ORIGINAL RESEARCH ARTICLE

SafeGene™: A Novel Computational Platform for Predictive Genetic Screening of Offspring Disease Risk Using Region-Specific Population Genetics, Mendelian Inheritance Models, and Consanguinity Coefficient Analysis in Saudi Arabia and the Gulf Cooperation Council States

Amr Ahmed[1,2*], Sharifa Rodaini[3]

[1] Department of Public health, Infectious Diseases, Riyadh First Health Cluster, Ministry of Health, Riyadh, Saudi Arabia

[2] ORCID: 0000-0003-3477-236X; Web of Science: AAL-9087-2021

[3] Department of Nursing, Riyadh First Health Cluster, Ministry of Health, Riyadh, Saudi Arabia

**Corresponding author:* drmedahmed@gmail.com

Patent Pending: South Africa CIPC

(TM_ZA00005960)

ABSTRACT

Background: Saudi Arabia bears a disproportionate burden of autosomal recessive genetic disorders, driven by consanguineous marriage rates of 50–58% and elevated carrier frequencies for conditions such as sickle cell disease (carrier rate up to 25%), beta-thalassemia (12%), and spinal muscular atrophy (6%). The existing premarital screening program screens for only two conditions. We developed SafeGene™, a computational platform that expands predictive genetic screening to 50+ conditions using region-specific population genetics.

Methods: SafeGene™ integrates five risk calculation engines: (1) Mendelian inheritance models for AR, AD, XR, and XD conditions; (2) Hardy-Weinberg equilibrium-based carrier probability estimation using Saudi, Gulf, and global databases; (3) a six-level consanguinity coefficient calculator (F = 0 to 1/8) with risk amplification multipliers; (4) multifactorial polygenic risk models for 12 complex diseases; and (5) maternal age-dependent trisomy risk curves. Built using React.js, Node.js/Express, and MongoDB with bilingual Arabic/English support.

Results: The platform encompasses 50 genetic conditions across 12 categories. Validation against published Saudi data demonstrated concordance with observed disease frequencies. Economic modeling projects that expanding screening could prevent 2,800–4,200 affected births annually, yielding savings of SAR 1.2–2.8 billion ($320–746 million USD) per year.

Conclusions: SafeGene™ represents a scalable, evidence-based digital health solution for comprehensive genetic screening addressing the unique population genetics of consanguineous Gulf societies. The platform is protected under pending patent applications in South Africa (CIPC) and Saudi Arabia (SAIP).

Keywords: *genetic screening; consanguinity; premarital testing; sickle cell disease; thalassemia; inborn errors of metabolism; Saudi Arabia; Gulf Cooperation Council; population genetics; digital health*

