

# 国际人用药品注册技术协调会

## ICH 协调指导原则

### 基因组采样和基因组数据管理指导原则

**E18**

**(中文翻译公开征求意见稿)**

**当前第 4 阶段版本**

**日期：2017 年 8 月 3 日**

本指导原则由相应ICH专家工作组制定，并经监管部门根据ICH流程进行协商。在该流程的第4步，建议ICH地区监管机构采纳最终草案。

## E18

### 文件历史

当前版本为第 4 步版本

编码	历史	日期
E18	在第 2 步中获得 ICH 大会批准，并且发表供公众参考。	2015 年 12 月 10 日
E18	在第 4 步中获得 ICH 大会管理成员批准，并推荐 ICH 监管机构采纳	2017 年 9 月 6 日

法律通告：该文件受到版权保护，可以持公共许可证使用、翻印、合并至其他著作中、改编、修正、翻译或传播，前提是文件中的 ICH 版权在任何时候均被承认。如果对文件进行改编、修正或翻译，必须采取合理措施来明确标明、区分或确认所作的变更或所依据的原始文件。应避免留下“原始文件更改、修正或翻译已经获得 ICH 背书或资助”的印象。

文件应照现在的样子提供，不应附带任何保证。ICH 或原文件的作者在任何情况下均不对索赔、损失或使用文件所产生的其他责任负责。

上述准许不适用于由第三方提供的内容。因此，对于版权归属第三方的文件，必须征得版权所有人的翻印许可。

ICH 协调指导原则  
基因组采样和基因组数据管理指导原则  
E18  
ICH 共识指南

目录

<b>1. 前言</b> .....	<b>1</b>
1.1.指导原则的目标.....	1
1.2. 背景 .....	1
1.3.指导原则的范围.....	1
1.4. 通则 .....	2
<b>2. 基因组采样</b> .....	<b>2</b>
2.1.样本采集和处理.....	2
2.1.1. 样品类型 .....	3
2.1.2. 样品采集时间.....	3
2.1.3. 样品保存条件.....	3
2.1.4. 样品稳定性和分解.....	3
2.1.5. 样品体积和组成 .....	4
2.1.6. 影响基因组样本质量和数量的参数.....	4
2.1.7. 干扰来源.....	4
2.2.样本运输和储存.....	5
2.2.1. 样本运输.....	5
2.2.2. 样本保存.....	5
2.2.3. 样本库存的管理.....	5
<b>3. 基因组数据</b> .....	<b>5</b>
3.1.基因组数据的生成 .....	6
3.2.基因组数据的处理和保存.....	6
<b>4. 隐私和保密</b> .....	<b>7</b>
4.1.样本和数据编码.....	7
4.2. 访问和透明度 .....	7
<b>5. 知情同意</b> .....	<b>8</b>
<b>6. 结果沟通</b> .....	<b>8</b>

1. 前言

1.1. 指导原则的目标

这份指导原则的主要目的是提供临床研究中基因组采样和基因组数据管理的协调性原则。通过对关键因素建立共识，本指导原则将有助于促进基因组研究的开展。这些关键因素涉及对基因组标本和数据的准确收集，储存及优化处理。该指导原则的另一个目标是增强和提醒研究人员依据当地法律法规考虑受试者隐私、数据生成保护、知情同意获取、结果的公开等问题的意识。

这份指导原则旨在建立利益相关者间的互动，包括药物研发者、研究者和管理者，并且鼓励在临床研究中开展基因组研究。

## 1.2. 背景

人们对于临床研究中生成的基因组数据的了解和兴趣与日俱增。特别是，基因组研究可应用于药物研发的所有阶段以评价基因组与药物疗效的相关性、理解疾病机理或药物药理学。鉴别导致药物疗效差异的基因组生物标志物对于优化患者治疗、设计更为有效的研究以及拟定药物说明书标签都是价值的。另外，基因组数据的生成和解释，无论是在临床试验期间还是药物开发项目进程中，均有助于更好地理解药理学和病理学机制，并且能够鉴别出新的药物靶点。

