XXXX-XX-XX发布

XXXX-XX-XX实施

|  |  |
| --- | --- |
| ICS | 35.030 |
| CCS | L 80 |

中华人民共和国国家标准

GB/T XXXXX—XXXX



信息安全技术 基因识别数据安全要求

Information security technology —Security requirements of genetic recognition data

（本稿完成时间：2021年5月6日）

`

目  次

[前  言 III](#_Toc71387969)

[1 范围 1](#_Toc71387971)

[2 规范性引用文件 1](#_Toc71387972)

[3 术语和定义 1](#_Toc71387973)

[4 缩略语 3](#_Toc71387984)

[5 概述 3](#_Toc71387985)

[5.1 文件结构 3](#_Toc71387986)

[5.2 基因识别数据及关联信息相关活动 4](#_Toc71387987)

[5.3 基因识别数据及关联信息应用场景 4](#_Toc71387988)

[6 基本安全要求 5](#_Toc71387989)

[7 安全处理要求 5](#_Toc71387990)

[7.1 通用要求 6](#_Toc71387991)

[7.2 个体服务场景安全要求 8](#_Toc71387992)

[7.3 研究开发场景安全要求 9](#_Toc71387993)

[7.4 其他场景安全要求 9](#_Toc71387994)

[8 安全管理要求 9](#_Toc71387995)

[附录A（资料性附录）基因识别数据各场景的典型活动 11](#_Toc71387996)

[附录B（资料性附录）基因识别数据各场景下的实体类型 13](#_Toc71387997)

[附录C（资料性附录）基因识别典型数据格式 14](#_Toc71387998)

[C.1 概述 14](#_Toc71387999)

[C.2 序列数据格式 14](#_Toc71388000)

[C.3 遗传变异数据格式 15](#_Toc71388001)

[附录D（资料性附录）医学服务类保存规则（摘选） 19](#_Toc71388002)

[附录E（资料性附录）知情同意书模板示例 20](#_Toc71388003)

[E.1医学服务场景知情同意书模板 20](#_Toc71388004)

[E.2法庭科学生物样本基因信息的鉴定服务场景知情同意书模板 21](#_Toc71388005)

[E.3消费级服务知情同意书模板 22](#_Toc71388006)

[E.4研究开发场景知情同意书模板 23](#_Toc71388007)

[E.5基因识别数据及关联信息存储（捐赠保藏）知情同意书模板 25](#_Toc71388008)

[E.6基因识别数据及关联信息共享知情同意书模板 27](#_Toc71388009)

[参考文献 30](#_Toc71388010)

前  言

本文件按照 GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由全国信息安全标准化技术委员会（SAC/TC260）提出并归口。

本文件起草单位：深圳华大基因科技有限公司、中国电子技术标准化研究院、神州网信技术有限公司、深圳华大法医科技有限公司、全知科技（杭州）有限公司、山西医科大学、西安交通大学、北京天融信网络安全技术有限公司、海信集团控股股份有限公司、中国信息通信研究院。

本文件主要起草人：郭小森、郝春亮、张建军、叶明芝、丁国徽、肖棉文、李潇、陈钢、朱丽地孜·叶尔欣、李鹰翔、唐川、郭云峰、魏凤玲、高升杰、胡影、沈悦生、姜先华、严江伟、赖江华、吴昊、李建彬、姜华艳、高雪松、王洪琦、尹烨、傅山、卢旗。

信息安全技术 基因识别数据安全要求

1. 范围

本文件规定了基因识别数据的基本安全要求、安全处理要求和安全管理要求。

本文件适用于基因识别数据控制者安全开展基因识别数据相关业务，也适用于第三方测评机构开展测试评估工作。

1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，标注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 35273—2020 信息安全技术 个人信息安全规范

GB/T 37864—2019 生物样本库质量和能力通用要求

GB/T AAAAA—AAAA 信息安全技术 网络数据活动安全要求

GB/T BBBBB—BBBB 信息安全技术 生物特征识别信息保护基本要求

1. 术语和定义

GB/T 35273—2020、GB/T 37864—2019、GB/T AAAAA—AAAA和GB/T BBBBB—BBBB界定的以及下列术语和定义适用于本文件。



基因识别数据 genetic recognition data

使用技术手段，从人类遗传物质中提取的表征个体或群体遗传信息的数据，该数据可以直接或间接识别到人类个体或群体。

注1：基因识别数据主要包括：基因组核酸序列数据、功能基因组数据，以及提取过程中生成的原始数据和中间数据。

注2：基因组核酸序列数据是指在基因组层面，利用各类测序技术（如：Sanger测序技术、高通量测序技术、质谱技术等）或基因分型技术（如：基因芯片技术、聚合酶链式反应技术等）获得的、描述核酸排列顺序的数据，各类遗传变异的数据（如：单核苷酸多态性（Single Nucleotide Polymorphism，SNP）、插入缺失（Insertions and Deletions，INDEL）突变、短串联重复序列(Short Tandem Repeat，STR)、拷贝数变异(Copy Number Variation, CNV)、基因组结构变异(Structural Variation, SV)等）；功能基因组数据指遗传物质中除基因组核酸序列数据外的表观遗传数据，基因空间位置数据等。典型基因识别数据格式参见附录C。



关联信息associated information

描述遗传物质样品及其数据主体的相关信息，如：采集时间、采集地点、采集对象、采集者、采集方式、样本类型、规格、单位、样品保存期限、测序平台信息、芯片信息、测序时间、数据量、数据类型、遗传疾病、姓名、性别、籍贯、民族、出生日期、身份证号、个体特征、家系信息、用药指南、健康风险等。



数据主体 data subject

基因识别数据所标识或者关联的自然人。

[来源：GB/T 35273-2020，3.3，有修改]



数据控制者 data controller

有能力决定（不论是单独决定还是共同决定）基因识别数据处理目的、方式等的组织或个人。

[来源：GB/T 35273-2020，3.4，有修改]



受托方 outsourced party

接受委托进行基因识别数据及关联信息的采集、提取、操作、使用及存储的法人、公共机构、规制机构、组织或个人。受托方属于数据控制者。



伦理委员会 ethics committee

为确保基因识别数据及关联信息安全、相关活动符合基本社会伦理道德和行为规范的要求，由具有生物医学、法学、社会学、生物信息学等专业背景和其他背景的人员组成的专业机构。该机构负责独立地审查、同意、跟踪审查试验方案及相关文件、获得和记录数据主体知情同意所用的方法和材料等，以保障数据主体的权益。

注：本文件是仅从个人信息及隐私安全保护方面给出的伦理委员会的补充定义，伦理委员会具有更全面的定义。



数据管理委员会 data management committee

为确保数据控制者的业务运转的连续性，保障基因识别数据及关联信息的保密性、完整性、可用性，在数据控制者内部设立的专业机构。该机构负责制定各项数据管理制度，实施数据安全技术防护，开展业务活动过程中的数据安全风险评估，制定数据安全事件应急预案，及时处置安全事件，组织数据安全教育及培训等，以保障业务活动的正常开展和数据主体的合法权益。



去标识化 de-identification

通过对基因识别数据及关联信息（如：姓名、身份证号等）的技术处理，使其在不借助额外信息的情况下，无法识别或者关联数据主体的过程。

[来源：GB/T 35273-2020，3.15，有修改]



匿名化 anonymization

指通过对基因识别数据及关联信息的技术处理，使得数据主体无法被识别或者关联，且处理后的信息不能被复原的过程。

注：匿名化需要平衡基因识别数据隐私性和可用性，匿名化是一种共享的基因识别数据不允许关联数据主体的约定。不同应用场景的匿名化机制有所差别，同时随着技术发展，需要定期对匿名化机制和对个人信息风险影响进行评估和更新。在任何场景下，明确匿名化的数据，不可以复原和关联数据主体，并且有相应的机制保障。

[来源：GB/T 35273-2020，3.14，有修改]



知情同意 informed consent

基因识别数据控制者和数据主体或数据主体的法律监护人之间关于基因识别数据采集和使用边界的约定。知情同意书也包括服务协议、委托书等。具体实施过程体现在知情同意书的签订过程。

注：知情同意由[知情权](https://baike.baidu.com/item/%E7%9F%A5%E6%83%85%E6%9D%83/3073586)和同意权两个密切相连的权利组成。知情权是同意权得以存在的[前提](https://baike.baidu.com/item/%E5%89%8D%E6%8F%90/10836999)和[基础](https://baike.baidu.com/item/%E5%9F%BA%E7%A1%80/32794)，同意权又是知情权的价值体现，强调数据主体的知情同意权，主要目的在于通过赋予基因识别数据控制者相应的[告知义务](https://baike.baidu.com/item/%E5%91%8A%E7%9F%A5%E4%B9%89%E5%8A%A1/7169067)，使数据主体在了解自己将面临的[风险](https://baike.baidu.com/item/%E9%A3%8E%E9%99%A9/2833020)、付出的[代价](https://baike.baidu.com/item/%E4%BB%A3%E4%BB%B7/7723860)和可能取得的[收益](https://baike.baidu.com/item/%E6%94%B6%E7%9B%8A/638677)的基础上自由作出[选择](https://baike.baidu.com/item/%E9%80%89%E6%8B%A9/980212" \t "https://baike.baidu.com/item/%E7%9F%A5%E6%83%85%E5%90%8C%E6%84%8F%E6%9D%83/_blank)，从而[维护](https://baike.baidu.com/item/%E7%BB%B4%E6%8A%A4/7097570)数据主体的利益，改变数据主体相对[弱势](https://baike.baidu.com/item/%E5%BC%B1%E5%8A%BF/7500228)地位。

