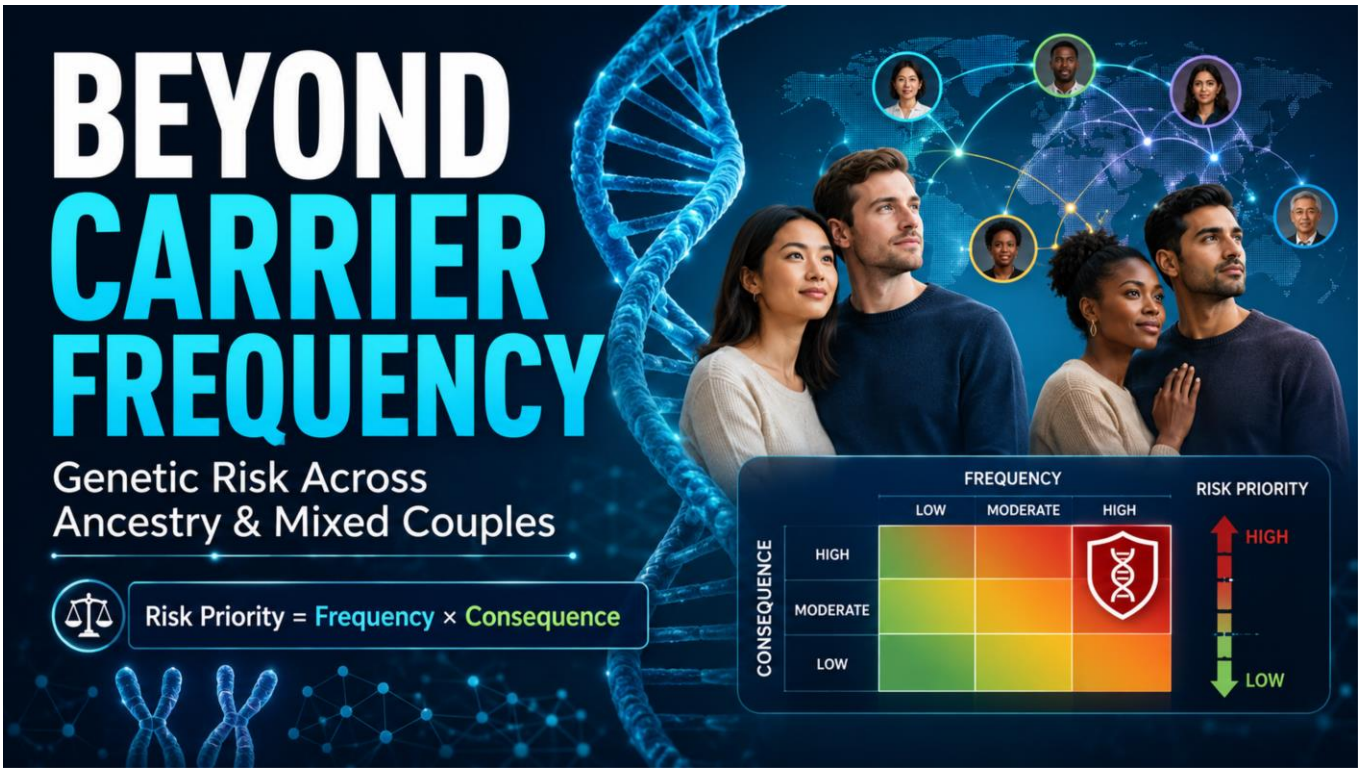

Beyond Carrier Frequency Across Human Ancestry

A Consequence-Weighted Framework for Genetic Screening Across Human Races, Ancestry Groups, and Inter-Racial Couples

[1]Boris Petrikovsky, MD, PhD

[2]Lev Neymotin, PhD

Draft date: July 3, 2026

Executive Summary

Part I of this project used Jewish founder populations as a well-documented model for genetic carrier screening. Its central conclusion was that carrier frequency alone is not an adequate guide to screening priority. A condition carried by many people but usually mild, treatable, or variably penetrant should not be treated as equivalent to a less frequent condition that predictably causes early death, severe neurodegeneration, blindness, deafblindness, major organ failure, or lifelong profound disability. Part I

[1] Boris Petrikovsky, MD, PhD, Professor of Obstetrics and Gynecology
[2] Lev Neymotin, PhD, former researcher at Brookhaven National Laboratory

therefore proposed a transparent screening-priority measure: Risk Priority Index = Carrier Frequency (%) × Severity of Consequence Score.

Part I also made an important genetic distinction. Reproductive risk is not limited to two parents carrying the identical mutation. Classical recessive disease may result from homozygosity for the same pathogenic variant, but it may also result from compound heterozygosity, in which two different pathogenic variants in the same gene converge functionally in the child. Beyond that, selected disorders may involve digenic, oligogenic, or modifier-gene effects. These more complex interactions are real, but they require stronger evidence before they can be used as practical compatibility rules.

This second paper extends the same logic from Jewish founder populations to the main human race and ancestry groups of the world. The purpose is not to convert race into a social hierarchy or a biological caste system. Race is used here only in the restricted scientific sense introduced in Part I: a broad descriptor of historically formed ancestry groups in which shared descent, geographic history, founder effects, genetic drift, endogamy, migration, selection, and admixture have produced statistically recognizable allele-frequency patterns. In clinical or commercial use, race-level language must always be refined into genetic ancestry, ethnicity, founder population, family origin, and individual test results.

