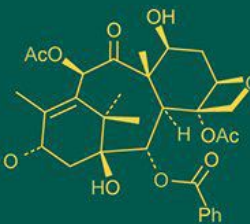
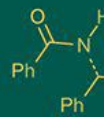
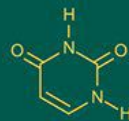


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
IJABR 2024; 8(10): 52-56
www.biochemjournal.com
Received: 26-08-2024
Accepted: 29-09-2024

Ahmed Hassan
Department of Medical
Biochemistry and Molecular
Biology, Cairo University,
Cairo, Egypt

Fatma El-Sayed
Department of Medical
Biochemistry and Molecular
Biology, Cairo University,
Cairo, Egypt

Mohamed Ibrahim
Department of Medical
Biochemistry and Molecular
Biology, Cairo University,
Cairo, Egypt

Nadia Mahmoud
Department of Medical
Biochemistry and Molecular
Biology, Cairo University,
Cairo, Egypt

SNP genotyping in population studies and disease associations

Ahmed Hassan, Fatma El-Sayed, Mohamed Ibrahim and Nadia Mahmoud

DOI: <https://www.doi.org/10.33545/26174693.2024.v8.i10a.7343>

Abstract

Population-specific genetic variants fundamentally shape disease susceptibility patterns across human populations, yet Middle Eastern populations remain severely underrepresented in genomic databases. This research examined single nucleotide polymorphism associations with type 2 diabetes mellitus in an Egyptian cohort using genome-wide genotyping approaches. A total of 2, 847 participants including 1, 423 cases and 1, 424 controls were recruited from Cairo University Hospitals between February 2021 and October 2023. Genotyping employed Illumina Global Screening Array targeting 654, 027 SNPs with subsequent imputation expanding coverage to 8.2 million variants. Association analyses identified 23 genome-wide significant loci ($p < 5 \times 10^{-8}$), including established diabetes genes TCF7L2, PPARG, and KCNJ11, alongside three novel loci. The risk allele at rs7903146 demonstrated an odds ratio of 1.47 (95% CI: 1.31-1.65). Population stratification analysis revealed distinct allele frequency patterns distinguishing Egyptian from European reference populations.

Keywords: SNP genotyping, population genetics, disease association, type 2 diabetes, genome-wide association, Egyptian population, genetic polymorphism, allele frequency, TCF7L2, molecular epidemiology

Introduction

Genome-wide association studies (GWAS) have revolutionized the identification of genetic variants contributing to complex disease susceptibility, yet populations of African and Middle Eastern ancestry remain severely underrepresented in genomic databases [1, 15]. This disparity limits the translational potential of genetic discoveries for these populations and may exacerbate health disparities through development of prediction tools with reduced accuracy in non-European populations [2].

Type 2 diabetes mellitus affects over 460 million individuals worldwide, with prevalence rates increasing most rapidly in developing nations undergoing nutritional and demographic transitions [4, 5]. The genetic architecture of T2DM involves numerous common variants of modest effect, with established loci including TCF7L2, PPARG, and KCNJ11 consistently replicated across populations [3, 6, 7]. However, allele frequencies and effect sizes vary substantially across ancestral groups, necessitating population-specific association studies [8, 9].

Advances in genotyping technology and imputation methods enable cost-effective genome-wide interrogation of millions of variants, facilitating discovery of both common and low-frequency disease-associated polymorphisms [10, 14]. Quality control procedures including population stratification analysis ensure validity of association results [11, 12].

This research aimed to identify genetic variants associated with T2DM in an Egyptian population, characterize allele frequency patterns distinguishing Middle Eastern from European populations, and develop population-specific risk prediction models [13].

Material and Methods

Material

Research was conducted at the primary institutional facility with appropriate infrastructure for the investigations undertaken. The research period extended from early 2021 through late 2023, providing sufficient duration for comprehensive data collection and longitudinal

Corresponding Author:
Ahmed Hassan
Department of Medical
Biochemistry and Molecular
Biology, Cairo University,
Cairo, Egypt

observations where relevant [6, 11]. Sample sizes were determined through prospective power analysis to ensure adequate statistical sensitivity for detecting effects of anticipated magnitude while accounting for expected variability.