1. 缩略语

下列缩略语适用于本文件。

BP 碱基对（Base Pair）

CMA 检验检测机构资质认定（China Inspection Body and Laboratory Mandatory Approval）

CNAS 中国合格评定国家认可委员会（China National Accreditation Service for Conformity Assessment）

CNV 拷贝数变异（Copy Number Variation）

INDEL 插入/缺失（Insertions and Deletions）

PCR 聚合酶链式反应（Polymerase Chain Reaction）

SNP 单核苷酸多态性（Single Nucleotide Polymorphisms）

STR 短串联重复序列（Short Tandem Repeat）

SV 基因组结构变异（Structural Variation）

1. 概述
   1. 文件结构

本文件描述了基因识别数据及关联信息的六个相关活动环节（见5.2节），说明了这些环节中与基因识别数据及关联信息相关的活动情况。

本文件描述了基因识别数据及关联信息的应用场景（见5.3节），说明了这些场景中所涉及到的数据控制者实体，这些实体是按照本文件履行相关保护责任的主体。

本文件在提出基本安全要求（见第6章）的基础上，针对每一个应用场景，提出安全保护要求（见第7章），并针对基因识别数据及关联信息提出管理要求（见第8章）。

* 1. 基因识别数据及关联信息相关活动

基因识别数据及关联信息相关活动包括：

1. 基因识别数据及关联信息的采集（简称：数据采集）：
   1. 提出基因识别数据及关联信息的使用需求（简称：提出需求）：是指数据控制者（组织机构或个人）为完成业务活动，对数据主体提出基因识别数据及关联信息的使用需求；
   2. 数据主体和数据控制者签订知情同意书（简称：知情同意书签订）：是指基因识别数据控制者提出基因识别数据及关联信息的使用需求、数据主体知情同意后，双方应完成的知情同意书的签订（知情同意书模板参见附件E）；
   3. 人类遗传物质采集（简称：遗传物质采集）：是指采集生物样品，并从中提取相应遗传物质的活动。生物样品包括血液、唾液、尿液、粪便、毛发、脑脊液和脱落细胞等。遗传物质采集活动中会产生相应的关联信息，例如：采集时间、采集地点、采集对象、采集者、采集方式、样本类型、规格、单位、样品保存期限等；
   4. 基因识别数据提取（简称：数据提取）：是指通过测序技术、基因分型技术等手段从遗传物质中获得对应的基因识别数据的活动。此活动会产生相关的关联信息，如：测序平台信息、芯片信息、测序时间、数据量、数据类型等。
2. 基因识别数据及关联信息的使用（简称：数据使用）：是指为满足知情同意约定的使用需求对所获取的基因识别数据及关联信息进行分析处理的活动。此活动会产生与数据主体直接相关的、涉及隐私的关联信息，如：遗传疾病、个体特征、用药指南、健康风险等。
3. 基因识别数据及关联信息的存储（简称：数据存储）：是指在数据主体知情同意的情况下，基因识别数据控制者对基因识别数据及关联信息进行存储的活动。

注：数据的存储也包括归档及保藏。

1. 特定场景下的基因识别数据及关联信息的共享、转让（简称：数据共享、转让）：是指为达到一定的服务目标或科研目标，在符合相关法律法规及数据主体同意的前提下，基因识别数据控制者将基因识别数据及关联信息共享给其他控制者的活动。
2. 基因识别数据及关联信息的销毁（简称：数据销毁）：是指在数据主体要求撤回授权情况下，不可逆地删除基因识别数据及关联信息的活动。
   1. 基因识别数据及关联信息应用场景

基因识别数据及关联信息应用场景包括：

1. 个体服务场景：针对数据主体，提取并使用基因识别数据及关联信息，得到结果，达到服务目标的场景。其特点是，在服务目标确定的情况下，所需的基因识别数据及关联信息种类和数量固定，使用方式、手段、流程相对固定。绝大多数情况下，相关服务业务的开展需要经过一定的行政审批或许可。常见的服务包括：
   1. 医疗服务：是指医疗服务机构使用基因识别数据及关联信息为数据主体进行辅助诊断、治疗、预后、护理等服务的场景。此场景中，部分机构具备完成5.2中所有活动的能力。部分机构会根据自身的情况，委托具备服务资质的受托方完成全部（或部分）遗传物质采集、数据提取（如医学检测机构）、数据使用、数据存储等活动；
   2. 法庭科学生物样本基因信息的鉴定服务：是指司法鉴定人以生物样本（如：血液、体液、组织、毛发等）为样品，运用科学技术（如：PCR、毛细管电泳、高通量测序等）和专门知识（如：遗传学、法医学等）对案、事件中涉及的专门性问题（如：种属、个体识别、亲权、亲缘关系等）进行鉴别和判断并提供鉴定意见的活动。涉及到的数据控制者主要为司法鉴定机构；
   3. 消费级服务：是指通过广告、电商、线下店等被消费者认知且直接购买的基因检测服务。该服务不涉及医疗服务机构的体外诊断。消费级服务机构直接向个人提供基因检测服务，结合已发表文献或其他研究结果，对个人的基因检测结果进行个性化解读，例如：营养健康服务、运动能力评估、遗传缺陷分析、溯源分析等。此场景中，部分机构具备完成5.2中所有活动的能力；部分机构会根据自身的情况，委托具备服务资质的受托方完成全部（或部分）活动。
2. 研究开发场景：是指利用基因识别数据及关联信息，以发现新知识、开发新技术或产品、改良现有技术或产品等为目标所进行的科学研究或者技术开发等活动。此场景中，基因识别数据的提取方式、使用方式、手段及流程等往往不固定。此场景涉及到的数据控制者为高等学校、医疗机构、科研机构、企业、非盈利机构等多种组织机构。此场景会存在不同个体、团队和组织之间的数据共享和交换使用的情况。常见的研究开发活动包括：
   1. 科学研究：以发现新知识为主要目的的活动，例如：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。科学研究成果一般会以文章、专利等形式公开发表，其中会涉及部分基因识别数据及关联信息；
   2. 技术开发：以开发新技术或升级已有技术为目的的活动，主要包括计算技术开发（如：算法、分析流程等）和实验技术开发（如：测序技术、临床实验方法等）两大类；
   3. 产品研发或升级：以研发或升级应用产品为目的的活动（多以盈利或实现具体落地功能为目的）。常见的活动包括临床检测类产品、药物研发、消费级服务等产品。
3. 其他场景包括：
   1. 公共卫生相关活动场景：一般由国家或地方公共卫生主管部门发起的活动，这些活动均会有一定的配套法律和行政授权，并由相应资质的组织机构完成（一家独立完成或多家协作完成），如：流行病调查、疾病筛查等；
   2. 遗传信息公共服务场景：在有关部门的行政授权下，所提供的基因识别数据及关联信息的存储、使用、共享等服务。此类服务具有公益性，用于支持相关领域的科学研究和产业发展，支撑国家或地方重大项目和工程的实施等。此类基因识别数据及关联信息通常来自数据主体的捐赠。
4. 基本安全要求

基因识别数据属于个人信息，应符合GB/T 35273—2020、GB/T AAAAA—AAAA和GB/T BBBBB—BBBB的要求。

数据控制者应论证使用基因识别数据及关联信息开展业务的必要性和合理性，明确对数据主体的影响，并依据标准要求确定相关业务流程、数据安全保护需求，获得相关的资质，包括但不限于：

1. 从事医疗服务的实体，应取得卫生主管部门批准并具有临床基因扩增检验实验室资质。
2. 从事法庭科学生物样本基因信息的鉴定服务的实体，应取得与从事的基因识别数据及关联信息相关的检验检测机构资质认定（CMA）或CNAS实验室认可以及当地司法行政部门许可。
3. 从事消费级服务的实体，应具有有关部门认证的实验室资质，并获得相应营业许可。
4. 从事研究开发的实体，应具备相关研究开发的能力，包括相应等级的实验室、相关研究经验及成果、研究团队等。
5. 从事公共卫生活动的实体，应获得相关法律法规和行政的许可。
6. 从事遗传信息公共服务的实体，应获得国家有关部门认定的资质。
7. 安全处理要求
   1. 通用要求
      1. 数据采集
         1. 提出需求

对数据控制者的要求包括：

1. 应根据自身业务的特点，提出需求，所需的基因识别数据及关联信息应与相关业务直接关联，且只收集和使用满足业务需要的最少数据。
2. 使用需求应得到许可后方可开展相关活动。依据不同的使用场景，相关使用需求应通过伦理委员会的审查，必要时还需有关部门的报备及审批同意。
   * + 1. 知情同意书签订

数据控制者应向数据主体展示知情同意书，并进行充分的解释说明，得到数据主体的知情同意，并签署该知情同意书（书面签署或可以防止假冒和篡改的技术保障的电子签署）。

应针对每一个场景准备相应的知情同意书，不应在一份知情同意书中涵盖多个应用场景。不应强制数据主体同时签署多个场景知情同意书。知情同意书应满足的要求包括（不同场景下的不同模板见附录E）：

