



# 中国不同民族永生细胞库的建立和应用

杨昭庆, 黄小琴, 孙浩, 边成, 林克勤, 褚嘉祐\*

中国医学科学院&北京协和医学院医学生物学研究所, 昆明 650118

\* 联系人, E-mail: [chujy@imbcams.com.cn](mailto:chujy@imbcams.com.cn)

收稿日期: 2023-09-10; 接受日期: 2023-12-04; 网络版发表日期: 2024-06-13

云南省重点研发计划(批准号: 202103AF140002)和云南省高层次卫生健康技术人才项目(批准号: L-2018003)

**摘要** 人类遗传资源具有重要的生命科学和医学研究及应用价值, 国际不同民族群体遗传多样性研究具有重要意义, 中国不同民族在民族渊源、语言、地理环境、生活习俗乃至生理表型和疾病易感性等方面存在差异, 为医学与健康研究提供丰富的遗传资源材料. EB病毒(Epstein-Barr Virus, EBV)转化的永生B淋巴细胞是保存人类遗传资源的一种主要方式“中国不同民族永生细胞库”历时二十多年, 已经建成包括49个法定民族, 90多个群体, 6126人份的永生细胞株的样本库. 作为重要的国家战略资源, 中国不同民族永生细胞库及其来源的核酸为中国人群遗传结构特征、基因变异、表达和功能等研究和应用提供研究材料, 展现在药物发现、疫苗、抗体和干细胞等领域的应用潜力, 显著促进中国在人类遗传资源的国际交流与合作, 并伴随着高通量测序和基因编辑等先进技术的发展, 将为人类生命科学研究和临床转化应用提供更具广阔前景的工具和资源.

**关键词** 人类遗传资源, 中国不同民族, 遗传多样性, 永生细胞库, 疾病基因, 疫苗与药物

遗传多态性是指生物基因组长期进化过程中积累起来的遗传变异. 这些变异可以是有益, 中性或有害, 其中一些被保存下来, 导致不同种族、群体和个体间基因组的差异, 这就是人类遗传多样性, 不同民族群体遗传多样性研究具有重要意义. 人类是一个具有多样性的群体, 人类基因组具有高度的变异性, 基因组多样性的研究是疾病基因组的重要组成部分, 对阐明不同人群或个体在疾病的易感性和抗性方面的差异具有重要意义. 研究、比较不同人种、人群的基因组, 这就是“人类基因组多样性计划(Human Genome Diversity Project, HGDP)”的主要内容.

中国历史悠久、地域辽阔, 是公认的人类遗传资

源大国, 有14亿人口, 56个民族, 汉族占总人口91.5%, 55种少数民族占人口8.5%. 这些民族在民族渊源、聚居地、语言、文化风俗等方面各有其特点. 中国不同民族群体大多具有百年甚至千年历史的聚居地区, 其中一些民族群体由于一些特殊的地理、历史原因, 形成多年的饮食习惯和生活习俗, 并多在群体内通婚, 从而形成许多独特的遗传隔离人群, 为进行隔离人群遗传研究提供丰富的资源. 由于近年来交通的改善和社会进步, 以及不断增加的与外界通婚及迁移等原因, 使不同民族群体独特的基因组特征面临消失的危险, 对这些宝贵遗传资源的保存成为一项抢救性的工作, 如果研究者们现在不做, 将永远失去保留完整的中国

引用格式: 杨昭庆, 黄小琴, 孙浩, 等. 中国不同民族永生细胞库的建立和应用. 中国科学: 生命科学, 2024, 54: 987-1001  
Yang Z Q, Huang X Q, Sun H, et al. The establishment and applications of the Immortal Cell Banks of Chinese ethnic groups (in Chinese). *Sci Sin Vitae*, 2024, 54: 987-1001, doi: [10.1360/SSV-2023-0035](https://doi.org/10.1360/SSV-2023-0035)

不同民族群体独特基因组多样性的机会。我国在人类遗传资源保存和利用方面已开展的工作与国际上的差距不大,是中国基因组研究中最有特色的一块,也是最受国际关注的一块,是我国参与乃至逐渐主导国际人类遗传资源和基因组学等国际交流与合作的基础。

## 1 中国不同民族永生细胞库的建设

“中国不同民族永生细胞库”是主要由国家自然科学基金重大项目和科学技术部支持的中国人类基因组项目的组成部分,从1994年启动的国家自然科学基金重大项目“中华民族基因组中若干位点基因结构的研究”第一子课题“中华多民族基因组的保存”到于1999年2月正式启动的国家自然科学基金重大项目“民族基因组的构和功能研究”第一分题“中华民族基因组的保存与多样性的研究”的完成,“中国不同民族基因组的保存”一直作为国家支持的重大项目。承担项目的三个单位是中国医学科学院&北京协和医学院医学生物研究所(以下简称“昆明所”)、中国科学院遗传与发育研究所和哈尔滨医科大学,项目负责人为昆明所褚嘉祐教授。

从1994年开始,在二十多年的时间里,以“中国不同民族基因组的保存”开始的10年获得国家启动经费,以后三个研究单位科技工作者继续艰苦努力,通过不同途径的科研经费获得对项目的支持。中国不同民族永生细胞库的建立和研究遵循“知情同意”的原则,依据各民族群体及其支系的历史源流和聚居地,按照严格的民族选样标准。目前所构建的“中国不同民族永生细胞库”已经建成包括49个中国法定民族,90多个民族分支群体,6126人份永生细胞株的样本库。为了更加安全、完整地保存中国各民族的基因组资源,避免因某个单位的永生细胞库发生意外情况而遭受损失,项目组进行三个单位的细胞库的部分细胞株交换和异地保存,成功地建立细胞株交换保存的工作程序<sup>[1]</sup>。昆明所永生细胞库,保存其中绝大多数民族群体的永生细胞株<sup>[2,3]</sup>。

在建立的民族群体细胞库和DNA库的基础上,项目组进行广泛、深入的遗传多样性研究,对部分民族群体进行一些民族源流及相互关系、中国不同民族之间的遗传关系和遗传多样性特征、以及疾病易感基因遗传多态性的研究,其中包括中国和亚洲民族。2002年, *Science* 率先报道包括中国不同民族永生细胞库在

内的全世界不同人群的类淋巴母细胞系(lymphoblastoid cell lines, LCLs)样本库建立工作<sup>[4]</sup>; 在此之后,基于中国不同民族永生细胞库也开展研究,在国际知名学术刊物发表一系列具有代表性的论文,例如1999年, *American Journal of Human Genetics* 上报道针对现代人到达东亚的时间和性质以及随后向东亚扩张的模式<sup>[5]</sup>; 1999年在 *Human Genetics* 报道利用Y染色体的遗传标记对三个不同支系的藏族群体进行遗传源流分析<sup>[6]</sup>; 2000年 *PNAS* 报道依据Y染色体对中国台湾及东南亚居民起源及迁徙的研究<sup>[7]</sup>; 2022年在 *BMC Biology* 上发表利用不同民族全外显子组测序数据,系统分析云南多个少数民族的遗传多样性,为三谱系起源假说提供新的证据<sup>[8]</sup>; 2023年 *Nature* 报道基于不包含36个中国少数民族的58个核心样本的人类泛基因组图谱,充分展示中国民族永生细胞库在复杂疾病定位和揭示人类进化的新线索中的巨大的应用潜力<sup>[9]</sup>。基于中国不同民族永生细胞库开展的研究已在国际和国内期刊累计发表200多篇论文,并出版3本专著<sup>[10]</sup>。相关工作还获得2005年和2007年国家自然科学二等奖,以及多个省部级科技奖励。2023年中国不同民族永生细胞库得到《中国人类遗传资源保藏审批决定书》,细胞库的采样、保藏和应用研究得以进一步标准化、规范化和顺利进行(表1)。

### 1.1 知情同意和采样

在项目实施过程中,为建立中国人永生细胞株开展的样本采集程序均遵守国家关于采集人类遗传资源的规定,在采样前得到相关部门伦理委员会的批准以及国家人类遗传资源管理部门的批准。与志愿者签署《知情同意书》时,一般均有同时精通本民族语言和汉语的医疗机构人员在场协助以保障完整准确的语言交流沟通及知情同意。

从早期的《人类遗传资源管理暂行办法》(国务院办公厅,1998年6月)到《中华人民共和国人类遗传资源管理条例》(国务院,2019年7月)及《人类遗传资源管理条例实施细则》(2023年7月),项目组不断遵照相应的要求,更新知情同意书等申报和审批材料,完善永生细胞库的采集和保藏要求,中国医学科学院的中国不同民族永生细胞库按最新要求于2023年6月获得细胞库保藏和关联性采集所需的中国人类遗传资源保

