



中华人民共和国国家标准

GB/TXXXXX—XXXX

基于高通量测序的染色体数目异常和结构 变异命名规范

Naming specification on chromosome numerical abnormalities and structural
variations based on high-throughput sequencing

(点击此处添加与国际标准一致性程度的标识)

草案版次选择

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX-XX-XX 发布

XXXX-XX-XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

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专利的责任。

本文件由国家药品监督管理局提出。

本文件由全国医用临床检验实验室和体外诊断系统标准化技术委员会（SAC/TC 136）归口。

本文件起草单位：

本文件主要起草人：

基于高通量测序的结构变异命名规范

1 范围

本文件规定了基于高通量测序的染色体数目异常和结构变异命名规范。

本文件适用于人类基因组学和遗传学领域基于高通量测序对染色体数目异常和结构变异检测结果的命名。

2 规范性引用文件

本文件参照《An International System for Human Cytogenomic Nomenclature2020》（ISCN 2020 版）和《产前遗传学诊断拷贝数变异和纯合区域数据分析解读及报告规范》。

3 术语和定义

下列术语和定义适用于本文件。

3.1 基因 gene

位于染色体上编码一个特定功能产物（如蛋白质或RNA分子）的一段核酸序列，是遗传信息的基本单位。

[GB/T 30989-2014, 定义 3.1]

3.2 基因组 genome

一种生物体具有的所有遗传信息的总和。

[GB/T 30989-2014, 定义 3.2]

3.3 核苷酸 nucleotide

构成核酸的基本单位，由三部分组成：五碳糖、磷酸和环状的含氮碱基。

[GB/T 30989-2014, 定义 3.3]

3.4 高通量测序 high-throughput gene sequencing

区别于传统 Sanger（双脱氧法）测序，能够一次并行对大量核酸分子进行平行序列测定的技术，通常一次测序反应能产出不低于 100Mb 的测序数据。

[GB/T 30989-2014, 定义 3.19]

3.5 结构变异 structural variation, SV

染色体上发生的大片段的改变。主要包括染色体缺失、重复、易位、倒位等，这些改变使染色体上基因的数目或排列顺序发生改变，从而导致生物性状的变化。

3.6 易位 translocation

一条染色体的断片移接到另一条非同源染色体臂上的一种结构畸变。常见的方式有相互易位、罗伯逊易位和插入易位等。相互易位指两条染色体同时发生断裂后相互交换无着丝粒断片形成两条衍生染色体的过程。罗伯逊易位又称“着丝粒融合”，发生于近端着丝粒染色体的一种易位形式，当两个近端着丝粒染色体在着丝粒部位或者着丝粒附近发生断裂后，二者的长臂在着丝粒处接合在一起，形成一条由长臂构成的衍生染色体。插入易位指两条非同源染色体同时发生断裂，但只有其中一条染色体的片段插入另一条染色体非末端部位的易位现象。

3.7 倒位 inversion

某一染色体发生两次断裂后，两个断裂点之间的片段旋转 180° 后重接，造成染色体上基因顺序重排的过程。有两种形式，臂内倒位和臂间倒位。前者两个断裂点在同一臂（长臂或短臂）内，后者两个断裂点在两臂之间。

3.8 衍生染色体 derivative chromosome

染色体断裂后形成的畸变染色体。分为平衡衍生染色体和不平衡衍生染色体。

3.9 环状染色体 ring chromosome

染色体长、短臂同时发生断裂，含有着丝粒的片段两断端发生重接，形成的环状的染色体。

3.10 GRCh37

是人类基因组的参考序列之一，也称为 Genome Reference Consortium Human Build 37。它是由 Genome Reference Consortium（基因组参考联盟）开发和维护的人类基因组参考序列之一。

3.11 嵌合体 mosaicism

体内同时存在两种或两种以上核型细胞系的个体。可以是染色体数目异常、结构异常及数目和结构异常之间的嵌合。

3.12 插入 insertion

一条染色体发生两次断裂，两个断裂点之间的片段插入到另一条染色体短臂或长臂中间的过程。插入片段可以在新的位置保持原始方向，也可以保持相反的方向。

3.13 纯合区域 regions of homozygosity, ROH

基因组区域中一定范围内连续呈现的杂合性丢失的现象。

说明：对于大部分的二倍体细胞如人类体细胞，拥有两份基因组，一份来自于父亲，另一份来自于母亲，在某一个等位基因位点上，如果来自父本和母本的碱基不同时，则该位点为杂合（heterozygous）。如果因为某种机制（如远亲关系或近亲关系婚姻或基因转换）导致在一定范围内连续的等位基因序列都是纯合子而无杂合子（拷贝数仍为 2 个），则该区域为基因组纯合区域（ROH）。