1. 告知方应为基因识别数据的需求提出方。
2. 应与相关业务对基因识别数据及关联信息的需求一致，不应要求超出需求的生物样品、基因识别数据及关联信息。
3. 应列明数据主体拥有的权利（参见6.2）。
4. 应告知遗传物质及所获取数据的用途、安全保护措施及应用的潜在风险、生物样品和数据存储时限。
5. 应列明业务过程中所涉及到的所有数据控制者。
6. 应描述基因识别数据及关联信息被共享和交换的内容、范围、对象。
7. 应描述数据主体撤回授权同意的方法。
   * + 1. 遗传物质采集

对数据控制者的要求包括：

1. 数据主体生物样品及遗传物质的采集应符合GB/T 37864—2019的要求。
2. 采集过程完成后，应对样品进行去标识化处理，去除样品与数据主体个人敏感信息的关联关系。
3. 如委托受托方采集生物样品，应签订委托协议，确保其在约定的目的和范围内采集遗传物质和关联信息；若需采集姓名、身份证号等个人敏感信息，则需具备相应安全能力资质，并保证个人敏感信息与基因识别数据分离、各自单独加密存储，并设置严格访问控制权限。
   * + 1. 数据提取

对数据控制者的要求包括：

1. 对基因识别数据的中的个人敏感信息予以明确区分，提取的数据格式应支持个人敏感信息与生物性特征的分离。
2. 基因识别数据提取应符合最小必要原则，严格按照行业推荐的基因识别类型和数量进行提取，不应提取与需求不直接相关的基因信息。
3. 数据提取过程中，应只将基因识别数据与生物样本建立关联，不应与任何个体或群体自然人的任何信息相关联。
4. 数据提取过程中，应严格按照知情同意书中的描述处理生物样本。数据提取完成并通过质量检测后，应销毁生物样本，有法律法规明确要求的情况除外。
5. 如委托提取数据，不应提供关联信息；应与受托方签订委托协议，确保其在约定的目的和范围内提取基因识别数据，且符合本文件对数据提取的数据安全要求。
6. 如委托提取数据，受托方应在协议约定的时限内将提取的数据交付给委托方，并使用适当的加密技术，保证数据交付安全。受托方应在数据存储期限内或协议约定期限内销毁数据。
   * 1. 数据使用

数据控制者在获取基因识别数据后，应严格按照知情同意书中的内容使用基因识别数据及关联信息，满足的安全要求包括：

1. 应根据相应场景的使用目的，正确使用基因识别数据及关联信息。
2. 基因识别数据使用时，可与符合使用目的的关联信息建立关联，达到使用目的后，应及时解除关联。
3. 如委托进行数据分析，应满足的数据安全要求包括：
4. 数据分析前，应对委托行为进行信息安全影响评估，以确保受托方具备足够的数据安全能力；
5. 数据分析前，应对受托方安全能力提出要求，通过知情同意形式明确双方的安全责任及安全措施，明确数据分析的规范，包括保存期限、销毁或返还期限等；
6. 数据分析过程中，应根据分析目的提供最少的关联信息；
7. 数据分析过程中，禁止受托方利用基因识别数据与其他来源数据关联数据主体的行为。
   * 1. 数据存储

对数据控制者的要求包括：

1. 应在按照7.1.1.2的要求取得数据主体的书面授权和同意后，进行基因识别数据及关联信息的存储。
2. 应在满足国家最低保存期限和知情同意书签订保存期限的基础上确定数据保存期限，进行存储。

注：医学服务场景数据保存规则可参考附录D。

1. 基因识别数据的存储设施应通过伦理委员会及数据管理委员会审查；存储设施的建设和运行应遵循国家相关安全要求标准和规范；应采用合理的数据加密等安全技术，保证数据保密性、可用性和完整性，通过相关安全评估。
2. 数据存储过程中，基因识别数据应经过去标识化处理。
3. 数据存储过程中，基因识别数据与关联信息应分开存储，并设置不同的访问和管理权限；
4. 基因识别数据及关联信息存储需达到审查批准或知情同意的存储时限且需继续存储的，应重新进行相关审查登记批准流程，并按照7.1.1.2的要求重新取得数据主体的同意。
5. 在取得数据主体同意的情况下，可委托其他具有基因识别数据及关联信息存储资质的机构依据相关法律法规采用加密技术代为存储。
   * 1. 数据共享、转让

对数据控制者的要求包括：

1. 基因识别数据及关联信息应先按照7.1.1.2的要求取得数据主体的知情同意、再通过伦理委员会和数据管理委员会的审核并向有关部门申报登记获得批准后，可进行最小限度的数据共享；该数据共享应符合知情同意书中明确描述的目的、共享对象、共享内容等。
2. 数据提供方应与接收方签署协议，确认数据接收方的数据应用场景及资质，确认接收方应承担相应的数据保护责任并约定违约条款。
3. 数据接收方应仅将所接收的数据用于提供方所约定的目的，不应用于其他目的，不应再次共享给其他实体。
4. 因业务目的确需扩展共享的范围和内容，且超出原知情同意书要求的，应单独向数据主体告知新的共享目的、涉及的数据类型、数据接收方的具体身份，并按照7.1.1.2的要求取得数据主体的知情同意。
5. 基因识别数据及关联信息原则上不应转让。数据控制者发生收购、兼并、重组和破产时的转让，应告知数据主体有关情况，并重新依照7.1.1.2的要求取得数据主体的知情同意，变更后的数据控制者应继续履行原数据控制者的责任和义务。
6. 基因识别数据及关联信息不应买卖。
7. 基因识别数据及关联信息不应出境，按照7.1.1.2的要求取得数据主体的知情同意并获得有关部门批准的除外。

7.1.5 数据销毁

对数据控制者的要求包括：

1. 国家有关部门明令销毁的，应及时销毁相关基因识别数据及关联信息。
2. 数据主体撤回授权或明示停止使用后，应及时销毁其基因识别数据及关联信息，经过匿名化处理进行公共卫生相关的群体分析或科学研究发表的数据除外。
3. 存储时间达到授权存储期限且无新的存储授权，应及时销毁基因识别数据及关联信息。
4. 如使用物理介质进行数据传输，传输完成后应及时销毁介质中的基因识别数据及关联信息。
5. 如委托进行数据分析，完成交付后，应确保受托方按照书面协议约定的期限及时销毁相关基因识别数据、关联信息、中间数据及分析结果。
6. 因机构吊销执照、取消资质、破产等原因无法开展相关业务，且存储的基因识别数据及关联信息无承接方的，应及时对数据进行销毁处理。
   1. 个体服务场景安全要求

个体服务场景中，应按照国家有关规定及有关部门批准的服务产品要求，确定样品种类、所需的基因识别数据及关联信息的内容，可能涉及的信息范围包括：

1. 医疗服务：
   1. 典型生物样品，如：血液、尿液、唾液、粪便、特定检材等；
   2. 单个产品基因检测位点数不应超过批准的服务产品中规定的位点数；
   3. 典型关联信息，如：姓名、年龄、性别、病史、检验报告、手术记录、用药记录、过敏史、生育信息、家族病史等。
2. 法庭科学生物物证鉴定：
   1. 典型生物样品，如：血液、唾液、骨骼、毛发、体液、脱落细胞等；
   2. 个体识别、亲权鉴定的基因检测位点数目及基因序列长度不应超过监管部门批准的数目和长度；
   3. 典型关联信息，如：姓名、身份证号、籍贯、年龄、性别、家系信息等。
3. 消费级服务中的其他产品：
   1. 典型生物样品，如：血液、唾液等；
   2. 单个产品基因检测位点数不应超过批准的服务产品中规定的位点数；
   3. 典型关联信息，如：姓名、身份证号、籍贯、年龄、性别、家系信息等。
   4. 研究开发场景安全要求

对数据控制者的要求包括：

1. 应制定科学的研究方案或计划，明确说明前期工作基础、研究目的、意义、必要性、相关实验设计、研究过程中可能涉及到的合作及相关数据共享需求（如：共享的数据类型、格式及其必要性等），并依照最小必要原则，制定采样数量、基因识别数据及关联信息的提取方法等。
2. 研究方案或计划应通过科学性审查、科研诚信调查、伦理委员会及数据管理委员会审批，如涉及特定地区人群、特定疾病等法律法规要求的，应向有关部门申报登记并获得批准。
3. 研究开发活动的知情同意书还应满足如下安全要求：
   1. 应明确告知数据主体科研项目名称、项目来源、研究目的及批准单位；
   2. 应告知数据主体基因识别数据及关联信息（包括中间数据）是否会长期存储；
   3. 不应使用宽泛的用途描述（如：“用于科研目的”等字样）。
4. 基于捐赠的基因识别数据及关联信息开展的研究开发活动，无需再次获取知情同意。
5. 所有环节中，基因识别数据及关联信息应经过去标识化处理后使用。
6. 研究开发过程中新发现的关联信息，应仅用于研究目的。
   1. 其他场景安全要求

对数据控制者的要求包括：

1. 公共卫生活动场景中，应按照有关部门的政策要求，提出合理的基因识别数据及关联信息的采集需求。
2. 公共卫生活动场景中，应准备专门的知情同意书，得到有关部门同意后使用。
3. 公共卫生活动场景中，应根据活动的内容和有关部门的批准，对指定范围内的基因识别数据及关联信息进行处理和使用。因业务需要，在结果交付或告知数据主体时，可恢复与数据主体的关联。
4. 遗传信息公共服务场景中，应根据有关部门批准的业务范围，存储基因识别数据及关联信息。
5. 遗传信息公共服务场景中，不应只以数据存储为目的提出基因识别数据及关联信息采集的需求，获得有关部门许可的除外。
6. 遗传信息公共服务场景中，应准备专门的知情同意书。知情同意书应符合如下安全要求：
   1. 应告知数据主体所采集的基因识别数据及关联信息属于捐赠，会用于广泛的科研目的，在国家有关部门授权范围内会被共享和交换。
   2. 应告知数据主体审批部门的名称及相关授权文件的名称或标识。
   3. 应告知数据主体对所获得的基因识别数据及关联信息进行的去标识化、匿名化的方法。