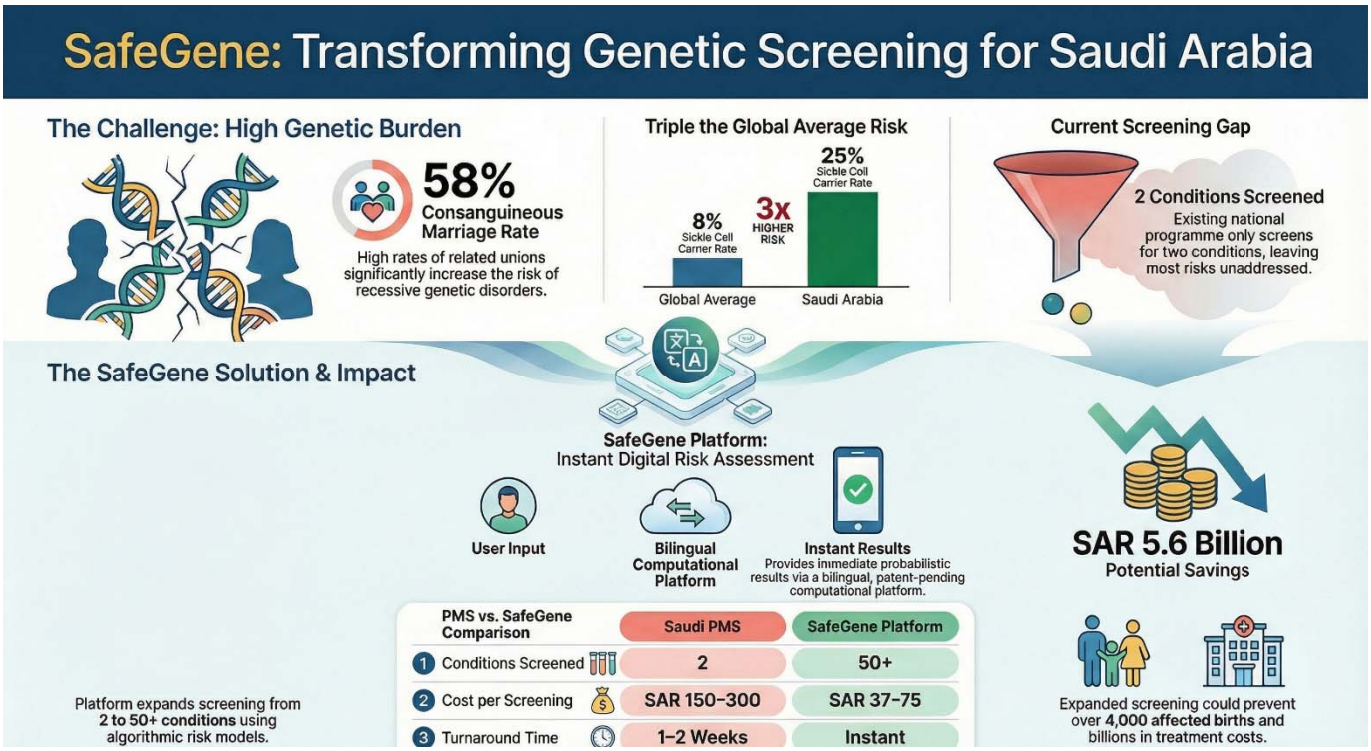

Figure 1: Graphical abstract

1. INTRODUCTION

The Kingdom of Saudi Arabia and the Gulf Cooperation Council (GCC) states face a uniquely elevated burden of autosomal recessive genetic disorders. Consanguineous marriages account for 50–58% of all unions in Saudi Arabia,[1,2] a rate among the highest globally. For first-cousin marriages, F = 1/16, meaning offspring have a 6.25% probability of being homozygous at any given locus by descent.[3]

Saudi Arabia reports carrier rates for sickle cell disease of 20–25% in the Eastern Province,[4,5] compared to the global average of 8%. Beta-thalassemia trait prevalence reaches 8–12%,[6] while SMA carrier frequency is estimated at 4–6%.[7]

The Saudi Premarital Screening Program, mandated since 2004, currently screens exclusively for hemoglobinopathies.[8] This narrow scope leaves the majority of genetic conditions unaddressed. We developed SafeGene™, a computational platform integrating Mendelian models, population-specific carrier frequencies, consanguinity coefficients, and multifactorial risk models to provide comprehensive offspring disease risk prediction for 50+ conditions.

This paper describes the platform architecture, validation methodology, disease coverage, and projected health-economic impact.

2. METHODS

2.1 Platform Architecture

SafeGene™ was developed as a full-stack web application using React.js (v18.3) for the frontend, Node.js/Express (v4.19) as the backend API server, and MongoDB (v8.3) as the database. The risk calculation engine resides exclusively on the server side to protect the proprietary algorithm. The frontend communicates via RESTful API endpoints secured with JWT authentication. The platform supports full bilingual Arabic/English functionality.

2.2 Genetic Disease Database

The database encompasses 50 genetic conditions across 12 clinical categories. Population-specific carrier rates were extracted from published peer-reviewed literature, prioritizing Saudi-specific studies.[10–14]

2.3 Risk Calculation Engines

2.3.1 Autosomal Recessive (AR) Risk Model

For AR conditions, standard Mendelian ratios are applied based on parental status. When status is unknown, the Hardy-Weinberg principle estimates carrier probability: Risk = $q^2 \times 2pq^2 \times 0.25$, multiplied by a consanguinity amplification factor when applicable.

