DELTA ECG: A GENETIC PERSPECTIVE

Zachary Levine
Faculty of Mathematics and Computer Science
Weizmann Institute of Science
Rehovot, Israel
{zachary.levine}@weizmann.ac.il

Hagai Rossman
Pheno.Ai
Tel Aviv, Israel
{hagai}@pheno.ai

Eran Segal
Faculty of Mathematics and Computer Science
Weizmann Institute of Science
Rehovot, Israel
{eran.segal}@weizmann.ac.il

METHODS

We analyzed 12-lead ECGs from The Human Phenotype Project, sampling each 10 s recording at 1000 Hz and segmenting each ECG into two four-second windows. In keeping with current state-of-the-art, we trained a contrastive convolutional/transformer neural network to embed ECG segments, enforcing similarity between embeddings from the same individual (both across and within visits) while maximizing separation from different individuals Kiyasseh et al. (2021); Diamant et al. (2022). Training curves are shown in Figure 1. As illustrated, embeddings from an individual's two visits cluster distinctly from a null distribution of unmatched visits, confirming that the self-supervised model learns individualized ECG signatures. After model validation, we computed a per-patient scalar by taking the cosine similarity between baseline and follow-up embeddings, defining this as *Delta ECG*—a single scalar quantifying patient-specific ECG changes over two years.

As input for genetic analyses, we used an imputed (from 0.6x) SNP array by the external third-party Gencove, removing first-degree relatives and adjusting for population structure through the use of the top 10 principal components as covariates. For transcriptomics, PBMC RNA-seq data from the same samples were normalized to counts-per-million per gene, per sample. A genome-wide association study (GWAS) on *Delta ECG* identified five genome-wide significant loci, highlighting a genetic basis for individual cardiovascular state changes (Fig. 2ii). Notably, none of these SNPs had been previously reported in any GWAS.

Among these, we identified variants in ***ARID5B*** (rs139222531, $P < 4 \times 10^{-9}$), previously associated with CVD, suggesting that *Delta ECG* may capture genetic predisposition to long-term cardiovascular risk Kichaev et al. (2019). ***SLC30A8*** (rs117871919, $P < 3 \times 10^{-8}$), a gene encoding a zinc transporter linked to glycemic variability, HbA1c, and BMI Masahiro et al. (2018); Saxena et al. (2007), was also identified. Given the established link between type 2 diabetes (T2D) and CVD Pearson-Stuttard et al. (2021), these findings suggest that *Delta ECG* shares genetic associations with both CVD and its key risk factors.

To further assess the biological relevance of *Delta ECG*, we examined its predictive capacity for gene expression using overrepresentation analysis of PBMC RNA-seq data. In men, our embeddings significantly predicted the expression of 57 genes ($P < 0.05$), with nine genes meeting a stricter threshold ($P < 0.01$). Predictions in women were weaker, with only nine genes associated at $P < 0.05$, and no significantly enriched pathways after multiple hypothesis correction. Notably, results in men these included genes encoding mitochondrial respiratory chain components (*MT-RNR1*, *MT-RNR2*, *MT-TV*), which agrees with previous work implicating oxidative phosphorylation in cardiovascular function Chistiakov et al. (2018).

Overall, by integrating deep learning-derived ECG representations with genetic and transcriptomic data, *Delta ECG* provides a scalable and label-free framework for quantifying temporal changes in cardiac health. Our findings demonstrate that leveraging deep learning for multi-omic analysis can uncover clinically relevant cardiovascular dynamics, offering a path toward personalized risk

assessment and monitoring. Future work should validate this approach across diverse populations to maximize its utility in precision cardiovascular medicine.

