

健康・医療・行動データを個人単位で連結した 東芝企業コホートの医療・ヘルスケア分野への応用

Application of Integrated Health, Medical, and Behavioral Data from Corporate Employee Cohort to Medical and Healthcare Fields

山口 泰平 YAMAGUCHI Taihei

高齢化や慢性疾患の増加に伴う医療費の拡大抑制に加え、感染症流行への備えが社会課題となっている。これに対応するため、実世界で継続的に得られる健康・医療データを統合・解析し、得られた知見を医療・ヘルスケア分野に還元して予防医療支援や施策改善に役立てる基盤が求められる。

東芝は、国内の東芝グループ従業員約19,000人の健康・医療データに、勤怠や購買といった職域・行動データなどを個人単位で連結して蓄積した企業コホート（以下、東芝企業コホートと呼ぶ）を構築し、複数の研究機関と共同研究を進めてきた。新型コロナウイルス感染症（COVID-19）ワクチンの追加接種後副反応調査では、蓄積データに追加問診データを接続して遺伝要因を探索し、減塩介入RCT（ランダム化比較試験）では、個別化予防の実装可能性と課題を明らかにした。これらは、企業コホートが平時の予防医療と有事の迅速な要因解析を支える実践的基盤となり得ることを示唆する。

Preparation for infectious disease outbreaks, rising healthcare costs associated with an aging society, and the increase in chronic diseases have become issues of vital importance. To address these social issues, a platform supporting preventive healthcare and improving healthcare measures through the continuous integration and analysis of real-world data and the application of the resulting knowledge to medical and healthcare fields is essential.

Toshiba Corporation has constructed a platform based on a corporate cohort of approximately 19 000 domestic employees of the Toshiba Group by continuously and individually integrating health and medical data with workplace behavioral data, such as attendance records and purchasing data, and conducted joint research with multiple research institutions. A study on adverse reactions following COVID-19 vaccine booster shots enabled exploratory analysis of genetic factors by combining data accumulated on the platform with additional questionnaire data. Another study involving a randomized controlled trial (RCT) of sodium reduction clarified both the feasibility of and the challenges associated with implementing personalized prevention. These studies demonstrate the potential of the platform based on the corporate cohort to support both preventive healthcare services in normal times and rapid factor analysis during times of crisis.

1. まえがき

医療・健康のデータは、個人の生活の質だけでなく、医療・介護負担、更には社会機能の継続可否にも影響する。高齢化の進展に伴う慢性疾患の増加は、医療費の継続的な増大を通じて社会保障の持続可能性に影響を及ぼし、感染症流行は短期間で医療需要と社会活動の双方に大きな変化をもたらす。このような複合課題に対しては、単発調査や経験則だけにに基づく対応には限界があり、継続的に収集される実世界データを活用して、リスクの把握や、対象の層別化、介入設計、効果検証、改善などを反復する必要がある。

このようなニーズに対応するため、東芝は、国内の東芝グループ従業員約19,000人の、体質データ（ゲノム）及び健康・医療データ（健診・問診・レセプト（診療報酬明細書）など）に加え、勤怠（残業・休暇・欠勤・休職など）や購買といった職域・行動データを、個人単位で連結した東芝企業コホートを構築した（図1）。コホートは、共通した因子を持

ち、一定期間にわたって観察対象となる集団を指す。東芝企業コホートは、“同じ企業に属すること”を共通因子とした観察対象集団であり、観察対象と継続的に接点を持ちやすい点が特徴である。これにより、平時から健康関連データを長期に蓄積できるだけでなく、感染症流行や介入実施などのイベント発生時にアンケートなどの追加データを取得し、蓄積データと接続して迅速に検証へ移行できる。また、健診や勤怠などのデータが継続的に更新されるため、時系列での変化把握や、介入前後の比較、層別化の検証などを実装しやすい。これらの観点から、東芝企業コホートは“平時の予防医療”と“有事の迅速な要因解析”を両立させる基盤として重要である。