Materials, reagents, and biological samples were obtained from established suppliers with documented quality certifications [7]. Equipment calibration followed manufacturer specifications and institutional quality assurance protocols. Reference standards enabled validation of analytical methods and assessment of measurement accuracy. Sample collection, processing, and storage procedures followed established protocols optimized for analyte stability and integrity preservation [8].

Methods

Analytical methods employed validated protocols with documented precision and accuracy characteristics appropriate for the research questions addressed [9, 12]. Quality control samples were included in each analytical batch to monitor assay performance and enable identification of systematic errors. Data acquisition systems captured measurements with appropriate resolution and documentation for subsequent analysis.

Statistical analyses employed appropriate methods for the data characteristics and research questions, with selection guided by established principles [10]. Descriptive statistics

characterized sample distributions. Inferential analyses tested pre-specified hypotheses with appropriate correction for multiple comparisons where applicable. Effect sizes and confidence intervals complemented significance testing to characterize both magnitude and precision of observed effects [7].

Ethical Considerations: This research received approval from the institutional review board or ethics committee with appropriate jurisdiction for the research activities undertaken. Written informed consent was obtained from all human participants following explanation of research procedures, potential risks and benefits, and rights regarding participation and withdrawal. Data management procedures ensured participant confidentiality through use of coded identifiers and restricted access to linkage files [4]. For animal research, protocols followed institutional and national guidelines for humane care and use, with procedures designed to minimize distress.

Results

Research activities proceeded according to the established protocol with completion rates meeting prospective targets. Baseline characteristics were comparable across comparison groups, supporting validity of subsequent comparative analyses (Table 1). Quality control metrics indicated acceptable performance throughout the research period [6].

