

A Precision Screening Framework for Thalassemia in a High-Mobility Population: Integrating 3D Phenotypic Mapping and Genotype-Specific Thresholds

Hou Qian

239022915@qq.com

Department of Laboratory Medicine Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University)

Weifeng Li

Department of Laboratory Medicine Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University)

Yike Wu

Department of Laboratory Medicine Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University)

Weihong Qin

Department of Laboratory Medicine Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University)

Weihua Zhao

Prenatal Diagnosis Center, Shenzhen Second People's Hospital

Dayong Gu

Department of Laboratory Medicine Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University)

Research Article

Keywords: $\alpha\beta$ -thalassemia, Phenotypic Masking, 3D Topographical Analysis, Reflex Testing, HbA2

Posted Date: April 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9213369/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective

In regions characterized by high hemoglobinopathy prevalence and significant population mobility, defining population-specific diagnostic thresholds is critical to overcoming "phenotypic masking" in complex thalassemia genotypes. This study evaluates the diagnostic utility of hematological indices and HbA_2 through a novel 3D topographical framework to refine screening protocols for the South China population.

Methods

A retrospective analysis was conducted on 1,672 genetically confirmed thalassemia carriers (1,090 α -thalassemia, 522 β -thalassemia, 60 $\alpha\beta$ -compound) and 5,051 healthy controls from January 2018 to June 2025. We utilized Receiver Operating Characteristic (ROC) curves and 3D spatial clustering analysis of Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), and Hemoglobin A_2 (HbA_2) to define optimal cut-off values and diagnostic performance.

Results

The molecular landscape was dominated by the $--^{SEA}/\alpha\alpha$ (54.5%) and $\beta^{CD41-42}/\beta^N$ (30.5%) genotypes, reflecting a distinct regional mutational spectrum. While HbA_2 functioned as a near-perfect biomarker for β -thalassemia (AUC 0.984) and $\alpha\beta$ -compound states (AUC 0.964), it lacked discriminatory power for silent α -carriers (AUC 0.563). Notably, $\alpha\beta$ -coinheritances demonstrated "phenotypic masking," where concurrent α -globin defects partially "normalized" erythrocyte indices compared to pure β -thalassemia (22.00 ± 2.11 vs. 20.60 ± 1.96 pg). Our 3D topographical model revealed significant spatial overlap between controls and silent α -carriers, representing a critical diagnostic "blind spot".

Conclusion

To mitigate the risk of occult carrier transmission, we propose a hierarchical precision framework: indices $HbA_2 > 3.75\%$ for primary β -thalassemia screening. $MCH < 28.5$ pg combined with $HbA_2 < 2.77\%$ as a trigger for reflex α -molecular testing to capture silent carriers. This multidimensional approach addresses the limitations of traditional single-parameter thresholds and provides a robust scientific basis for thalassemia prevention in high-prevalence urban environments.

Introduction

Thalassemia represents a profound global public health challenge, with a carrier burden exceeding 350 million individuals^{1,2}. While Mediterranean regions like Italy and Greece have established robust prevention pathways through the Italian Society for Thalassemia and Hemoglobinopathies (SITE), these frameworks are optimized for β -globin-dominant populations^{3,4}. In the "Thalassemia Belt" of South China, particularly Guangdong, the diagnostic landscape is significantly more complex, with carrier frequencies exceeding 21% in localized hotspots^{5,6}.

Unlike the Mediterranean spectrum, South China is characterized by a high prevalence of α -globin defects, particularly the $--^{SEA}$ deletion⁷. This genetic diversity introduces two critical diagnostic vulnerabilities: Silent α -Carrier Escape: Traditional screening thresholds often fail to capture silent α -carriers, who carry a 25% risk of conceiving offspring with Hb H disease when paired with a minor α -thalassemia partner. Phenotypic Masking: The co-inheritance of α - and β -thalassemia (accounting for 3.6% of our cohort) creates a "diagnostic paradox". The concurrent reduction in both globin chains partially restores intracellular balance, "normalizing" hematological indices and potentially concealing underlying β -globin defects during routine β -thalassemia screening⁸.

The established clinical cascade relies on Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), and Hemoglobin A_2 (HbA_2). However, as critics have noted, HbA_2 remains a non-standardized parameter across global laboratories, and reliance on single-parameter cut-offs can be misleading in regions with diverse mutational pressures⁹. This necessitates the calibration of diagnostic thresholds to local populations and the development of multi-parametric models to resolve spatial overlaps between healthy and pathological phenotypes.

To address these "blind spots," this study systematically correlates genotypic distributions with hematological phenotypes in a large Shenzhen-based cohort ($n = 6,723$). We introduce a 3D topographical model integrating MCV, MCH, and HbA_2 to visualize and quantify the phenotypic space of various thalassemia subtypes¹⁰. Our objective is to establish refined, population-specific screening thresholds and a reflex molecular testing framework that optimizes clinical counseling and prevents occult carrier transmission in high-prevalence urban environments.

Method and materials

Study Population and Ethical Considerations

This retrospective study was approved by the Ethics Committee of Shenzhen Second People's Hospital (Approval No: 2025-765-01PJ) and conducted in accordance with the principles of the Declaration of Helsinki. Between January 2018 and June 2025, we enrolled a cohort of 1,672 genetically confirmed thalassemia carriers (1,090 α -thalassemia, 522 β -thalassemia, and 60 $\alpha\beta$ -compound heterozygotes) and 5,051 healthy controls. Control subjects were verified as negative for both α - and β -globin mutations via molecular testing.

