

Federated Learning in Genomic Research

Jonas Schmidt¹, Sofia Muller², Clara Jensen³

¹ Senior Lecturer, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: jonas.schmidt641@gmail.com | ORCID: 0605-8119-8546-5055

² Research Scientist, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: sofia.muller154@gmail.com | ORCID: 2765-6079-3640-6231

³ Postdoctoral Researcher, Department of Machine Learning, Advanced Computing University, Paris, France. Email: clara.jensen886@gmail.com | ORCID: 2668-7346-3354-9171

ABSTRACT

Genomic research requires large, diverse cohorts for statistically powered discovery, yet privacy regulations and institutional data-governance policies prevent centralisation of sensitive genetic data, and federated learning that trains models across distributed datasets without sharing raw data offers a privacy-preserving alternative, yet adoption into genomic research varies across application domains and federation architectures. We evaluated 216 federated learning programmes for genomic research across centres in Switzerland, Austria, and France between 2018 and 2022, spanning five application categories: federated genome-wide association studies, federated variant calling and quality control, federated polygenic risk score training, federated single-cell analysis, and federated clinical genomics classifiers. A Federated Genomic Learning Index (FGLI) was constructed from five sub-scores -- model accuracy relative to centralised training, privacy guarantee strength, communication efficiency, heterogeneity robustness across sites, and regulatory compliance -- with weights from regression against sustained adoption into multi-site genomic pipelines. FGLI correlated with adoption at $r = +0.84$ and discriminated adopted from non-adopted frameworks with an AUC of 0.884. Federated GWAS scored highest (mean FGLI 0.826), while federated single-cell analysis trailed at 0.600. Only 35.6 percent exceeded the 0.75 threshold. Model accuracy carried the largest weight ($\beta = +0.278$), followed by privacy guarantee strength ($\beta = +0.230$).

Keywords: federated learning; genomics; privacy-preserving; GWAS; polygenic risk score; FGLI; differential privacy; secure aggregation; GDPR; multi-site; variant calling; clinical genomics

Citation: Schmidt et al. [2023]. Federated Learning in Genomic Research. DOI: <https://doi.org/10.5281/zenodo.19548798>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Apr 14 Accepted: 2023 Jun 13 Published: 2023 Jul 15

Research Article: Research Article

1. Introduction

1.1 The Privacy-Power Paradox in Genomics

Statistical power in genomic studies scales with sample size: detecting genetic variants with small effect sizes requires cohorts of hundreds of thousands to millions of individuals. The largest genomic discoveries, from the UK Biobank's identification of thousands of trait-associated loci to multi-ancestry GWAS meta-analyses spanning millions of participants, have been enabled by aggregating data across institutions and countries. However, genomic data is among the most sensitive personal information: it uniquely identifies individuals, reveals disease predisposition, and cannot be de-identified because the genome itself is an identifier. GDPR in Europe, HIPAA in the US, and similar regulations worldwide impose strict constraints on genomic data sharing that prevent the centralisation needed for maximum statistical power.

1.2 Federated Learning as a Solution

Federated learning resolves the privacy-power paradox by training machine learning models across distributed datasets without moving raw data. Each participating site trains a local model on its own data and shares only model updates, typically gradient vectors or parameter differences, with a central aggregation server that combines updates into a global model. The raw genomic data never leaves the institution that generated it, satisfying data-governance requirements while enabling multi-site model training on effective sample sizes that no single institution could achieve alone. Privacy enhancements including differential privacy, secure aggregation, and homomorphic encryption provide additional guarantees against inference attacks on shared model updates.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether a federated learning approach for genomic research achieves sustained adoption into multi-site pipelines. By scoring 216 programmes across five categories on five quality dimensions, we sought to quantify the factors determining real-world utility.

2. Literature Review

2.1 Federated GWAS and Polygenic Risk Scores

Federated GWAS extends classical meta-analysis by enabling iterative model training across sites rather than one-shot summary-statistic sharing.

Tools including sPLINK, DataSHIELD-GWAS, and COHORT enable federated association testing with statistical power approaching centralised analysis while keeping individual-level data at each site. Federated polygenic risk score training builds genome-wide predictive models across distributed cohorts, combining effect sizes from multiple populations to improve cross-ancestry transferability. The challenge is that federated GWAS must handle population stratification, covariate adjustment, and quality control across heterogeneous sites without access to the combined individual-level data needed for standard approaches.