ここでは、東芝企業コホートの特徴を整理し、二つの活用研究を通して明らかになった迅速性・実装性について述べ、更に今後の適用見通しを示す。

なお、東芝企業コホートのデータの取得・利用は、利用目的・範囲を明確化した上で一人一人の本人同意を得て行

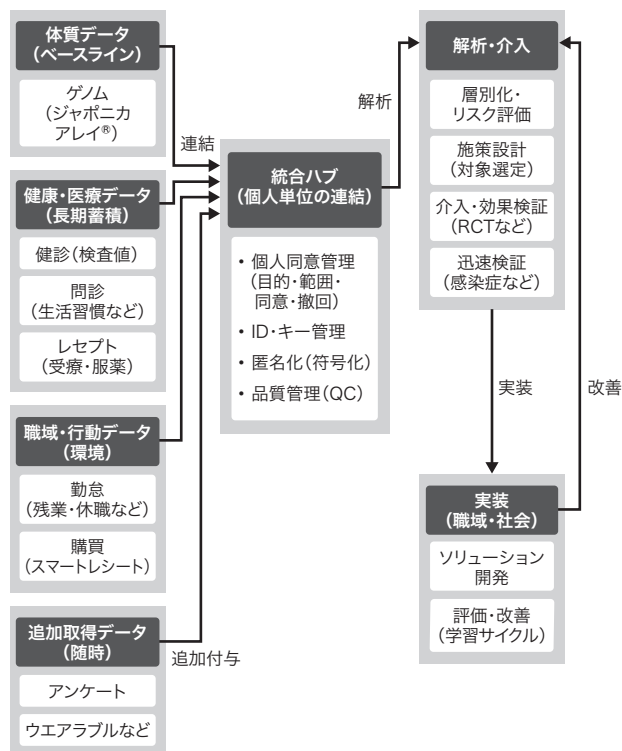


図1. 東芝企業コホートの概要

ゲノムなどの体質データ、健診・レセプトなどの長期時系列、勤怠・購買・追加取得データなどを、個人単位で連結する。

Overview of corporate cohort

い、必要に応じて追加の同意を取得する。個人を直接特定できる情報は研究利用データから分離し、研究用途に応じた仮名化・符号化などの保護措置を講じるとともに、研究は倫理審査などの手続きを経て実施する。更に、アクセス制御や監査などを通じて、安全管理を行う。このような情報ガバナンスを徹底した。

2. 東芝企業コホートの特徴

東芝企業コホートの特徴は、長期に追跡可能な約19,000人規模の母集団が形成されている点と、体質データ(ゲノム)と健康・医療データに加え、職域・行動データまでを個人単位で連結できる点である。“規模”と“多種データの統合”を同時に備えることにより、医療データ単独では捉えにくい健康変化を、就労状況や行動の文脈と合わせて多面的に解析できる。これにより、対象層別化、介入設計、効果検証、更に感染症などのイベント発生時の迅速な要因解析に資する。

2.1 データの範囲と規模

東芝企業コホートでは、ゲノムデータや健康・医療データを長期に蓄積している。ゲノムデータは、日本人集団に最適化されたSNP (Single Nucleotide Polymorphism: 一塩

基多型) アレイであるジャポニカアレイ®を用いて取得している。ジャポニカアレイ® NEOは、東北大学東北メディカル・メガバンク機構が構築した日本人全ゲノムリファレンスパネルなどを基に設計された日本人特化型アレイであり、約66万か所のSNPを搭載し、網羅的なジェノタイピングを可能にする^{(1), (2)}。また、健康・医療データは20年以上にわたり継続的に蓄積されており、時系列での健康変化や介入前後の比較を行う基盤となる。更に、東芝企業コホートでは、健康状態の背景要因や行動・環境要因を把握するために、職域・行動データも個人単位で連結している。これには、残業時間、休暇取得、欠勤・休職などが含まれる。購買データは、東芝テックの電子レシートサービス スマートレシートに基づく購買情報を活用し、本人同意の下で安全管理された形で解析に用いる。これにより、健康・医療データを生活行動や就労環境の情報と組み合わせで解釈できる。