RESULTS

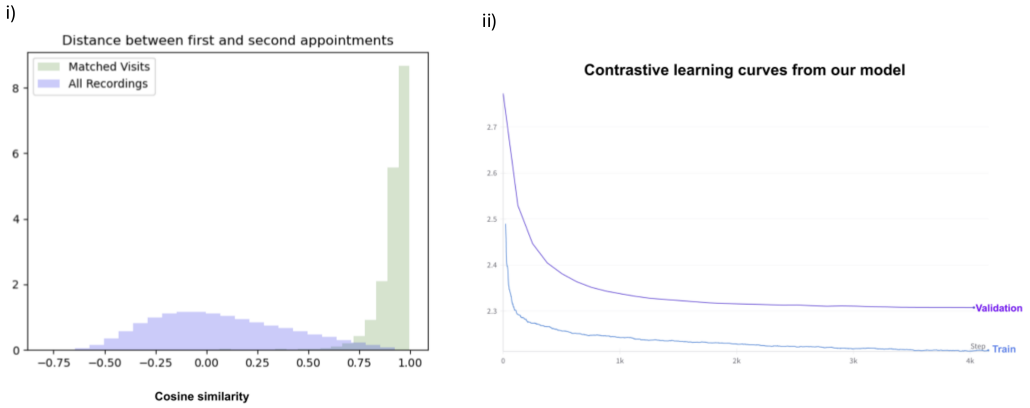
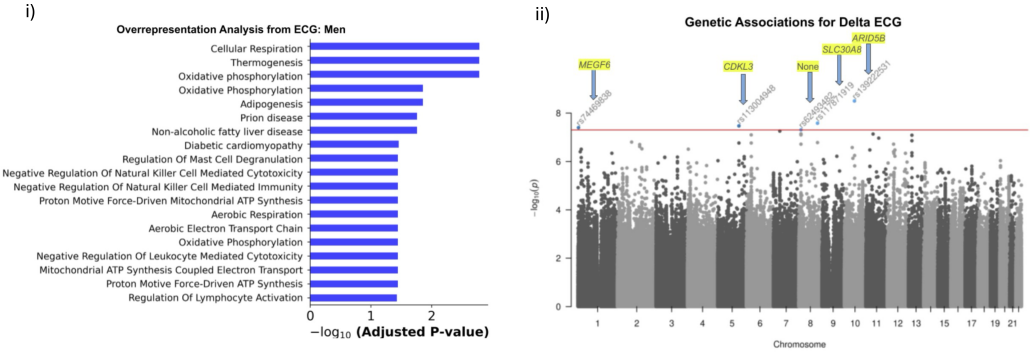


Figure 1: **Validating delta ECG** (i) Distribution of cosine similarity (Delta ECG) for matched vs. all unmatched visits, indicating a strong patient-specific signal. (ii) Contrastive learning curves, showing training (blue) and validation (purple) losses converging over steps.



iii)

Table 1: Significant Hits for Delta ECG. Delta ECG shares a genetic basis with CVD and its known risk factors and correlates.

SNP	Chromosome	P	Gene	Previous Associations
rs139222531	10	3.2e-9	*ARID5B*	CVD
rs117871919	8	2.66e-8	*SLC30A8*	A1C, BMI, T2D
rs113004948	5	3.42e-8	*CDKL3*	Eosinophil count
rs74469838	1	4.13e-8	*MEGF6*	Systemic mastocytosis
rs62493482	8	4.9e-8	*N.C.*	None

Figure 2: **Genetic Analyses. (i)** Overrepresentation analysis of genes significantly predicted by our ECG embeddings in men demonstrates enrichment for metabolic and immune-related pathways (e.g., mitochondrial respiration, mast cell degranulation). **(ii)** A Manhattan plot reveals five genomic loci surpassing the genome-wide significance threshold for *Delta ECG*, including *ARID5B* (associated with CVD), *SLC30A8* (glycemic control), *CDKL3* (eosinophil count), and *MEGF6* (mast cell pathways). **(iii)** A table highlighting these key variants, reinforcing that longitudinal ECG changes (captured via our self-supervised *Delta ECG* measure) share genetic underpinnings with CVD risk factors and immune processes.

MEANINGFULNESS STATEMENT

The genetic basis of cardiovascular disease remains largely unresolved. Existing deep-learning applications to electrocardiograms (ECGs) focus on single phenotypes or isolated timepoints Radhakrishnan et al. (2023); Libiseller-Egger et al. (2022); Wang et al. (2023), introducing biases. Here, we present *Delta ECG*, a metric based on pairs of embeddings that captures patient-specific changes in cardiovascular state over time. We demonstrate that this measure reliably differentiates within-patient variation from population-level differences and aligns with genome-wide significant loci for cardiovascular disease (CVD) and its risk factors.

REFERENCES

Dimitry A. Chistiakov, Tatiana P. Shkurat, Alexandra A. Melnichenko, Andrey V. Grechko, and Alexander N. Orekhov. The role of mitochondrial dysfunction in cardiovascular disease: a brief review. *Annals of Medicine*, 50(2):121–127, February 2018. ISSN 0785-3890. doi: 10.1080/07853890.2017.1417631.

Nathaniel Diamant, Erik Reinertsen, Steven Song, Aaron Aguirre, Collin Stultz, and Puneet Batra. Patient contrastive learning: a performant, expressive, and practical approach to ecg modeling. *PLOS Computational Biology*, 18(2):e1009862, February 2022. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009862. arXiv:2104.04569 [cs, eess].

