第九届省级教学成果奖（高等教育类）支撑材料明细表

成果代码：0411

成果名称：“一专多能”统合性特殊教育专业师资培养的路径与实践

成果主要完成人：王清路、王疆娜、田雪文、孙威、孙红梅、梁永胜、郭方玲、周彩霞、杨国昌、聂祥坤

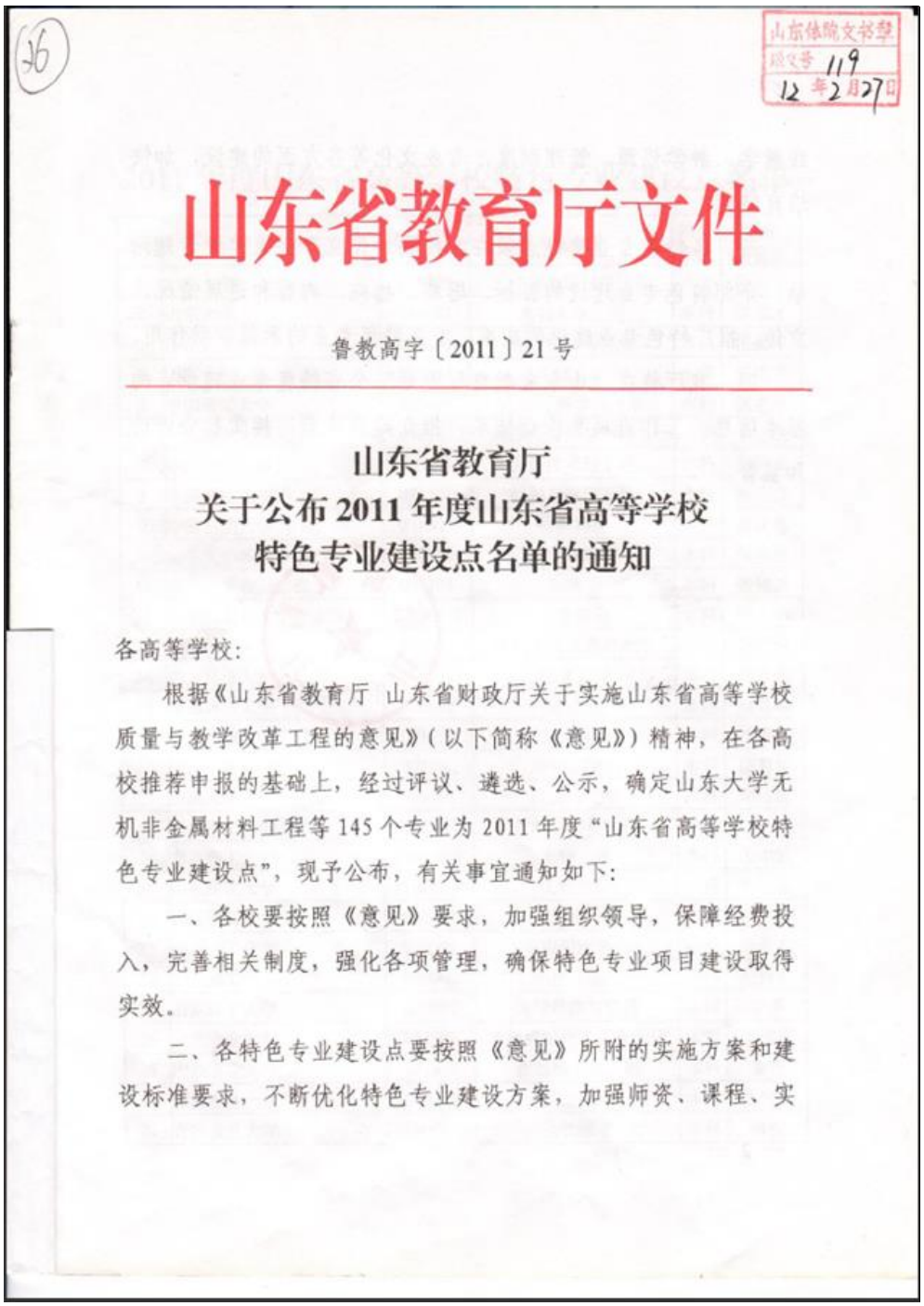
成果主要完成单位：山东新时代残疾人教育研究院
山东体育学院
齐鲁医药学院

山东省教育厅制
2021 年 10 月

目　录

1 专业建设取得的成就

1.1 山东省级特色专业建设点



山东体院文书室
登记号 119
12 年 2 月 27 日

山东省教育厅文件

鲁教高字〔2011〕21 号

山东省教育厅
关于公布 2011 年度山东省高等学校
特色专业建设点名单的通知

各高等学校：

根据《山东省教育厅 山东省财政厅关于实施山东省高等学校质量与教学改革工程的意见》（以下简称《意见》）精神，在各高校推荐申报的基础上，经过评议、遴选、公示，确定山东大学无机非金属材料工程等 145 个专业为 2011 年度“山东省高等学校特色专业建设点”，现予公布，有关事宜通知如下：

一、各校要按照《意见》要求，加强组织领导，保障经费投入，完善相关制度，强化各项管理，确保特色专业项目建设取得实效。

二、各特色专业建设点要按照《意见》所附的实施方案和建设标准要求，不断优化特色专业建设方案，加强师资、课程、实

践教学、教学资源、管理制度、专业文化等各方面的建设，加快培育特色。

三、各特色专业建设点须在学校网站设立专栏或建设专题网站，介绍特色专业建设的目标、思路、措施、内容和进展情况，宣传、推广特色专业建设的成果，发挥特色专业的示范带动作用。

四、我厅将在“山东省教育厅网站”公布特色专业建设点的基本信息、工作进展和检查结果，推介建设成就，接受社会评价和监督。

山东省教育厅

二〇一一年十二月二十日

序号	学校名称	专业代码	专业名称	培养层次	专业负责人
30	潍坊医学院	100801	药学	本科	高志芹
31	泰山医学院	081102	制药工程	本科	齐永秀
32	滨州医学院	100801	药学	本科	张树平
33	山东中医药大学	100804W	中草药栽培与鉴定	本科	张永清
34	山东中医药大学	100505W	中西医临床医学	本科	石作荣
35	济宁医学院	100801	药学	本科	上官国强
36	山东师范大学	070701	地理科学	本科	徐跃通
37	山东师范大学	070101	数学与应用数学	本科	傅希林
38	曲阜师范大学	080613W	网络工程	本科	曹宝香
39	曲阜师范大学	110201	工商管理	本科	张玉忠
40	聊城大学	070401	生物科学	本科	黄勇
41	聊城大学	070101	数学与应用数学	本科	孟广武
42	德州学院	050101	汉语言文学	本科	姜山秀
43	滨州学院	071201	电子信息科学与技术	本科	张循利
44	鲁东大学	080605	计算机科学与技术	本科	邹海林
45	鲁东大学	080204	高分子材料与工程	本科	陈厚
46	临沂大学	040105W	小学教育	本科	李同胜
47	临沂大学	070302	应用化学	本科	郑秀文
48	泰山学院	050406	美术学	本科	刘钢
49	济宁学院	070101	数学与应用数学	本科	朱松涛
50	菏泽学院	110204	财务管理	本科	胡尊东
51	山东财经大学（筹）	030101	法学	本科	赵信会
52	山东财经大学（筹）	020109W	金融工程	本科	安起光
53	山东体育学院	040103	特殊教育	本科	魏平
54	山东万杰医学院	100801	药学	本科	张平平
55	青岛滨海学院	050207	日语	本科	林成虎
56	枣庄学院	081101	化学工程与工艺	本科	刘雪静
57	山东工艺美术学院	050406	美术学	本科	唐家路
58	青岛大学	020104	金融学	本科	刘喜华
59	青岛大学	100304	医学检验	本科	刘成玉
60	烟台大学	080701	建筑学	本科	郝曙光

1.2 山东省骨干学科特殊教育与康复实验中心、山东省重点培育专业群

1.3 山东省高水平应用型立项建设专业---特殊教育专业群

山东省教育厅

鲁教高字〔2016〕11 号

山东省教育厅
关于公布高水平应用型立项建设专业（群）名单的通知

各省属公办本科高校:

根据《山东省教育厅山东省财政厅关于印发推进高水平应用型大学建设实施方案的通知》（鲁教高字〔2016〕8 号）要求，我厅组织专家对各省属公办本科高校推荐的高水平应用型立项建设专业（群）进行了审核认定，经商省财政厅同意，确定“十三五”期间对青岛大学高分子材料与工程专业（群）等 60 个重点专业予以立项支持，每个重点专业（群）资助建设经费 400 万

元/年；对青岛大学金融学专业（群）等40个培育专业予以立项支持，每个培育专业（群）资助建设经费200万元/年（2016年对每个培育专业（群）资助建设经费150万元）。现将名单予以公布。

各立项专业（群）建设高校要切实加强领导，按照实施方案要求，签订目标任务书，落实责任制，抓好项目管理，按期完成专业建设目标和建设任务，确保立项专业建设取得实效。

附件：1. 山东省高水平应用型重点立项建设专业（群）名单

2. 山东省高水平应用型培育立项建设专业（群）名单

山东省教育厅

2016年11月23日

序号	学校名称	培育专业（群）	
		核心专业	专业群
10	山东体育学院	**社会体育指导与管理**	公共事业管理、休闲体育、健康服务与管理
		特殊教育	运动康复、运动人体科学、应用心理学
11	济宁医学院	**法医学**	医学检验技术、卫生检验与检疫、药学、生物技术
12	山东政法学院	**法学**	监狱学、知识产权
13	临沂大学	**小学教育**	教育学、学前教育、心理学
		机械设计制造及其自动化	软件工程、网络工程、自动化、电气工程及其自动化
		物流管理	物流工程、国际经济与贸易、电子商务、工商管理、会计学
14	滨州学院	**计算机科学与技术**	通信工程、物联网工程、电子信息工程、光电信息科学与工程
15	德州学院	**物联网工程**	电子信息工程、网络工程、应用物理学
		服装设计与工程	纺织工程、非织造材料与工程、服装与服饰设计
16	菏泽学院	**化学工程与工艺**	能源化学工程、应用化学、环境科学与工程、制药工程
		生物工程	生物制药、生物科学、制药工程
17	济宁学院	**化学工程与工艺**	化学、应用化学、安全工程、高分子材料与工程
		机械设计制造及其自动化	电子信息工程、电气工程及其自动化

1.4 山东省一流本科专业建设点---特殊教育专业

山东体育学院一流本科专业建设点名单9个（国家级+省级）

更新：2021-5-8 11:56:42 发布：大学生必备网 纠错

高考填报志愿时，山东体育学院一流本科专业建设点名单有哪些是广大考生和家长朋友们十分关心的问题，根据山东体育学院官网发布的信息可知，山东体育学院共有2个专业入选了国家级一流本科专业建设点，7个专业入选了省级一流本科专业建设点（注：部分专业2019年度为省级，2020年度被评为了国家级），具体名单如下，供大家参考：

一流本科专业建设计划，简称“双万计划”，即2019-2020-2021年3年建设1万个左右国家级一流本科专业点和1万个左右省级一流本科专业点。

1、2020年度山东体育学院一流本科专业建设点名单（国家级+省级）

获批国家级一流本科专业建设点：休闲体育专业

获批省级一流本科专业建设点：应用心理学

2、2019年度山东体育学院一流本科专业建设点名单（国家级+省级）

1个专业入选国家级一流本科专业建设点名单：社会体育指导与管理专业

7个专业入选省级建设点名单：特殊教育、体育教育、运动训练、武术与民族传统体育、运动人体科学、运动康复、休闲体育

3、山东体育学院学科评估结果排名（第四轮）

以下是山东体育学院第四轮学科评估结果排名情况，也是最新评估结果。

序号	学校代码	学校名称	一级学科代码	一级学科名称	评选结果
1	10457	山东体育学院	0403	体育学	B-

4、学校简介

1 营养师月薪多少 2 营养师专业 3 二建 机电 4 中专升大专 5 本科自考 6 高铁乘务员要求 7 国有企业招聘网 8 助理工程师评审 9 工程文员 10 自考本科专业 11 助理医师报名 12 大额存单利率表
13 非全日制研 14 事业编制招 15 数字经济概 16 澳国立大学 17 埃默里大学 18 函授本科 19 资料员考试 20 高铁招收乘 21 利兹大学留 22 东莞入户条 23 本科专业目 24 资料员培训

全国高考
全国高考 江苏高考 浙江高考 山
福建高考 湖北高考 湖南高考 安
江西高考 河北高考 辽宁高考 吉
黑龙江 河南高考 陕西高考 山
甘肃高考 四川高考 贵州高考 云
青海高考 广东高考 广西高考 海
宁夏高考 北京高考 上海高考 天
重庆高考 新疆高考 西藏高考 内

大学名单
双一流 985大学 211大学 高
重点大学 全国大学 本科大学 专
公办大学 民办大学 公办本科 公

院校库
章程简章 录取分数 全国排名 广
院校评价 宿舍条件 收费标准 收
一流学科 一流专业 王牌专业 专

专业库
学科评估 专业认证 本科专业 专
专业介绍 专业排名 就业前景 专
热门专业 女生专业 男生专业

志愿填报
平行志愿 高考分数 一分一段 投

1.5 山东省一流本科课程---《运动心理学》、《临床分子生物学检验》

山 东 省 教 育 厅

鲁教高函〔2020〕3 号

山东省教育厅
关于公布 2019 年山东省一流本科课程认定结果的通知

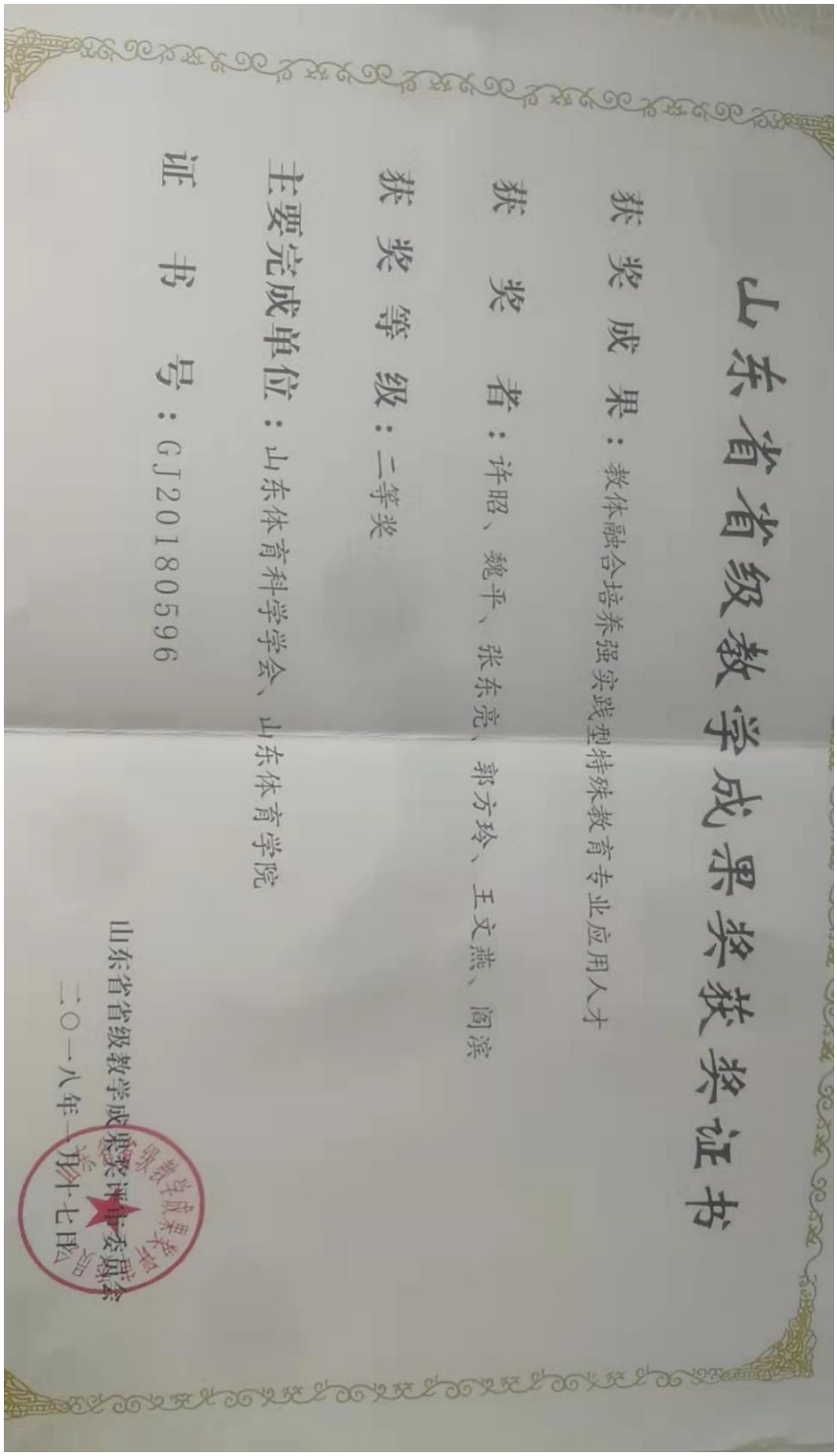
各本科高等学校:

根据《山东省教育厅关于印发<山东省一流本科课程建设实施方案>的通知》(鲁教高字〔2019〕6 号，以下简称《实施方案》)和《关于开展 2019 年山东省一流本科课程建设工作的通知》(鲁教高处函〔2019〕50 号)有关规定，经团队(个人)申报、学校推荐，专家评审与公示，我厅决定认定山东大学《病理生理学》等 798 门课程为 2019 年山东省一流本科课程。其中，线下一流课程 423 门，线上线下混合式一流课程 206 门，线上一流课程 27 门，虚拟仿真实验教学一流课程 113 门，社会实践一流课程 29 门。现将名单予以公布(详见附件)，并就有关事项通知如下:

序号	学校	课程名	负责人	团队成员
358	山东体育学院	体育概论	张志勇	李海霞，程卫波，牟艳，孔令鲁
359	山东体育学院	运动解剖学	费云芸	沙继斌，王传军，何伟
360	山东体育学院	运动生物力学	赖寒	杨杰，杨春荣
361	山东体育学院	运动心理学	许昭	魏平，张敏，张思
362	山东体育学院	运动训练学	石磊	陈艳萍，韩炜，王欣一男
363	山东协和学院	大学生创新创业教育	盛振文	吴伟伟，宋芹
364	山东协和学院	基础护理学	杨庆爱	林辉，许霞，原芸姿
365	山东协和学院	酒店管理概论	郭庆慧	林蒲，魏薇，邓革慧
165	齐鲁师范学院	高等代数	张克玉	江静，石啊莲，王建国
166	齐鲁师范学院	人体及动物生理学	孙洪兆	史远，王玢，王艺雅，于聪
167	齐鲁医药学院	口腔解剖生理学	董刚	姜春荣，高培福，曹雪梅，石岩
168	齐鲁医药学院	临床分子生物学检验	王清路	武天慧，张杰，王靖，王微

2 相关教研成果获奖

2.1 山东省教学成果奖---教体融合培养强实践型特殊教育专业应用人才（郭方玲）



山东省省级教学成果奖获奖证书

获奖成果：教体融合培养强实践型特殊教育专业应用人才

获奖者：许昭、魏平、张东亮、郭方玲、王文燕、阎滨

获奖等级：二等奖

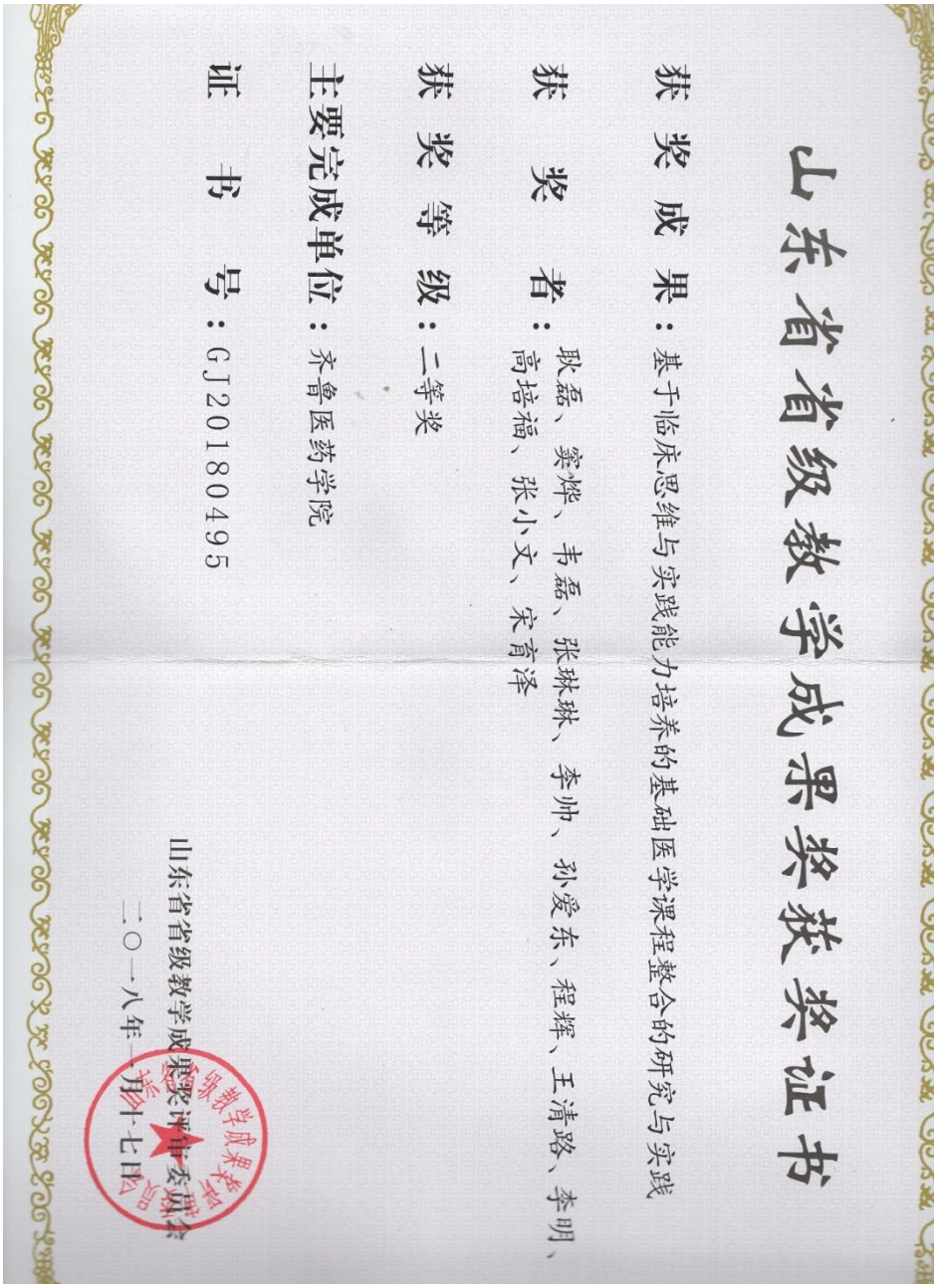
主要完成单位：山东体育科学学会、山东体育学院

证书号：GJ201805 96

山东省省级教学成果奖评审委员会

二〇一八年一月廿七日

2.2 山东省教学成果奖---基于临床思维与实践能力培养的基础医学课程整合研究与实践（王清路）



山东省省级教学成果奖

山东省省级教学成果奖获奖证书

获奖成果：基于临床思维与实践能力培养的基础医学课程整合的研究与实践

获 奖 者：耿磊、窦烨、韦磊、张琳琳、李帅、孙爱东、程辉、王清路、李明、高培福、张小文、宋育泽

获奖等级：二等奖

主要完成单位：齐鲁医药学院

证 书 号：GJ2018049 5

山东省省级教学成果奖评审委员会

二〇一八年一月十七日

2.3 山东省第七届“超星杯”高校青年教师教学比赛奖（王疆娜）

山东省第七届

“超星杯”高校青年教师教学比赛

获奖证书

姓　　名：王疆娜

比赛奖项：优秀奖

身份证号：610526198612170440

工作单位：山东体育学院

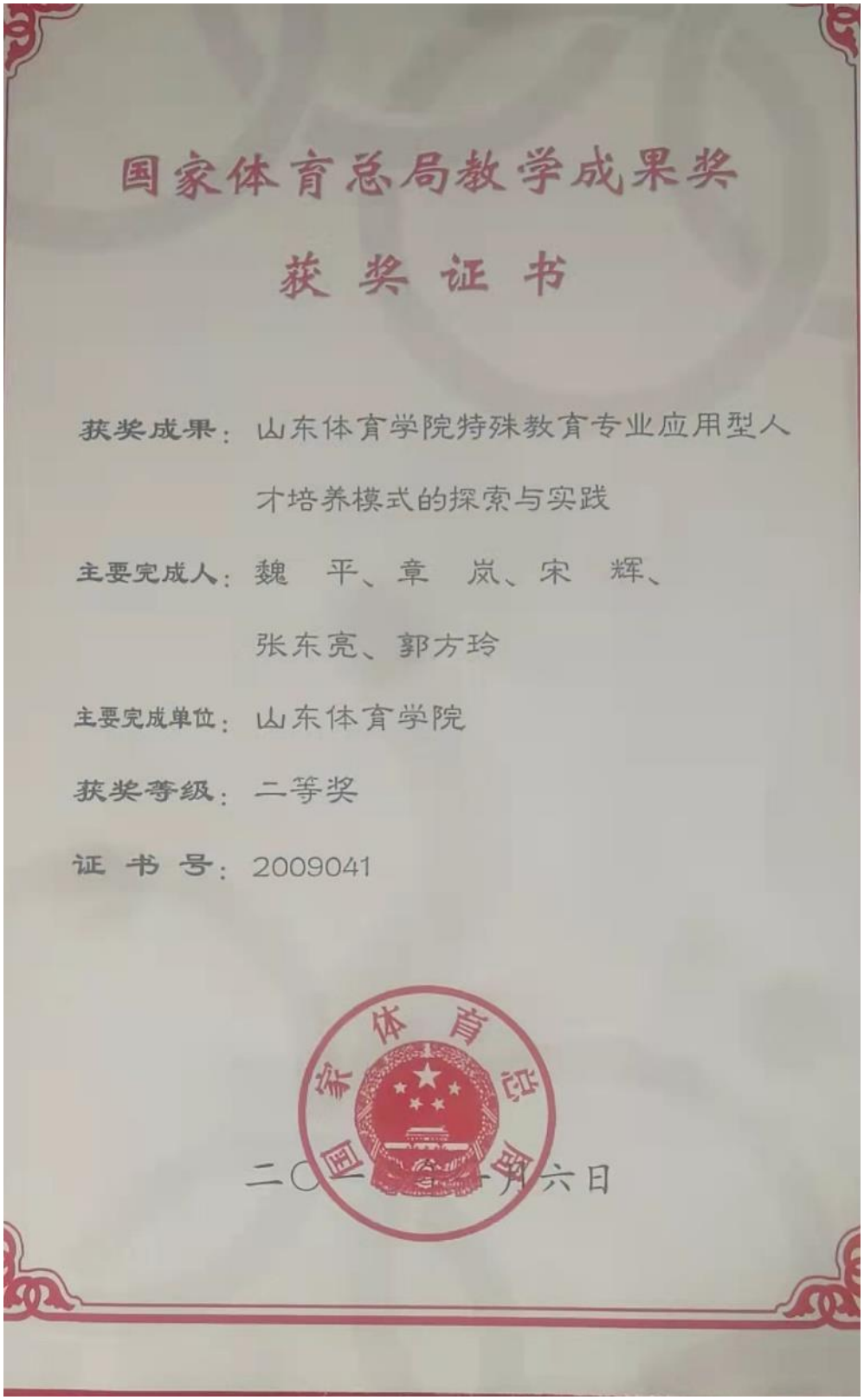
山东省教育厅

中国教育工会山东省委员会

编号：20201355

二〇二〇年八月

2.4 国家体育总局教学成果奖---山东体育学院特殊教育专业应用型人才培养模式的探索与实践（郭方玲）



国家体育总局教学成果奖

获奖证书

获奖成果：山东体育学院特殊教育专业应用型人才培养模式的探索与实践

主要完成人：魏　平、章　岚、宋　辉、张东亮、郭方玲

主要完成单位：山东体育学院

获奖等级：二等奖

证书号：2009041

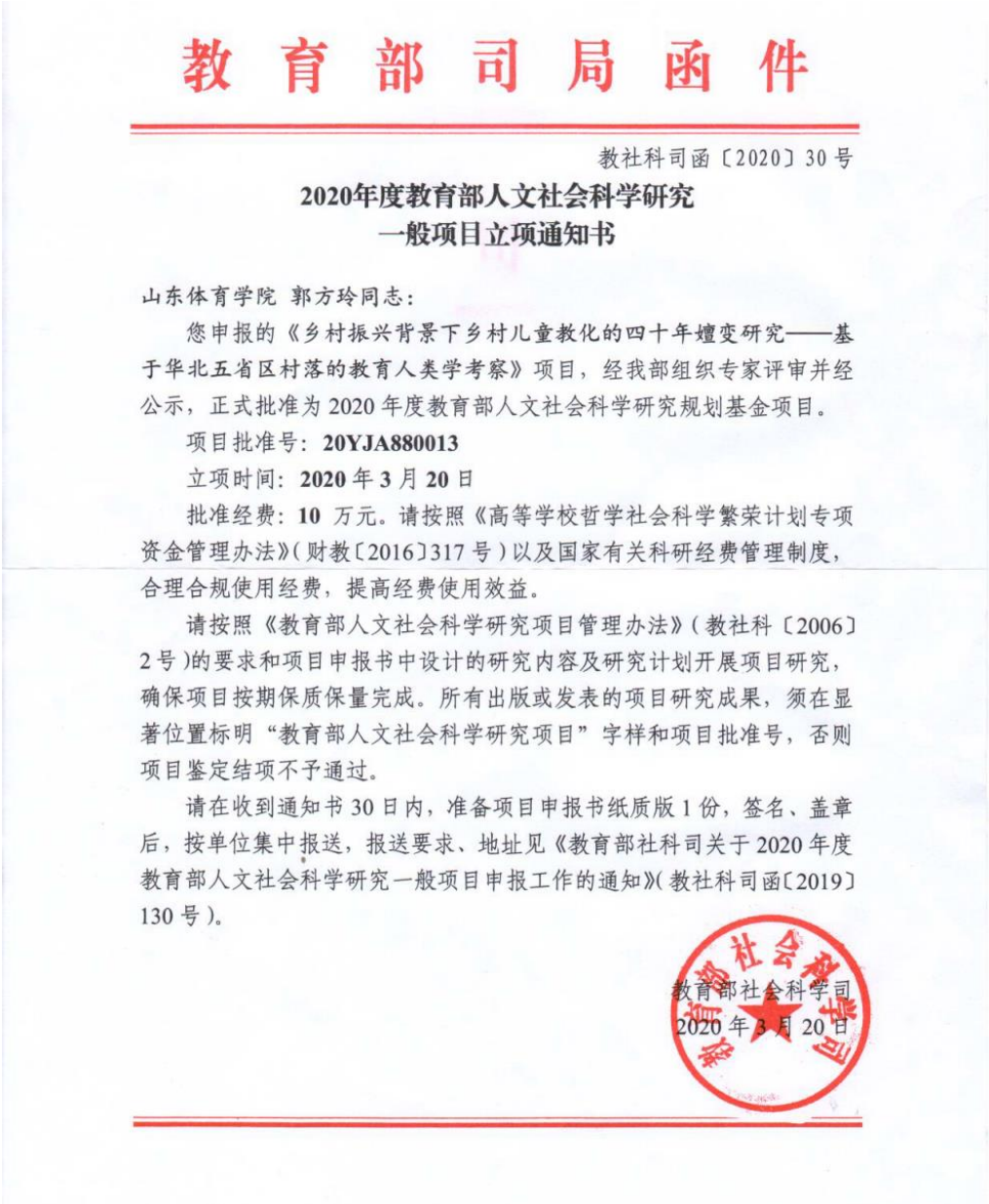
国家体育总局

二〇一〇年[illegible]月六日

3 教研成果相关教改项目、教材和论文

3.1 教学改革项目

3.1.1 乡村振兴背景下乡村儿童教化的四十年嬗变研究——基于华北五省区村落的教育人类学考察（郭方玲）



教 育 部 司 局 函 件

教社科司函〔2020〕30号

2020年度教育部人文社会科学研究
一般项目立项通知书

山东体育学院 郭方玲同志：

您申报的《乡村振兴背景下乡村儿童教化的四十年嬗变研究——基于华北五省区村落的教育人类学考察》项目，经我部组织专家评审并经公示，正式批准为2020年度教育部人文社会科学研究规划基金项目。

项目批准号：**20YJA880013**

立项时间：**2020**年**3**月**20**日

批准经费：**10**万元。请按照《高等学校哲学社会科学繁荣计划专项资金管理办法》（财教〔2016〕317号）以及国家有关科研经费管理制度，合理合规使用经费，提高经费使用效益。

请按照《教育部人文社会科学研究项目管理办法》（教社科〔2006〕2号）的要求和项目申报书中设计的研究内容及研究计划开展项目研究，确保项目按期保质保量完成。所有出版或发表的项目研究成果，须在显著位置标明“教育部人文社会科学研究项目”字样和项目批准号，否则项目鉴定结项不予通过。

请在收到通知书30日内，准备项目申报书纸质版1份，签名、盖章后，按单位集中报送，报送要求、地址见《教育部社科司关于2020年度教育部人文社会科学研究一般项目申报工作的通知》（教社科司函〔2019〕130号）。

教育部社会科学司
2020年3月20日

3.1.2 基础医学课程体系的优化与教学方法改革（王清路）

山 东 省 教 育 厅

鲁教高函〔2020〕20 号

山东省教育厅
关于公布 2020 年本科教学改革研究项目立项名单的通知

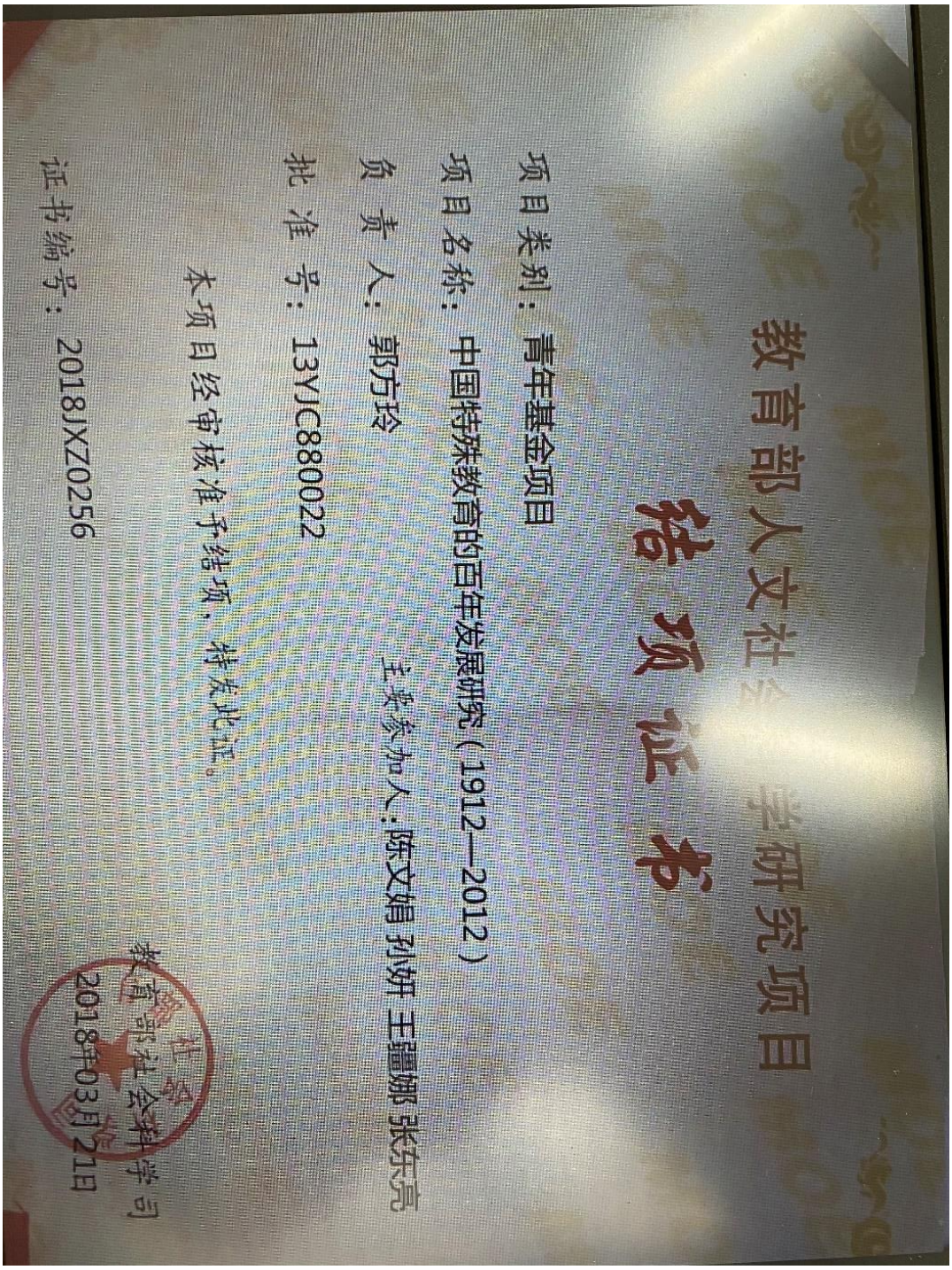
各普通本科高等学校:

根据《山东省教育厅关于做好 2020 年本科教学改革研究项目立项申报和管理工作的通知》（鲁教高函〔2020〕11 号，以下简称《通知》），我厅组织开展了 2020 年度山东省本科教学改革研究项目申报评审工作。经学校推荐、专家评审、社会公示等程序，确定立项重大专项 4 项、重大专项子课题 15 项、优秀教学成果培育项目 57 项、重点项目 91 项、面上项目 331 项。现将项目名单予以公布（详见附件），并就有关事项通知如下:

一、各高校要高度重视项目研究与实践，加强项目管理，组织项目团队认真开展项目研究，按照《通知》要求和项目研究需要提供研究经费，做好经费使用和项目进展监督检查，按时组织

项目编号	项目类型	项目名称	主持人姓名	主持单位	参与单位	团队成员
M2020103	面上项目	基础医学课程体系的优化与教学方法改革	王清路	齐鲁医药学院	无	韦磊,徐明,李采,张琳琳,李帅,程辉,王云飞,李明,张桥,张燕
M2020104	面上项目	基于“1+4”平台的地理信息科学专业创新型人才培养研究	张安定	鲁东大学	无	李国庆,孔祥生,吴孟泉,王涛,肖鲁湘,仲少云,于祥,张春华,王静璞
M2020105	面上项目	基于“金课”导向的药学专业药理学课程群案例库建设与实践	王巧云	滨州医学院	无	孙红柳,周玲,辛文妤,蒋王林,李爽
M2020106	面上项目	基于“科教融合、研创一体”多学科联合教学的乡村建设拔尖创新型人才培养模式研究	朱一荣	青岛理工大学	无	马鸿洋,申建红,李金成,董德坤,田华,王琳,赵芳超,聂彤,刘森
M2020107	面上项目	基于“课程思政、专业思政”的临床医学专业生物化学教学改革与探索	张媛英	山东第一医科大学（山东省医学科学院）	无	伊淑莹,于立娟,孙凌云,王涛,付晓艳,蒋汉明,孙贝贝,柏素云,翟静
M2020108	面上项目	基于“两性一度”标准下的《字体设计》课程教学改革与实践	李锋	齐鲁理工学院	无	张韬,宋润民,高良博,王丹,罗娟,张胜利,李要全,赵杨,姜露露
M2020109	面上项目	基于“免试认定教师资格改革”的公费师范生教育实践模式研究	韩涛	聊城大学	无	王桂清,卢军,黄春平,韩丽华,任永辉,王刚臣,马勇,王喆
M2020110	面上项目	基于“三链融合”的农业应用型人才培养实践教育体系构建研究	李敬锁	青岛农业大学	无	王玲玲,葛凤丽,温琳,杨焕玲,龚丽,王秀华,孔晨,孙瑜,辛德树
M2020111	面上项目	基于“三螺旋”理论的农业高校产学研结合人才培养机制创新研究	孙瑜	青岛农业大学	无	许秀梅,杨焕玲,张怡,郭燕茹,葛凤丽,黄宗堂,孟璐瑶,郭海红,王玲玲
M2020112	面上项目	基于“四个结合”的《植物学》课程思政教学体系的探索与实践	赵金辉	潍坊理工学院	无	杜晓映,赵玉翠,邱佳佳,刘燕燕,王新
M2020113	面上项目	基于“五育融合”的高校美育课程体系构建及实施路径研究	毛迎新	齐鲁师范学院	无	王岩,窦曼莉,谢文峰,万梅红,宋振军,桑潇,高杨
M2020114	面上项目	基于“项目驱动+翻转课堂”的微生物学 “三主式”混合教学模式构建与应用研究	李艳玲	山东第一医科大学（山东省医学科学院）	无	尹瑞法,田园,常正尧,郝岗平,史仁玖,潘国军,殷晓蕾,孟超,赵静
M2020115	面上项目	基于“虚拟仿真+互联网”技术的工程管理专业实验教学改革研究与实践	李晓冬	青岛理工大学	无	王志强,姜吉坤,申建红,陈传联,曹丙霞,崔庆宏,周凯

3.1.3 中国特殊教育的百年发展研究（郭方玲、王疆娜）



教育部人文社会科学研究项目

结项证书

项目类别：青年基金项目

项目名称：中国特殊教育的百年发展研究（1912—2012）

负责人：郭方玲　　主要参加人：陈文娟 孙妍 王疆娜 张东亮

批准号：13YJC880022

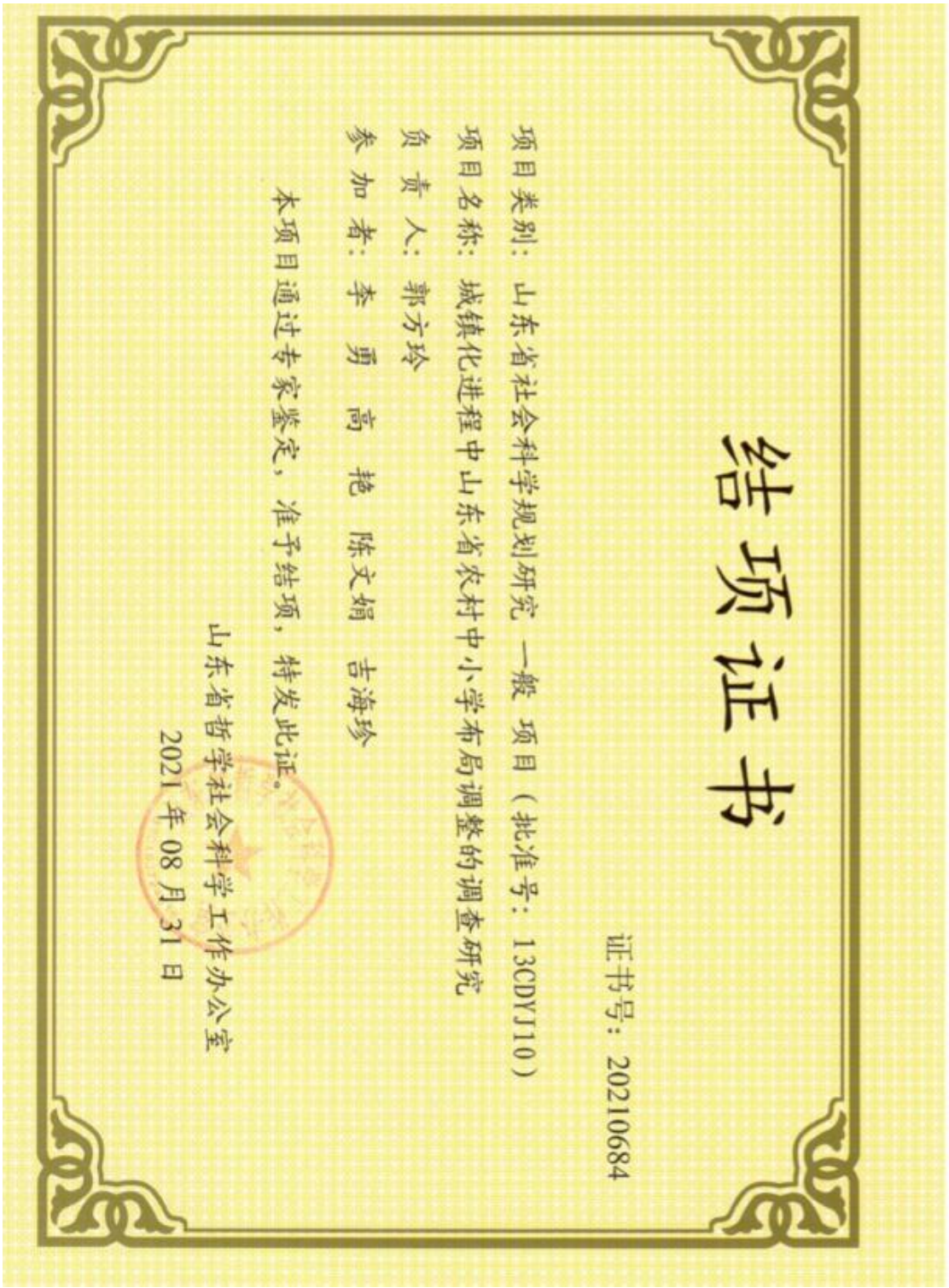
本项目经审核准予结项，特发此证。

教育部社会科学司

2018年03月21日

证书编号：2018JXZ0256

3.1.4 城镇化进程过程中山东省农村中小学布局调整的调查研究（郭方玲）



结项证书

证书号：20210684

项目类别：山东省社会科学规划研究 一般 项目（批准号：13CDYJ10）

项目名称：城镇化进程中山东省农村中小学布局调整的调查研究

负 责 人：郭方玲

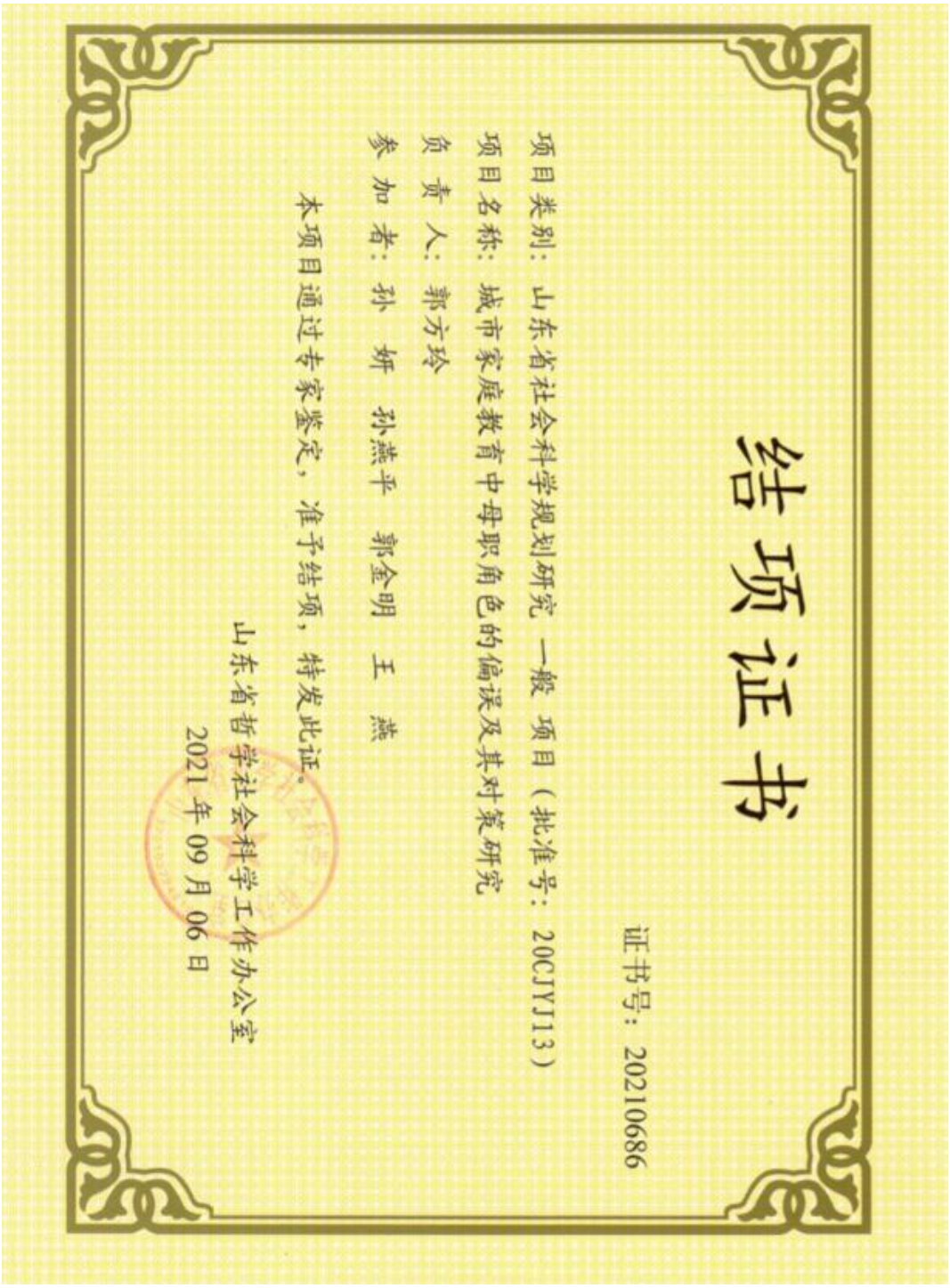
参 加 者：李 勇 高 艳 陈文娟 吉海珍

本项目通过专家鉴定，准予结项，特发此证。

山东省哲学社会科学工作办公室

2021 年 08 月 31 日

3.1.5 城市家庭教育中母职角色的偏误及其对策研究（郭方玲）



结项证书

证书号：20210686

项目类别：山东省社会科学规划研究 一般 项目（批准号：20CJYJ13）

项目名称：城市家庭教育中母职角色的偏误及其对策研究

负 责 人：郭方玲

参 加 者：孙 妍 孙燕平 郭金明 王 燕

本项目通过专家鉴定，准予结项，特发此证。

山东省哲学社会科学工作办公室

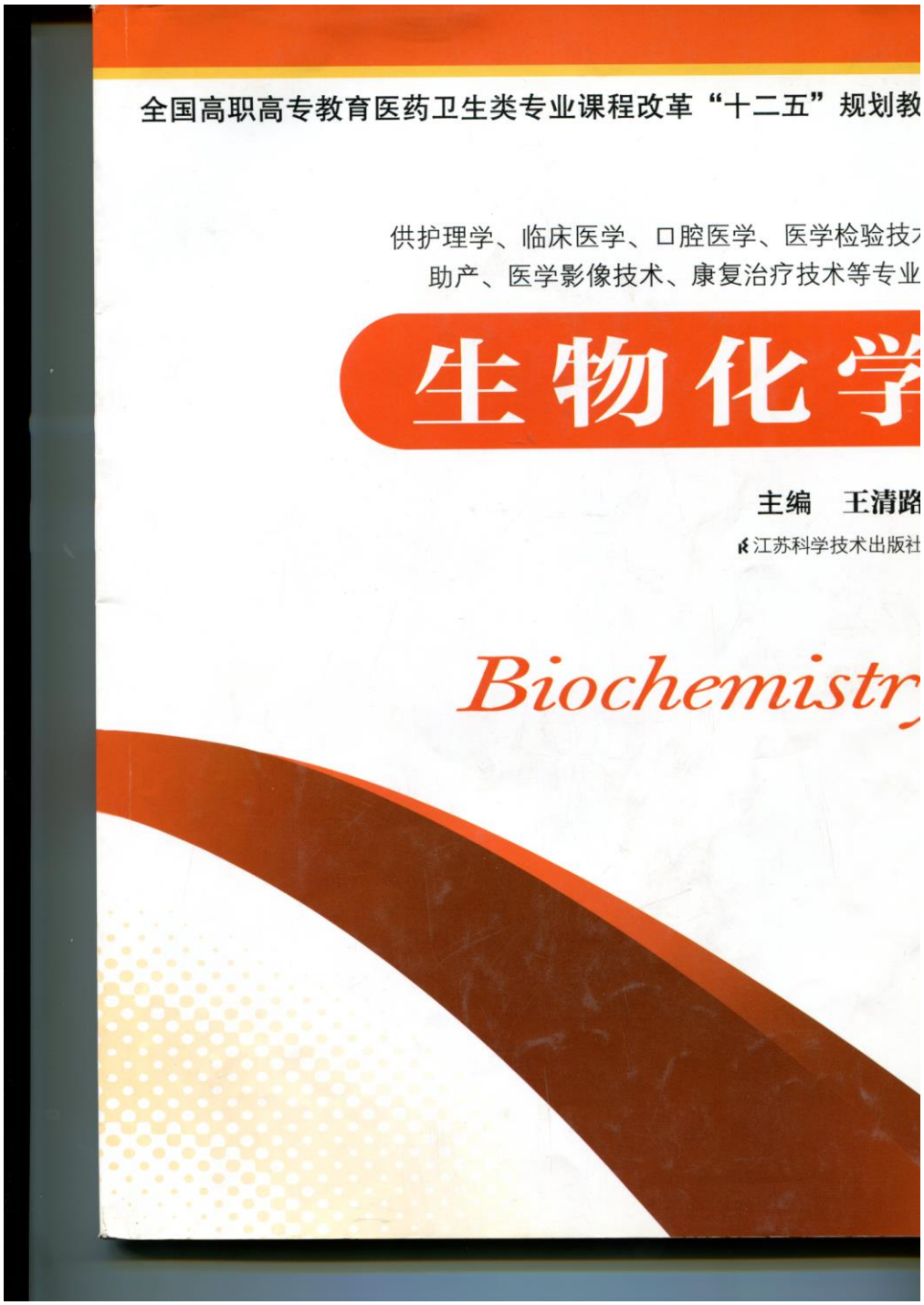
2021 年 09 月 06 日

3.2 主编教材

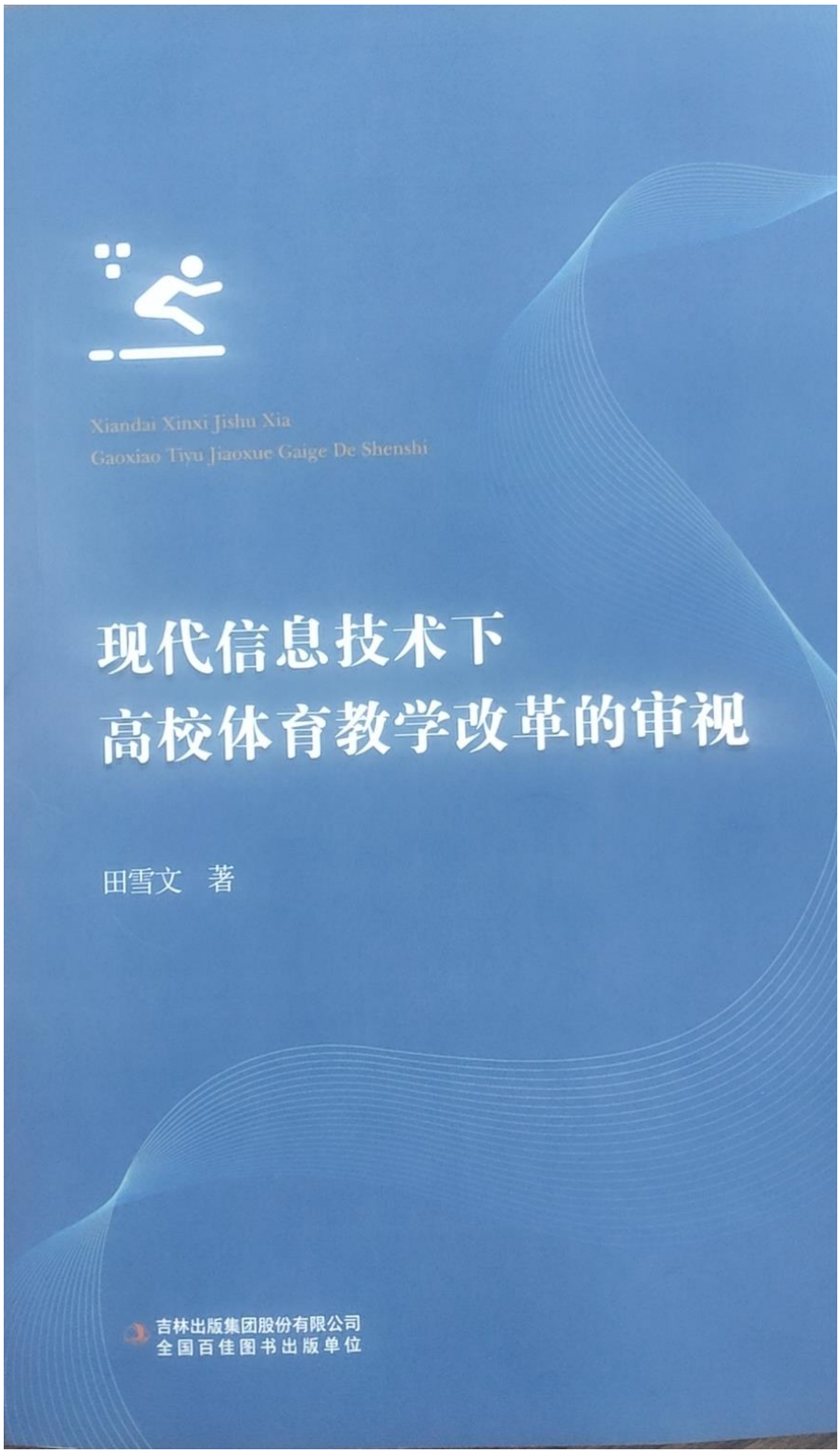
3.2.1 《生物化学》高等学校“十三五”创新规划教材（王清路，人民卫生出版社）

3.2.2 《生物化学》全国高等学校创新教材（王清路，人民卫生出版社）

3.2.3 《生物化学》全国高职高专教育医药卫生类专业课程改革“十二五”规划教材（王清路，江苏科学技术出版社）



3.2.3 《现代信息技术下高校体育教学改革的审视》（田雪文，吉林出版集团股份有限公司）



3.3 教学改革论文

3.3.1 Evaluation of Basic Medical Curriculum Integration Based on the Training of Chinese Excellent Category Doctors（王清路）

Medical
Communication

BBRC
Bioscience Biotechnology
Research Communications

Biosc.Biotech.Res.Comm. Vol 14 No (1) Jan-Feb-March 2021 Pp 08-11

Evaluation of Basic Medical Curriculum Integration Based on the Training of Chinese Excellent Category Doctors

Lei Wei, Linlin Zhang Cai Li, Ming Xu, Shuai Li, Yunfei Wang,
Rongjun Fan, Ming Li, Hui Cheng and Qinglu Wang*
DDepartment of Basic Medical Education, Qilu Medical University, Zibo, 255213, China

ABSTRACT

The traditional discipline-centered teaching mode is no longer adapt to the changes of current medical modes and the social demand for medical services owing to disjointed from clinical practice, knowledge separation and lack of connection between basic medical courses and clinical medical courses. Therefore, China has implemented the National Excellent Doctor Training plan and vigorously promoted the integrated reform of clinical medicine curriculum. In this study we compared the change of teaching effect between reform class and control class. We randomly selected 1 class from the 5-year clinical medicine major to carry out a series of pilot teaching reforms with curriculum integration, and at the same time, 1 class was selected for parallel control. Then the effect of the reform was evaluated from the aspects of test scores and 6-STATION OSCE. Student achievement and clinical skills are effectively improved through the integration of basic medical courses. The results showed that student achievement and clinical skills are effectively improved through the integration of basic medical courses It is concluded that we further need to integrate the various foundational and clinical disciplines into an organ-system based curriculum for the National Excellent Doctor Training plan.

KEY WORDS: CURRICULUM INTEGRATION, BASIC MEDICAL CURRICULUM, MEDICAL EDUCATION, DISCIPLINE-CENTERED TEACHING MODE.

INTRODUCTION

Under Flexner's influence, medical curricula around the world came to be structured into: Preclinical medicine learned in lecture theatres, laboratories, dissecting rooms, libraries and Clinical medicine learned in wards and operating theatres of teaching hospitals. Since the 1950s, medical colleges in Europe and The United States have proposed and implemented the teaching reform featuring the integration of medical curriculum. Curriculum integration involves the organization of teaching to interrelate or unify subjects frequently taught in separate academic courses or departments (Harden, et al. 1984, Scheffer, et al. 2012 Seethe and Khan 2019).

Most of the medical colleges in China are following the traditional system that is teacher centered, discipline based and opportunistic. With the development of global medical education and interdisciplinary integration, the model of Chinese medical education has also changed in the past decade. There were some defects in the traditional medical education pattern such as overlapping content of teaching, more time span, students learning burden, and comprehensive ability between various disciplines. Integration is an important means of dealing with overload of information, fragmented teaching of basic and clinical sciences, and the need for relevant and meaningful learning (Yamani and Rahimi 2016).

In this study, we have analyzed the problems in the process of integrated medical foundation course, and then really

Article Information:
*Corresponding Author: wql_zcq@126.com
Received 10/12/2020 Accepted after revision 23/03/2021
P-ISSN: 0974-6455 E-ISSN: 2321-4007
Thomson Reuters ISI Clarivate Analytics
Web of Science ESCI Indexed Journal

Identifiers & Pagination:
Vol 14(1) E-Pub 31st Mar 2021 Pp- 08-11

Published by Society for Science & Nature India
DOI: http://dx.doi.org/10.21786/bbrc/14.1/2

broke the barriers between disciplines, and integrated the systematic anatomy, histology and embryology, physiology, pathology, pathophysiology, pharmacology to "Two Introduction and Multiple Systems". We put forward the training of clinical ability, which is helpful to realize the training of students' clinical in the whole process of medical education. It is of great significance to realize the educational goal of "early clinical, multiple clinical and repeated clinical" and improve the training quality of medical talents.

MATERIAL AND METHODS

This study was granted an exemption by our Institutional Review Board since it evaluated outcomes of an existing mandatory component of the curriculum. We made an analysis the current research about the integration of basic medical courses based on published literature, and then carry out empirical research on medical colleges and universities that have integrated their courses, so as to provide scientific theoretical guidance and reference for the subject research.

Description of course integration: Six basic medical courses including *systematic anatomy, histology* and *embryology, physiology, pathology, pathophysiology* and *pharmacology* were selected as the subjects of course integration. The knowledge content of subject was decomposed into "knowledge points" to form "granulated resources". Guided by the cultivation of clinical thinking ability, the "granulated resources" will be centered on "organ-system", and the systematic teaching content will be reintegrated and reconstructed to form a modular curriculum system of "Two Introduction and Multiple Systems". Two Introduction is an introduction to human body form and function and the Multiple Systems are the respiratory system, circulatory system, digestive system, urinary system, blood system, nerve system, endocrine system, sensory system and reproductive system. According to the relevant requirements and procedures of the curriculum standards, the curriculum standards of the integrated curriculum were formulated. Then we organized the research team to compile "Introduction to human body morphology and mechanics" and 9 "organ-system" modules as case textbooks.

Teaching implementation of basic medical curriculum integration: In the five-year clinical medicine class of 2015, a 36-person pilot class for teaching reform was established based on the principle of mutual selection between the two sides, and other classes of the clinical medicine undergraduate course were taken as the control group. The pilot class of educational reform was organized for teaching according to the integrated teaching contents of basic medical courses, while the control group was taught according to the current talent training program. The teaching reform pilot class was taught based on integration of "teaching of early clinical probation" and the integrated course of basic medicine to form the idea of early clinical probation.

The process of "setting questions, seeking answer and basic clinical combination" was designed and directly applied to teaching. The teaching mode combining case-based teaching and PBL teaching was adopted to carry out teaching based on suitable early internship cases and combined with PBL or CBL teaching. The goal of the implementation of special teaching was to cultivate students' clinical thinking ability and improve students' ability to solve practical problems. The teaching methods of the control group were carried out according to the discipline-centered methods.

Evaluation of teaching effect: The teaching reform pilot class adopts the method of formative evaluation, summative evaluation and comprehensive evaluation. In the teaching process, the formative evaluation was highlighted, and the existing problems in the learning process are fed back to the students in time. The formative evaluation runs through the whole teaching activity. After the teaching activity of each course is finished, the summative evaluation based on comprehensive and case questions was adopted. After the learning of all the integrated courses of basic medicine, the comprehensive evaluation of learning effect was carried out by means of the basic stage assessment of simulated clinical practitioners. In addition to formative evaluation, summative evaluation and comprehensive evaluation were carried out in both the teaching reform class and the control class. The summative evaluation and assessment contents are generally consistent, but the teaching reform class was assessed according to the integrated curriculum, and the control class was assessed according to the unintegrated curriculum. Comprehensive evaluations were conducted in the same manner.

OSCE setting: A comprehensive 6-station OSCE was administered to the teaching reform and control class of five-year clinical medicine class of 2015. The examination was conducted after the clinical practice. The assessment of clinical skills includes the following aspects: Patient care skills, Interpersonal and communication skills, Professionalism skills, Practice-based learning and improvement skills, Systems-based practice skills and Medical knowledge skills; The OSCE consisted of 6 clinical problems; each clinical problem consists of six core competencies defined by the Accreditation council for Graduate Medical Education (ACGME) (Yang, et al. 2011).

Standardized patients should be used as a reference in the specific assessment. At each station, the summary scores were the sum of all the checklist items, and the six core competency sub scores were the sum of specific items for each competency. When presented, all scores were translated into 100 percentages. Please refer to the article of Yang et al. for more details (Yang, et al 2011).

Statistical analysis: All data were processed by SPSS 18.0 (SPSS Inc., Chicago, IL, USA). All data were presented as mean ± standard deviation. Comparison between groups was conducted using single-factor ANOVA followed by

Tukey's test. P<0.05 indicated significance, and P<0.01 indicated extreme significance.

RESULTS AND DISCUSSION

We have a teaching reform pilot class size of 36 students per year and control class size of 47 students. The students of teaching reform pilot class studied "Two Introduction and Multiple Systems" and the students of control class studied Six basic medical courses. Biochemistry and molecular biology as a comparative analysis course are taught in every class (see Table 1).

Average score (75.25) not including Biochemistry and molecular biology in teaching reform pilot class was higher than that (71.63) in control class. And however, *Biochemistry* and molecular biology was lower than that in control class. This results showed our teaching reform increased students' score. Next, the score of teaching reform pilot class and control class was further analyzed based on *Biochemistry* and *molecular biology* as a comparative analysis course are taught by same teacher. The calculating method is Relative performance= (Teaching reform subject $_{\text{Average score}}$/ *Biochemistry* and *molecular biology* score) / (Traditional subject $_{\text{Average score}}$/ *Biochemistry* and *molecular biology* score). The results from Figure 1 showed that relative performance in teaching reform pilot class was higher than that in control class.

Analysis of OSCE: In Figure 2, a significant difference in the performance between different aspects of core competency ($p < 0.05$) was noted. Teaching reform pilot class had the higher pass rate in the aspect of Practice-based learning and improvement skills (83%), Systems-based practice skills (75%) and Medical knowledge skills (91%), whereas the lower pass rate was noted in the aspect of professionalism (52%). Interestingly, teaching reform pilot class had the higher pass rate in the OSCE than that of control class.

Table 1. Average score in teaching reform pilot class and control class

Class	Students number	Course	Semester	Average score
Teaching reform pilot class	36	Introduction to Human Morphology	1,2	71.3
		Introduction to human mechanics	2	67.47
		Respiratory system	3	77
		Digestive system	3, 4	81.72
		Circulatory system	3	68.25
		Blood system	4	76.72
		Urinary system	4	79.94
		Sensory system	4	79.83
		Nerve system	5	73.19
		Endocrine system	5	75.25
		Reproductive system	5	77.11
		Biochemistry and molecular biology	2	71.8
Control Class	47	Systematic anatomy	2	51.98
		Histology and embryology	3	76.36
		Physiology	3	73.09
		Pathology	5	78.47
		Pathophysiology	5	72.3
		Pharmacology	5	77.55
		Biochemistry and molecular biology	4	77.91

In 1989, Shoemaker proposed a concept about integrated curriculum that is "Education that is organized in such a way that it cuts across subject matter lines, bringing together various aspects of the curriculum into meaningful association to focus upon broad areas of study (Betty 1989)." To this day, there is an ongoing discussion about whether medical curriculum should be discipline based or integrated. Abraham Flexner thought that students should first learn basic and biomedical sciences and then move to clinical sciences; however, a common criticism of this approach was that students would not see the relevance of basic and biomedical sciences applied to clinical practice, and it was preferable to encourage students to think as doctors from the day they enter medical school (Harden 1986).

Integration of medical curriculum was importance for medical education because basic science learning was placed in the context of clinical and professional practice and was considered by students to be more meaningful and relevant (Quintero, et al. 2016). After a discussion of the health-illness concept, we constructed a theoretical basis of this process that changed our traditional discipline-based learning perspective. The meaning of the health-illness process changed was defined as a social, cultural, biological, and psychological process

embedded and determined socially and culturally by group of human beings (Fanwei, et al. 2019).

Figure 1: Relative performance in teaching reform pilot class and control class

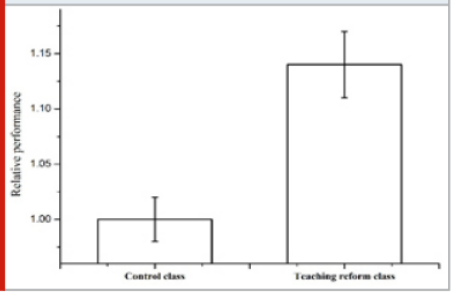
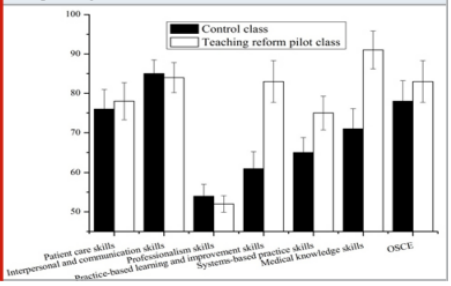


Figure 2: Overall pass rate (%) of each domain of ACGME competency and overall OSCE



As this approach implies that society and culture are no longer simply risk or etiological factors, our medical curriculum had to evolve into a new structure based on a "Two Introduction and Multiple Systems" concept of health and illness.In this study we randomly selected 1 class from the 5-year clinical medicine major to carry out a series of pilot teaching reforms with curriculum integration, and at the same time, 1 class was selected for parallel control. Then the effect of the reform was evaluated from the aspects of test scores and 6-STATION OSCE. The results showed that student achievement and clinical skills are effectively improved through the integration of basic medical courses.

CONCLUSION

The results demonstrated that student achievement and clinical skills are effectively improved through the integration of basic medical courses. We need further to integrate the various foundational and clinical disciplines into an organ-system based curriculum for a better National Excellent Doctor Training plan (China).

ACKNOWLEDGEMENTS

This work was supported by Research project on teaching reform of undergraduate colleges and universities in Shandong Province (Z2016M067, M2020103, M2018X0100).

Conflict of Interest: None. All authors contributed equally to this work.

REFERENCES

Anthroposophic Medicine (ICURAM). Patient Educ Couns.89(3): 447-454.

Betty S. (1989) Integrative Education: A Curriculum for the Twenty-First Century. OSSC Bulletin.33(n2.

Fanwei Q.U., Jin H.E., Hua M.A., Yanling J., Wenlan Z., et al. (2019) A Comparative Analysis of Medical Education Models and Curriculums of A Medical University and A Medical Education Center. JNMA J Nepal Med Assoc.57(215): 45-49.

Harden R.M., Sowden S., Dunn W.R. (1984) Educational strategies in curriculum development: the SPICES model. Med Educ.18(4): 284-297.

Harden R.M. (1986) Approaches to curriculum planning. Med Educ.20(5): 458-466.

Quintero G.A., Vergel J., Arredondo M., Ariza M.C., Gomez P., et al. (2016) Integrated Medical Curriculum: Advantages and Disadvantages. J Med Educ Curric Dev.3(

Sethi A. and Khan R.A. (2019) Curriculum integration: From Ladder to Ludo. Med Teach.1-3.

Scheffer C., Tauschel D., Neumann M., Lutz G., Cysarz D., et al. (2012) Integrative medical education: educational strategies and preliminary evaluation of the Integrated Curriculum for Medical Education. 65 No 3 670-678

Yamani N. and Rahimi M. (2016) The Core Curriculum and Integration in Medical Education.

Yang Y.Y., Lee F.Y., Hsu H.C., Huang C.C., Chen J.W., et al. (2011) A core competence-based objective structured clinical examination (OSCE) in evaluation of clinical performance of postgraduate year-1 (PGY(1)) residents. J Chin Med Assoc.74(5): 198-204.

3.3.2 协同创新理论视角下体医融合发展路径的模式构建（田雪文）

协同创新理论视角下体医融合发展路径的模式构建

魏夫超[1]，张传光[2]，孙志远[3]，田雪文[3]

（1.山东省体育科学研究中心，山东 济南 250102；2.山东省举重摔跤柔道运动管理中心，山东 济南 250102；

3. 山东体育学院运动与健康学院，山东 济南 250102）

摘 要：借助“协同创新”理论模型架构和美国“运动是良医”全球健康倡议对我国体医融合的启示，构建了基于体医融合的“医疗——社区——个人”三联动的健康促进模式：体医融合创新平台作为保障的协作平台，分别在三个层面实施干预力量，协同创新系统的构建为体医融合提供了源源不断的创新动力，保障了体医融合科学有序的运转。临床决策支持系统实现了体医融合的首要环节；社区科学健身指导系统为社区居民的健身提供了科学实践基础；个人模块的主动健康技术的应用保障了锻炼的健身效果。一个闭环的体医融合路径已经形成，为体医融合从理论走向实践提供了新思维和新模式，推动了体医融合新业态的发展。

关键词：协同创新；运动是良医；体医融合；发展路径

中图分类号：G80-05；G804.5 **文献标识码**：A **文章编号**：1007-6204(2021)06-0013-06

DOI:10.13940/j.cnki.lntykj.2021.06.003

A model construction of sports and medicine integrated development path from perspective of the collaborative innovation theory

WEI Fuchao[1], ZHANG Chuanguang[2], SUN Zhiyuan[3], TIAN Xuewen[3]

（1.Shandong Sport Research Center, Jinan 250102, Shandong , China; 2 Shandong Weightlifting, Wrestling and Judo Sports Management Center, Jinan 250102, Shandong , China; 3. Sports and Health Science School of Shandong Sports University, Jinan 250102, Shandong, China）

Abstract: Based on the theoretical framework of “collaborative innovation” and the inspiration of the American Global health initiative “Exercise is good Doctor” to the integration of physical medicine in China, this paper construct a three-linkage health promotion model of “medical-community-individual” based on the integration of physical medicine: As a guarantee collaboration platform, the sports medicine integration innovation platform ensures the orderly operation of the science of sports medicine integration; Clinical decision support system realizes the primary link of physical and medical integration; The community scientific fitness guidance system provides a scientific practice basis for the fitness of community residents, and the active health technology of individual module ensures the fitness effect of home exercise. A closed-loop path for the integration of sports and medicine has been formed, which provides new thinking and new models for the integration of sports and medicine from theory to practice, and promotes the development of new forms of integration of sports and medicine.

