Improved biorepository to support sickle cell disease genomics and clinical research: A practical approach to link patient data and biospecimens from Muhimbili Sickle Cell Program, Tanzania

Upendo Masamu[1*], Raphael Z Sangeda[1,2], Josephine Mgaya[1], Siana Nkya[3], Beatrice Octavian[1], Frank R Mtiiye[1], Joyce Nduguru[1], Agnes Jonathan[1], Daniel Kandonga[1], Irene K Minja[1,4], Paschal Rugajo[1,5], Emmanuel Balandya[6], Julie Makani[1,7]

[1]Muhimbili Sickle Cell Program, Department of Hematology and Blood Transfusion, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[2]Department of Pharmaceutical Microbiology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[3]Department Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[4]Department of Restorative Dentistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[5]Department of Internal Medicine, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[6]Department of Physiology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[7]Department of Hematology and Blood Transfusion, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

* Corresponding author: umasamu@blood.ac.tz

Abstract

Genetic modifiers underlying various sickle cell disease phenotypic expressions are largely unexplored in Africa due to lack of databases linking biospecimens with demographic and clinical data. The problem may be compounded by a complete lack of biorepositories in these settings. This article aims to document the physical verification of biospecimens stored in the biorepository and link them to patient clinical and demographic information to facilitate plans for genomic and related clinical research studies. We reviewed and updated the existing biorepository infrastructure at Muhimbili Sickle Cell Programme in Dar es Salaam, Tanzania. The database of archived biospecimens was updated with the location information of respective biospecimens following the physical verification of biospecimens and then mapping the patient demographic and clinical data with the biospecimen data using sickle cell patients' demographic identifiers. Three freezers maintained at -80°C store a total of 74,079 biospecimens, of which 63,345 were from 5,159 patients registered in the Muhimbili Sickle Cohort from 2004 to 2016. Out of stored biospecimens, follow-ups were 46,915 (74.06%), control 8,067 (12.74%), admission 5,517 (8.71%) and entry 2,846 (4.49%). Of these registered patients, females were 2,521 (48.87%) and males were 2,638 (51.13%). The age distribution was 1 to 59 years, with those above 18 years being 577 (11.18%) and children 4,582 (88.82%) of registered patients. The notable findings during the process include a lack of automated biospecimen checks, laboratory information management system and standardization of equipment used, biospecimens not linked to clinical and demographic data, date format inconsistencies, lack of regular updating of a database on exhausted biospecimens and updates when biospecimens are moved between positions within freezers. Well-organized biorepository plays a crucial role in answering future research questions. Enforcing strict standard operating procedures and quality control standards will ensure that laboratory scientists and other users adhere to the best biospecimen management procedures.

Introduction

Sickle cell disease (SCD) is an inherited disease caused by a single-gene mutation affecting the β-globin gene on chromosome 11. It results in an abnormal hemoglobin protein (HbS) which affects the shape and function of red blood cells (RBC) and subsequently impacts nearly all body organ systems [1,2]. SCD is one of the most prevalent inherited blood disorders. An estimated 300,000 individuals are born each year with SCD, the highest burden being in Africa, where up to 75% of SCD births occur [3]. It is estimated that 50–80% of infants born with SCD in Africa die before the age of 5 years [4]. Research on SCD is crucial, allowing the discovery of a cure that will reduce its burden on afflicted families. Therefore, it is imperative that the current and future research on SCD store and maintain biospecimens appropriately linked to their clinical and demographic information.

Biorepositories hold a crucial role in achieving biomedical progress and improving research capacity in Africa [5–7], given the advancements in technology in medical sciences. Biospecimens should be stored in a manner to support not only current use but also future clinical or research studies. Human biorepositories are used to understand diseases and diagnoses and ultimately provide information on the cure and treatment [8,9]. Several initiatives have been established to support African biorepository activities. Such initiatives include the Human, Heredity, and Health in Africa (H3Africa), which has helped several biorepositories such as the Integrated Biorepository at Makerere University College of Health Sciences (IBRH3AU), Institute of Human Virology (IHV) H3Africa Biorepository in Abuja Nigeria Clinical Laboratory Services (a Division of the Wits Health Consortium) in Johannesburg South Africa. According to the literature, Gambia was the first country in Africa to establish a national DNA bank [10,11]. Sustainability and the ability to answer research questions are critical for a standard biorepository [12–15]. A well-maintained biorepository is expected to have stored biospecimens to ensure quick biospecimen retrieval and the validity of

the analytical tests. This includes a clear link between the biospecimen data and clinical and demographic information, increasing the scientific value of stored biospecimens [16,17]. This can be achieved if clearly defined procedures govern biorepository processes. Several discussions have been around regarding the standards that any biorepository is supposed to have. One of the discussions is on standard operating procedures (SOPs) for biorepository processes such as sample collection, storage, processing, and information management such as Laboratory information management system (LIMS) [16,18,19]. There are few established biorepositories for SCD globally. These establishments would speed the discovery of mitigations against SCD, including discovering novel biological markers for treatment.

The Muhimbili Sickle Cell Programme at the Muhimbili University of Health and Allied Sciences (MUHAS) in Tanzania was established in 2004 and aimed to offer clinical care and build research capacity in SCD [1,3]. The program recruited a Muhimbili Sickle Cell Cohort (MSC) database of over 5,000 SS individuals, including their demographic, clinical, and biospecimens data. The main goal of the program was to conduct research on SCD that will identify the disease spectrum, causes of disease morbidity and mortality, and describe genetic determinants of the disease [20–25]. Following the research impact achieved with the MSC[3], MUHAS was awarded twice a grant named Sickle Pan African Consortium (SPARCo) by the National Institutes of Health (NIH). This was a collaborative grant between MUHAS in Tanzania as a research hub and a site, Cape Town University as a data coordinating center and two other research sites in Nigeria and Ghana in phase 1 between 2017 and 2021. In phase 2, granted between 2021 and 2026, additional three sites in four countries (Mali, Uganda and a shared site in Zimbabwe and Zambia) were established. SPARCo plans to conduct genomic and clinical studies in the second phase of the SPARCo grant award [26].