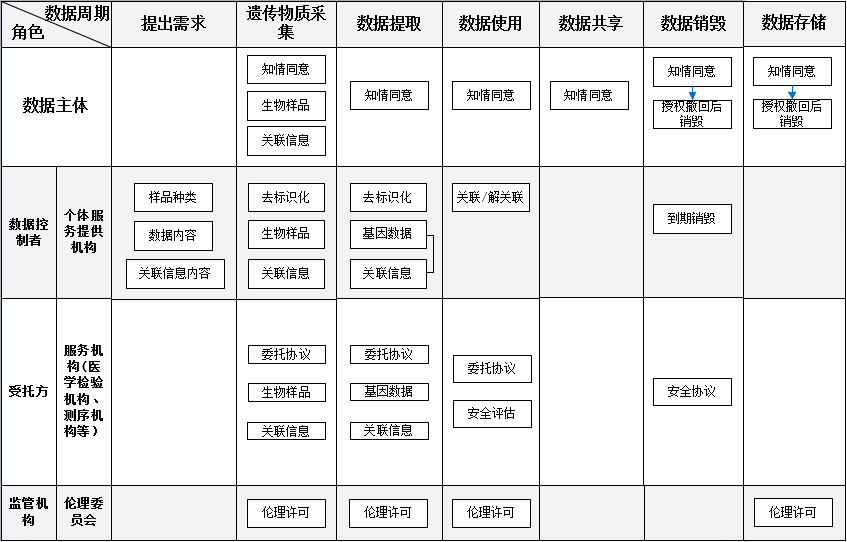
注：知情同意书内容模板可参照附录E。

1. 遗传信息公共服务场景中，如所采集的捐赠数据仅用于研究开发活动，该数据应经过去标识化处理。
2. 安全管理要求

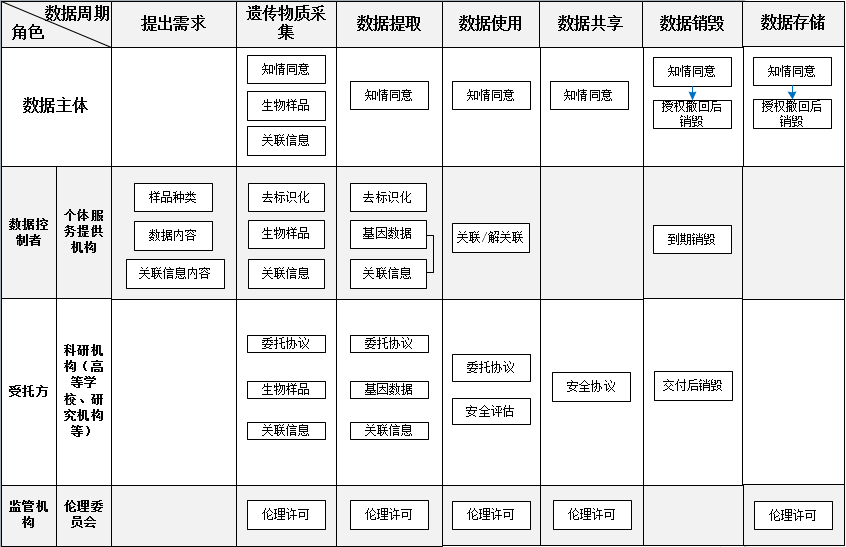
对数据控制者的要求包括：

1. 应参照GB/T 35273—2020、GB/T BBBBB—BBBB等标准的要求，设立专门数据管理机构（如：数据管理委员会），制定个人信息保护政策，建立数据安全管理机制，形成完整的信息安全管理体系，确保基因识别数据及关联信息全生命周期的安全性，保障数据主体的合法权益。
2. 应针对基因识别数据相关的特有智能设备（如：测序仪、质谱仪、生物计算服务器等)和业务特点，制定和实施相关的安全管理制度和技术措施（如：网络隔离措施、访问控制措施、入侵防范措施等）。
3. 当发生基因识别数据及关联信息被窃取、篡改、丢失、泄露、损毁等安全事件时，应按照规定向相关部门报告；及时采取应急处置及补救措施；应按照规定通过邮件、信函、电话、推送通知等方式及时告知数据主体；难以逐一告知数据主体时，应采取合理、有效的方式公开发布相关的警示信息。
4. 应建立数据主体撤回授权、投诉及跟踪处理机制，并在承诺时限内对数据主体的相关请求进行响应。
5. (资料性附录）  
   基因识别数据各场景的典型活动

个体服务及研究开发场景下基因识别数据的典型活动示例如图A.1、A.2所示：



图A.1 个体服务场景下的基因识别数据的活动图



图A.2 研究开发场景下的基因识别数据的活动图

1. (资料性附录）  
   基因识别数据各场景下的实体类型

表B.1 各场景下的典型实体类型

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 场景 | | 提出需求 | 遗传物质采集 | 数据提取 | 数据使用 | 数据共享 | 数据销毁 | 数据存储 |
| 法律规定的活动场景 | 与国家安全、国防安全直接相关 | 公安机关、人民检察院、人民法院、执法机构 | 行政执法部门认定的有资质机构、司法鉴定机构 | 司法鉴定机构、测序机构、医学检验机构 | 公安机关、人民检察院、人民法院 | 公安机关、人民检察院、人民法院 | 不涉及 | 司法鉴定机构、公安机关、人民检察院、人民法院 |
| 与公共安全、公共卫生、重大公共利益直接相关 |
| 与刑事侦查、起诉、审判和判决执行等直接相关 |
| 个体服务场景 | 医疗服务 | 各类医疗机构，医学检验机构 | 各类医疗机构、医学检验机构 | 各类医疗机构、医学检验机构 | 医疗合作机构 | 各类医疗机构、医学检验机构 | 各类医疗机构、医学检验机构 | 各类医疗机构、医学检验机构、数据保藏机构 |
| 消费级服务 | 基因检测机构 | 体检机构、基因检测机构 | 基因检测机构、测序机构 | 体检机构、商业合作机构、科研合作机构 | 体检机构、基因检测机构、测序机构 | 基因检测机构、测序机构 | 基因检测机构、数据保藏机构 |
| 研究开发场景 | 研究开发 | 高等学校、医疗机构、科研机构、非盈利机构、企业 | 高等学校、医疗机构、科研机构、企业 | 高等学校、医疗机构、科研机构、测序机构 | 高等学校、医疗机构、科研合作机构、公共数据库 | 高等学校、医疗机构、科研机构、  非盈利机构、企业 | 高等学校、科研机构、医疗机构等实体内的测序实验室测序机构 | 高等学校、医疗机构、科研机构、  非盈利机构、企业数据保藏机构 |
| 其他场景 | 公共卫生 | 国家或地方公共卫生主管部门 | 基因检测机构 | 基因检测机构、测序机构 | 国家或地方公共卫生主管部门 | 国家或地方公共卫生主管部门 | 不涉及 | 国家数据保藏机构 |
| 遗传信息公共服务 | 负责国家或地方重大项目和工程的国家有关部门部门 | 基因检测机构 | 基因检测机构、测序机构 | 负责国家或地方重大项目和工程的国家有关部门部门 | 负责国家或地方重大项目和工程的国家有关部门部门 | 不涉及 | 国家数据保藏机构 |

1. (资料性附录）  
   基因识别典型数据格式

C.1 概述

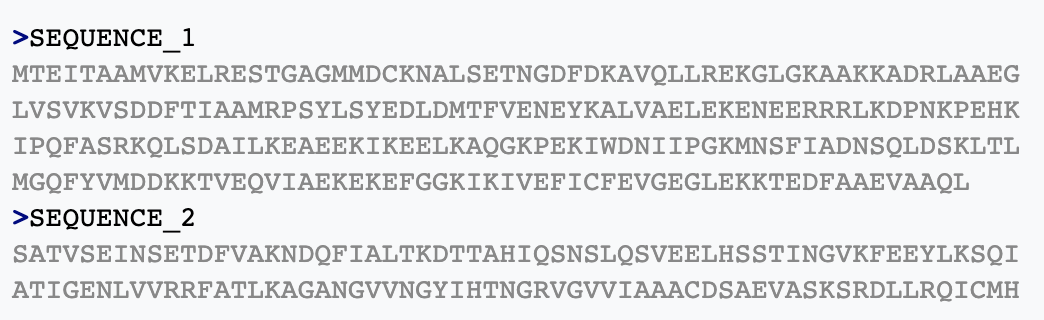
典型的基因识别数据格式包括序列数据格式和遗传变异数据格式两大类。常见的基因组序列数据格式包括FASTA、FASTQ、SAM/BAM等；常见的遗传变异数据格式包括VCF、GVCF、BED、PLINK、GEN/BGEN、VDS及MT等。