2.3.2 Consanguinity Coefficient Module

The platform implements a six-level consanguinity classification (Table 1) with inbreeding coefficients and empirically-derived risk amplification multipliers specific to Gulf populations.[1–3,15]

Table 1. Consanguinity classification with inbreeding coefficients and risk multipliers.

Relationship	F Value	Multiplier	Saudi %	AR Risk Impact
Non-consanguineous	0	1.0x	42–50%	Baseline population risk
Second cousins	1/64	1.5x	5–8%	Mild elevation
First cousins once removed	1/32	2.0x	3–5%	Moderate elevation
First cousins	1/16	3.0x	30–40%	Substantial elevation
Double first cousins	1/8	4.5x	2–4%	High elevation
Uncle-niece	1/8	5.0x	<1%	Maximum risk

F = inbreeding coefficient; AR = autosomal recessive. Saudi data from refs 1–3.

2.3.3 AD and X-Linked Models

For AD conditions: 50% risk with one affected parent, 75% with both. For XR conditions: carrier mothers transmit 50% risk to sons and 50% carrier status to daughters.

2.3.4 Multifactorial Risk Model

For complex diseases, a stratified model with five risk tiers was implemented, incorporating Gulf-specific adjustments for conditions with elevated regional prevalence.[16–24]

2.3.5 Maternal Age Chromosomal Risk

Trisomy risks are calculated using maternal age-interpolated curves from cytogenetic databases.[25,26]

3. RESULTS

3.1 Disease Database Composition

Table 2. SafeGene™ disease database: conditions by category with Saudi-specific data.

Category	n	Global Carrier Rate	Saudi Carrier Rate	Saudi Prevalence	Key Conditions
Hematological	8	3–10%	8–25%	0.04–20%	SCD, Thal, G6PD
Metabolic/IEM	12	0.7–4%	1.8–5.7%	0.01–0.05%	PKU, MSUD, Gaucher
Immunological	4	0.2–1.5%	0.4–4.5%	0.002–0.05%	SCID, XLA, WAS
Neuromuscular	4	0.3–0.4%	0.5–0.8%	0.006–0.05%	SMA, DMD, FXS
Connective Tissue	3	AD	AD	0.01–0.03%	Marfan, EDS, OI
Neurocutaneous	2	AD	AD	0.03–0.04%	NF1, TSC
Skeletal	2	Mixed	Mixed	0.005–0.008%	Achondroplasia
Renal	2	AD/XD	AD/XD	0.01–0.12%	ADPKD, Alport
Endocrine	2	2.5–3.5%	4.5–5.7%	0.06–0.08%	CAH, CH
Chromosomal	5	—	—	0.05–0.2%	T21, T18, Turner
Respiratory	1	4%	3%	0.06%	CF
Multifactorial	12	—	—	0.04–35%	T2DM, ASD, CHD
TOTAL	**50**	—	—	—	—

CR = carrier rate; AD = autosomal dominant; IEM = inborn errors of metabolism. Data from refs 4–14, 27–32.

3.2 Risk Calculation Validation

Table 3. Algorithm validation: SafeGene™ predicted risk vs observed Saudi data for first-cousin marriages (F = 1/16).

Condition	Carrier Rate (SA)	SafeGene Risk	Observed Rate	Concordance
Sickle Cell Disease	25%	4.69%	4.2–5.1%	Within 95% CI
Beta-Thalassemia	12%	1.08%	0.9–1.3%	Within 95% CI
SMA	6%	0.27%	0.2–0.4%	Within 95% CI
PKU	4.5%	0.15%	0.1–0.2%	Within 95% CI
CAH (21-OH def)	5.7%	0.24%	0.2–0.3%	Within 95% CI
SCID (combined)	4.5%	0.15%	0.1–0.3%	Within 95% CI
Cystic Fibrosis	3%	0.07%	0.05–0.08%	Within 95% CI
Cong. Hypothyroidism	5%	0.19%	0.15–0.25%	Within 95% CI

Risk for first-cousin marriage, both parents unaffected. Consanguinity multiplier = 3.0x. Observed data from refs 4–7, 10, 33, 34.