The challenge in Muhimbili Sickle Program and other centers in African counties is the ability to manage and trace biospecimens that move between different positions and locations in the

biorepository. In addition, linking the biospecimen data to the clinical and demographic data is tedious without an automated way of handling and storing biospecimens. Linking the biospecimens with their clinical and demographic data is essential because it may give insight into the pattern that may accelerate SCD research progress and discovery. Furthermore, automatic handling and storing of biospecimens is necessary to allow biospecimen integrity and reproducibility of results [19,27]. To achieve this goal, this study was designed to facilitate genomics research by preparing easily maintainable records of biospecimens from the previous MSC cohort and new biospecimens from patients. We describe the experience of conducting quality control and verifying all archived biospecimens from Muhimbili Sickle Program and Sickle Pan African Research Consortium (SPARCo) and integrate the biospecimen database with the clinical and demographic data using the Research Electronic Data Capture (REDCap). The system established will allow quick and intuitive data management that is scalable. The documentation of quality assured biorepository shall serve as the basis for other SSA countries to emulate in practice and pave the way for collaborative genomic studies utilizing the existing biospecimens.

MATERIALS AND METHODS

Study design and population

The study was conducted among patients who consented to participate in the MSC study and its nested studies, such as the Vascular Function Intervention Trial (VFIT) and Strategic Award Study [3,28]. The enrollment into the MSC included all individuals who attended a clinic or patients suspected clinically diagnosed to have SCD at Muhimbili National Hospital (MNH) in Dar es Salaam, Tanzania. Patients were registered mainly from the coastal regions and other regions in Tanzania.

Ethics Statement

The Senate Research and Publications Committee of the Muhimbili University of Health and Allied Sciences (MUHAS) in Tanzania approved the study (Ref. No.DA.282/298/01.C/). Written informed consent was obtained from all participants. For individuals 18 years or above, while for minors, a parent/ guardian consented and signed the consent on behalf of the patient; adolescents provided assent.

Biospecimens verification

The study reviewed, quality assured and updated the existing biorepository infrastructure at the program to link SCD biospecimen data with their clinical and demographic data. During the MSC, patients were identified at entry, admission, follow-up, and as a control which were recorded as visit types. Different biospecimens were collected, including plasma, buffy coat, serum, DNA and isolates. These biospecimens were stored in microtubes within a cryobox, whereby one cryobox would only store one biospecimen type for a specific study. The cryoboxes were labeled on top with the study name, biospecimen type and box number. The numbers on the cryobox were unique for each biospecimen type throughout the study. Basic clinical and demographic information was collected using case report forms (CRF).

Physical verification of biospecimen location within a freezer was done by verifying information such as freezer number, a compartment within a freezer, rack and box number. Information regarding the biospecimen, such as type and study, were also verified during the process. Data from physical verification were compared to the one on the database and updated corresponding records. The final database was cleaned up by maintaining all records of biospecimens found in the freezers during physical verification and having all the verified information about the biospecimens and freezer locations.

Tubes found empty and still in the freezers were removed and updated in the database. The clean biospecimen data was mapped and linked to its clinical and demographic information using demographic identifiers through which two categories were identified cohort-related for whom demographic identifiers were available and non-cohort for biospecimens from patients not admitted in the MSC cohort. The clean final biospecimen data were migrated to the REDCap database to allow easy record-keeping and retrieval. A REDCap (Research Electronic Data Capture) database was created to store curated biospecimens and patient data from this study. REDCap is a secure, web-based application designed to support data capture for research studies, providing (1) an intuitive interface for validated data entry; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to standard statistical packages; and (4) procedures for importing data from external sources [29,30]. The REDCap system and tools were hosted at Muhimbili University of Health and Allied Science.

Data analysis plan

Descriptive such as counts, averages and proportions were used to summarize patients' demographic and biospecimen data. Rstudio 2022.07.2 was used for summarizing the data.

Results

The Hematology Clinical and Research Laboratory (HCRL) at Muhimbili has three freezers (Fig1A) used to store the biospecimens from MSC. The freezer has three areas within a compartment, columns, and rows (Fig1B). Biospecimens of types of plasma, buffy coat, deoxyribonucleic acid (DNA), serum, urine, washed red blood cells and bacterial isolates were aliquoted in the 0.5-2uL microtubes and arranged in 9x9 cryoboxes. The freezers were monitored for temperature checks twice daily and adequately documented. A backup generator

is connected to the main power of the laboratory to observe the quality of the archived biospecimens in case of a power failure. Currently, the HCRL contains a total of 74,079 biospecimens. The space occupied by these biospecimens was 70% of the available space for archived biospecimens stored in the three -80°C freezers.

Out of 74,079 archived biospecimens, 63,345 biospecimens were from 5,159 patients registered under MSC. The remaining 10,734 biospecimens were from nested studies within the MSC. More patients were seen at the inception of the project in 2004, but in 2016 new clinics were opened at other referral hospitals that patients were referred to instead of coming to Muhimbili (Fig 2). Of 5,159 registered patients, the females were 2,521 (48.87%), while the males were 2,638 (51.13%). In terms of age distribution, the minimum age was less than one year while the maximum age was 59, with those above 18 years being 577 (11.18%), while children were 4,582 (88.82%) of registered patients under MSC "Table 1". The total number of biospecimens collected per patient is categorized in Fig 3. The group labeled 1 is the group with patients with only one biospecimen collected and others are in intervals of 10 biospecimens.