ICH 地区的管理部门均有各自独立发布的指导原则，鼓励在药物的生命周期内进行基因组样本采集。缺少临床研究基因组采样和数据管理的 ICH 协调指南使得申办方和研究者极难在全球临床研究中通过一致的方式来采集基因组样本和开展基因组研究。

基因组样本可应用在多种分析中，包括单一基因、成组基因和全基因组，在采集时可能在临床研究目标中预先规定或未作规定。

## 1.3. 指导原则的范围

该指导原则的范围与干预性和非干预性临床研究中的基因组采样和基因组数据管理有关。基因组研究可在临床研究中或在临床研究后开展，可这一点可能在临床研究方案中预先规定或不作规定。本文件阐述了基因组样本和数据的使用（无论分析在什么时间进行），包括预先规定和非预先规定的使用。在这份指导原则中描述的基因组样本和数据与 ICH E15 中定义的脱氧核糖核酸（DNA）和核糖核酸（RNA）特征相符<sup>1</sup>。

着重点是在知情同意范畴内采集、处理、运输、储存和处置基因组样本或数据的一般性原则。根据情况，对技术层面进行讨论，了解基因组采样和数据生成领域快速演变的技术进步。

本指导原则未包括与生物样本库法规或伦理学方面相关的详细指导原则，因为这些受赫尔辛基宣言的原则以及国别法律和法规的管辖。有关隐私和数据保护相关问题同样如此。然而，本指导原则中的原理适用于任何采用人源材料开展的基因组研究。

本指导原则中推荐的是一般性原则，应根据管辖国与从事基因组研究相关的法律、法规和政策做进一步解释。

## 1.4. 通则

随着科技进步以及对基因组影响的了解不断加深，有必要将采集的样本及其所生成的数据价值最大化。因此，强烈建议在所有阶段以及临床开发研究中采集基因组样本。而且，基因组研究的质量依赖于从所有受试者中无偏倚地开展系统性样本采集和分析以便完全代表研究人群。

维护样本完整性至关重要，这会对基因组样本的科研功用产生重大影响。这些样本的质量和数量及检测方法的技术表现（例如准确度、精密度、灵敏度、特异性、重复性）将决定基因组数据的可靠性。制定基因组样本操作和处理的标准规范可促进来自不同分析平台的数据整合，同时有助于做出临床决策。基因组样本和数据应按与非基因组样本和

健康信息相似的方式妥善地保存、维护和访问。

## 2. 基因组采样

基因组研究包括多种方法和应用。这些可能包括但不局限于核酸测序和基因分型；不同类型的 RNA 分析；基因表达或调控；表观遗传修饰的检测。不断演变的技术进步有望产生新的应用。研究范围将决定样品类型、待评价的分析物、用于提取分析物的方法学，以及用于稳定和保存供基因组试验用样本的方法学。样本质量和数量会影响生成数据的准确度和可靠性。因此，生物样本的采集，处理和制备是研究过程中的关键步骤。通过拟定和记录基因组样本采集、处理、运输和储存的标准程序，将分析前变异性最小化。此类程序和质量监测应与样品、分析物和所开展的试验类型相适应。应在实施前定义、记录和核对样品处理和制备等分析前流程。重要的是记录样本采集的时间、方法、地点（如，研究中心）和条件（如，保存温度和时间，固定时间）。程序的任何变化和偏离均应在每一个标本的生命周期中进行详细记录。记录所有样本在采集、处理和分析等所有阶段的保管链，包括每步的时间。强烈建议实施质量控制计划。一般来讲，应制定采集、处理、运输和储存的指导说明以确保从采集之时起至试验时每个步骤的生物学样本稳定性。

### 2.1. 样本采集和处理

当制定样本采集和处理策略以确保样本适用于基因组试验时，应考虑到多个分析前变量。如果参加临床试验的中心采用不同样本采集和处理程序，则后续试验的性能会随研究中心不同而有所差异。这可能影响数据的可判读性和可合并性，并且可能导致出现不可靠的结果。参加研究的所有中心的人员均应接受适当培训以了解如何遵循标准化程序。根据适当的生物安全性规范、受试者隐私法规和知情同意，采集样品并贴标。