The major new element in Part II is the inter-racial or cross-ancestry couple. Such couples introduce an additional layer of complexity. For some founder disorders, cross-ancestry pairing may reduce the probability that both partners carry the same founder variant. But that does not mean risk disappears. Different ancestry groups may carry different pathogenic variants in the same gene, and a child may still inherit a clinically important biallelic combination. The hemoglobin disorders are the most obvious example: sickle hemoglobin, beta-thalassemia, alpha-thalassemia, hemoglobin E, hemoglobin C, and other variants may combine across ethnic boundaries in ways that create severe disease.

The practical conclusion is that a modern partner-prechecking system should not be built around race labels alone. It should use a three-layer structure: a universal pan-ethnic screening layer offered to everyone; an ancestry/founder-population layer that adds conditions enriched in particular groups; and a couple-interaction layer that asks whether the two partners' variants converge on the same gene, same protein complex, same biologic pathway, or a known disease-modifying system. The output should be a risk-priority and counseling trigger, not an automatic judgment about whether two people should marry, reproduce, or continue a relationship.

This paper is therefore a design draft for a practical, consent-based preconception and premarital screening enterprise. It proposes representative race/ancestry categories, a working consequence scale, preliminary RPI tables, a cross-ancestry couple-risk model, and a step-by-step program design. It deliberately does not attempt to include every possible disease, every ethnic group, or every variant. The objective is to build a consistent framework that can later be expanded into a more comprehensive database and software-supported compatibility system.

Source note. This draft uses Part I as the internal model and combines it with current professional carrier-screening guidance, public-health sources, GeneReviews disease summaries, and ancestry-stratified allele-frequency resources. Carrier frequencies in the tables are approximate and should be verified condition-by-condition before any formal clinical or commercial deployment.

1. Introduction: From Jewish Founder Populations to Global Ancestry Screening

The first paper in this project, Beyond Carrier Frequency, developed a consequence-weighted method for genetic screening in Jewish founder populations. That model was appropriate because the Jewish screening experience is unusually well documented. Tay-Sachs disease, Canavan disease, familial dysautonomia, cystic fibrosis, Gaucher disease, Fanconi anemia, Niemann-Pick disease, mucolipidosis IV, Usher syndromes, and many Sephardic/Mizrahi founder disorders provide a rich and clinically meaningful test case.

The main lesson from Part I was simple but powerful: a screening system should not be driven by carrier frequency alone. Frequency describes how often a carrier state appears in a group. It does not describe what happens to a child who inherits two pathogenic alleles. A high-frequency but treatable disorder may deserve a different priority than a lower-frequency disorder that is fatal in infancy. This is why Part I introduced the Severity of Consequence Score (SCS) and the Risk Priority Index (RPI).

The second lesson was that ancestry labels must be handled carefully. Jewish genetic risk is not one uniform category. Ashkenazic, Moroccan, Iranian, Iraqi, Yemenite, Libyan, Georgian, Bukharan, Kurdish, Dagestani/Caucasian, Syrian, North African, and other Jewish lines have different founder effects and should not be collapsed into one medical bucket. The same logic applies even more strongly when the discussion is widened to the major races and ethnicities of the world.

The third lesson was that genetic compatibility is not merely a marker-to-marker match. Two identical pathogenic variants can cause recessive disease, but two different pathogenic variants in the same gene may do the same. This is the central concept of compound heterozygosity. At still greater complexity, variants in different genes can interact through common pathways or protein systems, but such digenic and oligogenic effects require careful evidence before being used in counseling.

Part II applies these principles to broader human genetic structure. It asks how a consequence-weighted system should treat European, African, Middle Eastern, North African, South Asian, East Asian, Southeast Asian, Indigenous American, Pacific Islander, Hispanic/Latino admixed, and other ancestry-linked groups. It also asks how the system should handle inter-racial couples, where the probability of identical founder-variant matching may decrease while the number of possible cross-gene and same-gene variant combinations may increase.

This is not a final clinical panel. It is a technical draft for designing a practical screening program. The paper therefore emphasizes structure: which groups should be represented, what data quality should be required, how severity should be scored, how partner combinations should be evaluated, and how a commercial or community program could operate while preserving consent, privacy, counseling, and ethical restraint.

2. Terminology: Race, Genetic Ancestry, Ethnicity, and Founder Line

Race, in the restricted scientific sense used here, refers to a broad historically formed ancestry group in which patterns of shared descent, geography, reproductive history, genetic drift, natural selection, migration, and admixture have produced statistically recognizable differences in allele frequencies. This definition is not social, hierarchical, moral, or cultural. It does not imply purity, superiority, inferiority, or fixed biological compartments. Human genetic variation is continuous and overlapping.

Genetic ancestry is more precise than race because it can be inferred from genetic data and can be partitioned by continental, regional, local, and sometimes founder-population components. Ethnicity is often a mixture of ancestry, language, geography, religion, culture, and social history. A founder line is narrower still: a population shaped by a limited number of ancestors, endogamy, or geographic or social isolation, leading to enrichment of particular variants.

