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

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

[1] Deparent of Internal Medicine, Infectious Diseases & Clinical Pharmacology, Riyadh First Health Cluster, Ministry of Health, Riyadh, Saudi Arabia

[2] ORCID: 000000033477236X; Web of Science: AAL90872021

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

**Corresponding author: Dr. Amr Kamel Khalil Ahmed, MD, MSc*

Patent Pending: South Africa

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%), betathalassemia (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 regionspecific population genetics.

Methods: SafeGene integrates five risk calculation engines: (1) Mendelian inheritance models for AR, AD, XR, and XD conditions; (2) HardyWeinberg equilibriumbased carrier probability estimation using Saudi, Gulf, and global databases; (3) a sixlevel consanguinity coefficient calculator (F = 0 to 1/8) with risk amplification multipliers; (4) multifactorial polygenic risk models for 12 complex diseases; and (5) maternal agedependent 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, evidencebased 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*

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%. Betathalassemia 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, populationspecific 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 healtheconomic impact.

2. METHODS

2.1 Platform Architecture

SafeGene was developed as a fullstack 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. Populationspecific carrier rates were extracted from published peerreviewed literature, prioritizing Saudispecific 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 HardyWeinberg 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 sixlevel consanguinity classification (Table 1) with inbreeding coefficients and empiricallyderived 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
Nonconsanguineous	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
Uncleniece	1/8	5.0x	<1%	Maximum risk

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

2.3.3 AD and XLinked 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 Gulfspecific adjusents for conditions with elevated regional prevalence.[16–24]

2.3.5 Maternal Age Chromosomal Risk

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

3. RESULTS

3.1 Disease Database Composition

Table 2. SafeGene disease database: conditions by category with Saudispecific 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 firstcousin 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
BetaThalassemia	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 (21OH 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 firstcousin 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 healtheconomic 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	Labbased	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 25fold 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 populationspecific 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 laboratorybased 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 Arabspecific GWAS data; (2) machine learning refinement using Saudi HIE data; (3) pharmacogenomic data incorporation; (4) pedigreedrawing 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.

1. ElHazmi MAF, Warsy AS, Al-Swailem AR, Al-Swailem AM, Bahakim HM. Consanguinity among the Saudi Arabian population. J Med Genet. 1995;32(8):623-6. doi: 10.1136/jmg.32.8.623
2. AlGazali L, Hamamy H, AlArrayad S. Genetic disorders in the Arab world. BMJ. 2006;333(7573):831-4. doi: 10.1136/bmj.38982.704931.AE
3. Hamamy H. Consanguineous marriages: preconception consultation in primary health care settings. J Community Genet. 2012;3(3):185-92. doi: 10.1007/s12687-011-0072-y
4. Jastaniah W. Epidemiology of sickle cell disease in Saudi Arabia. Ann Saudi Med. 2011;31(3):289-93. doi: 10.4103/0256-4947.81540
5. Nasserullah Z, AlJame A, Abu Srair H, Al Qatari G, Al Naim S, Ageel AR, et al. Neonatal screening for sickle cell disease, glucose-6-phosphate dehydrogenase deficiency and other hemoglobinopathies in the Eastern Province of Saudi Arabia. Ann Saudi Med. 2003;23(1-2):70-4. doi: 10.5144/0256-4947.2003.70
6. AlSultan A, AlZahrani A, Alfadhel M, AlOtaibi M, AlAbdulkareem I. Prevalence of beta-thalassemia carriers in the Eastern Province of Saudi Arabia. Saudi Med J. 2021;42(11):1183-9. doi: 10.15537/smj.2021.42.11.20210574
7. Ar Rochmah MA, Awano H, Harahap NIF, Putranti AH, Matsumoto M, Nishio H. Spinal muscular atrophy carriers in a Saudi Arabian population. Neurogenetics. 2017;18(3):167-73. doi: 10.1007/s10048-017-0521-4
8. Memish ZA, Saeedi MY. Six-year outcome of the national premarital screening and genetic counseling program in Saudi Arabia. Ann Saudi Med. 2011;31(3):229-35. doi: 10.4103/0256-4947.81527
9. Hartl DL, Clark AG. Principles of Population Genetics. 4th ed. Sunderland (MA): Sinauer Associates; 2007.
10. Alfadhel M, Al-Thihli K, Moubayed H, Eyaid W, Al-Jeraisy M. Drug treatment of inborn errors of metabolism: a systematic review. Arch Dis Child. 2013;98(6):454-61. doi: 10.1016/j.jsps.2012.08.001
11. Arafat M, Abdelhaleem M, Al-Gazali L. Genetic disorders in the Arab world: Saudi Arabia. In: Teebi AS, editor. Genetic Disorders Among Arab Populations. 2nd ed. Berlin: Springer; 2010. p. 451-502. doi: 10.1007/978-3-642-05080-0_14
12. AlOwain M, Al-Zaidan H, Al-Hassnan Z. Map of autosomal recessive genetic disorders in Saudi Arabia: concepts and future directions. Am J Med Genet A. 2012;158A(10):2629-40. doi: 10.1002/ajmg.a.35551
13. Abouelhoda M, Sobahy T, El-Kalioby M, Patel N, Shamseldin H, Monies D, et al. Clinical genomics can facilitate countrywide estimation of autosomal recessive disease burden. Genet Med. 2016;18(12):1245-9. doi: 10.1007/s00439-021-02295-y
14. Alkuraya FS. Genetics and genomic medicine in Saudi Arabia. Mol Genet Genomic Med. 2014;2(5):369-78. doi: 10.1002/mgg3.97
15. Bittles AH, Black ML. Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci USA. 2010;107(Suppl 1):1779-86. doi: 10.1073/pnas.0906079106
16. Ali O. Genetics of type 2 diabetes. World J Diabetes. 2013;4(4):114-23. doi: 10.4239/wjd.v4.i4.114

17. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010;363(13):1211-21. doi: 10.1056/NEJMoa0906312
18. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019;365(6460):eaav7188. doi: 10.1126/science.aav7188
19. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431-44. doi: 10.1038/s41588-019-0344-8
20. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro RV, McGuire AL, et al. Genetic basis for congenital heart disease: revisited. Circulation. 2018;138(21):e653-711. doi: 10.1161/CIR.0000000000000606
21. Helbig I, Scheffer IE, Mulley JC, Berkovic SF. Navigating the channels and beyond: unravelling the genetics of the epilepsies. Lancet Neurol. 2008;7(3):231-45. doi: 10.1016/S1474-4422(08)70039-5
22. Leslie RD, Atkinson MA, Notkins AL. Autoantigens IA-2 and GAD in type I (insulin-dependent) diabetes. Diabetologia. 1999;42(1):3-14. doi: 10.1007/s001250051124
23. Robert AA, Al Dawish MA, Braham R, Musallam MA, Al Hayek AA, Al Kahtany NH. Type 2 diabetes mellitus in Saudi Arabia: major challenges and possible solutions. Curr Diabetes Rev. 2017;13(1):59-64. doi: 10.2174/1573399812666160126142609
24. AlQuwaidhi AJ, Pearce MS, Critchley JA, Sobngwi E, O'Flaherty M. Trends and long-term projections of adult obesity in Saudi Arabia, 1992-2022. East Mediterr Health J. 2014;20(10):589-95. doi: 10.26719/2014.20.10.589
25. Hook EB, Cross PK, Schreinemachers DM. Chromosomal abnormality rates at amniocentesis and in live-born infants. JAMA. 1983;249(15):2034-8. doi: 10.1001/jama.1983.03330390038028
26. Morris JK, Mutton DE, Alberman E. Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. J Med Screen. 2002;9(1):2-6. doi: 10.1136/jms.9.1.2
27. Therrell BL, Padilla CD, Loeber JG, Kneisser I, Saadallah A, Borrajo GJ, et al. Current status of newborn screening worldwide: 2015. Semin Perinatol. 2015;39(3):171-87. doi: 10.1053/j.semperi.2015.03.002
28. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ. 2008;86(6):480-7. doi: 10.2471/BLT.06.036673
29. Orphanet. Prevalence and incidence of rare diseases: bibliographic data. Orphanet Report Series, Rare Diseases Collection. 2023;1:1-76. Available from: https://www.orpha.net
30. AlSayed M. Saudi Arabia: genetic disease and its impact. In: Emery and Rimoin's Principles and Practice of Medical Genetics. Oxford: Oxford University Press; 2012. p. 777-792. doi: 10.1093/med/9780195374759.003.0054
31. Tadmouri GO, Nair P, Obeid T, Al Ali MT, Al Khaja N, Hamamy HA. Consanguinity and reproductive health among Arabs. Reprod Health. 2009;6:17. doi: 10.1186/1742-4755-6-17

32. Alfadhel M, Al-Thihli K, Moubayed H, Eyaid W, Al-Jeraisy M. Spectrum of inborn errors of metabolism diagnosed in a tertiary children's hospital in Saudi Arabia. Orphanet J Rare Dis. 2016;11(1):126. doi: 10.1186/s13023-016-0510-3
33. Saudi Human Genome Program. SHGP Scientific Discoveries and Achievements 2013–2020. Riyadh: King Abdulaziz City for Science and Technology (KACST); 2020.
34. AlQahtani MH. The Saudi Human Genome Program: an in-depth look at its impact and achievements. Saudi J Biol Sci. 2021;28(1):693-6. doi: 10.1016/j.sjbs.2020.10.063
35. General Authority for Statistics (GaStat). Marriage and Divorce Statistics 2023. Riyadh: GaStat; 2023. Available from: https://www.stats.gov.sa
36. Kauf TL, Coates TD, Huazhi L, Mody-Patel N, Hartzema AG. The cost of health care for children and adults with sickle cell disease. Am J Hematol. 2009;84(6):323-7. doi: 10.1002/ajh.21408
37. Scalone L, Mantovani LG, Krol M, Dobrescu AM, Prophette B, Féron JM, et al. Costs and quality of life in transfusion-dependent thalassemia patients. Expert Rev Pharmacoecon Outcomes Res. 2008;8(5):469-76. doi: 10.1586/14737167.8.5.469
38. Armstrong EP, Malone DC, Yeh WS, Meltzer DO. Economic burden of spinal muscular atrophy. J Med Econ. 2016;19(8):822-6. doi: 10.1080/13696998.2016.1198355
39. Pai SY, Logan BR, Griffith LM, Buckley RH, Parrott RE, Dvorak CC, et al. Transplantation outcomes for severe combined immunodeficiency, 2000-2009. N Engl J Med. 2014;371(5):434-46. doi: 10.1056/NEJMoa1401177
40. AlHaggar M. Economic impact of inborn errors of metabolism treatment in low and middle income countries. Egypt J Med Hum Genet. 2013;14(2):123-8. doi: 10.1016/j.ejmhg.2013.03.003
41. Grosse SD, Rogowski WH, Ross LF, Cornel MC, Dondorp WJ, Khoury MJ. Population screening for genetic disorders in the 21st century: evidence, economics, and ethics. Public Health Genomics. 2010;13(2):106-15. doi: 10.1159/000226594
42. Langlois S, Benn P, Wilkins-Haug L. Current controversies in prenatal diagnosis: pre-conception expanded carrier screening should replace all current prenatal screening for autosomal recessive conditions. Prenat Diagn. 2023;43(9):1199-209. doi: 10.1002/pd.6389