#### 2.1.1. 样品类型

核酸可以从多种临床样品类型和基质中提取（例如全血、组织、颊粘膜拭子、唾液、骨髓穿刺液、尿液、粪便）。一些新型组织衍生型核酸来源（例如无细胞的游离 DNA 和循环肿瘤细胞）不断涌现，可能需要不同的分离方法。文中详细阐述的原则也适用于此类样品。待采集的样品类型应与预期用途相匹配。例如，一些类型的样品可用于 DNA 和 RNA 研究，而一些样品由于缺乏分析物稳定性，可能不适用于 RNA 分析。

在儿童受试者中，血液或其他组织的数量有限，因此可考虑非侵入性的样本采集来代替，如唾液、干血斑或皮肤碎屑可以考虑。对于某些类型的样本（如血液、肌肉活检）应注意无菌收集。当使用可能带有宿主外 DNA 和 RNA 污染风险的生物材料（如口腔拭子、唾液）时，应谨慎操作。

#### 2.1.2. 样品采集时间

当制定样本采集策略时，应结合临床研究目的来考量受试者间和受试者自身变异性。例如昼夜变化或给予的治疗会影响基因表达，在选择采样时间点时应予以考虑。表观遗传学，如 DNA 甲基化也可能随着时间的推移而改变（例如，受试者年龄）。尽管胚系 DNA 的序列是相对稳定的且不会随时间而变化，从肿瘤 DNA 和 RNA 中获取的信息可能受样本

采集来源、方法和/或样本采集时间的影响。

### 2.1.3. 样品保存条件

采集容器，是否需要添加剂、稳定剂或防腐剂及其类型将取决于目标核酸、样品类型、所需样本量或体积以及潜在分析法和/或技术。例如，血液或骨髓穿刺液样本应采集至含抗凝血剂或指定核酸类型适用的添加剂的试管中。组织样本可以在液氮中速冻或放置在适当的防腐剂中。

通常对组织进行固定以便长期保存。组织固定时应仔细考量的参数为固定剂的类型、固定时间、湿度、氧合和温度、以及与下游核酸提取方法的相容性。建议在临床研究样本采集前评价固定剂和添加剂对拟测定分析物以及所开展的试验类型的影响。另外，样品组织类别和体积会影响最佳固定时间，因此应纳入考虑范畴。初次固定后的样本处理方法还可能影响样品的完整性。

### 2.1.4. 样品稳定性和分解

应采取适当的处理措施来防止在样本采集和处理过程中发生核酸分解和基因组谱的改变。受 pH、缺氧、存在核酸内切酶和/或其他组织特异性参数的影响，可能发生核酸碎裂以及明显的基因表达变化。另外，从样本采集至冷冻、固定或处理的时间以及保存时间应根据需要进行优化。所采用的参数应记录在样本采集和处理指南、培训材料和样本报告中。建议对储存和处理条件进行监测。例如，应监测温度的可能变化，并记录以确保样本间一致性。

### 2.1.5. 样品体积和组成

液体样本的采集体积是一个需要审慎考量的问题。应考量指定用途所需的最少组织或细胞含量（例如，分析方法）以降低受试者的负荷。最优组织量取决于组织的细胞性（例如对于多细胞的组织类型，少量样本足以满足需求）以及特定细胞类型在整个样品中的相对比例（例如活检标本中的肿瘤面积/非肿瘤面积）。在可获取的组织量有限的情况下，可以考虑采集其他生物样本（见 2.1.1 部分）。由于肿瘤组织可能表现出分子异质性（嵌合现象）并且肿瘤活检通常包含部分正常组织，所以在基因组分析之前对样本进行有记录的病理学评价大有裨益。在采集成对样本时（例如肿瘤相对于正常组织、治疗前相对于治疗后样本或产前样品相对于产后母体样品），需要额外考量对比问题（例如匹配的样本、细胞类型）。

