

团 体 标 准

T/SZAS 47—2022

地中海贫血基因检测数据集

Data set of genetic test for thalassemia

2022 - 07 - 01 发布

2022 - 07 - 10 实施

深圳市标准化协会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 数据元目录	2
6 数据格式标准	3
7 数据归档目录	3
附录 A（资料性） 数据元目录.....	5
附录 B（资料性） 数据元值域代码表.....	8

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。

本文件由深圳华大医学检验实验室提出。

本文件由深圳市标准化协会归口。

本文件起草单位：深圳华大医学检验实验室、武汉华大医学检验所有限公司、深圳华大智造科技股份有限公司、深圳华大基因股份有限公司、云南省第一人民医院、郴州市第一人民医院。

本文件主要起草人：吕春杰、唐美芳、刘小燕、李翠翠、吴平、葛建敬、黄晓燕、陈仕平、张红云、单日强、李倩一、朱宝生、贺静、雷冬竹、张昊晴、吴昊、姜华艳。

地中海贫血基因检测数据集

1 范围

本文件规定了地中海贫血基因检测数据的范围以及数据元的规范化定义，数据集包括地中海贫血基因检测相关数据元和值域。

本文件适用于地中海贫血基因检测数据信息的存储、治理、交换与共享。

2 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，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 2261.1 个人基本信息分类与代码 第1部分：人的性别代码

GB/T 2261.2 个人基本信息分类与代码 第2部分：婚姻状况代码

GB/T 3304 中国各民族名称的罗马字母拼写法和代码

GB/T 34265—2017 Sanger法测序技术指南

GB/T 35890—2018 高通量测序数据序列格式规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

高通量测序 high-throughput sequencing

以一次并行对几十万到几百万条DNA分子进行序列测定和一般读长较短等为标志，适用于DNA的测序技术。

[来源：GB/T 35890—2018，3.1]

3.2

FASTQ 格式 FASTQ format

基于文本的，保存生物序列（通常是核酸序列）和其测序质量信息的，每四行表示一条序列的标准格式。

[来源：GB/T 35890—2018，3.9]

3.3

Sanger 测序 method of Sanger sequencing

用于基因测序的双脱氧链末端终止法，在链延伸过程中利用荧光标记双脱氧碱基随机阻断，产生以A、T、C、G结束的四组不同长度的核苷酸链，通过读取荧光信号实现对核酸碱基序列信息的读取。

[来源：GB/T 34265—2017，3.2]

3.4

AB1 格式 AB1 format

DNA电泳图文件格式，电泳图是电泳自动测序分析的结果图。

3.5

聚合酶链式反应 polymerase chain reaction, PCR

体外酶促合成特异DNA片段的一种方法，由高温变性、低温退火（复性）及适温延伸等几步反应组成一个周期，循环进行，使目的DNA得以迅速扩增。

3.6

实时荧光定量 PCR quantitative real-time PCR

一种在DNA扩增反应中，以荧光化学物质测每次PCR循环后产物总量的方法。

3.7

跨越断裂位点 PCR gap-PCR

以PCR技术为基础，设计出与缺失序列的两侧翼序列互补的引物，由于缺失两端连接，使本来在正常DNA序列中相距很远的这一对引物之间的距离，因断端连接而靠近，以至于能够扩增出特定长度的片段。

3.8

多重 PCR multiplex PCR

在同一PCR反应体系里加上两对以上引物，同时扩增出多个核酸片段的PCR反应，又称多重引物PCR或复合PCR。

3.9

多重连接探针扩增技术 multiplex ligation-dependent probe amplification, MLPA

以探针的扩增为基础的一种高通量、针对待测核酸中靶序列进行定性和定量分析的技术。

3.10

 α -地中海贫血 α -thalassemia

由于 α -珠蛋白基因的缺失或功能缺陷（点突变）而导致 α -珠蛋白链合成障碍所引起的一组溶血性贫血。

3.11

 β -地中海贫血 β -thalassemia

由于 β -珠蛋白基因的缺失或功能缺陷（点突变）而导致 β -珠蛋白链合成障碍所引起的一组溶血性贫血。

4 缩略语

下列缩略语适用于本文件。

D: 日期型 (date)

DNA: 脱氧核糖核酸 (deoxyribonucleic acid)

HbA: 成人血红蛋白 (adult hemoglobin)

HbA2: 血红蛋白A2 (hemoglobin A2)