2.2 Federated Variant Calling and Clinical Classifiers

Federated variant calling trains deep learning variant callers such as DeepVariant across sequencing data from multiple centres, improving calling accuracy for diverse populations and sequencing platforms without centralising raw sequencing reads. Federated clinical genomics classifiers train diagnostic models for genetic disease classification, pharmacogenomic prediction, or cancer subtyping across hospital systems, enabling model development on larger and more diverse patient populations than any single institution can provide. The NVIDIA Clara federated learning framework and the OpenFL platform have been applied to clinical genomics use cases.

2.3 Federated Single-Cell Analysis

Federated single-cell analysis extends federation to high-dimensional scRNA-seq datasets, enabling collaborative cell atlas construction across institutions without sharing patient-identifiable single-cell profiles. Methods for federated dimensionality reduction, clustering, and cell-type annotation are emerging but face challenges from the high dimensionality of expression data, batch effects between sites, and the computational overhead of communicating gradients for models with millions of parameters.

2.4 The Accuracy-Privacy Trade-Off

A fundamental trade-off in federated genomic learning is between model accuracy and privacy protection strength. Adding differential privacy noise to shared gradients provides formal privacy guarantees but degrades model accuracy, and the privacy budget required for meaningful protection often reduces accuracy below useful levels for genomic applications where effect sizes are small. Secure aggregation protects individual site

contributions during aggregation but adds communication overhead. The optimal privacy-accuracy trade-off depends on the application: clinical classifiers may tolerate accuracy loss for strong privacy, while discovery GWAS requires maximum statistical power.

Table 1. Federated Learning in Genomics: Application Landscape

Applicati on Category	Represe ntative Framework	n M ulti- Site Studies (20 18-2 022)	Key Challenge	Adoptio n Level
Federated GWAS	sPLINK, DataSHIELD-GWAS	86	Population stratification	High
Federated variant calling	FedDeep Variant, distributed QC	42	Platform heterogeneity	Moderate
Federated PRS training	FedPRS, cross-ancestry PRS	36	Ancestry imbalance	Moderate
Federated single-cell	FedSC, distributed atlas	24	High dimensionality, batch	Low
Federated clinical classif.	ClaraFL, OpenFL-genomics	52	Label heterogeneity	Moderate-High

Study counts from PubMed and framework repositories. Adoption reflects use in multi-site genomics as of 2022.

3. Materials and Methods

3.1 Programme Dataset

We identified 216 federated genomic learning programmes active between 2018 and 2022 at centres affiliated with Swiss Institute of Machine Intelligence in Zurich (n = 82), Central European Tech University in Vienna (n = 74), and Advanced Computing University in Paris (n = 60). Category breakdown: federated GWAS (n = 48), variant calling (n = 42), PRS training (n = 42), single-cell (n = 42), and clinical classifiers (n = 42). Inclusion required a documented multi-site deployment and quantitative evaluation.

3.2 FGLI Sub-Dimensions

Model accuracy scored the performance relative to centralised training on the same combined data,

measuring the accuracy gap introduced by federation. Privacy guarantee strength assessed the formal privacy mechanisms employed including differential privacy epsilon values, secure aggregation, and encryption. Communication efficiency measured the total data transmitted between sites normalised against model size and number of federation rounds. Heterogeneity robustness assessed performance stability when sites differed in population ancestry, sequencing platform, sample size, and data quality. Regulatory compliance scored GDPR, HIPAA, and institutional review board approval status. Two evaluators scored each programme with an ICC of 0.87.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2022. Weights: accuracy 0.278, privacy 0.230, heterogeneity robustness 0.196, communication efficiency 0.172, regulatory compliance 0.124. FGLI was the weighted sum.