### 2.1.6. 影响基因组样本质量和数量的参数

在其他众多因素中，来源样品的质量会影响提取的核酸质量和产量（见 2.1.5 部分）。因此，应针对处理条件和所采用的样品类型，定义和验证提取流程。样品类型的不同特征和成分会影响核酸的回收率，当选择核酸提取方法时，应考虑这些因素。例如，不同组织和体液样品的细胞裂解流程有所不同。除去特定细胞成分的工艺也有所不同，取决于样品的组成。值得注意的是，这些过程可能会影响基因表达，并导致假性结果。如果 DNA 和 RNA 均从同一样品中提取，应判定提取能否最好同时进行或组织样品是否应在采集时份。由于 RNA 相对于 DNA 的不稳定性所致，当分离 RNA 时需要采取额外预防

措施，例如使用不含 RNA 酶的仪器和试剂。在提取核酸前反复冷冻和解冻样品会影响基因组样本的完整性，应当尽量避免或者另外进行评价。为确定提取的目标核酸的质量和数量对于待开展计划的下游基因组试验是否充分，应采用适当的质量控制方法。

### 2.1.7. 干扰来源

潜在干扰和污染来源可能影响基因组试验的性能，这些包括内源性和外源性物质。鉴别样品类型中正常存在的内源性物质（例如会影响聚合酶链式反应（PCR）效率的血液血红蛋白或皮肤黑色素）以及会干扰特定试验方法的外源性物质（例如抗凝血剂、其他添加剂、固定剂、核酸分离所使用的试剂）至关重要。应在检测方法开发中阐明潜在干扰因素对该方法性能的影响。

## 2.2. 样本运输和储存

运输和储存条件将随样本类型和目标核酸而变化。一般来讲，样本在运输和储存过程中不应暴露于会影响目标核酸稳定性的环境下。

### 2.2.1. 样本运输

适当的运输条件应在样本运输前确立。为确保样品和/或提取的样本在可接受的条件下运输，应记录运输和接收日期以及接收时样品的近似温度。如果可能，样本应在样本类型和兴趣分析物适用的指定保存温度下运输。应记录相对于指定运输参数的偏离。

### 2.2.2. 样本保存

强烈建议样本在药物研发项目中以及之后长期保存以允许重复使用和/或未来使用。样品或提取的核酸的保存条件应适用于指定的基因组试验用途。建议将样本和提取的核酸分成多个小份保存以避免反复冷冻和解冻以及潜在污染。在不同的地方保存备份样本以避免同时全部丢失。如果样本反复使用以及经历冷冻/解冻循环，应记录每次冷冻/解冻循环，包括每个步骤的温度和时间。

样本保存需要物质基础设施以及稳健的实验室信息和数据管理系统。当将样本放置在生物储藏库中时应考量的因素包括遵守质量保证和质量控制计划、样本跟踪系统、信息安全、当地法律和知情同意。强烈建议将样本储存在配有适当备用电源系统和灾难应急计划的物质基础设施中。最为重要的是在任何时候均应明确样本的责任方，并且记录保管链。样本保存不应长于知情同意文件中描述的允许总保存时间。另外，应建立操作规程以确保当受试者撤回同意时或在声明的保存期结束时正确销毁样本。

### 2.2.3. 样本库存的管理

针对下述方面，监测和管理样本库存：样本使用知情同意、相对于样本留存政策的保留时间长度以及从生物储库中撤回样本的请求以及样本销毁记录。在使用每份样本之前，应对每个样本在前述方面的一致性进行确认。基因组的一致性可以用来在样本间进行身份识别（例如，样本备份，肿瘤/正常配对，治疗前和治疗后配对）。

### 3. 基因组数据

人类基因组数据可能源于胚系（从父代遗传）、体细胞（例如肿瘤组织突变）或线粒体（例如用于追踪母系遗传）来源。来自人的生物学样品可能带有非人类的基因组分子（例如微生物 DNA 或其他潜在传染病原）。生成的基因组数据类型取决于分析物和应用的技术平台。为进行全面的基因组对比，从单例受试者的健康和疾病组织，不同的组织类型和/或在不同时间点采集多份 DNA 或 RNA 样本是适当的。