Key words: collaborative innovation; exercise is good medicine; physical and medical integration; development path

随着人口红利消失殆尽、结构上趋于老龄化、 资源上供应短缺、就诊人数居高不下等问题的出

收稿日期：2021-05-12；**修回日期**：2021-10-27

基金项目：山东省重点研发计划（重大科技创新工程）“科学健身智能芯片和云服务平台的技术研究与集成应用”（项目编号：2020CXGC010902）；山东省研究生教育创新计划项目（项目编号：SDYY09075）。

作者简介：魏夫超（1986-），男，助理研究员，研究方向：体能康复训练。

通讯作者：张传光（1978-），男，副主任医师，研究方向：运动医学。

现[1]，加剧了国家和个人的医疗卫生负担，影响了经济和社会的健康发展[2]。解决这一问题必须从根源着手，转变思路，要从“医疗健康干预”向“非医疗健康干预”转变，体育锻炼的“防”，医疗卫生的“治”，体育和医学共同担负起提高人类健康素养的重任，体育健身侧重前端，医疗卫生侧重后端。因此，体医跨界融合是目前最有效的方法。基于此，政府通过出台一系列宏观政策来调控体医融合发展，其中包括《关于加快发展体育产业促进体育消费的若干意见》、《“健康中国2030”规划纲要》、《中国防治慢性病中长期规划（2017~2025年）》和《全民健康生活方式行动（2017~2025年）》等政策，从顶层设计层面为体医融合发展提供政策支持。本文基于协同创新理论，借鉴美国的运动是良医（Exercise is Medicine）全球倡议，为我国体医融合路径的构建提供参考。

1　运动是良医全球健康倡议解析

美国在“体医融合”运动促进健康服务体系建设方面很早就进行了探索与实践[3]。运动是良医（EIM）倡议是在2007年推出的，目的是推动实施基于体力活动水平证据的体医融合战略，以提高体力活动（Physical Activity，PA）在医疗保健中的地位。广泛实施体力活动咨询和转诊制度，作为护理的临床实践标准，通过补充和利用其他途径改善人口层面的体力活动水平，并有助于实现减少无活动和相关发病率及死亡率的全球目标[4]。该倡议由美国医学协会和美国运动医学学院（American College of Sports Medicine）制定，随后由ACSM协调，其总体目标是使体力活动水平成为在医疗系统中预防和治疗非传染性疾病的标准部分[5]。具体来说，呼吁所有医疗保健工作人员将PA视为每次为病人诊疗的重要依据[6-7]，并为患者提供有效的咨询和转诊，以满足他们的体育锻炼和健康需求。EIM倡议的核心原则：凭借有效的干预措施（包括医生的“绿色处方”和“运动转诊方法”），以医疗保健和社区为基础促进体力活动。EIM国家工作队（National Task Forces）由来自该国国家初级保健协会、其他医疗和保健协会（心脏学、内分泌逻辑学、体育医学、护理、营养/营养学和物理治疗）、体育、学术机构以及在可能的情况下卫生部或国家公共卫生研究所的代表组成。将体育活动生命体征（Physical Activity Vital Sign）纳入电子病历（Electronic Medical Records），发展社区网络PA方案和资源，并建立一个将两者联系起来的临床决策支持系统，便于医疗保健人员能够提供PA咨询，并将患者转介到社区履行其PA“处方”。EIM解决方案由五个步骤组成，这些步骤发生在三个模块中，通过使用健康保健技术和医疗决策支持系统，整合临床和社区资源以促进体力活动。

2　现行体医融合模式

目前，体医融合的研究主要有基于产业融合理论的产业融合模式和“三融合”理念模式、共生理论视角下的体医融合模式、双元创新融合模式以及特色地方实践模式。

2.1　产业融合模式

产业融合模式的过程是多维度的，整个融合过程需要经过三个阶段：技术融合（基础）、业务融合（核心）和市场融合（结果）[8]。虽然该模式促进了体医融合在试点地区的技术、资源、话语权的融合，推动体育与医疗的整体融合。但是，只是对技术、业务和市场三个维度的资源进行汇总集合，并没有发挥各维度资源的最大效能，也没有建立相应的公共服务体系，无法在重大突发公共卫生事件中发挥联动作用。“三融合”理念模式是将体育科学的医学理念、临床检测的测试技术和医学治疗的运动方法进行融合，这一模式在产业融合模式的基础上增加了理念的融合，只完成了对产业融合模式理念上的升级[9]。

2.2　共生融合模式

共生理论背景下的体医融合模式，均衡了体育和医疗卫生的利益分配，促进了技术、资源、话语权的融合，推动体育与医疗的整体融合[10]。通过理念、部门、人才、技术、策略融合共生等构建体医融合共生发展路径，以促进体育业和医疗服务业深度融合和共生发展[11]。虽然建立了互补、互通和互融阶段标准，但未从本质上解决两个系统行动一致和匹配度。

2.3　双元创新融合发展模式

双元创新融合发展模式在发展认知、整合空间和价值共鸣三个方面对体医融合的模式进行了创新型的研究，重新调整角色定位，明确身份边界关系，建构民众话语体系，形成优势互补机制、单元联动全局和塑造健康代言人[12]。促进了政府、社会及民众的参与实践基础，双元创新协同发展的渐进性为逐步实现出传统结合型向创新融合型的过渡转

变提供了理论基础。该模式未明确提出创新变革的基础构建，只揭示其具有优越的外显形式，缺乏一定的实践基础，导致实践效果并不明确。

2.4　地方实践模式

地方实践模式在实践层面上做了一些积极尝试，横向对比分析发现各地区所出现特有的“体医结合”发展模式表明：外在显著形式加快了“体医融合”模式表达和实现。如苏州“阳光健身卡”政策、上海“1+1+2”社区主动健康工程以及济南市全民健身指导中心，地方实践模式更加突出了地方实践的特色，模式设置灵活，对小范围和单一地域的体医融合发挥着不可替代的作用[13]。

就目前来看，体医融合模式尚未形成固定化的内容和形式，还没有一套满足于不同条件和环境的融合模式，体育和医疗卫生机构既要在原有的资源基础上进行融合，又要满足于市场不断变化，构建体育和医疗卫生部门能够高匹配和精诚合作的服务模式和服务体系是推进体医融合实现的关键所在。

3　协同创新理论解析及应用

协同创新通过将主体要素进行系统优化、合作创新，从整合与互动两个维度来分析（见图1）。整合维度包括：知识、资源、行动、绩效，互动维度指各个创新主体要素之间的互惠知识分享、资源优化配置、行动的最优同步和系统的匹配度。而根据两个维度上的不同位置，协同创新是一个沟通——协调——合作——协同的过程，是各个创新要素的整合以及创新资源在系统内的无障碍流动。协同创新具有整体性和动态性两大优势，各种要素是有机聚合而不是简单相加，在功能和结构上表现出统一的整体性，并且处于不断动态变化之中[14]。协同创新理论模型应用于体医融合主要体现在知识整合、资源整合、行动整合和绩效整合4个方面：

3.1　知识整合

理念融合，以健康为中心理念融合是体医融合模式革新的理念契合点。制度融合，即建立先进顺畅的融合制度，在体育和医疗之间的内部关系层面进行制度整合，实现内部有序融合和良性运转。规划融合，战略目标一致性，紧紧围绕全民健康这一条主线，全民健身规划与卫生健康规划统筹推进医疗与健康工作方针。话语权融合，因受到竞技体育体制的影响，人们对体育所具备的健康促进意识淡薄，国民体质健康素养比较薄弱，改变“重医轻体”被动健康观念，推动和加快主动健康技术的应用与推广，提高体育在医养健康中的作用。

3.2　资源整合

体质检测仪器与医学检测仪器的融合，研发新型体质医疗检测应用仪器。人才融合，推动体医融合复合型人才的培养，医生和社会体育指导员优势互补，加大对运动处方师的培训。评价指标融合，将现有的体质健康监测指标与医学指标融合。诊断疗法融合，通过整合体育科技手段和现代临床医学诊疗技术，将运动作为一种干预手段推广运用到疾病治疗及健康促进中。

3.3　行动整合

工作机制融合，加快建设健康促进协同机制，打破体医部门闭塞的壁垒，使其由条状化管理变为整合式服务。健康应用研究，建立运动处方库和康复处方库，丰富运动处方的类型，促进体育康复技术的研发与推广。临床——社区融合，其目的是在这两个关键领域开展非医疗干预手段。将体育活动生命体征（体力活动水平）纳入电子病历，发展国家支持共建的社区体力活动方案和资源网络。服务模式融合，延申体育和医疗卫生机构服务至社区和家庭。

3.4　绩效整合

研究创新体医融合新模式、新业态，建立协同创新科技平台，打造临床——社区融合网络系统，从被动治疗发展为完整的健康促进服务产业链条。目前来看，体医双方的整合已经具备了良好的理论和实践基础，在此基础上，加上协同创新原有的互动维度，使现有的知识可以互惠共享，资源得到最优化配置，融合后的行动可以达到最优同步以及两个系统可以建立良好匹配。互动维度良好运行的关键所在为协同创新平台的搭建。

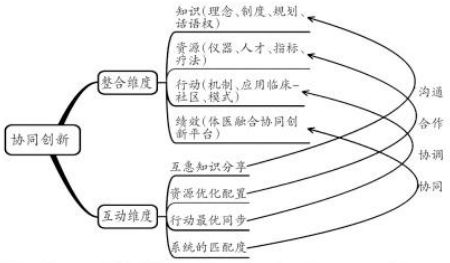


图1　基于协同创新理论体医融合维度

4　体医融合发展路径构想

新时代我国体医融合的发展逻辑是在健康需求的背景下，政府部门通过宏观调控引导体育和医疗部门协同联动指导全民健身，全民健身的顺利开展又在一定程度上满足了人们的健康需求[12]。医疗产业与体育产业同属服务产业，产业特征相似，并且医疗产业和体育产业的最终目的是为提高人们的健康素养而服务，在追求利润的同时，实现人们追求健康生活的需求。因此，在产业价值链扩展、消费需求和科技创新等因素的推动下，医疗产业为体育产业提供医疗服务市场，体育产业为医疗产业提供体育健身元素。体育产业和医疗产业都处于良好的政策环境，其市场前景十分广阔，两个产业相互渗透、相互交叉，最终融合为一体，逐步形成新产业的动态发展过程。

4.1　协同创新系统的构建

协同创新系统的构建有利于体医融合在整合和互动两个维度之间开展工作，保证了体医融合的服务和产品可以经过科学系统的生成和持续更新用以适应市场需求变化。该系统共分为三个周期：创始创新、显著创新和增量创新（见图2）。创始创新（周期A）始于阶段（1）识别和选择，其中战略创新目标是基于各种知识来源或基础研究而确定和生成的。新的想法往往产生于设计和开发团队、患者、融合组织和研究机构。然后，设计和开发团队构建想法，通常由患者和融合组织支持，评估和选择最可行的，而无用的被取消。此评估过程称为可行性分析。随后，选定的想法被传递到第二阶段（2）概念设计和项目规划。将选择的想法与知识和参与者的经验相结合，新的概念被创造出来。新创造概念的验证是以分析潜在的新概念的形式出现的，考虑其技术可行性和市场机会。信息将从第一阶段过渡到第二阶段，直到构思出一个成熟的概念。考虑到势能分析的结果，提出了方案计划。对下一个阶段（3）原型设计，对可能的产品或工艺进行设计和概述。该草案是原型的基础，由患者和技术专家进行评估，然后再进入下一阶段。因此，关于可能改进的信息，特别是关于补充产品的新服务的想法，可以生成在（4）原型更新，产品实施过程中产生的知识和信息（服务/过程）为生产过程的优化提供了基础。在（5）产品成功分析，成熟期/使用期和售后服务行为中，对产品或服务进行应用市场研究、调查以及收集患者的个人数据。售后服务可提供持续改进的潜力。在回收产品中存储的信息，带回给生产者，激发可能的变化或新的产品修改概念，创造新的服务或服务和产品的组合。除了创始创新，还有显著创新（周期B）和增量创新（周期C），特别是当市场反馈到阶段（2）概念设计和项目计划时。由此，循环B开始瞄准延伸产品/服务的生成通过阶段（3）（4）和（5）。增量创新（C），也经常被认为是持续改进，特别是当内部平衡直接传输到阶段（4）实施中，注意到的建议立即被引入到下一个系列的产品或服务或流程本身。对于显著的创新以及增量创新，重要的是要考虑到创新过程往往仍然是必要的，但不如在准备创始创新。所描述的创新过程结构和创新周期可能适用于组织内的协作活动，但它们是为组织间实现的创新活动而设计的。

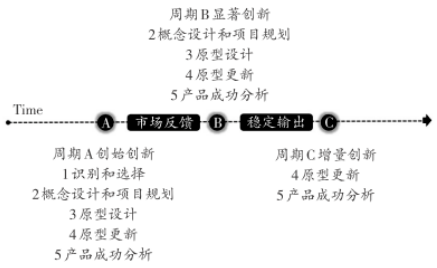


图2　协同创新系统

4.2　体医融合协同创新平台的构建

体医融合服务体系建设的基础在于体医融合创新平台的建立，平台的搭建有利于体育与医疗实现互惠知识的共用共享，利于优化所需的资源配置，以及无处不在的智能系统使得双方行动的最优同步，是两个系统高度匹配的中间桥梁。协同创新平台由临床模块、社区模块和个人模块三部分组成。

4.2.1　临床模块

临床模块共有三个步骤：诊断、运动处方或行为咨询、资源转介（将患者转介到自我管理或通过专业人员指导完成的体力活动方案）。诊断，检测患者的身体状况以及询问患者的患病史，并将结果录入到身体状况及基础疾病情况库；评估当前的体力活动水平，通过使用体力活动评估工具来实现，评估工具的使用旨在允许医生或他们的初级保健团队成员在短时间内评估和记录他们的病人的体力活动水平，实现这一步的关键是开发简便易行和

高效度的体力活动测试与评价工具或方案，并将患者的个人检测数据录入体力活动水平数据库。运动处方或行为咨询，基于运动项目及健身处方数据库资源向患者提供体力活动行为咨询和适当的体力活动处方，这一步的完成在于运动处方数据的构建。临床决策支持系统的灵活性成为了这一步骤的关键，数字终端的建立可大大提高利用率，该系统纳入了体医融合的专业人员，可以提供体力活动行为咨询和资源转介工作流程。这些专业人员可以称之为体力活动干预顾问，可以“桥接”医疗保健人员的体力活动咨询，并在资源转介系统中发挥重要作用，可为患者提供深入、个性化的咨询和随访，患者可以通过互联网自由访问和打印体力活动咨询和运动处方。在临床决策支持系统中，基于体力活动等级分类的计算机算法，对患者进行初级分类，并将信息录入到体质与运动能力数据库，并为之提供个性化的体力活动干预和运动方案。然后将这些信息嵌入到临床诊断支持系统数据库中，以促进个性化的运动处方和资源转介过程体力活动解决方案。资源转介，体力活动干预顾问为他们的患者提供了一个资源转介方案，可充分发挥目前所拥有的社区体力活动资源，以适应患者的具体需求和健康状况。患者可以得到三种资源转介选择，以满足他们的个人需要：(1)通过主动健康技术进行自我管理；(2)与通过电子认证的体力活动指导教练进行健康恢复训练；(3)通过第三方电子认证的在线课程以及各种运动教学视频进行线上健康学习。

4.2.2　社区模块

社区模块构建基于物联网的集运动健身、数据采集、指导与监控、评估、反馈、个性化健身咨询为一体的新型、闭环科学健身指导系统，共包含四部分：基于物联网的健身数据实时采集、健身过程的指导与监控、健身效果的评价与反馈、个性化健身咨询。社区居民通过物联网健身器械产生运动健身数据，数据采集与传输模块通过有线或无线数据采集接口(WiFi或RFID)，将健身数据通过移动互联网实时上传至后台健身数据云中心(体质测试数据也上传到后台数据云中心)。健身数据云中心对数据进行分类、加工、筛选和存储。运动健身监控指导专家团队对分类数据进行提取与专业分析，形成符合健身者的个性化运动处方，并将运动处方存储在健身数据云中心。健身指导人员通过调取存储在健身数据云中心的运动处方，分析和了解健身者的运动处方，对健身者进行现场指导与运动干预，以监督其健身计划的实施。运动健身监控管理云中心对健身和体质测试数据进行综合诊断分析，对健身效果进行评估，并将评估结果和指导意见通过手机短信、健身监控与指导系统、手机APP软件、智能终端等渠道反馈给健身者。

4.2.3　个人模块

个人模块基于主动健康技术的客观体力活动评估工具(计步器、加速度计和智能手机应用程序等)帮助监测患者体力活动计划取得进展。体力活动水平数据库的建立便于进行资源转介的患者同医生进行数据上的交流，即运动强度、持续时间和活动频率收集的数据可以通过上传到临床决策支持系统，能够讨论从第一次医疗访问到下一次医疗访问所做的健康恢复和健身改进方面的积极作用。

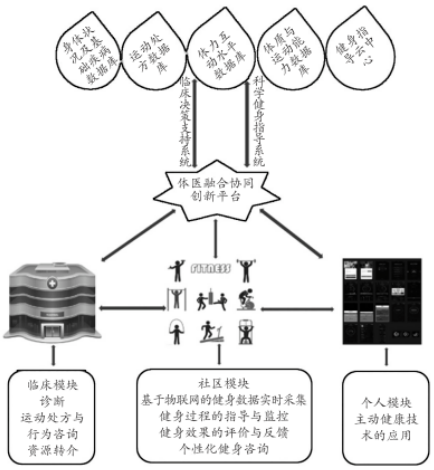


图3　体医融合协同创新平台简介

5　结束语

借鉴“协同创新”理论模型和美国的“运动是良医”健康促进模式，整合体育和医疗机构的优势资源，提出了构建基于体医融合的“医疗——社区——个人”三联动的健康促进模式。以体医融合创新平台为保障的协作平台、以体育机构和医疗机构为技术支撑的体医融合交叉专家团队、以“三联动”为运行核心，分别在三个层面实施干预力量。协同创新系统的构建为体医融合提供了源源不断的创

新动力，保障了体医融合科学有序的运转，临床决策支持系统，实现了体医融合的首要环节，社区科学健身指导系统，为社区居民的健身提供了科学实践基础。个人模块的主动健康技术的应用保障了居家锻炼的健身效果，至此，一个闭环的体医融合路径已经形成。随着全民健身和健康中国上升为国家战略，非医疗干预手段日渐深入人心。体医融合的培育与发展是新时代迎合国家战略的内在需求，也是医治未病的健康新理念。体医融合作为健康中国的重要实施路径，科学持久的发展模式是其核心，深入基层的服务理念与行动是考量体医融合模式成功的关键所在。今后，在政策和理论不断完善的同时，要深入一线通过足够的量的积累，使体医融合的步伐走得更为坚实稳重。

参考文献：

[1] 梁美富，郭文霞.“健康中国”战略背景下体医结合的发展路径探讨——基于PEST分析[J].河北体育学院学报，2018，32(3)：52-56.

[2] 我国首部健康管理蓝皮书发布[J].中国肿瘤临床与康复，2019(5)：636.

[3] 朱晓东，刘炎斌，赵　慎.健康中国建设背景下的“体医结合”实践路径研究——基于现代医学模式视角[J].山东体育学院学报，2019，35(4)：33-38.

[4] F. Lobelo，M. Stoutenberg，A. Hutber. The exercise is medicine global health initiative [J].Br J Sports Med，2014，48(22)：1627-1633.

[5] R. E. Sallis. Exercise is medicine and physicians need to prescribe it[J].British journal of sports medicine，2009，43：3-4.

[6] R. W. Grant，J. A. Schmittdiel，R. S. Neugebauer，et al. Exercise as a vital sign：a quasi-experimental analysis of a health system intervention to collect patient-reported exercise levels[J].Journal of general internal medicine，2014，29：341-348.

[7] K. J. Coleman. Initial validation of an exercise “vital sign” in electronic medical records[J].Medicine & Science in Sports & Exercise，2012 ，44：2071-2076.

[8] 马　健.产业融合理论研究评述[J].经济学动态，2002(5)：78-81.

[9] 李璟圆，梁　辰，高　璨，等.体医融合的内涵与路径研究——以运动处方门诊为例[J].体育科学，2019，39(7)：23-32.

[10] 汤立许，刘华强.体医融合的内外之困与化解之道——基于共生系统理论的视角[C].第十一届全国体育科学大会论文摘要汇编，2019.

[11] 冯振伟，王先亮.基于共生理论的体育业与医疗服务业融合共生路径构建研究[J].山东体育学院学报，2018，34(5)：1-7.

[12] 王春顺，姜方平，李国泰.新时代我国体医融合双元创新发展研究[J].体育文化导刊，2019(5)：6-11.

[13] 张剑威，汤卫东.“体医结合”协同发展的时代意蕴、地方实践与推进思路[J].首都体育学院学报，2018，30(1)：73-77.

[14] 陈　劲，阳银娟.协同创新的理论基础与内涵[J].科学研究，2012，30(2)：161-164.

（上接第12页）

参考文献：

[1] 李玉娟，高　义.“三生教育”研究述略[J].黑河学刊，2014(8)：80-82.

[2] 刘铁芳，刘艳侠，黄　鑫.疫情后的教育观念与实践变革(笔谈)[J].北京教育学院学报，2020，34(4)：21-28.

[3] 中国中央，国务院.“健康中国2030”规划纲要[EB/OL].http://www.gov.cn/gongbao/content/2016/content_5133024.htm，2016-10-25.

[4] 李小伟，袁圣敏.后疫情时期中国体育与健康教育的思考与发展策略研究[J].成都体育学院学报，2020，46(5)：15-20.

[5] 中共中央，国务院.中国教育现代化2035[EB/OL].https://baijiahao.baidu.com/s?id=1626257152189751688&wfr=spider&for=pc，2019-02-23.

[6] 李王杰，刘喜山.以新冠肺炎之殇触发体育锻炼机遇[J].沈阳体育学院学报，2020，39(3)：18-22.

[7] 翁锡全.体育·环境·健康[M].北京：人民体育出版社，2004.

[8] 冯　蕾.居家锻炼成常态　家庭体育时代即将到来[EB/OL].http://www.sport.gov.cn/n317/n344/c946577/content.html，2020-04-01.

[9] 抗击新冠肺炎疫情，习近平总书记这样谈“爱国卫生运动”[EB/OL].https://politics.gmw.cn/2020-06/06/content_33891626.htm，2020-06-06.

[10] 毛　迪，刘连发.重大灾疫情景下大众体育参与的认知判断及价值引导——基于习近平弘扬伟大抗疫精神重要讲话的思考[J].体育与科学，2020，41(6)：1-5.

3.3.3 平衡训练干预对痉挛型脑瘫儿童立位平衡控制能力的影响（王疆娜）

医学康复

·临床研究·

平衡训练干预对痉挛型脑瘫儿童立位平衡控制能力的影响

王疆娜[1]，杨敬暖[2]，毛敏[3]，宋祺鹏[1]，张翠[2]，田雪文[1]，孙威[2]

【摘要】 目的：探讨16周平衡训练干预对提高痉挛型脑瘫儿童身体站立位平衡控制能力的作用。方法：选取50名7～12岁粗大功能Ⅰ级的痉挛型脑瘫儿童，并随机分为观察组和对照组各25例。2组均给予常规康复训练，观察组在此基础上接受16周的平衡训练。受试者在干预前和干预后分别接受Berg平衡量表(BBS)、静态平衡及动态平衡测试。结果：治疗16周后，2组BBS评分均明显高于干预前($P<0.05$)；静态平衡测试在睁眼(DLO)与闭眼状态(DLC)下足底压力中心在左右方向最大动摇径(Dx)、前后方向最大动摇径(Dy)、移动总路程(Lng)和包络面积(Area)干预后值显著低于干预前($P<0.05$)；动态平衡测试干预后得分显著高于干预前($P<0.05$)、最大旋转角速度(MRS)、平均旋转角速度(ARS)显著小于干预前($P<0.05$)。组间平衡能力比较，干预后观察组和对照组Berg平衡量表得分无显著性差异；干预后观察组Dx-DLO、Dx-DLC、Dy-DLO、Dy-DLC、Lng-DLO、Area-DLO显著小于对照组($P<0.05$)；干预后观察组动态平衡得分显著高于对照组($P<0.05$)，ARS及MRS显著低于对照组($P<0.05$)。结论：16周的平衡仪和常规康复训练干预均可以有效地提高脑瘫儿童身体平衡控制能力，降低跌倒的风险；相比于常规康复训练，平衡仪训练改善静态及动态站立平衡能力效果更好。

【关键词】 平衡训练；脑瘫儿童；姿势控制

【中图分类号】 R49；R742 【DOI】 10.3870/zgkf.2020.02.009

Effects of 16-week balance training on standing balance ability in children with cerebral palsy

Wang Jiangna, Yang Jingnuan, Mao Min, et al. College of Sports and Health, Shandong Sports University, Jinan 250102, China

【Abstract】 Objective: To investigate the effects of 16-week balance intervention training on standing balance control ability in cerebral palsy children. **Methods:** Fifty participants with gross motor functional dysfunction level Ⅰ aged 7-12 years old were recruited and randomly divided into observation group ($n=25$) and control group ($n=25$). Both groups were given the regular training, and observation group was subjected to 16-week balance training additionally. The Berg balance scale, and static and dynamic balance ability tests were performed before and after intervention, respectively. **Results:** After 16-week treatment, compared with pre-intervention, there were significant differences in the functional Berg balance scores, and static balance and dynamic balance ability variables in both groups before and 16 weeks after interventions ($P<0.05$). After interventions, there was significant difference between observation group and control group in the maximal sway displacement of center of pressure (CoP) in anterior-posterior direction/medial-lateral direction with eyes open and closed condition, length and envelope area of CoP with eyes open ($P<0.05$), dynamic balance score, max rotation speed and average rotation speed ($P<0.05$). **Conclusion:** Both 16-week balance training and regular rehabilitations could improve the balance control ability and reduce the falling risk in cerebral palsy children. Compared with the regular rehabilitation, the balance training can improve the static and dynamic standing balance ability more effectively.

【Key words】 balance training; children with cerebral palsy; postural control

脑性瘫痪是胎儿或婴幼儿脑部非进行性损伤所致持续存在的中枢性运动和姿势发育障碍、活动受限症候群[1-2]。2018年我国约有600万脑瘫儿童，未来还将以每年4.6万的数量递增，60%为痉挛型脑瘫[3]。肌张力、前庭机能、肌肉力量、本体感觉、神经肌肉控制功能异常引起的平衡控制能力降低是导致患儿步态异常的主要原因[2,4-6]，长期发展将导致关节痉挛畸形，

基金项目：教育部人文社会科学研究青年基金项目(特殊教育学校中脑瘫儿童运动功能训练与精准体育教学的整合研究，19YJC880083)，山东省重点研发计划(脑瘫儿童步态运动功能障碍的精准量化评估关键技术研究，2019GSF108211；优秀运动员常见运动损伤的生物力学技术诊断及康复研究，2017G006043)
收稿日期：2019-07-09
作者单位：1.山东体育学院运动与健康学院，济南 250102；2.山东省体育科学研究中心，济南 250102)；3.美国北卡罗来那大学，美国北卡 300020
作者简介：王疆娜(1986-)，女，讲师，主要从事特殊人群运动康复方面的研究。
通讯作者：孙威，sunwei841024@163.com

甚至关节功能的完全丧失[7]。综述前人研究发现，改善平衡控制能力的训练方法，如悬吊训练[8]、感统训练[9]、中频电刺激疗法[10]、核心力量训练等[11]，虽一定程度上可提高肢体的运动、平衡及协调功能，但训练过程中不能充分调动患儿积极性、主动性。Biodex 训练作为结合趣味性的平衡训练方法，强调神经肌肉反馈控制的主动训练理念。研究发现，该训练方法具有提高脑卒中病人平衡能力，预防跌倒的良好功效[12]，但对脑瘫儿童的干预效果尚未可知。本研究拟运用专业的平衡训练系统对脑瘫儿童进行纵向干预训练，探讨平衡训练与常规康复训练在改善平衡控制能力的影响，为脑瘫儿童的康复提供参考依据，以指导患儿更科学更有效的康复训练。

1 资料与方法

1.1 一般资料 本研究选取 2018 年 1 月～2019 年 2 月接收的痉挛型瘫性患儿共 50 名为研究对象。纳入标准：符合 2015 年《中国脑性瘫痪康复指南》制定的痉挛型脑瘫诊断、分型标准[1]；年龄 7～12 岁；脑瘫儿童为粗大运动功能Ⅰ级，病症轻微；行走不需要辅助器械，能够独立完成一切日常生活能力；患儿监护人知情同意。排除标准：影响行走能力及步行姿态的骨关节疾病的脑瘫患者；严重的身体其他系统疾病、如先天性心脏病等不能完成实验者；严重智能发育障碍不能完成和配合实验者。50 名受试者被随机分为 2 组。2 组受试者在年龄、性别、病程、身高、体重等一般资料比较，均不具有显著性差异。见表 1。

表 1 2 组一般资料比较

组别	n	男/女（例）	年龄（岁，$\bar{x}\pm s$）	身高（cm，$\bar{x}\pm s$）	体重（kg，$\bar{x}\pm s$）	病程（年，$\bar{x}\pm s$）
观察组	25	12/13	9.01±1.23	137.21±9.76	30.36±8.04	8.46±0.98
对照组	25	12/13	9.34±1.65	135.03±7.98	29.34±9.48	8.50±0.87

1.2 方法 2 组均给予常规康复训练，观察组在此基础上接受 16 周的平衡训练。①常规康复训练：对照组脑瘫儿童采用常规康复训练手段，包括：15min 下肢关节活动度训练：摇髋法、分髋法、“骑马”训练、扶杠侧行、功率车训练；15min 下肢肌肉力量训练：楼梯行走、蹲起跳跃；10min 放松训练：下肢关节肌肉按摩放松；10min 躯干肌肉力量训练：双手搬箱、旋转推球训练；30min 步态训练：侧步走训练、倒步走训练、正常步行训练；10min 休息；30min 平衡训练：通过 Bobath 球进行前后方向重心移动、左右方向重心移动训练。所有的训练都在专业康复师的指导和保护下进行，确保患儿康复的质量和安全。常规康复训练每周 3 次，每次 2h，共 16 周。②平衡训练干预：观察组脑瘫儿童除了接受常规康复手段中的干预外，还接受 30min 的平衡能力训练。平衡能力训练采用美国 Biodex Medical System 公司生产的 Biodex 平衡训练系统中的平衡模式训练[13]。训练仪稳定性调至 8 级，选用的平衡训练模块包括稳定性极限、迷宫控制和随机控制训练三种模块。训练要求脑瘫儿童光脚站在 Biodex 平衡检测平台上，双手置于两侧扶手上，双脚之间的距离约 8～10cm，夹角约 10～15°，目视前方。平衡训练周期安排同常规康复训练，每周 3 次，每次 2h，共 16 周。

1.3 评定标准 于训练前、训练 16 周后分别进行如下评定：①功能性 Berg 平衡量表测试（Berg balance scale，BBS）[14]：总得分越高说明被测试者越能够独立完成动作，身体姿势控制能力较好[15]，整个测试一般在 10～15min 内完成。②静态平衡能力测试[16]：受试者按照测试要求光脚平行站立于测力台上，双脚内缘距离约与肩宽，手臂自然下垂，记录压力中心轨迹。静态平衡测试分为睁眼和闭眼两种状态，睁眼状态下受试者被要求注视身体正前方 2m，高 1.4m 的一目标点。测试指标包括：双脚站立睁眼（DLO）和双脚站立闭眼状态（DLC），足底压力中心在左右方向最大动摇径（Dx）、前后方向最大动摇径（Dy）、移动总路程（Lng）和包络面积（Area）。各项指标数值越小，其静态平衡能力越好。③动态平衡能力测试[17]：受试者双足站立在德国 Dr-Wolf 公司生产的 Blance-check 动态平衡测试仪，选择 Balance test 动态测试模式进行动态平衡能力测试，测试难度选为简单，测试控制选为 sensor十，测试时间为 20s。测试指标包括：平衡得分（Score）、最大旋转角速度（maximum rotation speed，MRS）、平均旋转角速度（average rotation speed，ARS）。

1.4 统计学方法 应用 SPSS 20.0 统计软件进行统计学分析，计量资料结果采用 $\bar{x}\pm s$ 表示，组间均数比较采用独立样本 t 检验，组内均数比较采用配对样本 t 检验，$P<0.05$ 为差异有统计学意义。

2 结果

治疗 16 周后，2 组功能平衡测试 Berg 评分均明显高于干预前（$P<0.05$）；静态平衡测试在睁眼与闭眼状态下 Dx、Dy、Lng、Area 干预后值显著低于干预前（$P<0.05$）；动态平衡测试干预后平衡得分显著高于干预前（$P<0.05$）、ARS、MRS 显著小于干预前（$P<0.05$）。组间平衡能力对比结果显示，干预后观察组和对照组 Berg 平衡量表得分无显著性差异；干预后观察组 Dx-DLO、Dx-DLC、Dy-DLO、Dy-DLC、Lng-DLO、Area-DLO 显著小于对照组（$P<0.05$）；干预后观察组动态平衡得分显著高于对照组（$P<0.05$），

ARS及MRS显著低于对照组($P<0.05$)。见表2。

表2 平衡测试各项指标2组患儿干预前后比较 $\bar{x}\pm s$

项目		观察组($n=25$)		对照组($n=25$)	
		干预前	干预后	干预前	干预后
功能平衡测试	Berg平衡得分(分)	26.6±3.4	36.7±4.1[a]	27.0±5.2	37.1±3.3[a]
静态平衡测试	Dx-DLC(mm)	23.5±5.2	18.3±4.3[ab]	24.1±3.9	22.1±3.1[a]
	Dx-DLC(mm)	27.1±3.5	22.5±3.3[ab]	26.9±3.6	25.0±2.8[a]
	Dy-DLC(mm)	30.1±3.6	25.2±4.5[ab]	31.1±3.2	29.7±4.1[a]
	Dy-DLC(mm)	36.6±2.9	30.1±3.0[ab]	35.9±2.5	34.0±4.4[a]
	Lng-DLC(mm)	330.5±32.1	300.2±28.8[ab]	340.8±40.2	322.1±37.2[a]
	Lng-DLC(mm)	402.6±38.8	389.3±45.4[a]	410.4±37.2	390.3±35.0[a]
	Area-DLC(mm^2)	553.2±40.1	502.5±45.2[ab]	545.8±50.5	540.2±43.7[a]
	Area-DLC(mm^2)	763.2±60.2	756.2±67.8[a]	770.2±70.1	757.2±70.3[a]
动态平衡测试	平衡(分)	932.2±54.1	1232.2±632.1[ab]	843.1±60.2	1032.8±58.5[a]
	ARS(°/s)	5.2±2.2	3.1±1.6[ab]	4.9±2.1	4.0±1.2[a]
	MRS(°/s)	10.2±4.4	7.2±2.5[ab]	9.9±5.1	8.7±3.0[a]

与干预前比较，[a]$P<0.05$；与对照组比较，[b]$P<0.05$

3 讨论

3.1 两种干预方式对脑瘫儿童身体平衡控制能力影响的分析 国内外众多研究表明，Berg平衡量表应用于脑瘫儿童的功能性平衡控制评估，具有较高的信效度[5,18]。本研究结果显示，经过16周的干预，2组受试者Berg量表得分均有显著提高，其跌倒风险降低。研究结果与国内外脑瘫儿童训练干预研究的结果相似。陈天聪等[18]研究同样显示12周的康复训练可以显著提高脑瘫儿童的Berg测试得分。我们研究发现，静态平衡测试结果说明，训练干预后两组患者足底压力中心晃动幅度减小，晃动轨迹较训练前更集中，说明身体平衡控制能力得到了改善。Ehab等[13]研究结果发现，干预后脑瘫患儿左右方向动摇径、前后方向动摇径和移动速度均出现显著性降低，说明训练干预有效的改善了平衡控制能力。动态平衡测试指标Score、MRS以及ARS，反映人体神经系统在不稳定状态下根据躯体感觉、视觉和前庭觉系统输入的信息作出应答反应，并通过肌骨系统的神经肌肉控制来调节人体重心，并及时调整身体姿势达到控制身体平衡的能力[17]。本研究结果证明，16周的平衡干预训练和常规康复训练，可以提高动态平衡得分，降低旋转角速度，可知干预后脑瘫儿童对姿势控制的调整及控制能力有所增强，可以更好的控制动态平衡能力。由上可知，平衡训练和常规康复训练均可以有效的改善脑瘫儿童功能性平衡能力、静态平衡能力以及动态平衡能力，起到提高身体平衡控制能力的作用。

3.2 两种干预方式对脑瘫儿童身体平衡控制能力的对比分析 本研究结果显示，相比于常规康复训练，16周的平衡训练可以更加有效地改善静态平衡能力和动态平衡能力，说明观察组脑瘫儿童平衡控制能力的康复效果相比于对照组效果更好。造成这种现象的原因可能与平衡训练仪特殊的训练模式有关。Biodex训练更多的是对视觉反馈和本体感觉的整合训练。Lord等[19]研究发现在控制视觉时(闭眼、戴眼罩)，人体闭眼站立时压力中心移动面积增加约30%，说明人体站立姿势稳定性明显下降，验证了视觉反馈对平衡控制的重要作用。脑瘫儿童在静态平衡仪上训练可能可以更多的依赖视觉反馈的信息，及时精确的对目标位置做出姿势调整反馈，进而更有效的起到改善平衡控制的作用。本体感觉是一种人体感知空间位置、姿势变化及身体各部位的运动变化的深感觉[21]。研究显示髋、踝关节本体感觉信息输入是维持平衡的重要因素，约占感觉信息输入的70%[22]。静态平衡仪训练主要通过踝关节和髋关节的调节来控制位置点向目标点靠近，这种训练对关节运动的精确度效果更佳。由以上分析可知，观察组的训练方法更有利于视觉反馈和本体感觉信息的整合提高，改善平衡控制能力。

值得注意的是，观察组和对照组功能性Berg平衡量表得分并没有显著性差异，这可能与不同的姿势控制策略有关。脑瘫儿童因其自身存在的神经肌肉系统控制障碍，前馈控制功能障碍，而更多的是依赖后馈控制系统。静态及动态平衡后馈控制属于小范围小幅度的精准动作姿势控制，而功能性Berg平衡后馈控制属于大范围、大幅度的粗大功能姿势控制[23]。前者需要膝关节和踝关节的小肌肉群快速协同控制；而后者需要全身所有肌群特别是核心控制肌群的参与控制[23]。本研究选取的平衡仪训练方式，主要是对下肢小肌群的协同控制进行训练，进而通过髋、膝、踝关节控制策略，起到改善精准小幅度的姿势控制能力，而无法对全身大肌群特别是核心肌群起到锻炼作用，无法更好地改善功能性平衡能力。本研究结果提示，康复医疗工作者在进行脑瘫儿童平衡仪训练康复时，应注重加强全身肌群特别是核心控制肌群的专业训练。

综上所述，16周的平衡仪干预和常规康复干预均可以有效地提高脑瘫儿童身体平衡控制能力，并能够降低跌倒的风险；而平衡仪训练对静态及动态平衡控制能力干预效果更好。本研究结果提示在脑瘫儿童的康复训练中可以尝试加入平衡仪训练以更好地改善站立平衡能力。

【参考文献】

[1] 唐久来，秦炯，邹丽萍，等.中国脑性瘫痪康复指南(2015)：第一部分[J].中国康复医学杂志，2015，30(7)：747-754.

