regents from the Multiple Myeloma Minimal Residual Disease kit from Cytognos (catalog code: CYT-MM-MRD) and the recommended drop-in antibodies of CD138-BV421 (BD Biosciences) and CD27-BV510 (BioLegend) according to protocols detailed by the kit manufacturer (www.cytognos.com/products/cyt-mm-mrd8). Samples were processed within 24 hours of collection and acquired on a single FACSCanto II flow cytometery (BD Biosciences).		
Instrument	FACSCanto II - 3 laser analyzer (4B/2R/2V) CE-IVD (P/N: 338962)		
	Stained cells were measured using the FACSDiva software (BD Biosciences). Flow cytometry data was analyzed using the nfinicyt software (Cytognos).		
	Flow cytometry was used in this study for immunophenotypical analysis of whole bone marrow samples. As samples were not pre-sorted, sample purity was not assessed.		
	The Automatic Gating & Classification tool in the Infinicyt software combined with the EuroFlow MM-MRD reference database was used for identification of cell population. Using this method, all events are automatically identified by clustering algorithms. The clusters are then automatically compared with the EuroFlow database and clusters that are identical with cel populations in the database are directly identified and clusters that are not identical are assigned to check populations for manual review and classification as normal or abnormal cells.  No flow cytometry figures or plots are included in the paper.		
Magnetic resonance im	naging		
Design type	Indicate task or resting state; event-related or block design.		
Design specifications	Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.		
Behavioral performance measure	State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).		
Acquisition			
Imaging type(s)	Specify: functional, structural, diffusion, perfusion.		
Field strength	Specify in Tesla		
Sequence & imaging parameters	Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.		
Area of acquisition	State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.		
Diffusion MRI Used			
	Not used		

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization	If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.		
Normalization template	Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.		
Noise and artifact removal	Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).		
Volume censoring	Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.		
Statistical modeling & infe	rence		
Model type and settings	Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).		
Effect(s) tested	tested  Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.		
Specify type of analysis:	Whole brain ROI-based Both		
Statistic type for inference (See <u>Eklund et al. 2016</u> )	Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.		
Correction	Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).		
Models & analysis  n/a   Involved in the study			
Functional and/or effective co	nnectivity Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).		
Graph analysis	Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).		

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.