#### 3.1. 基因组数据的生成

基因组数据可以由多种不同的速发展中的技术平台和方法产生。受试者的广泛基因组图分析技术上是可行的，并且生成的数据可以保存并随时间推移重复使用。重要的是根据基因组数据的指定用途来选择适当的平台和方法。因此需了解数据生成过程中是使用研究级别的方法还是经过了多方验证的方法。当基因组数据将被用于临床决策时，应根据当地法规和政策，考虑进行适当水平的检测方法验证。在探索性环境下，基因组数据可通过使用研究级别的试剂和器材获得，这些试剂与器材是未经过临床验证的。然而，即使在不需要进行临床级别验证的探索性环境下，也应进行充分的分析验证，以确保足够的准确性和一致性，以便对结果进行可靠的解释。

对于基因组研究，对所有分析阶段的处理和分析工作流程应进行详细记录，其详细程度如同对上游样本采集和处理的记录要求一样。该文档应该包括分析的每个阶段所使用的工具、版本和参数、所使用的参考基因组数据的来源和版本以及所使用的数据库（例如基因组构建、转录组装配、变异注释数据库）以及用于处理数据的计算环境和资源。应为工作流程的所有阶段设置适当的质量控制（QC）程序和度量阈值。在进行下游分析之前，应该对工作流的每个阶段进行这种 QC 程序和阈值的确定。此外，在可能的情况下，对于给定类型的检测和预期输出，应确定 QC 程序和阈值，以确保不同数据集之间的一致性。同样地，生物信息学工具、算法和相关参数应在下游分析之前选择，并且尽可能对给定类型的检测、分析物和实验目的的输出保持一致。

虽然基因组学中的信息学方法不断变化和发展，但可能有许多工具、资源和分析方法可供选择，该领域已经定义了“金标准”的方法或资源，在设计工作流程时应考虑这些问题。强烈推荐使用公开可用的注释资源，以便跨平台比较和整合来自不同研究的基因组学和非基因组（如蛋白质组学）的结果。最后，对于任何算法（包括研究人员或临床医生在解释结果时使用的简单启发式）或用于将基因组数据与临床或生物学数据相结合用于治疗决策或研究目的的统计检验，应进行适当记录。

申办方应确保标准地使用样本和基因组数据，遵循样本用于生成基因组数据这一有目的的和允许的用途。基因组数据的使用应遵守方案、以及每一个地区/国家的知情同意流程。

#### 3.2. 基因组数据的处理和保存

重要的是理解不同类型的基因组数据是如何生成、处理、分析和储存的。一般来讲，1 台仪器会生成 1 个或更多个原始数据文件，随后进行处理并转化为可与临床或生物学数据整合的格式。除了最后处理的数据集之外，建议保留可维持原始数据完整特征的数据文件夹；这些可能是在工作流程中记录的原始素材或者衍生出来用于分析的数据，都可



以用来还原原始数据。基因组数据文件应存储在可长期保存的安全介质中。另外，应有链接将基因组数据与其他临床数据相联系以允许现在或未来使用（根据情况）。基因组样本可应参与者要求销毁，但是数据销毁有悖于科研可信用度原则，特别是在临床研究背景下。事实上，一旦数据被分析并纳入研究结果，数据就不会被破坏而不会危及临床研究的科学完整性。因此，建议一旦数据产生及在研究中应用，则保存数据。在适用的情况下，可能需要制定程序以根据受试者要求使基因组数据得到适当的处理。

## 4. 隐私和保密

基因组样本和数据的处理和操作应以保护受试者隐私的方式进行。对于基因组数据，与其他临床数据相同，编码方法以及安全性和访问程序均有助于保密。在样本采集，运输，分析和保存的每一步中应执行以编码计划和访问限制为蓝本的适当安全性措施。应适当考虑每个辖区的数据保护和保密法规和政策。