Inclusion criteria for the study group were: (1) age 18–60 years; (2) availability of comprehensive clinical, hematological, and genomic data; and (3) definitive diagnosis supported by laboratory findings and pedigree analysis. Exclusion criteria included: (1) concomitant iron-deficiency anemia (to eliminate interference with microcytic indices); (2) β -thalassemia major or intermedia; and (3) significant systemic organ dysfunction or malignancy. No statistically significant demographic differences were observed between the study and control cohorts ($P > 0.05$), ensuring inter-group comparability.

Hematological and Phenotypic Analysis

Peripheral blood (2–3 mL) was collected in *EDTA-K₂* anticoagulant tubes. Hematological indices, specifically Mean Corpuscular Volume (MCV) and Mean Corpuscular Hemoglobin (MCH), were determined using a Sysmex XN-1000 automated analyzer (Sysmex Co., Kobe, Japan) with manufacturer-certified reagents.

To address concerns regarding the standardization of Hemoglobin *A₂* (*HbA₂*), quantification was performed via capillary electrophoresis on the Helena V8 system (Helena Biosciences, USA). The system underwent daily multi-level calibration and rigorous internal quality control using manufacturer-certified reagents to ensure analytical precision and inter-laboratory comparability, mitigating the impact of non-standardized parameters on diagnostic thresholds.

Molecular Genotyping

Genomic DNA was isolated from peripheral blood leukocytes using an automated nucleic acid extraction platform (Kaishuo Biological Co., Xiamen, China).

α -Thalassemia Analysis: Common deletional variants ($-\alpha^{3.7}$, $-\alpha^{4.2}$, $--^{SEA}$, $--^{THAI}$) were identified using Gap-Polymerase Chain Reaction (Gap-PCR). Non-deletional mutations, including Hb Constant Spring (Hb CS), Hb Quong Sze (Hb QS), and Hb Westmead, were characterized by Reverse Dot Blot PCR (RDB-PCR).

β -Thalassemia Analysis: A panel of 19 β -globin gene mutations (including – 28, –29, CD17, CD26, CD41–42, and IVS-II-654, among others) was screened using reverse dot blot PCR (RDB-PCR). All assays were performed using dedicated thalassemia detection kits (Yilifang Biological Products Co., Ltd., Shenzhen, China) and a Bio-Rad C-1000 thermal cycler. The comprehensive list of the 19 mutations and their corresponding HGVS nomenclature is provided in Supplementary Table S1.

3D Topographical and Statistical Analysis

To resolve the diagnostic challenge of "phenotypic masking" in complex genotypes, we employed a 3D topographical spatial analysis. This multidimensional framework integrated MCV, MCH, and *HbA₂* into a unified diagnostic space to visualize the clustering of different thalassemia phenotypes. ROC curve

analysis was performed using MedCalc® (v20.217) and Python (v3.13) to define optimal cut-off values. Statistical significance was established at a two-tailed $P < 0.0001$.

Data were analyzed using MedCalc® Statistical Software (v20.217) and Python (v3.13). Normality was assessed via the D'Agostino-Pearson test. Given the non-normal distribution of *HbA2*, MCV, and MCH ($P < 0.0001$), continuous variables are presented as medians with interquartile ranges (IQR). Inter-group comparisons were performed using the Kruskal-Wallis test followed by Conover's post-hoc analysis for pairwise differences. Receiver Operating Characteristic (ROC) curves were constructed to determine optimal diagnostic thresholds, sensitivity, and specificity. Statistical significance was established at a two-tailed $P < 0.0001$.

Results

Molecular Landscape and Allelic Heterogeneity

Systematic molecular profiling of 1,672 thalassemia carriers identified 19 distinct variants, reflecting a highly heterogeneous mutational spectrum (Table 1). Rare genotypes, including non-deletional α -variants and complex $\alpha\beta$ -globin coinheritances accounting for $< 1\%$, were collectively categorized as 'Other' (comprehensive breakdown provide in Supplementary Table S2).

Table 1
Genotype Distribution of 1672 Thalassemia Carriers

Diagnostic Category and Genotype	HGVS Nomenclature / Variant Description	No. of Cases (n)	Frequency (%)
α-Thalassemia		1090	65.19
<i>Silent</i> ($-\alpha/aa$)			
$-\alpha^{3.7}/aa$	NG_000006.1: g.34164_37967del3804	281	16.81
$-\alpha^{4.2}/aa$	NG_000006.1: g.30131_34352del4222	90	5.38
$\alpha^{CS}\alpha/aa$	<i>HBA2</i> : c.427T > C; p.Ter143Glnext*31	37	2.21
<i>Minor</i> ($--/aa$ or $-\alpha/-\alpha$)			
$--^{SEA}/aa$	NG_000006.1:g.26264_45564del19301	594	35.53
<i>Hb H Disease</i>			
$-\alpha^{3.7}/--^{SEA}$	Compound heterozygosity (α^+/α^0)	17	1.02
<i>Other^d</i>		71	4.25
β-Thalassemia		522	31.22
$\beta^{CD41-42}/\beta^N$	<i>HBB</i> : c.124_127delTTCT;	159	9.51
$\beta^{IVS-654}/\beta^N$	<i>HBB</i> : c.316-197C > T	151	9.03
β^{CD17}/β^N	<i>HBB</i> : c.52A > T	82	4.90
β^{-28}/β^N	<i>HBB</i> : c.-78A > G	73	4.37
<i>Other^d</i>		57	3.41
$\alpha\beta$-Compound		60	3.59
$--^{SEA}/aa; \beta^{IVS-654}/\beta^N$		10	0.60
<i>Other^d</i>		50	2.99
Total		1672	100.00