Table 1: Demographic characteristics of patients under the Muhimbili Sickle Cohort

Attribute	**N (%)**	**Test statistics (p-value)**
Gender		
Female	2,521 (48.87%)	
Male	2,638 (51.13%)	χ^2= 2.65(0.1033)
Age (years)		
Minimum	< 1 year	
Maximum	59	
Average	8.29	
Age category (N (%))		
Adults	577 (11.18%)	
Children	4,582 (88.82%)	χ^2= 3109.10 (< 0.001)

Out of 63,345 biospecimens that were collected per each visit type follow-ups were 46,915 (74.06%), control 8,067 (12.74%), admission 5,517 (8.71%) and entry 2,846 (4.49%) "Table 2". From the physical verification, biospecimen types present include plasma 38,008 (60.00%), buffy coat 15,323 (24.19%), serum 4,671 (7.37%), DNA 4,528 (7.15%) and bacterial isolates 258 (0.41%) "Table 3". The 557 biospecimens "Table 3" were from a nested study on Amino Acids in Tanzanian Children with Sickle Cell Disease under Vascular Function Intervention Trial (V-FIT)[31]. One of the outcomes of the verification process was to generate and update freezer plan maps, usually placed in front of the freezer showing the position of boxes with biospecimens within the freezers. The map captures the biospecimen type, box number, compartment, rack position within a freezer and the visit type (Fig 1C).

Table 2: Number of archived biospecimens per year and visit type under Muhimbili Sickle Cohort

Year	Admission	Control	Entry	Follow-up	Year total
2004	34	10	965	1,144	2,153 (3.40%)
2005	442	154	255	4,261	5,112 (8.07%)
2006	224	275	381	4,785	5,665 (8.94%)
2007	359	859	336	7,575	9,129 (14.41%)
2008	619	1,221	266	6,270	8,376 (13.22%)
2009	564	423	80	4,409	5,476 (8.64%)
2010	529	462	63	3,847	4,901 (7.74%)
2011	857	1,025	106	5,258	7,246 (11.44%)
2012	607	696	139	3,162	4,604 (7.27%)
2013	568	1,032	154	3,255	5,009 (7.91%)
2014	391	1,232	78	2,393	4,094 (6.46%)
2015	323	637	23	524	1,507 (2.38%)
2016		41		32	73 (0.12%)
Visit type total	**5,517 (8.71%)**	**8,067 (12.74%)**	**2,846 (4.49%)**	**46,915 (74.06%)**	**63,345**

Table 3: Number of each archived biospecimen type under the Muhimbili Sickle Cohort

Year	Buffy coat	DNA	Isolates	Plasma	Serum	V-FIT
2004	589	20		1,153	391	
2005	2,407	93		2,420	192	

2006	2,121	240	69	2,933	302	
2007	1,449	1,829	37	4,699	1,115	
2008	738	673	95	6,292	578	
2009	603	205	54	3,857	757	
2010	254	296	3	4,139	209	
2011	1,339	1,172		4,365	370	
2012	1,624			2,559	253	168
2013	1,993			2,487	223	306
2014	1,604			2,267	140	83
2015	550			816	141	
2016	52			21		
Biospecimen type total	**15,323 (24.19%)**	**4,528 (7.15%)**	**258 (0.41%)**	**38,008 (60.00%)**	**4,671 (7.37%)**	**557 (0.88%)**

Discussion

This paper documents best practices and challenges from verifying archived biospecimens and experience at the Muhimbili Sickle Cell Program in Dar es Salaam, Tanzania. Physical verification of archived biospecimens was carried out whereby a total of 74,079 biospecimens were archived, of which 63,345 biospecimens were from 5,159 patients registered under MSC. Different biospecimens, such as buffy coat, plasma, DNA, serum and bacterial isolates, were archived.

During the physical verification process, some observations needed to be addressed to have a sustainable biorepository to answer future research questions. The absence of an automated way of handling biospecimens made the verification process tedious and manual. The lack of automatic biospecimen checks added to the long, tedious biospecimen verification process. Institutions planning to have biorepositories should consider having an automated way of handling biospecimens, such as a Laboratory Management Information System (LIMS), to allow automatic biospecimen checks that will enable fast and efficient tracking of archived biospecimens. The use of an automated way of handling and storing biospecimens was also recommended by other researchers to preserve biospecimen integrity and allow the reproducibility of results [19,27,32,33].

Some of the tubes with archived biospecimens were missing date labels that indicate when the biospecimens were taken. The recording of the date on the tube needs to be enforced because biospecimens are only kept for a specific period and those with no dates may need to be discarded. Proper date annotation is critical to allow biospecimens' data integrity and quality [17,19]. In addition, some of the archived biospecimens were exhausted in the freezer, but their information was not updated in the database. This calls for laboratory scientists to regularly update the database with information on the exhausted biospecimens. This will allow other users to be informed of the finished biospecimens. The process can be enforced and make laboratory scientists more accountable by implementing a LIMS that tracks users' activities by their identifiers [32].

Lack of standards on the type of tube to be used, whereby some tubes had volume labels while some did not. The absence of a volume label caused difficulties verifying the volume for the archived biospecimens. This calls for developing an SOP on the standard of the types of equipment to be used for biospecimen archiving, which will guide the procurement of respective equipment. Other researchers also recommended the need to have SOPs for the procurement of biospecimen archiving equipment [8].

Different recorded date formats from multiple users who entered data in Microsoft Excel. The date format challenge can also occur when a proper database system is in place. However, the database management engine usually handles date format issues more conveniently. Internet connectivity is another challenge in Africa, preventing real-time access to LIMS. In such cases, laboratory scientists may use MS Excel to temporarily log data and then later import it to an internet database via LIMS. Hence to overcome these challenges, there is a need to have an SOP in place that map the process of recording information for the biospecimen [8]. Such SOP includes a clause defining a date format to be used and adhered to.