In a screening program, race can serve only as the first broad organizing layer. It is useful because disease-related alleles are not randomly distributed across humanity. It is insufficient because the actual clinical question is never, 'What race is this person?' The clinical question is, 'Which pathogenic variants are present in this person and in this potential reproductive partner, and what is the consequence if a child inherits the relevant combination?'

For that reason, the practical hierarchy used in this paper is: race/continental ancestry as the broadest layer; regional ancestry and ethnicity as the second layer; founder population and family origin as the third layer; and actual genetic test results as the decisive layer.

3. Objectives

- To extend the consequence-weighted genetic-screening framework of Part I beyond Jewish founder populations to the major human race and ancestry groups of the world.
- To create a practical representative list of race, ancestry, ethnicity, and founder-population categories that are sufficiently documented to support preliminary screening-priority analysis.
- To preserve the 1–5 Severity of Consequence Score used in Part I and to use the same Risk Priority Index formula for initial screening prioritization.
- To add a cross-ancestry couple layer, recognizing that inter-racial couples may have reduced probability of identical founder-variant matching but may still have meaningful same-gene, hemoglobin, compound-heterozygous, and modifier-gene risks.
- To design a step-by-step prechecking program that could be developed into a commercial or community screening service, while remaining practical, ethically restrained, consent-based, and compatible with genetic counseling.
- To identify a research line on the biological mechanisms by which different pathogenic variants may reinforce one another at the gene, protein, pathway, and developmental-threshold levels.

4. Analytical Framework

4.1 Core Definitions

Carrier frequency is the approximate proportion of unaffected individuals in a defined group who carry one pathogenic variant for a condition. For example, a carrier frequency of 1 in 25 equals 4%.

Known-carrier couple risk is different from population carrier frequency. For most autosomal recessive disorders, when both reproductive partners are confirmed carriers of pathogenic variants in the same condition, each pregnancy carries a 25% risk of an affected child.

Random same-group affected-child risk can be approximated as carrier frequency × carrier frequency × 1/4. For a condition with carrier frequency 1 in 25 in both partners' group, the pretest random-couple risk is 1/25 × 1/25 × 1/4 = 1/2,500.

Random cross-group affected-child risk can be approximated as carrier frequency in Partner A's group × carrier frequency in Partner B's group × 1/4, provided the variants converge on the same gene or same clinically validated disease mechanism.

Severity of Consequence Score (SCS) ranks the gravity of the affected-child outcome from 1 to 5. It includes lethality, disability burden, age of onset, treatability, reversibility, progressive decline, sensory loss, organ failure, cancer predisposition, and ethical complexity.

Risk Priority Index (RPI) remains the first-layer screening-priority formula: RPI = Carrier Frequency (%) × SCS. It is not a disease probability. It is a way to rank conditions for screening attention within a defined group.

4.2 Severity of Consequence Scale

SCS	Interpretation	Practical meaning
5	Catastrophic	Usually fatal, profoundly disabling, or devastating early in life
4	Severe	Major lifelong disability, organ failure, deafblindness, severe metabolic disease, severe immune/bleeding disease, or major multisystem disease
3	Serious but variable/actionable	Major condition with treatment, surveillance, variable prognosis, or reduced but meaningful life impact
2	Moderate	Meaningful medical consequence, usually manageable, often nonlethal, sometimes episodic or trigger-dependent

SCS	Interpretation	Practical meaning
1	Mild/low medical severity	Usually benign, mild, situational, or primarily social/identity-sensitive rather than medically devastating

Half-steps may be used when a condition spans categories. Hearing loss, for example, should not be scored as one uniform disease because nonsyndromic, syndromic, progressive, vestibular, deafblindness-associated, and medically complex forms have different consequences.

5. Representative Race/Ancestry Categories

The following list is intentionally practical rather than philosophical. It does not claim that humanity is divided into rigid boxes. It identifies broad ancestry layers and nested ethnic/founder groups for which statistically meaningful data often exist or should be developed. The program should treat these categories as intake prompts, not final diagnoses.