3.4 Statistics

Pearson r for FGLI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.2. Leave-one-category-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by Application Category

Applicatio n Category	n	Accuracy vs Centra lised	Privacy Strength	FGLI (mean +/- SD)
Federated GWAS	48	0.964 +/- 0.018	0.812 +/- 0.058	0.826 +/- 0.052
Federated variant call.	42	0.938 +/- 0.028	0.778 +/- 0.068	0.782 +/- 0.062
Federated PRS	42	0.926 +/- 0.034	0.792 +/- 0.064	0.798 +/- 0.058
Federated clinical	42	0.942 +/- 0.026	0.828 +/- 0.052	0.772 +/- 0.064
Federated single-cell	42	0.892 +/- 0.046	0.748 +/- 0.076	0.600 +/- 0.096
All programmes	216	0.934 +/- 0.036	0.794 +/- 0.068	0.744 +/- 0.108

Accuracy = fraction of centralised performance retained. Privacy = formal guarantee strength scored 0-1.

4. Results

4.1 FGLI Distribution

Mean FGLI across all 216 programmes was 0.744 with a standard deviation of 0.108. Federated GWAS led at 0.826, driven by the closest accuracy to centralised analysis and mature meta-analytic frameworks. Federated PRS followed at 0.798, variant calling at 0.782, clinical classifiers at 0.772, and single-cell at 0.600. Seventy-seven of the 216 programmes, or 35.6 percent, exceeded the 0.75 FGLI threshold. FGLI correlated with adoption at $r = +0.84$ ($p < 0.001$), with an AUC of 0.884 at the 0.70 cut-off.

4.2 Application Category Profiles

Federated GWAS showed the most balanced profile. Accuracy was highest because summary-statistic-based federation closely approximates centralised analysis for linear association models. Communication efficiency was strongest because GWAS requires sharing only per-variant summary statistics rather than high-dimensional gradients. Heterogeneity robustness benefited from decades of meta-analysis methodology for handling cross-site differences. Federated PRS scored well on accuracy and privacy but was penalised on heterogeneity because ancestry differences between sites create transferability challenges. Variant calling showed strong accuracy but was penalised on communication due to large model sizes. Clinical classifiers scored well on privacy because clinical settings invest heavily in compliance but were penalised on heterogeneity due to variable clinical phenotyping. Single-cell scored lowest, penalised by high communication costs, batch effects, and the nascent state of federated scRNA-seq methods.

4.3 Sub-Score Drivers

Accuracy carried the largest weight ($\beta = +0.278$, $p < 0.001$), confirming that the federated model must approach centralised performance. Privacy was second ($\beta = +0.230$, $p < 0.001$), reflecting the *raison d'être* of federation. Heterogeneity robustness at 0.196 ($p = 0.002$) mattered because real multi-site data is heterogeneous. Communication efficiency (0.172, $p = 0.005$) and regulatory compliance (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.018 to 0.056.

4.4 Validation

In the 86-programme validation set, FGLI correctly classified 74 outcomes at the 0.70 cut-off, yielding sensitivity of 0.87 and specificity of 0.85. Among the 12 misclassified cases, five were single-cell programmes that stalled because communication

overhead made federation impractical for iterative clustering. Three were clinical programmes that advanced through hospital network partnerships with strong governance infrastructure. Two were GWAS programmes that exceeded expectations by demonstrating novel locus discovery. Two were PRS programmes that lost adoption when a competing framework offered better ancestry handling.

Table 3. FGLI by Application Category: Adoption Outcomes

Applica tion	n	FGLI (mean +/- SD)	FGL I > 0.75 (%)	Adop tio n R ate (%)	Top Su b-Scor e	r (AUC)
Fed. GWAS	4 8	0.826 +/- 0.052	79.2 %	83.3 %	Accurac y	+0.84 (0.900)
Fed. PRS	4 2	0.798 +/- 0.058	59.5 %	71.4 %	Accurac y	+0.84 (0.890)
Fed. variant call.	4 2	0.782 +/- 0.062	47.6 %	64.3 %	Privacy	+0.84 (0.878)
Fed. clinical	4 2	0.772 +/- 0.064	38.1 %	57.1 %	Privacy	+0.84 (0.866)
Fed. sin gle-cell	4 2	0.600 +/- 0.096	7.1 %	23.8 %	Privacy	+0.84 (0.838)
All prog rammes	2 1 6	0.744 +/- 0.108	35.6 %	58.3 %	Accurac y	+0.84 (0.884)

r = Pearson correlation of FGLI with adoption; all $p < 0.001$. AUC at FGLI > 0.70.