It was also observed that some biospecimens were moved to new positions within a freezer, but that information was not updated in the biospecimens' database. This calls for logging to track biospecimens' movement within the freezer to account for biospecimen misplacement. Furthermore, a few tubes were empty and still found in the freezer. Strict rules and SOP should be implemented to guide laboratory scientists on biospecimen management. Such rules include removing empty tubes once exhausted and updating the database. In addition, there should be ownership of the biospecimen management systems among the users. This means a person who oversees and enforces all the rules should be answerable to all queries [34]. Staff training and periodic review of processes are essential to ensure that SOPs are being adhered to [33,34].

Conclusion

The biorepository plays a crucial role in answering future research questions to allow the discovery of novel genomic markers for diagnosing and treating SCD. However, this will be achieved if biospecimens are adequately managed. Strict controlled standard operating procedures and quality assured and quality controlled standards must be enforced to ensure that laboratory scientists and other users adhere to the best biospecimen management procedures. This includes SOP for the procurement of standard equipment, biospecimen storage (such as biospecimen labeling during archiving), and movement from and to the freezers. Regular training on best practices and SOPs for new staff and refresher training for existing ones. Furthermore, one should have LIMS and automatic checks that will allow fast and efficient tracking of archived biospecimens.

Acknowledgments

We would like to thank SCD patients whose data were used in this study and all staff at the Muhimbili Sickle Cell Program for their support.

Author Contributions

Conceptualization: Raphael Z Sangeda

Investigation: Upendo Masamu, Josephine Mgaya, Beatrice Octavian and Frank R Mtiiye

Formal analysis and Visualization: Upendo Masamu and Raphael Z Sangeda

Writing – Original Draft Preparation: Upendo Masamu, Raphael Z Sangeda, Josephine Mgaya, Siana Nkya, Beatrice Octavian, Frank R Mtiiye, Joyce Nduguru, Agnes Jonathan, Daniel Kandonga, Irene K Minja, Paschal Rugajo, Emmanuel Balandya and Julie Makani

Writing – Review & Editing: Upendo Masamu, Raphael Z Sangeda, Josephine Mgaya, Siana Nkya, Beatrice Octavian, Frank R Mtiiye, Joyce Nduguru, Agnes Jonathan, Daniel Kandonga, Irene K Minja, Paschal Rugajo, Emmanuel Balandya and Julie Makani.

References

1. Tluway F, Makani J. Sickle cell disease in Africa: an overview of the integrated approach to health, research, education and advocacy in Tanzania, 2004-2016. Br J Haematol. 2017 Jun;177(6):919–29.
2. Masamu U, Sangeda RZ, Kandonga D, Ondengo J, Ndobho F, Mmbando B, et al. Patterns and patient factors associated with loss to follow-up in the Muhimbili sickle cell cohort, Tanzania. BMC Health Serv Res. 2020 Dec 14;20(1):1141.
3. Makani J, Tluway F, Makubi A, Soka D, Nkya S, Sangeda R, et al. A ten year review of the sickle cell program in Muhimbili National Hospital, Tanzania. BMC Hematol. 2018;18(1):1–13.
4. Aygun B, Odame I. A global perspective on sickle cell disease. Pediatr Blood Cancer. 2012 Aug;59(2):386–90.
5. Staunton C, Moodley K. Challenges in biobank governance in Sub-Saharan Africa. BMC Med Ethics. 2013 Dec 11;14(1):35.

6. Edwards T, Cadigan RJ, Evans JP, Henderson GE. Biobanks containing clinical specimens: Defining characteristics, policies, and practices. Clin Biochem. 2014 Mar;47(4–5):245–51.

7. Hewitt RE. Biobanking: the foundation of personalized medicine. Curr Opin Oncol. 2011 Jan;23(1):112–9.

8. Harati MD, Williams RR, Movassaghi M, Hojat A, Lucey GM, Yong WH. An Introduction to Starting a Biobank. In: Methods in Molecular Biology [Internet]. 2019. p. 7–16. Available from: http://link.springer.com/10.1007/978-1-4939-8935-5_2

9. Siwek M. An Overview of Biorepositories—Past, Present, and Future. Mil Med. 2015 Oct;180(10S):57–66.

10. Sirugo G, Loeff MS Van Der, Sam O, Nyan O, Pinder M, Hill A V, et al. A national DNA bank in The Gambia, West Africa, and genomic research in developing countries. Nat Genet. 2004 Aug;36(8):785–6.

11. Staunton C, Moodley K. Challenges in biobank governance in Sub-Saharan Africa. BMC Med Ethics. 2013 Dec 11;14(1):35.

12. Kelly SM, Wiehagen LT, Schumacher PE, Dhir R. Methods to Improve Sustainability of a Large Academic Biorepository. Biopreserv Biobank. 2017;15(1):31–6.

13. Doucet M, Yuille M, Georghiou L, Dagher G. Biobank sustainability: current status and future prospects. Journal of Biorepository Science for Applied Medicine. 2017;Volume 5:1–7.

14. Adedokun BO, Olopade CO, Olopade OI. Building local capacity for genomics research in Africa: Recommendations from analysis of publications in Sub-Saharan Africa from 2004 to 2013. Glob Health Action. 2016;9(1):1–9.

15. De Souza YG, Greenspan JS. Biobanking past, present and future. AIDS. 2013 Jan;27(3):303–12.

16. Harati MD, Williams RR, Movassaghi M, Hojat A, Lucey GM, Yong WH. An Introduction to Starting a Biobank. In: Methods in Molecular Biology. 2019. p. 7–16.

17. Greenberg B, Christian J, Meltzer Henry L, Leavy M ichelle, Moore H. Biorepositories: Addendum to Registries for Evaluating Patient Outcomes: A User's Guide,. https://www.ncbi.nlm.nih.gov/books/NBK493632/#!po=12.5000; 2018.