Broad layer	Representative subgroups	Screening relevance
African / African diaspora	West African, Central African, East African, African American, Afro-Caribbean, Afro-Latino	Sickle cell trait/disease, alpha-thalassemia, G6PD deficiency, some hemoglobin C patterns, variable beta-thalassemia; data quality varies by subregion.
European	Northern, Western, Southern, Eastern, Finnish, Irish, French Canadian, Cajun, Amish/Mennonite/Hutterite, Roma	Cystic fibrosis, SMA, MCAD deficiency, PKU, Tay-Sachs in French Canadian/Cajun and selected Irish lines, Finnish disease heritage, founder effects in isolates.
Middle Eastern / North African / Mediterranean	Arab, Turkish, Armenian, Iranian, Kurdish, Greek, Italian, Cypriot, Sardinian, Sephardic/Mizrahi lines	Beta-thalassemia, alpha-thalassemia, familial Mediterranean fever, G6PD deficiency, selected founder disorders; high heterogeneity.
South Asian	Indian, Pakistani, Bangladeshi, Sri Lankan, Nepali, regional/tribal/caste/endogamous groups	Beta-thalassemia, alpha-thalassemia in some groups, hemoglobin E in selected eastern/northeastern groups, G6PD deficiency, high importance of endogamy and regional structure.
East Asian	Chinese, Japanese, Korean, Mongolian, regional East Asian groups	Alpha-thalassemia in southern China and related regions, GJB2/GJB3-related hearing loss patterns, selected metabolic and founder variants; lower CF frequency than Europeans but not zero.
Southeast Asian	Cambodian, Thai, Lao, Vietnamese, Hmong, Burmese/Myanmar, Filipino, Malay, Indonesian	Hemoglobin E, alpha-thalassemia, beta-thalassemia, Hb Constant Spring, G6PD deficiency; major hemoglobin-interaction burden.
Hispanic / Latino / Admixed American	Mexican, Central American, Caribbean, Andean, Southern Cone, mixed Native American/European/African ancestry	Admixed risk cannot be inferred from label alone. CF, hemoglobinopathies, G6PD, SMA, and region-specific Indigenous/founder variants require ancestry-aware testing.
Indigenous American / First Nations	North, Central, and South American Indigenous groups; Inuit/Arctic groups; tribal/community-specific lines	Underrepresented in databases. Some founder and metabolic signals exist, but broad claims should be avoided without community-specific evidence.
Oceanian / Pacific Islander	Melanesian, Polynesian, Micronesian, Māori, Aboriginal Australian and Torres Strait Islander communities	Hemoglobinopathies and G6PD in some regions; data are uneven and often underrepresented. Local data and community partnership are essential.
Jewish founder populations	Ashkenazic, Sephardic, Mizrahi, North African, Iranian, Iraqi, Yemenite, Georgian, Bukharan, Kurdish, Libyan, Dagestani/Caucasian, Syrian, others	Covered in Part I but should be carried forward as a nested founder-population layer, not as one uniform genetic category.

The strongest practical approach is not to choose between ancestry-specific and pan-ethnic screening. The program should use both. A universal screening core prevents false reassurance in mixed or unknown ancestry. An ancestry/founder add-on prevents loss of high-signal conditions that may be rare globally but important in a specific group.

6. Preliminary Consequence-Weighted Risk-Priority Table Across Major Groups

The table below is a working draft, not a final clinical panel. It intentionally mixes broad group signals and founder-line signals to demonstrate the method. Frequencies are approximate, may differ by region and testing method, and must be verified before publication or commercial deployment. RPI is calculated as carrier percentage × SCS. For ranges, the table uses a representative high-signal value to show why the condition deserves attention.

Condition / gene group	Highest-signal group or broad ancestry	Carrier frequency used	Carrier %	SCS	RPI	Interpretive note
β-thalassemia trait	Mediterranean / Middle Eastern / South Asian / Southeast Asian hotspots	up to 20% in selected high-frequency regions	20.0	4.0	80.0	Severe beta-thalassemia major/intermedia if both partners contribute clinically significant HBB variants.
Hemoglobin E trait / HbE–β-thalassemia risk	Cambodian / Thai / Lao / Southeast Asian	Cambodian about 1 in 4; Thai about 1 in 9	25.0	3.5	87.5	HbE trait is mild, but HbE combined with beta-thalassemia can produce serious transfusion-dependent disease.
Familial Mediterranean fever, MEFV	Sephardic/Mizrahi Jewish, Armenian, Arab, Turkish, Mediterranean/Middle Eastern	as high as about 1 in 5 in several affected groups	20.0	2.0	40.0	Common and treatable; penetrance and severity vary. Should not dominate solely by frequency.
Sickle cell trait / sickle cell disease	African / African diaspora; also Middle Eastern, Indian, Mediterranean, Caribbean, Latin American	about 1 in 13 Black/African American babies with SCT	7.7	4.0	30.8	Major cross-ancestry issue because HbS can combine with HbC, beta-thalassemia, or other HBB variants.
α-thalassemia, including α0 cis deletions	Southeast Asian; also African, Mediterranean, Middle Eastern, South Asian	highly regional; α0 common in Southeast Asia	5.0+	5.0	25.0+	Hb Bart hydrops fetalis can be catastrophic when both parents carry cis two-gene deletions.
G6PD deficiency	African, Mediterranean, Middle Eastern, Asian; malaria-belt ancestry	about 1 in 10 African American males in the U.S.; higher in some regions	10.0	2.0	20.0	X-linked and trigger-dependent; important for neonatal jaundice and drug/food exposures, but not a standard autosomal-recessive couple score.
Tay-Sachs disease, HEXA	Ashkenazic Jewish; French Canadian; Louisiana Cajun; selected Irish lines	about 1 in 27 in classic high-risk founder groups	3.7	5.0	18.5	Founder-line risk can persist outside Jewish populations; cross-ancestry risk is lower but not zero.
Cystic fibrosis, CFTR	European / Ashkenazic Jewish; also Hispanic, African American, Asian American at lower frequencies	European about 1 in 25	4.0	4.0	16.0	Pan-ethnic screening is needed because CFTR risk is not exclusive to Europeans.
Spinal muscular atrophy, SMN1	Pan-ethnic	about 1 in 40 to 1 in 60	2.0	4.0	8.0	Universal core condition; residual risk differs by ancestry and testing method.
GJB2-related nonsyndromic hearing loss	Many populations; variant spectrum differs by ancestry	about 3% carrier rate reported in one U.S. control sample	3.0	2.5	7.5	Ethically sensitive; genotype class influences severity. Separate nonsyndromic from syndromic/progressive forms.
MCAD deficiency, ACADM	Northern European; selected founder lines	about 1 in 40 to 1 in 100 for common variant	2.5	3.0	7.5	Potentially fatal metabolic crises but highly actionable with newborn screening and avoidance of fasting.
Phenylketonuria / PAH deficiency	Northern European, Irish, Turkish, some East Asian and other groups	about 1 in 50 in Northern European ancestry	2.0	3.0	6.0	Serious if untreated; newborn screening and diet strongly modify outcome.
Cystic fibrosis, CFTR	Hispanic American	about 1 in 58	1.7	4.0	6.9	Lower than European frequency but clinically important; panel sensitivity differs by ancestry.
Cystic fibrosis, CFTR	African American / Black	about 1 in 61	1.6	4.0	6.6	Lower than European frequency; residual risk after limited panels can be higher.
Wilson disease, ATP7B	Global; some regional/founder enrichment	about 1 in 90 general carrier frequency	1.1	3.5	3.9	Treatable but potentially severe liver/neurologic disease if missed.