Note: N = 1672. Genotypes are described according to the Human Genome Variation Society (HGVS) nomenclature. α -thalassemia deletions ($-\alpha^{3.7}$, $-\alpha^{4.2}$, $--^{SEA}$) are referenced to NG_000006.1. Gene names (*HBA2*, *HBB*) are italicized per standard nomenclature.

^d **Other:** Includes rare non-deletional α -globin variants, β -point mutations, and complex $\alpha\beta$ -globin configurations with an individual frequency of <1%. A comprehensive breakdown of the mutational spectra for all categories is provided in **Supplementary Table S2**.

Abbreviations: CS, Constant Spring; QS, Quong Sze; WS, Westmead; Hb, Hemoglobin; n, number of cases.

α -Thalassemia Spectrum: Among 1,090 α -thalassemia carriers, deletional variants were the primary molecular drivers. The $--^{SEA}/\alpha\alpha$ genotype was most prevalent (54.5% of α -thalassemia cases), followed by $-\alpha^{3.7}/\alpha\alpha$ (25.8%) and $-\alpha^{4.2}/\alpha\alpha$ (8.3%) (Fig. 1A). Non-deletional variants, although less frequent, were dominated by $\alpha^{CS}/\alpha\alpha$ (3.4%) and $\alpha^{WS}/\alpha\alpha$ (2.6%) alleles, which typically correlate with more pronounced clinical phenotypes.

β -Thalassemia Profile: Analysis of 522 β -thalassemia subjects identified a highly concentrated mutational spectrum. Four recurrent mutations accounted for 89.1% of all β -globin defects: $\beta^{CD41-42}/\beta^N$ (30.5%) and $\beta^{IVS-1-654(A>G)}/\beta^N$ (28.9%) were the predominant alleles, followed by $\beta^{Codons17(A>T)}/\beta^N$ (15.7%) and $\beta^{Codons-28(A>G)}/\beta^N$ (14.0%) (Fig. 1B).

Complex $\alpha\beta$ -Coinheritance

Concurrent α - and β -globin mutations were identified in 3.6% of the cohort (n = 60). The most frequent configuration was $--^{SEA}/\alpha\alpha$ co-inherited with $\beta^{IVS-1-654}/\beta^N$ (16.7%; Fig. 1C).

Quantifying Phenotypic Masking in $\alpha\beta$ -Compound States

Our results demonstrate a significant "phenotypic masking" effect in $\alpha\beta$ -compound heterozygotes (n = 60). While HbA_2 levels remained diagnostically elevated ($5.37 \pm 0.63\%$), the co-inheritance of α -globin defects partially normalized red cell indices compared to pure β -thalassemia carriers. Specifically, MCH levels were significantly higher in the $\alpha\beta$ -compound group than in the β -thalassemia group (22.00 ± 2.11 vs. 20.60 ± 1.96 pg, $P < 0.0001$). This partial restoration of α/β globin chain balance suggests that relying solely on microcytic thresholds established in other populations may lead to the underdiagnosis of underlying α -globin defects during routine β -carrier screening.

Phenotypic Clustering and 3D Topographical Analysis

To elucidate genotype-phenotype correlations, we mapped the cohort within a three-dimensional (3D) diagnostic space defined by MCV, MCH, and HbA_2 (Fig. 2). Statistical analysis confirmed profound hematological divergence across genotypes ($P < 0.0001$; Table 2). β -thalassemia and $\alpha\beta$ -compound carriers formed a distinct, high-density cluster characterized by significantly elevated HbA_2 levels compared to healthy controls (2.54 ± 0.42 ; $P < 0.0001$).

Table 2
Comparison of hematological indices across various thalassemia genotypes.

Genotype Group	n	HbA ₂ (%)	MCV (fL)	MCH (pg)
Healthy controls	5051	2.54 ± 0.42	87.89 ± 6.86	29.14 ± 2.94
Silent α -thalassemia	450	2.48 ± 0.34	81.76 ± 5.18 ^a	26.51 ± 2.03 ^a
Minor α -thalassemia	609	2.38 ± 0.29	68.33 ± 3.64 ^a	21.84 ± 1.25 ^a
Hb H disease	31	1.65 ± 0.61	64.87 ± 8.89 ^{ac}	18.82 ± 1.95 ^{ac}
β -thalassemia	522	5.20 ± 0.63 ^b	64.94 ± 5.25 ^a	20.60 ± 1.96 ^a
$\alpha\beta$ -Compound	60	5.37 ± 0.63 ^b	68.99 ± 5.84 ^a	22.00 ± 2.11 ^a

Note: Data expressed as Mean ± SD. Statistical analysis was performed using MedCalc® Statistical Software version 20.217 (MedCalc Software Ltd, Ostend, Belgium);

^a*P* < 0.0001 vs. Healthy Controls: Indicates profound microcytosis and hypochromia, consistent with quantitative defects in globin chain synthesis.