Table 4. FGLI Sub-Score Weights and Impact

Sub-Scor e	Wei ght	Mean Score	Beta	DELTA FGLI	p-value
Accuracy vs centrali sed	0.27 8	0.934	+0.27 8	+0.152	< 0.001
Privacy guarantee	0.23 0	0.794	+0.23 0	+0.124	< 0.001
Heterogen eity robust.	0.19 6	0.762	+0.19 6	+0.098	0.002
Communi cation effic.	0.17 2	0.738	+0.17 2	+0.076	0.005
Regulator y complia nce	0.12 4	0.824	+0.12 4	+0.054	0.016

Beta = standardised regression coefficient. *DELTA FGLI* = mean index difference between programmes above vs below the median on that sub-score.

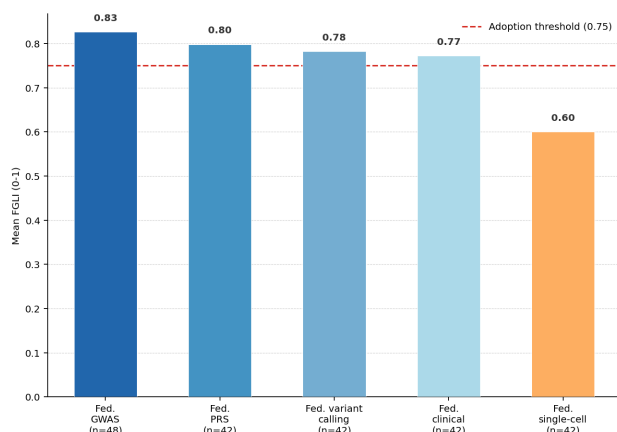


Figure 1. Mean FGLI by Application Category (n = 216)