[2] 王军，郎永斌，杜江华.悬吊运动训练对痉挛型脑瘫患儿运动及平衡功能的影响[J].中国当代儿科杂志，2018，20(6)：465-469.

[3] 张丽华，金彩君，王立苹，等.感觉统合训练对痉挛型脑瘫儿童立位平衡功能的影响[J].中国康复理论与实践，2011，17(1)：72-

74.
[4] Grande-Alonso M, Moral SB, Minguez ZA, et al. Biobehavioural analysis of the vestibular system and posture control in patients with cervicogenic dizziness. A cross-sectional study[J]. Neurologia, 2018, 33(2): 98-106.
[5] 姜积华，张新斐，陈小芳，等. 本体感觉训练对跟腱延长术后痉挛型脑瘫肢体运动功能的影响[J]. 按摩与康复医学，2015，6(8)：47-48.
[6] 章马兰，刘振寰. 针刺对痉挛型脑瘫患儿下肢肌张力的影响[J]. 中国针灸，2018，38(6)：27-32.
[7] 漆带丽，颜华，李婵. 股四头肌训练仪在脑瘫患儿康复中的应用.[J]. 当代护士，2010，8(1)：75-76.
[8] 岑红燕，黄军祥，彭林海，等. 悬吊运动训练对脑卒中偏瘫患者下肢功能障碍的康复效果[J]. 中国医学创新，2019，16(2)：15-18.
[9] 董卓瑛，吴谦，雷晓辉，等. 感统训练联合骑马术对脑瘫患儿功能恢复的临床研究[J]. 中国社区医师，2018，34(15)：177-178.
[10] 杜颖，娄丹. 运动疗法结合电刺激治疗痉挛型脑性瘫痪患儿疗效分析[J]. 中国中西医结合儿科学，2018，10(1)：44-47.
[11] 查磊，鲁雷. 核心力量训练对不同人群的应用效果研究[J]. 商丘师范学院学报，2018，28(6)：95-99.
[12] 潘翠环，叶正茂，缪萍，等. 平衡功能训练对脑卒中后步行功能恢复的影响[J]. 中国康复医学杂志，2014，29(3)：263-264.
[13] Abd El-Kafy, El-Basatiny. Effect of postural balance training on gait parameters in children with cerebral palsy[J]. Am J Phys Med Rehabil, 2014, 93(11): 938-947.
[14] Christovao TC, Pasini LA, Grecco LA, et al. Effect of postural insoles on static and functional balance in children with cerebral palsy: a randomized controlled study[J]. Braz J Phys Ther, 2015, 19(1): 44-51.
[15] Kembhavi G, Darrah J, Magill-Evans J, et al. Using the berg balance scale to distinguish balance abilities in children with cerebral palsy[J]. Pediatr Phys Ther, 2002, 14(2): 92-99.
[16] 袁空军，吴加弘. 广场舞和太极拳锻炼对中老年女性动态平衡能力的影响[J]. 北京体育大学学报，2018，41(3)：82-86.
[17] 何璐，徐开寿，邱晒红，等. Berg 平衡量表对痉挛性脑瘫儿童平衡功能评定的信度研究[J]. 中国康复，2010，25(1)：21-23.
[18] 陈天聪，叶一卫，程佩峰，等. 运动平板训练对学龄期脑瘫患儿平衡功能与步态的影响[J]. 中国康复医学杂志，2014，29(7)：633-636.
[19] Lord SR, Ward JA, Williams P. Exercise effect on dynamic stability in older women: a randomized controlled trial[J]. Arch Phys Med Rehabil, 1996, 77(3): 232-236.
[20] Fitzpatrick R, McCloskey DI. Proprioceptive, visual and vestibular thresholds for the perception of sway during standing in humans[J]. J Physiol, 1994, 478 (Pt 1): 173-186.
[21] Deursen RW, Simoneau GG. Foot and ankle sensory neuropathy, proprioception, and postural stability[J]. J Orthop Sports Phys Ther, 1999, 29(12): 718-726.
[22] Thorstensson AH, Roberthson. Adaptations to changing speed in human locomotion: speed of transition between walking and running[J]. Acta Physiologica, 2010, 131(2): 211-214.
[23] Gome uka NA, Bona RL, Rosa RGD, et al. Adaptations to changing speed, load, and gradient in human walking: Cost of transport, optimal speed, and pendulum[J]. Scandinavian Journal of Medicine & Science in Sports, 2014, 24(3): e165-e173.

·外刊拾粹·

经皮神经电刺激治疗头痛

在发达国家，约有2%的人口每天都会经历慢性头痛。在过去的二十年中，针对头痛治疗的侵入性和无创性神经刺激技术以及无创性神经调控技术得到了长足的发展。本研究旨在探究经皮神经电刺激(PENS)对原发性头痛的疗效。

本研究为回顾性研究，纳入了在2012年至2016年间的36例慢性难治性头痛患者。在这些患者中25例患有原发性头痛，其中14例为慢性偏头痛(CM)，9例为慢性丛集性头痛(CCH)，2例是新发的每日持续性头痛(NDPH)。研究者应用21号针头将经皮神经电刺激治疗传递至疼痛同侧的枕神经。治疗每次25到30分钟，每12秒钟刺激一次，频率为2Hz至100 Hz，一天三次，电压根据患者个人的耐受度而定，介于1.2伏与至2.5伏之间。

在患有慢性丛集性头痛的患者中，9例中有6例在第一次治疗后症状明显改善，表现为发作频率和/或严重程度降低至少持续4周。通过进一步的治疗，其中4例患者在后续治疗中获得了相似的改善。但是，在CM / NDPH患者中，仅有4例在PENS治疗中获益。对枕大神经阻滞的治疗反应并不能预测对PENS的治疗反应。

结论：这项针对难治性原发性头痛患者的回顾性研究发现，经皮神经电刺激可能对慢性丛集性头痛的患者有益。

(樊蕴辉 译)(陆蓉蓉审译)

Weatherall M, et al. Percutaneous Electrical Nerve Stimulation (PENS) Therapy for Refractory Primary Headache Disorders: A Pilot Study. Br J Neurosurg. 2019. doi.org/10.1080/02688697.2019.1671951.

中文翻译 由 WHO 康复培训与研究合作中心(武汉)组织
本期由复旦大学华山医院 吴毅教授主译编

3.3.4 OSCE 考核模式对护理在校生临床操作技能水平影响的 Meta 分析（王清路）

OSCE 考核模式对护理在校生临床操作技能水平影响的 Meta 分析

刘园园[1]，朱　霞[2]，王清路[3*]
（1.中国石油大学胜利学院，山东　东营　257061；2.东营众创能源科技股份有限公司，山东　东营　257000；
3.齐鲁医药学院，山东　淄博　255300）

摘　要：目的　系统性评价 OSCE 考核方式对护理在校生临床操作技能水平的影响。方法　计算机检索中国知网、万方数据库、维普数据库、中国生物医学文献数据库、PubMed、EBSCO、Springer Link 数据库中关于 OSCE 教学法对护理在校生临床操作技能水平影响的随机对照试验，由 3 名研究者对相关文献进行筛选和资料提取，并对符合 Cochrane 偏倚风险判断标准的文献进行 Meta 分析。结果　共纳入 14 篇 RCT 文献。Meta 分析结果显示：(1) 临床操作技能考核理论成绩中 OSCE 考核模式与传统考核方式相比具有统计学差异（$P<0.05$）；(2) 临床操作技能实训考核中 OSCE 考核模式与传统考核方式相比具有统计学差异（$P<0.001$），但在内科护理学实训成绩中 OSCE 考核模式与传统考核方式相比并无统计学差异（$P=0.15>0.05$）。结论　OSCE 考核模式可提高护理在校生总体的临床理论成绩和临床操作技能水平，但是对护理专科生临床操作技能的影响效果不显著。

关键词：OSCE 考核模式；护理在校生；临床操作技能；Meta 分析
中图分类号：G424.28　　文献标识码：A　　文章编号：1671-1246(2020) 16-0148-04

护理临床操作技能是为了治疗疾病、减轻病人痛苦、促进病人康复、满足病人需要，帮助病人以最佳的身心状态度过疾病各个阶段所实施的实践性较强的护理临床基本技能[1]。护理临床操作技能作为一门综合性高、应用性强的专业课程，其特点是将多种临床常见疾病联系在一起，相关内容多且细，知识点相互穿插且繁杂，加上临床上的护理操作并非独立存在，因此需要护理在校生将理论知识和临床操作技能结合起来才能具备综合处理各种疑难病症的能力。调查发现[2-3]，以传统考核方式培养的护理专业学生的基础理论知识较扎实，但护理临床知识的应用、操作能力较低。因此，选择更好的考核方式对促进护生临床护理操作技能提升具有必要性和重要性。1975 年英国 Dundee 大学的 Harden 博士提出的客观结构化临床考试（Objective Structured Clinical Examination，OSCE）是一种全新的客观临床能力考核模式，它突破了医学教育中传统的书面考试方式，可以客观地评价护理学生的临床技能和临床决策能力[4-5]。目前国内护理院校对护理在校生 OSCE 考核模式的研究已经达到一定的数量，但是单一研究的样本量相对较少。因此，本研究以循证为依据，运用 Meta 分析的方法分析 OSCE 考核模式对护理在校生护理临床操作技能水平的影响，探讨适合护理在校生护理临床操作技能考核的最佳模式，以此为护理临床教学考核模式的选择提供科学依据。

1　资料与方法

1.1　检索策略

检索中国知网、万方数据库、维普数据库、中国生物医学文献数据库、PubMed、EBSCO、Springer Link 数据库等，检索时限为建库至 2019 年 12 月，检索 OSCE 考核模式对护理在校生临床操作技能水平影响的文献。检索方式为主题词加自由词的组合方式，如“OSCE”“Nursing procedures or nursing operation”“Nursing students”“护理在校生”“护生”“OSCE”“客观结构化临床考试”等，并追踪已发表的相关 Meta 分析的参考文献。

1.2　纳入标准和排除标准

纳入标准：(1) 护理在校生作为研究对象；(2) 对照组采用传统的护理临床操作考核模式；(3) 干预组采用 OSCE 考核方式；(4) 研究效应指数为计量资料。排除标准：(1) 非 RCT 类的文献；(2) 对照组的考核方式为非传统法；(3) 研究对象是已工作的护理人员；(4) 重复发表的文献。

1.3　资料提取

由 3 名研究者通过互盲的形式，按照纳入与排除的标准对文献进行筛选，首次筛选题目和摘要部分，然后筛选正文部分，最后再进行交叉核对，并在意见不统一时通过征求第三方的意见予以解决。纳入的文献通过表格的方式对资料进行提取，包括作者、发表年份、纳入例数、研究对象、研究内容、评价指标等。

1.4　文献质量评价

本研究根据 Cochrane 系统评价手册的偏倚风险评估方法对纳入的 RCT 文献进行评价和分级[3]。分级质量分为 A、B、C 3 个级别，评价内容包括是否盲法、结局指标的完整性、是否随机产生序列等。其中 C 级需满足 4 项标准中的一个，说明发生偏倚的可能性很高；B 级需满足一个或一个以上的标准，说明发生偏倚的标准为中度；A 级偏倚较低，需要完全满足 4 项质量标准。

1.5　统计学处理

基金项目：山东省自然科学基金（ZR2018MH038）

本研究采用 RevMan5.1 软件进行数据的 Meta 分析，因所纳入文献均属于计量资料，故采用 $(\bar{x}\pm s)$ 为效应统计指标进行资料的 Meta 分析。采用 χ^2 检验法对异质性进行判断，如 $P>0.05$，$I^2<50\%$，选用固定相应模型；$P<0.1$，$I^2\geq 50\%$，则无临床异质性，故采用随机效应模型。如无法进行异质性判断，则需放弃 Meta 分析，采用描述性研究。

2 结果

2.1 纳入文献情况

初检文献 737 篇，其中中文 727 篇，英文 10 篇。经各数据联合查重、查看摘要和正文内容，最终纳入文献 14 篇。纳入文献的基本特征见表 1。

2.2 文献质量评价

表 1 纳入文献的基本特征

作者	发表年份	研究类型	纳入例数		研究对象	研究内容	评价指标
			干预组	对照组			
钟志璃[9]	2019	RCT	43	43	专科	儿科护理学	理论成绩、实训成绩
曹永军[7]	2019	RCT	80	70	本科	妇产科护理学	实训成绩
刘燕[10]	2012	RCT	60	60	本科	基础护理学	实训成绩
赵丽红[18]	2015	RCT	156	152	本科	基础护理学	实训成绩
邓辉[19]	2009	RCT	56	46	专科	急救护理学	理论成绩、实训成绩
洪霞[11]	2016	RCT	50	50	专科	内科护理学	理论成绩、实训成绩
曹迎东[12]	2018	RCT	60	60	本科	内科护理学	理论成绩、实训成绩
朱平[13]	2015	RCT	58	58	本科	内科护理学	实训成绩
徐倩[14]	2019	RCT	50	50	专科	内科护理学	理论成绩、实训成绩
杨茜[15]	2015	RCT	152	148	本科	内科护理学	实训成绩
吴娅玲[16]	2018	RCT	40	40	专科	内科护理学	理论成绩、实训成绩
韩江红[17]	2019	RCT	240	240	本科	综合实训	实训成绩
邓雪英[18]	2019	RCT	245	235	本科	综合实训	实训成绩
穆彤[19]	2015	RCT	53	52	专科	综合实训	理论成绩、实训成绩

本研究共纳入 14 篇 RCT 文献，且均为中文文献。多采用随机分组的方式，并且对是否采用盲法均未提及。质量评价结果显示：纳入的 14 篇文献均为 B 级。

2.3 Meta 分析结果

2.3.1 OSCE 考核模式对临床操作技能考核理论成绩的影响　在纳入的 7 篇文献[6,10-12,14,16,19]中研究对象共 693 例，其中干预组 352 例，对照组 341 例。调查 OSCE 考核模式对护理在校生临床操作技能考核理论成绩的影响，异质性检验结果为 $P<0.1$，$I^2=98\%$，故采用随机效应模型进行 Meta 分析。分析结果显示，差异有统计学意义[WMD=7.33，95%CI (2.20，12.46)，$P=0.005<0.05$]。见图 1。

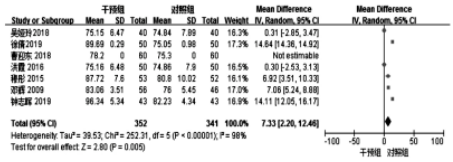


图 1 OSCE 考核模式对护理在校生临床操作技能理论成绩的 Meta 分析

2.3.2 OSCE 考核模式对临床操作技能实训考核成绩的影响　在纳入的 14 篇文献[6-19]中研究对象 2 647 例，干预组 1 343 例，对照组 1 304 例。调查 OSCE 考核模式对护理在校生临床操作技能实训成绩的影响，异质性结果为 $P<0.1$，$I^2=100\%$，故采用随机效应模型进行 Meta 分析。分析结果显示，差异具有统计学意义[WMD=8.29，95%CI (5.14，11.44)，$P<0.000\ 01$]。见图 2。

2.3.3 OSCE 考核模式对不同学历水平的护理在校生临床操作技能实训成绩的影响　本研究分别对在校护理本科生[7,9,12-13,15,17-18]和在校护理专科生[6,10-11,14,16,19]进行临床操作技能实训成绩进行 Meta 分析。分析结果显示，OSCE 考试模式对护理本科生的影响差异具有统计学意义[WMD=8.70，95%CI (1.59，15.82)，$P=0.02<0.05$]，而 OSCE 考试模式对护理专科生的影响差异并无统计学意义[WMD=7.77，95%CI (-0.01，15.55)，$P=0.05$)]。见图 3~4。

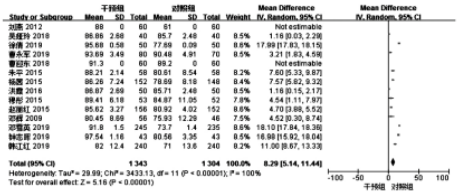


图 2 OSCE 考核模式对临床操作技能实训成绩影响的 Meta 分析

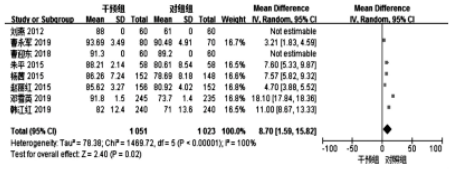


图 3 OSCE 考核模式对护理本科生临床操作技能实训成绩影响的 Meta 分析

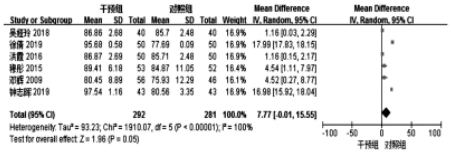


图 4 OSCE 考核模式对护理专科生临床操作技能实训成绩影响的 Meta 分析

2.3.4 OSCE 考核模式对在校护生临床课程操作实训成绩的影响　(1) OSCE 考核模式对基础护理学实训成绩的影响。纳入 2 篇文献[8-19]的研究对象共 428 例，其中干预组 216 例，对照组 212 例。调查 OSCE 考核模式对在校护生基础护理学实训成绩的影响，结果显示，差异具有统计学意义[WMD=4.70，95%CI

(3.88,5.52) ,*P*<0.000 01]。见图 5。

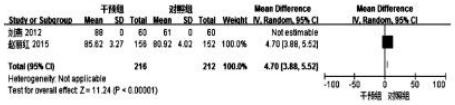


图 5 OSCE 考核模式对在校护生基础护理学实训成绩影响的 Meta 分析

(2) OSCE 考核模式对内科护理学实训成绩的影响。共纳入 6 篇文献[11-16],研究对象共 816 例,其中干预组 410 例,对照组 406 例。调查 OSCE 考核模式对在校护生内科护理学实训成绩的影响,异质性结果为 *P*<0.1,*I*²=100%,故采用随机效应模型进行 Meta 分析。结果显示,差异无统计学意义[WMD=7.10,95%CI (-2.49,16.69) ,*P*=0.15>0.05]。见图 6。

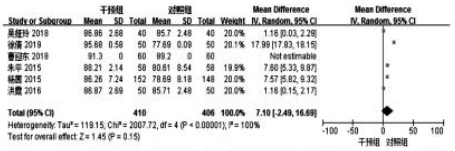


图 6 OSCE 考核模式对在校护生内科护理学实训成绩影响的 Meta 分析

(3) OSCE 考核模式对护理综合实训成绩的影响。共纳入 3 篇文献[17-19],研究对象共 1 065 例,其中干预组 538 例,对照组 527 例。调查 OSCE 考核模式对在校护生护理综合实训成绩的影响,异质性结果为 *P*<0.1,*I*²=98%,故采用随机效应模型进行 Meta 分析。结果显示,差异有统计学意义[WMD=11.36,95%CI (3.63,19.09) ,*P*=0.004<0.05]。见图 7。

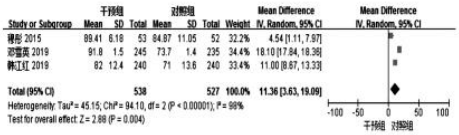


图 7 OSCE 考核模式对在校护生综合护理实训成绩影响的 Meta 分析

3 讨论

3.1 OSCE 考核模式对在校护生临床护理操作技能考核理论成绩的影响分析

临床护理操作技能在为病人治疗疾病、促进康复、减轻痛苦方面具有重要作用,而精湛、娴熟的护理操作技能需要护理在校生将理论知识和临床操作技能结合起来才能更好地综合处理各种疑难病症[20]。本研究对纳入的文献进行 Meta 分析,结果发现,OSCE 考核模式相对于传统的临床护理操作理论考核,存在统计学差异(*P*<0.05) 。分析原因可能是 OSCE 考核模式作为临床课程的考核模式,其内容讲解更加偏重临床理论课知识的教学,故干预组与对照组在临床护理操作技能考核理论成绩中差异具有统计学意义。

3.2 OSCE 考核模式可提高在校护生临床护理操作技能实训成绩

OSCE 考核模式可以模拟真实的医院场景,营造真实的医院氛围,引导护生运用护理程序对病人进行整体优质护理,在此过程中提高了护生分析问题及解决问题的能力[15]。在 OSCE 考核模式下,在校护理学生能够做到角色明确、职能分明,将各项临床护理操作有条不紊地进行下去。本研究共有 14 篇文献参与 Meta 分析,研究结果发现,OSCE 考核模式可以提高在校护理本科生临床操作技能实训成绩。OSCE 考核模式可以将临床常见的案例通过标准化病人的方式展现出来,同时将平时所学的护理知识与其他临床护理操作技能结合起来,能够帮助护生融会贯通地运用知识。但是,OSCE 考核模式对在校护理专科生临床操作技能实训成绩的影响产生的差异并无统计学意义(*P*>0.05) ,分析原因可能为在校护理专科生需要用 3 年的时间学完本科 4 年的相关课程,因此临床护理实训练习时间相对较少,同时实践动手能力相比较本科生来说会差一些,故 OSCE 考核模式对在校护理本科生和专科生会产生不同的效果。

3.3 OSCE 考核模式对护理在校生临床实训课程的影响具有差异性

本研究结果发现,OSCE 考核模式对基础护理学和护理综合实训的影响差异均具有统计学意义 (*P*<0.05) ,而对内科护理学的影响并无统计学差异 (*P*>0.05) 。分析原因可能为:(1) 本研究仅纳入 RCT 文献进行分析,对非 RCT 文献未做研究,因此出现结果的相对偏倚。(2) 内科护理学因知识点相互联系较少、知识点较多或护生对知识点之间的理解未融会贯通,因而在进行 Meta 分析时,干预组与对照组的差异无统计学意义 (*P*>0.05) 。同时,因儿科护理学、妇产科护理学、急救护理学所纳入的文献仅有 1 篇,故未做 Meta 分析,但经查阅研究结果可知,OSCE 考核模式可以明显提高护生儿科护理学与妇产科护理学的实训考核成绩。

参考文献:

[1]杨文晴,李晓杰,周谊霞,等.案例式立体教学联合微课在护理临床技能实训课中的效果观察[J].护理学报,2019,26(1) :12-16.

[2]李霞,莫霖,张萍,等.不同教学法在护生临床护理操作技能培养中的应用效果的网状 Meta 分析[J].解放军护理杂志,2018,35(3) :7-11.

[3]Vassallo M,Grey J,Hemsley A,et al. Developing an objective structured clinical examination in comprehensive geriatric assessment - A pilot study [J].Education for health (Abingdon, England) ,2019,32(2) :95-98.

[4]徐志平,夏立平.基于 OSCE 模式的《多站式护理综合实训》项目课程开发与实践[J].中国护理管理,2016,16(12) :1666-1669.

[5]Yousuf F.Agreement between simulated patients and faculty: Assessment of communication skills during objective structured clinical examination[J]. Pakistan journal of medical sciences,2019,35(6) :1570-1574.

[6]钟志晖,于浩.OSCE 模式在《儿科护理》实训考试中的应用[J].心理月刊,2019,14(10) :229.

[7]曹永军,位淑梅,高立硕,等.改良的客观结构化临床考试模式在《妇产科护理学》实验考核中的应用初探[J].天津护理,2019,27(4) :379-384.

[8]刘燕,袁晓丽,兰天,等.OSCE 应用于《基础护理学》对护生综合能力的影响[J].中外健康文摘,2012,9(8) :275-276.

[9]赵丽红,廖璎珍.简化 OSCE 在护生基础护理技能考核中的应用[J].全科护理,2015(15) :1469-1470.

[10]邓辉,左凤林,汪芝碧,等.客观结构化临床考试模式在急救护理学综合实训教学中的应用[J].护理学报,2009,16(3) :37-39.

[11]洪霞.OSCE 模式情景模拟教学法在内科护理实训教学中的应用[J].广西中医药大学学报,2016,19(1) :143-146.

基于可视化的新生儿坏死性小肠结肠炎研究分析

刘雪薇[1]，董荣芝[2]，刘希杰[3]，张晓丽[4*]

（1.滨州医学院公共卫生与管理学院，山东　烟台　264003；2.滨州医学院烟台附属医院，山东　烟台　264100；
3.滨州医学院附属医院，山东　滨州　256603；4.滨州医学院护理学院，山东　烟台　264003）

摘　要：目的　了解2000—2018年新生儿坏死性小肠结肠炎（Necrotizing EnteroColitis，NEC）的研究热点和前沿。方法　利用CiteSpace可视化分析软件对PubMed数据库2000—2018年收集的1 320篇新生儿NEC相关文献进行关键词共现分析和聚类分析，绘制出国内外新生儿NEC研究热点知识图谱；对关键词进行突变分析和时区视图分析，探索国内外新生儿NEC研究前沿，追踪其研究发展趋势。结果　新生儿NEC的研究热点涉及危险因素研究、临床表现及诊断研究、预后与结局研究、治疗及预防研究4个方面；三大前沿话题是母乳喂养、营养及肠道菌群。结论　新生儿NEC病因复杂，危险因素多样化，不论是诊断、治疗还是预防的方法都十分有限，迫切需要新的安全、有效的疗法，以满足临床需求。

关键词：新生儿；坏死性小肠结肠炎；CiteSpace；可视化分析
中图分类号：R722　　文献标识码：A　　文章编号：1671-1246（2020）16-0151-04

坏死性小肠结肠炎（Necrotizing EnteroColitis，NEC）是早产儿和新生儿中最常见的危及生命的胃肠道疾病[1]，它的发病率和死亡率类似于脑膜炎和白血病等其他严重的儿童疾病，并日益被认为是造成小儿神经发育迟缓的主要原因[2]。尽管经过40多年的研究，但NEC的病因仍不清楚[2]，早期症状不典型[3]，通常是非离散性和非特异性的，主要表现为肠壁各层的出血性和坏死性炎症[4]，且目前国际上尚无统一的有效治疗方案[5]。因此关于新生儿NEC的发病机制、诊断、预防和治疗等方面的研究受到了医务人员和科研人员的持续关注。为了准确把握历年来新生儿NEC的研究成果，促进其进一步的临床研究，本文采用信息可视化方法——科学知识图谱，并结合定性分析与计量学分析，对2000—2018年国内外有关新生儿NEC的研究文献进行系统分析。

1　材料与方法

1.1　数据来源与检索策略

在生物医学文献数据库PubMed中进行检索，检索策略为：（Necrotizing Enterocolitis[Title/Abstract]）AND（（newborn[Title/Abstract]）OR（infant[Title/Abstract]）），时间跨度限定为2000—2018年，没有语言限制，共检索到1 320篇相关文献。

1.2　分析工具与统计方法

采用CiteSpace（5.3.R4，64 bit）可视化软件为主要分析工具。在CiteSpace生成的知识图谱中，节点表示分析元素，如国家、作者、关键词及共引文献等。CiteSpace作为现在最受欢迎的可视化工具之一，可以根据引文生成共引网络，以揭示特定研究领域的结构[6]，其所绘制的知识图谱既是可视化的知识图形，又是序列化的知识谱系，具有“图”和“谱”的双重性质与特征，可以显示知识单元之间动态、交叉、演化等诸多隐含的复杂关系[7]。目前，知识图谱已被广泛应用于医学[8-9]、经济学[10-11]、管理科学[12]、信息科学与图书馆学[13-14]、教育教学研究[15]等各个学科领域。本研究通过对关键词的共现分析及聚类分析来探讨新生儿NEC的研究热点，通过关键词突变分析及时区视图分析来追踪新生儿NEC的研究趋势。

2　结果

2.1　研究热点分析

2.1.1　关键词计量分析及共现分析　对纳入的1 320篇新生儿NEC文献中的关键词进行词频统计，去除检索主题词及语义重复词等不能提示研究热点的关键词后，出现频率及中心性排名前10的关键词如表1所示。关键词共现图谱显示，节点数N=

基金项目：山东省重点研发项目（2018GSF118207）

[12]曹迎东，李翠凡，李欢，等.OSCE情景模拟教学法在内科护理学实践教学中的应用[J].实用临床护理学电子杂志，2018（30）：6，15.
[13]朱平，吴广霞，郭秀珍，等.教师标准化病人在内科护理学客观结构化临床考试中的应用[J].齐鲁护理杂志，2015（21）：113-114.
[14]徐倩.探讨OSCE模式情景模拟教学法在内科护理实训教学中的应用[J].健康大视野，2019（10）：186.
[15]杨茜，李雨欣，张世友，等.以客观结构化临床考试为导向的《内科护理学》教学模式研究[J].齐鲁护理杂志，2015，21（16）：114-115.
[16]吴娅玲，刘敏.基于OSCE模式的情景模拟教学法在内科护理实训教学中的实践探讨[J].西部素质教育，2018，4（17）：204，211.
[17]韩江红，王静，吴亦男，等.OSCE考核在护理本科生毕业技能考核中的应用效果评价[J].全科护理，2019，17（10）：1165-1167.
[18]邓雪英.OSCE培养本科护士临床能力的研究与实践[J].家庭医药，2019（4）：252.
[19]穆彤，陈莉军.客观结构化临床考试模式在中职院校护理综合实训教学中的应用[J].中国医药导报，2015，12（34）：134-137.
[20]钟伟，刘园园，尹崇高.外科护理学理论教学方法效果比较的Meta分析[J].卫生职业教育，2017，35（18）：57-60.

（*通讯作者：王清路）

3.3.5 临床检验技术模拟临床检验过程实验教学的效果探索（王清路）

●临床教学

临床检验技术模拟临床检验过程实验教学的效果探索

张 杰[1]，刘 悦[2]，王清路[1]，朱 丽[1]，邢 杰[1]

（1. 山东万杰医学院；2. 山东万杰医学院护理系；山东 淄博 255213）

［摘要］ 对我校 2010 级医学检验技术专业临床检验技术课程，进行了“模拟临床检验过程”实验课的教学探索，通过 2-3 人为一组的分组实验使学生在模拟临床环境中获得了临床检验技术的操作技能，并产生了主动学习兴趣，由 QQ 网络对参加实验与未参加实验两届毕业生调查统计，效果良好。

［关键词］ 临床检验技术；实验教学改革；模拟临床检验过程 DOI：10.3969/j.issn.1002-4701(2014)01-0089-02

［中图分类号］ G420 ［文献标识码］ A ［文章编号］ 1002-4701(2014)01-0089-02

临床检验技术是医学检验技术专业的专业课程之一，实验课采用的是“理论讲授-实验验证”的方法，实验内容与临床实际有一定差距，不能激发学生自主学习的兴趣，因而培养的学生动手操作能力不足。为使实验与临床实际更加接近，“与临床接轨”[1]，让学生“学中做、做中学”，培养真正的高技能应用型人才，笔者在 2010 级医学检验技术专业专科学生中进行了“模拟临床检验过程”实验课的教学探索[2]，提高了实验教学效果。

一、对象和方法

（一）对象。

2010 级医学检验技术专业专科班 53 名学生。

（二）方法。

1. 将 126 总教学时数的理论课与实验课时之比 80∶46 调整为 64∶62，增加了实验 16 学时，并编写了《临床检验技术实验指导》。

2. 根据临床检验技术技能操作的特点采用 2-3 人为一组的实验教学活动，每个学生都有机会模拟患者也有机会模拟检验科大夫进行实验。

3. 制定模拟临床检验的 4 个环节：(1) 标本采集：模拟临床时一个学生模拟患者持教师先填好的申请单，一个学生模拟检验科大夫按申请单内容做标本采集。例如，进行红细胞计数和血红蛋白测定时，模拟患者把申请单交给模拟大夫后，模拟大夫按照实验指导要求对模拟患者进行静脉血采集，然后对换角色进行模拟训练；(2) 标本检测：在教师指导下，学生按照实验指导要求进行项目检测。例如，实验红细胞计数时，教师在指导学生实验的过程中讲解实验原理、操作及注意事项，教师不但让学生熟练掌握“直接显微镜红细胞计数”这种经典的操作方法，而且让学生知道目前临床最常应用“血细胞分析仪”检测红细胞数量；(3) 结果报告：依据标本检测项目和报告方式的不同，设计了相应的临床检验模拟报告单，教师指导学生进行相应的结果计算及准确填写报告单。发放给患者报告单之前，教师模拟为上级大夫对报告单进行审核并签字确认，然后由模拟大夫的学生再把报告单发放给模拟患者的学生；(4) 实验过程评价：教师根据每个学生的操作是否规范，是否正确及报告单填写的是否规范进行评分。

4. 设计学生期末成绩的多元评价体系。期末成绩由 4 部分组成：(1) 期末理论及应用知识考核占总成绩的 60%；(2) 学生实验成绩占 15%；(3) 独立实验操作考核成绩占 20%；(4) 日常出勤占 5%。

5. 采用 QQ 网络，对已毕业的 2008 级、2009 级及正在实习的 2010 级学生进行评价比较。对已毕业的 2008 级 18 人、2009 级 22 人和正在实习的 2010 级学生 53 人，采用 QQ 网络（QQ 群）发出调查信息，调查统计如表 1 和表 2 所示：

表 1 08 级、09 级及 10 级学生到医院检验科临检室实习的适应时间统计

适应时间	2008 级(%)（18 人）	2009 级(%)（22 人）	2010 级(%)（53 人）
1 周	1 人(5.6)	2 人(9.2)	43 人(81.2)
2 周	0 人(0)	1 人(4.6)	5 人(9.4)
3 周	3 人(16.6)	7 人(31.7)	3 人(5.7)
4 周	14 人(77.8)	12 人(54.5)	2 人(3.7)

表 2 2010 级学生对模拟临床实验教学效果评价

项目	优(%)	良好/中等(%)	一般(%)	差(%)
实验总体效果	43(81.1)	8(15.1)	2(3.8)	0(0)
动手能力培养	48(90.6)	3(5.6)	2(3.8)	0(0)
创新意识培养	46(86.8)	2(3.8)	3(5.6)	2(3.8)
运用能力培养	42(79.3)	6(11.3)	5(9.4)	0(0)
团队协作培养	53(100)	0(0)	0(0)	0(0)

二、结 果

由表 1、2 中统计数据显示：

第一，2008 级、2009 级及 2010 级学生在医院检验科实习 1 周适应人数分别为 5.6%、9.2%、 （下转第 145 页）

操作，也可观看有关制度与技术操作示范的幻灯片，最后要对培训做好总结，通过互动了解护生的培训接受情况。

3. 对护生进行安全方面的重点教育。护生的到来一定程度上都给科室带来了事故差错的安全隐患，因此对护生加强安全教育十分重要。要帮助护生加强自我保护意识，认识安全的重要性。告知其必须在执业护士的指导监督下才能为病人进行护理，严禁擅自单独对病人进行操作。叮嘱其必须以严禁的态度对待护理工作，在执行操作时必须严格按照相关规章制度进行。以曾出现的安全隐患情况为例，告诫护生注意自我保护，避免发生安全问题[4]。同时注重提高患者沟通技巧，多带护生深入病房中，对护生示范如何与患者有效交流，帮助护生掌握沟通交流技巧，改善护患关系。

总之，护生在临床实习中由于各种原因会存在护理安全隐患，医院应加强这方面的重视，做好相应的防范工作。

[参考文献]

[1] 胡惠惠. 护生实习过程中潜在的护理安全问题及对策[J]. 西北医学教育，2008，16(4)：798-799.
[2] 王美兰，黄崇敏，朱佩兰，等. 护生临床实习中潜在的护理安全问题及对策[J]. 中国农村卫生事业管理，2010，30(1)：49-50.
[3] 何晓兰，陈桂丽. 护生临床实习安全隐患分析及对策[J]. 护理实践与研究，2010，07(23)：101-103.
[4] 孙爱玲，肖晓玲. 临床实习护生存在的护理安全隐患与对策[J]. 中国医疗前沿(上半月)，2010，05(9)：89-90.