[T/GDPMAA 0001—2020, 定义 3.4]

3.14 单亲二体 uniparental disomy, UPD

指一个个体的两条同源染色体都来自同一亲本，或来自父母一方的染色体片段被另一方的同源部分取代。UPD 从产生机制分析主要包含单亲异二体（isodisomy, isoUPD）和单亲同二体（heterodisomy, hetUPD）及混合型三种情况。

单亲同二体是指两条同源染色体都来自同一亲体的同一染色体，母源单亲同二体和父源单亲同二体分别指某一染色体的两条同源染色体均来源于母亲或父亲的同一染色体。单亲异二体是指两条同源染色体分别来自同一亲体的两条同源染色体，母源单亲异二体和父源单亲异二体分别指两条同源染色体来源于母亲或父亲的两条同源染色体。

[T/GDPMAA 0001—2020, 定义 3.5, 分类 9.3]

[染色体异常与遗传咨询 第 5 版 第 18 章, 有修改]

3.15 三倍体 triploidy

如果染色体的数目变化是单倍体（ n ）的整倍数，即以 n 为基数，成倍的增加或减少，则称为整倍性改变。在 $2n$ 的基础上，增加一个染色体组（ n ），则染色体为 $3n$ ，即三倍体。

[医学遗传学 第 6 版 第 9 章 第 2 节]

3.16 单倍体 haploidy

如果染色体的数目变化是单倍体（ n ）的整倍数，即以 n 为基数，成倍的增加或减少，则称为整倍性改变。在 $2n$ 的基础上，减少一个染色体组（ n ），即单倍体。

[医学遗传学 第 6 版 第 9 章 第 2 节]

3.17 单体

二倍体中缺少两条同源染色体中一条的细胞或个体。

3.18 三体

二倍体中某一对同源染色体增加了一条的细胞或个体。

3.19 参考序列 reference sequence

测序片段对应的物种基因组序列。

[GB/T 35890-2018, 定义 3.11]

4 结构变异的规范描述

结构变异的规范描述包括检测方法、基因组版本号、变异类型、细胞遗传学定位（染色体编号和细胞遗传学条区带名称、参考序列、核苷酸范围）、嵌合比例等。

各类结构变异类型的规范描述示例如下。

4.1 结果正常

4.1.1 正常男性

seq[GRCh37] (1-22)×2, (X,Y)×1

基于 GRCh37 版本的基因组参考序列，高通量测序检测显示正常男性拥有 2 个拷贝数的 22 条常染色体，1 个拷贝数的 X 染色体和 Y 染色体。

4.1.2 正常女性

seq[GRCh37] (1-22,X)×2

基于 GRCh37 版本的基因组参考序列，高通量测序检测显示正常女性拥有 2 个拷贝数的 22 条常染色体，2 个拷贝数的 X 染色体。

4.1.3 不提示性别的样本

seq[GRCh37] (1-22)×2, (X,N)×1

基于 GRCh37 版本的基因组参考序列，高通量测序检测显示该样本拥有 2 个拷贝数的 22 条常染色体，2 个拷贝数的性染色体：1 条染色体为 X 染色体，另一条染色体未知（X 染色体或 Y 染色体）。

4.2 数目异常

4.2.1 单体

seq[GRCh37] (22)×1

基于 GRCh37 版本的基因组参考序列，高通量测序检测显示 22 号染色体拷贝数减少 1 个。

4.2.2 三体

seq[GRCh37] (22)×3

基于 GRCh37 版本的基因组参考序列，高通量测序检测显示 22 号染色体拷贝数增加 1 个。

4.2.3 嵌合单体

seq[GRCh37] (22)×1[0.51]

基于GRCh37版本的基因组参考序列，高通量测序检测显示51%的细胞为22号染色体拷贝数减少1个。

4.2.4 嵌合三体

seq[GRCh37] (22)×3[0.51]

基于GRCh37版本的基因组参考序列，高通量测序检测显示51%的细胞为22号染色体拷贝数增加1个。

4.2.5 三倍体

seq[GRCh37] (1-22)×3, X×2, Y×1

基于GRCh37版本的基因组参考序列，高通量测序检测显示1-22号常染色体均重复了1个拷贝数，性染色体（X）重复了1个拷贝数，导致多了1套染色体组。

4.2.6 单倍体

seq[GRCh37] (1-22)×1, X×1, Y×0

基于GRCh37版本的基因组参考序列，高通量测序检测显示1-22号常染色体均只有1个拷贝数，性染色体（X）也只有1个拷贝数。

4.3 结构异常

4.3.1 缺失

seq[GRCh37] del(10)(q21.3)