HbF: 胎儿血红蛋白 (fetal hemoglobin)

HGB: 血红蛋白 (hemoglobin)

L: 布尔型 (boolean)

MCV: 平均红细胞体积 (mean cell volume)

MCH: 平均红细胞血红蛋白量 (mean cell hemoglobin)

N: 数值型 (number)

S: 字符串型 (string)

T: 时间型 (time)

5 数据元目录

5.1 数据元目录公用属性

数据元目录公用属性如表1所示。

表1 数据元目录公用属性

属性名称	描述
版本	V1.0
注册机构	注册机构名称
相关环境	生物信息、生物大数据
分类模式	分类法
主管机构	主管机构名称
注册状态	标准状态

表1 数据元目录公用属性（续）

属性名称	描述
提交机构	提交机构名称

5.2 数据元目录专用属性

5.2.1 地中海贫血基因检测数据元目录专用属性包括个人信息元数据目录、检测过程信息元数据目录、表型信息元数据目录、送检信息元数据目录、检查结果信息元数据目录五部分。具体每个数据元的标识符、名称、定义、信息保护、单位、数据类型见附录 A。数据元允许值见附录 B。

5.2.2 个人信息元数据目录描述见表 A.1。

5.2.3 检测过程信息元数据目录描述见表 A.2。

5.2.4 表型信息元数据目录描述见表 A.3。

5.2.5 送检信息元数据目录描述见表 A.4。

5.2.6 检查结果信息元数据目录描述见表 A.5。

6 数据格式标准

6.1 原始数据格式

原始数据格式如表2所示。

表2 原始数据格式描述

数据格式标识符	数据格式名称	适用范围	数据格式允许值
DF01.01.001	FASTQ	地中海贫血高通量检测的原始数据信息。	SF01.01.001 FASTQ
DF01.01.002	ABI	地中海贫血基因检测二次验证的原始数据信息。	SF01.01.002 ABI
DF01.01.003	结果图片	地中海贫血基因检测的结果图片。	SF01.01.003 结果图片
DF01.01.004	下机原始文件	具体格式以仪器自动生成的下机文件为准。	—

6.2 原始数据格式代码表

原始数据格式代码如表3所示。

表3 原始数据格式代码描述

数据格式说明标识符	数据格式名称	规格说明
SF01.01.001	FASTQ	GB/T 35890—2018 高通量测序数据序列格式规范。
SF01.01.002	ABI	ABI 文件是一个 DNA 电泳图文件，电泳图是电泳自动测序分析的结果图。
SF01.01.003	结果图片	结果图片是检测中所产生的原始结果文件。
SF01.01.004	下机原始文件	下机原始文件，以仪器自动生成的下机文件格式为准。