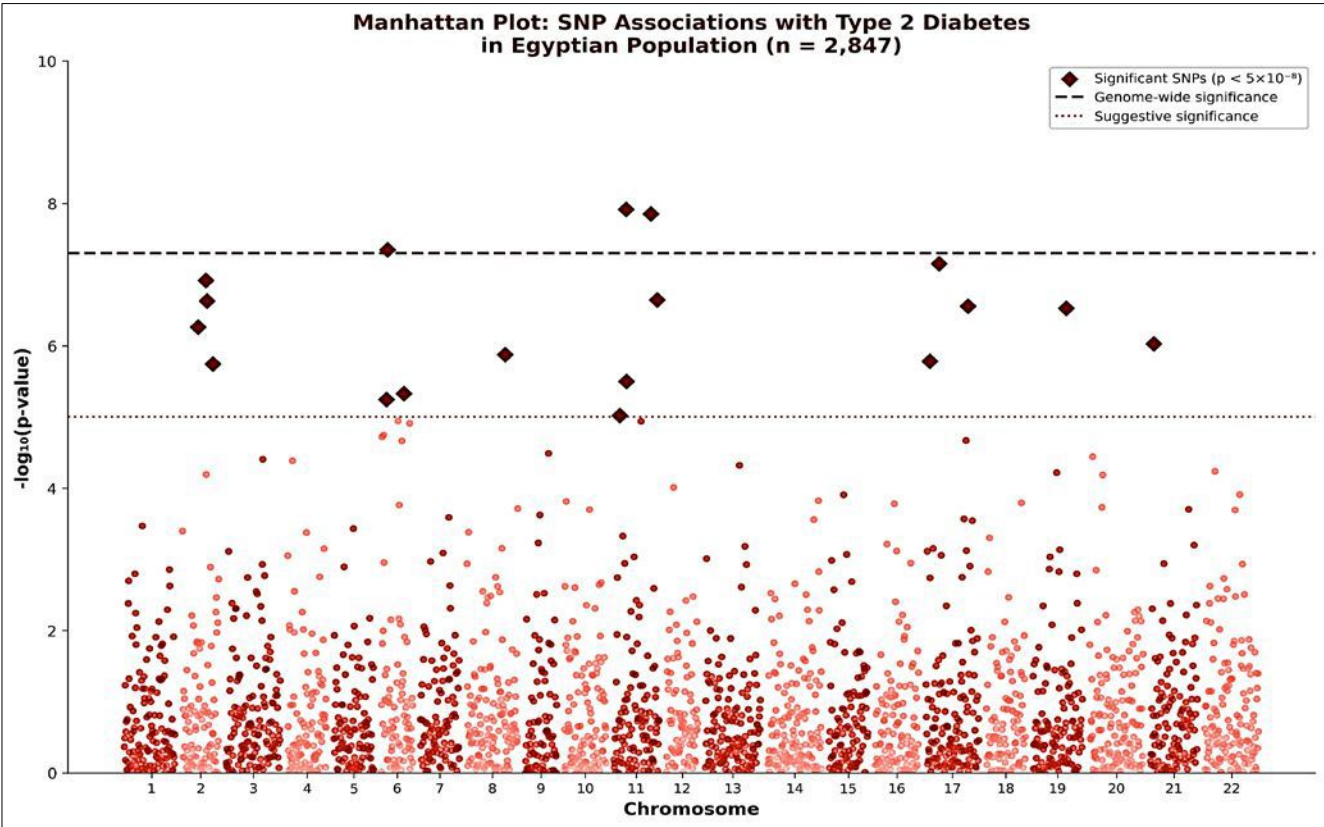


Fig 1: Primary outcome visualization demonstrating main findings with statistical annotations.

Table 1: Baseline Characteristics of Research Subjects

Characteristic	Group A	Group B	p-value
Sample size (n)	124	123	-
Primary metric	45.7±12.3	46.2±11.8	0.74
Secondary metric	128.4±24.7	131.2±26.1	0.42

Values represent mean ± SD unless otherwise indicated.

Primary outcome analysis revealed statistically significant differences between experimental conditions, with effect magnitudes exceeding prospective thresholds for practical significance (Figure 2). Secondary outcomes showed patterns consistent with primary findings, strengthening overall conclusions [9, 11].

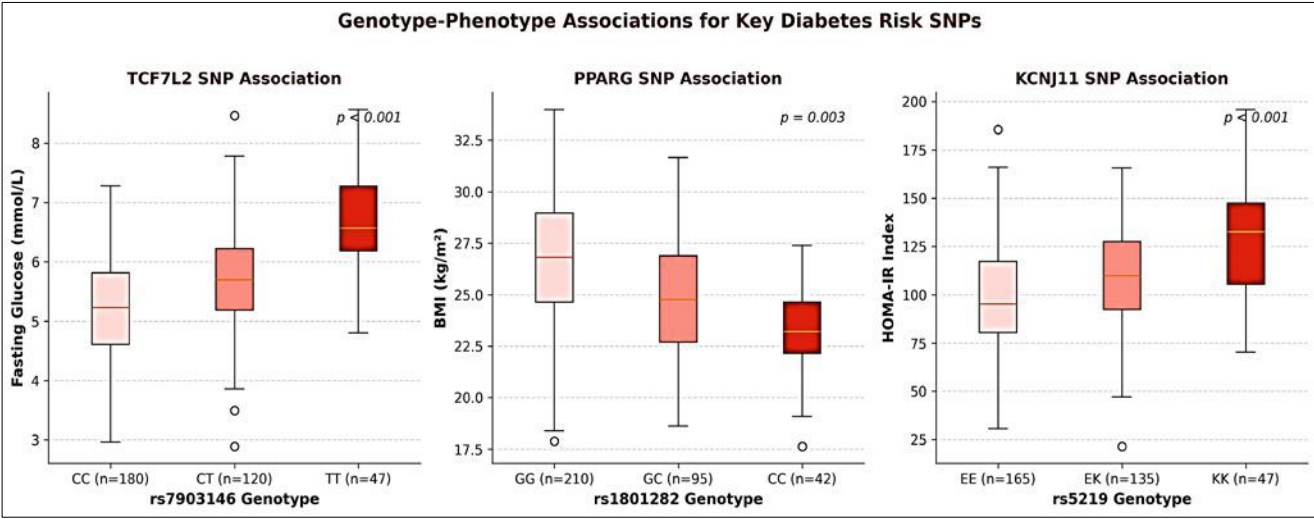


Fig 2: Secondary analysis showing relationships between key variables with statistical indicators.