表1 部分已建立的中国不同民族永生细胞和采样地点(截止到2022年4月)

Table 1 Established immortal cell lines and sampling location of different Chinese ethnic groups (April 2022)

序号	民族群体	三机构细胞株数 <sup>a)</sup>	昆明所保存细胞株数	保存地点 <sup>b)</sup>	采样地点
1	阿昌族1	93	93	昆明	云南省芒市
2	阿昌族2	75	75	昆明	云南省梁河县
3	白族1	33	33	昆明	云南省洱源县
4	白族2	62	62	昆明	云南省大理市
5	布朗族1	54	54	昆明	云南省勐海县
6	布朗族2	136	136	昆明	云南省勐海县
7	布朗族(克木人)	148	148	昆明	云南省勐腊县
8	布依族1	64	49	哈尔滨、北京、昆明	贵州省黔南布依族自治州
9	布依族2	258	258	昆明	贵州省册亨县
10	朝鲜族	58	55	哈尔滨、昆明	吉林省延边朝鲜族自治州
11	达斡尔族	56	40	哈尔滨、昆明	内蒙古自治区呼伦贝尔市
12	傣族1	39	39	昆明	云南省勐海县
13	傣族2	17	17	昆明	云南省景洪市
14	傣族3	74	74	昆明	云南省芒市
15	傣族4	55	55	昆明	云南省瑞丽市
16	傣族(花腰傣)	101	101	昆明	云南省新平县
17	德昂族	122	122	昆明	云南省芒市
18	东乡族	133	133	昆明	甘肃省榆中县
19	东乡族	65	50	北京、昆明	甘肃省东乡县
20	侗族	50	49	北京、昆明	广西壮族自治区三江县
22	独龙族	72	72	昆明	云南省贡山县
23	鄂伦春族	38	38	哈尔滨、昆明	内蒙古自治区阿里河自治县
24	鄂温克族	43	41	哈尔滨、昆明	黑龙江省讷河市
25	仡佬族	115	115	昆明	贵州省务川县
26	哈尼族	63	0	北京	云南省墨江县
27	哈尼族(爱伲人)	75	75	昆明	云南省勐海县
28	哈萨克族	64	52	哈尔滨、昆明	新疆伊宁哈萨克族自治州
29	汉族(福建)	58	0	哈尔滨	福建省闽清县
30	汉族(甘肃)	62	50	北京、昆明	甘肃省武威市
31	汉族(广东)	43	43	昆明	广东省韶关市
32	汉族(广东)	57	0	北京	广东省广宁县
33	汉族(河南)	60	0	北京	河南省方城县
34	汉族(湖南)	74	74	昆明	湖南省临武县
35	汉族(湖南)	59	0	北京	湖南省双峰县
36	汉族(山东)	53	53	昆明	山东省周平县
37	汉族(陕西)	58	0	北京	陕西省麟游县
38	汉族(四川)	56	51	哈尔滨、昆明	四川省乐玉县
39	汉族(浙江)	59	0	北京	浙江省缙云县
40	赫哲族	18	17	哈尔滨、昆明	黑龙江省同江市
41	回族(宁夏)	63	0	哈尔滨	宁夏回族自治区同心县

(表1续1)

序号	民族群体	三机构细胞株数 <sup>a)</sup>	昆明所保存细胞株数	保存地点 <sup>b)</sup>	采样地点
42	回族(宁夏)	73	73	昆明	宁夏回族自治区吴忠市
43	回族(云南)	46	46	昆明	云南省通海县
44	基诺族	74	74	昆明、北京、哈尔滨	云南省景洪市
45	京族	94	94	昆明	广西壮族自治区东兴市
46	景颇族1	33	33	昆明	云南省芒市
47	景颇族2	139	139	昆明	云南省芒市
48	柯尔克孜族	57	50	北京、哈尔滨、昆明	新疆维吾尔自治区乌恰县
49	拉祜族1	46	46	昆明	云南省勐海县
50	拉祜族2	44	44	昆明	云南省勐海县
51	黎族	57	57	昆明	海南省白沙县
52	傈僳族	21	21	昆明	云南省福贡县
53	傈僳族	18	18	昆明	云南省贡山县
54	傈僳族	68	68	昆明	云南省泸水市
55	满族	65	50	哈尔滨、北京、昆明	辽宁省岫岩县
56	毛南族	21	21	昆明	广西壮族自治区环江县
57	蒙古族1	49	49	昆明	内蒙古自治区包头市
58	蒙古族2	54	54	昆明	内蒙古自治区东胜区
59	蒙古族3	61	0	哈尔滨	内蒙古自治区海拉尔区
60	蒙古族(云南)	109	109	昆明	云南省通海县
61	苗族(贵州)	43	43	北京、昆明	贵州省松桃县
62	仡佬族	63	63	昆明	广西壮族自治区罗城县
63	纳西族	37	37	昆明	云南省丽江市
64	纳西族(摩梭人)	45	45	昆明	云南省宁蒗县
65	纳西族	146	146	昆明	云南省丽江市
66	怒族1	65	65	昆明	云南省贡山县
67	怒族2	22	22	昆明	云南省福贡县
68	普米族1	82	82	昆明	云南省兰坪县
69	普米族2	98	98	昆明	云南省丽江市
70	羌族	48	48	昆明、北京	四川省茂县
71	撒拉族	71	71	昆明	青海省循化县
72	畲族	46	46	昆明	福建省福安市
73	水族	138	138	昆明	贵州省三都县
74	塔吉克族	63	50	北京、昆明	新疆维吾尔自治区塔什库尔干县
75	土家族	56	52	北京、昆明	湖北省来凤县
76	土族1	39	39	昆明	青海省互助县
77	土族2	60	60	昆明	青海省互助县
78	佤族1	57	57	昆明	云南省沧源县
79	佤族2	69	69	昆明	云南省西盟县
80	维吾尔族	60	50	北京、昆明、哈尔滨	新疆维吾尔自治区伊宁市
81	锡伯族	43	39	哈尔滨、昆明	新疆维吾尔自治区查布察尔县
82	瑶族(广西)	58	0	北京	广西壮族自治区都安县
83	瑶族(云南)	41	41	昆明	云南省勐腊县

(表1续2)

序号	民族群体	三机构细胞株数 <sup>a)</sup>	昆明所保存细胞株数	保存地点 <sup>b)</sup>	采样地点
84	彝族(四川)	49	48	北京、昆明	四川省布拖县
85	彝族(宁蒗)	47	47	昆明	云南省宁蒗县
86	彝族(云南石林)	116	116	昆明	云南省石林县
87	彝族(云南丽江)	128	128	昆明	云南省丽江市
88	裕固族1	55	50	北京、昆明	甘肃省肃南县
89	裕固族2	35	35	昆明	甘肃省肃南县
90	藏族(青海)	46	46	昆明	青海省贵南县
91	藏族(西藏)	62	62	昆明	西藏自治区拉萨市
92	藏族(云南中甸)	23	23	昆明	云南省中甸县
93	藏族(云南贡山)	48	48	昆明	云南省贡山县
94	壮族	55	55	昆明	广西壮族自治区百色市
	合计	6126	5389		

a) 昆明所、中国科学院遗传与发育生物学研究所、哈尔滨医科大学三个共同承担项目单位所保藏的永生细胞株合计; b) 有多个保存地点表示该民族群体细胞株存在异地交换和共同保存。

## 藏审批决定书。

对于中国不同民族, 遵循着行政区域的划分及相关文献资料记载, 并咨询民族学家、语言学家和历史学家意见后选择民族聚居地。寻找三代以上(含三代)都是该民族的志愿者, 这些志愿者应无明显遗传性疾病、传染性疾病及心血管等疾病。志愿者之间应无亲缘关系。

汉族群体选择陕西省、广东省、湖南省、浙江省和福建省等地的汉族。在少数民族选择中包括中国法定的56个民族, 并且考虑一些重要的支系。涉及到主要分布在西南、西北、中南地区的藏缅语系, 主要分布在广西、云南、贵州、广东、海南和湖南南部的汉藏语系下的侗台语族, 主要分布于贵州、广西、湖南、云南、四川、广东、海南、湖北、江西等省的苗瑶语系, 主要分布于云南等地的孟高棉语族中佤德昂语支: 佤语、德昂语、布朗语, 以及分布于新疆、甘肃、青海、内蒙古中的阿尔泰语系等。

每一样本有唯一编号和基本信息, 采样日期、采样地信息, 并录入电子管理系统。样本个人基本信息包括: 研究对象序号、姓名、性别、出生年月、出生地点、民族、居住地、第一语言及其他能说或能听懂的语言、研究对象家族其他成员信息等。其他特殊信息如具有的特殊表现型、家族是否有相关的疾病等。