C.2 序列数据格式

C.2.1 FASTA格式

FASTA格式是一种用于表示核苷酸序列或氨基酸（蛋白质）序列的文本格式，常用于存储基因组信息，其中核苷酸或氨基酸使用单字母编码表示。

每个FASTA文件可存储多条序列信息，每条序列信息的存储格式由包括两部分构成：第一部分为以“>”开头的描述行或标题行/标识符行，给出了序列的名称和唯一标识符，还可能包含其他信息。第二部分是实际序列本身，以单字母字符串表示。有效字符以外的任何字符都将被忽略（包括空格、制表符及星号等）。文件扩展名通常是fasta或fa，另外特殊类型序列也有对应的文件扩展名，如ffn对应基因编码区间（nucleotide of gene regions），faa指氨基酸（amino acid），frn对应非编码RNA（non-coding RNA）序列。具体示例参见图C.1。



图C.1 FASTA格式示例图

C.2.2 FASTQ格式

FASTQ格式是FASTA格式的扩展，用于存储生物序列（通常是核苷酸序列）及其相应的质量分数，常用于存储下一代测序（Next Generation Sequencing）数据。为了简洁，序列字母和质量分数都用一个ASCII字符进行编码。

FASTQ文件中每个序列用四行表示：

第1行以“@”字符开头，后面跟着序列标识符和可选的描述；

第2行是核苷酸或氨基酸序列信息，使用单字母编码表示；

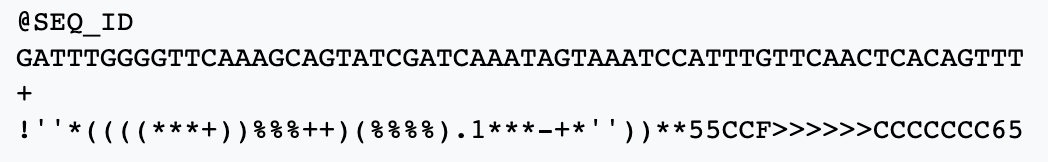
第3行以“+”字符开头，其后视需要跟随相同的序列标识符（和任何描述）；

第4行为序列对应的编码质量，每一个符号对应相应核苷酸或氨基酸的质量；表示质量的字节从0x21开始运行（最低质量；ASCII中为“!”）到0x7e（最高质量；ASCII中为“~”）。图C.2提供了按质量从左到右递增的质量值字符示例。



图C.2质量值字符排列示例

FASTQ格式示例参见图C.3。



图C.3 FASTQ格式示例

C.2.3 SAM/BAM格式

Sequence Alignment/Map （SAM）格式是一种制表符分隔纯文本格式，主要用来存储比对到参考序列（reference sequence）的序列，也可以用来存储未经过比对的序列数据。Binary Alignment/Map （BAM）格式是SAM文件通过区块压缩（block gzipped）而产生的二进制格式。SAM/BAM格式通常用于存储由下一代测序技术产生的核苷酸序列。

SAM格式由标题部分（header section）和比对部分（alignment section）组成。标题部分必须在比对部分之前。标题以“@”符号开始，将其与比对部分区别开来。

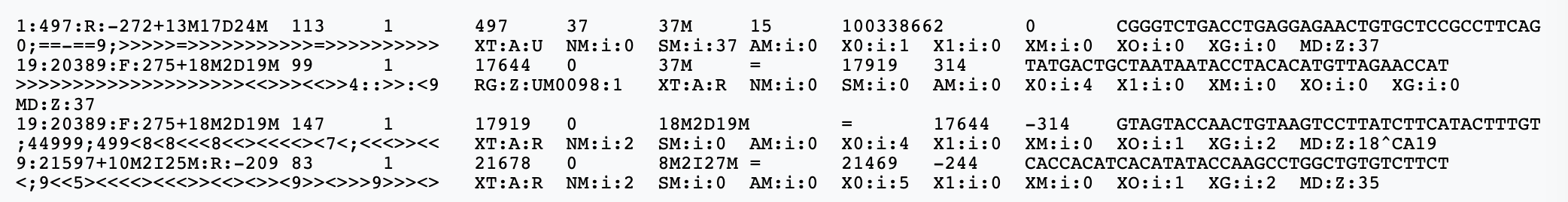
比对部分有序列名、标记（Flag）、参考序列名（RNAME）、比对位置（POS）, 序列比对质量（MAPQ），CIGAR字符串，配对序列名（RNEXT），配对序列的位置（PNEXT）比对序列长度（TLEN）、序列（SEQ）及碱基对质量（QUAL）等11个强制字段。除强制字段外，还有多个自定义区域。具体细节可参见（https://samtools.github.io/hts-specs/SAMv1.pdf）。对于以下序列比对示例参见图C.4。

Graphical user interface, text, application

Description automatically generated

图C.4 序列比对示例

对应的比对部分示例参见图C.5。

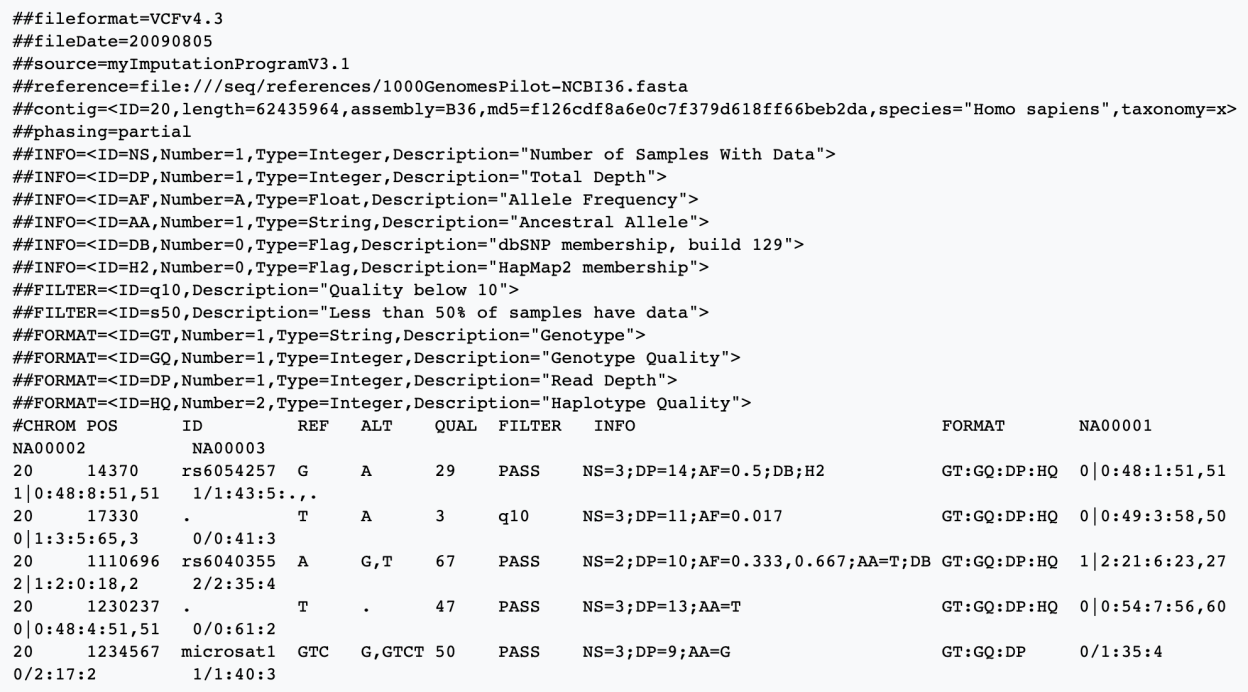


图C.5 对应的比对部分示例

C.3 遗传变异数据格式

C.3.1 VCF/BCF格式

Variant Call Format（VCF）是一种制表符分隔的文本格式，用于存储个体或群体的遗传变异数据。BCF格式是VCF格式的区块压缩二进制格式（Block gzipped）。VCF格式分为文件头（header）和变异数据两部分，文件头由#标识，标识了文件中的相关信息。变异数据内容包括染色体（CHROM），物理位置（POS）、变异名（ID）、参考碱基（REF）、变异碱基（ALT）、质量（QUAL）、过滤（FILTER）和详细信息（INFO）等八个强制字段。除此以外，第九列格式（FORMAT）用来记录每个样本的变异信息，后续每一列代表一个样本的变异信息。具体内容参见https://samtools.github.io/hts-specs/VCFv4.2.pdf，具体示例参见图C.6。



图C.6 VCF/BCF格式示例

C.3.2 GVCF格式

GVCF格式（Genomic Variant Call Format）是一种特殊的VCF格式，由Genome Analysis Toolkit的HaplotypeCaller v3引入（Van der Auwera 2021）。与普通VCF关键区别在于，GVCF记录所有位点的碱基情况。GVCF中的记录包括我们对某一位点是否为纯合参考位点的可信程度的准确估计。GVCF格式具体示例参见图C.7。

图C.7 GVCF格式示例

##fileformat=VCFv4.2

##ALT=<ID=NON\_REF,Description="Represents any possible alternative allele at this location">

##FORMAT=<ID=AD,Number=R,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">

##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">

##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">

##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">

##FORMAT=<ID=PL,Number=G,Type=Integer,Description="Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification">

##FORMAT=<ID=SB,Number=4,Type=Integer,Description="Per-sample component statistics which comprise the Fisher's Exact Test to detect strand bias.">

##INFO=<ID=BaseQRankSum,Number=1,Type=Float,Description="Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities">

##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some reads may have been filtered">

##INFO=<ID=END,Number=1,Type=Integer,Description="Stop position of the interval">