Gleb Kichaev, Gaurav Bhatia, Po-Ru Loh, Steven Gazal, Kathryn Burch, Malika K. Freund, Armin Schoech, Bogdan Pasaniuc, and Alkes L. Price. Leveraging polygenic functional enrichment to improve gwas power. *The American Journal of Human Genetics*, 104(1):65–75, January 2019. ISSN 0002-9297. doi: 10.1016/j.ajhg.2018.11.008.

Dani Kiyasseh, Tingting Zhu, and David A. Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients. (arXiv:2005.13249), May 2021. URL `http://arxiv.org/abs/2005.13249`. arXiv:2005.13249 [cs, eess, stat].

Julian Libiseller-Egger, Jody E. Phelan, Zachi I. Attia, Ernest Diez Benavente, Susana Campino, Paul A. Friedman, Francisco Lopez-Jimenez, David A. Leon, and Taane G. Clark. Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Scientific Reports*, 12(1):22625, December 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-27254-z.

Kanai Masahiro, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Kubo M, Okada Y, and Kamatani Y. Genetic analysis of quantitative traits in the japanese population links cell types to complex human diseases. *Nature genetics*, 50(3), March 2018. ISSN 1546-1718. doi: 10.1038/s41588-018-0047-6. URL `https://pubmed.ncbi.nlm.nih.gov/29403010/`.

Jonathan Pearson-Stuttard, James Bennett, Yiling J Cheng, Eszter P Vamos, Amanda J Cross, Majid Ezzati, and Edward W Gregg. Trends in predominant causes of death in individuals with and without diabetes in england from 2001 to 2018: an epidemiological analysis of linked primary care records. *The Lancet Diabetes Endocrinology*, 9(3):165–173, March 2021. ISSN 2213-8587. doi: 10.1016/S2213-8587(20)30431-9.

Adityanarayanan Radhakrishnan, Sam F. Friedman, Shaan Khurshid, Kenney Ng, Puneet Batra, Steven A. Lubitz, Anthony A. Philippakis, and Caroline Uhler. Cross-modal autoencoder framework learns holistic representations of cardiovascular state. *Nature Communications*, 14(1):2436, April 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-38125-0.

Richa Saxena, Benjamin F. Voight, Valeriya Lyssenko, Noël P. Burtt, Paul I. W. de Bakker, Hong Chen, Jeffrey J. Roix, Sekar Kathiresan, Joel N. Hirschhorn, Mark J. Daly, Thomas E. Hughes, Leif Groop, David Altshuler, Peter Almgren, Jose C. Florez, Joanne Meyer, Kristin Ardlie, Kristina Bengtsson Boström, Bo Isomaa, Guillaume Lettre, Ulf Lindblad, Helen N. Lyon, Olle Melander, Christopher Newton-Cheh, Peter Nilsson, Marju Orho-Melander, Lennart Råstam, Elizabeth K. Speliotes, Marja-Riitta Taskinen, Tiinamaija Tuomi, Candace Guiducci, Anna Berglund, Joyce Carlson, Lauren Gianniny, Rachel Hackett, Liselotte Hall, Johan Holmkvist, Esa Laurila, Marketa Sjögren, Maria Sterner, Aarti Surti, Margareta Svensson, Malin Svensson, Ryan Tewhey, Brendan Blumenstiel, Melissa Parkin, Matthew Defelice, Rachel Barry, Wendy Brodeur, Jody Camarata, Nancy Chia, Mary Fava, John Gibbons, Bob Handsaker, Claire Healy, Kieu Nguyen, Casey Gates, Carrie Sougnez, Diane Gage, Marcia Nizzari, Stacey B. Gabriel, Gung-Wei Chirn, Qicheng Ma, Hemang Parikh, Delwood Richardson, Darrell Ricke, and Shaun Purcell. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science (New York, N.Y.)*, 316(5829):1331–1336, June 2007. ISSN 1095-9203. doi: 10.1126/science.1142358.

Xin Wang, Shaan Khurshid, Seung Hoan Choi, Samuel Friedman, Lu-Chen Weng, Christopher Reeder, James P. Pirruccello, Pulkit Singh, Emily S. Lau, Rachael Venn, Nate Diamant, Paolo

Di Achille, Anthony Philippakis, Christopher D. Anderson, Jennifer E. Ho, Patrick T. Ellinor, Puneet Batra, and Steven A. Lubitz. Genetic susceptibility to atrial fibrillation identified via deep learning of 12-lead electrocardiograms. *Circulation: Genomic and Precision Medicine*, 16(4): 340–349, August 2023. doi: 10.1161/CIRCGEN.122.003808.