2.2 東芝企業コホートの価値

東芝企業コホートの価値は、単に多種のデータ項目を保有していることではなく、多種データが個人単位で連結され、時系列で継続的に更新される点にある。この価値は、第一に、複数年にわたるデータ蓄積によって健康状態の推移やイベント前後の変化を捉えられるという“長期性”をもたらす。第二に、他項目のデータを個人単位で連結することで、多面的な要因解析を可能にするという“統合性”を持つ。第三に、必要時に追加アンケートや介入ログなどを取得し、蓄積データに接続して検証に利用できるという“拡張性”を備える。第四に、感染症などのように速度が価値を生む局面で、追加取得から接続・解析までの工程を短縮できるという“迅速性”を持つ。これら四つの特性の組み合わせが、東芝企業コホートを、研究だけでなく実装評価にも活用可能な基盤として位置付けている。

2.3 データ連結と品質管理

多種データを長期にわたり統合するためには、個人単位の連結キー管理、データ辞書やコード体系の整備、及び欠測値や外れ値への対応が不可欠である。特に、健診項目の変更、制度改定、レセプト記載の揺らぎ、勤怠制度の変更などは、時系列解析に影響する。このため、東芝企業コホートでは、データ取り込み時のバリデーション、時間軸の正規化、及び解析目的に応じた特徴量設計を行い、再現可能な解析を支える品質管理を重視している。例えば、直近3か月の残業時間、欠勤日数、購買パターンなどの特徴量を設計することにより、健康状態と行動・環境との関連を評価しやすくなった。こうした継続的なデータ整備と品質管理が、東芝企業コホート活用の前提条件である。データ連結と品質管理(QC)の流れを図2に示す。