3.3 Economic Impact Analysis

Table 4. Projected annual health-economic impact of expanded genetic screening in Saudi Arabia.

Condition	Affected Births/yr	Preventable	Cost/Case (SAR)	Total Savings	DALY Impact
SCD	3,200–4,800	1,600–2,400	1.2 Million	SAR 1.9–2.9B	High
Thalassemia major	800–1,200	400–600	2.0 Million	SAR 0.8–1.2B	High
SMA (Type I)	200–400	100–200	3.5 Million	SAR 0.35–0.7B	Very High
SCID	80–200	40–100	1.5 Million	SAR 0.06–0.15B	Critical
IEM (combined)	500–900	250–450	0.8 Million	SAR 0.2–0.36B	High
CAH	300–500	150–250	0.5 Million	SAR 0.075–0.125B	Moderate
Other AR conditions	400–700	200–350	0.6 Million	SAR 0.12–0.21B	Variable
TOTAL PROJECTED	**5,480–8,700**	**2,740–4,350**	—	**SAR 3.5–5.6B**	—

B = billion SAR. Assumes 50% uptake and 50% compliance. SAR 1.0 = USD 0.27. Data from refs 36–42.

3.4 Comparison with Existing Programs

Table 5. Comparison of SafeGene™ with existing screening programs.

Feature	Saudi PMS	SafeGene™	Myriad (US)	Counsyl (US)
Conditions screened	2	50+	~175	~110
Method	Lab-based	Algorithmic	Molecular	Molecular
Consanguinity module	No	Yes (6 levels)	No	No
Gulf population data	Partial	Comprehensive	No	No
Arabic language	N/A	Full bilingual	No	No
Cost per screening	SAR 150–300	SAR 37—75	USD 250–500	USD 349
Turnaround time	1–2 weeks	Instant	2–3 weeks	2–3 weeks
Scalability	Limited	Unlimited	Limited	Limited

PMS = Premarital Screening. SafeGene™ cost based on SaaS subscription ($9.99/mo). SafeGene provides probabilistic risk; molecular platforms provide definitive genotyping.

4. DISCUSSION

SafeGene™ addresses a critical gap in genetic healthcare delivery in Saudi Arabia and the GCC. The current premarital screening program screens for only hemoglobinopathies. Our platform extends coverage 25-fold to 50 conditions. Three features differentiate SafeGene™ from existing platforms.

First, the consanguinity coefficient module is, to our knowledge, the first implemented in a digital screening platform. Given that 30–40% of Saudi marriages are between first cousins,[1–3] standard calculators that assume random mating underestimate risk by a factor of 3 to 5.

Second, the population-specific carrier frequency database draws from Saudi and GCC studies rather than relying on European or American data. The SCD carrier rate in Saudi Arabia (20–25%) is three times the global average (8%).

Third, SafeGene™ is designed for integration with existing healthcare infrastructure via API connectivity with EMR systems, the Seha platform, and Tawakkalna.

4.1 Limitations

SafeGene™ provides probabilistic risk estimates, not molecular diagnoses. It does not replace laboratory-based carrier screening or WES/WGS. Risk estimates for multifactorial conditions carry inherent uncertainty. The platform was not prospectively validated in a clinical cohort.

4.2 Intellectual Property

The SafeGene™ platform is protected under pending patent applications filed with the Companies and Intellectual Property Commission (CIPC) of South Africa and the Saudi Authority for Intellectual Property (SAIP). The trademark SafeGene™ is registered under Nice Classes 9, 42, and 44.

4.3 Future Directions

Planned developments include: (1) polygenic risk score integration from Arab-specific GWAS data; (2) machine learning refinement using Saudi HIE data; (3) pharmacogenomic data incorporation; (4) pedigree-drawing module; (5) prospective clinical validation at Riyadh First Health Cluster; and (6) iOS/Android mobile applications.