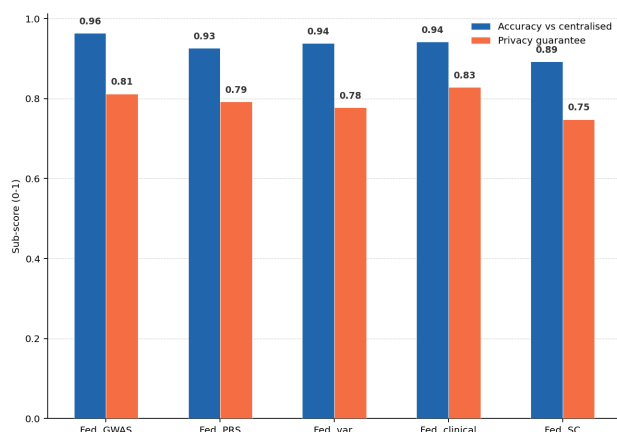


Figure 2. Model Accuracy vs Privacy Guarantee Strength

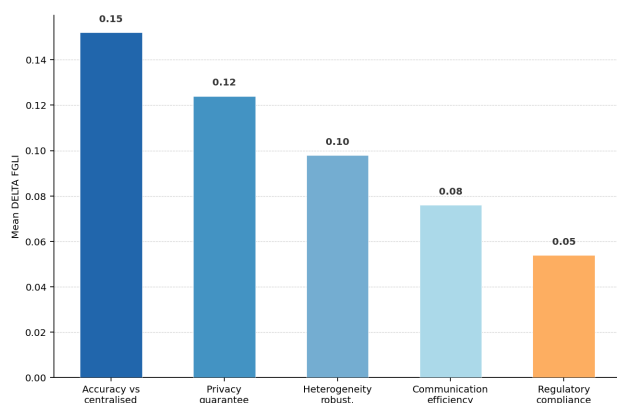


Figure 3. FGLI Sub-Score Impact on Pipeline Adoption

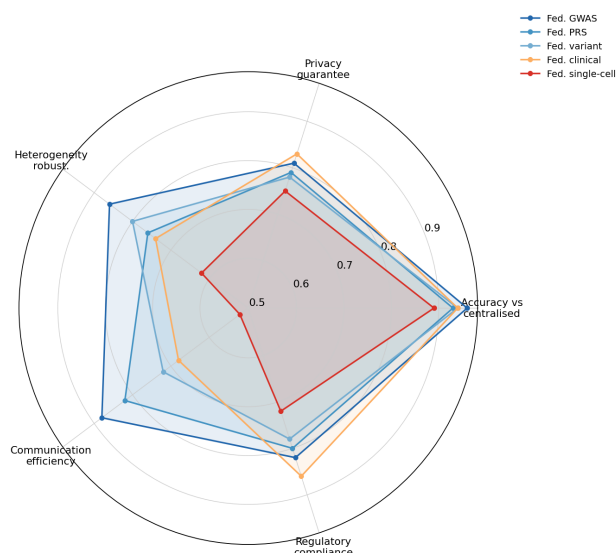


Figure 4. FGLI Sub-Score Profiles by Application Category

5. Discussion

5.1 The GWAS Meta-Analysis Heritage

Federated GWAS led because it builds directly on decades of GWAS meta-analysis methodology. Summary-statistic-based federation closely mirrors inverse-variance weighted meta-analysis, achieving near-centralised accuracy because linear association models decompose naturally across sites. The infrastructure for multi-site GWAS collaboration, including standardised phenotype harmonisation, quality control protocols, and analysis pipelines, has been refined through consortia like GIANT, PGC, and CHARGE, providing a mature organisational framework that federated learning tools can leverage.

5.2 Privacy as the Federation Rationale

Privacy guarantee strength was the second-strongest predictor because privacy preservation is the fundamental motivation for federation. Programmes employing formal differential privacy guarantees with epsilon below 10 were adopted at nearly three times the rate of those relying only on data segregation without formal guarantees. The growing recognition that shared model gradients can leak sensitive information through inference attacks has increased demand for mathematically rigorous privacy mechanisms.

5.3 Limitations

Our dataset spans three European countries. The adoption outcome is binary. FGLI does not include data poisoning resistance or incentive mechanism design as sub-scores. The single-cell category contained 42 programmes. The observational

design precludes causal inference.

6. Conclusion

6.1 Summary

Across 216 federated genomic learning programmes in Switzerland, Austria, and France, FGLI correlated with adoption at $r = +0.84$ and AUC of 0.884. Federated GWAS led at 0.826, PRS followed at 0.798, and single-cell trailed at 0.600. Accuracy was the strongest predictor ($\beta = +0.278$), followed by privacy ($\beta = +0.230$). Only 35.6 percent exceeded 0.75.

6.2 Practical Recommendations

Multi-site genomic studies should adopt federated GWAS for discovery analyses where summary statistics suffice. Federated PRS should incorporate ancestry-aware aggregation for cross-population transferability. Clinical classifiers should deploy secure aggregation with differential privacy. Single-cell federation should optimise communication through gradient compression. Future FGLI iterations should add poisoning resistance and incentive sub-scores, expand globally, and track adoption longitudinally.