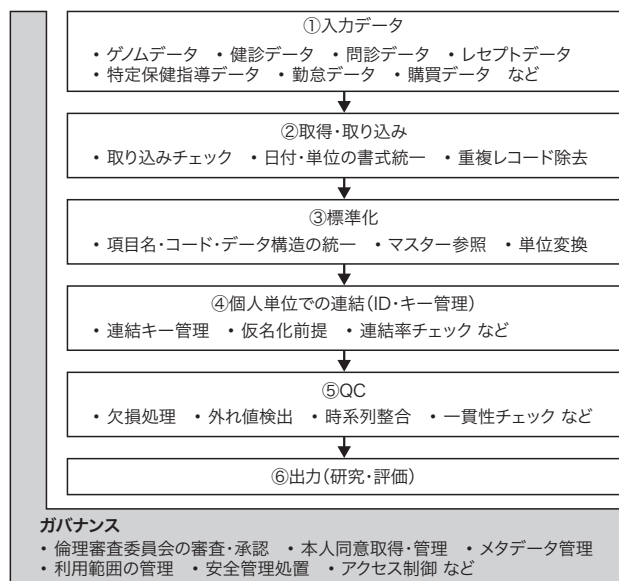


図2. データ連結とQCの流れ

ガバナンス管理の下、解析用データを作成する。

Flowchart of data linkage and quality control

3. 東芝企業コホートの活用研究

3章では、東芝企業コホートの蓄積データに、必要な追加情報を接続して、迅速に解析へ移行できるという特徴を活用した二つの研究について述べる。

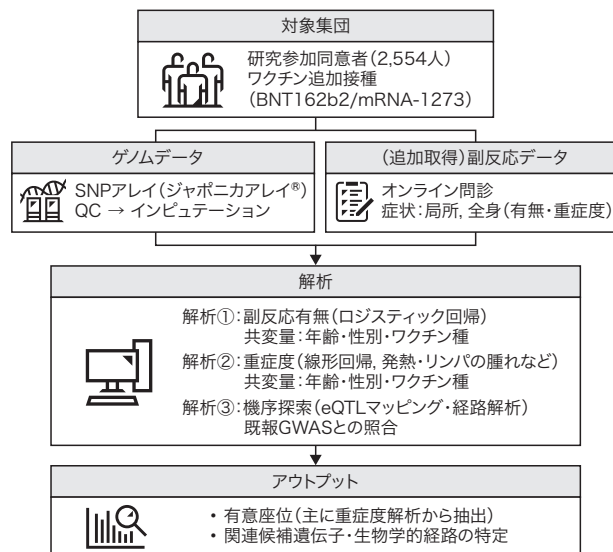
3.1 COVID-19ワクチン追加接種後副反応の遺伝要因の探索

mRNA(メッセンジャーリボ核酸)ワクチン3回目を接種した2,554人を対象として、副反応に関するゲノムワイド関連解析(GWAS)を実施した。感染症局面において、蓄積データに最小限の追加情報を接続することで、迅速に遺伝要因の探索へ移行できることを示す⁽³⁾。

3.1.1 研究目的・デザイン

この研究の目的は、COVID-19ワクチン追加接種後に生じる副反応の個人差に関与する遺伝要因を探索し、免疫応答の個体差理解と安全性評価、更に将来的なリスク層別化に資する手掛かりを得ることである。

対象は、BNT162b2又はmRNA-1273のワクチンを追加接種した2,554人である。接種後アンケート(オンライン問診)で副反応情報を取得し、蓄積されたゲノムデータと接続して解析した。アンケートでは、注射部位の発赤・腫脹(しゅちょう)・硬結・疼痛(とうつう)・灼熱(しゃくねつ)感・掻痒(そうよう)などの局所症状、発熱・倦怠(けんたい)感・頭痛・筋肉痛などの全身症状、及び欠勤・休職を把握した。重症度は、症状の性質に応じて段階化した。例えば、発赤、腫



eQTL: Expression Quantitative Trait Locus(発現量の形質遺伝子座)

*カッコ内の数字は被験者数

図3. COVID-19ワクチン追加接種後副反応の遺伝要因の探索

コホートデータに追加取得データを接続することで、大規模なクロス解析を迅速に実施できる。

Flowchart of exploratory analysis of genetic factors related to adverse reactions after COVID-19 vaccine booster shots

脹、硬結は直径に基づき、疼痛・掻痒・灼熱感は薬剤使用の可否などに基づき、発熱は自己申告体温に基づく0.5℃刻みのカテゴリーとして、それぞれ定義した。

ゲノムデータについては、属性不一致、近縁、回答欠損などを除外した上で、MAF(Minor Allele Frequency)や、欠損率、ハーディーワインベルグ平衡などの基準による品質管理を実施し、その後、全ゲノムインピュテーションを適用して、既存の遺伝子型情報を基に解析対象とする遺伝子変異の範囲を拡張した。統計解析では、症状の有無に対するロジスティック回帰と、重症度に対する線形回帰を基本とし、年齢・性別・ワクチン種を共変量とした。探索的解析としての性格を持つため、結果の解釈には慎重さが必要であるが、感染症流行下における迅速な知見創出の枠組みとして重要である。研究の概要を図3に示す。