7 数据归档目录

7.1 数据归档目录结构

数据归档目录结构如表4所示。

表4 数据归档目录结构

第一级	第二级	第三级
Item_id	Raw_data	Rawdata_id
Item_id	Result_data	

注：每级目录情况说明见7.2。第一级目录和第二级目录为强制性目录，其他级目录为推荐性目录。

7.2 数据归档目录详情

7.2.1 概述

7.2是对7.1归档目录结构具体每级目录的情况说明。共3级目录，每一级目录均包含目录标识符、目录名称、目录定义及父目录。

7.2.2 数据归档第一级目录

目录标识符：DI01.01.01

目录名称：Item_id

定义：该目录用于存档地中海贫血基因检测相关数据。

7.2.3 数据归档第二级目录

目录标识符：DI01.02.01

目录名称：Raw_data

定义：该目录用于存档地中海贫血基因检测过程中检测仪器生成的原始数据。

父目录：DI01.01.01

目录标识符：DI01.02.02

目录名称：Result_data

定义：该目录用于存档地中海贫血基因检测的结果。

父目录：DI01.01.01

7.2.4 数据归档第三级目录

目录标识符：DI01.03.01

目录名称：Rawdata_id

定义：存放原始数据的指定文件。

父目录：DI01.02.01、DI01.02.02

附录 A
(资料性)
数据元目录

A.1 个人信息元数据目录

个人信息如表A.1所示。

表A.1 个人信息

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE01.01.001.00	姓名	受检者在公安管理部门正式登记注册的姓氏和名称	保护		S	
DE01.01.002.00	性别	受检者生理性别在特定编码体系中的代码	不保护		S	GB/T 2261.1 个人基本信息分类与代码 第1部分:人的性别代码
DE01.01.003.00	年龄	受检者从出生当日公元纪年日起到计算当日止生存的时间长度,计量单位为岁	不保护	岁	N	
DE01.01.004.00	民族	受检者所属民族类别在特定编码体系中的代码	不保护		S	GB/T 3304 中国各民族名称的罗马字母拼写法和代码
DE01.01.005.00	手机号码	受检者的移动电话号码	保护		S	
DE01.01.006.00	身份证件号码	受检者的身份证件上的唯一法定标识符	保护		S	
DE01.01.007.00	婚姻状况代码	受检者当前婚姻状况代码	不保护		S	GB/T 2261.2 个人基本信息分类与代码 第2部分:婚姻状况代码
DE01.01.008.00	籍贯	为受检者出生时祖父的居住地(户口所在地);祖父去世的,填写祖父去世时的户口所在地;祖父未落常住户口的,填写祖父应落常住户口的地方。	不保护		S	

A.2 检测过程信息元数据目录

检测过程信息如表A.2所示。

表A.2 检测过程信息

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE02.01.001.00	检测名称	检测项目或产品名称	不保护		S	
DE02.01.002.00	检测技术	检测技术代码	不保护		S	B.1 检测技术代码表
DE02.01.003.00	检测来源	检测来源代码	不保护		S	B.2 检测来源代码表
DE02.01.004.00	样本类型	样本类型代码	不保护		S	B.3 样本类型代码表
DE02.01.005.00	费用类型	进行检测的费用类型在特定编码体系中的代码	不保护		S	B.4 费用类型代码表
DE02.01.006.00	样本状态	样本的状态在特定编码体系中的代码	不保护		S	B.5 样本状态代码表
DE02.01.007.00	样本检测状态	样本的检测状态在特定编码体系中的代码	不保护		S	B.6 样本检测状态代码表
DE02.01.008.00	样本存储条件	样本存储的条件	不保护		S	
DE02.01.009.00	样本外观	样品外观是否完整有效	不保护		S	

表A.2 检测过程信息（续）

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE02.01.010.00	样本审核时间	样本审核当日时间	不保护		D	
DE02.01.011.00	样本审核代码	样本是否审核的代码	不保护		S	B.7 审核代码表

A.3 表型信息元数据目录

表型信息如表A.3所示。

表A.3 表型信息

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE03.01.001.00	一年内接受过异体输血标志	受检者有无一年内接受过异体输血	不保护		L	
DE03.01.002.00	四周内接受过异体细胞免疫治疗标志	受检者有无四周内接受过异体细胞免疫治疗	不保护		L	
DE03.01.003.00	移植手术标志	受检者有无移植手术	不保护		L	
DE03.01.004.00	干细胞治疗标志	受检者有无干细胞治疗	不保护		L	
DE03.01.005.00	临床检查 HGB（血红蛋白）值	临床检查 HGB 值	不保护	g/L	N	
DE03.01.006.00	临床检查 MCV（平均红细胞体积）值	临床检查 MCV 值	不保护	fL	N	
DE03.01.007.00	临床检查 MCH（平均红细胞血红蛋白量）值	临床检查 MCH 值	不保护	pg	N	
DE03.01.008.00	临床检查 HbA（成人血红蛋白）值	临床检查 HbA 值	不保护	%	N	
DE03.01.009.00	临床检查 HbA2（血红蛋白A2）值	临床检查 HbA2 值	不保护	%	N	
DE03.01.010.00	临床检查 HbF（胎儿血红蛋白）值	临床检查 HbF 值	不保护	%	N	
DE03.01.011.00	临床检查其他结果	临床检查其他结果	不保护		S	
DE03.01.012.00	既往病史	对个体既往健康状况和疾病的详细描述	不保护		S	WS 364.4-2011 表 21 CV02.10.005 既往患病种类代码表
DE03.01.013.00	现病史	对患者当前所患疾病情况的详细描述	不保护		S	
DE03.01.014.00	临床症状描述	临床症状的详细描述	不保护		S	