The table demonstrates the main point of the framework. Hemoglobin E and beta-thalassemia can rise to the top because their carrier frequencies are very high in specific populations. Familial Mediterranean fever also ranks high by frequency, but its consequence score is deliberately moderated because penetrance, treatability, and clinical severity differ from catastrophic infantile neurodegenerative disorders. Tay-Sachs, by contrast, remains high-priority despite a lower carrier frequency because its consequence score is maximal.

7. Inter-Racial and Cross-Ancestry Couples: The Added Layer

Inter-racial couples require a separate analytic layer because the two partners may not share the same founder variants. This can reduce risk for some narrow founder disorders, but it does not eliminate reproductive genetic risk. The actual question is whether both partners carry pathogenic variants that converge on the same functional genetic unit or validated disease mechanism.

For founder-private variants, cross-ancestry pairing often lowers the probability of two copies of the same founder mutation. For example, an Ashkenazic-specific founder variant is less likely to be found in a partner with no Jewish, French Canadian, Cajun, or related founder background. However, if the non-Jewish partner carries a different pathogenic variant in the same gene, the couple may still face compound-heterozygous risk. A compatibility system that checks only for exact identical markers could therefore miss clinically meaningful risk.

For globally distributed genes, inter-racial status may be less protective than people assume. CFTR, SMN1, HBB, HBA1/HBA2, GJB2, PAH, and ATP7B are not restricted to one race. Frequencies, variant spectra, and test sensitivity differ by ancestry, but the gene-level risk can cross racial boundaries.

The hemoglobin disorders are the clearest example of cross-ancestry complexity. One partner may carry sickle hemoglobin and another may carry beta-thalassemia or hemoglobin C. A Southeast Asian partner may carry hemoglobin E, while a partner from South Asia, the Middle East, or the Mediterranean may carry beta-thalassemia. In such cases, inter-racial pairing may create exactly the kind of gene-level convergence that a superficial race-based screen would fail to predict.

Couple type	Example	Practical implication
Same ancestry, same founder condition	Both partners from the same founder line; same high-frequency variant pool	Higher probability of exact founder-variant matching; RPI table useful for initial prioritization
Same broad race, different ethnic/founder lines	Both partners “European” or both “South Asian,” but from different regional or endogamous backgrounds	Broad race may mislead. Use regional/founder history and actual variants.
Inter-racial, founder-private condition	One partner has a founder-line condition; the other has no known related ancestry	Risk usually decreases but should not be set to zero; check full gene, not only founder marker.
Inter-racial, same gene with different variant spectra	Examples: CFTR, GJB2, PAH, ATP7B, SMN1	Compound heterozygosity can occur; testing must be gene-aware and not only marker-aware.
Inter-racial hemoglobin combination	HbS × β-thalassemia; HbE × β-thalassemia; HbC × HbS; α-thalassemia modifiers	High priority because clinically severe disease can arise from different variants in the same globin system.
Admixed ancestry or unknown ancestry	Hispanic/Latino, African American, Caribbean, Middle Eastern, Central Asian, many modern families	Use pan-ethnic expanded screening plus ancestry-informed interpretation; avoid false reassurance from self-label alone.

8. Illustrative Couple-Level Calculations

The RPI ranks conditions for screening attention. For couple-level risk, the more useful pretest calculation is carrier frequency in Partner A’s group × carrier frequency in Partner B’s group × 1/4. The examples below are simplified and should be replaced by actual test results whenever available.