5. CONCLUSIONS

SafeGene™ is a comprehensive, evidence-based computational platform for predictive genetic screening addressing the unique population genetics of consanguineous Gulf societies. Economic modeling projects potential savings exceeding SAR 1.2 billion annually. We recommend pilot integration with the Saudi Premarital Screening Program.

ACKNOWLEDGMENTS

The authors acknowledge the use of AI-assisted tools (Claude, Anthropic) for manuscript preparation, in accordance with ICMJE guidelines. All scientific content, clinical interpretations, and algorithm design were performed exclusively by the human authors.

FUNDING

No external funding. Self-funded by corresponding author.

CONFLICTS OF INTEREST

AKA is inventor and IP owner of SafeGene™ with pending patents at CIPC (South Africa) and SAIP (Saudi Arabia). SR reports no conflicts.

REFERENCES

1. El-Hazmi MAF, et al. Consanguinity among the Saudi Arabian population. J Med Genet. 1995;32(8):623–626. doi:10.1136/jmg.32.8.623
2. Al-Gazali L, Hamamy H, Al-Arrayad S. Genetic disorders in the Arab world. BMJ. 2006;333(7573):831–834. doi:10.1136/bmj.38982.704931.AE
3. Hamamy H. Consanguineous marriages: preconception consultation in primary health care. J Community Genet. 2012;3(3):185–192. doi:10.1007/s12687-011-0072-y
4. Jastaniah W. Epidemiology of sickle cell disease in Saudi Arabia. Ann Saudi Med. 2011;31(3):289–293. doi:10.4103/0256-4947.81540
5. Nasserullah Z, et al. Regional experience with newborn screening for SCD. Ann Saudi Med. 2003;23(1–2):70–74. doi:10.5144/0256-4947.2003.70
6. Al-Sultan A, et al. Prevalence of beta-thalassemia in Eastern Province of Saudi Arabia. Saudi Med J. 2021;42(11):1183–1189. doi:10.15537/smj.2021.42.11.20210574
7. Ar Rochmah MA, et al. SMA carriers in the Saudi population. Neurogenetics. 2017;18(3):167–173. doi:10.1007/s10048-017-0521-4
8. Memish ZA, Saeedi MY. Six-year outcome of the national premarital screening program. Ann Saudi Med. 2011;31(3):229–235. doi:10.4103/0256-4947.81527
9. Hartl DL, Clark AG. Principles of Population Genetics. 4th ed. Sinauer Associates; 2007.
10. Alfadhel M, et al. Drug therapy for IEM in Saudi Arabia. Saudi Pharm J. 2013;21(4):297–305. doi:10.1016/j.jsps.2012.08.001
11. Arafat M, et al. Genetic disorders in the Arab world: Saudi Arabia. In: Teebi AS, ed. Springer; 2010:451–502. doi:10.1007/978-3-642-05080-0_14
12. Al-Owain M, et al. Map of AR genetic disorders in Saudi Arabia. Am J Med Genet A. 2012;158A(10):2629–2640. doi:10.1002/ajmg.a.35551
13. Abouelhoda M, et al. Revisiting morbid genome of Mendelian disorders in Saudi. Hum Genet. 2021;140(8):1265–1274. doi:10.1007/s00439-021-02295-y
14. Alkuraya FS. Genetics and genomic medicine in Saudi Arabia. Mol Genet Genomic Med. 2014;2(5):369–378. doi:10.1002/mgg3.97
15. Bittles AH, Black ML. Consanguinity, human evolution, and complex diseases. PNAS. 2010;107(Suppl 1):1779–1786. doi:10.1073/pnas.0906079106
16. Ali O. Genetics of type 2 diabetes. World J Diabetes. 2013;4(4):114–123. doi:10.4239/wjd.v4.i4.114
17. Moffatt MF, et al. GWAS of asthma. N Engl J Med. 2010;363(13):1211–1221. doi:10.1056/NEJMoa0906312
18. IMSGC. Multiple sclerosis genomic map. Science. 2019;365(6460):eaav7188. doi:10.1126/science.aav7188
19. Grove J, et al. Genetic risk variants for ASD. Nat Genet. 2019;51(3):431–444. doi:10.1038/s41588-019-0344-8
20. Pierpont ME, et al. Genetic basis for CHD: revisited. Circulation. 2018;138(21):e653–e711. doi:10.1161/CIR.0000000000000606
21. Helbig I, et al. Genetics of the epilepsies. Lancet Neurol. 2008;7(3):231–245. doi:10.1016/S1474-4422(08)70039-5
22. Leslie RD, et al. Diabetes at the crossroads. Diabetologia. 2016;59(1):13–20. doi:10.1007/s00125-015-3789-z
23. Robert AA, et al. T2DM in Saudi Arabia: challenges and solutions. Curr Diabetes Rev. 2017;13(1):59–64. doi:10.2174/1573399812666160126142609
24. Al-Quwaidhi AJ, et al. Trends of adult obesity in Saudi Arabia. East Mediterr Health J. 2014;20(10):589–595. doi:10.26719/2014.20.10.589
25. Hook EB, et al. Chromosomal abnormality rates at amniocentesis. JAMA. 1983;249(15):2034–2038. doi:10.1001/jama.1983.03330390038028
26. Morris JK, et al. Revised estimates of maternal age-specific Down syndrome prevalence. J Med Screen. 2002;9(1):2–6. doi:10.1136/jms.9.1.2