建立中华民族永生细胞库样本的采集对象是中国56个民族的健康个体。细胞库建立早期采样数量较少, 近年来按计划, 每一民族采集100人左右。原则上男女各半, 考虑到男性同时具有X和Y染色体, 男性样本可偏多些。理论上讲, 应不限志愿者的年龄。但出于对样本量的要求和伦理等因素考虑, 本研究的志愿者年龄多为18~55岁。样本采集由具有从业资格的医护人员进行, 采集志愿者外周静脉血, 用EB病毒转化外周血B淋巴细胞。永生细胞株转化成功后, 置于液氮冰箱中保存。

## 1.2 永生细胞株的建立

永生细胞株的建立步骤主要包括淋巴细胞的转化、冻存、复苏, 中国不同民族永生细胞库中对此建立较为完整的技术体系<sup>[10]</sup>。

永生细胞株指的是原代培养物培养成功后, 可继续传代下去, 并且通过传代可无限增殖持续生存的细胞系。人B淋巴细胞是建立永生性细胞系的极好对象, 主要是因为B淋巴细胞容易获得, 转化程序简单, 转化效率高。目前国内外最常用的方法是使用EB病毒(Epstein-Barr Virus, EBV)转化外周血B淋巴细胞<sup>[11,12]</sup>, 使其成为一种能连续分裂、永久生存的类淋巴母细胞系(LCLs), 即永生细胞株。人类和部分灵长类动物的成熟B淋巴细胞上具有EB病毒的受体, EB病毒可以通过

结合B淋巴细胞上的特异受体而感染B淋巴细胞, 导致细胞无限制增殖。在建立中国不同民族永生细胞库的过程中, 建株的成功率超过95%。一般采用含10% 二甲亚砜(dimethyl sulfoxide, DMSO)和20%胎牛血清的细胞冻存液在液氮中冻存细胞, 每个细胞株一般冻存2~3个代次, 每次10支, 作为种子细胞。此后每1~2年随即抽取细胞样本进行复苏和传代培养, 观察细胞的生长状态和活性。

### 1.3 永生细胞的质量控制

永生细胞实行细胞代次控制: 目的是保持永生细胞株的生物学和遗传学的正常特性, 本团队建立永生细胞株三级种子库的概念并按此标准进行质量管理。

永生细胞的支原体检测<sup>[13,14]</sup>: 支原体污染是永生细胞库的最主要危险, 严重时可使整个细胞库毁于一旦。除在操作中注意每一步骤外, 适时进行检测也是十分必要的。检测支原体污染的方法有很多, 如细胞培养法、免疫检测法、间接DNA染色(Hoechst33258染色)法及PCR(polymerase chain reaction)法。目前常用的检测方法是细胞培养法和PCR法, 前者操作较为麻烦, 但结果准确可靠。后者快速且灵敏, 但只能扩增引物所涵盖的特定支原体种类。在实践中往往这两种方法结合使用, 交叉验证<sup>[10]</sup>。

永生细胞的遗传稳定性: 细胞株的遗传特征是否保持稳定关系到使用永生细胞株进行科学研究的可靠性, 这常常通过染色体分析和分子标记来进行检测, 包括: (i) 染色体显带技术: 核型分析采用通用的染色体G带显色方法(G-banding), 也称为Giemsa染色。本团队对已建的细胞株前50代随机抽样, 经染色体核型G带分析, 均为正常的二倍体核型。有学者将EB病毒转化人B淋巴细胞6~58代划为前永生细胞株, 绝大部分具有正常的二倍体核型; 而后永生细胞株(167~509代)拥有异常核型的克隆<sup>[15,16]</sup>。(ii) 应用微卫星检测永生细胞的遗传稳定性<sup>[15,17]</sup>: 微卫星DNA又被称为短串联重复序列(short tandem repeat, STR), STR位点选取文献报道的高突变(突变率>0.09%)的13个STR位点, 对核基因组进行微卫星不稳定性检测。本团队对已建的细胞株前50代随机抽样, STR位点进行扫描检测, 证明其质量稳定可靠。(iii) 在长期传代的永生细胞株中的检测端粒酶活性: 端粒酶是一种特殊的核糖核蛋白逆转录酶, 能复制染色体末端的端粒结构, 保持染色体的

稳定性<sup>[18,19]</sup>。本团队对所建立的永生细胞株检测结果为细胞株在80代以前未有端粒酶活性的检出。

## 2 永生细胞的应用

实验证明, 永生细胞可以无限制地不断提供研究用的DNA, 以及淋巴细胞中表达的RNA和蛋白质, 永生细胞在传代180代以内仍然保持二倍体核型, 保留着供体对象的遗传稳定性。国际上公认应用100代以内的永生细胞进行的研究能够代表供体的基因组。

转化成功的永生细胞株有着广泛的用途。除参与国际国内合作外, 还是进行基因结构与表达和功能、遗传多样性和进化起源、分子病因学、免疫学、肿瘤疫苗、药物基因组学研究及药物筛选评价模型、再生医学和转化医学、全人单克隆中和抗体及EBV致病机理等研究领域的好材料。

### 2.1 中国不同民族永生细胞库用于遗传多样性研究和国际合作

中国不同民族永生细胞库向国内多家研究机构提供细胞、DNA和技术支持, 此外, 经国家人类遗传资源有关部门批准, 2001年中国人类基因组项目与位于巴黎的欧洲人类基因组多样性研究中心(Center d'Etude du Polymorphisme Humain, CEPH)签订合作协议, 中国提供149株永生细胞给CEPH。欧洲人类基因组多样性研究中心项目(HGDP-CEPH)始于20世纪90年代初, 聚焦于人类种群的遗传多样性, 并确定与不同疾病相关的遗传标记。中国提供的这149株LCLs, 经检测都达到国际认可的细胞质量标准, 由这些细胞提取的DNA已为许多国际研究机构所使用, 参与国际合作研究, 2011年CEPH经不完全统计, 有超过150篇论文中使用中国提供到CEPH的永生细胞株。

中国科学家作为人类基因组组织机构泛亚SNP协会(The HUGO Pan-Asian SNP Consortium)成员, 参与泛亚SNP协会的国际合作研究, 2009年, 采用中国不同民族永生细胞株开展的研究在*Science*发表论文“Mapping human genetic diversity in Asia”, 该论文通过SNP分析获得亚洲人类SNP遗传多样性图谱, 并以此分析亚洲人群的历史迁移路线, 其中东亚人群中14个群体共计132株永生细胞来自于中国不同民族永生细胞库<sup>[20]</sup>。

通过CEPH提供的上述中国不同民族14个群体共计132株永生细胞所提取的DNA, 中国不同民族永生细胞库实质性参与国际人群结构和人群多样性研究等国际合作. 2008年, Li等人<sup>[21]</sup>用全基因组多态性分析世界人群间关系, Jakobsson等人<sup>[22]</sup>使用来自HGDP-CEPH的485个个体样本分析各人群的全基因组SNP基因型、单倍型和拷贝数目多态性. 同样采用上述包括中国不同民族永生细胞库的样本, 论文发表于*Science*和*Nature*.

中国研究团队也对不同民族永生细胞库中的一些群体也进行全外显子测序, 昆明所与复旦大学团队合作进行民族源和流的分析, 证明云南氏羌起源的人群可能是在6700年前从藏区迁移过来, 百越与氏羌的遗传距离估计为7000年, 百越与百濮的遗传距离估计为5500年, 百濮居群相对孤立. 但云南少数民族内部可能因生活环境和生活习惯长期影响引起遗传差异, 这些遗传上的变异可能与疟疾适应相关, 相关结果在2022年*BMC Biology*发表<sup>[8]</sup>. 此后, 应用中国不同民族永生细胞库, 复旦大学、西安交通大学、中国医学科学院等26家单位联合发布中国人群泛基因组联盟(Chinese pangenome consortium, CPC)一期研究进展“A pangenome reference of 36 Chinese populations”, 2023年发表于*Nature*<sup>[9]</sup>. 该研究初步构建我国人群的泛基因组参考图谱, 新鉴定580万个点突变和小变异, 并发现目前通用参考基因组上缺失的参考序列, 进一步揭示中国人群中与适应性演化、DNA修复、免疫反应、寿命等表型或功能相关的基因序列.