Scenario	Pretest calculation	Approximate risk	Interpretive note

Scenario	Pretest calculation	Approximate risk	Interpretive note
European-European couple, CFTR	1/25 × 1/25 × 1/4	1/2,500 per pregnancy before testing	Classic same-ancestry example; severe but now more treatable than historically.
European-African American couple, CFTR	1/25 × 1/61 × 1/4	about 1/6,100 before testing	Lower than European-European, but not zero; test sensitivity differs by ancestry.
Black/African American same-group couple, HbS trait approximation	1/13 × 1/13 × 1/4	about 1/676 before testing	Illustrates why hemoglobinopathy screening is a central public-health issue.
Cambodian HbE carrier-frequency signal × partner with beta-thalassemia carrier signal	1/4 × 1/20 × 1/4	about 1/320 before testing	Shows cross-ethnic hemoglobin risk; exact risk depends on partner ancestry and HBB result.
Known carrier × untested partner	1 × partner carrier frequency × 1/4	depends on partner's tested or estimated frequency	Once one partner is known to be a carrier, the other partner's testing becomes urgent.
Both partners confirmed carriers in same relevant gene	1 × 1 × 1/4	1/4 per pregnancy	At this point, the couple needs genetic counseling and reproductive options, not only risk ranking.

9. Practical Program Design for Partner Prechecking

The proposed enterprise should be designed as a consent-based reproductive-risk prechecking system. It should not be marketed as a mate-selection authority. Its purpose is to identify genetic combinations that deserve counseling before pregnancy or before major marital/reproductive decisions are made. A staged design is recommended.

Step 1. Intake, consent, and scope definition. Obtain explicit consent from each participant. Explain that the system screens for reproductive genetic risk, not personal worth, identity, social fitness, or general health. Document whether the participant wants an open medical report, a compatibility-only report, or both.

Step 2. Structured ancestry and family-history collection. Collect self-identified race, ethnic background, countries/regions of family origin, known founder populations, language/religious/endogamous community history, consanguinity, adoption/unknown ancestry, and known family disease history. Use three-generation family history when possible.

Step 3. Data-quality classification. Assign each ancestry component to a data-quality tier: robust population data; moderate regional/founder data; limited or underrepresented data; unknown/admixed. The lower the data quality, the more the program should rely on broad sequencing rather than a narrow ethnicity panel.

Step 4. Universal screening core. Offer a pan-ethnic core panel to all participants. This should include major autosomal recessive and selected X-linked conditions with sufficient severity, carrier frequency, and clinical validity. Hemoglobinopathy testing should be included as a core element, not treated as an optional race-only add-on.

Step 5. Ancestry/founder add-on layer. Add variants or genes enriched in specific ancestry/founder lines. Examples include Tay-Sachs in Ashkenazic/French Canadian/Cajun contexts, Finnish disease heritage conditions, hemoglobin E in Southeast Asian groups, and region-specific thalassemia panels.

Step 6. Partner-combination algorithm. Compare the two partners at four levels: identical pathogenic variant; different pathogenic variants in the same gene; different genes in a validated disease mechanism or protein/pathway system; and modifier-gene situations with evidence-based clinical effect. The first two are strongest for routine counseling.

Step 7. Consequence-weighted prioritization. For each flagged condition, calculate RPI and, where possible, couple-specific pretest or post-test risk. Add the 1–5 SCS so that common but treatable conditions do not automatically outrank less common catastrophic conditions.

Step 8. Classification of output. Use clear categories: no high-risk reproductive match detected; carrier detected but partner negative with residual risk; same-gene carrier couple requiring counseling; uncertain or limited-evidence interaction; urgent hemoglobinopathy combination; X-linked maternal carrier issue; mitochondrial/maternal-line issue.

Step 9. Genetic counseling and confirmatory testing. Require genetic counseling for high-risk, uncertain, or emotionally sensitive results. Confirm pathogenicity, phase, zygosity, and variant classification before any major decision. Do not rely on direct-to-consumer raw data alone.

Step 10. Reproductive-options discussion. For confirmed high-risk couples, discuss options such as natural conception with prenatal diagnosis, IVF with PGT-M, donor gametes, adoption, embryo/fetal testing, preparation for an affected child, or choosing a different reproductive pathway. The program should present options, not dictate choices.

Step 11. Privacy and compatibility reporting. Separate clinical data from compatibility output. Consider a confidential compatibility model for communities that prefer not to disclose individual carrier states, but also allow open medical reporting when participants want their own results.

Step 12. Database updating and quality control. Update variant classification, carrier-frequency estimates, severity scores, penetrance data, and treatment effects continuously. Maintain a review board with medical genetics, genetic counseling, bioethics, data privacy, and community representation.

10. Proposed Output Format for a Commercial Prechecking Report

A practical report should be short enough to be understood but detailed enough to support counseling. The main report should not overwhelm users with long variant tables. A technical appendix can contain the full gene and variant information.

The report should avoid language such as 'good match' or 'bad match' as a moral judgment. Better language is 'no high-risk reproductive carrier pairing detected,' 'same-gene carrier pairing detected,' 'hemoglobinopathy combination requiring counseling,' or 'uncertain evidence; genetics review recommended.'

The system should preserve a distinction between compatibility and diagnosis. A compatible result does not guarantee a healthy child. An incompatible or high-risk result does not mean the relationship is wrong. It means that reproductive options should be discussed before pregnancy.