^b*P* < 0.0001 vs. Healthy Controls: Denotes diagnostic elevation of HbA₂ levels, a hallmark of β -thalassemia and $\alpha\beta$ -compound phenotypes.

^c*P* < 0.0001 vs. α -thalassemia minor: Reflects the significantly exacerbated erythrocyte index depressions characteristic of the clinically severe Hb H disease phenotype.

Abbreviations: HbA₂, hemoglobin A₂; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin.

Statistical Methodology: Normality was assessed via the D'Agostino-Pearson test. Inter-group significance was determined using the Kruskal-Wallis test followed by Conover's post-hoc analysis for all pairwise comparisons to account for non-parametric distributions.

Within the α -globin spectrum, a clear gradient of microcytic hypochromic severity was observed. While a 27 pg MCH threshold effectively sequestered most pathological genotypes, "silent" α -carriers showed substantial phenotypic overlap with healthy controls, representing a critical diagnostic "blind spot" for conventional screening.

Comparative Diagnostic Performance and High-Risk Genotypes

ROC curve analysis was performed to quantify the diagnostic utility of hematological markers (Figures S1-S3, Fig. 3, Table 3). HbA₂ demonstrated near-perfect diagnostic accuracy for classical β -thalassemia (AUC 0.984; 95% CI: 0.980–0.987) and $\alpha\beta$ -compound states (AUC 0.964; 95% CI: 0.959–0.968), validating it as the primary screening metric for β -globin defects.

Table 3
Diagnostic performance, optimal cut-off values, and 95% confidence intervals of HbA_2 , MCV, and MCH across thalassemia genotypes.

Diagnosis Group	Marker	Cut-off	AUC (95%CI)	Sensitivity	Specificity
Silent α-thalassemia (n = 450)	HbA_2	≤ 2.77	0.563 (0.551–0.575)	88.0	27.8
	MCV	≤ 86.3	0.659 (0.648–0.670)	85.8	57.8
	MCH	≤ 28.5	0.683 (0.671–0.694)	89.1	59.8
Minor α-thalassemia (n = 609)	HbA_2	≤ 2.62	0.684 (0.673–0.695)	90.0	38.8
	MCV	≤ 75.4	0.903 (0.895–0.910)	97.5	85.1
	MCH	≤ 24.1	0.894 (0.887–0.902)	99.0	84.0
Hb H disease (n = 31)	HbA_2	≤ 2.00	0.850 (0.842–0.859)	74.2	94.6
	MCV	≤ 79.6	0.904 (0.897–0.911)	93.5	73.1
	MCH	≤ 22.0	0.954 (0.948–0.959)	100	83.1
β-thalassemia (n = 522)	HbA_2	> 3.75	0.984 (0.980–0.987)	97.9	98.5
	MCV	≤ 76.2	0.953 (0.948–0.958)	96.0	83.0
	MCH	≤ 24.4	0.937 (0.931–0.943)	95.4	82.2
$\alpha\beta$-Compound (n = 60)	HbA_2	> 3.9	0.964 (0.959–0.968)	98.3	91.9
	MCV	≤ 79.6	0.859 (0.851–0.867)	96.7	73.4
	MCH	≤ 25.0	0.842 (0.833–0.851)	95.0	75.5

Note: Bold values indicate the highest AUC within each diagnostic group. AUC, area under the curve; CI, confidence interval.

For the most clinically severe phenotype, Hb H disease, MCH demonstrated exceptional diagnostic accuracy with an AUC of 0.954 and 100% sensitivity at a cut-off of ≤ 22.0 pg. This robust sensitivity ensures that high-risk genotypes are effectively captured, directly addressing the necessity for reliable screening markers to prevent severe birth defects in high-prevalence regions. For minor α -thalassemia, MCV provided the highest diagnostic yield (AUC 0.903; 95% CI: 0.895–0.910).

Discussion and Conclusion

Divergence from Mediterranean Models: Regional Genetic Specificity

While established guidelines from Mediterranean regions, such as those provided by the Italian Society for Thalassemia and Hemoglobinopathies (SITE)^{3,4,11}, offer a foundational framework for β -thalassemia management, they frequently underrepresent the unique mutational pressures found in the South China "Thalassemia Belt"⁷. Our cohort's molecular landscape is dominated by the $--^{SEA}$ deletion (54.5% of α -cases) and $\beta^{CD41-42}$ (30.5% of β -cases), reflecting a distinct regional signature. Unlike Mediterranean populations where β -thalassemia is the primary concern, the high prevalence of large-fragment α -globin deletions in Shenzhen creates a high risk for Hb H disease and Hb Bart's hydrops fetalis. Consequently, applying generic Mediterranean cut-offs to this population is not only suboptimal but potentially misses high-risk carriers⁸.