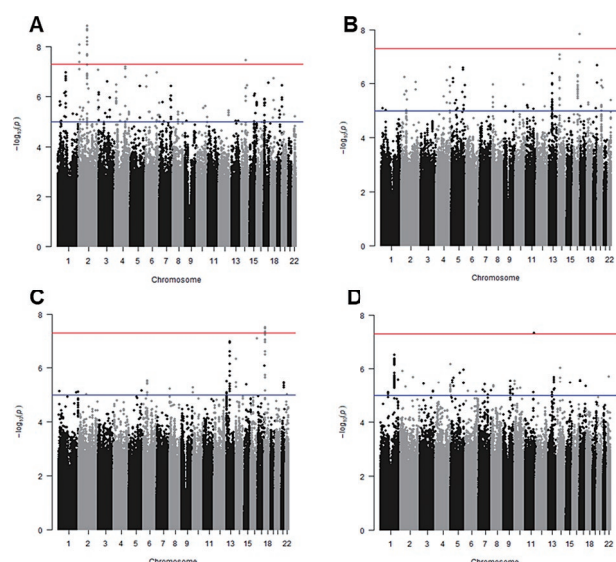
3.1.2 主な結果と考察

副反応の有無を目的変数としたGWASでは、いずれの症状についてもゲノムワイド有意水準に達する関連は認められなかった。一方、症状の重症度を目的変数としたGWASでは、リンパ節腫脹、掻痒、硬結、注射部位の発赤などに関して複数の有意座位が認められた。図4⁽³⁾は、4種類の副反応重症度に対するGWASの結果である。左上から、Aはリンパ節腫脹(サンプル数2,459人)、Bは掻痒(同2,503人)、Cは接種部位の硬結(同2,213人)、Dは接種部位の発赤(同

2,201人)である。

最も強い関連は、リンパ節腫脹の重症度と染色体2q12領域のlead variant rs76152249との関連であり、この領域にはIL1RL1, IL18R1, IL18RAPなどの免疫関連遺伝子が含まれていた。更に、発熱ではHLA(ヒト白血球抗原)を含むMHC(主要組織適合複合体)領域に関連する免疫応答経路が、掻痒ではNF- κ B(転写因子として働いたんばく質複合体)結合関連経路が示唆された。

これらの結果は、ワクチン追加接種後副反応において、“発現するか否か”よりも“どの程度強く現れるか”という重症度の方が、宿主遺伝要因の寄与を検出しやすい可能性を示している。副反応の有無は、回答バイアスやしきい値設定の影響を受けやすいのに対し、重症度は量的な差異を含むため、遺伝的感受性差が統計的に現れやすいと解釈できる。一方で、この研究は単一企業集団を対象とした探索的解析であり、副反応情報は自己申告に基づくため、測定誤差や選択バイアスの影響を完全には排除できない。また、独立した外部集団での再現性検証も今後の重要課題である。したがって、この研究の意義は、蓄積したゲノム情報と接種後アンケートという簡易な追加情報の組み合わせにより、感染症局面で迅速に関連探索を実施し得ることを示した点にある。



*Omae, Y. et al. Genome-wide association study of common side effects following COVID-19 booster vaccination in a cohort of corporate employees in Japan. Scientific Reports. 2025;15:12728. <https://doi.org/10.1038/s41598-025-90787-6> (Fig. 1)から転載。CC BY-NC-ND 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Chromosome: 染色体上の位置 P: 関連の優位性
—: ゲノムワイド有意水準, 5×10^{-8} —: 示唆的有意水準, 1×10^{-5} ●: ゲノム変異

図4. 4種類の副反応重症度に対するGWASの結果

重症度解析では複数の有意座位が認められ、特にリンパ節腫脹重症度では2q12領域に最も強い関連が認められた。

Results of Genome-wide association studies (GWAS) for four types of adverse reactions

3.2 遺伝情報とAIアプリを用いた減塩介入のランダム化比較試験

食塩の過剰摂取は高血圧及び心血管疾患の主要な修正可能リスク要因であり、減塩は重要な生活習慣介入である。一方で、食塩摂取に対する血圧上昇反応性(食塩感受性)には個人差があり、遺伝的背景が関与し得る。そこでこの研究では、遺伝情報に基づいて減塩が有効である可能性が高い層を想定し、AI健康アプリケーション(以下、AIアプリと略記)による行動変容支援を組み合わせた個別化介入の有効性を、職域をフィールドとするランダム化比較試験により検証した。東芝企業コホートの遺伝情報を活用した個別化予防の実装可能性を評価する⁽⁴⁾。