Category	Meaning	Action
Green	No high-risk pairing detected in screened genes	Explain residual risk; no test eliminates all risk.
Yellow	One partner carrier; other negative or untested	Recommend partner testing if not completed; explain residual risk by ancestry and panel type.
Orange	Both partners carry variants in same gene, but consequence/phase/pathogenicity requires review	Genetic counselor or medical geneticist review before conclusions.
Red	Both partners carry pathogenic variants in the same condition with established 25% recessive risk	Confirm results; discuss reproductive options.
Purple	Hemoglobinopathy or globin-system combination	Special review because different HBB/HBA variants can combine across ancestries.
Blue	X-linked or maternal-line issue	Different inheritance logic; evaluate fetal sex risk, maternal carrier status, or mitochondrial transmission.

11. Analysis and Discussion

The global extension of the Part I framework confirms that carrier frequency by itself is a weak guide. Hemoglobin E trait can be extremely common in Cambodian and other Southeast Asian groups, but the trait alone is usually mild. The danger appears when hemoglobin E is paired with beta-thalassemia. Familial Mediterranean fever may have a high carrier frequency in several Mediterranean and Middle Eastern populations, but treatability and reduced penetrance moderate its consequence score. Cystic fibrosis is less frequent outside European-derived populations, but it remains serious and should not be treated as a race-exclusive condition.

The hemoglobin disorders deserve special treatment because they violate the simplistic idea that one screens only for one disease in one race. Sickle hemoglobin, hemoglobin C, hemoglobin E, beta-thalassemia, alpha-thalassemia, and other variants are part of a shared globin system. The clinically important question is not only whether two partners share the same ethnic background, but whether their variants interact within the same oxygen-carrying system. In a modern mixed population, hemoglobinopathy screening should be universal or near-universal.

Inter-racial couples do not create a simple reduction of risk. They change the shape of risk. Founder-private identical-variant matching may be reduced, but same-gene compound heterozygosity may remain possible. In genes such as CFTR, GJB2, PAH, ATP7B, and HBB, different populations may carry different pathogenic variants in the same gene. The program therefore needs gene-level matching rather than marker-level matching alone.

The modern movement toward pan-ethnic expanded carrier screening is appropriate because it reduces dependence on self-reported race and catches risk in mixed or unknown ancestry. But ancestry remains medically useful. It helps interpret residual risk, panel sensitivity, founder variants, and the likelihood of specific pathogenic alleles. The goal is not to erase race/ancestry information, but to move through it: from broad group, to founder line, to gene, to variant, to couple-specific biological consequence.

The ethical risk is real. A prechecking system can easily be misunderstood as a system for ranking people. That must be avoided in design, language, and reporting. The program should rank reproductive genetic combinations, not human beings. It should produce counseling triggers, not social verdicts. It should support informed choices, not pressure people into predetermined reproductive decisions.

A commercial program will succeed only if it is both scientifically rigorous and emotionally acceptable. It must be practical enough for real couples and communities, but careful enough to avoid false reassurance, false alarm, misuse of race labels, and overinterpretation of uncertain gene-gene interactions.

12. Research Line: How Do Mutant Variants Reinforce One Another?

The phrase 'genes recognizing each other' is useful metaphorically but should be translated into molecular language. Genes do not recognize one another in a conscious or intentional sense. The problem arises when the products of genes—RNA, proteins, enzymes, receptors, channels, structural molecules, or regulatory elements—fail to supply enough functional activity for a biological system to remain above a critical threshold.

The simplest mechanism is biallelic loss of function in the same gene. Each parent contributes one pathogenic variant. The child inherits two damaged copies. If both variants severely reduce or abolish function, the child may have little or no working protein. This is the classical autosomal recessive model.

The second mechanism is compound heterozygosity with unequal allele severity. One variant may be catastrophic and the other partially functional. The resulting phenotype may depend on the residual activity of the milder allele. This explains why not all variant combinations in a gene have identical consequences. CFTR and GJB2 provide clear examples in which genotype class strongly influences phenotype.

The third mechanism is cis/trans configuration. In some cases, two variants are dangerous only when they are on opposite chromosomes, because each allele is compromised. In other cases, two variants on the

same chromosome leave the other chromosome functional and produce a different risk. This is important in CFTR interpretation and in alpha-thalassemia, where cis two-gene deletions carry a very different reproductive risk than trans deletions.

The fourth mechanism is pathway convergence. Two different genes may participate in the same protein complex, metabolic pathway, developmental sequence, immune system, sensory pathway, or structural tissue system. A partial defect in each may be tolerated separately, but together they may push the system below a functional threshold. This is the biologic basis for some digenic or oligogenic inheritance.

The fifth mechanism is modifier-gene reinforcement. A primary disease genotype may be made worse or milder by variants elsewhere. Examples include fetal hemoglobin modifiers in sickle cell disease, SMN2 copy number in spinal muscular atrophy, and modifier systems that influence hemoglobin E / beta-thalassemia severity. These are scientifically important, but most are not yet suitable for simple compatibility scoring unless the evidence is strong.

The research program should therefore add an Interaction Evidence Score in the future. Same-gene pathogenic biallelic convergence would receive the highest evidence score. Established hemoglobin-system combinations would also rank high. Suspected pathway interactions, modifier-only effects, and computational predictions would rank lower until supported by clinical data.