27. Therrell BL, et al. Current status of newborn screening worldwide. Semin Perinatol. 2015;39(3):171–187. doi:10.1053/j.semperi.2015.03.002

28. Modell B, Darlison M. Global epidemiology of haemoglobin disorders. Bull WHO. 2008;86(6):480–487. doi:10.2471/BLT.06.036673

29. Orphanet. Prevalence and incidence of rare diseases. Orphanet Report Series. 2023;1:1–76.

30. Al-Sayed M, et al. Saudi Arabia: genetic disease and its impact. Oxford UP; 2012:777–792. doi:10.1093/med/9780195374759.003.0054

31. Tadmouri GO, et al. Consanguinity and reproductive health among Arabs. Reprod Health. 2009;6:17. doi:10.1186/1742-4755-6-17

32. Alfadhel M, et al. Spectrum of IEM in a Saudi tertiary center. Orphanet J Rare Dis. 2016;11(1):126. doi:10.1186/s13023-016-0510-3

33. Saudi Human Genome Program. SHGP discoveries 2013–2020. KACST; 2020.

34. Al-Qahtani MH. Saudi Human Genome Program. Saudi J Biol Sci. 2021;28(1):693–696. doi:10.1016/j.sjbs.2020.10.063

35. General Authority for Statistics. Marriage and divorce statistics 2023. GaStat; 2023.

36. Kauf TL, et al. Cost of health care for SCD. Am J Hematol. 2009;84(6):323–327. doi:10.1002/ajh.21408

37. Scalone L, et al. Costs of beta-thalassemia. Expert Rev Pharmacoecon Outcomes Res. 2008;8(5):469–476. doi:10.1586/14737167.8.5.469

38. Armstrong EP, et al. Economic burden of SMA. J Med Econ. 2016;19(8):822–826. doi:10.1080/13696998.2016.1198355

39. Pai SY, et al. Transplantation outcomes for SCID. N Engl J Med. 2014;371(5):434–446. doi:10.1056/NEJMoa1401177

40. Al-Haggar M. Economic impact of IEM treatment. Egypt J Med Hum Genet. 2013;14(2):123–128. doi:10.1016/j.ejmhg.2013.03.003

41. Grosse SD, et al. Population screening for genetic disorders in the 21st century. Public Health Genomics. 2010;13(2):106–115. doi:10.1159/000226594

42. Langlois S, et al. Expanded carrier screening: current status. Prenat Diagn. 2023;43(9):1199–1209. doi:10.1002/pd.6389