18. Vaught JB, Henderson MK. Biological sample collection, processing, storage and information management. IARC Sci Publ. 2011;(163):23–42.

19. Oluwole OG, Oosterwyk C, Anderson D, Adadey SM, Mnika K, Manyisa N, et al. The Implementation of Laboratory Information Management System in Multi-Site Genetics Study in Africa: The Challenges and Up-Scaling Opportunities. Journal of Molecular Pathology [Internet]. 2022;3(4):262–72. Available from: https://www.mdpi.com/2673-5261/3/4/22

20. Makani J, Cox SE, Soka D, Komba AN, Oruo J, Mwamtemi H, et al. Mortality in Sickle Cell Anemia in Africa: A Prospective Cohort Study in Tanzania. Schrijver I, editor. PLoS One. 2011 Feb 16;6(2):e14699.

21. Makani J, Komba AN, Cox SE, Oruo J, Mwamtemi K, Kitundu J, et al. Malaria in patients with sickle cell anemia: burden, risk factors, and outcome at the outpatient clinic and during hospitalization. Blood. 2010 Jan 14;115(2):215–20.

22. Mmbando BP, Mgaya J, Cox SE, Mtatiro SN, Soka D, Rwezaula S, et al. Negative epistasis between Sickle and foetal haemoglobin suggests a reduction in protection against malaria. PLoS One. 2015;10(5).

23. Makani J, Mgaya J, Balandya E, Msami K, Soka D, Cox SE, et al. Bacteraemia in sickle cell anaemia is associated with low haemoglobin: A report of 890 admissions to a tertiary hospital in Tanzania. Br J Haematol. 2015;171(2):273–6.

343 24. Cox SE, Makani J, Fulford AJ, Komba AN, Soka D, Williams TN, et al. Nutritional
344 status, hospitalization and mortality among patients with sickle cell anemia in Tanzania.
345 Haematologica. 2011 Jul;96(7):948–53.
346 25. Makani J, Menzel S, Nkya S, Cox SE, Drasar E, Soka D, et al. Genetics of fetal
347 hemoglobin in Tanzanian and British patients with sickle cell anemia. Blood. 2011 Jan
348 27;117(4):1390–2.
349 26. Makani J, Sangeda RZ, Nnodu O, Nembaware V, Osei-Akoto A, Paintsil V, et al.
350 SickleInAfrica. Lancet Haematol [Internet]. 2020 Feb;7(2):e98–9. Available from:
351 https://linkinghub.elsevier.com/retrieve/pii/S2352302620300065
352 27. Linsen L, Van Landuyt K, Ectors N. Automated Sample Storage in Biobanking to
353 Enhance Translational Research: The Bumpy Road to Implementation. Front Med
354 (Lausanne). 2020 Jan 9;6.
355 28. Cox SE, Makani J, Ellins E, Walter G, Mtunguja S, Kirkham FJ, et al. Endothelial
356 Function In Tanzanian Children With Sickle Cell Disease: Baseline Results From The
357 Vascular Function Intervention Trial (VFIT). Blood. 2013 Nov;122(21):984–984.
358 29. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap
359 consortium: Building an international community of software platform partners. J
360 Biomed Inform [Internet]. 2019 Jul;95:103208. Available from:
361 https://linkinghub.elsevier.com/retrieve/pii/S1532046419301261
362 30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic
363 data capture (REDCap)—A metadata-driven methodology and workflow process for
364 providing translational research informatics support. J Biomed Inform [Internet]. 2009
365 Apr;42(2):377–81. Available from:
366 https://linkinghub.elsevier.com/retrieve/pii/S1532046408001226

31. Cox SE, Makani J, Ellins E, Walter G, Mtunguja S, Kirkham FJ, et al. Endothelial Function In Tanzanian Children With Sickle Cell Disease: Baseline Results From The Vascular Function Intervention Trial (VFIT). Blood. 2013 Nov;122(21):984–984.

32. Labarga A, Beloqui I, Martin AG. Information Management. In 2017. p. 29–39. Available from: https://link.springer.com/10.1007/978-1-4939-6921-0_4

33. Cicek MS, Olson JE. Mini-Review of Laboratory Operations in Biobanking: Building Biobanking Resources for Translational Research. Front Public Health [Internet]. 2020 Jul 28;8. Available from: https://www.frontiersin.org/article/10.3389/fpubh.2020.00362/full

34. Vaught J, Lockhart NC. The evolution of biobanking best practices. Clinica Chimica Acta [Internet]. 2012 Oct;413(19–20):1569–75. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0009898112002215

