

附件 2

《基因识别数据分类分级指南（征求意见稿）》编制说明

一、工作简况

（一）任务来源

2022 年 10 月，由深圳市标准化协会批准团体标准《基因识别数据分类分级指南》立项。由深圳华大基因股份有限公司、深圳华大基因科技有限公司、深圳华大生命科学研究院、深圳华大法医科技有限公司、杭州美创科技股份有限公司、北京炼石网络技术有限公司、深圳市第二人民医院、北京新航城智慧生态技术研究院有限责任公司、北京大兴国际机场临空经济区（大兴）管理委员会、上海国际人类表型组研究院、深圳市网安计算机安全检测技术有限公司、深圳永安在线科技有限公司共同组成。

（二）编制背景、目的和意义

近年来，随着新冠疫情的爆发，以核酸检测为代表的基因检测产业蓬勃发展，产业数据随之快速增长。随着各种基因应用范围的扩大，数据量爆发增长，大规模的数据资源汇聚共享模式不断被开启。如何确保基因数据合规利用尤为重要。而数据分类分级则是确保合规的基础。

从法规上看，数据分类分级制度是《中华人民共和国数据安全法》第三章“数据安全制度”中要求的首条制度。国

家标准《GB/T 37988—2019 信息安全技术—数据安全能力成熟度模型》，提出针对数据生命周期安全的定义，数据分类分级是数据采集安全过程的第一个基本实践，是数据生命周期安全管理的第一步，经由数据分类分级，确定数据类别与级别的数据在其生命周期中各个环节需落实的安全控制措施。制定分类分级制度并依规管理，是实现基因数据安全治理目标的重要工作。

与此同时，数据分类分级是平衡数据安全成本的依据。数据分类分级从隐私安全与保护成本的角度出发，对数据进行分类和等级划分，进而根据不同需要对关键数据进行重点防护。传统的数据分级对于现在的大数据时代已不再适用，现阶段需对于不同行业做出定向的、有针对性和可实施性的分类分级标准。

另外，目前基因数据市场规模庞大，但是乱象丛生，没有基因识别数据分类分级的标准指南，缺少定向、专业性、有针对性和可实施性的参考标准。因此亟需标准来对市场进行规范，促进产业健康发展。大量采集、使用基因数据，给人们的生活带来便利及保障生命健康的同时，也出现了基因数据非法收集、滥用、泄露等问题，基因数据在经济社会发展中一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益都将造成极大的危害，因此，制定基因数据分类分级要求标准，对主要的安全隐患进行控制，从而保证基因数据安全的需要已迫在眉睫。目前国内外还没有任何针对个人基因数据分类分级的

标准。站在为推进国家数据安全治理的角度来看，基因识别数据分类分级是具有重要意义的。

（三）主要编制过程

1. 前期准备

2022 年 9 月 28 日，深圳华大基因股份有限公司、深圳华大基因科技有限公司、深圳华大生命科学研究院、深圳华大法医科技有限公司、杭州美创科技股份有限公司、北京炼石网络技术有限公司、深圳市第二人民医院、北京新航城智慧生态技术研究院有限责任公司、北京大兴国际机场临空经济区（大兴）管理委员会、上海国际人类表型组研究院、深圳市网安计算机安全检测技术有限公司、深圳永安在线科技有限公司成立了标准编制小组。标准编制小组对国内外基因识别数据分类分级标准的发展状况、发展趋势及存在问题等进行全面系统调研。对调研资料加以整理分析，起草标准的初步框架。