##INFO=<ID=ExcessHet,Number=1,Type=Float,Description="Phred-scaled p-value for exact test of excess heterozygosity">

##INFO=<ID=MLEAC,Number=A,Type=Integer,Description="Maximum likelihood expectation (MLE) for the allele counts (not necessarily the same as the AC), for each ALT allele, in the same order as listed">

##contig=<ID=20,length=63025520,assembly=GRCh37>

##source=HaplotypeCaller

#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA12878

20 10001567 . A <NON\_REF> . . END=10001616 GT:DP:GQ: PL 0/0:38:99:0,101,1114

20 10001617 . C A,<NON\_REF> 493.77 . BaseQRankSum=1.632; DP=38;ExcessHet=3.0103;MLEAC=1,0 GT:AD:DP:GQ:PL:SB 0/1:19,19,0:38:99:522,0,480,578,538,1116:11,8,13,6

20 10001618 . T <NON\_REF> . . END=10001627 GT:DP:GQ:MIN\_DP:PL 0/0:39:99:37:0,105,1575

20 10001628 . G A,<NON\_REF> 1223.77 . DP=37;ExcessHet=3.0103;MLEAC=2,0 GT:AD:DP:GQ:PL:SB 1/1:0,37,0:37:99:1252,111,0,1252,111,1252:0,0,21,16

20 10001629 . G <NON\_REF> . . END=10001660 GT:DP:GQ:PL 0/0:43:99:0,102,1219

C.3.3 PLINK格式

PLINK格式是群体遗传学分析工具PLINK使用的遗传变异存储格式，在人类遗传学领域广泛应用。PLINK格式通过谱系文件PED（Pedigree）和MAP两个空白字符分隔文件，或对应的二进制BED格式（Binary Pedigree）、BIM格式（Binary Map）和 FAM（Family）格式三个空白字符分隔文本来存储群体遗传变异数据。

PED文件是一个空白字符分隔文本文件，用来存储个体的信息（如家庭ID、个体ID、父亲/母亲ID、性别、性状以及遗传突变信息），包含至少六列信息，分别是：

Family ID

Individual ID

Paternal ID

Maternal ID

Sex (1=male; 2=female; other=unknown)

Phenotype

第七列之后可以添加遗传变异，例如：

FAM001 1 0 0 1 2 A A G G A C

FAM001 2 0 0 1 2 A A A G 0 0

MAP文件记录遗传突变的位置信息，包含染色体、突变ID、遗传距离和物理距离，四列信息。

chromosome (1-22, X, Y or 0 if unplaced)

rs# or snp identifier

Genetic distance (morgans)

Base-pair position (bp units)

C.3.4 GEN/BGEN格式

GEN格式又被称为Oxford或者Chiamo格式，最初被设计用来存储impute2产生的基因型概率数据。BGEN格式是GEN格式的区块压缩形式。GEN文件包含文件头以及数据两部分。文件头的前N项必须对应着数据的变异信息，通常是家庭ID+个体ID或者个体ID的方式。当GEN文件不包含头文件时，个体信息必须存储在一个单独的表格内。

GEN的数据部分的格式：前N列必须包含变异的信息，包括变异ID（SNP ID），位置（Location），变异碱基（Alternative Allele）及参考碱基（Reference Allele）。

C.3.5 VDS格式

Variant Dataset（VDS）格式（https://hail.is/docs/0.1/overview.html#overview-vds）是Hail 1使用的一种基于Hadoop框架的分布式遗传变异数据存储格式，该格式将突变位点（variants）及注释（variant annotations）以及群体样本（sample）及注释（sample annotations）按照行和列存储；同时存储数据集的元数据（global annotation）。具体示例参见图C.8。

A picture containing chart

Description automatically generated

图C.8 VDS格式示例

C.3.6 MatrixTable格式

MatrixTable格式是一般化的VDS，是Hail2使用的基于Hadoop的分布式遗传变异数据格式。MatrixTable包括全局（global field）、行 （row field）、列（column field）以及数据（entry field）四个部分,具体内容参见（https://hail.is/docs/0.2/overview/matrix\_table.html）。

1. (资料性附录）  
   医学服务类保存规则（摘选）

|  |  |  |  |
| --- | --- | --- | --- |
| 法规归类项目 | 保存期限 | 相关法规 | 对应条款 |
| 临床项目（包括癌症基因检测等） | 2年 | 《医疗机构临床实验室管理办法(2006年卫生部)》卫医发〔2006〕73号，第三十二条 | 医疗机构临床实验室应当建立质量管理记录,包括标本接收、标本储存、标本处理、仪器和试剂及耗材使用情况、校准、室内质控、室间质评、检验结果、报告发放等内容。质量管理记录保存期限至少为2年。 |
| 产前筛查与诊断  （包括胎儿染色体异常无创产前基因检测、单基因病检测等） | 3年 | 开展高通量基因测序产前筛查与诊断临床应用试点通知《国卫妇幼妇卫便函（2015）4号》，八 | 标本、信息和资料的保存期限应不少于3年。 |
| 新生儿筛查  （包括新生儿耳聋基因检测、遗传病筛查等） | 10年 | 新生儿疾病筛查相关记录《新生儿疾病筛查技术规范(2010年版)》卫妇社发〔2010〕96号，（五） | （五）有完整的实验室检测信息资料，存档保留至少10年 |
| 胚胎冻存  （包括胚胎植入前染色体异常检测等） | 5年 | 生殖医学杂志2018年10月第27卷第10期《冷冻胚胎保存时限的中国专家共识》 | 有胚胎冻存年限的建议，没有数据保存的相关文件，但可结合胚胎冻存年折中对数据进行保存（5年）。参考文献：《冷冻胚胎保存时限的中国专家共识》 |
| 病原微生物  （包括感染病原高通量基因检测等） | 20年 | 《病原微生物实验室生物安全管理条例》第三十七条 | 实验室应当建立实验档案，记录实验室使用情况和安全监督情况。实验室从事高致病性病原微生物相关实验活动的实验档案保存期，不得少于20年。 |

1. (资料性附录）  
   知情同意书模板示例

E.1医学服务场景知情同意书模板

表 E.1 医学服务场景知情同意书模板

|  |  |
| --- | --- |
| 医学服务场景知情同意书模板 | 编写要求 |
| **检测（全称）协议书**  **（版本号及版本日期）** | 该部分医学服务场景知情同意书标题。包括医学检测服务的标题、版本号及版本日期。 |
| **须知页**  1.检测项目介绍  2.检测局限性  3.风险提示  4.检测结果与报告  5.数据安全及隐私保护说明  6.医师告知与同意 | 该部分为数据主体须知范围。包括检测项目介绍、检测局限性、风险提示、检测结果与报告、数据安全及隐私保护说明。  检测局限性应包括检测范围、无法检测内容告知等。  风险提示应包括检测结果说明、可能出现假阳性或假阴性结果告知、是否委托处理及处理方式说明等。  检测结果与报告应包括检测的结果、仅对本次送检样本负责、本检测仅为筛查结果，不作为诊断结果告知、客户服务联系方式等。  数据安全及隐私保护说明应包括数据全生命周期的所有环节中涉及到的数据控制者、遗传物质及所获取数据的用途、生物样品和数据存储时限、数据隐私保护手段、安全保护措施及应用的潜在风险等。 |
| **签字页**  检者在充分知晓上述情况的基础上，承诺以下事项：   * 我已阅读《xxx检测知情同意书》，已充分了解该检测的性质、适用人群、目的、风险和必要性。医方已履行了告知义务，对其中疑问进行了解答。并承诺保护受检者个人隐私信息。 * 我承诺提供的资料真实、完整。 * 经本人及家属慎重考虑，自愿进行检测，同意回访。 * 我授权院方处理本次检测涉及的血液、血浆和医疗废物。   **双方确认上述内容为双方意思的真实表达，院方已履行了告知义务，受检者已享有充分知情和选择的权利，签字生效。**  送检医师（签字）： 签字日期：  受检者/监护人（签字）： 与受检者关系： 签字日期：  （当受检者知情同意能力欠缺或不足时，其监护人签字生效） | 该部分为数据主体或监护人的签字页，包括送检医师及受检人的签字及日期。 |

E.2法庭科学生物样本基因信息的鉴定服务场景知情同意书模板

法庭科学生物样本基因信息的鉴定服务知情同意书模板可参考《司法部关于印发司法鉴定文书格式的通知 司发通[2016]112号》文件要求进行编写。

表 E.2 法庭科学生物样本基因信息的鉴定服务场景知情同意书模板

|  |  |
| --- | --- |
| 法庭科学生物样本基因信息的鉴定服务场景知情同意书模板 | 编写要求 |
| **司法鉴定委托书**  **（编号）** | 该部分法庭科学生物样本基因信息的鉴定服务场景知情同意书标题及编号。 |
| 1.委托方信息  2.司法鉴定机构信息  3.委托鉴定事项描述  4.是否重新鉴定  5.参考的技术标准  6.特殊情况描述  7.鉴定用途  8.与鉴定有关的基本案情描述  9.鉴定材料  10.费用及意见书发送方式  11.约定事项  12.风险提示  13.其他事项 | 该部分为数据主体须知范围，包括委托方（数据主体）信息、鉴定机构信息、鉴定事项、案情描述、鉴定意见等。  委托鉴定事项应描述需要解决的专门性问题，如：亲权鉴定、亲缘鉴定等。  特殊情况应说明有无半年内输血、是否做过骨髓移植、是否双胞胎或近亲等。  鉴定材料应包括名称、种类、样品来源、数量、性状、保存状况、收到时间等。  约定事项应包括鉴定时限、材料使用、回避、材料真实性等。  风险提示应描述鉴定中可能存在的风险告知。 |
| **签字页**  委托方： 日期：  司法鉴定机构： 日期：  身份真实性承诺： 日期：  委托事项变更：  邮寄单号/领取人签名： 发放日期： | 该部分为签字页，包括委托方、司法鉴定机构的签字及日期。 |