[收稿日期] 2013-06
[作者简介] 姜香银，女，本科，主管护师。

(上接第 89 页)

81.2%；2 周的适应人数分别为 0、4.6%、9.4%，3 周的适应人数分别为 16.6%、31.7%、5.7%，4 周的适应人数分别为 77.8%、54.5%、3.7%，数据表示：利用模拟临床检验实验教学的 2010 级学生到医院实习 1 周就适应检验科工作岗位的高达 81.2%，而未进行模拟临床检验实验教学的 2008 级和 2009 级仅分别达到 5.6%、9.2%，其效果是显而易见的。

第二，2010 级检验学生对模拟临床实验教学总体效果、培养动手能力、创新意识、运用能力方面达到优和良好程度的占学生总数的 90% 以上，团队协作培养方面达到优的占 100%。

三、讨 论

“模拟临床检验过程”设计的实验与检验科工作实际更加接近，大大激发了学生的学习兴趣。让学生在模拟临床检验工作的环境中模拟临床检验的 4 个工作环节系统化的进行各种实验技能的操作，不但培养了学生的“临床服务”意识和让学生体会到对“患者负责”的态度，而且也让学生端正了实事求是的科学的工作态度，在很大程度上激发了学生的学习兴趣，培养了学生自主学习和自主操作的能力，使学生的临床技能得到了逐步强化。

通过调整理论与实验的课时比，实验课时增加了 34.7% 的学时数，这使得学生的临床操作技能时间得到了有效的保障；由于增加实验项目，与医院检验科的检验项目更趋一致，学生在模拟临床环境中学习临床检验技术的理论和操作技能知识，学生可以更多的反复模拟临床检验操作技能，学习的内容就是工作内容，融“教、学、做”于一体[3]，因而使本课程实验教学更贴近临床检验实验的实际情况，实验效果就更好。

期末进行独立的实验操作考核，有助于增强岗位需要的职业能力。期末的实验考核是连续动态的过程，要求学生按临床常规程序进行独立操作，锻炼了学生动手、动脑能力和理论联系实际的能力，使学生具备临床检验实际岗位要求的基本操作能力[4]。

正确对待调查结果。由表 1、表 2 统计结果可以看出 2010 级学生的实验教学总体效果、动手能力、创新意识等方面达到优和良好程度的占学生总数的 90% 以上，团队协作培养达到优的占 100%。到临检室实习的适应时间缩短到 1 周的占到全体学生的 81.2%。这说明模拟临床检验过程进行的实验教学与医院检验科工作是一一对应的，学生的基本操作技能提高了，对实习工作会更加得心应手，更加感兴趣，更加愿意学，学生对这样的实验教学更加满意，因此说“模拟临床检验过程”的实验教学改革是成功的。

进行“模拟临床检验过程”实验课的教学改革虽然能提高学生的专业基本操作技能，能提高实验课的教学效果，但本实验教学方法用时偏多，教师工作量加大等问题还有待进一步完善。

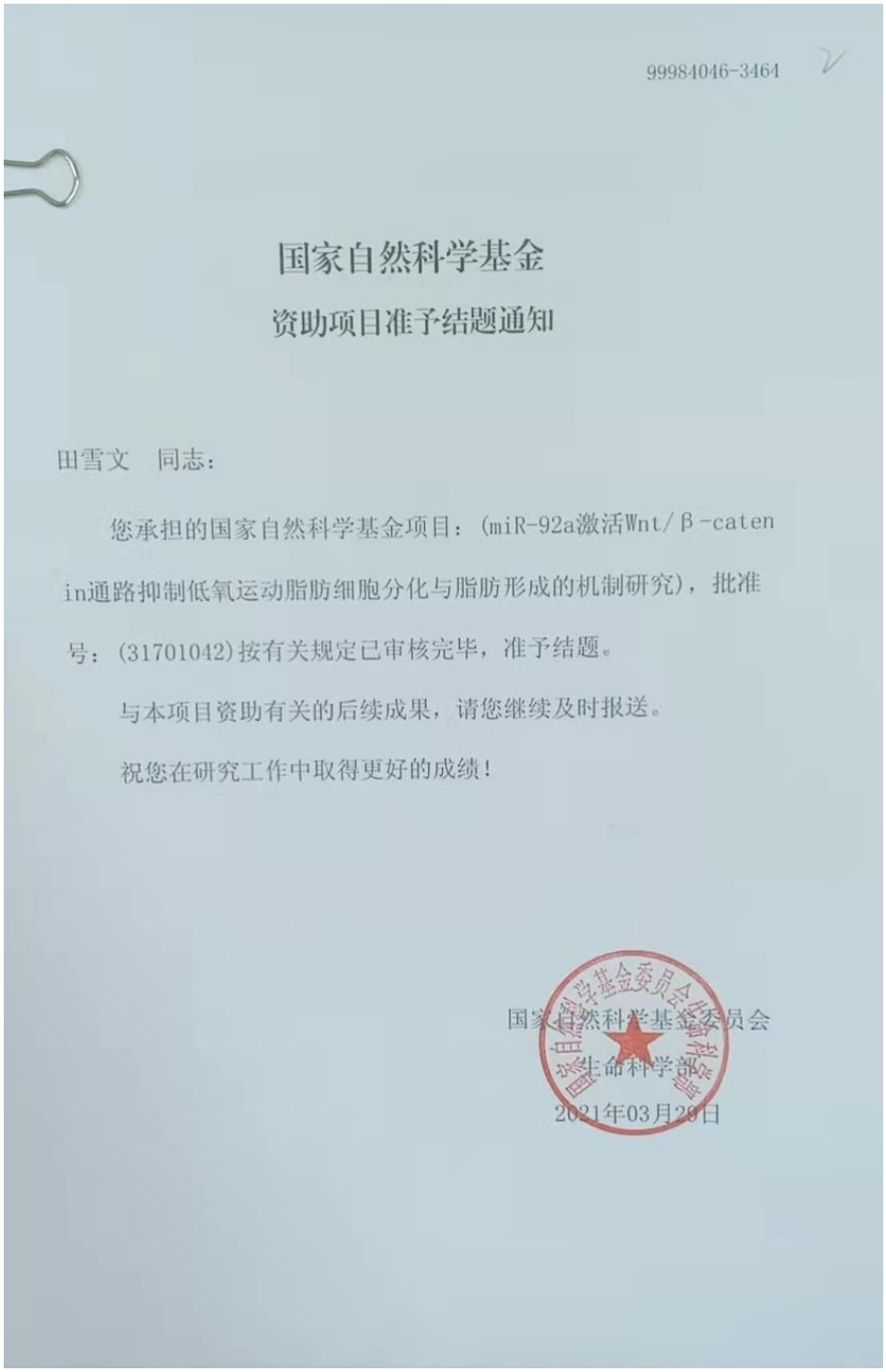
[参考文献]

[1] 袁学杰，袁金涛. 高职高专医学检验技术专业实验实训教学的改革探讨[J]. 中国误诊学杂志，2010，10(27)：6655-6656.
[2] 钱丽丽，李振江，靳 静. 高职高专医学检验专业实用性人才培养教改探索[J]. 河南职工医学院学报，2009，21(6)：655-656.
[3] 于洪帅，孙 蕊，柳 叶，等. 浅谈开放实验室对大学生综合素质的培养[J]. 检验医学教育，2008，15(3)：37-38.
[4] 侯振江，李红岩，李吉勇. 高职高专院校检验医学技术专业教学改革的探索[J]. 国际检验医学杂志，2009，30(4)：415-416.

[收稿日期] 2013-03
[作者简介] 张 杰，男，硕士，讲师，研究方向：临床检验技术及血液学检验技术的教学和研究工作。

4 高水平科研成果（部分）

4.1 国家自然基金（田雪文、孙威）



99984046-3464

国家自然科学基金

资助项目准予结题通知

田雪文　同志：

您承担的国家自然科学基金项目：(miR-92a激活Wnt/β-catenin通路抑制低氧运动脂肪细胞分化与脂肪形成的机制研究)，批准号：(31701042)按有关规定已审核完毕，准予结题。

与本项目资助有关的后续成果，请您继续及时报送。

祝您在研究工作中取得更好的成绩！

国家自然科学基金委员会

生命科学部

2021年03月29日

99984046-3464

国家自然科学基金

资助项目准予结题通知

孙威　同志：

您承担的国家自然科学基金项目：(太极拳运动预防老年人跌倒的神经力学机制及计算机仿真研究)，批准号：(31700815)按有关规定已审核完毕，准予结题。

与本项目资助有关的后续成果，请您继续及时报送。

祝您在研究工作中取得更好的成绩！

国家自然科学基金委员会

生命科学部

2021年03月20日

4.2 山东省科技进步奖二等奖（田雪文、王清路）

山东省科学技术奖

证 书

为表彰山东省科学技术奖获得者，特颁发此证书。

项目名称：全民健身一体化智能平台的构建
获奖等级：贰等
获 奖 者：田雪文（第壹位）
身份证号：3706021978****4388
工作单位：山东省体育科学研究中心
类　　别：科学技术进步奖

山东省人民政府

2018 年 03 月 23 日

证书号：JB2017-2-71-R01

山东省科学技术奖

证 书

为表彰山东省科学技术奖获得者，特颁发此证书。

项目名称：全民健身一体化智能平台的构建
获奖等级：贰等
获 奖 者：王清路（第肆位）
身份证号：3724211976****4614
工作单位：齐鲁医药学院
类　　别：科学技术进步奖

山东省人民政府

2018 年 03 月 23 日

证书号：JB2017-2-71-R04

4.3 中国轻工业联合会科学技术进步奖一等奖（王清路）

中国轻工业联合会科学技术进步奖

证　书

为表彰中国轻工业联合会科学技术进步奖获得者，特颁发此证书。

项目名称：全民健身一体化智能平台的构建

奖励等级：一等奖

获 奖 者：徐培明　师丹玮　田雪文　王清路
宋祺鹏　程静静　杨敦暖　[illegible]

2018 年 2 月 18 日

证书号：2017-J-1-14

4.4 2018 年山东省高等学校科学技术奖三等奖（王清路）

2018 年度山东省高等学校科学技术奖

证　书

为表彰我省高等学校科研工作者取得的科研成果，经山东省高等学校科学技术奖评审委员会审定，特发此证，以资鼓励。

成果名称：金属硫蛋白功能解析与开发利用

奖励等级：本科高校类　　叁等奖

获 奖 者：王清路　周彩霞　张　杰　程　静　崔为同
成洪达　范荣军　武天慧　窦　烨

山东省教育厅
二〇一八年十月

证书编号：2018BK30080

4.5 2019 年山东省高等学校科学技术奖三等奖（田雪文、王清路）

2019 年度山东省高等学校科学技术奖

证　书

为表彰我省高等学校科研工作者取得的科研成果，经山东省高等学校科学技术奖评审委员会审定，特发此证，以资鼓励。

成果名称： 过度运动导致免疫失衡的机制及蛹虫草减缓运动疲劳的应用研究

奖励等级： 本科高校类　　叁等奖

获 奖 者： 田雪文　王清路　赵学峰　宋祺鹏　程静静
沙继斌　王疆娜　钟　诚　张玉军

山东省教育厅

二〇一九年十二月

证书编号：2019BK30066

4.6 2019 年山东省高等学校科学技术奖三等奖（王清路）

2019 年度山东省高等学校科学技术奖

证　书

为表彰我省高等学校科研工作者取得的科研成果，经山东省高等学校科学技术奖评审委员会审定，特发此证，以资鼓励。

成果名称： 基于荧光可视病毒模型构建中药抗病毒谱的研究

奖励等级： 本科高校类　　叁等奖

获 奖 者： 王清路　程　静　田雪文　王　微　张　璟
范荣军　王云飞　崔为同　罗　莹　类承斌

山东省教育厅
二〇一九年十二月

证书编号：2019BK30124

4.7 2019 年山东省药学会科学技术奖二等奖（王清路）

证书

为表彰在促进我省药学科技方面中做出重要贡献的科技工作者，特颁发此证书，以资鼓励。

获奖等级：二等奖

获 奖 者：王清路

获奖位次：第壹位（共柒位）

项目名称：

基于荧光可视病毒模型构建中药抗病毒谱的研究

山东省药学会科学技术奖励委员会

二〇二九年八月

4.8 代表性论文（部分）

4.8.1 Increasing the amount of phosphoric acid enhances the suitability of Bradford assay for proteomic research（王清路）

Weitong Cui[1]
Huaru Xue[1]
Hongda Cheng[1]
Haibin Zhang[1]
Jinghua Jin[2]
Qinglu Wang[1]

[1]Key Laboratory of Biomedical Engineering & Technology of Shandong High School, Qilu Medical University, Zibo P. R. China
[2]Environmental Protection Research Institute of Light Industry, Beijing, P. R. China

Received October 12, 2018
Revised November 28, 2018
Accepted December 11, 2018

Research Article

Increasing the amount of phosphoric acid enhances the suitability of Bradford assay for proteomic research

The Bradford assay is one of the most commonly used methods for protein quantification. However, in proteomic research, the lysis buffer generally used for dissolving proteins can cause some interference to the assay. The dye reagent of classical Bradford assay contains 8.50% (w/v) phosphoric acid, which is an important factor relating to the color yield of the assay. In this study, the phosphoric acid content in dye reagent was increased to 9.35% (w/v), 10.20% (w/v), and 11.05% (w/v) to evaluate the changes of interference and the effects of lysis buffer on the interaction between proteins and dye reagent. Results show that lysis buffer not only causes background interference but also affects the protein–dye chromogenic process. Analysis of different components in the lysis buffer showed that carrier ampholyte is the main factor that introduces interference to the Bradford assay. Detergents are well-known interfering compounds in the Bradford assay, but CHAPS and octyl b-D-glucopyranoside only cause slight interference. When the amount of phosphoric acid was increased from 8.50%(w/v) to 10.20% (w/v), the sensitivity of the Bradford assay to proteins in lysis buffer was increased, and the interference delivered by lysis buffer was considerably reduced.

Keywords:
Bradford assay / Carrier ampholyte / Lysis buffer / Phosphoric acid / Two-dimensional electrophoresis

DOI 10.1002/elps.201800430

1 Introduction

Protein quantification is important for unbiased comparison of electrophoretograms in comparative proteomic analysis. Many methods are available for protein quantification [1, 2], including Kjeldahl method [3], ultraviolet absorption method [4], Biüret assay [5], Lowry assay [6], bicinchoninic acid assay [7], and Bradford assay [8]. Among these methods, the Bradford assay is simple, rapid, cheap, sensitive, and comparably compatible with the lysis buffer, which is widely used in 2DE experiments for dissolving proteins [9–12]. Additionally, a modified Lowry method from Peterson is compatible with the lysis buffer for 2DE [13].

Although the Bradford assay has some advantage over other colorimetric protein assays in 2DE experiments, the interfering effects of lysis buffer cannot be ignored. Lysis buffer must contain large amounts of detergents, carrier ampholytes, and urea to increase the solubility of proteins. These compounds can interact with the Coomassie dye and produce a remarkable background interference to the quantification process [10]. Numerous research works on the interference of lysis buffer have been conducted [14–16], but only a few focus on the reduction of the interference.

The Bradford assay is based on the binding of CBB G-250 to proteins, which results in a protein–dye complex with increased molar absorbance at 595 nm [8]. The dye reagent in the original method comprises 0.01% (w/v) CBB G-250, 4.75% (v/v) ethanol, and 8.50% (w/v) phosphoric acid. Among these components, phosphoric acid is a major factor because it regulates the protonation degree of the CBB G-250 molecules [17]. Insufficient phosphoric acid results in large numbers of unprotonated dye molecules, whereas an excess of phosphoric acid complicates equilibrium displacement due to the excess protons [18]. Some researchers have found that decreasing the phosphoric acid concentration could increase the absorbance and decrease the variability in the chromogenic responses of different proteins [17, 19, 20].

In routine protein quantitation processes, increasing the content of phosphoric acid in the dye reagent was found to lower the interference of lysis buffer. Therefore, the phenomenon was systematically investigated by augmenting the content of phosphoric acid to different proportions. The result demonstrates that the increase in phosphoric acid not only reduces the interference of lysis buffer but also increases the

Correspondence: Dr. Qinglu Wang, Key Laboratory of Biomedical Engineering & Technology of Shandong High School, Qilu Medical University, Zibo 255300, P. R. China
Fax: +86-533-2829166
E-mail: wql_zcq@126.com

Abbreviations: **DDM**, n-Dodecyl β-D-maltoside; **GE**, General Electric; **NP-40**, Nonidet P-40; **OGP**, octyl b-D-glucopyranoside

sensitivity of the Bradford assay to samples containing lysis buffer, making the Bradford assay more compatible with 2DE experiments. In addition, the interfering effect of different kinds of carrier ampholytes and detergents was analyzed.

2 Materials and methods

2.1 Main reagents and equipment

BSA and γ-globulin were obtained from Roche Applied Science (Roche Diagnostics (Shanghai) Limited, Shanghai, China). Ovalbumin, urea, thiourea, octyl b-D-glucopyranoside (OGP), n-Dodecyl β-D-maltoside (DDM), CBB G-250, Triton X-100, DTT, Nonidet P-40 (NP-40), and CHAPS were purchased from Sigma-Aldrich (Sigma-Aldrich Shanghai Trading Co Ltd, Shanghai, China). Pharmalyte (pH 3–10 and pH 5–8), ampholine (pH 3.5–10 and pH 4–6), and IPG buffer (pH 4–7) were obtained from GE Healthcare (General Electric Co. (China), Shanghai, China). All chemicals used were of analytical grade or the best grade available. The H_2O used in this paper refers to Milli-Q water.

2.1.1 Dye reagent

CBB G-250 (100 mg) was dissolved in 50 mL 95% ethanol. A total of 100, 110, 120, or 130 mL of 85% (w/v) phosphoric acid was added to this solution. The resulting solution was diluted to a final volume of 1 L with H_2O and immediately filtered. The dye reagent was stored in brown bottles. For convenience, the dye reagent with 100, 110, 120, or 130 mL of 85% (w/v) phosphoric acid per liter was designated as Dye (8.50%), Dye (9.35%), Dye (10.20%), and Dye (11.05%), respectively.

2.1.2 Lysis buffer

The standard urea/thiourea lysis buffer [21], which contains 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 2% (w/v) DTT, and 2% (v/v) pharmalyte (pH 3–10), was used. The buffer was aliquoted into Eppendorf tubes (1 mL/tube) and stored at −20°C before use.

2.1.3 Standard solutions of proteins

Three kinds of protein solutions (BSA, ovalbumin, and γ-globulin) were used as standard, all at a concentration of 1 mg/mL in H_2O. The protein concentration in the standard solutions was determined spectrophotometrically based on the absorbance at 280 nm of 1 mg/mL solution of BSA, ovalbumin, and γ-globulin in 1 cm light path, resulting in 0.66, 0.75, and 1.35, respectively [22]. These solutions were also aliquoted and stored frozen at −20°C.

The absorbance of all samples was measured using a Shimadzu dual beam path spectrophotometer UV-1800 (Shimadzu Corp., Kyoto, Japan).

2.2 Methods

2.2.1 Measurement of the interaction between lysis buffer and different dye reagents

Certain amounts of lysis buffer (20, 40, 60, 80, and 100 μL) were pipetted into test tubes (15 × 100 mm, the same below). The volumes in the test tubes were adjusted to 100 μL with H_2O. Then, 5 mL of Dye (8.50%) was added to each test tube and were mixed by shaking. After incubation for 5 min, the absorbance at 595 nm (OD595nm) was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O and 5 mL of Dye (8.50%). The same process was repeated with the three other kinds of dye reagents.

2.2.2 Sensitivity evaluation of the Bradford assay to standard proteins with or without lysis buffer

Solutions of BSA at the concentrations of 0.20, 0.40, 0.60, 0.80, and 1.00 mg/mL were prepared, and 100 μL of each of the preceding solutions was added to a test tube with 5 mL of Dye (8.50%). After 5 min, OD595nm was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O and 5 mL of Dye (8.50%). When evaluating the effect of lysis buffer on the interaction between the dye and proteins, the same process as above was performed, except for the addition of 100 μL of lysis buffer to the reagent blank and the standard protein sample group. The same performance was repeated using the two other kinds of standard proteins (ovalbumin and γ-globulin) and the three other kinds of dye reagents (Dye (9.35%), Dye (10.20%), and Dye (11.05%)).

2.2.3 Evaluation of the interference of the individual components in lysis buffer

Solutions containing only one component of the lysis buffer were prepared in H_2O. The concentrations of individual components were the same as those in the lysis buffer. Sample volumes were all 100 μL, and the measurement process was the same as indicated above.

3 Results and discussion

3.1 Effect of phosphoric acid on the interference of lysis buffer

Lysis buffer contains large amounts of interfering substances, and at least one of these compounds interferes with the Bradford assay [22]. Phosphoric acid is a key component in the dye reagent of the Bradford assay [17, 19, 20]. The relationship between the interference of lysis buffer and the content of phosphoric acid in the dye reagent has not been studied. Results shown in Fig. 1 demonstrate that the increase in phosphoric acid concentration in the dye reagent could markedly reduce

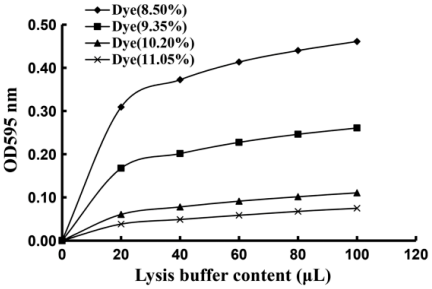


Figure 1. Interaction of different amounts of lysis buffer with dye reagents containing different amounts of phosphoric acid. (A) Sample volumes were all 100 μL. (B) Different volumes of lysis buffer were adjusted to 100 μL with H_2O. (C) Absorbance at 595 nm (OD595nm) was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O and 5 mL of corresponding dye reagent. d) The values represented the $\bar{x} \pm s$ of OD595nm for three replicate samples.

the OD595nm of lysis buffer. When using Dye (10.20%) or Dye (11.05%), the interference of lysis buffer drops to only 20 or 13% of that using the Dye (8.50%), respectively, indicating that the increase in phosphoric acid improves the compatibility of the Bradford assay with chemicals in the lysis buffer.

3.2 Effect of phosphoric acid on the sensitivity of the Bradford assay

Sensitivity to proteins is another important factor that must be considered for evaluating a dye reagent in protein quantification. The sensitivity of different dye reagents to various kinds of standard proteins without the interference of lysis buffer is first examined. As shown in Fig. 2, with the increase in the content of phosphoric acid, the sensitivity of the Bradford assay decreases with all three kinds of standard proteins. When phosphoric acid constitutes 11.05% of the dye reagent, the absorbance for all three standard proteins drops to approximately half of the value obtained with the original dye reagent Dye (8.50%). Guo [23] also reported that the sensitivity for BSA decreases as the concentration of phosphoric acid in the dye reagent increases from 6.80 to 13.60%.

When lysis buffer is added to the protein–dye system, with the increase in phosphoric acid content in the dye reagent from 8.50% to 11.05%, the OD595nm for all three standard proteins first rises then drops, as illustrated in Fig. 3, which is quite contrary to the patterns in Fig. 2. The dye reagent containing 10.20% of phosphoric acid shows the highest OD595nm, which indicates the highest sensitivity to proteins. The result from Figs. 2 and 3 indicates that lysis buffer not only introduces some background absorbance reading but also affects the complex formation between CBB G-250 and proteins.

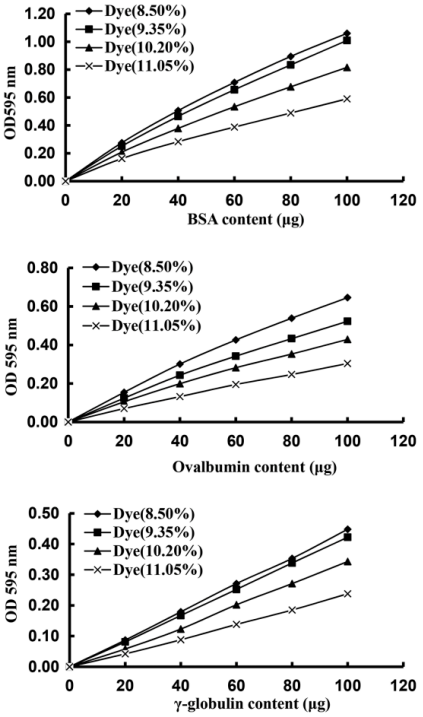


Figure 2. Interaction of different standard proteins with dye reagents containing different amounts of phosphoric acid without the interference of lysis buffer. (A) Sample volumes were all 100 μL. (B) Absorbance at 595 nm (OD595nm) was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O and 5 mL of corresponding dye reagent. (C) The values represented the $\bar{x} \pm s$ of OD595nm for three replicate samples.

Notably, when interacting with a certain kind of dye reagent, the OD595nm of different standard proteins at a certain concentration is quite different. BSA is extensively used as a standard in protein quantification because it is cheap and readily available in a pure form. The two other commonly used standards are ovalbumin and γ-globulin. The results illustrated in Figs. 2 and 3 show that all the three standard proteins display similar trends in the change of sensitivity with the increase in phosphoric acid in the dye reagent in the absence or presence of lysis buffer. However, a large difference in the sensitivity exists among different standard proteins, with BSA as the most sensitive and γ-globulin with the least sensitivity. Several studies have reported the importance of lysine and arginine residues in the binding of CBB G-250 to proteins [24–27]. The diverse amino acid composition may be the cause of different sensitivities among BSA, ovalbumin,

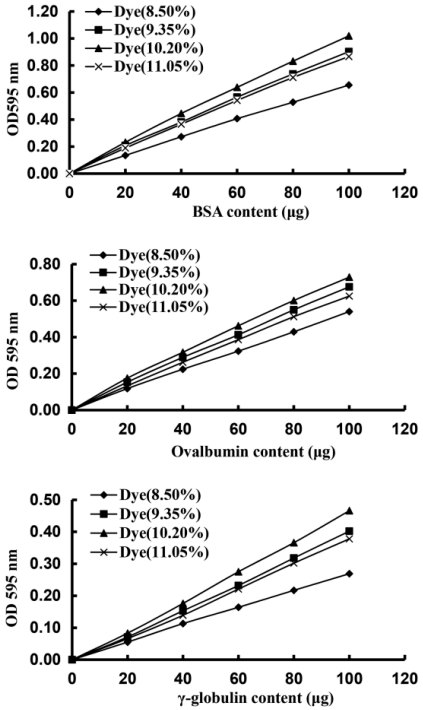


Figure 3. Effect of lysis buffer on the interaction of different standard proteins with dye reagents containing different amounts of phosphoric acid. (A) Sample volumes were all 100 μL. (B) Lysis buffer (100 μL) was added to the standard protein samples and the reagent blank. (C) Absorbance at 595 nm (OD595nm) was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O, 100 μL of lysis buffer, and 5 mL of corresponding dye reagent. (D) The values represented the $\bar{x} \pm s$ of OD595nm for three replicate samples.

and γ-globulin. Accordingly, providing a clear indication of the standard proteins used, which was not performed in most related literature, is essential.

One of the functions of lysis buffer is the denaturation of proteins, which extends the structure of proteins and exposes amino acids, completing the complex formation between CBB G-250 and proteins. This phenomenon may be the reason for the enhanced sensitivity of proteins due to lysis buffer; however, the exact mechanism must still be further researched. The theory [18] that excessive amount of phosphoric acid, which supplies an excess of protons and disturbs the equilibrium displacement, may be an appropriate explanation for the lower sensitivity in Dye (11.05%) than that in Dye (10.20%) in Fig. 3.

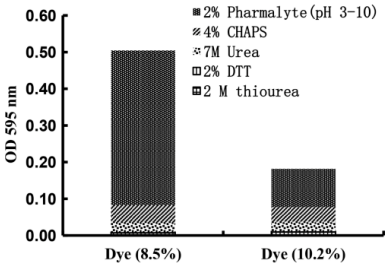


Figure 4. Interference of different components in the lysis buffer to the Bradford assay. (A) Sample volumes were all 100 μL. (B) The concentrations of individual components were the same as those in the lysis buffer. (C) Absorbance at 595 nm (OD595nm) was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O and 5 mL of corresponding dye reagent. (D) The values represented the $\bar{x} \pm s$ of OD595nm for three replicate samples.

As shown in Figs. 2 and 3, regardless of what kind of standard protein is used, the OD595nm of the same amount of protein without the effect of lysis buffer is remarkably different from that with lysis buffer. The result suggests that the effect of lysis buffer should be considered when constructing a standard curve for quantitating protein samples containing lysis buffer; otherwise, the measured value will have a large deviation.

3.3 Interference of individual components in lysis buffer

Denaturants, detergents, reductants, or other chemicals are generally utilized in 2DE experiments for denaturing proteins or enhancing their solubility [10]. The aforementioned results demonstrate that Dye (10.20%) has a considerable advantage in quantitating 2DE samples because it can lower the interference of lysis buffer and increase the sensitivity of the Bradford assay. However, the compatibility of Dye (10.20%) with all or only some of the constituents in the lysis buffer remains unknown. Therefore, the interference of individual components of the lysis buffer was studied using Dye (8.50%) and Dye (10.20%).

Figure 4 shows that the interference caused by 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, or 2% (w/v) DTT is little and does not change much in Dye (8.50%) and Dye (10.20%). By contrast, the interference of 2% (v/v) pharmalyte (pH 3–10) remarkably reduces with the increase in phosphoric acid in the dye reagent.

According to the instructions of the manufacturer (e.g., Handbook 80-6429-60AC, GE Healthcare), in 2DE experiments, protein samples loaded onto IPG strips should be prepared in lysis buffer containing various kinds of carrier ampholytes (pharmalyte reagents, ampholines, or IPG buffers). Pharmalyte (pH 5–8), ampholine (pH 3.5–10),

Table 1. Interference of different carrier ampholytes and detergents to the Bradford assay

Category	Reagent name	Absorbance at 595 nm	
		Dye (8.50%)	Dye (10.20%)
2% (v/v) carrier ampholytes	Ampholine (pH 3.5–10)	0.126 ± 0.004	0.011 ± 0.001
	Ampholine (pH 4–6)	0.113 ± 0.004	0.010 ± 0.001
	Pharmalyte (pH 3–10)	0.421 ± 0.004	0.104 ± 0.002
	Pharmalyte (pH 5–8)	0.479 ± 0.005	0.122 ± 0.002
	IPG buffer (pH 4–7)	0.425 ± 0.003	0.083 ± 0.003
4% (v/v) detergents	CHAPS	0.047 ± 0.002	0.040 ± 0.001
	Triton X-100	2.466 ± 0.007	2.502 ± 0.006
	NP-40	2.456 ± 0.004	2.486 ± 0.004
	OGP	0.009 ± 0.002	0.015 ± 0.002
	DDM	1.736 ± 0.003	1.498 ± 0.004

a) Concentrations of different kinds of carrier ampholytes and detergents were the same as those in the lysis buffer.
b) Sample volumes were all 100 μL.
c) Absorbance at 595 nm was measured in 3 mL cuvettes against a reagent blank prepared from 100 μL of H_2O and 5 mL of the corresponding dye reagent.
d) The values represented the $\bar{x} \pm s$ of OD595nm for three replicate samples.

ampholine (pH 4–6), and IPG buffer (pH 4–7) were also examined to check whether some other kinds of carrier ampholytes have a similar response as that of pharmalyte (pH 3–10). As expected, the result showed that these carrier ampholytes have similar response as pharmalyte (pH 3–10) (Table 1), that is, Dye (10.20%) is more compatible with these carrier ampholytes than Dye (8.50%).

The data in Table 1 also illustrated that interference effects of ampholine (either pH 3.5–10 or pH 4–6) are much smaller than those of pharmalyte (either pH 3–10 or pH 5–8). Carrier ampholytes are a heterogeneous mixture of synthetic polymers incorporating a variety of acidic and basic buffering groups. The net charges of ampholyte molecules partly depend on the pH of the environment [28]. When the concentration of phosphoric acid increases, the dye reagent becomes more acidic, and more ampholyte molecules have a net positive charge. The interference of carrier ampholytes is reduced probably because ampholyte molecules with a net positive charge do not easily bind with dye molecules.

The incompatibility of detergent- or surfactant-containing samples with the Bradford assay is commonly recognized [2]. Bradford reported that 0.1 mL of 1% Triton X-100 or 1% SDS could present abnormalities that are difficult to overcome [8]. Non-ionic detergents, such as Triton X-100, NP-40, and DDM, are found to have interfered severely with the assay using either kind of dye reagent (Table 1), which is consistent with what Bradford found. OGP is also a non-ionic detergent applied to isoelectric focusing and 2DE [29] but has almost no interfering effect on any dye reagent. The interfering effect of CHAPS, a kind of zwitterionic detergent, is also small. When quantitating 2DE samples with the Bradford assay, selecting proper carrier ampholytes and detergents, which can meet requirements for research and also interfere as little as possible, is better to obtain accurate quantification results.

4 Concluding remarks

Based on the aforementioned results, the optimized dye reagent recipe for 2DE samples is as follows: 0.01% (w/v) CBB G-250, 4.75% (v/v) ethanol, and 10.20% (w/v) phosphoric acid. The dye reagent is made by dissolving 100 mg of CBB G250 in 50 mL of 95% (v/v) ethanol. The solution is then mixed with 120 mL of 85% (w/v) phosphoric acid and made up to 1 L with distilled water. Then, the reagent should be filtered and stored in an amber bottle at room temperature, remaining stable for several weeks. The alternative quantification protocol contains three parts. First, a standard curve in the range of 0–100 μg standard protein is constructed. The addition of 100 μL of lysis buffer to the standard protein samples and the reagent blank must be ensured. Second, practical 2DE samples are quantitated by pipetting 100 μL of practical sample and 5 mL of dye reagent to a test tube and mixing well by inversion or shaking. After incubation for 5 min, the absorbance at 595 nm of the mixtures can be measured against the reagent blank prepared from 100 μL of lysis buffer and 5 mL of dye reagent. Finally, the protein concentration is calculated according to the standard curve. If the calculated protein concentration exceeds 1 μg/μL, then assay is performed for a range of dilutions.

Overall, by increasing the amount of phosphoric acid in the dye reagent from 8.50 to 10.20% (w/v), the interference of lysis buffer is effectively reduced, and the relatively highest sensitivity of the Bradford assay is realized in quantitating 2DE samples, making the Bradford assay more compatible with 2DE experiments. The decrease in the interference caused by carrier ampholytes is the main reason why increasing the proportion of phosphoric acid can evidently reduce the interference of lysis buffer. The selection of proper standard protein and less interfering components and the addition of lysis buffer in the construction of a standard curve are important factors to be considered for an accurate 2DE protein sample quantification, which is essential for subsequent electrophoretogram analysis and loading quantity references among different researchers and laboratories.

This work was supported by grants from the Natural Scientific Foundation of Shandong Province in China (ZR2017PC021 and ZR2018MH038), the Technology Development Project Plan of Shandong Education Department (J15LM63 and J13LL61), and the platform construction program of college-city integration of Zibo City (2018ZBXC008).

The authors have declared no conflict of interest.

5 References

[1] Sapan, C. V., Lundblad, R. L., Price, N. C., *Biotechnol. Appl. Biochem.* 1999, *29*, 99–108.

[2] Olson, B. J., Markwell, J., *Curr. Protoc. Protein Sci.* 2007, *48*, 3.4.1–3.4.29.

[3] Ballentine, R., *Methods Enzymol.* 1957, *3*, 984–995.

[4] Aitken, A., Learmonth, M., in: Walker, J. M. (Ed.), *The Protein Protocols Handbook*, Humana Press, Totowa 1996, 3–6.

[5] Gornall, A. G., Bardawill, C. J., David, M. M., *J. Biol. Chem.* 1949, *177*, 751–766.

[6] Lowry, O. H., Rosebrough, N. J., Farr, A. L., Randall, R. J., *J. Biol. Chem.* 1951, *193*, 265–275.

[7] Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., Klenk, D. C., *Anal. Biochem.* 1985, *150*, 76–85.

[8] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[9] Weist, S., Eravci, M., Broedel, O., Fuxius, S., Eravci, S., Baumgartner, A., *Proteomics* 2008, *8*, 3389–3396.

[10] Kao, S. H., Wong, H. K., Chiang, C. Y., Chen, H. M., *Proteomics* 2008, *8*, 2178–2184.

[11] Krieg, R. C., Dong, Y., Schwamborn, K., Knuechel, R., *J. Biochem. Biophys. Methods* 2005, *65*, 13–19.

[12] Okutucu, B., Dincer, A., Habib, O., Zihnioglu, F., *J. Biochem. Biophys. Methods* 2007, *70*, 709–711.

[13] Peterson, G. L., *Anal. Biochem.* 1977, *83*, 346–356.

[14] Aminian, M., Nabatchian, F., Vaisi-Raygani, A., Torabi, M., *Anal. Biochem.* 2013, *434*, 287–291.

[15] Qian, X., Dong, H., Hu, X., Tian, H., Guo, L., Shen, Q., Gao, X., Yao, W., *Anal. Biochem.* 2014, *465*, 50–52.

[16] Eddhif, B., Guignard, N., Batonneau, Y., Clarhaut, J., Papot, S., Geffroy-Rodier, C., Poinot, P., *Data Brief* 2018, *17*, 938–953.

[17] Read, S. M., Northcote, D. H., *Anal. Biochem.* 1981, *116*, 53–64.

[18] da Silva, M. A., Arruda, M. A., *Anal. Biochem.* 2006, *351*, 155–157.

[19] Chan, J. K., Thompson, J. W., Gill, T. A., *Anal. Biochem.* 1995, *226*, 191–193.

[20] Stoscheck, C. M., *Anal. Biochem.* 1990, *184*, 111–116.

[21] Weiss, W., Görg, A., in: Hagen, J. V. (Ed.), *Proteomics sample preparation*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim 2008, 129–142.

[22] Kruger, N. J., *Methods Mol. Biol.* 1994, *32*, 9–15.

[23] Guo, M., Jiang, Y., *Prog. Biochem. Biophys.* 1996, *23*, 558–561.

[24] de Moreno, M. R., Smith, J. F., Smith, R. V., *J. Pharm. Sci.* 1986, *75*, 907–911.

[25] Congdon, R. W., Muth, G. W., Splittgerber, A. G., *Anal. Biochem.* 1993, *213*, 407–413.

[26] Atherton, B. A., Cunningham, E. L., Splittgerber, A. G., *Anal. Biochem.* 1996, *233*, 160–168.

[27] Chial, H. J., Splittgerber, A. G., *Anal. Biochem.* 1993, *213*, 362–369.