2022 年 10 月 8 日，召开第一次起草工作会议，初步确定标准编制的原则和标准的框架内容。

2. 标准立项

2022 年 10 月 9 日，深圳市标准协会批准《基因识别数据分类分级指南》的立项。

3. 编制标准草案

2023年4月28日，召开第二次起草组会议，主要讨论本标准的范围，此次研讨会后，经过修改形成工作组讨论初稿。

2023年6月9日，召开第三次起草组会议，主要研究讨论各条款的准确性、实用性和可维护性。经过这次讨论，编制小组根据讨论结果进行修改，形成工作组讨论第二稿。

2023年8月4日，召开第四次起草组会议，起草小组就标准各章节进行讨论，确认文件内容。此外，对标准的语言与格式进行了规范。

2023年9月13日，召开第五次起草组会议，各共同起草单位达成意见一致。

其间，编制小组也不断完善草案，进行内容上的更新，格式上的修改，以保证草案的质量。

4. 公开征求意见

2023年9月至2023年10月，起草工作组将标准草案发送给基因识别数据及关联信息的处理者、监管部门、第三方评估机构等各单位的专家，就草案进行讨论，征求意见和建议。同时，深圳市标准化协会将征求意见稿公布在深圳市标准化协会官方微信公众号，广泛征求建议。

二、标准原则/依据和主要内容

（一）制标原则/依据

1.1 协调一致、依法原则

以现行法律法规的规定和相关标准为基础。本标准中的

约束性条款首先应与这些法律法规、强制性标准的要求保持一致，并在必要情况下予以细化和延伸，但总体上不得有悖于法律法规和强制性标准的要求。

1.2 数据分类原则

a) 科学性原则：按照基因数据的多维特征及其相互间客观存在的逻辑关联进行科学和系统化的分类。

b) 规范性原则：所使用的词语或短语能确切表达数据类目的实际内容范围，在表达相同的概念时，保证用语一致性。

c) 稳定性原则：基因识别数据的分类应以选择体现分类数据对象的本质特征，且不易发生变化的维度和视角作为数据分类的基础和依据，以确保由此产生的分类结果稳定。

d) 扩展性原则：数据分类方案在总体上应具有概括性和包容性，能够实现各种类型数据的分类，以满足将来可能出现的数据类型。

1.3 数据分级原则

a) 合规性原则：数据级别划分应满足相关法律法规及监管要求。

b) 符合伦理原则：数据分级时，涉及人的生命科学和医学研究应符合伦理原则的要求。

c) 综合判定原则：数据分级时应结合数据的应用场景、组合、取值、数据量的大小等，力求数据分级准确合理。

d) 就高原则：一个数据集包含多个不同级别的数据项时，按照数据项的最高级别对数据集进行定级。

e) 动态调整原则：数据的类别级别可能因时间变化、政

策变化、安全事件发生、不同业务场景的敏感性变化或相关行业规则不同而发生改变，因此需要对数据分级进行定期审核并及时调整。

（二）主要内容

2.1 规范了基因识别数据分类分级指南的相关术语和定义；

2.2 规范了基因识别数据分类分级的原则；

2.3 规范了基因识别数据分类分级体系建立的方法；

2.4 规范了基因识别数据分类规则的认定；

2.5 规范了基因识别数据分类分级规则的认定；

2.6 规范了基因识别数据分类规则和分级规则调整的路径；

2.7 规范了基因识别数据中重要数据的保护要求；

2.8 规范了基因识别数据出境合规的要求。

三、产业化情况和预期的经济效益

数据作为现代社会发展的核心要素，也是科技进步的源头。基因数据的科研和应用价值非常高，科学研究、医疗保健、农业、环境等应用场景不胜枚举，有基础生物学研究、人类遗传学研究、医学遗传学研究、药物药理研究以及其他涉及人类遗传信息的研究等，对生命健康影响重大。