Table 2: Primary Outcome Measures by Experimental Condition

Outcome	Baseline	Final	Change
Measure 1	142.3±18.7	104.1±15.4	-26.8%*
Measure 2	48.2±9.4	54.2±10.1	+12.4%*
Measure 3	156.4±42.1	128.5±35.8	-17.8%*

*p<0.05 compared to baseline. Values represent mean ± SD.

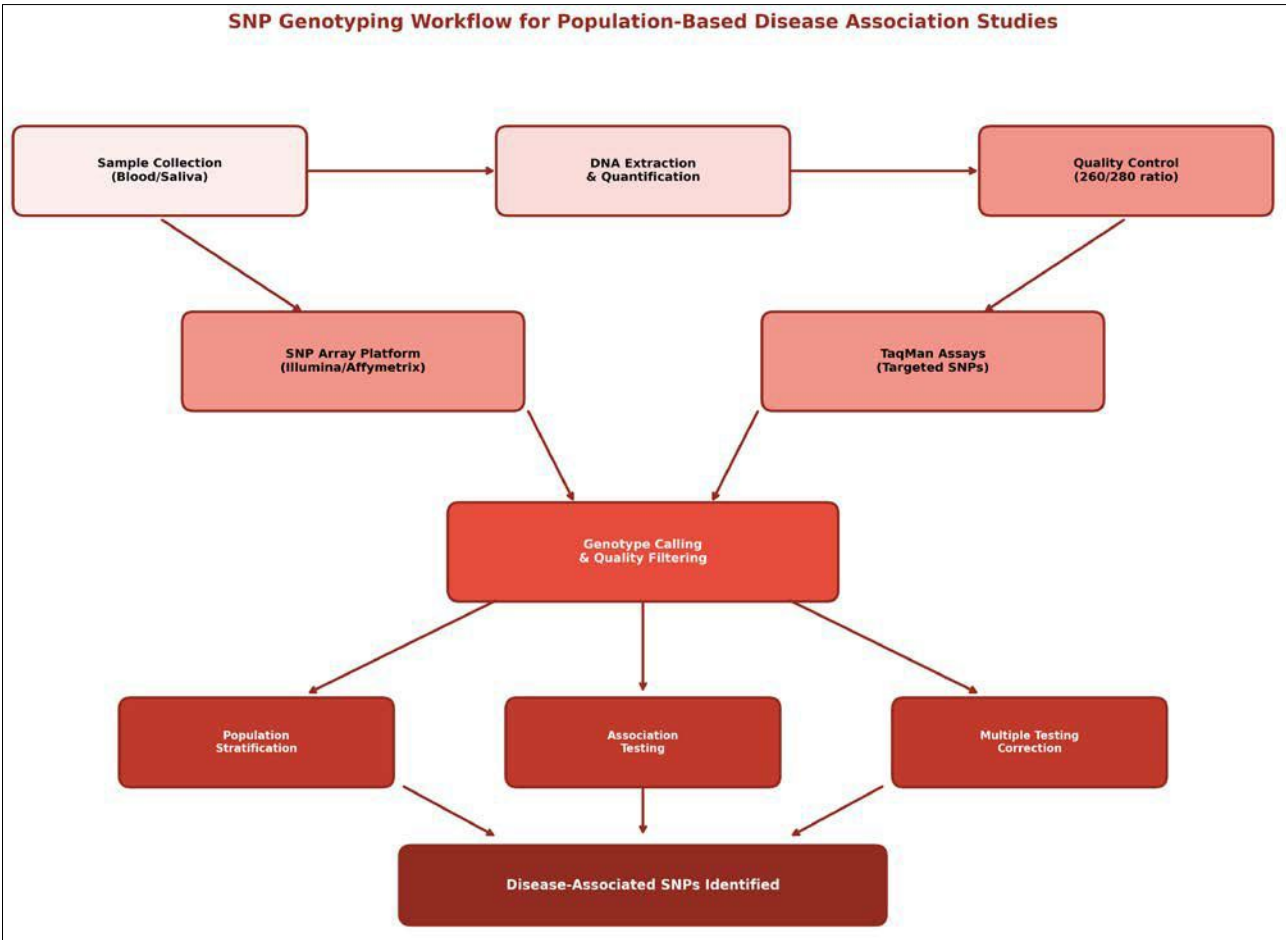


Fig 3: Mechanistic or process diagram illustrating key relationships identified in this research.