## 2.2 利用永生细胞进行疾病基因的研究

由于细胞库的多群体性, LCLs非常适合用于为关联分析提供对照, 进行疾病致病基因的寻找. CEPH的细胞库样本曾被作为对照, 用以发现伊朗群体中区别于世界其他群体的特殊的SNP. 并结合功能讨论这些SNP和疾病的相关性, 以求发现潜在的疾病关联位点<sup>[23]</sup>. 在研究一些特定居住地群体与自然条件相关的基因多样性也有重要作用. 如在印度群体中, 永生细胞作为对照样本, 用于发现低氧适应相关基因<sup>[24]</sup>. 在世界范围内, 研究环境温度对人类种群发育和免疫力的影响, 并寻找影响作用的基因和产生选择的多态性位点<sup>[25]</sup>. 同时这一细胞库也用于改进全基因组关联分析时的策略, 以期能更好地发现基因和疾病的关联.

如: 研究人群特点对关联分析和连锁情况的影响, 为疾病的关联和分析提供前期基础<sup>[26-28]</sup>. 随着二代测序分析技术的发展, HGDP-CEPH永生细胞产生的数据, 也用于探索染色体的结构改变和疾病的关系. 如用LCLs探讨染色体缺失和疾病的相关性<sup>[29]</sup>, 或进行基因拷贝数变异(copy number variation, CNV)及与疾病发生的相关性研究<sup>[30,31]</sup>. 同时, 由于HGDP-CEPH细胞株包含世界性群体的信息, 还用于在世界范围内分析与疾病相关的SNP在全球范围内的分布情况, 如糖尿病的遗传风险位点筛选<sup>[32]</sup>. 并通过比较世界范围内不同群体中发生的自然选择作用, 来阐述疾病和遗传变异的关联关系<sup>[33,34]</sup>.

中国团队利用中国不同民族永生细胞库中的不同民族样本作为基本背景与采集病例对照, 对蒙古族、傣族和东亚群体的高血压易感位点进行研究<sup>[35-38]</sup>, 发现民族易感的一些高血压基因, 如*FAM110D*和*ADD1*<sup>[39]</sup>. 其中*FAM110D*和血压的相关性没有在其他民族群体中被观察到. 同样利用细胞库资源对云南民族群体的β珠蛋白基因进行民族多样性研究, 为中国少数民族地中海贫血的筛查和诊断分析提供遗传基础<sup>[36]</sup>.

此外, LCLs还可用于研究癌症的发生机制、肿瘤免疫逃逸机制. 在研究某些类型的癌症时, 特别是和EB病毒感染相关肿瘤如淋巴瘤和鼻咽癌等方面具有特别的价值. LCLs可用于发现EB病毒导致的淋巴细胞增生疾病的发病机制和治疗方案: 如激活某种特异T细胞可有效杀死EBV转化的自体淋巴母细胞B细胞系, 而某些T细胞却可以诱导EBV转化的淋巴母细胞B细胞系增生<sup>[40,41]</sup>. 也有学者用基因编辑的方法对EB病毒引起的肿瘤进行治疗<sup>[42]</sup>, 或用特异性抗体对EBV引起的淋巴肿瘤进行靶向治疗<sup>[43]</sup>. LCLs也可用于对自身免疫相关疾病的研究: 如对多发性硬化的发生机理的探讨. 多发性硬化症(multiple sclerosis, MS)被认为是由遗传易感性和个体的环境因素引发的. EB病毒感染是MS后续发展的主要危险因素, 研究来自MS感染个和对照组的EBV-LCLs的差异基因表达可发现MS的致病遗传位点和致病机理<sup>[44]</sup>. LCLs还可以用于研究针对自身免疫病的一些单克隆抗体, 如用系统性红斑狼疮(systemic lupus erythematosus, SLE)患者外周血的单个核细胞被Epstein-Barr病毒转化, 并测试由此产生的多克隆B淋巴母细胞系对某些T细胞系膜抗原的抗体

活性, 以此研究对抗系统性红斑狼疮的单克隆抗体<sup>[45,46]</sup>.

### 2.3 应用永生细胞进行药物、疫苗和干细胞的研究

EBV转化的永生技术能够快速、高效、可靠地生产无限数量的个性化细胞. 因此, LCLs被广泛用作分子和细胞功能研究的生物材料来源, 能提供基因组DNA, 以及大多RNA、蛋白质和系统生物学分析的无限材料, 可用作各种医学研究的细胞培养模型, 除用于分子和遗传学基础研究外, 在药物、疫苗和干细胞等众多研究领域已有较广泛的应用, 尤其在高通量生物学研究和大规模药物筛选领域展现出巨大的应用潜力和良好的应用前景, 为生物科学研究提供强有力的支撑作用<sup>[3]</sup>.

(1) 用于药物基因组学研究和筛选药物反应相关基因. 在日趋发展的精准医学领域, 遗传变异是患者对药物产生不同反应性的主要原因, 在个体和群体中均表现明显, 如抗肿瘤药物和治疗心血管疾病的药物的有效性和毒性常常呈现出个体和群体差异, 从群体水平可发现遗传变异与药物反应性之间的关系和机制. LCLs源于外周血B淋巴细胞, 代表宿主基因组的种系变异, 采用LCLs进行的药物基因组学研究已经证明人体对药物的反应性也是一种遗传性状. 与常见临床试验样本相比, LCLs细胞模型具备可用样本量大、可以覆盖多个种族和人群、细胞株的基因测序和基因分型数据可以公开获得等优点, 是经济、高效的药物关联基因筛选和验证系统. 基于LCLs的人类基因组高通量测序和关联分析(genome wide association study, GWAS)可以发现和识别影响药物反应性的新基因和变异, 探究和描述药物剂量效应关系的分子基础.

(2) 用于发现抗肿瘤药物遗传标记. 抗肿瘤药物反应性的体外研究以往主要集中在使用肿瘤细胞系来发现体细胞突变与药物反应性之间的关联性, 虽然长期以来已经确定肿瘤中的体细胞系突变(somatic mutations)会影响药物的疗效, 但大量的研究积累表明生殖系变异(germline variants)同样会影响抗肿瘤药物疗效, 甚至可能比体细胞系突变有更大的影响. 因此在肿瘤药物基因组学中, 这两类突变均需要进行检测和鉴定, 从而最大限度地提高药物疗效, 并且最大限度地降低药物的毒性和副作用.

尽管存在因组织表达差异性所致的一些药物转运

酶在永生淋巴细胞中表达不高等局限性, LCLs模型在早期就应用于卡铂、顺铂、酪氨酸激酶抑制剂等众多抗癌药物的基因组学研究, 取得一系列重大发现. 例如在早期的研究中, Wheeler等人<sup>[47]</sup>在2013年就使用608株LCLs细胞模型, 通过对超过3000万个SNPs位点的检测和GWAS研究, 分析抗肿瘤系列铂制剂在不同人群中反应差异相关的遗传变异, 发现*NBAS*和*KRT16P2*分别与卡铂和顺铂的毒性相关联, 并进一步证实*BCL2*, *GSTM1*, *GSTT1*, *ERCC2*和*ERCC6*等基因变异与铂制剂反应性的关联. 此外, 一些研究还使用LCLs模型陆续发现*MGMT*基因的SNP与替莫唑胺药效的关联性, *FKBP5*表达水平与阿糖胞苷细胞毒性的关联等. 研究人员使用680株来自千人基因组计划(1000 Genomes Project)的LCLs, 采用高通量筛选、GWAS和基因表达分析, 方法, 识别与44种FDA(U.S. Food and Drug Administration)批准的常见抗癌药物反应性密切相关的基因, 结果发现*NQO1*基因的SNP rs1800566多态性可通过影响基因的表达和蛋白的活性, 进而影响常见的药物反应信号通路, 从而与三氧化二砷、厄洛替尼、曲美替尼, 以及紫杉醇+表柔比星联合等常见抗癌药物的耐药性和患者的剂量反应性密切相关, 可作为调节药效的治疗靶点. 该研究表明LCLs可用于筛选和发现药物协同作用效应的生物学标记物<sup>[48]</sup>.

(3) 用于发现抗心血管疾病药物遗传标记. LCLs中存在有胆固醇稳态途径并发挥功能作用, 这是LCLs适用他汀类药物遗传学研究的基础. 研究证明他汀类药物抑制胆固醇生物合成作用中的关键转录因子*SREBF2*通路的成分和效应受体LDLR存在于LCLs中; 此外, 他汀类药物如辛伐他汀可导致LCLs细胞内胆固醇降低, 并引起*HMGR*和其他*SREBF2*靶基因的显著上调, 还有, LCLs中LDLR细胞表面蛋白表达增加且与LDL中载脂蛋白*APOB*减少相关<sup>[49]</sup>. 使用LCLs的GWAS分析发现*MYLIP*等基因位点与他汀类药物的LDL-C反应性相关, 通过LCLs表达谱分析发现长非编码RNA*RNP1-13D10.2*等他汀类药物诱导基因转录物表达与体内他汀类药物反应存在相关性.