[28] Lopez, M. F., *Methods Mol. Biol.* 1999, *112*, 111–127.

[29] Holloway, P. J., Arundel, P. H., *Anal. Biochem.* 1988, *172*, 8–15.

4.8.2 High heterogeneity undermines generalization of differential expression results in RNA-Seq analysis（王清路）

Cui et al. Human Genomics (2021) 15:7
https://doi.org/10.1186/s40246-021-00308-5

Human Genomics

PRIMARY RESEARCH Open Access

High heterogeneity undermines generalization of differential expression results in RNA-Seq analysis

Weitong Cui[1†], Huaru Xue[1†], Lei Wei[1], Jinghua Jin[2], Xuewen Tian[3] and Qinglu Wang[1*]

Abstract

Background: RNA sequencing (RNA-Seq) has been widely applied in oncology for monitoring transcriptome changes. However, the emerging problem that high variation of gene expression levels caused by tumor heterogeneity may affect the reproducibility of differential expression (DE) results has rarely been studied. Here, we investigated the reproducibility of DE results for any given number of biological replicates between 3 and 24 and explored why a great many differentially expressed genes (DEGs) were not reproducible.

Results: Our findings demonstrate that poor reproducibility of DE results exists not only for small sample sizes, but also for relatively large sample sizes. Quite a few of the DEGs detected are specific to the samples in use, rather than genuinely differentially expressed under different conditions. Poor reproducibility of DE results is mainly caused by high variation of gene expression levels for the same gene in different samples. Even though biological variation may account for much of the high variation of gene expression levels, the effect of outlier count data also needs to be treated seriously, as outlier data severely interfere with DE analysis.

Conclusions: High heterogeneity exists not only in tumor tissue samples of each cancer type studied, but also in normal samples. High heterogeneity leads to poor reproducibility of DEGs, undermining generalization of differential expression results. Therefore, it is necessary to use large sample sizes (at least 10 if possible) in RNA-Seq experimental designs to reduce the impact of biological variability and DE results should be interpreted cautiously unless soundly validated.

Keywords: RNA sequencing, Differential expression, Heterogeneity, Reproducibility, Outlier, Tumor

Background

RNA-Seq has become an indispensable tool for transcriptome-wide analysis of differential gene expression in oncology to elucidate the mechanism of tumorigenesis and metastasis [1–3]. Due to the high cost [4] and the advantage of low technical variation [5–7] of RNA-Seq technology, many RNA-Seq experiments were performed with very small sample sizes, even with no replicates, but broader biological statements have been drawn on these experiments, discounting the influence of biological variability [8–10].

Extensive genetic intertumoral and intratumoral heterogeneity has long been recognized [11–14]. High genetic heterogeneity may greatly affect differentially expressed gene (DEG) detection in RNA-seq analysis and therefore undermine the reliability of differential expression (DE) results. However, the impact of tumor heterogeneity on the reliability of DE results obtained from RNA-seq data has rarely been studied.

* Correspondence: wql_zcq@126.com
†Weitong Cui and Huaru Xue contributed equally to the work and should be regarded as co-first authors.
[1]Key Laboratory of Biomedical Engineering & Technology of Shandong High School, Qilu Medical University, Zibo 255300, China
Full list of author information is available at the end of the article

Scientists of a biotechnology firm had tried to confirm published preclinical research findings related to their research, but they were shocked to find that the best-known scientific findings from cancer biology were confirmed in only 6 cases out of 53 [15, 16]. Poor reproducibility of ovarian cancer microRNA profiles has also been reported [17]. The findings above reveal the severity of the reproducibility problem in cancer research, which is probably caused by tumor heterogeneity. As drug development relies heavily on literatures, the problem of irreproducible data may increase the costs of drug development along with the number of late-stage clinical-trial failures [15]. Since RNA-Seq has been used extensively in cancer research, it is urgent to study the potential effect of tumor heterogeneity on the reliability of DE results in RNA-seq analysis.

Normally, it is arduous for researchers to verify their own or other people's findings due to the difficulty of sampling and limited budget. However, with the help of public large-scale projects which have plenty of samples, such as the Cancer Genome Atlas (TCGA) [18], reproducibility verification of DE results is possible. RNA-Seq data in TCGA database have been extensively employed in studies for understanding genetic changes in tumors [19–23].

In this work, the raw RNA-Seq count data for the three cancer types that have the most samples, namely breast cancer (BRCA), kidney renal clear cell carcinoma (KIRC), and lung adenocarcinoma (LUAD), were obtained from TCGA database. First, we investigated the reproducibility of DE results among the four repeated differential expression analysis, each using totally different samples, for any given number of biological replicates between 3 and 24. Then, we investigated the detection power depending on the number of biological replicates. Finally, we explored why a great many DEGs were not reproducible. All DE analyses were performed using edgeR [24], the most popular R package for DE analysis of RNA-Seq data [9]. The edgeR tool has been proved to have superior specificity and sensitivity as well as good control of false-positive errors [9, 25–27].

Results

Number of DEGs depending on the number of biological replicates

As shown in Fig. 1, just in terms of quantity, it seems that the more biological replicates used, the more DEGs will be identified. All the three curves in Fig. 1 show an increasing dynamic, but the rate of increase seems to diminish after around 10 biological replicates. It can also be inferred from the error bars that the number of DEGs for a given number of biological replicates generally differs greatly.

Reproducibility of DE results among the four repeats for a given number of biological replicates

As is shown in Fig. 2a, c, and e, for a given number of biological replicates, the number of reproducible DEGs is much less than the mean total number of DEGs, and the more repeats being performed, the lower the number of common DEGs becomes. The results indicate the poor reproducibility of DE results, which can be clearly seen from the changes of overlap rate in Fig. 2b, d, and f as well.

Both the number of common DEGs and the overlap rate increase with the elevated number of biological replicates, but the increasing rate slows down after around 10 biological replicates. For all three cancer types studied, the overlap rates among four repeats are all below 40% for the maximum number of biological replicates, and the percentage drops to below 10% for 3 biological replicates, which implies that the DE results for relatively large sample sizes are not reliable, and the reliability of DE results for small sample sizes are even poorer.

The evolution of power depending on the number of biological replicates

As it is difficult to choose one repeat to represent the four repeats for any given number of biological replicates, the common DEGs (intersection) and all detected DEGs (union) of the four repeats were used to calculate

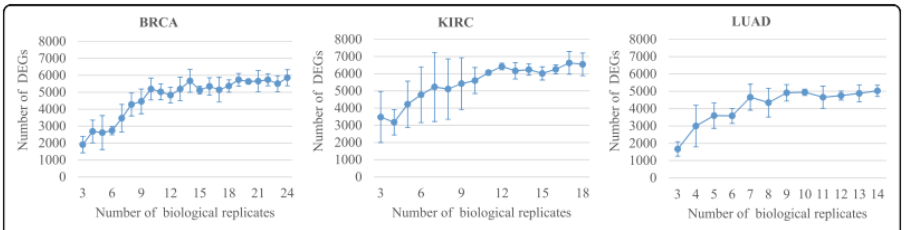


Fig. 1 The relationship between the mean number of DEGs and the number of biological replicates. The maximum biological replicate numbers vary depending on the total sample numbers for each cancer type in TCGA. The values represent the $M \pm SD$ of the number of DEGs for any given number of biological replicates

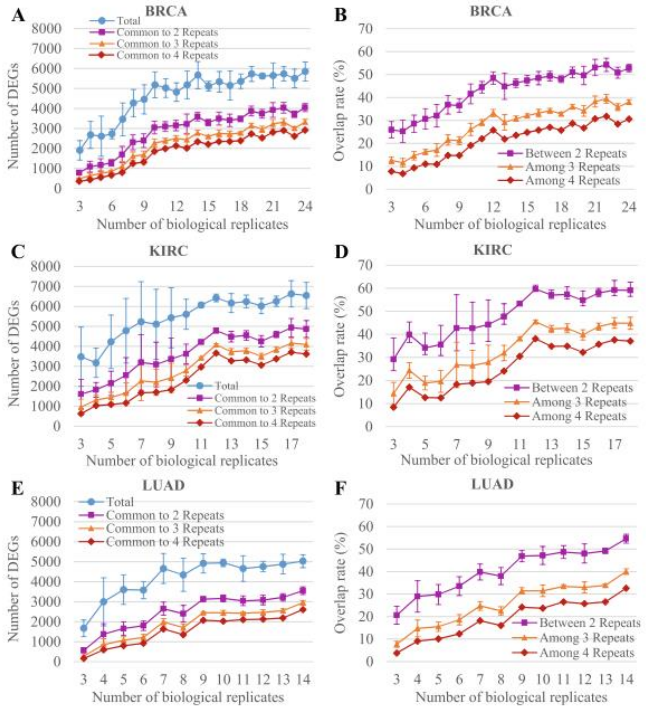


Fig. 2 Reproducibility of DE results among the four repeats for a given number of biological replicates. **a**, **c**, **e** The mean number of common DEGs for any two (purple line), three (orange line), or four (red line) repeats for each cancer type depending on the number of biological replicates, and the mean total number of DEGs for any given number of biological replicates (blue line) is also shown for reference. **b**, **d**, **f** The overlap rate of DE results for any two (purple line), three (orange line), or four (red line) repeats for each cancer type depending on the number of biological replicates

the power and intersection/union ratio (see the "Materials and methods" section).

As shown in Fig. 3a, b, and c, for each cancer type, when the number of biological replicates is between 3 and 10, both the number of DEGs and the power of the intersections grow rapidly, but the increasing rate is quite slow after about 10 biological replicates, which is similar to the trend of overlap rate in Fig. 2. As to the unions in Fig. 3d, e, and f, the number of DEGs and the power has similar trends with those of the intersections, but the two indicators reach plateaus faster than those of the intersections do.

The low power of DEG detection for small sample sizes can also be seen from the three curves in Fig. 3a, b, and c. For instance, the power of intersection for 3 biological replicates was below 16% for all three cancer types (as low as 6 % for LUAD), which means that more than 84% of the DEGs in the referential intersection cannot be detected using 3 biological replicates. Our findings clearly reveal that using more biological replicates is not only desirable but needed to improve the DE detection power using RNA-Seq.

As shown in Fig. 3a–f, both the intersection and the union for a given number of biological replicates contain some specific DEGs (i.e., DEGs that do not match with the reference), which means that the DEGs obtained using large sample sizes do not necessarily include all the DEGs obtained using small sample sizes.

As can be seen in Fig. 3g, h, and i, for any given number of biological replicates, the number of DEGs of the union is far larger than that of the intersection, which indicates that most of the DEGs detected in the four

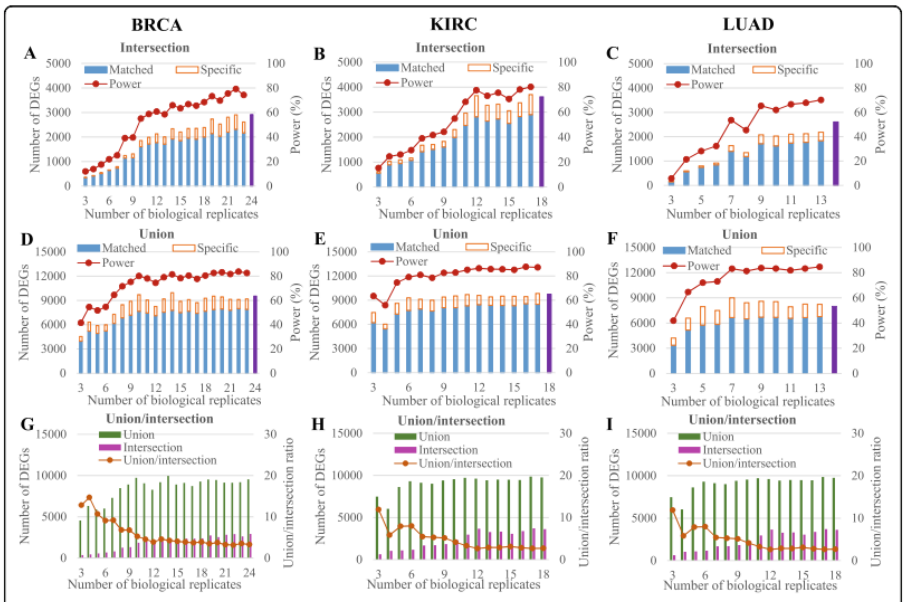


Fig. 3 Evolution of detection power and union/intersection depending on the number of biological replicates. **a–c** The number of DEGs and the power of the intersection for a given number of biological replicates for each cancer type. **d–f** show the number of DEGs and the power of the union for a given number of biological replicates for each cancer type. For all the stacked bars in charts **a–f**, the blue part represents the number of DEGs that match with the corresponding referential intersection or union, while the orange part represents the specific DEGs (not match with the corresponding reference) of the intersection or union. Purple color bars in charts **a–f** represent the number of DEGs in the referential intersections or unions for each cancer type. **g–i** The number of DEGs in the union (green bar) and intersection (pink bar), as well as the union/intersection ratio (orange line), for a given number of biological replicates for each cancer type

repeats for a given number of biological replicates are specific to the samples studied, rather than "true" DEGs for the tumor and normal tissues of each cancer type. This effect is much more intense for small sample sizes, which also reflects the poorer reproducibility of DE results obtained using small sample sizes.

Dispersion of normalized read counts for non-common DEGs

The results above demonstrate that a large number of DEGs for one repeat are not DEGs for another and these DEGs are referred to as non-common DEGs in this paper. Although non-common DEGs have also been found in previous literatures [9], the cause of the non-common DEGs has rarely been investigated before.

Ten non-common DEGs in repeat II and repeat III for 10 biological replicates in BRCA were used as examples to illustrate the phenomenon, as shown in Table 1. Among the ten genes, *IBSP*, *SGCG*, *DCT*, *APCDD1*, and *DPP6* were identified as DEGs only in repeat III, while

Table 1 Detailed %CV, Log_2FC, and FDR values for the 10 non-common DEGs in BRCA

Gene symbol	Repeat II				Repeat III			
	%CV		Log_2FC	FDR	%CV		Log_2FC	FDR
	N	T			N	T		
IBSP	122	167	− 8.14	0.09	115	115	− 6.71	3.80E− 12
SGCG	138	241	− 0.32	0.85	234	116	7.38	2.70E− 10
DCT	186	307	− 0.49	0.78	152	83	4.79	8.40E− 09
APCDD1	70	201	0.57	0.49	45	84	2.57	1.44E− 08
DPP6	71	300	− 0.61	0.68	54	143	3.91	4.62E− 08
SLC16A3	53	43	− 2.52	3.34E− 09	208	99	− 1.07	0.18
CDH23	83	55	2.79	1.12E− 08	91	155	1.38	0.07
FOXJ1	59	200	− 5.97	1.82E− 08	207	194	− 1.47	0.16
FGF10	63	91	3.27	2.25E− 08	90	178	0.38	0.76
BMP5	76	92	4.30	1.41E− 07	93	313	− 2.66	0.08

Capital letters "T" and "N" represent the tumor group and the normal group of each repeat, respectively. The numbers of biological replicates in either tumor groups or normal groups are 10. %CV indicates the percent coefficient of variation

SLC16A3, *CDH23*, *FOXJ1*, *FGF10*, and *BMP5* were identified as DEGs only in repeat II. The %CV, Log_2FC, and false discovery rate (FDR) values for the 10 non-common DEGs in KIRC and LUAD are shown in Supplementary Table S1 and S2, respectively.

As shown in Table 1, the values of FDR for these genes are all smaller than 2×10^{-7} when they are DEGs, which definitely means that they are, statistically, significant DEGs between the tumor and normal group, even if a threshold of 0.0001 is applied to control the FDR. Even so, the ten genes are not identified as DEGs in the other repeat. The result indicates that statistically significant DEGs are not as reliable as commonly believed.

In order to explore the reasons behind the non-common DEGs, we analyzed the dispersion of normalized read counts of these genes for some clues. As shown in Table 1, more than half of the %CVs are above 100. On the whole, there are many more %CVs over 100 in the tumor groups than in the normal groups, with three %CVs in the tumor groups even higher than 300, which probably means that gene expression levels in tumor groups have greater variability than those in normal groups. As CV is the ratio of the standard deviation to the mean, the high %CV reflects great dispersion of normalized read counts. In DE analysis, the high dispersion of read counts for a given gene can cause remarkable changes to the values of log_2FC and FDR and sometimes may even lead to opposite results. The high variation of expression levels for the same gene in different samples may be the main cause of the poor reproducibility of DE results.

In addition, we noticed in Table 1 that 4 out of the 10 non-common DEGs have opposite regulating trends in different repeats, i.e., upregulated in one repeat, but downregulated in the other, as demonstrated by the values of Log_2FC. By checking the Log_2FC values of the 3079 common DEGs between repeat II and III, we found that 35 of them (about 1.14%) also show opposite regulating trends, which indicates that the common DEGs are not reliable either.

It is clearly shown by the boxplots in Fig. 4 that outlier counts commonly exist in both the tumor and the normal groups, which is also true for the non-common DEGs in KIRC (Supplementary Figure S1) and LUAD (Supplementary Figure S2). Combining the read count dispersion in Fig. 4 with the %CV values in Table 1, we find that the high %CVs are mainly caused by the outlier counts, especially the extreme outliers. By excluding the outlier counts from analysis, 8 out of the 10 non-common DEGs become common DEGs, but the remaining 2 are still non-common DEGs, which implies that the problem of non-common DEGs can be partially

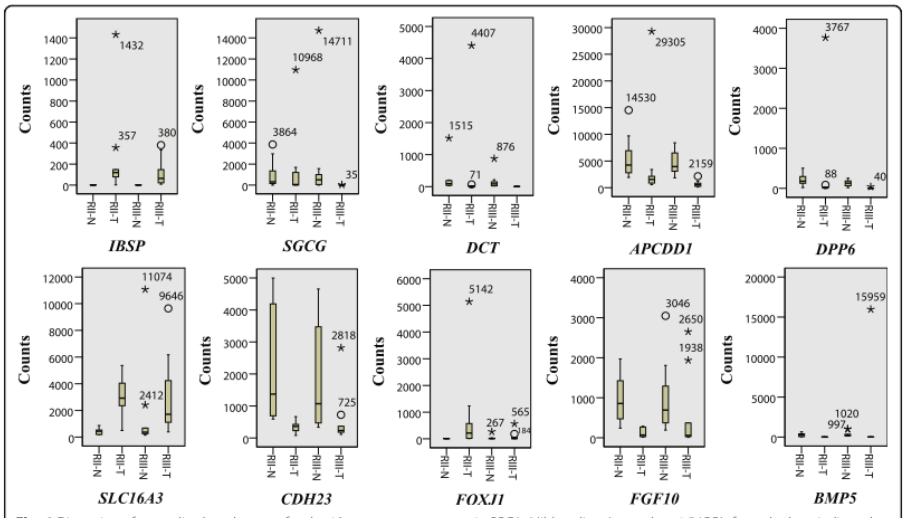


Fig. 4 Dispersion of normalized read counts for the 10 non-common genes in BRCA. Mild outliers (more than 1.5 IQR's from the box, indicated by a circle) and extreme outliers (more than 3 IQR's from the box, indicated by an asterisk) are shown. The number beside the marker shows the normalized read count value of the point. RII and RIII refer to repeat II and repeat III, respectively. Capital letters "T" and "N" represent the tumor group and the normal group, respectively. IQR indicates the interquartile range

resolved by excluding the outliers from analysis. We also confirmed that the opposite regulating trends of the 4 genes described above can be corrected by excluding outlier counts from analysis.

Discussion

The results above are based on the RNA-Seq data of human tumor and adjacent normal samples. Nevertheless, the issue of low reliability of DEGs for very small sample sizes has also been found in studies using RNA-Seq data from mouse [28, 29], *Saccharomyces cerevisiae* [10], and tomato [9], which implies that the problem of reproducibility is common in RNA-seq analysis.

The maximum number of biological replicates studied here is already much larger than recommended in the literatures [9, 10], but still a large proportion of the DEGs detected are irreproducible. While results obtained using 3 biological replicates for each condition in experimental designs are generally accepted as reliable, our results show quite the opposite, at least in cancer research using RNA-Seq.

As shown in Table 1 and Fig. 4, outlier counts, especially the extremes, account for much of high variation of expression levels. It should be emphasized that outlier counts are commonly scattered in different samples, rather than focused in one sample, in which case the outlier counts can be eliminated by excluding the aberrant sample. As DE analyses with small sample sizes are more susceptible to outliers, the poor reproducibility of DE results for small sample sizes is understandable.

The authenticity of outlier counts is beyond the scope of this research. Nevertheless, figuring out whether the extreme counts are true or not is the prerequisite to properly deal with them. The popular edgeR [24] takes raw read counts as input and provides its own normalization approach [30] but does not handle the outlier counts. Given the enormous influence of outliers on DE analysis, the problem of outlier counts should be properly addressed in future versions of edgeR.

Since low technical variation is one of RNA-Seq's potential advantages [5–7], most of the variations might be attributed to biological variations which can be reflected in extensive genetic intertumoral and intratumoral heterogeneity [11–14]. Biological variation, unlike measurement error, cannot be reduced with technology improvements, but can only be measured by considering expression measurements taken from multiple biological samples within the same group [8]. Therefore, large sample sizes should be considered when designing RNA-Seq experiments for DGE detection to reduce the effect of biological variability. However, based on our findings, it is impossible to determine an optimal number of biological replicates which can guarantee all detected DEGs are reliable for a given RNA-Seq experiment, but approximately at least 10 replicates per condition should be used to achieve relatively high reproducibility and detection power.

One goal of DE analysis in cancer research by RNA-Seq is to identify and catalog expression of new or alternative transcripts between tumor and normal tissues, which is essential for understanding the mechanism of tumorigenesis and developing effective therapies. Apparently, given the high heterogeneity of tumor and normal samples, it is hard to achieve that goal using small sample sizes, let alone with no biological replicates. Moreover, as demonstrated by our findings, incorporating a relatively larger sample size than recommended for RNA-Seq experiments in previous literatures [9, 27] does not mean the DE results are fully credible.

Conclusions

In conclusion, both tumor tissue samples and normal tissue samples show high heterogeneity. DE results of small sample sizes are more susceptible to heterogeneity, compared with those of large sample sizes. As a result, reproducibility of DE results and DEG detection power for small sample sizes are far lower than those for large sample sizes. Even if large sample sizes are utilized, a large proportion of the detected DEGs are irreproducible. Therefore, large sample sizes (at least 10 if possible) should be considered in RNA-Seq experimental designs to reduce the interfering effect of sample heterogeneity and DEGs of interest should be validated before making generalized statements.

Similarly, since it is difficult to distinguish which DEGs are specific to the samples in use and which are common to the studied populations, DE results from published RNA-Seq literatures, especially those with very small sample sizes or no biological replicates, should be consulted with caution. With regard to the reproducibility crisis which is particularly severe in cancer biology [15, 16] and has remarkably hindered the translation of cancer research to clinical success [31], much remains to be done to discern the DEGs caused by biological variability and to improve the reproducibility of DE results.

Materials and methods

Raw count data collection

Raw RNA-Seq read count data for all available BRCA, KIRC, and LUAD tumors and available adjacent normal tissues were downloaded from The Cancer Genome Atlas (TCGA) database. To ensure sample consistency, data from metastatic or formalin-fixed paraffin-embedded tissue samples [32, 33], as well as repeated data for the same samples, were excluded. After exclusion, total numbers of tumor and normal tissue samples included in BRCA, KIRC, and LUAD datasets were 1177, 610, and 592, respectively.

DE analysis of the collected raw count data

This work was designed to investigate the evolution of reproducibility of DE results and the detection power depending on the number of biological replicates *n*. Although there are algorithms specially developed for DE analysis of RNA-Seq data without biological replicates [34–36], the results obtained are debatable as it is impossible to estimate the level of biological variability. If there are only two biological replicates, it is difficult to detect an outlier (bad) expression value. Therefore, the minimum *n* was set at 3 for each cancer type.

For each value of *n*, four sets of *n* tumor samples and *n* normal samples were randomly chosen without replacement from datasets of each cancer type to simulate four different experimental repeats, which were denoted as repeat I, II, III, and IV, respectively, for ease of description. For any given *n*, samples in the four repeats were all different. Limited by the number of normal samples in BRCA, KIRC, and LUAD datasets (i.e., 99, 72, and 58, respectively), the maximum *n* for BRCA, KIRC, and LUAD was accordingly set at 24, 18, and 14, respectively. The sampling process was shown in Supplementary Figure S3. Raw read count data of samples in each set of tumor and normal groups were used to construct gene expression matrices for subsequent analyses.

All DE analyses were done with R software (version 3.5.3) and the edgeR package [24] (version 3.22.5). Trimmed-mean *M* values (TMM) normalization was performed to normalize the counts among the different samples [37–40]. As high dispersion of low counts interfered with some of the statistical approximations used in edgeR, genes with low counts were filtered out using the *filterByExpr* function as recommended in the user's guide. Genes were marked as DEGs if the absolute value of $\log_2$ transformed fold change ($\log_2$FC) ≥ 1 and the false discovery rate (FDR) < 0.05.

Reproducibility of DE results among the four repeats for a given number of biological replicates

As described above, four repeated DE analyses were performed for each number of biological replicates *n*; therefore, four lists of DEGs were obtained for each *n*. To analyze the reproducibility of DE results, we compared the four lists of DEGs in terms of overlap rate which was defined as the ratio of the number of common DEGs (i.e., DEGs that were common to the compared repeats) to the total number of DEGs of the corresponding repeats. For instance, the total number of DEGs identified in repeat II and III for 10 biological replicates in BRCA was 6528, 3079 of which are common to both of the two repeats; therefore, the overlap rate of the DE results for the two repeats was 47.17%. The overlap rate of DE results for any two, three, or four repeats for a given *n* was computed in the same way as exemplified above. Overlap rate was calculated using VENNY (version 2.1.0) (Oliveros, J.C. (2007–2015) Venny; an interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Power analysis for a given number of biological replicates

We have four lists of DEGs for each *n*, but the number of DEGs and DEG composition of the four lists are quite different, so it is difficult to choose one of the lists as a representative. Therefore, the intersection (i.e., DEGs that are common to all four repeats) and the union (i.e., all DEGs identified for all four repeats) for any given *n* were used for power analysis. Note that the power was defined as the ability for a given sample size to detect "true" DEGs. Obviously, we needed a reference list of "true" DEGs. As is generally accepted that results obtained using larger sample sizes are more robust, the intersection and union of the maximum *n* for each cancer type were used as references.

The power was calculated by the ratio of the number of DEGs in the intersection (or union) for any given *n* to the number of DEGs in the corresponding referential intersection (or union). The ratio of the number of DEGs in the union to that in the intersection for any given *n* was also calculated and marked as union/intersection.

Read count dispersion analyses for non-common DEGs

The non-common DEGs, as opposed to the common DEGs, were the DEGs that could be identified in one repeat, but not in another. To explore the cause of non-common DEGs, we selected 10 non-common DEGs from the DEG lists of repeat II and repeat III for 10 biological replicates in BRCA dataset and analyzed the characteristics of raw count data of these genes. Although the number of DEGs was close between the two repeats, about 53% of the DEGs were non-common DEGs.

In order to eliminate the interference of different sequencing depth, TMM normalized read counts were used for analysis. The percent coefficient of variation (%CV) of normalized read counts in both tumor and normal groups of repeat II and repeat III was calculated for each non-common DEG. Similarly, dispersion of normalized read counts was analyzed and displayed in boxplots using IBM SPSS Statistics (19.0). The read count dispersion analyses for the non-common DEGs in KIRC and LUAD were conducted in the same way as in BRCA.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40246-021-00308-5.

Additional file 1: Supplementary Table S1. Detailed %CV, Log_2FC, and FDR values for the 10 non-common DEGs in KIRC

Additional file 2: Supplementary Table S2. Detailed %CV, Log_2FC, and FDR values for the 10 non-common DEGs in LUAD

Additional file 3: Supplementary Figure S1. Dispersion of normalized read counts for the 10 non-common genes in KIRC. Mild outliers (more than 1.5 IQR's from the box, indicated by O) and extreme outliers (more than 3 IQR's from the box, indicated by *) are shown. The number beside the marker shows the normalized count value of the point. RII and RIII refer to repeat II and repeat III, respectively. Capital letters "T" and "N" represent the tumor group and the normal group, respectively. IQR indicates the interquartile range.

Additional file 4: Supplementary Figure S2. Dispersion of normalized read counts for the 10 non-common genes in LUAD. Mild outliers (more than 1.5 IQR's from the box, indicated by O) and extreme outliers (more than 3 IQR's from the box, indicated by *) are shown. The number beside the marker shows the normalized count value of the point. RII and RIII refer to repeat II and repeat III, respectively. Capital letters "T" and "N" represent the tumor group and the normal group, respectively. IQR indicates the interquartile range.

Additional file 5: Supplementary Figure S3. Diagram of experimental design for the BRCA dataset. The processes of sampling and analysis for the KIRC and LUAD datasets were the same as that of BRCA. The numbers of tumor tissue samples and adjacent normal tissue samples for KIRC were 526 and 72, respectively, while the two numbers were 509 and 58, respectively, for LUAD. Restrained by the number of normal tissue samples which was far less than the number of tumor samples for each cancer type, the maximum number of biological replicates for BRCA, KIRC, and LUAD was accordingly set at 24, 18, and 14, respectively.

Abbreviations

BRCA: Breast cancer; CV: Coefficient of variation; DE: Differential expression; DEGs: Differentially expressed genes; FC: Fold change; FDR: False discovery rate; IQR: Interquartile range; KIRC: Kidney renal clear cell carcinoma; LUAD: Lung adenocarcinoma; RNA-seq: RNA sequencing; SD: Standard deviation; TCGA: The Cancer Genome Atlas; TMM: Trimmed-mean *M* values

Acknowledgements

The results shown here are in whole based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Authors' contributions

QW, WC, and HX conceived and designed the experiments. WC, HX, and LW performed statistical analyses and analyzed the data. WC and HX wrote the paper. JJ, XT, and QW advised on the analyses and revised the manuscript. The authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of Shandong Province, China (grant No. ZR2019PC053), the National Natural Science Foundation of China (grant No. 31701042), the Key Research and Development Program of Zibo (grant No. 2018kj040045), and the Zibo Platform for Gene Editing and Cell Application (grant No. 2018ZBXC008).

Availability of data and materials

The datasets supporting the conclusions of this article are available in TCGA repository, https://www.cancer.gov/tcga.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Key Laboratory of Biomedical Engineering & Technology of Shandong High School, Qilu Medical University, Zibo 255300, China. [2]Environmental Protection Research Institute of Light Industry, Beijing 100089, China. [3]Shandong Sport University, Jinan 250102, China.

Received: 26 June 2020 Accepted: 19 January 2021
Published online: 28 January 2021

References

1. Wang E, Zou J, Zaman N, Beitel LK, Trifiro M, Paliouras M. Cancer systems biology in the genome sequencing era: part 1, dissecting and modeling of tumor clones and their networks. Semin Cancer Biol. 2013;23:279–85.
2. Wang E, Zou J, Zaman N, Beitel LK, Trifiro M, Paliouras M. Cancer systems biology in the genome sequencing era: part 2, evolutionary dynamics of tumor clonal networks and drug resistance. Semin Cancer Biol. 2013;23:286–92.
3. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. 2019;20:631–56.
4. Hitzemann R, Bottomly D, Darakjian P, Walter N, Iancu O, Searles R, et al. Genes, behavior and next-generation RNA sequencing. Genes Brain Behav. 2013;12:1–12.
5. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18:1509–17.
6. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621–8.
7. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;11:94.
8. Hansen KD, Wu Z, Irizarry RA, Leek JT. Sequencing technology does not eliminate biological variability. Nat Biotechnol. 2011;29:572–3.
9. Lamarre S, Frasse P, Zouine M, Labourdette D, Sainderichin E, Hu G, et al. Optimization of an RNA-seq differential gene expression analysis depending on biological replicate number and library size. Front Plant Sci. 2018;9:108.
10. Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016;22:839–51.
11. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 1805;2010:105–17.
12. Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. Mol Oncol. 2014;8:1095–111.
13. Wei L, Wang J, Lampert E, Schlanger S, DePriest AD, Hu Q, et al. Intratumoral and intertumoral genomic heterogeneity of multifocal localized prostate cancer impacts molecular classifications and genomic prognosticators. Eur Urol. 2017;71:183–92.
14. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15:81–94.
15. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. Nature. 2012;483:531–3.
16. Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016;533:452–4.
17. Wan YW, Mach CM, Allen GI, Anderson ML, Liu Z. On the reproducibility of TCGA ovarian cancer microRNA profiles. PLoS One. 2014;9:e87782.
18. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19:A68–77.
19. Zhang Y, Sui J, Shen X, Li C, Yao W, Hong W, et al. Differential expression profiles of microRNAs as potential biomarkers for the early diagnosis of lung cancer. Oncol Rep. 2017;37:3543–53.
20. Choi W, Ochoa A, McConkey DJ, Aine M, Hoglund M, Kim WY, et al. Genetic alterations in the molecular subtypes of bladder cancer: illustration in the Cancer Genome Atlas Dataset. Eur Urol. 2017;72:354–65.
21. Goovaerts T, Steyaert S, Vandenbussche CA, Galle J, Thas O, Van Criekinge W, et al. A comprehensive overview of genomic imprinting in breast and its deregulation in cancer. Nat Commun. 2018;9:4120.
22. Bakhoum MF, Esmaeli B. Molecular characteristics of uveal melanoma: insights from the Cancer Genome Atlas (TCGA) Project. Cancers (Basel). 2019;11.
23. Zhao B, You Y, Wan Z, Ma Y, Huo Y, Liu H, et al. Weighted correlation network and differential expression analyses identify candidate genes associated with BRAF gene in melanoma. BMC Med Genet. 2019;20:54.

24. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
25. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform. 2015; 16:59–70.
26. Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. 2013;14:R95.
27. Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, et al. Erratum: how many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016;22:1641.
28. Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics. 2013;14:91.
29. Burden CJ, Qureshi SE, Wilson SR. Error estimates for the analysis of differential expression from RNA-seq count data. PeerJ. 2014;2:e576.
30. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.
31. Hutchinson L, Kirk R. High drug attrition rates--where are we going wrong? Nat Rev Clin Oncol. 2011;8:189–90.
32. Esteve-Codina A, Arpi O, Martinez-Garcia M, Pineda E, Mallo M, Gut M, et al. A comparison of RNA-seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. PLoS One. 2017; 12:e0170632.
33. Kwong LN, De Macedo MP, Haydu L, Joon AY, Tetzlaff MT, Calderone TL, et al. Biological validation of RNA sequencing data from formalin-fixed paraffin-embedded primary melanomas. JCO Precis Oncol. 2018;2018.
34. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011;21:2213–23.
35. Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics. 2012;28:2782–8.
36. Low JZB, Khang TF, Tammi MT. CORNAS: coverage-dependent RNA-Seq analysis of gene expression data without biological replicates. BMC Bioinformatics. 2017;18:575.
37. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.
38. Maza E, Frasse P, Senin P, Bouzayen M, Zouine M. Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments: a matter of relative size of studied transcriptomes. Commun Integr Biol. 2013;6:e25849.
39. Maza E. In papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-Seq experimental design. Front Genet. 2016;7:164.
40. Li X, Cooper NGF, O'Toole TE, Rouchka EC. Choice of library size normalization and statistical methods for differential gene expression analysis in balanced two-group comparisons for RNA-seq studies. BMC Genomics. 2020;21:75.

Publisher's Note

4.8.3 Effect of high variation in transcript expression on identifying differentially expressed genes in RNA-seq analysis（王清路）

Received: 7 February 2021 | Revised: 4 July 2021 | Accepted: 15 July 2021

DOI: 10.1111/ahg.12441

ORIGINAL ARTICLE

Annals of human genetics WILEY

Effect of high variation in transcript expression on identifying differentially expressed genes in RNA-seq analysis

Weitong Cui[1] | **Huaru Xue**[1] | **Yifan Geng**[1,2] | **Jing Zhang**[1] | **Yajun Liang**[1] | **Xuewen Tian**[3] | **Qinglu Wang**[1,3]

[1] Key Laboratory of Biomedical Engineering & Technology of Shandong High School, Qilu Medical University, Zibo, P. R. China

[2] Xuzhou Medical University, Xuzhou, P. R. China

[3] Shandong Sport University, Jinan, P. R. China

Correspondence
Qinglu Wang, Key Laboratory of Biomedical Engineering & Technology of Shandong High School, Qilu Medical University, Zibo, 255300, P. R. China.
Email: wql_zcq@126.com

Weitong Cui and Huaru Xue contributed equally to the study and should be regarded as cofirst authors.

Funding information
the Natural Science Foundation of Shandong Province, China, Grant/Award Number: ZR2019PC053; the Key Research and Development Program of Zibo, Grant/Award Number: 2018kj040045; the Zibo City Integration Development Project, Grant/Award Number: 2019ZBXC320; the Taishan Scholars Program of Shandong Province, Grant/Award Number: tsqn201909148; the National Natural Science Foundation of China, Grant/Award Number: 31701042.

Summary

Great efforts have been made on the algorithms that deal with RNA-seq data to enhance the accuracy and efficiency of differential expression (DE) analysis. However, no consensus has been reached on the proper threshold values of fold change and adjusted *p*-value for filtering differentially expressed genes (DEGs). It is generally believed that the more stringent the filtering threshold, the more reliable the result of a DE analysis. Nevertheless, by analyzing the impact of both adjusted *p*-value and fold change thresholds on DE analyses, with RNA-seq data obtained for three different cancer types from the Cancer Genome Atlas (TCGA) database, we found that, for a given sample size, the reproducibility of DE results became poorer when more stringent thresholds were applied. No matter which threshold level was applied, the overlap rates of DEGs were generally lower for small sample sizes than for large sample sizes. The raw read count analysis demonstrated that the transcript expression of the same gene in different samples, whether in tumor groups or in normal groups, showed high variations, which resulted in a drastic fluctuation in fold change values and adjusted*p*-values when different sets of samples were used. Overall, more stringent thresholds did not yield more reliable DEGs due to high variations in transcript expression; the reliability of DEGs obtained with small sample sizes was more susceptible to these variations. Therefore, less stringent thresholds are recommended for screening DEGs. Moreover, large sample sizes should be considered in RNA-seq experimental designs to reduce the interfering effect of variations in transcript expression on DEG identification.