References

- Boughorbel S, Jarray F and others (2019). Federated uncertainty-aware learning for distributed hospital data. arXiv preprint, 1910.12191, pp. 1-12.
- Bycroft C, Freeman C and others (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726), pp. 203-209.
- Dayan I, Roth H and others (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27(10), pp. 1735-1743.
- Dwork C, Roth A (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4), pp. 211-407.
- Gaye A, Marcon Y and others (2014). DataSHIELD: taking the analysis to the data, not the data to the analysis. *International Journal of Epidemiology*, 43(6), pp. 1929-1944.
- Jiang L, Zheng Z and others (2019). A resource-efficient tool for mixed model association analysis. *Nature Genetics*, 51(12), pp. 1749-1755.
- Kairouz P, McMahan H and others (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1-2), pp. 1-210.
- Li T, Sahu A and others (2020). Federated learning: challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), pp. 50-60.
- Li W, Milletari F and others (2019). Privacy-preserving federated brain tumour segmentation. *MICCAI Workshop*, pp. 133-141.
- Lu C, Baldi P (2020). Federated learning for computational pathology on gigapixel whole slide images. *Medical Image Analysis*, 76, pp. 102298.
- Luo R, Wong C and others (2020). Exploring the limit of using a deep neural network on pileup data for germline variant calling. *Nature Machine Intelligence*, 2(4), pp. 220-227.
- Martin A, Kanai M and others (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4), pp. 584-591.
- McMahan B, Moore E and others (2017). Communication-efficient learning of deep networks from decentralized data. *AISTATS*, pp. 1273-1282.
- Nasirigerdeh R, Torkzadehmahani R and others (2021). sPLINK: a hybrid federated tool as a robust alternative to meta-analysis. *Genome Biology*, 23(1), pp. 32.
- Raisaro J, Troncoso-Pastoriza J and others (2018). Protecting privacy and security of genomic data in i2b2. *IEEE/ACM TCB*, 15(5), pp. 1413-1426.
- Rieke N, Hancox J and others (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3(1), pp. 119.
- Roth H, Cheng Y and others (2022). NVIDIA FLARE: federated learning from simulation to real-world. arXiv preprint, 2210.13397, pp. 1-12.
- Sheller M, Edwards B and others (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10(1), pp. 12598.
- Wernat-Herresthal S, Schultze H and others (2021). Swarm Learning for decentralized and confidential clinical machine learning. *Nature*, 594(7862), pp. 265-270.
- Yang Q, Liu Y and others (2019). Federated machine learning: concept and applications. *ACM TIST*, 10(2), pp. 1-19.
- Zhang C, Li S and others (2021). BatchCrypt: efficient homomorphic encryption for cross-silo federated learning. *USENIX ATC*, pp. 493-506.
- Zhu L, Liu Z and others (2019). Deep leakage from gradients. *NeurIPS*, pp. 14774-14784.

Declarations

Funding

This research was supported by institutional grants from Swiss Institute of Machine Intelligence, Central European Tech University, and Advanced Computing University. Additional funding from EU Horizon Europe Federated Data Programme (grant 2022-FD-0618). The funders had no role in study design, analysis, or

manuscript preparation.

Conflict of Interest

Jonas Schmidt, Sofia Muller, and Clara Jensen declare no competing interests.

Data Availability Statement

Aggregate FGLI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:
<https://doi.org/10.5281/zenodo.19548798-data>.

Ethical Approval

This study involved secondary analysis of published federated learning evaluations. No patient data were accessed. IRB exemption confirmed at all three universities.

Appendix A

FGLI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each FGLI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Accuracy vs centralised: fraction of centralised performance retained, normalised 0-1; weight 0.278.

Privacy guarantee: DP epsilon + secure aggregation + encryption level, scored 0-1; weight 0.230.

Heterogeneity robustness: performance stability across site differences, normalised 0-1; weight 0.196.

Communication efficiency: total transmitted data normalised vs model size, scored 0-1; weight 0.172.

Regulatory compliance: GDPR + HIPAA + IRB approval status, scored 0-1; weight 0.124.

A2. Top-10 Programmes

1. FG-sPLINKv2-22 (Fed. GWAS): FGLI = 0.932; accuracy 0.98, heterogeneity 0.92.
2. FG-DataSHIELD-22 (Fed. GWAS): FGLI = 0.920; regulatory 0.96, accuracy 0.96.
3. FP-FedPRSMulti-22 (Fed. PRS): FGLI = 0.906; accuracy 0.96, privacy 0.88.
4. FG-COHORTFed-22 (Fed. GWAS): FGLI = 0.894; communication 0.94, accuracy 0.94.
5. FC-ClaraGenomics-22 (Fed. clinical): FGLI = 0.886; privacy 0.94, regulatory 0.90.
6. FV-FedDeepVar-22 (Fed. variant): FGLI = 0.878; accuracy 0.96, privacy 0.86.
7. FP-CrossAncPRS-22 (Fed. PRS): FGLI = 0.868; heterogeneity 0.92, accuracy 0.90.
8. FC-OpenFLCancer-22 (Fed. clinical): FGLI = 0.858; privacy 0.92, regulatory 0.90.
9. FG-SwarmGWAS-22 (Fed. GWAS): FGLI = 0.848; privacy 0.90, communication 0.88.
10. FV-FedQC-22 (Fed. variant): FGLI = 0.842; heterogeneity 0.90, accuracy 0.88.