3.2.1 研究目的・デザイン

この研究の目的は、食塩感受性に関連する遺伝情報を本人に提示し、AIアプリによる食事記録支援と減塩に特化した情報提供を組み合わせることが、3か月間で食塩摂取量の低下に結び付くか否かを検証することである。試験デザインは3群ランダム化比較試験であり、被験者を、遺伝情報の提示・AIアプリ利用・減塩特化情報を組み合わせた介入群、介入を行わない対照群、並びに遺伝情報提示や減塩特化情報を伴わないAIアプリ限定群に割り付けた。被験者は20～65歳で、正常高値以上の血圧を満たし、AGT M235TのTアレルを一つ以上持つ者とした。主要アウトカムは、随時尿からインターソルト式による1日食塩摂取量であり、二次アウトカムとして、BMI(体格指数)、行動変容意図、収縮期・拡張期

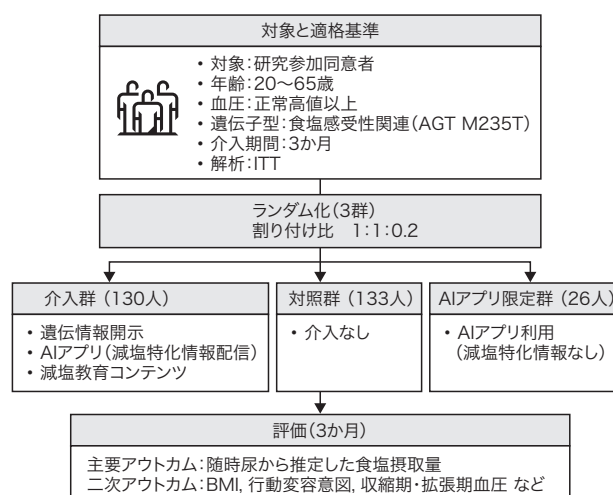


図5. 遺伝情報とAIアプリを用いた減塩介入のランダム化比較試験

東芝企業コホートデータを用いて事前スクリーニングを行うことで、被験者の試験開始時点の状態をそろえる。

Flowchart of randomized controlled trial of sodium reduction intervention using genetic information and artificial intelligence (AI) application

表1. 介入開始から3か月時点の各評価項目の平均値及び群間差

Mean values and between-group differences of outcomes at three months after start of intervention

アウトカム	被検者数(人)	平均(SD)			群間差(95%信頼区間)	
		介入群	対照群	AIアプリ限定群	介入群-対照群	介入群-AIアプリ限定群
食塩摂取量 (g/日)	268	10.8(2.0)	11.0(2.0)	10.9(2.1)	-0.2(-0.7~0.3)	-0.04(-0.9~0.8)
BMI	276	25.5(4.6)	25.5(4.0)	24.9(4.1)	0.1(-1.0~1.1)	0.7(-1.2~2.5)
行動変容意図	286	2.9(1.2)	2.7(1.1)	3.0(1.2)	0.2(-0.1~0.4)	-0.1(-0.6~0.4)
収縮期血圧 (mmHg)	128	130.2(19.9)	132.8(11.5)	134.2(12.9)	-2.6(-8.7~3.4)	-4.1(-13.1~4.9)
拡張期血圧 (mmHg)	128	83.4(14.6)	85.7(8.3)	88.3(9.7)	-2.3(-6.7~2.1)	-4.9(-11.5~1.7)

*Sato, K. et al. Sodium Reduction Program Incorporating Genetic Profile and an AI-Based App: A Randomized Clinical Trial. JAMA Netw Open. 2025;8(10):e2537540. doi:10.1001/jamanetworkopen.2025.37540 を基に改変して作成

血圧などを設定した⁽⁵⁾。解析は意図した治療(ITT: Intention to Treat)に基づいて実施し、職域における実装可能性の観点も含めて評価した。研究の概要を図5に示す。