KEYWORDS
Differential expression, false discovery rate, fold change, RNA-seq, sample size, threshold

Abbreviations: BRCA, breast cancer; DE, differential expression; DEGs, differentially expressed genes; FC, fold change; FDR, false discovery rate; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; RNA-seq, RNA sequencing; SD, standard deviation; TCGA, The Cancer Genome Atlas; TMM, trimmed-mean M values.

1 | INTRODUCTION

RNA-seq is currently the most popular technology for cataloging and comparing genome-wide gene expression and has been extensively applied in identifying DEGs under different conditions (Stark et al., 2019, Ozsolak & Milos, 2011, Wang et al., 2009). Generally, both expression level ratio (fold change [FC]) and statistical significance (*p*-value) should be considered to determine whether a gene is differentially expressed (Poplawski & Binder, 2018, Schurch et al., 2016). FC is usually base 2 logarithm transformed (denoted as $\log_2$FC) to treat up- and down-regulated genes in a similar fashion (Quackenbush, 2002). Note that FC has no meaning in a statistical point of view and is a kind of "biological" threshold. If multiple testing is involved, as in RNA-seq analysis, the adjusted *p*-value or false discovery rate (FDR) should be applied to control the number of false discoveries (Sampson et al., 2018).

Although $\log_2$FC and FDR are commonly used to filter DEGs in RNA-seq experiments, no fixed values exist for the two parameters. The most commonly used cutoff value for FC is 2 (i.e., $|\log_2 FC| \geq 1$), and 0.05 for FDR (i.e., FDR < 0.05)(Poplawski & Binder, 2018, Lamarre et al., 2018). However, one can empirically set the cutoff values of the two parameters based on their particular results (Robles et al., 2012, Al Seesi et al., 2014, Schurch et al., 2016, Poplawski & Binder, 2018, Yin et al., 2020). If too many DEGs are detected using the common filtering threshold, more stringent thresholds can be applied to reduce the number of DEGs. Similarly, the threshold can be lowered if too few DEGs are detected.

It is generally believed that the more stringent the threshold, the more reliable the results. However, the aforementioned hypothesis should be established on the premise that gene expression patterns are consistent between individuals of the same group. In fact, true gene expression levels vary among individuals because gene expression is inherently a stochastic process and may be influenced by many factors such as age, sex, nutrition, temperature, chemicals, and infectious agents (Kim & Marioni, 2013, Jia et al., 2019). In cancer research, intertumoral and intratumoral heterogeneity, which has become the main obstacle in treating malignant diseases, has long been recognized (Marusyk & Polyak, 2010, Burrell & Swanton, 2014, Wei et al., 2017, Dagogo-Jack & Shaw, 2018). Nevertheless, the impact of sample heterogeneity on DE analysis has not been taken seriously.

Although RNA-seq technology provides various advantages over microarray (Hansen et al., 2011, Ozsolak & Milos, 2011), issues related to experimental design and data analysis in microarray also exist in RNA-seq. Many researchers have focused on the issues of RNA-seq experimental design and subsequent statistical analysis methods such as the number of biological replicates (Ching et al., 2014, Liu et al., 2014, Gierlinski et al., 2015, Schurch et al., 2016, Lamarre et al., 2018, Poplawski & Binder, 2018), normalization methods (Li et al., 2012, Dillies et al., 2013, Li & Tibshirani, 2013, Yu et al., 2013, Lin et al., 2016, Maza, 2016), and statistical tools (Soneson & Delorenzi, 2013, Zhang et al., 2014, Moulos & Hatzis, 2015, Conesa et al., 2016, Rapaport et al., 2013, Seyednasrollah et al., 2015, Lin & Pang, 2019), but few have paid attention to the effect of different threshold values on DEG identification.

Poor reproducibility of DEGs has been shown in several studies using different datasets from tomato plants (Lamarre et al., 2018), yeast (Schurch et al., 2016), mouse strains (Soneson & Delorenzi, 2013), and human tissues(Cui et al., 2021), and cell lines (Liu et al., 2014); however, the findings of all these studies were based on DE analyses using the common threshold value for FC ($|\log2FC| \geq 1$) and FDR (FDR < 0.05). The question was whether the reproducibility of DE results could be improved by increasing significance stringency. To investigate whether more stringent thresholds could yield more reliable results in RNA-seq analysis, the present study analyzed the impact of both FDR and LFC thresholds on DE analyses, using RNA-seq read count data of breast cancer (BRCA), kidney renal clear cell carcinoma (KIRC), and lung adenocarcinoma (LUAD) obtained from TCGA (Tomczak et al., 2015). The total number of samples for BRCA, KIRC, and LUAD was 1177, 598, and 567, respectively. Additionally, the high transcript expression variability was illustrated by showing gene ranks and normalized read counts of top 10 DEGs in the four DE analyses of BRCA with 20 samples per condition. The objective of this study was to explore the relationship between threshold stringency and reliability of DE results, and help RNA-seq practitioners choose an appropriate threshold for DEG identification and better understand and interpret DE results.

2 | MATERIALS AND METHODS

2.1 | Raw count data collection

BRCA, KIRC, and LUAD are the three cancer types with the largest number of samples in TCGA database, which is especially advantageous in demonstrating variations in transcript expression. Raw RNA-seq read count data of all available BRCA, KIRC, and LUAD tumor and adjacent normal tissues were obtained from TCGA. Data from metastatic or formalin-fixed paraffin-embedded tissue samples (Esteve-Codina et al., 2017, Kwong et al., 2018), as well as repeated data for the same samples sequenced more than once, were excluded to ensure

sample consistency. After exclusion, the number of tumor samples in BRCA, KIRC, and LUAD datasets was 1078, 526, and 509, respectively, while the corresponding number of normal samples was 99, 72, and 58, respectively. The batch effects in all datasets were checked and corrected using the Empirical Bayes method (http://bioinformatics.mdanderson.org/tcgambatch). The FPKM data of all available BRCA samples were also obtained from TCGA. Single cell RNA-seq data for naïve HeLa cells (90 replicates) and priming HeLa cells (89 replicates) induced with interferon gamma were obtained from the GEO database (GSE150198), and a parallel analysis was performed on the cell line with the two conditions (Naïve vs. Priming).

2.2 | DE analysis using different threshold levels

This study examined the relationship between different threshold levels and the reliability of DE results depending on the number of biological replicates n per condition. Estimating the level of biological variability with only one biological replicate was impossible, and detecting an outlier (bad) expression value was difficult if only two biological replicates were present. Hence, the minimum n was set at 3.

For each value of n, four sets of n tumor samples and n normal samples were randomly selected without replacement from each dataset to simulate four different experimental repeats. For any given n, the samples in the four repeats were all different. The raw read count data of samples in each set of tumor and normal groups were used to construct gene expression matrices for subsequent DE analyses. For all the three cancer types studied, as well as the other cancer types in TCGA, the number of normal samples was far less than the number of tumor samples. Restrained by the sampling method and the number of normal samples, the maximum n for BRCA, KIRC, and LUAD was accordingly set at 24, 18, and 14, respectively.

All DE analyses were done with the R software (version 3.5.3) and the edgeR package (Robinson et al., 2010) (version 3.22.5), which is the most popular tool for DE analysis of RNA-seq data (Lamarre et al., 2018) and has superior specificity and sensitivity as well as effective control of false positive errors (Rapaport et al., 2013, Seyednasrollah et al., 2015, Lamarre et al., 2018, Schurch et al., 2016). Trimmed-mean M values (TMM) normalization was performed to normalize the counts among the different samples (Robinson & Oshlack, 2010, Maza, 2016, Li et al., 2020). Three different combinations of FDR and $\log_2$FC values, namely, (FDR < 0.05, $|\log_2 FC| \geq 1$), (FDR < 0.01, $|\log_2 FC| \geq 2$), and (FDR < 0.001, $|\log_2 FC| \geq 4$), were used as thresholds to screen DEGs. The three thresholds were denoted as (FDR(0.05), LFC(1)), (FDR(0.01), LFC(2)), and (FDR(0.001), LFC(4)), respectively, for convenience.

2.3 | Testing reliability of DE results for different threshold levels

The overlap rate was used as an indicator to reflect the reliability of DE results for different threshold levels. As described earlier, four repeated DE analyses were performed for each number of biological replicates n and each threshold level; therefore, four lists of DEGs were obtained. To analyze the reliability of DE results among the four repeats for each n and each threshold level, the four lists of DEGs were compared in terms of the overlap rate, which was defined as the ratio of the number of overlap DEGs (i.e., DEGs that were common to all four repeats) to the total number of DEGs detected in the four repeats. The overlap rate was analyzed using "vennCounts" and "vennDiagram" functions from the limma package. Analysis of sensitivity and specificity was carried out according to Lamarre et al. (2018). The R codes used in the study can be found in the Supporting information.

3 | RESULTS

3.1 | Effect of increasing significance stringency on DEG screening

The mean number of DEGs detected using different threshold levels for each cancer type is shown in Figure 1. For all values of n, when a more stringent threshold is adopted, the number of DEGs drops dramatically. It is worth noting that the trends of the three curves for different threshold levels are inconsistent. When the lowest threshold (FDR(0.05), LFC(1)) is applied, the mean number of DEGs first increases rapidly for the number of biological replicates between 3 and 10, and then remains relatively stable. The curves for (FDR(0.01), LFC(2)) also show mild upward trends when $n < 10$. When applying the most stringent threshold (FDR(0.001), LFC(4)), the curves are almost flat, implying that the number of DEGs for different sample sizes is close.

To explore the possible interactions between the two thresholds of FDR and FC, the effect of each threshold on DEG identification was also investigated separately. Supporting information Figure S1 shows the average number of DEGs for different FC thresholds (LFC(1), LFC(2), and LFC(4)), with a fixed threshold of 0.05 to control the FDR, while Supporting information Figure S2 displays the average number of DEGs for different FDR

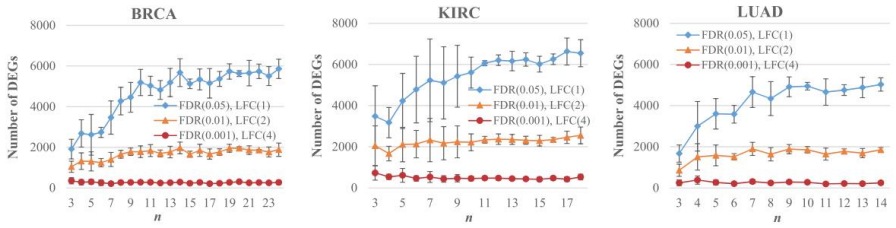


FIGURE 1 **Mean (±SD) number of DEGs detected using different threshold levels for the three cancer types**. The values represent the mean (±SD) number of DEGs of four independent DE analyses for each number n of biological replicates per condition. As the total numbers of samples in the three datasets are different, the maximum n varies depending on the number of normal samples for each cancer type. FDR and LFC refer to the adjusted p-value and $\log_2$FC, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

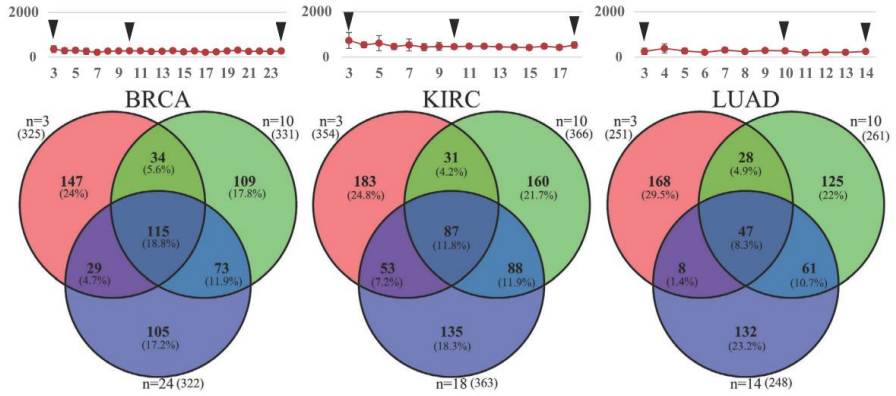


FIGURE 2 Venn diagrams of DEGs filtered by (FDR(0.001), LFC(4)) for $n = 3$, 10, and n(max) for each cancer type [Colour figure can be viewed at wileyonlinelibrary.com]

thresholds (FDR(0.05), FDR(0.01), and FDR(0.001)), with a fixed threshold of 1 to control the absolute value of $\log_2$FC. As shown in Figure 1, Supporting information Figures S1 and S2, the threshold of FC clearly plays a key role in DEG identification by the three different combinations of FC and FDR thresholds used in this study.

To investigate whether the DEGs detected by (FDR(0.001), LFC(4)) were the same for different sample sizes, the DEG lists of different sample sizes were compared for each cancer type. As the tables produced are too large and obscure (the vennCounts results for each cancer type can be found in Supporting information Tables S1–S3), $n = 3$, 10, and n (max) are selected to represent small, medium, and relatively large sample sizes, respectively, and the overlapping DEGs among the three sample sizes for each cancer type are shown using Venn diagrams. As shown in Figure 2, for each cancer type, the percentages of overlap DEGs among the three sample sizes are all less than 20%, indicating that although the number of DEGs for different sample sizes is close, DEG compositions vary greatly, consistent with the results shown in Supporting information Tables S1–S3.

3.2 | Reliability of DE results for different threshold levels

The overlap rate was employed as the indicator to reflect the reliability of DE results (see Materials and Methods). For each value of biological replicates n, four DE analyses using totally different sets of tumor and normal samples were performed, resulting in four lists of DEGs. The

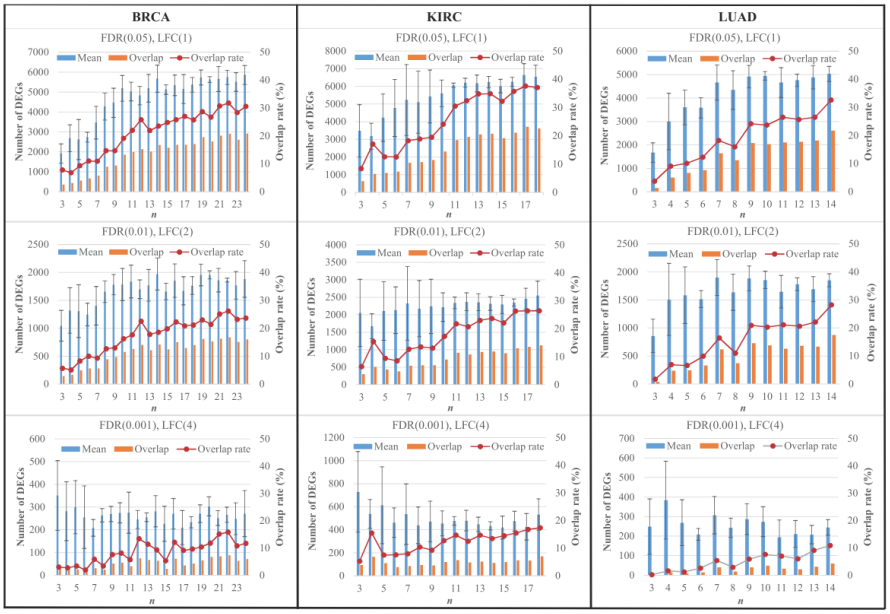


FIGURE 3 Evolution of overlap rates depending on biological replicate numbers for different threshold levels. The mean number of DEGs and the number of overlap DEGs for each *n* are also shown [Colour figure can be viewed at wileyonlinelibrary.com]

reliability of DE results was analyzed by comparing the four lists of DEGs for a given sample size.

The study examined the effect of increasing significance stringency on the overlap rate of the four lists of DEGs for each *n*, and the result is shown in Figure 3. The three curves of the overlap rate for different threshold levels all have similar trends, that is, low overlap rate for small sample sizes and high overlap rate for relatively large sample sizes. Apparently, no matter which threshold is used, the mean number of DEGs far exceeds the number of overlap DEGs for each *n*. More importantly, the overlap rates for all values of *n* gradually decrease with increasing significance stringency. When the most stringent threshold (FDR(0.001), LFC(4)) is applied, the overlap rates for the *n* (max) shockingly drop to less than 20%. Similar results are obtained when each threshold is investigated separately (Supporting information Figures S3 and S4), but the decrease of the overlap rate for different values of *n* with the raising stringency of FC threshold in Supporting information Figure S3 is more dramatic than that in Supporting information Figure S4 with the raising stringency of FDR threshold. Although the results demonstrate that the values of the overlap rate decrease with increasing significance stringency, RNA-seq practitioners should know that DEGs identified using a particular dataset are correct for the samples studied.

3.3 | Reliability of DEGs with the most dramatic changes in mRNA expression

Generally, DEGs with the most dramatic changes in expression are the primary concern; however, based on the aforementioned results, these genes might have poor reproducibility. To prove this point, the DEG compositions (Table 1) of the top 10 up- and down-regulated DEGs were also analyzed among the four repeats for $n = 20$ in BRCA. As can be observed in Table 1, different sets of samples result in different DEG lists. No common DEG exists among the four repeats, and the overlap DEGs between any two repeats are scarce. The poor reproducibility of the DEGs shown in Table 1 indicates that the top up- and down-regulated DEGs identified in one particular study are highly unreliable.

As the amount of RNA-seq data is huge, the process of DE analysis is usually automated. Researchers always

TABLE 1 Top 10 up- and down-regulated DEGs in each repeat for $n = 20$ in BRCA

Repeat I			Repeat II			Repeat III			Repeat IV		
Gene	**Log_2FC**	**FDR**	**Gene**	**Log_2FC**	**FDR**	**Gene**	**Log_2FC**	**FDR**	**Gene**	**Log_2FC**	**FDR**
MYOC	9.28	5.39E-19	*MYOC*	8.76	2.51E-22	*CSN1S1*	7.82	4.37E-08	*MYH2*	15.02	3.27E-13
PENK	8.41	1.51E-17	*LEP*	7.43	1.64E-18	*LEP*	7.51	1.47E-17	*ACTN2*	11.17	1.94E-10
DLK1	7.09	3.27E-11	*GLYAT*	7.12	9.29E-22	*BMP3*	7.36	1.71E-11	*ACTA1*	11.00	1.78E-10
LEP	6.88	8.40E-12	*GYS2*	6.92	1.27E-15	*TRHDE-AS1*	7.00	5.47E-18	*TRDN*	10.90	1.72E-14
MTND1P23	6.65	3.73E-09	*CSN1S1*	6.67	1.79E-15	*NPY2R*	6.90	7.15E-10	*TNNC2*	10.52	1.09E-13
CEACAM5	−7.29	4.44E-10	*CRISP3*	−7.01	3.11E-08	*GLYATL1P4*	−7.26	1.17E-06	*CBLN2*	−6.70	4.80E-11
S100P	−7.51	6.74E-18	*CST2*	−7.35	9.78E-12	*INSM1*	−7.91	9.15E-09	*EPYC*	−6.74	1.20E-15
SNORD17	−7.60	3.34E-11	*S100A7*	−7.96	2.75E-08	*PCSK1*	−8.06	6.94E-12	*CST1*	−6.58	5.81E-09
IBSP	−7.75	1.23E-21	*IBSP*	−8.51	3.54E-21	*TRPA1*	−8.67	2.24E-13	*COL10A1*	−7.57	1.03E-33
MMP1	−8.55	1.57E-20	*CYP2A7*	−8.96	1.81E-08	*SLC30A8*	−8.76	5.84E-09	*DCD*	−7.85	2.09E-05

Positive log_2FC values indicate that the mRNA expression of these genes are up-regulated in normal tissue groups, while negative log_2FC values correspond to up-regulated genes in tumor tissue groups. The ranks of the 10 top DEGs for each repeat in the three other repeats are shown in Supporting information Table S4.

focus on the LFC and FDR values of each gene in the results, while neglecting the examination of the raw data. To better understand why the reproducibility of DE results is so poor, even for large sample sizes, the present study examined the TMM-normalized read counts of the 10 up- and down-regulated DEGs of repeat I in Table 1. Among these genes, *CEACAM5*, *S100P*, *SNORD17*, *IBSP*, and *MMP1* are up-regulated DEGs in tumor tissue samples, while *MYOC*, *PENK*, *DLK1*, *LEP*, and *MTND1P23* are up-regulated DEGs in normal tissue samples. Figure 4 shows that the transcript expression of each of the genes in different samples, whether in tumor groups or in normal groups, varies greatly, which reflects the high heterogeneity of both tumor and normal samples. The FPKM values of each of the top 10 DEGs also show high variations in different samples of each group (Supporting information Figure S5), consistent with the results shown in Figure 4.

We also ran principal components analysis (PCA) for each cancer type comprising all samples used in the study to better illustrate the heterogeneity of both tumor and normal samples. As shown in Supporting information Figure S6, samples from the tumor group and the normal group cluster separately with limited overlaps for each cancer type. Notably, the PCA scatter plots show that samples in both tumor and normal groups are quite dispersed, representing high sample variability.

The results of the parallel analysis performed on HeLa cell line can be found in Supporting information Figures S7-S9. Although the DEG numbers for each n of the HeLa cell dataset are significantly less than those of the three human tissue datasets studied above, the trends of the three curves under different threshold conditions in Supporting information Figure S7 are consistent with those in Figure 1. Moreover, as shown in Supporting information Figure S8, the overlap rates obtained for the most stringent threshold are generally the lowest among the three threshold conditions. Furthermore, Supporting information Figure S9 also shows that the transcript expression of each gene has a character of high variation between different samples in either groups.

3.4 | Analysis on the optimal threshold controlling the FDR depending on replicate number

Generally, a fixed threshold of FDR <0.05 is used. Nevertheless, Lamarre et al. (2018) found that choosing a threshold for FDR around 2^{-n} (with n the number of biological replicates per condition) should be optimal to enhance the sensitivity (true positive rate) and specificity (true negative rate) of DE analysis, but as they stated, their result had only been shown for their Tomato Ovary Gene Expression data.

To investigate whether 2^{-n} was the optimal threshold controlling the FDR for the TCGA data, we first compared the performance of FDR(0.05) and FDR(2^{-n}) in DEG filtration using the RNA-seq count data of BRCA, KIRC, and LUAD, and the result is presented in Figure 5. It is obvious in Figure 5A, C, and E that the number of DEGs detected by FDR (2^{-n}) is much less than that by FDR (0.05) for any given n, except for $n = 3$ and 4. As $2^{-n} > 0.05$ if $n = 3$ or 4, the result is understandable because fewer DEGs are left when more stringent thresholds are applied. When FDR (0.05) is applied, the number of DEGs continues to increase rapidly with the increase in sample size. Nevertheless, for FDR(2^{-n}), the number of DEGs for different n is close, but DEG compositions are quite different, similar to the features of the curves for (FDR(0.001), LFC(4)) in Figure 1. Furthermore, for most values of n, the overlap rates for FDR (2^{-n}) are much lower than those for FDR (0.05) as shown in Figure 5 B,D, and F, indicating that using 2^{-n} as the threshold for FDR can keep the number of DEGs

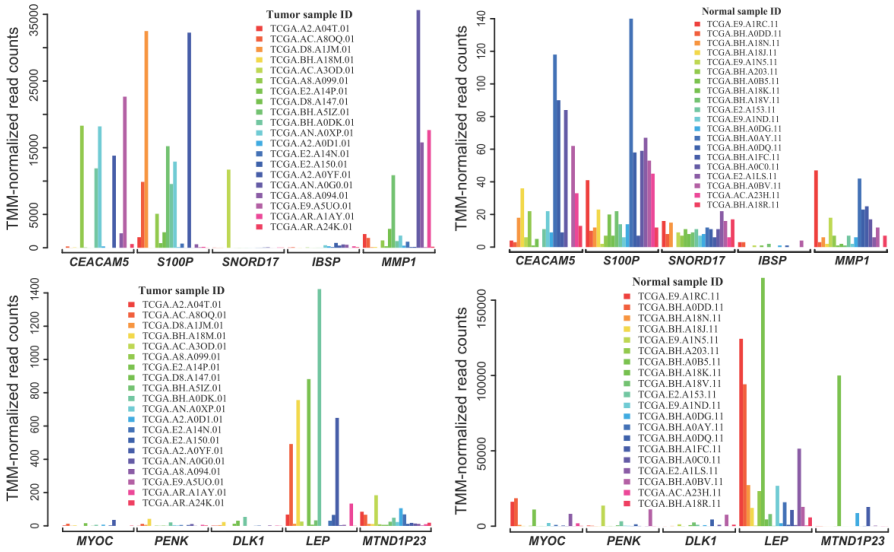


FIGURE 4 TMM-normalized read count values of the top 10 up- and down-regulated DEGs of repeat I for $n = 20$ in BRCA [Colour figure can be viewed at wileyonlinelibrary.com]

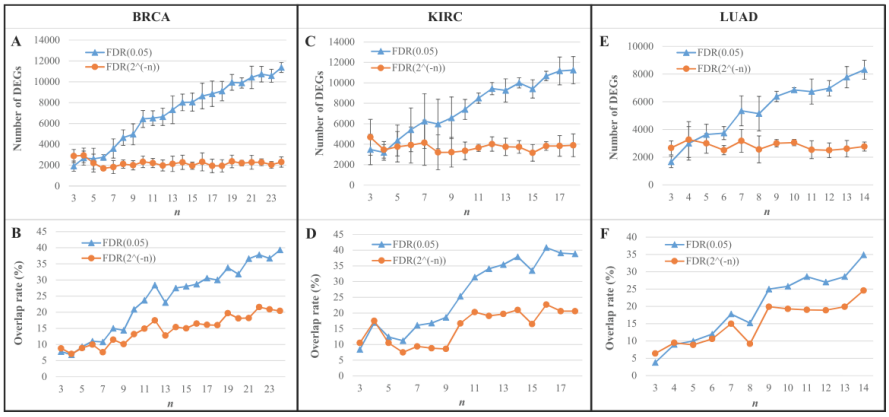


FIGURE 5 Comparison of DEG numbers (A, C, and E) and overlap rates (B, D, and F) for FDR(0.05) and FDR(2^{-n}) [Colour figure can be viewed at wileyonlinelibrary.com]

relatively stable for different n, but cannot improve the reliability of DE results.

Next, the sensitivity and specificity of DE results were analyzed for FDR(0.05) and FDR(2^{-n}), respectively, using the datasets of the three cancer types studied, according to the method proposed by Lamarre et al. (2018). As shown in Supporting information Figure S10, different datasets generate slightly different results, but generally the sensitivity

of FDR(2^{-n}) is higher than that of FDR(0.05) for small sample sizes and lower than that of FDR(0.05) for large sample sizes, and vice versa for the specificity of FDR(2^{-n}) and FDR(0.05).

Finally, receiver operating characteristic (ROC) analysis was performed on the DE analyses for the three cancer types to estimate the optimal FDR thresholds depending on replicate number as Lamarre et al. (2018) did. However, our results are quite different from those of Lamarre et al. (2018). As shown in Supporting information Figure S11, the arrangement of the ROC curves for different n is not as orderly as that of Lamarre et al. (2018). Additionally, the ROC curves for different repeats of each n are distinct from one another as shown in Supporting information Figures S12-S14. Furthermore, for a given n, the optimal cutpoint values of FDR threshold for different curves are quite different, especially for $n < 10$ (Supporting information Figure S15).

Lamarre et al. (2018) also performed multiple DE analyses for each n from 2 to 7; however, only one ROC curve was shown for each n in Figure 4 of their article. Apparently, the optimal FDR threshold calculated from only one ROC curve cannot represent the optimal FDR thresholds of all the curves for each n. As reflected in Supporting information Figure S15, although the mean of the four optimal FDR thresholds for each n shows a downward trend with increasing numbers of biological replicates, the optimal FDR thresholds for each n are clearly not equal to 2^{-n}, which is inconsistent with the result of Lamarre et al. (2018).

4 | DISCUSSION

The present study thoroughly analyzes the effect of the variations in transcript expression on the reliability of DE results in RNA-seq analysis, using different threshold levels of FC and FDR (either combined or separately) and different sample sizes. The results demonstrate that no matter which threshold is applied, the reliability of DE results for different sample sizes is quite poor. Moreover, the more stringent the threshold, the more unreliable the DE results, which is contrary to what we usually believe. An arbitrary FDR threshold of 0.05 for a DE analysis, regardless of the number of biological replicates per condition, would not be optimal; however, it is difficult to assign an optimal FDR threshold for a particular DE analysis based on the number of biological replicates per condition, as the optimal FDR thresholds for different DE analyses for a given n are quite different. How to choose the optimal FDR threshold for a particular DE analysis remains to be further investigated.

The substantial difference in read count values for the same gene among different samples, whether in tumor groups or in normal groups, indicates that gene expression at mRNA level varies tremendously, which is also illustrated by the PCA plots (Supporting information Figure S6) for each cancer type comprising all samples used in the study. Although tumor heterogeneity has long been recognized, normal sample heterogeneity has been underestimated. According to the results above, normal tissue samples also have significant heterogeneity at mRNA level. Furthermore, many genes show extremely high transcript expression in certain samples (Figure 4); however, the reason for these extreme expression values remains unclear.

The mean and standard deviation (SD) are essential statistical parameters used in DE analysis (Robinson et al., 2010, Love et al., 2014). However, the values of these parameters can be significantly altered if extremely high variations in transcript expression exist, causing considerable differences in the results obtained from different sets of samples. Extreme read count values can also lead to very large LFC values and extremely small FDR values, which may explain why the more stringent the threshold, the worse the reliability of DE results. The extreme read count values seem to be randomly distributed across different samples; therefore, it is difficult to improve the reliability of DE results by excluding certain samples.

Compared with large sample sizes, small ones are more susceptible to high variations in mRNA expression, leading to poorer reliability of DE results for small sample sizes. Theoretically, if mRNA expression levels are consistent among samples in the same group, then the DEGs obtained using different sets of samples should be the same or similar, regardless of whether small or large sample sizes are used. However, as shown in Figure 1, when (FDR(0.05), LFC(1)) is applied, the number of DEGs detected positively correlates with n if n is below 10. Similar phenomena have been reported in previous studies using RNA-seq read count data from mouse strains (Soneson & Delorenzi, 2013), yeast (Schurch et al., 2016), tomato plants (Lamarre et al., 2018), and human tissues (Cui et al., 2021), and cell lines (Liu et al., 2014). The results of the parallel analysis using HeLa cells (Supporting information Figures S7-S9) are also consistent with those obtained from human tumor and normal tissue samples. These findings indicate that a high variation in transcript expression is probably a universal feature of biological samples.

Taken together, DEGs obtained using more stringent thresholds have poorer reproducibility due to a high variation in transcript expression. Therefore, less stringent thresholds are recommended for filtering DEGs, in case "true" DEGs are excluded by too stringent thresholds. Large sample sizes should be considered in RNA-seq experimental designs to reduce the interfering effect of the variation in transcript expression. Although RNA-seq is a promising technology for monitoring transcriptome

changes, DE results should be cautiously interpreted, as a large proportion of the DEGs identified are irreproducible. A high variation in transcript expression also illustrates the necessity and importance of large genomics projects such as TCGA.

ACKNOWLEDGEMENTS

The results shown here are in whole based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This work was supported by the Natural Science Foundation of Shandong Province, China (No. ZR2019PC053); the Key Research and Development Program of Zibo (No. 2018kj040045); the Zibo City Integration Development Project (2019ZBXC320); the Taishan Scholars Program of Shandong Province (No. tsqn201909148), and the National Natural Science Foundation of China (No. 31701042).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHORS' CONTRIBUTIONS

QW, WC, and HX participated in the conception and design of the study. WC and HX acquired and prepared the testing data. WC, HX, YG, and JZ performed DE analysis. WC, YG, and YL compared DE results. WC and HX wrote the manuscript. XT and QW advised on the analyses and revised the manuscript. All authors read and approved the final manuscript.

AVAILABILITY

The datasets supporting the conclusions of this study are available in TCGA, https://www.cancer.gov/tcga.

REFERENCES

Al Seesi, S., Tiagueu, Y.T., Zelikovsky, A. & Mandoiu, Ii (2014) Bootstrap-based differential gene expression analysis for RNA-Seq data with and without replicates. *BMC Genomics*, *15*, S2.

Burrell, R.A. & Swanton, C. (2014) Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol*, *8*, 1095–1111.

Ching, T., Huang, S. & Garmire, L.X. (2014) Power analysis and sample size estimation for RNA-Seq differential expression. *RNA*, *20*, 1684-1496.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., Mcpherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X. & Mortazavi, A. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol*, *17*, 13.

Cui, W., Xue, H., Wei, L., Jin, J., Tian, X. & Wang, Q. (2021) High heterogeneity undermines generalization of differential expression results in RNA-Seq analysis. *Hum Genomics*, *15*, 7.

Dagogo-Jack, I. & Shaw, A.T. (2018) Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*, *15*, 81–94.

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloe, D., Le Gall, C., Schaeffer, B., Le Crom, S., Guedj, M., Jaffrezic, F. & French Statomique, C. (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*, *14*, 671–683.

Esteve-Codina, A., Arpi, O., Martinez-Garcia, M., Pineda, E., Mallo, M., Gut, M., Carrato, C., Rovira, A., Lopez, R., Tortosa, A., Dabad, M., Del Barco, S., Heath, S., Bague, S., Ribalta, T., Alameda, F., De La Iglesia, N., Balana, C. & Group, G. (2017) A Comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS One*, *12*, e0170632.

Gierlinski, M., Cole, C., Schofield, P., Schurch, N.J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., Blaxter, M. & Barton, G.J. (2015) Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics*, *31*, 3625–3630.

Hansen, K.D., Wu, Z., Irizarry, R.A. & Leek, J.T. (2011) Sequencing technology does not eliminate biological variability. *Nat Biotechnol*, *29*, 572–573.

Jia, C., Wang, L.Y., Yin, G.G. & Zhang, M.Q. (2019) Single-cell stochastic gene expression kinetics with coupled positive-plus-negative feedback. *Phys Rev E*, *100*, 052406.

Kim, J.K. & Marioni, J.C. (2013) Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol*, *14*, R7.

Kwong, L.N., De Macedo, M.P., Haydu, L., Joon, A.Y., Tetzlaff, M.T., Calderone, T.L., Wu, C.J., Kwong, M.K., Roszik, J., Hess, K.R., Davies, M.A., Lazar, A.J. & Gershenwald, J.E. (2018) Biological validation of RNA sequencing data from formalin-fixed paraffin-embedded primary melanomas. *JCO Precis Oncol*. https://doi.org/10.1200/PO.17.00259

Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Le Berre-Anton, V., Bouzayen, M. & Maza, E. (2018) Optimization of an RNA-seq differential gene expression analysis depending on biological replicate number and library size. *Front Plant Sci*, *9*, 108.

Li, J. & Tibshirani, R. (2013) Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res*, *22*, 519–536.

Li, J., Witten, D.M., Johnstone, I.M. & Tibshirani, R. (2012) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics*, *13*, 523–538.

Li, X., Cooper, N.G.F., O'toole, T.E. & Rouchka, E.C. (2020) Choice of library size normalization and statistical methods for differential gene expression analysis in balanced two-group comparisons for RNA-seq studies. *BMC Genomics*, *21*, 75.

Lin, B. & Pang, Z. (2019) Stability of methods for differential expression analysis of RNA-seq data. *BMC Genomics*, *20*, 35.

Lin, Y., Golovnina, K., Chen, Z.X., Lee, H.N., Negron, Y.L., Sultana, H., Oliver, B. & Harbison, S.T. (2016) Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. *BMC Genomics*, *17*, 28.

Liu, Y., Zhou, J. & White, K.P. (2014) RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*, *30*, 301–304.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*, 550.