Figure captions

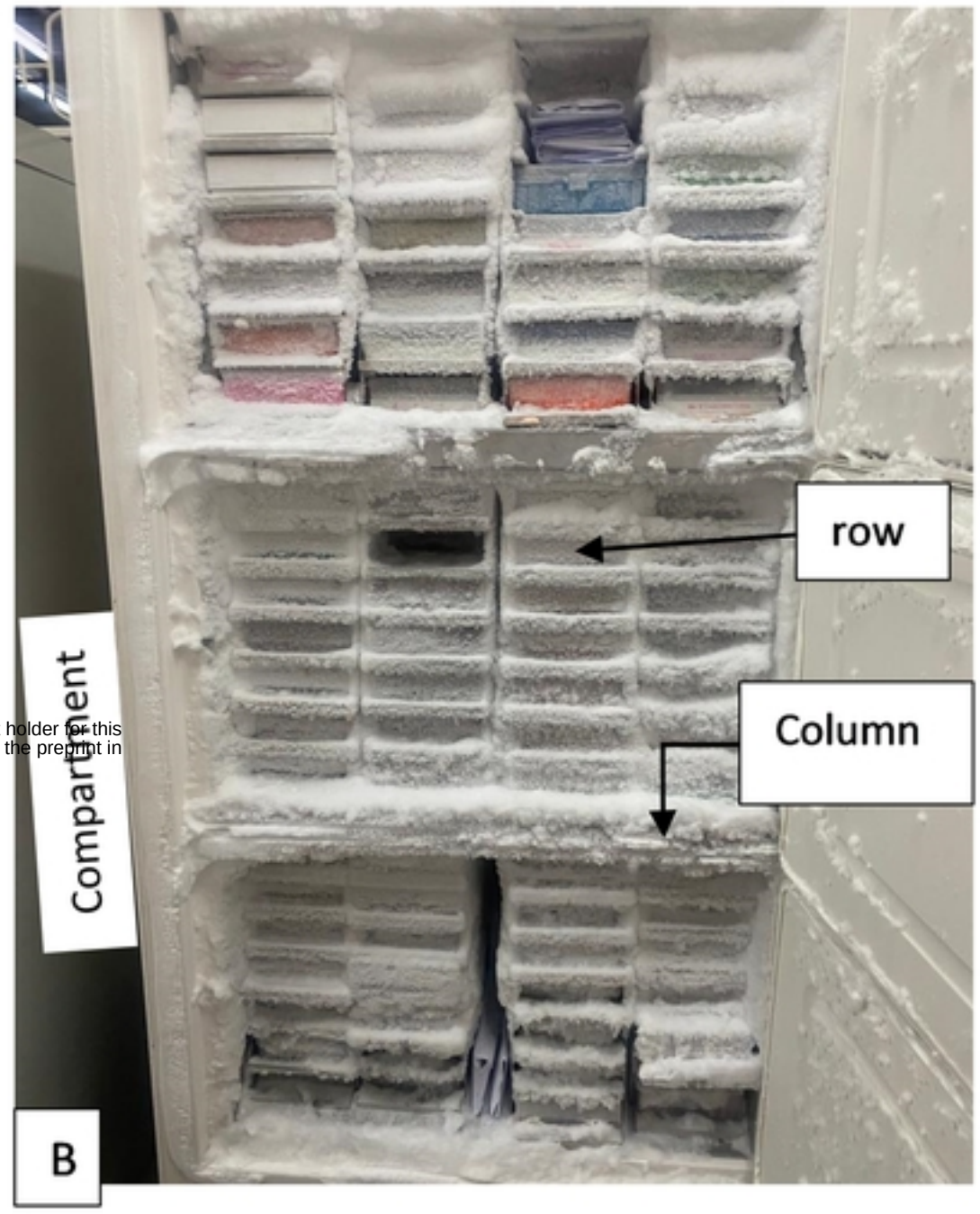
Fig 1: Freezers with the archived biospecimens. (A) -80C freezers for storing archived biospecimen. (B) A picture showing the inside of one of the freezers. (C) Freezer plan map showing cryobox positions within the freezer (Compartment, C-Column and R-Rack, Biospecimen type: P-Plasma, FU-Follow-up, E-Entry, S-Serum)

Fig 2: Patients enrollment trend and biospecimen collection (A) Number of unique visiting patients and biospecimens collected over the years 2004-2016 in Muhimbili Sickle Cohort (MSC); (B) indicates the number of new patients enrolled each year.

391 **Fig 3:** Number of patients with a given number of biospecimens stored in the biorepository.

392 . The group labeled 1 is the group with patients with only one biospecimen collected and others

393 are in intervals of 10 biospecimens.