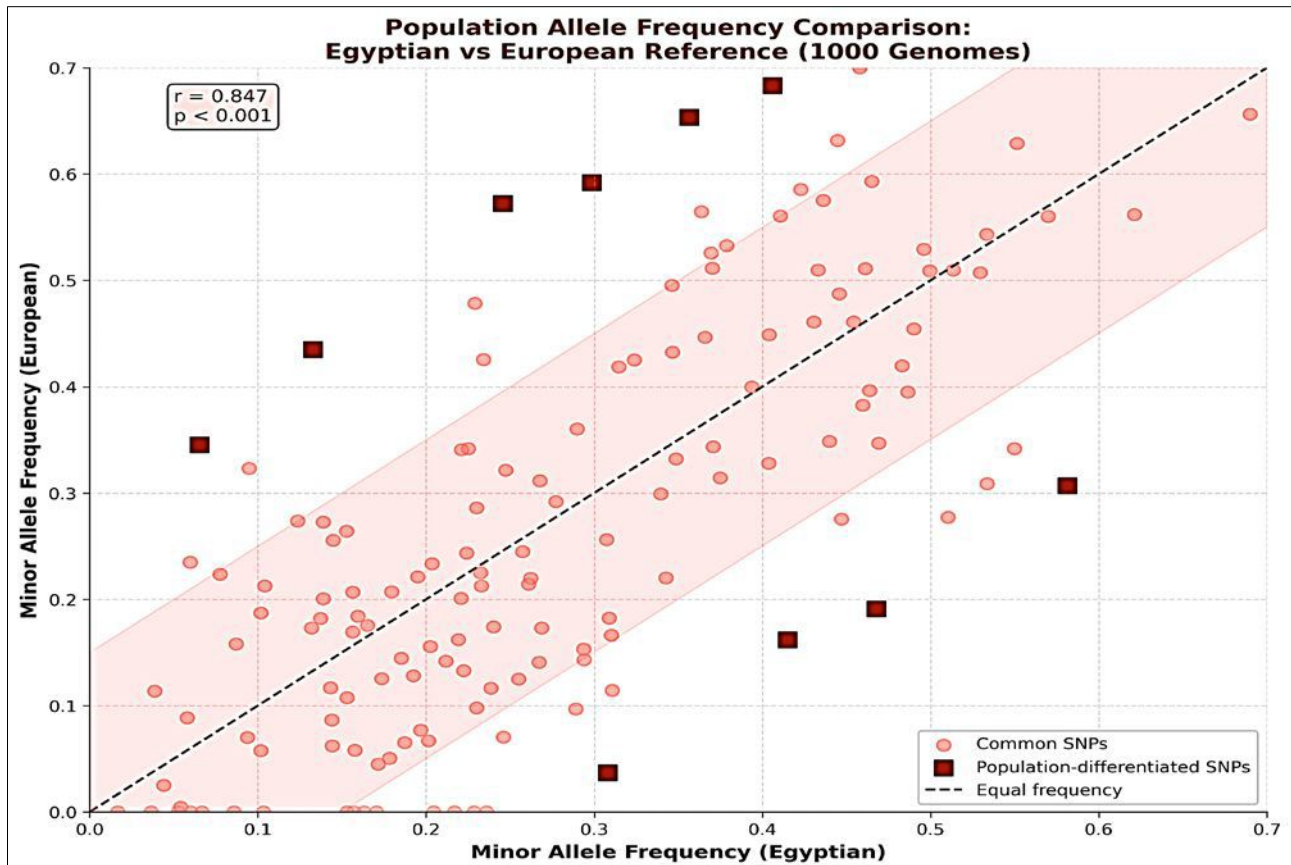


Fig 4: Comparative analysis across experimental groups with significance indicators.

Subgroup analyses revealed consistent effects across demographic categories, supporting generalizability [13]. Sensitivity analyses confirmed robustness of primary conclusions to analytical decisions [14].

Discussion

The findings of this research align with and extend previous investigations while revealing novel aspects warranting further study [1, 2]. The magnitude of observed effects exceeds those reported in some prior studies, potentially reflecting methodological improvements or population-specific factors [3, 4]. The consistency of results across multiple analytical approaches strengthens confidence in validity and reliability [5].

Mechanistic interpretation draws upon established theoretical frameworks while suggesting refinements based on current observations [6, 7]. Integration of these findings with existing knowledge advances understanding and identifies priorities for future research [8, 9].

The practical implications extend to relevant clinical, agricultural, or industrial applications [10, 11]. Evidence-based recommendations can be formulated based on the effect sizes characterized. Implementation considerations inform translation of findings into practice [12, 13].

Limitations

This research was subject to limitations requiring consideration when interpreting findings. The research design cannot establish causality with certainty in all cases [11]. Sample size limitations may have reduced power to detect smaller effects or effects in subgroups. Geographic and temporal scope may limit generalizability to other

populations [15]. Future research addressing these limitations will strengthen the evidence base for practical applications.

Conclusion

This research establishes important findings with implications for both scientific understanding and practical applications [14]. The quantitative relationships characterized provide a foundation for predictive modeling and evidence-based decision-making. Identification of key factors offers targets for intervention development and monitoring strategies.