13. Limitations

- Carrier frequencies are approximate and often population-specific. A value taken from one country, founder line, or clinical database may not apply to all people who share a broad racial label.
- Race and ethnicity labels are imperfect. Self-identified identity, social race, genetic ancestry, and family origin may not align, especially in admixed populations.
- The RPI is a screening-priority index, not a disease probability and not a substitute for genetic counseling.
- Some high-frequency conditions are not classical autosomal recessive diseases. G6PD deficiency, Fragile X premutations, mitochondrial disease, adult-onset cancer predisposition, and pharmacogenetic risks require separate scoring logic.
- The present draft emphasizes conditions with meaningful existing databases. Underrepresented Indigenous, African, Pacific Islander, and local founder populations require additional data before strong conclusions can be made.
- Ethically sensitive conditions, especially hearing loss and adult-onset disorders, should not be handled as simple 'prevent disease' categories without attention to identity, disability culture, treatability, penetrance, and personal values.

14. Conclusions and Recommendations

The consequence-weighted framework developed in Part I can be extended beyond Jewish founder populations, but only if it is adapted rather than copied. Global screening cannot be reduced to one table of races and diseases. It requires a layered model: broad race/ancestry, regional ethnicity, founder line, family history, individual genetic result, partner result, and biological consequence of the couple’s combined variants.

The most important conceptual addition in Part II is the inter-racial couple. Cross-ancestry pairing may reduce some founder-private risks, but it does not eliminate genetic risk. It may also create gene-level convergence that is invisible to a simplistic race-label system. Hemoglobin disorders demonstrate this most clearly, but the same principle applies to many genes with population-specific variant spectra.

A practical partner-prechecking program should combine universal pan-ethnic screening with ancestry/founder add-ons and a couple-specific interaction algorithm. The program should check exact

variant matches, different variants in the same gene, validated hemoglobin and pathway interactions, X-linked risks, and data-quality limitations.

The 1–5 Severity of Consequence Score should be retained for consistency with Part I. It prevents common but treatable conditions from automatically dominating the screening discussion and prevents catastrophic but less frequent conditions from being neglected.

The program should be designed as a counseling trigger, not as a mate-selection verdict. Its ethical center should be informed reproductive choice, disease prevention when appropriate, privacy, respect for individuals and communities, and transparency about uncertainty.

The next stage should be development of a structured database. For each condition, the database should record gene, variants, inheritance mode, carrier frequency by ancestry/founder line, consequence score, treatability, penetrance, age of onset, evidence level, screening method, residual risk, and validated gene-gene interactions. Once that database exists, the commercial or community program can become algorithmic without becoming careless.

Appendix A. Proposed Database Fields for the Next Project

Database field group	Recommended contents
Condition identity	Condition name; OMIM/GeneReviews/ClinGen identifiers; disease category
Gene/variant layer	Gene(s); variant(s); pathogenicity class; founder variant status; sequencing vs targeted detection
Population layer	Race/continental ancestry; regional ancestry; ethnicity; founder line; data-quality tier
Frequency layer	Carrier frequency; allele frequency; disease prevalence; newborn-screening incidence; confidence interval if available
Consequence layer	SCS; age of onset; lethality; disability burden; treatability; reversibility; penetrance; ethical sensitivity
Couple layer	Exact same variant; same gene different variants; hemoglobin combination; X-linked risk; mitochondrial/maternal risk; validated digenic/oligogenic interaction
Reporting layer	Green/yellow/orange/red/purple/blue category; counseling urgency; residual risk; recommended confirmatory tests
Governance layer	Source date; evidence grade; review date; reviewer identity; update history

References / Source Base for Draft Development

1. Petrikovsky B, Neymotin L. Beyond Carrier Frequency: A Consequence-Weighted and Interaction-Aware Framework for Genetic Screening in Jewish Founder Populations. Part I draft supplied by authors.

2. Gregg AR, et al. Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics. Genetics in Medicine. 2021.

3. American College of Obstetricians and Gynecologists. Carrier Screening for Genetic Conditions; Carrier Screening in the Age of Genomic Medicine. Committee opinions and guidance.

4. ARUP Consult. Cystic Fibrosis (CFTR) Expanded Variant Panel. Test fact sheet with carrier frequencies and testing limitations.

5. Centers for Disease Control and Prevention. Data and Statistics on Sickle Cell Disease. Updated public-health statistics.

6. World Health Organization, Regional Office for Africa. Sickle Cell Disease and hemoglobin-disorder key facts.

7. GeneReviews. Alpha-Thalassemia. NCBI Bookshelf.

8. UCSF Northern California Comprehensive Thalassemia Center. Hemoglobin E background data and carrier rates by Asian subgroup.

9. GeneReviews. GJB2-Related Autosomal Recessive Nonsyndromic Hearing Loss. NCBI Bookshelf.

10. GeneReviews. Medium-Chain Acyl-Coenzyme A Dehydrogenase Deficiency. NCBI Bookshelf.

11. MedlinePlus Genetics. Glucose-6-Phosphate Dehydrogenase Deficiency.

12. gnomAD. Genetic Ancestry and ancestry-stratified allele-frequency resources.

13. National Human Genome Research Institute. Familial Mediterranean Fever overview.

14. GeneReviews. Wilson Disease. NCBI Bookshelf.

15. NCBI Bookshelf / StatPearls and other clinical summaries for phenylketonuria carrier-frequency context.