（一）顺应市场需求

农业经济时代的核心生产要素是土地，工业经济时代的核心生产要素是技术和资本，数字经济时代的核心生产要素则是数据。得益于基因检测技术的快速发展，基因行业也随

之不断发展，从全球基因测序市场规模来看，市场迅速扩容，根据 BCC Research 发布的数据，2021 年全球基因测序市场规模 157.22 亿美元，将以 19.1% 的 CAGR 增长至 2026 年的 377.21 亿美元，其中基因测序服务领域 2021 年市场规模约为 82.38 亿美元，将以 20.7% 的 CAGR 增长至 2026 年的 210.66 亿美元。从国内市场规模来看，2021 年中国基因测序市场规模为 15.90 亿美元，将以 21.6% 的 CAGR 增长至 2026 年的 42.35 亿美元，预计中国基因测序服务领域规模也约占全球的 1/10。

从基因识别数据的应用场景来看，在医疗健康领域，基因识别数据可以用于个性化医疗、疾病风险评估、药物研发和精准治疗等方面。许多医疗机构和个人都需要基因识别数据来指导诊断和治疗决策；在农业领域，基因识别数据可以用于作物品种改良、疾病抗性培育和农作物适应性研究等方面。农业企业和研究机构需要基因识别数据来提高农作物的产量和质量；在生物科研领域，基因识别数据可以用于基因功能研究、物种起源和进化研究、生物多样性保护等方面。科研机构和学术界都需要基因识别数据来推动科学研究的进展。在基因识别数据应用范围扩大的情况，基因数据的增长呈现井喷的态势。因此，该标准的制定顺应当前基因行业的市场需求。

（二）降低安全成本

基因数据具有体量大、分析杂、汇集难的特点，基因数据安全治理工作任务重，成本高。全球各国的数据安全和隐私保护法在相继陆续推出，如《中华人民共和国网络安全法》《中华人民共和国数据安全法》等均明确提出了要采取数据分类、重要数据备份和加密等安全保护措施，该标准可以提供基因识别数据分类分级的指导，通过遵循该标准的分类分级要求后，企业可以根据数据的敏感程度和风险等级来制定相应的安全保护策略，会降低基因识别数据的安全成本，防止数据泄露和滥用，同时也可以满足相关法规和隐私要求，以提高企业基因识别数据的安全性和合规性。由此会推动相关技术和工具的发展，为厂商和企业提供商机，推动相关新技术的商业化，加速基因行业与应用的产业化进程。

四、国内外相关研究依据、技术标准

针对基因行业的基因识别数据分类分级的标准，目前国内外并无完整规范。个人信息涉及的标准情况如下：

1、国际标准情况：

经检索得知，国际标准中关于基因识别数据分类分级的，有一个相关的标准，《ISO/IEC 27001:2022 Information security management systems Requirements 信息技术 安全技术 信息安全管理体系 要求》该标准规定了一个综合的信息安全管理体系，帮助组织建立、实施、监控和改进信息安全控制措施，以保护其信息资产免受各种威胁的影响，适用于任何类型、规模和行业的组织。

2、国内标准情况：

关于国内标准检索得知有以下 3 个相关标准,详情如下,

《GB/T 38667—2020 信息技术 大数据 数据分类指南》规定了大数据的分类方法和标准。该标准适用于大数据的分类和管理,旨在提供一种统一的分类框架,以便组织和管理大数据资源。

《GB/T 35273—2020 信息安全技术 个人信息安全规范》本标准规定了开展收集、存储、使用、共享、转让、公开披露、删除等个人信息处理活动的原则和安全要求。适用于规范各类组织的个人信息处理活动。

《GB/T 4754—2017 国民经济行业分类》标准规定了全社会经济活动的分类与代码。本标准适用于在统计、计划、财政、税收、工商等国家宏观管理中,对经济活动的分类,并用于信息处理和信息交换。

这些标准并没有专门针对基因识别数据的分类分级要求进行完整的规定。

五、重大分歧意见的处理经过和依据

本标准在编写过程中无重大分歧意见。

六、贯彻深标协标准的要求和措施建议

在本标准通过审核、批准发布之后,由相关部门组织力量对本标准进行宣贯,在行业内进行推广。建议本标准自发布 6 个月之后开始实施。

七、其他应予说明的事项

无。