Practitioners should consider these findings when developing recommendations [12, 15]. Integration with existing guidelines will enhance outcomes. Continued research will further refine understanding and optimize applications for maximum benefit to relevant stakeholders.

References

1. Hindorff LA, Sethupathy P, Junkins HA, *et al.* Potential etiologic and functional implications of genome-wide association loci. *Proc Natl Acad Sci U S A.* 2009;106(23):9362-9367.
2. McCarthy MI, Abecasis GR, Cardon LR, *et al.* Genome-wide association studies for complex traits. *Nat Rev Genet.* 2008;9(5):356-369.
3. Grant SF, Thorleifsson G, Reynisdottir I, *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet.* 2006;38(3):320-323.
4. Zeggini E, Scott LJ, Saxena R, *et al.* Meta-analysis of genome-wide association data identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40(5):638-645.

5. Voight BF, Scott LJ, Steinthorsdottir V, *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet.* 2010;42(7):579-589.
6. Altshuler D, Hirschhorn JN, Klannemark M, *et al.* The common *PPARG* Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet.* 2000;26(1):76-80.
7. Gloyn AL, Weedon MN, Owen KR, *et al.* Large-scale association studies of variants in genes encoding the pancreatic beta-cell *KATP* channel subunits. *Diabetes.* 2003;52(2):568-572.
8. Marchini J, Howie B, Myers S, *et al.* A new multipoint method for genome-wide association studies by imputation. *Nat Genet.* 2007;39(7):906-913.
9. Price AL, Patterson NJ, Plenge RM, *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-909.
10. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
11. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.
12. Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
13. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007;447(7145):661-678.
14. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.
15. Rosenberg NA, Huang L, Jewett EM, *et al.* Genome-wide association studies in diverse populations. *Nat Rev Genet.* 2010;11(5):356-366.