LCLs可同时提供遗传变异、RNA和蛋白基因表达和细胞表型三个维度的信息及两两之间的关联性. 有研究团队利用480株LCLs细胞, 发现*RHOA*是一个新的他汀类药物效应基因, 与LCLs供体的LDL-C水平相

关联。通过分析基因变异与细胞表型的关联性, 揭示6个与辛伐他汀作用效应关联的表达数量性状位点(differential expression quantitative trait loci, expression quantitative trait loci, eQTL)和差异表达数量性状位点(differential expression quantitative trait loci, deQTL), 发现影响肌酸代谢限速酶*GATM*的遗传变异位点和可能的机制<sup>[50]</sup>。此外, 该团队在随后的工作中, 使用259源于欧洲人和153个非裔美国人的LCLs的RNA-seq(RNA sequencing)数据, 进一步确定*TBC1D4*, *OAS1*, *GLUL*等15个基因的重要顺式*cis*-deQTL, 其中许多基因不仅与他汀类药物临床效应相关, 还在人类健康和疾病中具有功能, 例如病毒防御、葡萄糖调节和对化疗药物的反应等, 表明他汀类药物对多种生理和疾病状态可发挥影响作用<sup>[51,52]</sup>。

LCLs在胆固醇药物基因组研究应用中还有一些局限性, 包括LCLs中不存在参与他汀类药物激活和代谢的*SLCO1B1*转运蛋白和酶, 如*CYP3A4*和*CYP3A5*, 这使得LCLs不适合研究他汀类药物的药代动力学; 此外LCLs不表达*APOB*和胆汁酸生成的基因, 因此不能分别模拟含APOB脂蛋白的合成或分泌或胆固醇降解和排泄。还有, 药物对LCLs细胞的作用和影响可能受到EBV的对细胞的转化效应影响。尽管有这些局限性, LCLs已被证明是他汀类药物基因组研究中有价值的细胞模型, 可用于识别与胆固醇和脂蛋白代谢相关的个体差异, 用于识别与他汀类药物反应细胞和临床相关的遗传标记, 为评估遗传变异的功能提供独特的遗传资源。

(4) 用于线粒体疾病药物筛选。由于LCLs具有容易从外周血获得和可无限增殖的特点, 尤其是携带mtDNA突变的LCLs可以在细胞水平上永久保持其生物学特性, 已获批或正在研发的线粒体药物中, 很大一部分是小分子化合物。因此LCLs作为线粒体疾病的药物筛选模型具有明显的优点, 使用这种疾病细胞模型可以大大提高线粒体药物发现的效率和准确性。Chin等人构建五个含有mtDNA点突变的永生淋巴瘤母细胞系, 以评估使用患者来源的LCLs在线粒体疾病中发现小分子药物的能力, 艾地苯醌作为一种辅酶Q(CoQ)的类似物, 可增加有MT-ND1 m.3460G>A突变或MT-ND4 m.11778G>A突变的Leber遗传性视神经病变(Leber's hereditary optic neuropathy, LHON)患者来源LCLs的基础呼吸能力, 但在有MT-ATP6 m.8993T>G变

体的Leigh病细胞中则未观察到类似的现象<sup>[51]</sup>。GNE-7915是一种富含亮氨酸的重复激酶2(leucine-rich repeat kinase 2, *LRRK2*)抑制剂, 可以修复由与帕金森病相关的*LRRK2* p.G2019S突变引起的线粒体DNA损伤<sup>[53]</sup>。这些结果表明携带mtDNA突变患者来源的LCLs可能对同一药物表现出不同的敏感性, 也揭示淋巴瘤母细胞作为线粒体药物筛选细胞模型的潜在价值。

(5) 用于疫苗和抗体研究。早期的研究发现LCLs具有较强的抗原递呈能力, 激活细胞毒性T淋巴细胞(CTL)从而激活细胞免疫, 在控制感染、肿瘤疫苗研究和肿瘤细胞治疗中发挥作用。由于几乎所有的EBV基因表达产物都能在LCLs中找到, 这些EBV转化的B淋巴细胞可作为有效的抗原递呈细胞(antigen-presenting cells, APC)尤其是在靶向EBV抗原的肿瘤治疗中有重要的应用价值。此外, LCLs细胞库丰富的遗传多样性, 为肿瘤、自身免疫和抗感染免疫研究提供丰富的HLA基因型资源和活的免疫细胞材料, 且LCLs具有作为肿瘤睾丸抗原(cancer-testis Antigen, CTA)的潜在用途。早期的研究发现LCLs中具有CTA基因*SSX4*, *GAGE*, *SSX1*等的表达, 尽管这些基因在LCLs中表达水平不高, 但使用DNA甲基转移酶1抑制剂或与组蛋白脱乙酰酶抑制剂联合使用可显著增加CTA的表达, 从而使LCLs能够诱导CTA特异性T细胞反应, 由此可以看出, CTA基因表达在LCLs细胞长期培养期间的稳定性使LCLs成为血液肿瘤和实体瘤中癌睾丸抗原疫苗的良好候选者<sup>[54]</sup>。

LCLs还广泛用于人的单克隆抗体制备, 尤其是采用EBV转化人的记忆B细胞, 成功筛选多种病原体的单克隆中和抗体, 例如抗SARS冠状病毒、流感、HCV、基孔肯雅病毒等病原的单抗。此外, 在自身免疫疾病治疗中, 还可用于以获得抗Rh-抗原D的特异性抗体。

(6) 用于制备诱导性多功能干细胞和精准医学研究。LCLs具有快速生长、可分裂、不受有限生命周期限制等特点, 可持续地获取大量同质的人类细胞系, 是解决生物医药领域众多方向研究和应用的基础材料。在此基础上, 通过诱导多能干细胞(induced pluripotent stem cells, iPSCs)和基因编辑等技术使LCLs得以向更多应用领域进行延伸。其中一个重要的目的是设计和制备具有某些遗传特征的细胞模型, 进而在疾病发生机理和药物研发等领域开展更为全面、准确和系统的

研究,这不仅为再生医学提供强有力的研究工具,也为遗传病研究、个性诊疗、药物筛选等领域带来更多机遇。

可再生体外细胞培养物如LCLs有助于人们研究遗传对人类特征的影响。然而,细胞系在多大程度上能忠实地保持供体的特异性表型仍存在争议,尤其是LCLs在经过EBV转化和传代培养中可能导致供体特异性基因表达特征的丢失。LCLs具有通过使用细胞分化方案模拟组织特异性生理学的潜力,iPSC系统是一种替代方法,LCLs能无限地提供研究所需的DNA,在一定的代次内保持基因组的稳定性,而且,由于是活细胞,还能提供RNA,并开展在活细胞上进行的功能研究。有研究通过对6个供体的成熟LCLs培养物重新编程为iPSC,表明重编程过程会导致LCLs供体特异性基因调控特征的恢复,LCLs是生成iPSC的合适起始材料<sup>[55]</sup>。对帕金森病患者的LCLs衍生iPSCs(LiPSCs)的研究表明,LiPSCs与LCLs细胞相比,基因组中体细胞变异的出现没有显著差异,并且LiPSCs可以通过使用直接神经球转换方法(dNS方法)分化为功能神经元,显示出与成纤维细胞诱导的多功能干细胞(dermal fibroblast-induced pluripotent stem cells, DF-iPSCs)相似的帕金森病表型,表明LCLs是生成iPSC和模拟神经系统疾病的强大工具,LCLs存储库可用作生成iPSC和疾病模型的全球性资源<sup>[56]</sup>。

大量研究表明,LCLs展现出其在诱导多能干细胞方面的潜力,LCLs可经过诱导后在保留细胞的基因组特征的前提下,形成具有分化为多组织能力的干性细胞<sup>[55-60]</sup>。Barrett等人<sup>[57]</sup>建立基于LCLs诱导iPSCs的高效方法,证明LCLs来源的iPSCs表现出与成纤维细胞来源的iPSCs相同的特征,在保留其基因型的同时,表现出正常的多能性,并且易于分化为所有三种胚层细胞类型。Choi等人<sup>[59]</sup>的研究进一步证实基于LCLs的iPSCs没有发生EBV相关的基因整合,也未在iPSCs中检测到关键病毒蛋白的表达。Fujimori等人<sup>[56]</sup>利用从两个来自健康供体和携带*PARK2*突变患者的LCLs克隆中建立iPSCs,且可分化为功能神经元,并表现出与皮肤成纤维细胞诱导的iPSCs相似的几种帕金森病表型。这表明可以基于病人来源的LCLs,通过诱导成为iPSCs,从而可开展进行特定组织中的致病机制研究。