Quantifying "Phenotypic Masking": Resolving the Diagnostic Paradox

A significant contribution of this study is the rigorous quantification of "phenotypic masking" in $\alpha\beta$ -compound heterozygotes. It is widely recognized in clinical hematology that red cell indices alone may be insufficient for a definitive risk assessment of complex thalassemia genotypes⁸. Our data supports this: $\alpha\beta$ -compound carriers exhibit significantly higher MCH levels compared to pure β -thalassemia carriers (22.00 ± 2.11 vs. 20.60 ± 1.96 pg). This partial restoration of globin chain balance effectively "cloaks" the underlying α -defect¹². By integrating HbA_2 (which remains elevated at 5.37 ± 0.63) with these "normalized" indices, our 3D topographical model provides a visual and statistical mechanism to identify these complex genotypes that would otherwise bypass traditional 2D screening filters.

HbA2 Reliability and "Precision Trigger" Strategy

The non-standardization of HbA_2 across different analytical platforms is a recognized challenge in hemoglobinopathy screening⁸. However, within the standardized environment of the Helena V8 capillary electrophoresis system, our results demonstrate that HbA_2 remains the gold standard for β -globin defects (AUC 0.984)⁹. We propose that our established cut-offs (e.g., $HbA_2 > 3.75\%$ for β -thalassemia) should be viewed not as universal reference intervals, but as "Precision Triggers" tailored to the local

genetic background^{8,9}. This approach mitigates the risk of "misleading" results by ensuring that molecular testing is triggered even in cases with borderline microcytosis.

Overcoming the "Blind Spot" of Silent α -Thalassemia

Silent α -thalassemia carriers ($-a/aa$) remain the most elusive targets in primary screening, with MCV and MCH AUCs falling below 0.70⁸. The extensive density overlap between these carriers and healthy controls in our 3D model underscores the systemic vulnerability of index-only screening^{10,13}. To address this, we advocate for a lowered threshold for molecular intervention (MCH < 28.5 pg) in high-prevalence urban centers. This reflex testing strategy is essential for preventing the conception of children with Hb H disease, particularly in high-mobility populations like Shenzhen where pedigree analysis may be incomplete⁷.

Conclusion: Toward a Precision Screening Framework

In conclusion, this study establishes a precision screening framework tailored to the unique mutational landscape of South China, moving beyond the limitations of generic international guidelines. This reflex testing strategy is essential for preventing the conception of children with Hb H disease, particularly in high-mobility populations like Shenzhen where pedigree analysis may be incomplete.

Our proposed regional protocol utilizes: $HbA_2 > 3.75\%$ serves as the primary trigger for β -thalassemia. MCH < 28.5 pg combined with $HbA_2 < 2.77\%$ as a mandatory threshold for reflex α -molecular testing to capture silent carriers.

This multidimensional approach provides a rigorous scientific basis for primary medical institutions to minimize occult carrier transmission and ensure the safety of prenatal counseling in high-mobility urban environments.

Abbreviations

- CS
Constant Spring
- QS
Quong Sze
- WS
Westmead
- Hb
Hemoglobin
- n
number of cases.

Declarations

Funding

This work was supported by the Clinical Research Special Project of the Medical and Health Technology Development Research Center, National Health Commission (Grant No. WKZX2024DN0181), the Guangdong Science and Technology Foundation (Grant No. 2023B0101200003), and the Shenzhen Science and Technology Foundation (Grant No. KJZD20230923115359001).

Ethics approval and consent to participate

The study protocol was in line with the Declaration of Helsinki (as revised in Brazil 2013). This study was approved by the Ethics Committee of Shenzhen Second People's Hospital (Approval No: 2025-765-01PJ) and all participants were informed and signed a written informed consent.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Availability of data and materials

Data is provided within the supplementary information files.

Acknowledgements

not applicable.

Authors' contributions

D.G. designed and supervised the overall study, interpreted data, and performed final revision of the manuscript. **H.Q.** conceptualized the project, performed experiments, and drafted the initial manuscript. **W.L.** and **Y.W.** collected and analyzed clinical samples. All authors reviewed and approved the final version of the manuscript.

References

1. Weatherall DJ (2010) The inherited diseases of hemoglobin are an emerging global health burden. *Blood* 115:4331–4336. 10.1182/blood-2010-01-251348
2. Taher AT, Musallam KM, Cappellini M (2021) D. beta-Thalassemias. *N Engl J Med* 384:727–743. 10.1056/NEJMra2021838
3. Pinto VM et al (2025) Thalassemias and Sickle Cell Diseases in Pregnancy: SITE Good Practice. *J Clin Med* 14. 10.3390/jcm14030948