### 4.1. 样本和数据编码

基因组数据应按与其他临床数据相同的高保密标准对待。ICH E15 描述了多种基因组样本和数据编码方式，包括单一编码和双重编码。为降低复杂性和错误概率，建议基因组样本和数据采用单一编码，但是应与当地法规或法律保持一致。ICH E15 中定义的匿名化具有限制，在这种编码方式下，无法再次识别受试者，因为编码代码已被清除。这是因为随着基因组信息和分析方法的日益增多，不能完全通过删除受试者的标识符和唯一代码之间的联系来防止重新识别每个研究参与者。此外，匿名有两方面的影响：1) 该过程使得先前未曾链接的基因组数据无法与表型数据相联系，和 2)，匿名化不允许在撤除同意后销毁样本或开展长期临床监测。在处理基因组数据时，研究人员和发起组织应尊重适用的隐私和数据保护条例和法规。

### 4.2. 访问及透明度

使用基因组样本和数据可能涉及根据知情同意随时间推移反复访问（建议广泛同意允许分享和转发：见第 5 部分）。这种访问可以在申办方，或申办方监督下的合作者或外部研究人员中进行。这将包括单个级别的数据和/或汇总结果。应制定政策和程序，以确保使用户访问日志对所有基因组样本和数据设置严格的访问权限，类似于其他临床数据。有关访问的政策和程序应考虑危及个体研究参与者隐私的风险，以及危及数据质量和解释的风险。

在经过同意的研究机构以及共享样本与数据的外部机构都应采用基于风险分级的基因组样本和数据访问方法。当样本储存、基因组分析或数据保存外包时，合同应规定责任方通过适当方式监督外包机构以确保样本和/或数据受到适当保护。

与外部组织或研究人员共享数据和/或样本可以促进医学科学进步且有多个方面的好处，并最大限度地提高研究成果的透明度。应充分遵守不同国家和地区关于与第三方共享个体水平数据包括公共数据库的法律法规。

## 5. 知情同意

根据 ICH E6， 获取知情同意是药品临床试验质量管理规范（GCP）的一部分。基因组

研究的知情同意书或者纳入临床研究同意书中或者单独获取。基因组研究应依照适用当地法规在知情同意的范围内开展，这包括采集和储存基因组样本和数据。知情同意应当用简单的语言描述采集的生物材料的种类和数量、采集程序和返回基因组数据的位置。如果有机会获得遗传咨询，最好将信息包括在知情同意书中。应特别考虑仅在征得受试者法定代表或监护人同意后才能入组至研究中的受试者（例如年幼受试者、严重痴呆受试者）。

虽然当下知情同意实践由各当地法规指引，但建立全球认可的基因组样本采集和使用的知情同意共识要素可以极大地助力基因组研究。

基因组样本采集和使用的知情同意应允许广泛地对样本进行分析（例如基因组、转录组、全基因组测序），无论分析在什么时间进行。在理想情况下，知情同意应允许广泛地对样本进行使用，例如检测方法开发、疾病研究，药物反应或药物警戒。应当遵守和尊重地方法规和政策。

## 6. 结果沟通

在临床研究背景下的基因组研究，旨在评估药物反应的基因组相关性，促进对疾病生物学的理解和/或确定药物药理学的机制。这类研究有时可能会产生研究课题主要目的之外的意外发现，但这些意外发现的数据是可能具有临床意义的。其中一些数据还可能指导临床治疗。例如在并非旨在研究癌症风险的研究中，从全基因组测序数据中鉴定出 BRCA1 突变。

因此，鼓励在研究中生成基因组数据的研究机构和申办方持有将研究结果发回给受试者的立场并采取相应机制。这种立场应明确表明是传达计划研究的主要研究结果还是意外发现的结果，是两者都沟通还是都不沟通。在理想情况下，还应根据情况阐明传达时间（在临床研究中或临床研究后），由谁传达（研究人员，医生，遗传顾问）以及向谁传达（受试者或如果是儿童和失能个体，则为主要护理者或法定监护人）。在将结果传达给受试者时，应评估遗传咨询的针对性；应在临床上解释结果对治疗决策的影响，并与受试者（或主要护理人员或法定监护人）进行讨论。应尊重受试者是否希望并同意收到此类信息。此外，应考虑所应用的检测方法及其验证水平，因为这可能影响结果的准确性和可靠性。