个性化治疗和精准医学常常要求具有源于患者自身的有活性的生物学样本,例如血液和手术样本,但是

除肿瘤学研究外,常常难以获得高质量的生物样本,严重制约个性化诊疗的发展,尤其是对神经精神系统疾病的个性化医疗,脑组织的获取是困难的,而且个体死亡后所获得的脑组织常常具有不需要的生物学信号影响。iPSC和LCLs是最常见的两种体外细胞来源和细胞模型,LCLs与人iPSC衍生的神经元相比,具有体细胞突变少,多克隆特性,使用简便、经济,培养和运输条件要求低,有大型健康和疾病队列样本,有长期研究和历史等,每年有百篇以上文献报道。因此,尽管LCLs具有与体外培养神经元细胞系或iPSC衍生的神经元样细胞明显不同,但LCLs保留一些原始工体的表观基因组特征和表达谱特征,仍然可以用于神经系统疾病的基因组、蛋白组、代谢组和药物反应生物标记物等研究<sup>[61]</sup>。重要的是,患者特异性LCLs可用于生成诱导多能干细胞(iPSC),而iPSC又可用于创建特定细胞类型以用于机制研究。LCLs诱导的iPSC在罕见病如Charcot-Marie-Tooth病(Charcot-Marie-Tooth disease type 1B, CMT1B),以及药物代谢CYP基因多态性人诱导多能干细胞(hiPSC)系的建立和研究中得以成功应用<sup>[62,63]</sup>。

LCLs到iPSC重编程效率可通过方法的优化得以明显提高,例如使用表达OCT3/4, SOX2, KLF4, L-MYC, LIN28,小鼠多能转录因子p53DD的质粒和商业化重编程介质,可实现100%的重编程效率,并且生成的iPSC具有与人类ESC非常相似的转录和功能特征<sup>[64,65]</sup>。基于oriP/EBNA-1的附加型载体的电穿孔也是一种有效的基因传递方法,可以将LCLs有效地重编程为iPSC<sup>[66]</sup>。使用优化的CRISPR激活(CRISPR activation, CRISPRa)激活内源性多能性因子可以减少这些问题,额外靶向胚胎基因组活性富集的Alu基序和miR-302/367基因座可支持使用CRISPRa对人类细胞进行高质量的多能重编程<sup>[67]</sup>。

## 2.4 应用永生细胞进行基因编辑

LCLs细胞可利用以CRISPR/Cas9系统或PE系统(prime editing system)为代表的基因编辑技术对LCLs细胞进行基因敲除或基因敲入,通过引入特定的基因变异或多态性位点,研究这些变异对基因功能和人类疾病的影响,建立疾病模型进而深入了解各种遗传疾病的发病机制和研发治疗方法。LCLs和基因编辑技术也可以进行药物基因组学细胞模型的研究,通过对相关通路甚至免疫相关基因的编辑,高效地对药物或疫

苗进行筛选,以寻找特定亚群对药物的敏感性和耐药性。更重要的一点是,LCLs建立时所使用的B淋巴细胞是来源于特定个体的血液样本,可以通过基因编辑或细胞重编程等策略,形成诱导多功能细胞,具有从淋巴细胞扩展到其他组织细胞类型的能力和潜力,因此在LCLs细胞库的基础上开展的研究和测试将会有助于开发个性化的医疗方案,对个体的精准医疗有着重要的意义。

除可以针对多种研究目的构建基于LCLs构建细胞模型之外,利用基因编辑技术也可以在LCLs上开展EBV致肿瘤机制的研究。EBV感染全世界95%的成年人,与每年超20万例人类恶性肿瘤相关,也是免疫抑制相关淋巴瘤的主要感染病因<sup>[68,69]</sup>。但EBV诱导B淋巴细胞的机制尚不完全清楚, Ma等人<sup>[42]</sup>利用CRISPR/Cas9系统地进行全基因组功能缺失筛查,识别对EBV感染LCLs细胞生长和生存至关重要的宿主依赖因子,为EB与宿主相互作用的研究提供资源,并强调基于基因编辑技术的筛查在人类肿瘤病毒研究中的效用。

LCLs具有大量提供稳定基因组材料的特点,可以作为基因检测相关研究中标准质控品的制备。Zhou等人<sup>[70]</sup>利用CRISPR-Cas9技术,在LCLs细胞中定向引入在 $\alpha$ -地中海贫血中观察到的基因型缺失,经编辑的淋巴瘤母细胞系可作为基因检测的标准质控材料。

在LCLs基础上进行基因编辑的技术也存在一些挑战并经历发展。B淋巴细胞系通常由于不具有脂质转染性而难以转染,因此使得基因编辑甚至异位基因表达或基因沉默都较为困难<sup>[71,72]</sup>。B细胞、淋巴瘤细胞和白血病细胞的细胞系通常被认为是难转染的细胞。通过逆转录病毒或非整合型病毒载体转染的方法通常会存在生物安全以及伴随炎症样反应的问题<sup>[73-76]</sup>,而常规的电转法又常受到质粒大小及转染效率较低的影响<sup>[77,78]</sup>。Canoy等人<sup>[79]</sup>通过优化细胞培养

条件及优化电转染介质中的盐浓度和使用的质粒量,实现分别约79%转染率和58%的活性率。Jiang等人<sup>[80]</sup>则是通过优化慢病毒包装及优化sgRNAs的生成和传递,优化在EBV转化中进行CRISPR/Cas9编辑的方法,该方案在从多个不同的人类供体建立的LCLs中具有高度的可重复性。这些方法显著降低应用LCLs进行基因编辑实验的技术难度,使LCLs与基因编辑等先进技术的结合具有更宽广的前景。

### 3 总结与展望

随着人类基因组研究不断引领科技发展,生物样本库在健康和医学事业中日益受到重视,与此同时,国内外对于人类遗传资源的保护与开发研究势头强劲,在众多生物样本库中,人类永生细胞库具有巨大的优势。从无限量提供保持遗传特性的DNA、RNA的角度,其实更具有“生物银行(Biobank)”的价值。人类几乎所有疾病都是遗传与环境共同作用的结果,中国不同民族永生细胞库提供不同民族群体的遗传背景生物材料,对于疾病基因的研究具有很大意义。在遵循国家人类遗传资源管理要求的基础上,细胞库的建设可由健康人群向疾病人群扩展,建立疾病家系的永生细胞系进行深入的研究。

但建立人类永生细胞库对技术和设施仍有一定的要求,要求具备全年维持工作人员和稳定液氮供应,还须有严格的质量控制和管理,因此在一些地区很难扩大规模。此外,永生细胞系在研究上也有一些局限性,如不适合做甲基化的研究等。

中国不同民族永生细胞库20多年来为国内科研院所和大专院校提供一些民族群体细胞株提取的DNA,但总的来说数量规模仍然做得非常局限,希望随着我国对生物样本库建设的投入加强,中国不同民族永生细胞库在健康与疾病研究中发挥更大的作用。

### 参考文献

- 1 Chu J Y, Xu J J, Fu S B, et al. The establishment of immortal cell bank of Chinese nation (in Chinese). *Int J Genet*, 2008, 31: 8-25 [褚嘉祐, 徐玖瑾, 傅松滨, 等. 中华民族永生细胞库的建立. *国际遗传学杂志*, 2008, 31: 8-25]
- 2 Huang X Y, Liu A, Yu Y, et al. Establishment and preservation of ten national immortal cell lines in China (in Chinese). *Hereditas*, 2002, 24: 643-645 [黄小义, 刘岸, 于旸, 等. 中国10个民族永生细胞系的建立与保存. *遗传*, 2002, 24: 643-645]
- 3 Xu C F and Duan Z Y. A supporting role of Chinese National Immortalized Cell Bank in life science research (in Chinese). *Hereditas*, 2017, 39:

- 75–86 [许崇凤, 段子渊. 中华民族永生细胞库在生命科学研究中的支撑作用. *遗传*, 2017, 39: 75–86]
- 4 Cann H M, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science*, 2002, 296: 261–262
  - 5 Su B, Xiao J, Underhill P, et al. Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice Age. *Am J Hum Genet*, 1999, 65: 1718–1724
  - 6 Qian Y, Qian B, Su B, et al. Multiple origins of Tibetan Y chromosomes. *Hum Genet*, 2000, 106: 453–454
  - 7 Su B, Jin L, Underhill P, et al. Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA*, 2000, 97: 8225–8228
  - 8 Yang Z, Chen H, Lu Y, et al. Genetic evidence of tri-genealogy hypothesis on the origin of ethnic minorities in Yunnan. *BMC Biol*, 2022, 20: 166
  - 9 Gao Y, Yang X, Chen H, et al. A pangenome reference of 36 Chinese populations. *Nature*, 2023, 619: 112–121
  - 10 Chu J Y. The establishment of immortal cell bank of Chinese nation—theory and practice (in Chinese). Shanghai: Shanghai Scientific & Technical Publishers, 2009 [褚嘉祐. 中华民族永生细胞库的建立: 理论与实践. 上海科学技术出版社, 2009]
  - 11 Miller G, Lipman M. Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. *Proc Natl Acad Sci USA*, 1973, 70: 190–194
  - 12 Neitzel H. A routine method for the establishment of permanent growing lymphoblastoid cell lines. *Hum Genet*, 1986, 73: 320–326
  - 13 Lu J H. Mycoplasma infective in cell culture and its detection technique Medical biologics. Beijing: People's Medical Publishing House, 1995 [卢锦汉. 细胞培养感染支原体及其检测技术医学生物制品学, 北京: 人民卫生出版社, 1995]
  - 14 Uphoff C C, Drexler H G. Detection of mycoplasma contamination in cell cultures. *CP Mol Biol*, 2014, 106
  - 15 Mohyuddin A, Ayub Q, Siddiqi S, et al. Genetic instability in EBV-transformed lymphoblastoid cell lines. *Biochim Biophys Acta Gen Subj*, 2004, 1670: 81–83
  - 16 Shirley M D, Baugher J D, Stevens E L, et al. Chromosomal variation in lymphoblastoid cell lines. *Hum Mutat*, 2012, 33: 1075–1086
  - 17 Li Y H, Huang X Q, Lin K Q, et al. Study on long-term genetic stability of immortal cell line established by EB virus transformed B lymphocytes. *Chin J Med Genet*, 2008, 25: 276–279 [李彦涵, 黄小琴, 林克勤, 等. EB病毒转化B淋巴细胞建立的永生细胞株长期传代遗传稳定性的研究. 中华医学遗传学杂志, 2008, 25: 276–279]
  - 18 Kataoka H, Tahara H, Watanabe T, et al. Immortalization of immunologically committed Epstein-Barr virus-transformed human B-lymphoblastoid cell lines accompanied by a strong telomerase activity. *Differ*, 1998, 62: 203–211
  - 19 Sugimoto M, Tahara H, Ide T, et al. Steps involved in immortalization and tumorigenesis in human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. *Cancer Res*, 2004, 64: 3361–3364
  - 20 Abdulla M A, Ahmed I, Assawamakin A, et al. Mapping human genetic diversity in Asia. *Science*, 2009, 326: 1541–1545
  - 21 Li J Z, Absher D M, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 2008, 319: 1100–1104
  - 22 Jakobsson M, Scholz S W, Scheet P, et al. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, 2008, 451: 998–1003
  - 23 Tong P, Prendergast J G, Lohan A J, et al. Sequencing and analysis of an Irish human genome. *Genome Biol*, 2010, 11: R91
  - 24 Aggarwal S, Gheware A, Agrawal A, et al. Combined genetic effects of EGLN1 and VWF modulate thrombotic outcome in hypoxia revealed by Ayurgenomics approach. *J Transl Med*, 2015, 13: 1
  - 25 Ji L, Wu D, Xie H, et al. Ambient temperature is a strong selective factor influencing human development and immunity. *Genom Proteom BioInf*, 2020, 18: 489–500
  - 26 Bosch Fusté E, Laayouni H, Morcillo Suárez C, et al. Decay of linkage disequilibrium within genes across HGDP-CEPH human samples: most population isolates do not show increased LD. *BMC Genom*. 2009 10: 338
  - 27 Conrad D F, Jakobsson M, Coop G, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet*, 2006, 38: 1251–1260
  - 28 Huyghe J R, Fransen E, Hannula S, et al. A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. *Eur J Hum Genet*, 2011, 19: 347–352
  - 29 Holmes L V, Strain L, Staniforth S J, et al. Determining the population frequency of the CFHR3/CFHR1 deletion at 1q32. *PLoS ONE*, 2013, 8: e60352
  - 30 Colobran R, Comas D, Faner R, et al. Population structure in copy number variation and SNPs in the CCL4L chemokine gene. *Genes Immun*, 2008, 9: 279–288
  - 31 Ghahramani Seno M M, Kwan B Y, Lee-Ng K K M, et al. Human PTCHD3 nulls: rare copy number and sequence variants suggest a non-

- essential gene. *BMC Med Genet*, 2011, 12: 1–9
- 32 Guinan K J. Worldwide distribution of type II diabetes-associated TCF7L2 SNPs: evidence for stratification in Europe. *Biochem Genet*, 2012, 50: 159–179
- 33 López Herráez D, Bauchet M, Tang K, et al. Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS ONE*, 2009, 4: e7888
- 34 Patillon B, Luisi P, Blanche H, et al. Positive selection in the chromosome 16 VKORC1 genomic region has contributed to the variability of anticoagulant response in humans. *PLoS ONE*, 2012, 7: e53049
- 35 Huang L, Chu Y, Huang X, et al. Association between gene polymorphisms of voltage-dependent  $\text{Ca}^{2+}$  channels and hypertension in the Dai people of China: a case-control study. *BMC Med Genet*, 2020, 21: 44
- 36 Sun H, Liu H, Huang K, et al.  $\beta$ -globin gene cluster haplotypes in ethnic minority populations of southwest China. *Sci Rep*, 2017, 7: 42909
- 37 Sun H, Yang Z Q, Liu S H, et al. Validation of the correlation between three GWAS positive transcription factor genes and hypertension in Dai and Mongolian populations in China (in Chinese). *Chin J Birth Health Heredity*, 2014, 22: 15–19 [孙浩, 杨昭庆, 刘舒媛, 等. 中国傣族及蒙古族群体中3个GWAS阳性转录因子基因与高血压的相关性的验证研究. *中国优生与遗传杂志*, 2014, 22: 15–19]
- 38 Sun H, Yang Z Q, Wang X Y, et al. Validation of 9 SNPs positive for hypertension in East Asian population in Dai and Mongolian population in China (in Chinese). *Chin J Birth Health Heredity*, 2014, 22: 8–13 [孙浩, 杨昭庆, 王秀云, 等. 东亚群体中高血压GWAS阳性9个SNP在中国傣族及蒙古族群体中的验证研究. *中国优生与遗传杂志* 2014, 22: 8–13]
- 39 Zhang L, Li W, Weng Y, et al. A novel splice site variant in the *POPDC3* causes autosomal recessive limb-girdle muscular dystrophy type 26. *Clin Genet*, 2022, 102: 345–349
- 40 Lacerda J F, Ladanyi M, Louie D C, et al. Human Epstein-Barr virus (EBV)-specific cytotoxic T lymphocytes home preferentially to and induce selective regressions of autologous EBV-induced B cell lymphoproliferations in xenografted C.B-17 scid/scid mice. *J Exp Med*, 1996, 183: 1215–1228
- 41 Xiang Z, Liu Y, Zheng J, et al. Targeted activation of human  $\text{V}\gamma 9\text{V}\delta 2$ -T cells controls Epstein-Barr virus-induced B cell lymphoproliferative disease. *Cancer Cell*, 2014, 26: 565–576
- 42 Ma Y J, Walsh M J, Bernhardt K, et al. CRISPR/Cas9 Screens reveal Epstein-Barr Virus-transformed B cell host dependency factors. *Cell Host Microbe*, 2017, 21: 580–591.e7
- 43 Lai J, Tan W J, Too C T, et al. Targeting Epstein-Barr virus-transformed B lymphoblastoid cells using antibodies with T-cell receptor-like specificities. *Blood*, 2016, 128: 1396–1407
- 44 Wieland L, Schwarz T, Engel K, et al. Epstein-Barr virus-induced genes and endogenous retroviruses in immortalized B cells from patients with multiple sclerosis. *Cells*, 2022, 11: 3619
- 45 Numasaki M, Fukuoka Y, Kudo T, et al. A novel human monoclonal antibody, TONO-1, reactive with T-lymphocytic leukemia cells. *Int J Cancer*, 1995, 62: 42–47
- 46 Sasaki T, Muryoi T, Sekiguchi Y, et al. Monoclonal human anti-DNA antibodies from EB virus-transformed lymphocytes of systemic lupus erythematosus (SLE) patients. *J Clin Immunol*, 1985, 5: 246–253
- 47 Wheeler H E, Gamazon E R, Stark A L, et al. Genome-wide meta-analysis identifies variants associated with platinating agent susceptibility across populations. *Pharmacogenomics J*, 2013, 13: 35–43
- 48 Akhtari F S, Green A J, Small G W, et al. High-throughput screening and genome-wide analyses of 44 anticancer drugs in the 1000 Genomes cell lines reveals an association of the *NQO1* gene with the response of multiple anticancer drugs. *PLoS Genet*, 2021, 17: e1009732
- 49 Theusch E, Kim K, Stevens K, et al. Statin-induced expression change of *INSIG1* in lymphoblastoid cell lines correlates with plasma triglyceride statin response in a sex-specific manner. *Pharmacogenomics J*, 2017, 17: 222–229
- 50 Mangravite L M, Engelhardt B E, Medina M W, et al. A statin-dependent QTL for *GATM* expression is associated with statin-induced myopathy. *Nature*, 2013, 502: 377–380
- 51 Chin R M, Panavas T, Brown J M, et al. Patient-derived lymphoblastoid cell lines harboring mitochondrial DNA mutations as tool for small molecule drug discovery. *BMC Res Notes*, 2018, 11: 205
- 52 Theusch E, Chen Y D I, Rotter J I, et al. Genetic variants modulate gene expression statin response in human lymphoblastoid cell lines. *BMC Genomics*, 2020, 21: 1–8
- 53 Howlett E H, Jensen N, Belmonte F, et al. LRRK2 G2019S-induced mitochondrial DNA damage is LRRK2 kinase dependent and inhibition