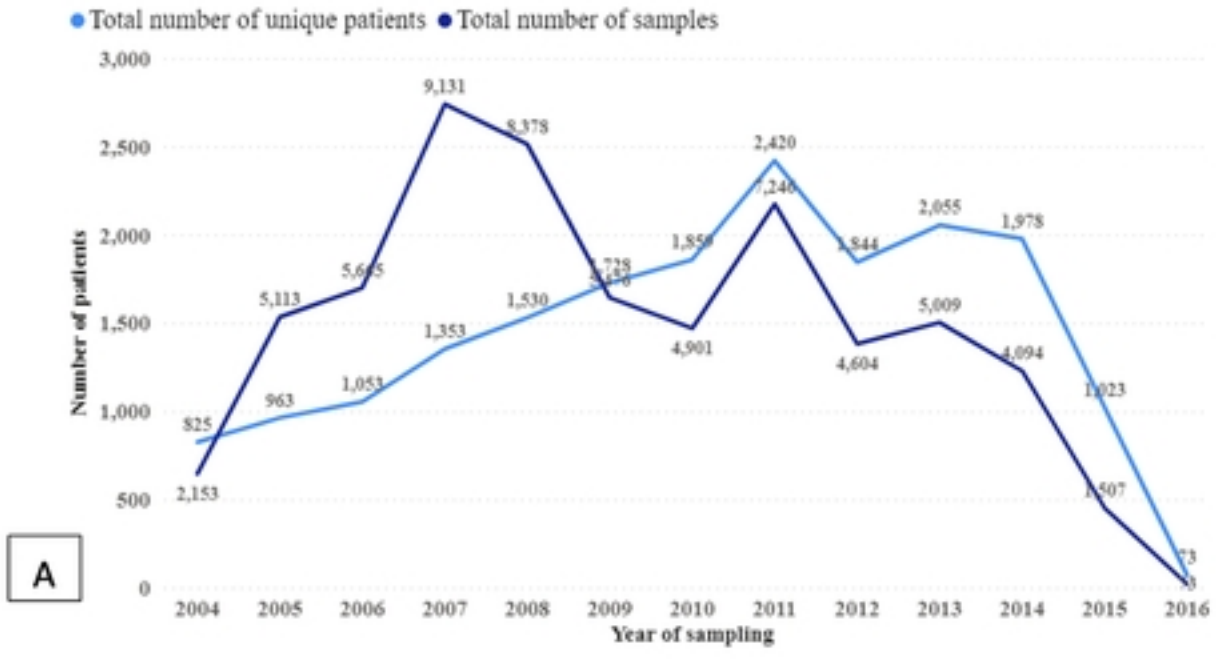


C

						FREEZER 1							
	COLUMN	C1		C2		C3		C4		C5		C6	
COMPARTMENT	RACK	box labe	sample visits	box labe	sample	box label	sample visits	box labe	sample	box labe	sample	box labe	sample
Upper	R1	P-B-77	CONTROL	P-C-1	all	P-C-25	all	P-4	CONTROL	P-9	FU	P-33	FU
		P-B-78	CONTROL	P-C-2	all	P-C-26	all	P-5	CONTROL	P-10	FU	P-34	FU
		P-B-79	CONTROL	P-C-3	all	P-C-27	all	P-6	CONTROL	P-11	FU	P-35	FU
		P-B-80	CONTROL	P-C-4	all	P-C-28	all	P/P-A-7	CONTROL	P-12	FU	P-36	FU
	R2	P-B-81	CONTROL	P-C-5	all	P-C-29	all	P-8	CONTROL	P-13	FU	P-37	FU
		P-B-82	CONTROL	P-C-6	all	P-C-30	all	P-A-9	CONTROL	P-14	FU	P-38	FU
		B-85	CONTROL	P-C-8	all	P-C-31	all	P-A-10	CONTROL	P-15	FU	P-39	FU
		B-86	CONTROL	HCV-9		P-C-32	all	P-A-11	CONTROL	P-16	FU	P-40	FU
	R3	B-81	CONTROL	P-C-9	all	P-C-33	all	P-A-12	CONTROL	P-17	FU	P-41	FU
		B-87	CONTROL	P-C-10	all	P-C-34	all	P-A-13	CONTROL	P-18	FU	P-42	FU
		B-88	CONTROL	P-C-11	all	P-C-35	all	P-A-14	CONTROL	P-19	FU	P-43	FU
		B-89	CONTROL	P-C-12	all	P-1	ADMISSION	P-A-15	CONTROL	P-20	FU	P-44	FU
	R4	P-B-40	ADMISSION	P-C-13	all	P-2	ADMISSION	P-A-16	CONTROL	P-21	FU	P-45	FU
		P-B-41	ADMISSION	P-C-14	all	P-3	ADMISSION	P-A-17	CONTROL	P-22	FU	P-46	FU
		P-B-43	ADMISSION	P-C-15	all	P-4	ADMISSION			P-23	FU	P-47	FU
		S-42	ADMISSION	P-C-16	all	P-5	ADMISSION			P-24	FU	P-48	FU
	R5	B-B-91	CONTROL	P-C-17	CONTROL	P-6	ADMISSION	P-1	FU	P-25	FU	P-49	FU
		P-B-257	FU+E	P-C-18	CONTROL	P-7	ADMISSION	P-2	FU	P-26	FU	P-50	FU
		S-43	ADMISSION	P-C-19	CONTROL	P-8	ADMISSION	P-3	FU	P-27	FU	P-51	FU
		S-45	ADMISSION	P-C-20	CONTROL	P-1	CONTROL	P-4	FU	P-28	FU	P-52	FU
	R6	P-B-258	FU+E	P-C-21	CONTROL	P-2	CONTROL	P-5	FU	P-29	FU	P-53	FU
				P-C-22	CONTROL	P-3	CONTROL	P-6	FU	P-30	FU	P-54	FU
				P-C-23	CONTROL			P-7	FU	P-31	FU	P-55	FU
				P-C-24	CONTROL			P-8	FU	P-32	FU	P-56	FU
	R7	P-57	FU	P-A-77	FU	P-A-K	FU+E	B-5	FU	B-25	FU	B-45	FU
		P-58	FU	P-A-78	FU	P-A-L	FU+E	B-6	FU	B-26	FU	B-46	FU
		P-59	FU	P-A-79	FU	P-A-M	FU+E	B-7	FU	B-27	FU	B-47	FU
		P-60	FU	P-A-80	FU	P-A-N	FU+E	B-8	FU	B-28	FU	B-48	FU
	R8	P-61	FU	P-A-81	FU	P-A-O	FU+E	B-9	FU	B-29	FU	B-49	FU
		P/P-A-62	FU	P-A-82	FU	P-A-P	FU+E	B-10	FU	B-30	FU	B-50	FU
		P-A-63	FU	P-A-83	FU	P-A-Q	FU+E	B-11	FU	B-31	FU	B-51	FU
		P-A-64	FU	P-A-84	FU			B-12	FU	B-32	FU	B-52	FU

FREEZER 2 | FREEZER 1 | FREEZER 7

Figure 1


Total number of unique patients
Total number of samples
Number of patients
Year of sampling
2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
825 963 1,053 1,353 1,530 1,728 1,858 2,420 1,844 2,055 1,978 1,023 73
2,153 5,113 5,665 9,131 8,378 5,476 4,901 7,246 4,604 5,009 4,094 1,507 0
A