NC_000010.10:g.67900001_68650000del

基于GRCh37版本的基因组NC_000010.10的参考序列，高通量测序检测到10号染色体长臂q21.3的缺失。缺失了第67,900,001位核苷酸至第68,650,000位核苷酸之间的片段，第67,900,000位核苷酸至第68,650,001位核苷酸相连。

4.3.2 重复

seq[GRCh37] dup(10)(q21.3)

NC_000010.10:g.67900001_68650000dup

基于GRCh37版本的基因组NC_000010.10的参考序列，高通量测序检测到10号染色体长臂q21.3的重复。重复了第67,900,001位核苷酸至第68,650,000位核苷酸之间的片段，其方向和原始序列方向相同。

4.3.3 嵌合缺失

seq[GRCh37] del(10)(q21.3)

NC_000010.10:g.67900001_68650000del[0.51]

51%的细胞发生了上述 4.3.1 的变化。

4.3.4 嵌合重复

seq[GRCh37] dup(10)(q21.3)

NC_000010.10:g.67900001_68650000dup[0.51]

51%的细胞发生了上述 4.3.2 的变化。

4.3.5 插入

seq[GRCh37] ins(4;X)(q28.3;q21.31q22.1)

NC_000023.10:g.89555676_100352080del

NC_000004.11:g.134850793_134850794ins[NC_000023.10:g.89555676_100352080]

基于 GRCh37 版本的基因组参考序列，高通量测序检测到染色体间平衡性插入，X 染色体长臂第 89,555,676 位核苷酸到第 100,352,080 位核苷酸之间的 DNA 序列插入到 4 号染色体的长臂第 134,850,793 位核苷酸和第 130,850,794 位核苷酸之间。X 染色体的插入片段的方向与参考序列的方向相同。

4.3.6 易位

seq[GRCh37] ins(4;X)(q28.3;q21.31q22.1)

NC_000023.10:g.89555676_100352080del

NC_000004.11:g.134850793_134850794ins[NC_000023.10:g.89555676_100352080]

基于 GRCh37 版本的基因组参考序列，高通量测序检测到染色体间平衡性插入，X 染色体长臂第 89,555,676 位核苷酸到第 100,352,080 位核苷酸之间的 DNA 序列插入到 4 号染色体的长臂第 134,850,793 位核苷酸和第 130,850,794 位核苷酸之间。X 染色体的插入片段的方向与参考序列的方向相同。

4.3.7 倒位

seq[GRCh37] inv(2)(p22.3q31.1)mat

NC_000002.11:g.[32310435_32310710delinsG;32310711_171827243inv]

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 2 号染色体存在遗传自母亲的臂间倒位（第 32,310,711 位核苷酸到第 171,827,243 位核酸之间的片段），同时伴有 276bp 的缺失（第 32,310,435 核酸到第 32,310,710 位核苷酸之间的片段）和短臂断裂位点上插入一个碱基(G)。

4.3.8 衍生染色体

seq[GRCh37] der(2)t(2;11)(p25.1;p25.2)

NC_000002.11:g.pter_8247756delins[NC_000011.9:g.pter_15825272]

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 2 号和 11 号染色体易位衍生而来的 2 号衍生染色体，断裂位点分别位于 2p25.1（位于第 8,247,756 位核苷酸和第 8,247,757 位核苷酸之间）和 11p15.2（位于第 15,825,272 位核苷酸和第 15,825,273 位核苷酸中间）。

4.3.9 环状染色体

seq[GRCh37] r(8)(p23.2q24.3)

NC_000008.10:g.pter_3300000del::14000000_qterdel

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 8 号染色体在 p23.2 和 q24.3 处发生断裂后，第 3,300,001 位核苷酸与第 139,999,999 位核苷酸连接，形成环状染色体。

4.4 其他特殊情况

4.4.1 整条染色体 ROH

seq[GRCh37] (2)×2 hmz

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 2 号染色体的整条染色体均为纯合状态。

4.4.2 片段型 ROH

seq[GRCh37] 7q11.21q22.1 (64800001-100200000)×2 hmz

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 7 号染色体长臂 q11.21q22.1 区域（第 64,800,001 位核苷酸和第 100,200,000 位核苷酸之间的片段）为纯合状态。

4.4.3 母源单亲同二体

seq[GRCh37] (2)×2 hmz mat

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 2 号染色体整条染色体为纯合状态，且 2 号染色体的 2 条同源染色体均来源于母亲的同一染色体。

4.4.4 父源单亲同二体

seq[GRCh37] 7q11.21q22.1 (64800001-100200000)×2 hmz pat

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 7 号染色体长臂 q11.21q22.1 区域（第 64,800,001 位核苷酸和第 100,200,000 位核苷酸之间的片段）为纯合状态，且均来源于父亲的同一染色体。