- restores mtDNA integrity in Parkinson's disease. *Hum Mol Genet*, 2017, 26: 4340–4351
- 54 Neumann F, Kaddu-Mulindwa D, Widmann T, et al. EBV-transformed lymphoblastoid cell lines as vaccines against cancer testis antigen-positive tumors. *Cancer Immunol Immunother*, 2013, 62: 1211–1222
- 55 Thomas S M, Kagan C, Pavlovic B J, et al. Reprogramming LCLs to iPSCs results in recovery of donor-specific gene expression signature. *PLoS Genet*, 2015, 11: e1005216
- 56 Fujimori K, Tezuka T, Ishiura H, et al. Modeling neurological diseases with induced pluripotent cells reprogrammed from immortalized lymphoblastoid cell lines. *Mol Brain*, 2016, 9: 1–4
- 57 Barrett R, Ornelas L, Yeager N, et al. Reliable generation of induced pluripotent stem cells from human lymphoblastoid cell lines. *Stem Cells Transl Med*, 2014, 3: 1429–1434
- 58 Bueno C, Sardina J L, Di Stefano B, et al. Reprogramming human B cells into induced pluripotent stem cells and its enhancement by C/EBP $\alpha$ . *Leukemia*, 2016, 30: 674–682
- 59 Choi S M, Liu H, Chaudhari P, et al. Reprogramming of EBV-immortalized B-lymphocyte cell lines into induced pluripotent stem cells. *Blood*, 2011, 118: 1801–1805
- 60 Rajesh D, Dickerson S J, Yu J, et al. Human lymphoblastoid B-cell lines reprogrammed to EBV-free induced pluripotent stem cells. *Blood*, 2011, 118: 1797–1800
- 61 Gurwitz D. Human iPSC-derived neurons and lymphoblastoid cells for personalized medicine research in neuropsychiatric disorders. *Dialogues Clin Neurosci*, 2016, 18: 267–276
- 62 Lee J, Woo D H, Park H J, et al. Human induced pluripotent stem cell line with cytochrome P450 enzyme polymorphism (CYP2C19\*2/CYP3A5\*3C) generated from lymphoblastoid cells. *Stem Cell Res*, 2018, 27: 34–37
- 63 Son D, Kang P J, Yun W, et al. Generation of induced pluripotent stem cell (iPSC) line from Charcot-Marie-Tooth disease patient with MPZ mutation (CMT1B). *Stem Cell Res*, 2017, 24: 5–7
- 64 Kumar S, Curran J E, Espinosa E C, et al. Highly efficient induced pluripotent stem cell reprogramming of cryopreserved lymphoblastoid cell lines. *J Biol Methods*, 2020, 7: e124
- 65 Kumar S, Curran J E, Glahn D C, et al. Utility of lymphoblastoid cell lines for induced pluripotent stem cell generation. *Stem Cells Int*, 2016, 2016: 1–20
- 66 Kim M, Park J, Kim S, et al. Generation of induced pluripotent stem cells from lymphoblastoid cell lines by electroporation of episomal vectors. *Int J Stem Cells*, 2023, 16: 36–43
- 67 Sokka J, Yoshihara M, Kvist J, et al. CRISPR activation enables high-fidelity reprogramming into human pluripotent stem cells. *Stem Cell Rep*, 2022, 17: 413–426
- 68 Cohen J I, Fauci A S, Varmus H, et al. Epstein-Barr virus: an important vaccine target for cancer prevention. *Sci Transl Med*, 2011, 3: 107fs107
- 69 Lieberman P M. Epstein-Barr virus turns 50. *Science*, 2014, 343: 1323–1325
- 70 Zhou L, Li R, Zhang R, et al. Utilizing CRISPR/Cas9 technology to prepare lymphoblastoid cell lines harboring genetic mutations for generating quality control materials in genetic testing. *Clin Lab Anal*, 2020, 34: e23256
- 71 Meacham J M, Durvasula K, Degertekin F L, et al. Physical methods for intracellular delivery: practical aspects from laboratory use to industrial-scale processing. *SLAS Tech*, 2014, 19: 1–18
- 72 Zhao N, Qi J, Zeng Z, et al. Transfecting the hard-to-transfect lymphoma/leukemia cells using a simple cationic polymer nanocomplex. *J Control Release*, 2012, 159: 104–110
- 73 Glover D J, Lipps H J, Jans D A. Towards safe, non-viral therapeutic gene expression in humans. *Nat Rev Genet*, 2005, 6: 299–310
- 74 Kaestner L, Scholz A, Lipp P. Conceptual and technical aspects of transfection and gene delivery. *Bioorg Med Chem Lett*, 2015, 25: 1171–1176
- 75 Kim T K, Eberwine J H. Mammalian cell transfection: the present and the future. *Anal Bioanal Chem*, 2010, 397: 3173–3178
- 76 Mosier D E. Introduction for “Safety considerations for retroviral vectors: a short review”. *Appl Biosaf*, 2004, 9: 68–75
- 77 Chicaybam L, Barcelos C, Peixoto B, et al. An efficient electroporation protocol for the genetic modification of mammalian cells. *Front Bioeng Biotechnol*, 2017, 4: 99
- 78 Machy P, Lewis F, McMillan L, et al. Gene transfer from targeted liposomes to specific lymphoid cells by electroporation. *Proc Natl Acad Sci USA*, 1988, 85: 8027–8031
- 79 Canoy R J, André F, Shmakova A, et al. Easy and robust electrotransfection protocol for efficient ectopic gene expression and genome editing in

human B cells. *Gene Ther*, 2023, 30: 167–171

80 Jiang S, Wang L W, Walsh M J, et al. CRISPR/Cas9-mediated genome editing in Epstein-Barr virus-transformed lymphoblastoid B-cell lines. *CP Mol Biol*, 2018, 121: 12–31

## The establishment and applications of the Immortal Cell Banks of Chinese ethnic groups

YANG ZhaoQing, HUANG XiaoQin, SUN Hao, BIAN Cheng, LIN KeQin & CHU JiaYou

*Department of Medical Genetics, Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Kunming 650118, China*

Human genetic resources are of great research significance. Different ethnic groups in China have their characteristics in terms of ethnic origin, settlement, language, culture, and customs. Due to the difference in founder effect, geographical environment and lifestyle, there are differences in physiological phenotype and disease susceptibility among the Chinese ethnic groups, which provides rich materials for medical and health research. For more than 20 years, the Immortal Cells Bank of Chinese Ethnic Groups has been supported by the National Natural Science Foundation of China and by the Ministry of Science and Technology of the People's Republic of China. So far the cell bank includes a total of 6126 immortalized cell lines from 49 Chinese ethnic groups, which include more than 90 Chinese ethnic subgroups. As an important national strategic resource, the Chinese immortal cell bank and the nucleic acids derived from it provide research materials for the applied researches of genetic structural characteristics, gene variation, expression and function of the Chinese populations, demonstrating the progress in discovering drug, vaccines and antibodies, and the application potential in stem cells research. In summary, the Immortal Cells Bank of Chinese Ethnic Groups has significantly promoted China's international exchanges and cooperation on human genetic resources. With the development of advanced technologies such as high-throughput sequencing and gene editing, it will provide more resources and tools for human life science research and clinical translation applications.

**human genetic resources, ethnic groups, genetic diversity, lymphoblastoid cell bank, disease genes, vaccines and medicines**

doi: [10.1360/SSV-2023-0035](https://doi.org/10.1360/SSV-2023-0035)