DNAChecker Whitepaper

Overview

DNAChecker is a lightweight, privacy-preserving genomic insights tool designed to run entirely on a user's device. It enables individuals to upload raw genetic data files—such as those from 23andMe, AncestryDNA, or other DTC genotyping services—and receive accurate interpretations of key traits, health conditions, and wellness markers.

This whitepaper outlines the core methodology behind DNAChecker, including SNP parsing, trait interpretation, phasing considerations, confidence handling, and the biological context in which these checks operate.

Scientific Foundation

DNAChecker interprets Single Nucleotide Polymorphisms (SNPs) that have been associated in scientific literature and genome-wide association studies (GWAS) with various traits, wellness factors, and disease risk markers. SNPs are the most common type of genetic variation among people and represent a single base-pair change at a specific position in the genome.

Each check is based on curated sets of SNPs relevant to a phenotype (e.g., caffeine metabolism, lactose intolerance, obesity risk), including:

- The rsID (e.g., rs9939609)
- The gene context (e.g., FTO)
- The risk or trait-associated allele (e.g., A)
- An optional zygosity consideration (i.e., whether one or both alleles need to be risk variants)

Key Concepts

1. Genotype Classification

Each SNP in a user's DNA data is matched and interpreted based on the following categories:

- **Homozygous Risk**: Both alleles (e.g., AA) match the risk allele. Often considered the strongest indicator of trait or condition presence.
- **Heterozygous Risk**: One allele matches the risk allele (e.g., AT). This may indicate a partial or reduced expression of the associated trait.

- **Homozygous Non-risk**: Both alleles differ from the risk allele (e.g., TT). Suggests minimal or no predisposition for the condition.
- **No-call/Not Found**: SNP is absent in the data or unreported. Treated neutrally but flagged.

2. Zygosity Weighting

Certain conditions have dose-dependent relationships with specific alleles. For example:

- **Obesity and the FTO gene (rs9939609)**: Homozygous carriers (AA) may exhibit significantly higher BMI than heterozygous (AT) or non-carriers (TT).

DNAChecker handles these cases using a zygosity model to display results in a nuanced way (e.g., “Homozygous risk: High predisposition” vs. “Heterozygous: Moderate predisposition”).

Technical Approach

1. Parsing and Matching

- DNAChecker parses raw text-based genotype files, identifying rsIDs and corresponding genotype pairs.
- For each rsID, the tool checks whether the user carries the associated risk allele and its zygosity.
- Multi-SNP traits are aggregated into a single score or description based on the number and severity of risk alleles found.

2. Local Trait Database (checks.json)

- The core of DNAChecker’s functionality is driven by a local JSON file containing phenotype definitions.
- Each trait or condition includes:
    - name: The human-readable label.
    - genes: A list of associated genes (e.g., FTO, LCT, CYP1A2).
    - snps: An array of SNP objects, each with:
        - rsid
        - riskAllele
        - gene
        - (optional) impactWeight or zygosityType

This approach enables rapid analysis that can be reproduced for sharing.

3. Risk Aggregation

For multi-SNP traits, the tool uses one of three aggregation strategies:

- **Binary Match**: Any match to a risk allele signals increased risk.
- **Weighted Sum**: SNPs contribute a score based on homozygosity and known effect size.
- **Threshold Model**: A user must match a minimum number of risk alleles to trigger a "positive" classification.

Application Examples

1. Caffeine Metabolism

- SNP: rs762551
- Gene: CYP1A2
- Risk Allele: A (slow metabolizer)
- Outcome: Determines whether a person is a fast or slow caffeine metabolizer, impacting alertness and anxiety response.

2. Alzheimer's Disease Risk

- SNPs: rs429358, rs7412
- Gene: APOE
- Genotype: E4 allele presence formed by combinations of C/T alleles
- Outcome: E4/E4 homozygotes show a greatly increased risk, while E3/E4 show moderate risk.

3. Predisposition for Narcissism

- SNP: rs53576
- Gene: OXTR (Oxytocin metabolism)
- Risk Allele: A
- Outcome: Homozygous A indicates likely narcissistic tendencies.

Accuracy and Limitations

Accuracy

DNAChecker uses high-confidence SNP-phenotype associations drawn from:

- GWAS Catalog
- SNPedia
- OpenSNP
- PubMed-indexed literature

Trait associations are filtered based on statistical significance, population size, and replication across cohorts.

Limitations

- **Trait Complexity**: Many traits are polygenic and environment-influenced. A risk allele does not guarantee phenotype expression.
- **Ethnic Variability**: Risk factors validated in one ancestry group may not apply universally.
- **Incomplete Data**: Some SNPs may be missing from older or partial genotype files.

Peer Comparisons and Family Sharing

DNAChecker introduces a secure and optional *Comparisons* feature that allows users to privately share and compare their genetic reports with selected individuals—such as family members, partners, or friends. This functionality enhances the interpretability of genetic traits and conditions by enabling users to observe patterns across shared ancestry or relationships.

Key Features:

- **Invite-Based Sharing**: Users can invite others to view specific parts of their genetic results, with full control over what is shared.
- **Side-by-Side SNP View**: DNAChecker presents allele-by-allele comparisons of relevant SNPs between individuals, highlighting differences and shared risk factors.
- **Family Trait Exploration**: Families can track the inheritance of traits like behavioral tendencies or cancer risk across generations.
- **Condition Concordance**: Compare risk profiles for complex traits like obesity, Alzheimer's risk, or vitamin metabolism to assess potential shared genetic predispositions.

Conclusion

DNAChecker represents a new paradigm in personal genomics: one that is privacy-first, user-controlled, and scientifically grounded. By combining transparent SNP-based analysis with local processing and intuitive visualization, DNAChecker empowers individuals to explore their genetic predispositions without the need to surrender data to centralized services.

As the demand for personalized health insights grows, so does the importance of tools that respect user autonomy. DNAChecker fills this gap by enabling users to ask and answer meaningful questions about their DNA—whether they're curious about caffeine metabolism, trait inheritance, or risk factors for chronic conditions.

The platform is built to scale scientifically, technically, and ethically, supporting ongoing additions of new markers, evolving interpretation frameworks (e.g., polygenic scores), and peer-to-peer comparison features. Its offline-first design makes it accessible, trustworthy, and inherently secure.

By democratizing access to reliable genetic interpretation, DNAChecker serves as both an educational resource and a stepping stone toward a future of proactive, data-literate health exploration—where each person remains the primary guardian of their genomic identity.