E.3消费级服务知情同意书模板

表 E.3 消费级服务场景知情同意书模板

|  |  |
| --- | --- |
| 消费级服务场景知情同意书模板 | 编写要求 |
| **xxx检测（全称）协议书**  **（版本号及版本日期）** | 该部分为消费级服务场景知情同意书标题，包括消费级服务的标题、版本号及版本日期。 |
| **须知页**  1.服务定义  2.服务介绍  3.接受条款  4.服务条件  5.服务范围  6.风险告知  7.数据安全及隐私声明  8.用户知悉  9.用户责任  10.购买方式  11.禁止转售  12.中止服务 | 该部分为数据主体须知范围。包括服务定义、服务内容介绍、服务条件及范围、风险提示、数据及隐私安全告知等。  服务定义应包括针对该协议中的数据主体、数据控制者（一般为XX公司）、提供的服务内容以及涉及到的个人相关信息进行定义。  接受条款应包括数据主体的权利告知，告知内容应包括接受服务、知情同意、签署条件、撤回授权等。  风险告知，包括可能产生的健康或心理问题的风险、样本质量导致的检测失败的风险、误差风险、个人基因信息主动分享的风险等。  数据安全及隐私保护说明应包括数据全生命周期的所有环节中涉及到的数据控制者、遗传物质及所获取数据的用途、生物样品和数据存储时限、数据隐私保护手段、安全保护措施及应用的潜在风险等。 |
| **签字页**  检者在充分知晓上述情况的基础上，承诺以下事项：   * 我已阅读《xxx检测服务协议》，已充分了解该检测的性质、适用人群、目的、风险和必要性。服务方已履行了告知义务，对其中疑问进行了解答。并承诺保护受检者个人隐私信息。 * 我承诺提供的资料真实、完整。 * 经本人及家属慎重考虑，自愿进行检测。   **双方确认上述内容为双方意思的真实表达，服务方已履行了告知义务，受检者已享有充分知情和选择的权利，签字生效。**  受检者/监护人（签字）： 与受检者关系： 签字日期：  （当受检者知情同意能力欠缺或不足时，其监护人签字生效） | 该部分为数据主体或监护人的签字页，包括签字和日期。 |

E.4研究开发场景知情同意书模板

表 E.4 研究开发场景知情同意书模板

|  |  |
| --- | --- |
| 研究开发场景知情同意书模板 | 编写要求 |
| **xxx研究（全称）知情同意书**  **（版本号及版本日期）** | 该部分为研究开发场景知情同意书标题，包括知情同意书的标题、版本号及版本日期。 |
| **须知页**  我们将要开展一项xxx研究，现在邀请您参加这项研究。这项研究的负责人是xxx单位的xxx，资助方是xxx。  请仔细阅读本知情同意书后慎重做出是否参加研究的决定。参与这项研究与否，完全取决于您自己的决定。当研究人员或者您的主管医生和您讨论知情同意书时，您可以随时提问并让他/她向您解释您不明白的地方。您可以与家人、朋友、病友、医生或其他您信任的人讨论之后再决定是否参加。  如果您认为您有兴趣参加这项研究，请认真阅读以下材料。我们首先声明：您在阅读后作出不参加这项研究的决定，不会影响您日常的医疗或参与项目发起方组织的其他科研、教育活动。详细情况请阅读本《知情同意书》。  （根据情况增补：若您目前正参加其他某项研究，或使用某些药物，请告知研究人员或您的主管医生）   1. 为什么开展这项研究? （需简要阐明研究前期工作基础、目的、意义、必要性） 2. 本研究项目来源（需阐明项目承担单位、通过的相关伦理、学术的讨论审批等）？ 3. 这项研究怎样进行，我该怎样参与？ 4. 哪些研究人员可以使用我的样本，他们可以拥有关于我的哪些信息？ 5. 我的样品及数据是如何储存的？ 6. 我的样本和信息将被保存多久？ 7. 我参加这项研究有无任何费用或补偿？ 8. 我参加这项研究有何受益？ 9. 我参加这项研究有何风险和不适？对我的家庭、所在社群有何风险？ 10. 这项研究如何保护我的个人信息和隐私？ 11. 如果我因为参加这项研究而受伤将如何处理？ 12. 我能获得使用我样本和数据的研究结果吗？ 13. 这项研究结束后剩余样本（如有）和产生的数据如何处理？ 14. 我能退出这项研究，并终止对我样本及数据的存储和使用吗？ 15. 如果我有问题或困难，想了解研究进展或了解自己作为研究对象的权利，该与谁联系？   如果您有与这项研究相关的任何问题，请联系（项目负责人）博士/医生（电话、邮箱）。（建议同时提供另一名研究者的联系方式）  如果您有与自身权利/权益相关的任何问题，或者您想反映参与这项研究过程中遭遇的困难、不满和忧虑，请联系xxx生命伦理委员会，联系电话：xxxx-xxxxxxxx，电子邮件：xxx@xxx.xxx | 该部分为数据主体须知范围。主要包括研究的目的及必要性等、项目来源、数据如何存储、存储时限、风险及收益、个人信息及隐私如何保护、如何退出等。 |
| **签字页**  我已告知该参与者xxx（项目名称）的研究背景、目的、步骤、风险及获益情况，给予他/她足够的时间阅读知情同意书、与他人讨论，并解答了其有关研究的问题；我已告知该参与者当遇到相关问题时的联系方式；我已告知该参与者他/她可以退出这项研究；我已告知该参与者他/她将得到这份知情同意书的副本，上面包含我和他/她的签名。  研究者：（签名） 日期：  **知情同意声明**  我已被告知这项研究的背景、目的、步骤、风险及获益情况。我有足够的时间和机会进行提问，问题的答复我很满意。我也被告知，当我有问题、想反映困难、顾虑、对研究的建议，或想进一步获得信息，或为研究提供帮助时，应当与谁联系。我已经阅读这份知情同意书，并且同意参加这项研究。我知道我可以在研究期间的任何时候无需任何理由退出这项研究及退出方式。我被告知我将得到这份知情同意书的副本，上面包含我和研究者的签名。我声明，   * 我同意参与“研究题目全称”； * 我 □不愿意获知这项研究中关于我个人的研究发现（此句话视项目研究情况决定是否需要）。   （正楷）  参与者姓名（年满14岁） 参与者签名（年满14岁） 日期  （当样本提供者不满18岁时，或知情同意能力欠缺、不足时，增加或替换以下方式）  （正楷）  法定监护人姓名 法定监护人签名 与参与者关系 日期  \*参与者：指样本提供者或其法定监护人。  \*参与者/法定监护人必须亲自注明其签名日期。  \*研究者或研究者指定进行知情同意程序的人（要求有适当资质且经过培训），必须在参与者签署同意书的同一次访谈期间在同意书上签字，并注明日期。 | 该部分为数据控制者（研究者）及数据主体（参与者）的签字页，包括签字和日期。 |