4. Casale M et al (2022) Good Clinical Practice of the Italian Society of Thalassemia and Haemoglobinopathies (SITE) for the Management of Endocrine Complications in Patients with Haemoglobinopathies. *J Clin Med* 11. 10.3390/jcm11071826
5. Lai K, Huang G, Su L, He Y (2017) The prevalence of thalassemia in mainland China: evidence from epidemiological surveys. *Sci Rep* 7:920. 10.1038/s41598-017-00967-2
6. Xu XM et al (2004) The prevalence and spectrum of alpha and beta thalassaemia in Guangdong Province: implications for the future health burden and population screening. *J Clin Pathol* 57:517–522. 10.1136/jcp.2003.014456
7. Li Y et al (2025) Unveiling the molecular landscape of delta-thalassemia and delta-globin variants in southern China: novel mutations, gene spectrum, and implications for thalassemia diagnosis. *Front Genet* 16:1584310. 10.3389/fgene.2025.1584310
8. Thilakarathne S, Jayaweera UP, Premawardhena A (2024) Unresolved laboratory issues of the heterozygous state of beta-thalassemia: a literature review. *Haematologica* 109:23–32. 10.3324/haematol.2022.282667
9. Qian H et al (2025) Analysis of thalassemia genotypes and HbA2 test results in pregnant women in Shenzhen, China. *Sci Rep* 15:7483. 10.1038/s41598-025-91908-x
10. Huang J et al (2023) Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nat Biomed Eng* 7:797–810. 10.1038/s41551-022-00991-2
11. in *Guidelines for the Management of Transfusion-Dependent beta-Thalassaemia (TDT)* (eds A. T. Taher et al.) (2025)
12. Chetta M et al (2025) The Masked Thalassemia: A Rare Case of a Patient with Normal HbA2 Levels, beta-Thalassemia Pathogenic Variant (CD39 C > T), and a Novel delta-Globin Gene Deletion. *Appl Clin Genet* 18:233–241. 10.2147/TACG.S544633
13. Xie S, Liu S, Gao Y, Tang J, Cao L (2022) Application of Hemostatic Devices in Laparoscopic Hepatectomy. *J Vis Exp*. 10.3791/63368

Figures

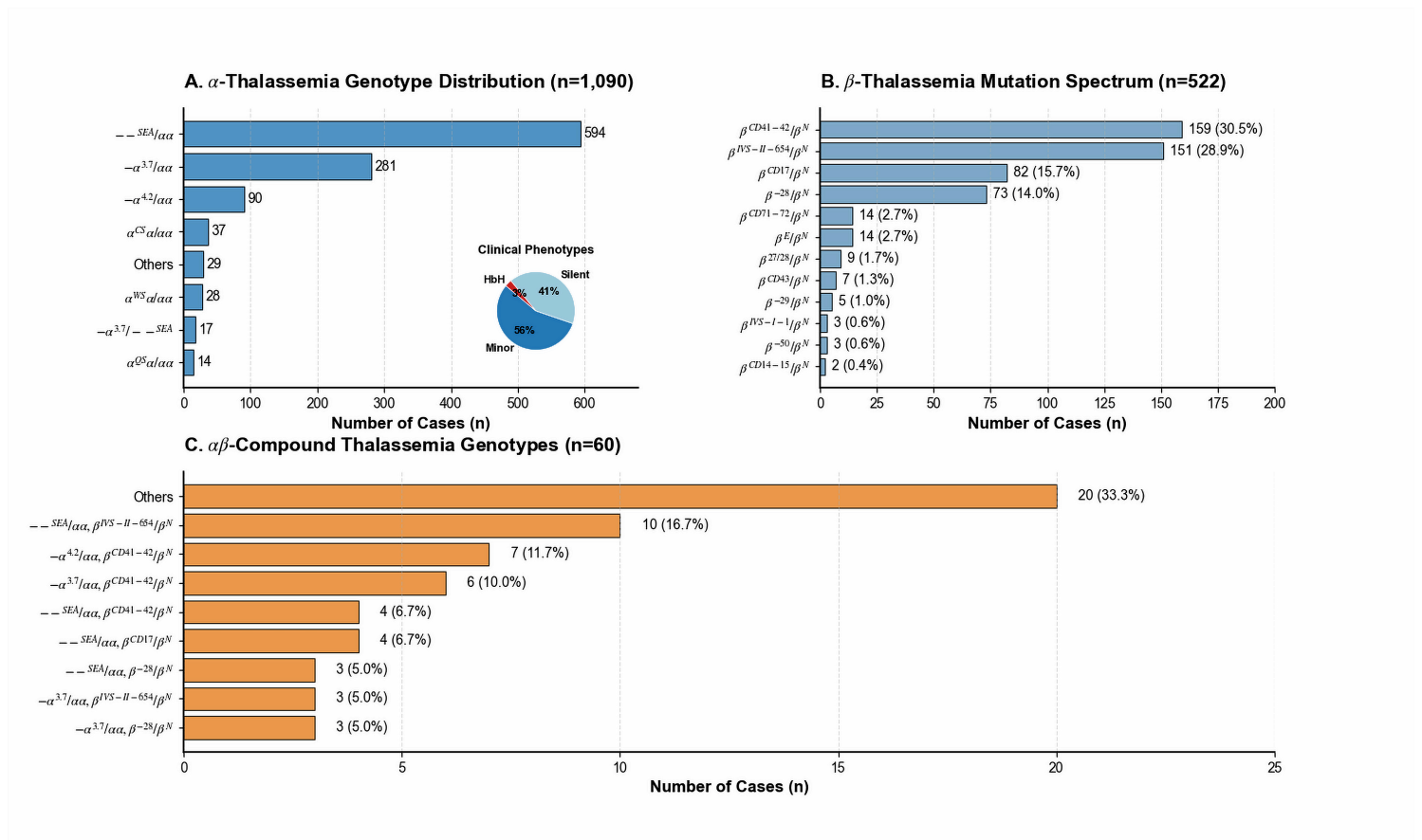


Figure 1

Molecular spectrum of thalassemia carriers. A: Genotype distribution of α -thalassemia; B: Mutation spectrum of β -thalassemia; C: Genotype combinations of $\alpha\beta$ -compound thalassemia.