3.2.2 主な結果と考察

312人の被験者をランダム化し、289人の追跡を完了した。試験開始時点の推定食塩摂取量は平均約11 g/日であった。主要アウトカムである3か月後の推定食塩摂取量については、介入群と対照群、介入群とAIアプリ限定群のいずれの比較においても、統計学的に有意な差は認められなかった。3か月時点における主要アウトカム及び二次アウトカムの平均値と群間差を表1に示す⁽⁴⁾。

二次アウトカムについても、群間の有意差はなく、遺伝情報提示とAIアプリを組み合わせた介入は、3か月の観察期間では、食塩摂取量の明確な低下を示さなかった。

この研究の意義は、遺伝情報に基づく個別化予防を職域で実装し、ランダム化比較試験という厳密な枠組みで評価した点にある。加えて、有意な効果が示されなかったことから、行動変容は、遺伝情報やAIアプリによる情報提供だけで十分に成立するとは限らず、継続支援の設計、フィードバック頻度、食環境や組織施策との連動など、周辺要因との相互作用が大きい可能性を示唆している。また、介入期間が3か月と比較的短いこと、随時尿推定値に測定誤差が含まれること、更に職域集団で得られた結果の一般化可能性には留意が必要である。今後は、介入コンポーネントの再設計に加え、食事ログ、購買ログ、血圧の縦断データなどを組み合わせた多層のアウトカム評価や、環境介入との組み合わせ、対象層別化の高度化などが論点になる。

4. あとがき

東芝企業コホートは、職域における長期健康データ基盤として、平時の予防医療支援と有事の迅速な要因解析を両立させ、研究、実装、評価のサイクルを回す“学習する基盤”となり得る。重要なのは、単に多種のデータを蓄積することではなく、情報がバランスの下で、継続的に更新される多種

データを統合し、必要時に追加情報を接続して解析と実装へつなげられる点にある。

今後は、東芝企業コホートの価値を更に高めるために、他集団との連携を個票集約に依存しない分散解析やフェデレーションとして実装し、共通データモデルや評価指標の標準化を進めることが重要である。その際には、情報ガバナンスの徹底を前提とし、再現性と社会還元の確度を高めながら、学習サイクルのスケール拡大を図る必要がある。東芝企業コホートは、健康・医療データと行動データを継続的に統合し、その知見を医療・ヘルスケア分野へ還元するための実践的基盤として、今後ますます重要になる。

文 献

- (1) Sakurai-Yageta, M. et al. Japonica Array NEO with increased genome-wide coverage and abundant disease risk SNPs. The Journal of Biochemistry. 2021, 170, 3, p.399-410. DOI: 10.1093/jb/mvab060.
- (2) Tadaka, S. et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. Human Genome Variation. 2019, 6, 28. DOI: 10.1038/s41439-019-0059-5.
- (3) Omae, Y. et al. Genome-wide association study of common side effects following COVID-19 booster vaccination in a cohort of corporate employees in Japan. Scientific Reports. 2025, 15, 12728. DOI: 10.1038/s41598-025-90787-6.
- (4) Sato, K. et al. Sodium Reduction Program Incorporating Genetic Profile and an AI-Based App: A Randomized Clinical Trial. JAMA Network Open. 2025, 8, 10, e2537540. DOI: 10.1001/jamanetworkopen.2025.37540.
- (5) Brown, I. J. et al. Estimating 24-Hour Urinary Sodium Excretion From Casual Urinary Sodium Concentrations in Western Populations: The INTERSALT Study. American Journal of Epidemiology. 2013, 177, 11, p.1180-1192. DOI: 10.1093/aje/kwt066.

・ジャポニカアレイ®は、国立大学法人 東北大学の登録商標。



山口 泰平 YAMAGUCHI Taihei
Nextビジネス開発部
イノベーション事業推進室 プロジェクト推進部
Innovation Business Promotion Office