E.5基因识别数据及关联信息存储（捐赠保藏）知情同意书模板

表 E.5 基因识别数据及关联信息存储（捐赠保藏）知情同意书模板

|  |  |
| --- | --- |
| 基因识别数据及关联信息存储（捐赠保藏）知情同意书模板 | 编写要求 |
| **基因识别数据及关联信息捐赠知情同意书**  **（版本号及版本日期）** | 该部分为基因识别数据及关联信息存储（捐赠保藏）知情同意书标题，包括知情同意书的标题、版本号。 |
| **须知页**  **关于本知情同意书**  请您仔细阅读此知情同意书。同意书中包含将样本和数据捐献给xxx（数据保藏机构名称）保藏和使用。我们团队的成员将与您沟通有关参与此捐献的细节。同意参与捐献的人员被称为“捐献者”。该术语将在本同意书中通篇使用。如果您对捐献或此文档有任何疑问，请询问我们。是否参与捐献取决于您本人的决定。如果您决定参与，则必须签署此同意书以示您的许可。我们将为您提供此表格的签名副本供保留。  若有资格捐献的人由于健康状况或年龄不足18岁（未成年）而无法给与同意许可，我们将要求该人的授权代表或者孩子的父母或法定监护人表示同意。在整个同意书中，“您”始终指代参与研究的捐献者。  1.为什么要捐赠生物样本及信息?  （捐献生物样品及信息的意义、存储机构的位置、样品类型、经费来源等信息）  2.我该怎样参与捐献？  （捐赠的步骤，包括知情同意、采集的关联信息、样品采集方式、测序机构。还包括其他事项，如：用于科研的照片或录像、家庭成员参与意愿、其他类型的样本、过去的生物样本、未成年人知情同意、研究成果的返还等。）  3.哪些人员可以使用我的样本，他们可以拥有关于我的哪些信息？  （使用人员包括学术机构的研究人员和非营利性公司的研究人员。数据可能会在特定研究项目中与研究人员合作的营利性公司共享，不会被出售给任何人以获取利润，只提供去标识化的样本。）  4.我的样品及数据是如何储存的？  （样本及数据的保存方式，如：容器、编码、加密等）  5.我的样本和信息将被保存多久？  （没有销毁的时间，可能被存储并无限期地用于研究，以便用于将来的研究）  6.我的隐私将如何被保护？  （隐私信息加密并隔离存储，只有相关人员进行存储、使用和共享，如：相关的研究及工作人员、生物样本库相关人员、伦理委员会、供独立认证和监督的组织的人员、政府有关部门人员等）  7.我的捐献有无任何费用或补偿？  您的捐献是免费的（即使基于您的样本和信息进行的研究导致了产品或技术的开发或销售）。  8.参与捐献对我有何受益？  （有助于了解与自己相关的前沿研究进展，直接或间接改善捐献者及家庭成员的健康）  9.参加此研究对我有何风险？  （包括采样时的身体不适、家系信息滥用的担忧、新发现的遗传突变的心理负担、信息被非授权使用的风险等）用针采血时，您可能会有一些身体不适（可能会出现瘀伤、头晕、昏厥或针头处感染等）  10.我能获得使用我样本和数据的研究结果吗？  （通常情况下，不会获得参与的研究的结果）  11.如果我因为参与捐献而受伤将如何处理？  （告知发生以外伤害，将不会得到赔偿）  12.我能退出这项研究，并终止对我样本及数据的存储和使用么？  （有权在任何时候撤回许可，样品及信息将被销毁，提供联系人员及方式）  13.如果我有问题或困难，想了解研究进展或了解自己作为研究对象的权利，该与谁联系？  （提供联系人员及方式）  如果您想与不直接参与生物样本库项目的人交谈，请联系伦理委员会办公室：\_\_\_\_\_\_ | 该部分为数据主体须知范围。主要包括研究的目的及必要性等、项目来源、数据如何存储、存储时限、风险及收益、个人信息及隐私如何保护、如何退出等。 |
| **签字页**  我已经阅读这份知情同意书，知晓我的各项权益，并且同意捐献我的样本和数据。  （正楷）  参与者姓名（年满14岁） 参与者签名（年满14岁） 日期  （当样本提供者不满18岁时，或知情同意能力欠缺、不足时，增加或替换以下方式）  （正楷）  法定监护人姓名 法定监护人签名 与参与者关系 日期  （当样本提供者不能读写时，增加或替换以下方式）  （正楷）  证人姓名 证人签名 与参与者关系 日期  \*参与者：指样本提供者或其法定监护人。  \*参与者/法定监护人必须亲自注明其签名日期。  \*指定进行知情同意程序的人（要求有适当资质且经过培训），必须在参与者签署同意书的同一次访谈期间在同意书上签字，并注明日期 | 该部分为数据主体（参与者）的签字页，包括签字和日期。 |

E.6基因识别数据及关联信息共享知情同意书模板

表 E.6 基因识别数据及关联信息共享知情同意书模板

|  |  |
| --- | --- |
| 基因识别数据及关联信息共享知情同意书模板 | 编写要求 |
| **基因识别数据及关联信息共享知情同意书**  **（版本号及版本日期）** | 该部分为基因识别数据及关联信息共享知情同意书标题，包括知情同意书的标题、版本号。 |
| **须知页**  请您仔细阅读此知情同意书。我们团队的成员将与您沟通您的基因识别数据及关联信息共享的细节。同意基因识别数据及关联信息被共享的人员被称为“捐献者”。该术语将在本同意书中通篇使用。如果您对您的基因识别数据及关联信息或此文档有任何疑问，请询问我们。是否参与数据及信息的共享，取决于您本人的决定。如果您决定参与，则必须签署此同意书以示您的许可。我们将为您提供此表格的签名副本供保留。  若有资格捐献的人由于健康状况或年龄不足18岁（未成年）而无法给与同意许可，我们将要求该人的授权代表或者孩子的父母或法定监护人表示同意。在整个同意书中，“您”始终指代参与数据及信息共享的捐献者。  1.为什么要共享？  （描述数据共享的意义）  2.我该怎样参与？  （参与的步骤，包括签署知情同意、样品的类型及数量、基因识别数据及信息的类型及数量、研究成果的返还等。）  3.哪些人员可以使用我的样本？他们可以获得关于我的哪些信息？  （使用人员包括学术机构的研究人员和非营利性公司的研究人员。数据可能会在特定研究项目中与研究人员合作的营利性公司共享，不会被出售给任何人以获取利润，只提供去标识化的样本。根据最小必要原则，接受方只会获得与使用目的直接相关的基因识别数据及关联信息（不包括姓名等与身份直接相关的个人敏感信息））。  4.如何共享给其他人？  （评估数据及信息使用者的使用目的、合理性及数据安全能力，签署共享协议，明确接收方只能自己使用）  4.我的样品及数据是如何储存的？  （样本及数据在数据接收方的保存方式，如：容器、编码、加密等）  5.我的样本和信息将被保存多久？  （没有预定的销毁时间，可能被存储并无限期地用于研究，以便用于将来的研究）  6.我的隐私将如何被保护？  （隐私信息加密并隔离存储，只有相关人员进行存储、使用和共享，如：相关的研究及工作人员、生物样本库相关人员、伦理委员会、供独立认证和监督的组织的人员、政府有关部门人员等）  7.数据共享有无任何费用或补偿？  免费的（即使基于您的样本和信息进行的研究导致了产品或技术的开发或销售）。  8. 数据共享对我有何受益？  （有助于了解与自己相关的前沿研究进展，直接或间接改善捐献者及家庭成员的健康）  9.数据共享对我有何风险？  （包括家系信息滥用的担忧、新发现的遗传突变的心理负担、信息被非授权使用的风险等）  10.我能获得使用我样本和数据的研究结果吗？  （通常情况下，不会获得参与的研究的结果）  12.我能退出数据共享，并终止对我样本及数据的存储和使用么？  （有权在任何时候撤回许可，数据控制者处的样品及信息可以继续存储或将被销毁，提供联系人员及方式；将不可能销毁已经交给研究人员的数据） | 该部分为数据主体须知范围。主要包括数据共享的意义、参与方式、共享何人、如何共享、风险及收益、个人信息及隐私如何保护、如何退出等。 |
| **签字页**  我已经阅读这份知情同意书，知晓我的各项权益，并且同意共享我的样本和数据。  （正楷）  参与者姓名（年满14岁） 参与者签名（年满14岁） 日期  （当样本提供者不满18岁时，或知情同意能力欠缺、不足时，增加或替换以下方式）  （正楷）  法定监护人姓名 法定监护人签名 与参与者关系 日期  （当样本提供者不能读写时，增加或替换以下方式）  （正楷）  证人姓名 证人签名 与参与者关系 日期  \*参与者：指样本提供者或其法定监护人。  \*参与者/法定监护人必须亲自注明其签名日期。  \*指定进行知情同意程序的人（要求有适当资质且经过培训），必须在参与者签署同意书的同一次访谈期间在同意书上签字，并注明日期 | 该部分为数据主体（参与者）的签字页，包括签字和日期。 |

参 考 文 献

[1] 中国医师协会医学遗传医师分会，中华医学会儿科学分会内分泌遗传代谢学组，中国医师协会青春期医学专业委员会临床遗传学组，等。全基因组测序在遗传病检测中的临床应用专家共识[J]。中华儿科杂志，2019，57 （6）：419-423

[2] Adriana Deac, 2018. "Regulation (Eu) 2016/679 Of The European Parliament And Of The Council On The Protection Of Individuals With Regard To The Processing Of Personal Data And The Free Movement Of These Data," Perspectives of Law and Public Administration, Societatea de Stiinte Juridice si Administrative (Society of Juridical and Administrative Sciences), vol. 7(2), pages 151-156, December.

[3] 中华人民共和国国务院.中华人民共和国人类遗传资源管理条例[EB/OL].2019-06-10. http://www.gov.cn/zhengce/content/2019-06/10/content\_5398829.htm.

[4]卫生和计划生育委员会.涉及人的生物医学研究伦理审查办法[EB/OL].2016-10-12.http://www.nhc.gov.cn/cmssearch/xxgk/getManuscriptXxgk.htm?id=84b33b81d8e747eaaf048f68b174f829.

[5] Band, Gavin, and Jonathan Marchini. 2018. "BGEN: A Binary File Format for Imputed Genotype and Haplotype Data." BioRxiv, April, 308296.

Cock, Peter J. A., Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. 2010. “The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants.” Nucleic Acids Research 38 (6): 1767-71.

[6] Van der Auwera, G. GVCF - Genomic Variant Call Format. Accessed February 8, 2021. https://gatk.broadinstitute.org/hc/en-us/articles/360035531812-GVCF-Genomic-Variant-Call-Format

[7] Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and and David Haussler. 2002. “The Human Genome Browser at UCSC.” Genome Research 12 (6): 996–1006.

[8] Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” Bioinformatics 25 (16): 2078–79.

[9] Lipman, D. J., and W. R. Pearson. 1985. “Rapid and Sensitive Protein Similarity Searches.” Science 227 (4693): 1435–41.

[10] Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” American Journal of Human Genetics 81 (3): 559–75.

[11] Samtools/Hts-Specs. (2012) 2021. TeX. samtools. https://github.com/samtools/hts-specs.

[12] 姜先华。DNA分析技术在法庭科学领域的进展。现代仪器与医疗。2004，（6）。

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_