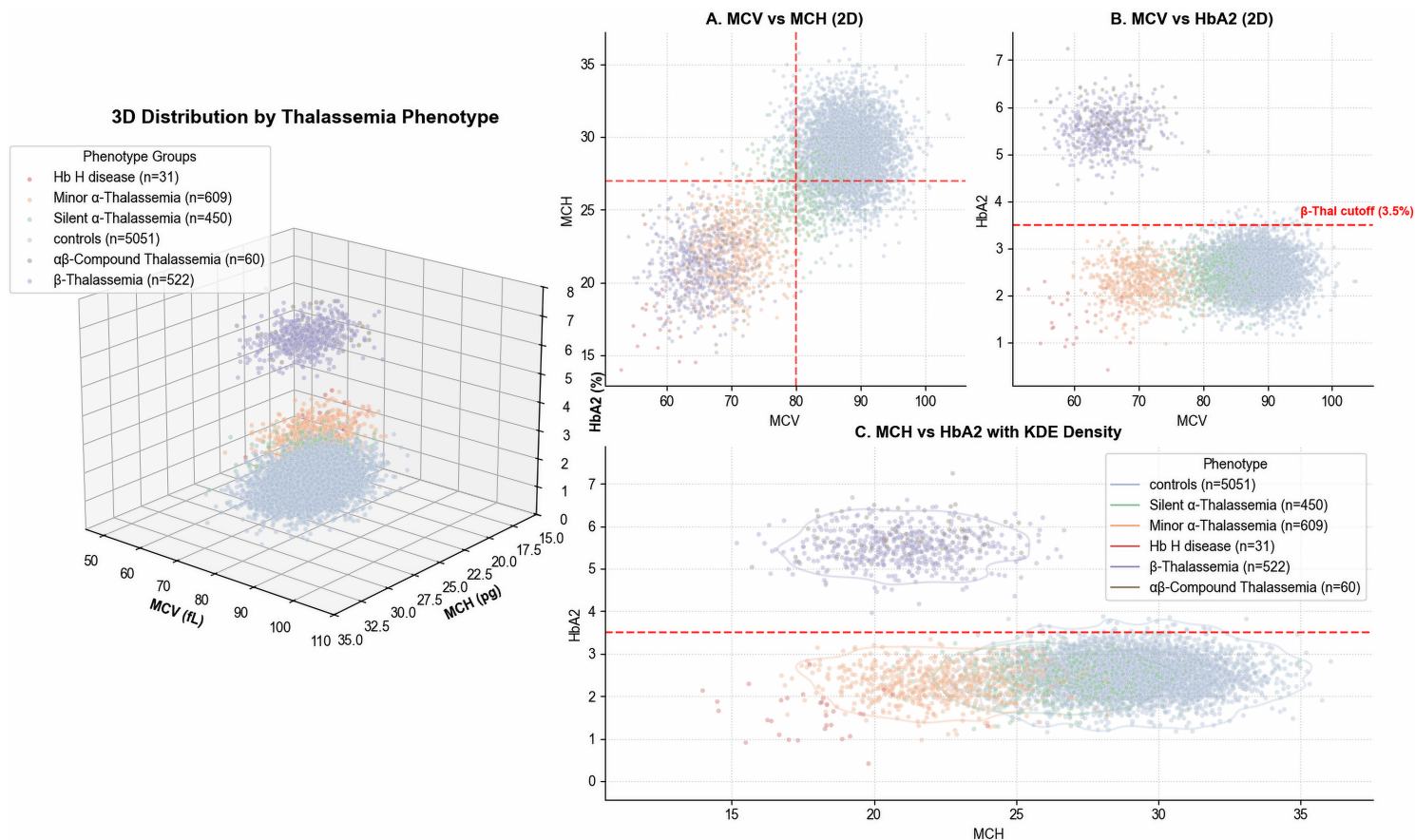


Figure 2

Multiparametric 3D and 2D Spatial Distribution of Hematological Indices Across Thalassemia

Genotypes. The integration of Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), and Hemoglobin A₂ (HbA_2) demonstrates the distinct phenotypic landscape of controls (n=5,051) versus five thalassemia subgroups. (1): Global 3D Topographical View (Left): Illustrates the robust spatial separation of β -thalassemia (purple) and $\alpha\beta$ -compound heterozygotes (brown), driven primarily by HbA_2 elevation. The α -thalassemia spectrum (green/orange/red) demonstrates a progressive downward trajectory in the MCV-MCH plane. (2): Panel A (MCV vs. MCH 2D Projection): Highlights the varying degrees of microcytosis and hypochromia. The dashed red lines (MCV 80 fL, MCH 27 pg) represent conventional screening thresholds, showing significant control overlap with silent α -carriers (n= 450). (3): Panel B (MCV vs. HbA_2 2D Projection): Demonstrates the absolute diagnostic power of HbA_2 in segregating β -globin defects. The red dashed line denotes the 3.5% β -thalassemia screening cutoff. (4): Panel C (MCH vs. HbA_2 with KDE Density Contours): Kernel Density Estimation (KDE) contours reveal the high-density centers of each phenotype. Note the extensive density overlap between controls and silent α -carriers, contrasting with the discrete, isolated cluster of β -thalassemia (n=522) and the extreme deviation of Hb H disease (n=31).