4.4.5 母源单亲异二体

seq[GRCh37] (2)×2 htz mat

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 2 号染色体整条染色体为纯合状态，且 2 号染色体的 2 条同源染色体来源于母亲的 2 条同源染色体。

4.4.6 父源单亲异二体

seq[GRCh37] 7q11.21q22.1 (64800001-100200000)×2 htz pat

基于 GRCh37 版本的基因组参考序列，高通量测序检测到 7 号染色体长臂 q11.21q22.1 区域（第 64,800,001 位核苷酸和第 100,200,000 位核苷酸之间的片段）为纯合状态，且来源于父亲的 2 条同源染色体。

4.4.7 大片段拷贝数变异的测序结果描述

如果通过外显子组测序或靶向测序检测到 CNV，无法精确检出位于内含子的断裂位点。以下示例展示了断裂位点不确定时的描述方法。

seq[GRCh37]12q21.32q21.32(88928628×2,88939469_88939743×0,88973993×2)mat pat

NC_000012.11:g[(88928629_88939469)_(88939743_88973992)del];[(88928629_88939469)_(88939743_88973992)del]

通过外显子组数据的 CNV 分析，检测到 12 号染色体上的 *KITLG* 基因发生基因内纯合缺失。由于该方法的技术限制，不能得到精确的断裂位点。已知父母双方都是这种缺失的杂合子携带者。

5 附录

附件 A 缩略词

表 A. 缩略词说明

序号	缩略词	说明
1	seq	高通量测序 (High-throughput sequencing)
2	del	缺失 (Deletion)，表示某一片段的缺失
3	dup	重复 (Duplication)，表示某一片段的重复。
4	ins	插入 (Insertion)，表示某一片段的插入。
5	t	易位 (Translocation)，表示染色体之间的片段交换。
6	rob	罗伯逊易位 (Robertsonian translocation)
7	ter	染色体末端 (Terminal)
8	inv	倒位 (Inversion)，表示某一片段的倒位
9	der	染色体衍生物 (Derivative)，当染色体发生结构变异时，新形成的染色体。

10	r	环状染色体 (Ring chromosome)
11	hmz	纯合性 (Homozygosity)
12	htz	杂合性 (Heterozygosity)
13	mat	母系来源 (Maternal)
14	pat	父系来源 (Paternal)

附件 B GRCh37 染色体长度和参考序列

表 B. GRCh37 染色体长度和参考序列

Chromosome	参考序列	Size (bp)
1	NC_000001.10	249,250,621
2	NC_000002.11	243,199,373
3	NC_000003.11	198,022,430
4	NC_000004.11	191,154,276
5	NC_000005.9	180,915,260
6	NC_000006.11	171,115,067
7	NC_000007.13	159,138,663
8	NC_000008.10	146,364,022
9	NC_000009.11	141,213,431
10	NC_000010.10	135,534,747
11	NC_000011.9	135,006,516
12	NC_000012.11	133,851,895
13	NC_000013.10	115,169,878
14	NC_000014.8	107,349,540
15	NC_000015.9	102,531,392
16	NC_000016.9	90,354,753
17	NC_000017.10	81,195,210
18	NC_000018.9	78,077,248
19	NC_000019.9	59,128,983
20	NC_000020.10	63,025,520
21	NC_000021.8	48,129,895
22	NC_000022.10	51,304,566
X	NC_000023.10	155,270,560

Y	NC_000024.9	59,373,566
---	-------------	------------

参 考 文 献

- [1] McGowan-Jordan J, Hastings RJ, Moore S. An International System for Human Cytogenomic Nomenclature (2020)[J].Cytogenetic and genome research, 2020.
- [2] Newman S , Hermetz K , Weckselblatt B ,et al. Next-generation sequencing of duplication CNVs reveals that most are tandem and some create fusion genes at breakpoints.[J].American Journal of Human Genetics, 2015, 96(2):208-220.
- [3] Ordulu Z, Wong K, Currall B, et al. Describing Sequencing Results of Structural Chromosome Rearrangements with a Suggested Next-Generation Cytogenetic Nomenclature[J].The American Journal of Human Genetics, 2014, 94(5):695-709.
- [4] Schluth-Bolard,Caroline,Labalme,et al.Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations[J]. 2013, 50:144-150.
- [5] GB/T 30989-2014 高通量基因测序技术规程
- [6] T/GDPMAA 0001—2020 产前遗传学诊断拷贝数变异和纯合区域数据分析解读及报告规范
- [7] 全国科学技术名词审定委员会审定.血液学名词[M].北京：科学出版社.2022.
- [8] 全国科学技术名词审定委员会审定.生殖医学名词[M].北京：科学出版社.2022.
- [9] 全国科学技术名词审定委员会审定.医学遗传学名词[M].北京：科学出版社.2021