A.4 送检信息元数据目录

送检信息如表A.4所示。

表A.4 送检信息

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE04.01.001.00	受检者编号	受检者编号	不保护		S	
DE04.01.002.00	采样日期	采集样本的日期	不保护		D	
DE04.01.003.00	签署知情同意书	阅读并充分理解本知情同意书内容，受检者/监护人完成签名	不保护		S	
DE04.01.004.00	送检医院名称	受检者送检所在医院的名称	不保护		S	
DE04.01.005.00	送检医生姓名	送检医生在公安管理部门正式登记注册的姓氏和名称	保护		S	
DE04.01.006.00	科室名称	标识受检者在医院就诊的科室名称	不保护		S	
DE04.01.007.00	住院号	按照某一特定编码规则赋予住院就诊对象的顺序号	不保护		S	

表A.4 送检信息 (续)

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE04.01.008.00	送检市(地区、州)	送检市、地区或州的名称	不保护		S	
DE04.01.009.00	送检县(区)	送检县或区名称	不保护		S	

A.5 检查结果信息元数据目录

检查结果信息如表A.5所示。

表A.5 检查结果信息

标识符	名称	定义	信息保护	单位	数据类型	数据元允许值
DE05.01.001.00	α 地中海贫血检测结果	α 地中海贫血检测结果,用突变基因格式描述	不保护		S	
DE05.01.002.00	β 地中海贫血检测结果	β 地中海贫血检测结果,用突变基因格式描述	不保护		S	
DE05.01.003.00	α 地中海贫血检测结果说明	α 地中海贫血检测结果说明名称	不保护		S	B.8 α 地中海贫血检测结果说明代码表
DE05.01.004.00	β 地中海贫血检测结果说明	β 地中海贫血检测结果说明名称	不保护		S	B.9 β 地中海贫血检测结果说明代码表
DE05.01.005.00	其他异常血红蛋白描述	对检出异常血红蛋白的描述	不保护		S	
DE05.01.006.00	新发突变描述	有无其他新发突变表达	不保护		S	
DE05.01.007.00	检测者名字	检测实验人员签名	不保护		S	
DE05.01.008.00	审核者名字	具备实验师和检验师资质的审核人员签名	不保护		S	
DE05.01.009.00	报告时间	检测结果出报告时间	不保护		D	

附录 B
(资料性)
数据元值域代码表

B.1 检测技术代码

检测技术代码规定了检测技术名称的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.1。

表B.1 检测技术代码表

值	值含义
1	聚合酶链式反应和高通量测序技术 (PCR-NGS)
2	Sanger 测序
3	实时荧光定量 PCR (qPCR)
4	多重 PCR 和反向点杂交 (PCR-RDB)
5	跨越断裂位点 PCR (Gap-PCR)
6	多重连接探针扩增法 (MLPA)
7	其他

B.2 检测来源代码

检测来源代码规定了检测来源的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.2。

表B.2 检测来源代码表

值	值含义
1	临床
2	科研

B.3 样本类型代码

样本类型代码规定了样本类型的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.3。

表B.3 样本类型代码表

值	值含义
1	外周血
2	干血片
3	DNA
4	唾液
5	其他

B.4 费用类型代码

费用类型代码规定了费用类型的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.4。

表B.4 费用类型代码表

值	值含义
1	收费
2	免费
3	退费

B.5 样本状态代码

样本状态代码规定了样本状态的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.5。

表B.5 样本状态代码表

值	值含义
1	未到样
2	已到样可出库
3	出库锁定
4	已出库
5	归还锁定
6	销毁
7	返还
8	暂存

B.6 样本检测状态代码

样本检测状态代码规定了样本检测状态的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.6。

表B.6 样本检测状态代码表

值	值含义
1	未出库
2	已出库
3	待出报告
4	已出报告
5	暂缓生产

B.7 审核代码

审核代码规定了审核的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.7。

表B.7 审核代码表

值	值含义
1	未审核
2	已审核
3	存疑

B.8 α 地中海贫血检测结果说明代码

α 地中海贫血检测结果说明代码规定了 α 地中海贫血检测结果说明的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.8。

表B.8 α 地中海贫血检测结果说明代码表

值	值含义
1	静止型携带者
2	标准型携带者
3	HBH病
4	Hb Bart's 胎儿水肿综合征

B.9 β 地中海贫血检测结果说明代码

β 地中海贫血检测结果说明代码规定了 β 地中海贫血检测结果说明的代码。

采用1位数字顺序代码，从“1”开始编码，按升序排列。见表B.9。

表B.9 β 地中海贫血检测结果说明代码表

值	值含义
1	携带者
2	双重杂合子
3	纯合子