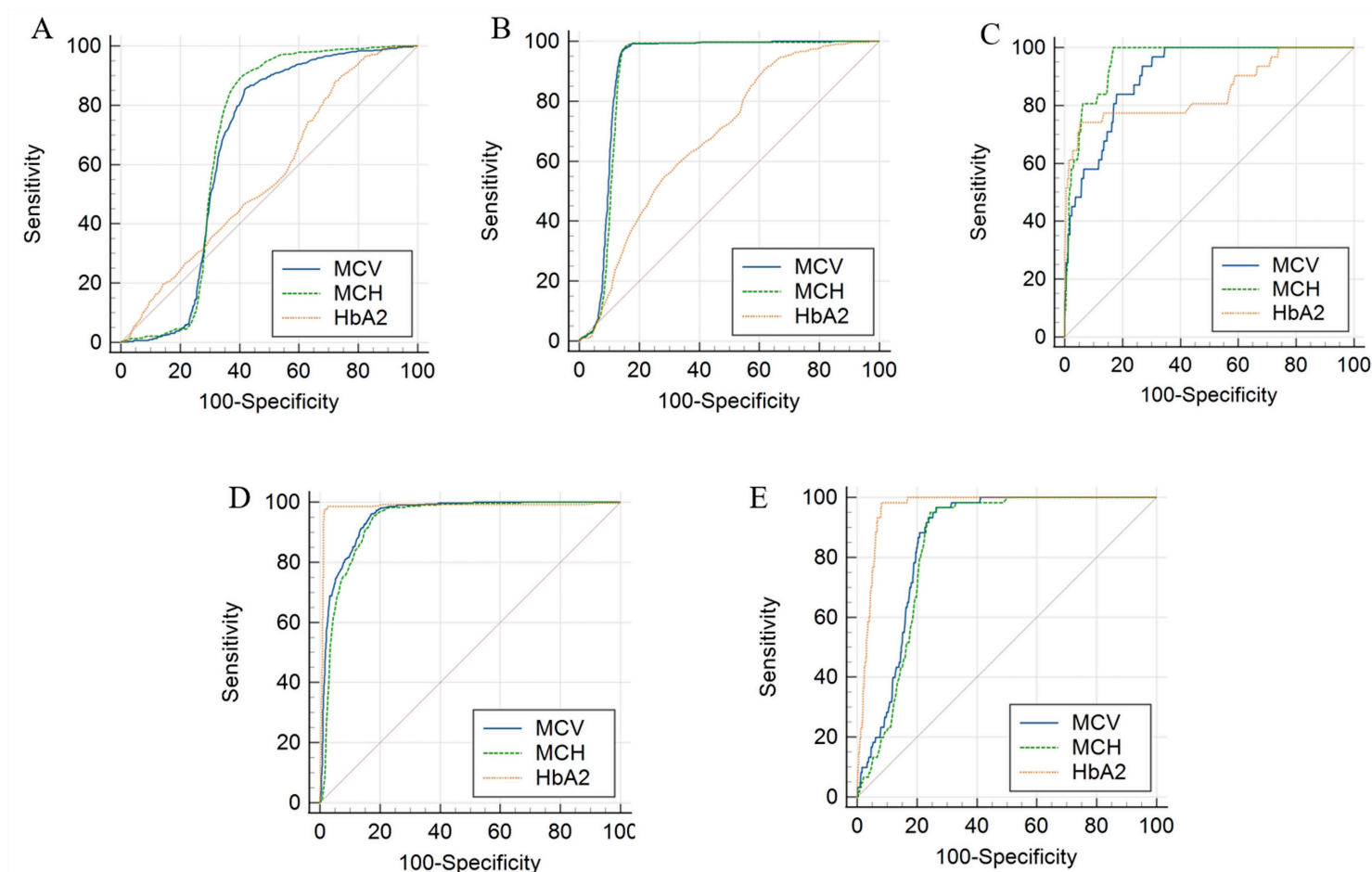


Figure 3

Comparative diagnostic efficacy of MCV, MCH, and HbA_2 across thalassemia genotypes. Overlaid ROC curves comparing the discriminatory power of the three hematological markers for each clinical subtype. The curves represent MCV (blue solid line), MCH (green dashed line), and HbA_2 (orange dotted line). (A) Silent α -thalassemia: All markers showed suboptimal performance ($AUC < 0.70$). (B) Minor α -thalassemia: MCV and MCH significantly outperformed HbA_2 . (C) Hb H disease: MCH demonstrated the highest diagnostic accuracy ($AUC = 0.954$), followed closely by MCV. (D) β -thalassemia: HbA_2 proved to be the superior marker ($AUC = 0.984$). (E) $\alpha\beta$ -compound thalassemia: HbA_2 maintained high diagnostic efficiency, outperforming red cell indices. The diagonal line represents the line of no discrimination ($AUC = 0.5$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.tif](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)

- [FigureS1S3legends.docx](#)
- [TableS1S2.docx](#)
- [rawdate.xlsx](#)
- [cofeforfigure1.docx](#)
- [codeforfigure2.docx](#)