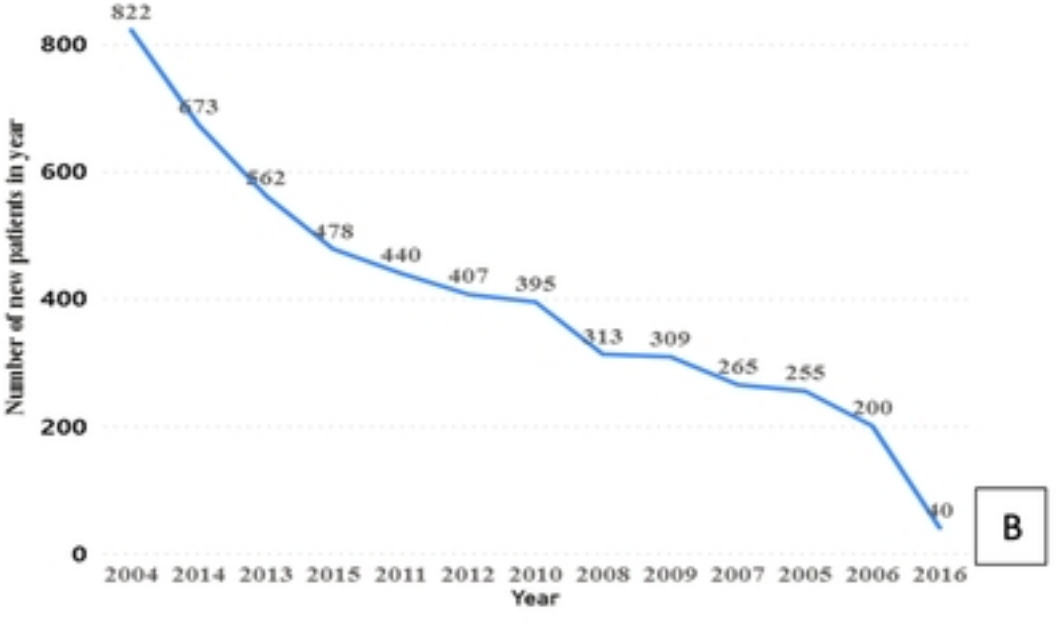

Number of new patients in year
Year
2004 2014 2013 2015 2011 2012 2010 2008 2009 2007 2005 2006 2016
822 673 562 478 440 407 395 313 309 265 255 200 40
B

Figure 2

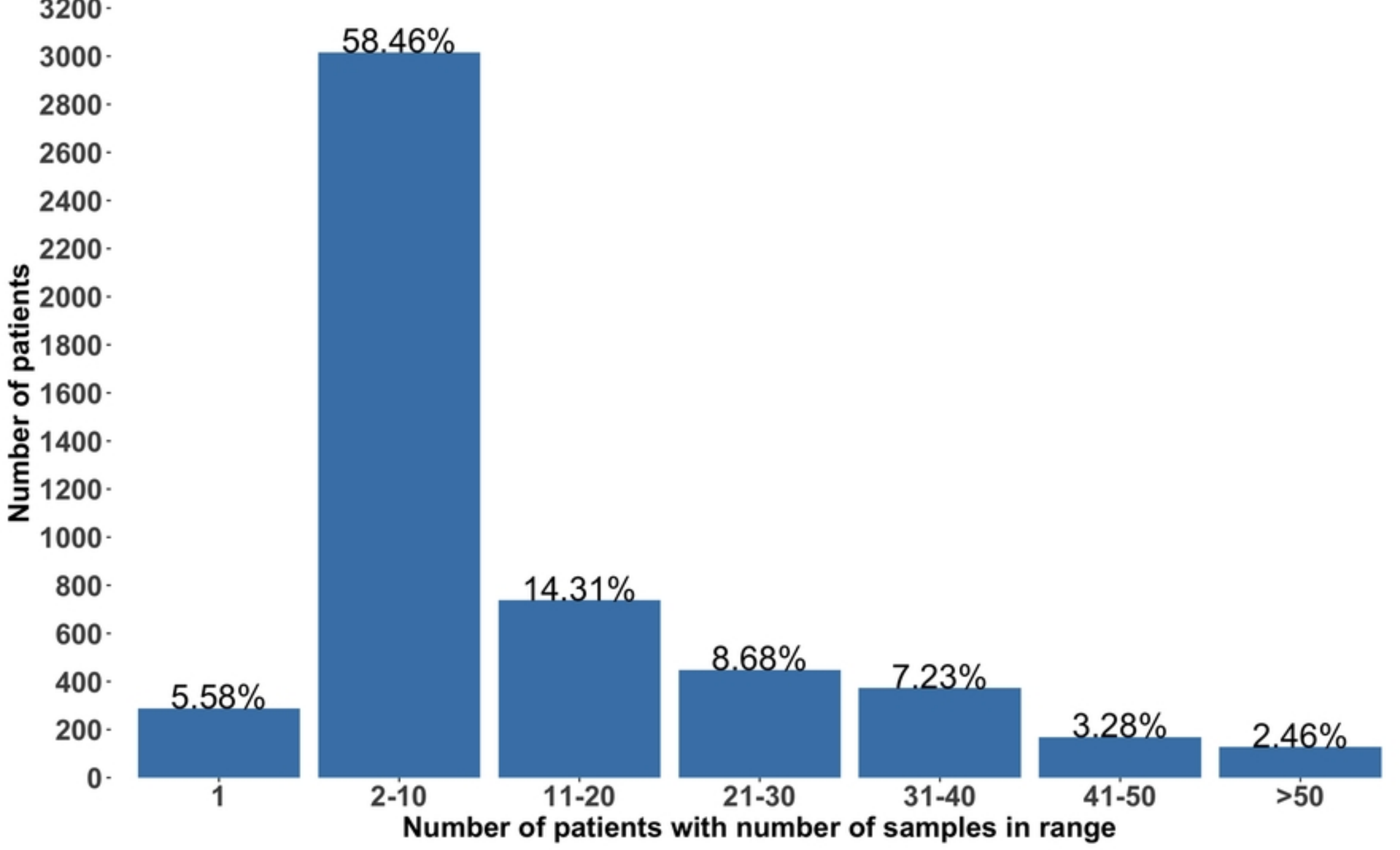

3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Number of patients
58.46%
14.31%
8.68%
7.23%
5.58%
3.28%
2.46%
1
2-10
11-20
21-30
31-40
41-50
>50
Number of patients with number of samples in range

Figure 3