Marusyk, A. & Polyak, K. (2010) Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*, *1805*, 105–117.
Maza, E. (2016) Papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-Seq experimental design. *Front Genet*, *7*, 164.
Moulos, P. & Hatzis, P. (2015) Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. *Nucleic Acids Res*, *43*, e25.
Ozsolak, F. & Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*, *12*, 87–98.
Poplawski, A. & Binder, H. (2018) Feasibility of sample size calculation for RNA-seq studies. *Brief Bioinform*, *19*, 713–720.
Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat Genet*, *32*, 496–501.
Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D. & Betel, D. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol*, *14*, R95.
Robinson, M.D., Mccarthy, D.J. & Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140.
Robinson, M.D. & Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, *11*, R25.
Robles, J.A., Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J. & Taylor, J.M. (2012) Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*, *13*, 484.
Sampson, J.N., Boca, S.M., Moore, S.C. & Heller, R. (2018) FWER and FDR control when testing multiple mediators. *Bioinformatics*, *34*, 2418–2424.
Schurch, N.J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., Blaxter, M. & Barton, G.J. (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, *22*, 839–851.
Seyednasrollah, F., Laiho, A. & Elo, L.L. (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform*, *16*, 59–70.
Soneson, C. & Delorenzi, M. (2013) A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, *14*, 91.
Stark, R., Grzelak, M. & Hadfield, J. (2019) RNA sequencing: the teenage years. *Nat Rev Genet*, *20*, 631–656.
Tomczak, K., Czerwinska, P. & Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*, *19*, A68-77.
Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, *10*, 57–63.
Wei, L., Wang, J., Lampert, E., Schlanger, S., Depriest, A.D., Hu, Q., Gomez, E.C., Murakam, M., Glenn, S.T., Conroy, J., Morrison, C., Azabdaftari, G., Mohler, J.L., Liu, S. & Heemers, H.V. (2017) Intratumoral and intertumoral genomic heterogeneity of multifocal localized prostate cancer impacts molecular classifications and genomic prognosticators. *Eur Urol*, *71*, 183–192.
Yin, L., Xiao, L., Gao, Y., Wang, G., Gao, H., Peng, Y., Zhu, X., Wei, J., Miao, Y., Jiang, K. & Lu, Z. (2020) Comparative bioinformatical analysis of pancreatic head cancer and pancreatic body/tail cancer. *Med Oncol*, *37*, 46.
Yu, D., Huber, W. & Vitek, O. (2013) Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. *Bioinformatics*, *29*, 1275–1282.
Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R. & Zhao, Q.Y. (2014) A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS One*, *9*, e103207.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Cui W, Xue H, Geng Y, Zhang J, Liang Y, Tian X, Wang Q. Effect of high variation in transcript expression on identifying differentially expressed genes in RNA-seq analysis. *Ann Hum Genet*. 2021;1–10.
https://doi.org/10.1111/ahg.12441

4.8.4 Delayed peroneal muscle reaction time in male amateur footballers during a simulated prolonged football protocol（孙威）

RESEARCH IN SPORTS MEDICINE
https://doi.org/10.1080/15438627.2020.1868467

Routledge
Taylor & Francis Group

Check for updates

Delayed peroneal muscle reaction time in male amateur footballers during a simulated prolonged football protocol

Wei Sun[a], Edwin C. H. Chan[b] and Daniel T. P. Fong[b]

[a]College of Sports and Health, Shandong Sport University, Jinan, Shandong, China; [b]National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

ABSTRACT

Peroneal muscle fatigue could result in ankle inversion sprain injuries. This study investigated the peroneal muscle reaction time during a simulated prolonged football protocol. Nine male footballers completed a 105-minute simulated prolonged football protocol. The peroneal muscle reaction time to an ankle inversion perturbation was measured every 15 minutes by a surface electromyography system sampling at 1000 Hz. One-way repeated ANOVA with post-hoc paired t-test showed a steady upward trend starting from 48.9 ms at baseline to 57.1 ms at the end of the first half, followed by a recovery back to 50.9 ms at the start of the second half and a further delay in the last 30 minutes to 60.2 ms at the end of the protocol. Delayed peroneal muscle reaction was found after 30 minutes of the first half and 15 minutes of the second half of a football match. The risk of ankle sprain could increase in the latter minutes in each half protocol. Thus, prevention injury training strategies should focus on these specific durations in football matches.

ARTICLE HISTORY
Received 12 August 2020
Accepted 18 December 2020

KEYWORDS
Ankle sprain; syndesmotic injuries; ligamentous injuries; neuromuscular control; soccer; sports medicine; ankle injuries; biomechanics

Introduction

Ankle sprain was the most common injury in football, accounting for approximately 84% of all sports-related injuries (Doherty et al., 2014; Fong et al., 2007). Football players suffering from repeated ankle sprains are more likely to develop chronic ankle instability (Gribble et al., 2016). The recurrence rate for a lateral ankle sprain is reportedly as high as 80% among athletes, and it is responsible for the longest absenteeism from participation compared with that in other sports injuries (Fong et al., 2007). Ankle injuries also have a drastic effect on the healthcare system, with an estimation of 1–1.5 million people in the UK attending emergency rooms and clinics yearly (Lamb et al., 2009) and medical costs amounting to £1–2 billion annually. In the US and the Netherlands, such costs were estimated to be US$ 6.2 billion and €208 million per year, respectively (Gribble et al., 2016).

A previous research pointed out that one common cause of ankle sprain injury is delayed peroneal muscle reaction time to an ankle inversion perturbation (Fong et al.,

CONTACT Daniel T. P. Fong d.t.fong@lboro.ac.uk National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

2009). Peroneal muscle reaction time was defined as the time between ankle perturbation and the onset of peroneal muscle activity; it could be obtained by measuring the timing of peroneal activation in response to a sudden inversion perturbation by using electromyography and a trapdoor mechanism (Hoch & McKeon, 2014). When the ankle joint is forced into excessive inversion, the peroneal muscle is the first to react by providing a reflexive contraction as a dynamic defence mechanism to protect the ankle joint (Hertel, 2002). Some studies showed that people with ankle sprain exhibited delayed peroneal reaction time to inversion perturbation compared with healthy ones (Hoch & McKeon, 2014). The peroneal muscle reaction time was found to be around 55–80 ms (Konradsen & Ravn, 1991), which is still too slow to react to a sudden inversion ankle sprain that occurred in 40–50 ms (Fong et al., 2009). Further understanding of the peroneal muscle reaction time during football games could help introduce ankle training strategies for preventing injuries.

Previous studies also suggested that ankle sprains are more likely to occur during the latter minutes of the first half and during the second half in a football game (De Noronha et al., 2019). Football players were reported to experience muscular fatigue after a 90-minute football match as they were required to perform a lot of running, changing direction, jumping, and kicking that would lead to muscular fatigue (Reilly et al., 2008). Fatigue from prolonged matches may affect the muscle responses, motor control and joint position sense for maintaining joint stability (Gribble et al., 2007), induce a change in balance strategy and increased the ankle muscles reaction time and as a result (Valldecabres et al., 2020), increase the risk of ankle sprain and decrease the level of performance (Almonroeder et al., 2020). Those findings may be able to explain the reason why more ankle sprain injuries were being recorded towards the end of each half in real football matches (De Noronha et al., 2019). However, the delayed reaction time with prolonged football game is still unclear, it is very essential to explore the correlation between delayed reaction time and prolonged football game for designing the appropriate training methods for preventing ankle sprain.

Some studies have tried to use isokinetic exercise (Jackson et al., 2009) and football-specific intermittent exercise (Rahnama et al., 2006) protocols to introduce fatigue to the ankle muscles. However, these protocols do not truly represent the nature of football, which is a prolonged exercise with a mixture of different exercise intensities, i.e. running, jogging, and sprinting. Therefore, the present study aimed to investigate the peroneal muscle reaction time in male amateur football players during a simulated prolonged football protocol. The peroneal muscle reaction time to an ankle inversion perturbation was hypothesised to significantly increase from baseline in a simulated prolonged football protocol.

Materials and methods

Participant

Nine healthy male football players (age: 23.3 ± 0.8 years; height: 1.77 ± 0.04 m; body mass: 72.5 ± 8.4 kg) from a local amateur football league were recruited using invitation letters, advertisements and personal invitations in this study. All participants had at least 5 years of recreational football experience and were members of several football teams in a local

amateur league with teams from local universities and colleges. The exclusion criteria included balance disorders, serious lower extremity injury within 1 year, ankle instability, and sensory impairments, as evaluated by an orthopaedic specialist. All the participants did not perform any vigorous exercise 24 hours before testing. All the participants did not perform any vigorous exercise 24 hours before testing. They completed a health screen questionnaire and a written informed consent form approved by the Loughborough University Research Ethics Committee (R18-P069).

Sample size calculation

Sample size estimation was performed on G*Power software (Germany) on the basis of a previous study, which reported that the peroneal reaction time increased from 40.76 ± 10.90 ms to 53.74 ± 4.76 ms after a football game (Fong et al., 2020). By setting the level of significance to 0.05 and the statistical power to 0.80 in a two-tailed test on matched pairs, the effect size and estimated required sample size were calculated to be 1.37 and 7, respectively.

Ankle peroneal muscle reaction test

This test required the participant to stand with their dominant leg on a trapdoor platform which has been widely used in studies to produce a simulate an ankle sprain motion and the other leg on a fixed block with their body weight placed equally on their feet. The dominant limb was defined as the preferred limb to kick a ball as verbally reported by the participant (Sun et al., 2015). The sudden simulated ankle inversion perturbation was introduced to the dominant limb for EMG data collection for three trials. The sEMG data collections of the peroneal muscle were conducted at 1000 Hz, using the Trigno Wireless System (Delsys Inc., Boston, MA, USA). The use of a 1000 Hz sampling rate was suggested to have a moderate to high test–retest reliability for testing muscle latency, with interclass correlation coefficients ranging from 0.68 to 0.94 (Xu et al., 2005). The amplifier bandwidth frequency ranged from 20 to 450 Hz, with an input voltage of 9 VDC at 0.7A and the common-mode rejection ratio was 80 Db. A wireless Trigno EMG sensor was attached on the skin surface of the peroneal muscle belly (upper 1/4 position between the tip of the head of the fibula to the tip of the lateral malleolus) on the lateral side of the lower limb by a method in an EMG manual (Figure 1) (Fong et al., 2013; Perotto, 2005). Recording sites were prepared by shaving the area and wiping with alcohol pads to decrease electrical impedance, Electrodes (41 × 20 × 5 mm, D.E. 2.3, Delsys Inc., Boston, MA) were placed along the longitudinal axis of peroneal muscle on the dominant leg of the participant's body (Gullett et al., 2009).

EMG data were analysed in accordance with the procedures from the International Society of Electrophysiology and Kinesiology (Merletti & Torino, 1999) by using the EMG Works Analysis 4.0 (Delsys Inc., Boston, MA, USA). To calculate the mean normalised EMG values, the raw EMG signals were subsetted, filtered (passband: 3, response: band pass, corner F1: 10 Hz, corner F2: 500 Hz), rectified, integrated root mean square (window length: 0.100, window overlap: 0.08, remove offset) to calculate the mean normalised EMG values. The onset time of the peroneal muscle was determined by a sudden increase of EMG signal which exceeded 5% of the maximum signal value of the muscle (Fong et al., 2020; Konrad, 2006). The time between the start of pulling the trigger of the trapdoor

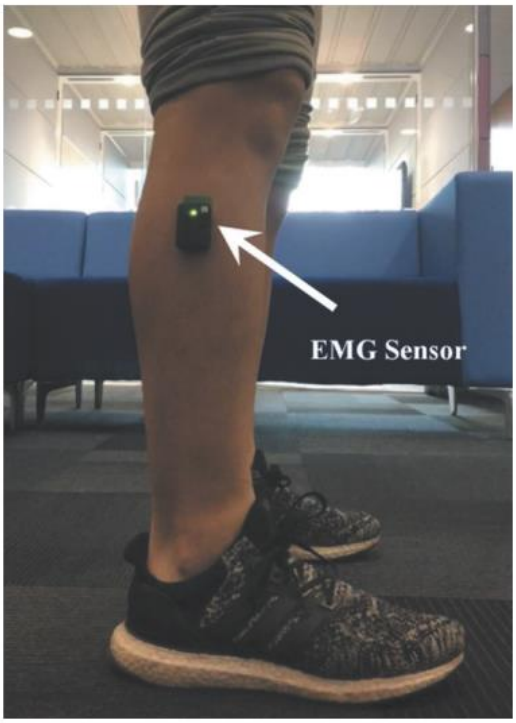


Figure 1. The position of EMG sensor during reaction time test.

platform and the onset of the EMG signal was the reaction time of the peroneal muscle. The average value of the reaction time from three trials was used for statistical analysis.

Prolonged simulated football protocol

The prolonged simulated football protocol was modified from the Loughborough intermittent shuttle test to simulate a football game (Nicholas et al., 2000). The shuttle running test includes two 45 minutes half with a 15 minutes rest period in between. The test was set up with two cones 20 m apart and each cycle of the test includes 60 m walking. 20 m sprinting, 4 seconds recovery walk, and 60 m jogging and 60 m sprinting (Fong et al., 2020). Each participant started with a dynamic warmup for 15 minutes and then moved on to the ankle muscle reaction test for the EMG data collection at 0 minutes before starting the prolonged simulated football protocol. The peroneal reaction test was then repeated at every 15 minutes interval, i.e. at 0, 15, 30, 45, 60, 75, 90, and 105 minutes for EMG data collection. The participants were monitored by the researcher throughout the shuttle running test and required to repeat the cycle until each 15 minutes interval was reached. No practice trials were allowed to ensure the unexpected nature of an ankle sprain.

Table 1. The reaction time of peroneal muscle throughout the prolonged simulated football protocol.

Time (min)	Reaction Time (ms)	p-value of t-test with baseline	95% CI for difference	Cohen's d
0 min	48.9 ± 1.3	–	–	–
15 mins	51.8 ± 2.2	1.000	−8.219, 2.564	1.57
30 mins	54.9 ± 1.6	0.001*	−8.853, −3.044	4.05
45 mins	57.1 ± 1.7	<0.001*	−9.627, −6.606	5.37
60 mins	50.9 ± 1.3	0.522	−5.127, 1.187	1.53
75 mins	56.6 ± 3.0	0.005*	−13.024, −2.401	3.37
90 mins	59.2 ± 1.6	<0.001*	−14.430, −6.178	6.91
105 mins	60.2 ± 1.1	<0.001*	−14.152, −8.391	9.13

*Significant difference from baseline (0 min), $p < 0.05$.

Statistical analysis

Statistical analysis was performed with SPSS software (version 20.0, SPSS Inc, USA) in this study. All the data were reported as means ± standard deviations. One-way repeated ANOVA was conducted on the dependent variables over time. If a significant time effect was found, a post-hoc Bonferroni test was conducted to evaluate the significant difference between reaction time at each time and baseline. The significance level was set at $p < 0.05$. Partial eta squared (n^2p) was used to represent the effect of the time effect of repeated ANOVA. The thresholds for partial eta squared were as follows: 0.01–0.06, small; 0.06–0.14, moderate; >0.14, large (Pierce et al., 2004). Cohen's d was used to represent the effect size of post-hoc comparison. The thresholds for Cohen's d were as follows: <0.2, trivial; 0.21–0.50, small; 0.510.80, medium; >0.81, large (Cohen, 2013).

Results

As shown in Table 1 and Figure 2, one-way repeated ANOVA suggested the significant time effect for the reaction time ($F = 38.937$, $p < 0.001$, $n^2p = 0.848$). The post-hoc results

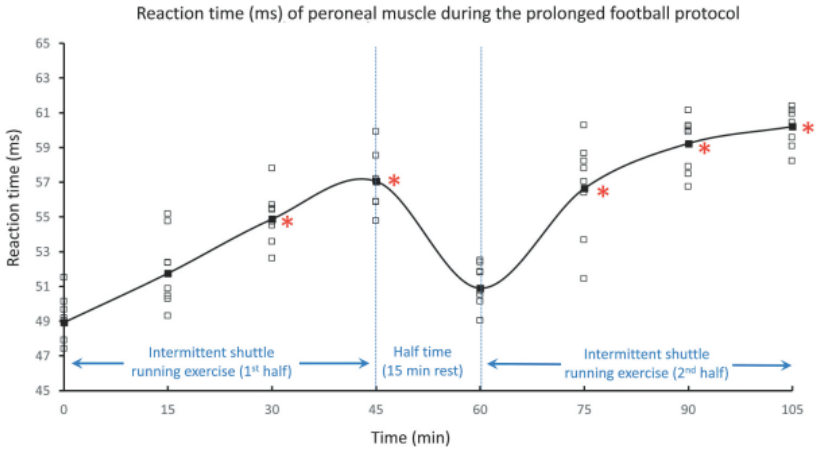


Figure 2. Reaction time of peroneal muscle during the prolonged football protocol (*Indicated significant difference from baseline (0 min), $p < 0.05$).

showed that the reaction time at 30 ($p = 0.001$, Cohen's $d = 4.05$), 45 ($p < 0.001$, Cohen's $d = 5.37$), 75 ($p = 0.005$, Cohen's $d = 3.37$), 90 ($p < 0.001$, Cohen's $d = 6.91$), and 105 min ($p < 0.001$, Cohen's $d = 9.13$) throughout the prolonged simulated football protocol increased significantly compared with that at baseline.

Discussion

The results showed that the peroneal muscle reaction time significantly increased at 30, 45, 60, and 90 minutes from the beginning, thereby supporting the hypothesis. The results were similar and supportive to those of most previous studies that tested peroneal muscle reaction time in a sudden simulated ankle inversion sprain situation (Henry et al., 2010). These findings suggested that muscular or physiological fatigue may have been induced at the later stage of each half of the prolonged exercise protocol and caused an increase in peroneal muscle reaction time. Although the range of the peroneal muscle reaction time in the present study stayed in the healthy range at 55–80 ms (Konradsen & Ravn, 1991), it was still not quick enough to react to a sudden ankle inversion that occurred within 50 ms (Fong et al., 2009). Moreover, the injury was expected to occur faster in dynamic motions, such as in football games, and with a higher twisting force; thus, an even faster peroneal reaction time was needed for protection. The increase in peroneal muscle reaction time towards the end of each half could also increase the ankle injury risk at the later stage of each half.

A notable detail that the reaction time significantly increased at 75 (56.6 ± 3.0 ms), 90 (59.2 ± 1.6 ms), and 105 (60.2 ± 1.1 ms) minutes in the second half in this study, while the previous study (Fong et al., 2020) found the reaction time significantly increased only at 105 minutes, the end of a football game. The gender might lead to a discrepancy in the increasing reaction time of the second half. The male participants were recruited in this study, female in a previous study. A recent study showed that the maximal isometric peak force recovered in the male group was significantly slower during the first hour of rest than that in the female group after prolonged exercise-induced fatigue (Hakkinen, 1993). This finding may imply that the potential risk of ankle sprain was larger in males than in females during the second half of the football game.

Another interesting finding was that the peroneal reaction time (from 48.9 ms to 60.2 ms) of each time point in the present study was longer than that in the previous study (40.76–53.74 ms). The different methods to define the starting time of the reaction time may lead to the discrepancy between the two studies. Previous studies found that two definitions were adopted: first was between the start of pulling the trigger of the trapdoor platform and the first rising response of EMG signals (Sun et al., 2016) and the second one was between the start of ankle inversion motion and the first rising response of EMG signals (Fong et al., 2020).

In addition, the peroneal muscle reaction time on the dominant leg was examined in the present study. The study that investigated the delayed reaction time during prolonged football protocol in female amateur footballers (Fong et al., 2020) showed a significantly shorter peroneal latency in the nondominant leg than in the dominant leg. The difference may be the result of different demands being placed on the dominant and nondominant legs during exercise (Beynnon et al, 2002). For example, football players shoot, pass and jump with their dominant leg more than their nondominant leg; therefore, the dominant

leg may experience fatigue faster and result in a longer peroneal latency (Knight & Weimar, 2011). This phenomenon could explain why the ankle sprain incident was 2.4 times more on the dominant leg than on the non-dominant leg in a competitive football season (Yeung et al., 1994). However, the reaction time of the nondominant leg during prolonged football protocol is still unclear. Further study could focus on the difference in the reaction time between these legs during prolonged football exercise protocol for differentiated training strategy of ankle sprain injury prevention. Peroneal muscle endurance should be evaluated in the preseason screening test (Delvaux et al., 2020). Some training interventions, such as the recent successful attempts of the use of kinesiology tape (Farquharson & Greig, 2017) and compression stockings (Pavin et al., 2019), could also be conducted to reduce the peroneal muscle reaction time and prevent ankle sprain injuries.

This study has three limitations. First, the trapdoor platform for an ankle sprain simulation has caused some concern about its validity as most of the ankle sprain injuries in football normally happened in dynamic motions such as landing from a jump and running instead of standing with both feet flat on a flat surface. However, this is the most common way among the very few methods to mimick an ankle sprain motion for the peroneal muscle reaction test. Second, there was potential human response error on the synchronisation of EMG data collection for calculation of the peroneal muscle reaction time as there may be a minor time error between pulling the trigger of the trapdoor platform and the real starting time of ankle inversion motion. This finding implies that the standard methods of data collection and definition of reaction time are very important to assess the neuromuscular reaction function in the future study. Therefore, the researcher should focus on standardising the testing methods on neuromuscular reaction time for conclusions from many studies, such as sampling frequency, warming up programme, data processing. Thirdly, the control group is lacked. In this study, the control condition was the data collect at the initial time point, as we were looking at the time effect. To further enrich the study design, future studies can have a resting group added to further show the effect of participating and not participating in the prolonged football protocol.

In conclusion, the delayed peroneal muscle reaction was found after 30 minutes of the first half and 15 minutes of the second half of a football match. The risk of ankle sprain could increase in the latter minutes in each half protocol, so the prevention injury training strategies should focus on the specific duration in soccer match.

Acknowledgments

We acknowledged the participants of this study.

Disclosure statement

The authors declared no conflict of interest.

ORCID

Wei Sun http://orcid.org/0000-0003-2706-7814
Daniel T. P. Fong http://orcid.org/0000-0001-7314-189X

References

Almonroeder, T. G., Tighe, S. M., Miller, T. M., & Lanning, C. R. (2020). The influence of fatigue on decision-making in athletes: A systematic review. *Sports Biomechanics*, *19*(1), 76–89. https://doi.org/10.1080/14763141.2018.1472798

Beynnon, B. D., Murphy, D. F., & Alosa, D. M. (2002). Predictive factors for lateral ankle sprains: A literature review. *Journal of Athletic Training*, *37*(4), 376–380.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

De Noronha, M., Lay, E. K., McPhee, M. R., Mnatzaganian, G., & Nunes, G. S. (2019). Ankle sprain has higher occurrence during the latter parts of matches: Systematic review with meta-analysis. *Journal of Sport Rehabilitation*, *28*(4), 373–380. https://doi.org/10.1123/jsr.2017-0279

Delvaux, F., Schwartz, C., Rodriguez, C., Forthomme, B., Kaux, J. F., & Croisier, J. L. (2020). Preseason assessment of anaerobic performance in elite soccer players: Comparison of isokinetic and functional test. *Sports Biomechanics*. https://doi.org/10.1080/14763141.2020.1750681

Doherty, C., Delahunt, E., Caulfield, B., Hertel, J., Ryan, J., & Bleakley, C. (2014). The incidence and prevalence of ankle sprain injury: A systematic review and meta-analysis of prospective epidemiological studies. *Sports Medicine*, *44*(1), 123–140. https://doi.org/10.1007/s40279-013-0102-5

Farquharson, C., & Greig, M. (2017). Kinesiology tape mediates soccer- simulated and local peroneal fatigue in soccer players. *Research in Sports Medicine*, *25*(3), 313–321. https://doi.org/10.1080/15438627.2017.1314294

Fong, D. T. P., Chan, Y. Y., Mok, K. M., Yung, P. S. H., & Chan, K. M. (2009). Understanding acute ankle ligamentous sprain injury in sports. *Sports Medicine, Arthroscopy, Rehabilitation, Therapy and Technology*, *1*, 14. https://doi.org/10.1186/1758-2555-1-14

Fong, D. T. P., Hong, Y., Chan, L. K., Yung, P. S. H., & Chan, K. M. (2007). A systematic review on ankle injury and ankle sprain in sports. *Sports Medicine*, *37*(1), 73–94. https://doi.org/10.2165/00007256-200737010-00006

Fong, D. T. P., Leung, W. C., Mok, K. M., & Yung, P. S. H. (2020). Delayed ankle muscle reaction time in female amateur footballers after the first 15 min of a simulated prolonged football protocol. *Journal of Experimental Orthopaedics*, *7*(1), 54–62. https://doi.org/10.1186/s40634-020-00275-1

Fong, D. T. P., Wang, D., Chu, V. W. S., & Chan, K. M. (2013). Myoelectric stimulation on peroneal muscles with electrodes of the muscle belly size attached to the upper shand gives the best effect in resisting simulated ankle sprain motion. *Journal of Biomechanics*, *4*(6), 1088–1091. https://doi.org/10.1016/j.jbiomech.2013.01.019

Gribble, P. A., Bleakley, C. M., Caulfield, B. M., Docherty, C. L., Fourchet, F., Fong, D. T. P., Hertel, J., Hiller, C. E., Kaminski, T. W., McKeon, P. O., Refshauge, K. M., Verhagen, E. A., Vicenzino, B. T., Wikstrom, E. A., & Delahunt, E. (2016). Evidence review for the 2016 International Ankle Consortium consensus statement on the prevalence, impact and long-term consequences of lateral ankle sprains. *British Journal of Sports Medicine*, *50*(24), 1496–1505. https://doi.org/10.1136/bjsports-2016-096189

Gribble, P. A., Hertel, J., & Denegar, C. R. (2007). Chronic ankle instability and fatigue create proximal joint alterations during performance of the star excursion balance test. *International Journal of Sports Medicine*, *28*(3), 236–242. https://doi.org/10.1055/s-2006-924289

Gullett, J. C., Tillman, M. D., Gutierrez, G. M., & Chow, J. W. (2009). A biomechanical comparison of back and front squats in healthy trained individuals. *Journal of Strength and Conditioning Research*, *23*(1), 284–292. https://doi.org/10.1519/JSC.0b013e31818546bb

Hakkinen, K. (1993). Neuromuscular fatigue and recovery in male and female athletes during heavy resistance exercise. *International Journal of Sports Medicine*, *14*(2), 53–59. https://doi.org/10.1055/s-2007-1021146

Henry, B., McLoda, T., Docherty, C. L., & Schrader, J. (2010). The effect of plyometric training on peroneal latency. *Journal of Sport Rehabilitation*, *19*(3), 288–300. https://doi.org/10.1123/jsr.19.3.288

Hertel, J. (2002). Functional anatomy, pathomechanics, and pathophysiology of lateral ankle instability. *Journal of Athletic Training*, *37*(4), 364–375.

Hoch, M. C., & McKeon, P. O. (2014). Peroneal reaction time after ankle sprain: A systematic review and meta-analysis. *Medicine and Science in Sports and Exercise*, *46*(3), 546–556. https://doi.org/10.1249/MSS.0b013e3182a6a93b

Jackson, N. D., Gutierrez, G. M., & Kaminski, T. (2009). The effect of fatigue and habituation on the stretch reflex of the ankle musculature. *Journal of Electromyography and Kinesiology*, *19*(1), 75–84. https://doi.org/10.1016/j.jelekin.2007.06.016

Knight, A. C., & Weimar, W. H. (2011). Difference in response latency of the peroneus longus between the dominant and nondominant legs. *Journal of Sport Rehabilitation*, *20*(3), 321–332. https://doi.org/10.1123/jsr.20.3.321

Konrad, P. (2006). *The ABC of EMG: A practical introduction to kinesiological electromyography*. Version 1.0. Noraxon Inc.

Konradsen, L., & Ravn, J. B. (1991). Prolonged peroneal reaction time in ankle instability. *International Journal of Sports Medicine*, *12*(3), 290–292. https://doi.org/10.1055/s-2007-1024683

Lamb, S. E., March, J. L., Hutton, J. L., Nakash, R., & Cooke, M. W., & Collaborative Ankle Support Trial (CAST Group). (2009). Mechanical supports for acute, severe ankle sprain: A pragmatic, multicentre, randomised controlled trial. *Lancet*, *373*(9663), 575–581. https://doi.org/10.1016/S0140-6736(09)60206-3

Merletti, R., & Torino, P. D. (1999). Standards for reporting EMG data. *Journal of Electromyography and Kinesiology*, *9*(1), III-IV.

Nicholas, C. W., Nuttall, F. E., & Williams, C. (2000). The Loughborough intermittent shuttle test: A field test that simulates the activity pattern of soccer. *Journal of Sports Sciences*, *18*(2), 97–104. https://doi.org/10.1080/026404100365162

Pavin, L. N., Leicht, A. S., Gimenes, S. V., Da Silva, V. C., De Moura Simim, M. A., Marocolo, M., & Da Moto, G. R. (2019). Can compression stockings reduce the degree of soccer match-induced fatigue in females? *Research in Sports Medicine*, *27*(3), 351–364. https://doi.org/10.1080/15438627.2018.1527335

Perotto, A. O. (2005). *Anatomical guide for the electromyographer: The limb and trunk* (4th ed.). Charles C Thomas Publisher Ltd.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, *64*(6), 916–924. https://doi.org/10.1177/0013164404264848

Rahnama, N., Lees, A., & Reilly, T. (2006). Electromyography of selected lower-limb muscles fatigued by exercise at the intensity of soccer match-play. *Journal of Electromyography and Kinesiology*, *16*(3), 257–263. https://doi.org/10.1016/j.jelekin.2005.07.011

Reilly, T., Drust, B., & Clarke, N. (2008). Muscle fatigue during football match-play. *Sports Medicine*, *38*(5), 357–367. https://doi.org/10.2165/00007256-200838050-00001

Sun, W., Song, Q., Yu, B., Zhang, C., & Mao, D. (2015). Test-retest reliability of a new device for assessing ankle joint threshold to detect passive movement in healthy adults. *Journal of Sports Sciences*, *33*(16), 1667–1674. https://doi.org/10.1080/02640414.2014.1003589

Sun, W., Zhang, C., Song, Q., Li, W., Cong, Y., Chang, W., Mao, D., & Hong, Y. (2016). Effect of 1-year regular Tai Chi on neuromuscular reaction in elderly women: A randomized controlled study. *Research in Sports Medicine*, *24*(2), 145–156. https://doi.org/10.1080/15438627.2015.1126280

Valldecabres, R., Richards, J., & De Benito, A. M. (2020). The effect of match fatigue in elite badminton players using plantar pressure measurements and the implications to injury mechanisms. *Sports Biomechanics*. https://doi.org/10.1080/14763141.2020.1712469

Xu, D. Q., Li, J. X., & Hong, Y. (2005). Effect of regular Tai Chi and jogging exercise on neuromuscular reaction in older people. *Age and Ageing*, *34*(5), 439–444. https://doi.org/10.1093/ageing/afi114

Yeung, M. S., Chan, K. M., So, C. H., & Yuan, W. Y. (1994). An epidemiological survey on ankle sprain. *British Journal of Sports Medicine*, *28*(2), 112–116. https://doi.org/10.1136/bjsm.28.2.112

4.8.4 Detraining effects of regular Tai Chi exercise on postural control ability in older women: A randomized controlled trial（孙威）

Journal of Exercise Science & Fitness 16 (2018) 55–61

Contents lists available at ScienceDirect

Journal of Exercise Science & Fitness

journal homepage: www.elsevier.com/locate/jesf

Detraining effects of regular Tai Chi exercise on postural control ability in older women: A randomized controlled trial

Wei Sun [a, b], Lin Wang [a], Cui Zhang [a, b], Qipeng Song [b], Houxin Gu [b], Dewei Mao [a, c, *]

[a] School of Kinesiology, Shanghai University of Sport, Shanghai, China
[b] Shandong Institute of Sport Science, Jinan, Shandong, China
[c] Shandong Sport University, Jinan, Shandong, China

ARTICLE INFO

Article history:
Received 4 December 2017
Received in revised form
9 February 2018
Accepted 4 June 2018

Keywords:
Tai Chi
Aging
Postural control
RCT design

ABSTRACT

Objective: This study aimed to investigate the training and detraining effects of Tai Chi (TC) on postural control ability in single leg stance (SLS) by conducting a single-blind randomized controlled trial.

Method: Forty-eight older women were randomly divided into the TC, brisk walking (BW), and control(C) groups by using computer-generated program. The participants completed a 16-week intervention training and 8-week detraining program. Postural control ability in SLS was tested at the baseline, 16 t h, 20 t h, and 24 t h weeks. The primary outcomes included single-leg stance time (Time) and secondary outcomes included maximal displacement of the center of pressure (COP) in the anterior−posterior (AP) direction (D-ap), maximal displacement of the COP in the medial−lateral (ML) direction (D-ml), total length of the COP trajectories (Lng), and 95% confidence ellipse area of the COP movements (area), mean AP total excursion velocities (V-ap), and mean ML total excursion velocities (V-ml).

Results: Significant within-group difference compared with the baseline and between-groups difference compared with control group were found at 16 t h, 20 t h, and 24 t h weeks in the TC group and at the 16 t h and 20 t h weeks in the BW group in all the primary and secondary outcomes. Most of secondary outcomes including Lng, D-ml, V-ml, Area increased significantly at the 24 t h week compared with that at the 16 t h week in BW group.

Conclusions: TC was effective in improving postural control ability and maintaining intervention gains, and was recommended as an appropriate exercise to prevent falls in the older adults.

Introduction

The risk of falling in the older adults increases with aging. Approximately one-third of older adults over 65 years of age fall at least once a year.[1] Falls could result in severe injuries, such as fractures, head injuries, and even death.[2] Moreover, the high costs of health care resulting from falls have placed an enormous burden on families. The total costs reached about 23.3 billion in the USA and 1.6 billion dollars in the UK.[3] Declining postural control ability in single-leg stance (SLS), which is profoundly challenging for older adults, is a significant predictor of falls[4] in the elderly. Nearly 50% of falls occur during the single-leg support phase, such as stepping over obstacles and climbing stairs.[5,6]

Regular Tai Chi (TC) could improve postural control ability.[7,8] A cross-sectional study reported that long-term TC practitioners performed well in SLS tests with their eyes closed,[9] possessed less body sway in perturbed single-leg stance,[10] leaned further without losing stability, and showed a good control of their leaning trajectory.[11] Longitudinal studies also provided evidence of the benefits of TC for postural control ability. After a 24-week intervention, the TC group showed significantly shorter total, medial−lateral, and anterior−posterior center of pressure (COP) sway paths compared with the control group.[7] Similarly, another study also corroborated that a 10-week TC training could decrease the COP path and area during postural control tests in the older adults.[12] Furthermore, TC exercise could improve joint kinesthesia,[13] muscle strength in

* Corresponding author. School of Kinesiology, Shanghai University of Sport, Shanghai, China.
E-mail addresses: sunwei841024@163.com (W. Sun), wangling@sus.edu.cn (L. Wang), gracejoyzc@163.com (C. Zhang), songqipeng@163.com (Q. Song), 94761346@qq.com (H. Gu), m_d_wei@sina.com (D. Mao).

https://doi.org/10.1016/j.jesf.2018.06.003

lower extremities,[14] and neuromuscular reaction in older women.[15]

Although TC has been recognized as an effective exercise to improve postural control in older adults, few detraining effects on postural control ability were known. Some older people have stopped training for various reasons, including diseases, injuries, and travels, and they may go on retraining. After post-exercise, some intervention effects on the physical function could start to diminish.[13] Nevertheless, few data offered the magnitude and retention of the maintenance of postural control ability during detraining periods.

Alternatively, brisk walking (BW) was one of the prevalent moderate-intensity aerobic exercise forms across all ages. Although some longitudinal studies have proven that BW could improve static and dynamic balance abilities and lead to the reduction of fall risk in the older adults,[16,17] others affirmed the inconsistent results on postural stability.[18] To our knowledge, TC and BW are safe methods of exercise for older women and require an equivalent energy expenditure.[19] Nonetheless, the detraining effects of both exercises on postural control ability in older women remained unclear.

The present study aims to compare the detraining effects of TC and BW on postural control in the older adults. The following hypotheses are formulated: (1) after the 16-week intervention, the postural control with SLS will improve in both groups, and (2) TC will be effective for maintaining SLS during a detraining period.

Methods

Study design

A single-blinded randomized controlled trial was designed to compare the effects of TC and BW on body balance in SLS during a 16-week training and an 8-week detraining (Figure 1). Both TC and BW groups participated in one 60-min intervention exercises at 5 times a week for 16 weeks. The control group attended group session with the same schedule as the two intervention groups. After stopping the exercises, all participants were prohibited to perform regular exercises for 8 weeks. Postural control ability was tested at the baseline and at the 16 t h, 20 t h, and 24 t h weeks.

Participants

Sample size estimation

G*Power software was used to calculate the sample size with the formula by Hopkins.[20] The following data were determined: effect size = 0.35, two-tailed significance, statistical power = 0.8, α value = 0.05, and drop-out rate = 25%.[18] So three groups of 48 participants were the required sample size.

Participant recruitment and randomization

48 older women aged 60–70 years were recruited through newspapers, leaflets, and community advocacy from Jinan city, China. The exclusion criteria were as follows: having any regular exercise experience and any records of cardiovascular, neurological, falling history, and musculoskeletal diseases. All participants were randomly divided into the TC (n = 16), BW (n = 16), and control (C) groups (n = 16) by using computer-generated program. This study was approved by the ethics committee of Shandong Sport University (No.201613). All the participants were requested to sign a written informed consent statement. The total study period was 6 months.

Exercise intervention

During the 16-week training periods, each group participated in a 60-min session at 5 times a week for 16 weeks. In addition, at least 64 attendance sessions of 80 (80%) were required for each participant among the three groups.

The participants were individually taught to perform a24-form TC exercise by a qualified TC master in the first 3 weeks. Each session included a 10-min warm-up, 20-min learning new movement forms, 20-min reviewing learned movements before, and 10-min cool-down. Subsequently, they practiced with master supervision for the 13 weeks. Each session included a 10-min warm-up, 40-min TC, and 10-min cool-down.

Brisk walking was defined as walking at a 1.79 m/s speed value.[21] During this exercise, the participants perceived that their breathing significantly accelerated, that their body got extremely hot, and that their sweat streamed down.[18] A professional instructor asked the participants to regulate their pace and speed on a pedestrian road. The time of walking increased from 10 to 40 min progressively over the first 3 weeks and then remained constant at 40-min for the later 13 weeks. A session consisted of a 10- minutes warm-up, 40- minutes BW, and 10- minutes cool-down.

The control group was asked to watch TV programs, read newspapers, or attend healthy education lectures with the same schedule as the two other groups. However, they were prohibited to perform any regular exercise and were allowed to maintain their dietary habits.

During the 8-week detraining, the participants of the three groups were asked to stop the intervention exercise and any regular exercise. The researchers called all participants on a weekly basis to confirm whether they participated in any programmed exercises.

Outcomes

Primary outcomes

The SLS tests were performed to assess postural control ability in a quiet testing room, which reported good interclass correlation coefficient (ICC = 0.95 to 0.99) and within the rater interclass correlation coefficient (ICC = 0.73 to 0.93).[22] This measurement procedure asked the participants to stand on the ground in SLS with eyes open and closed, arms hanging on the sides of their relaxed bodies while the other leg was flexed 90° at hip and knee joint. When the balances with eyes open were tested, participants were required to gaze at a dot on the wall 2.5 m away. The length of time was recorded from the moment the participants' foot was off the floor until it touched the floor again. The SLS with the participants' eyes open and closed were performed thrice, and the longest one was selected for analysis. A 1-min break was given between trials.

Secondary outcomes

The tests were performed with a foot pressure plate (RSscan footscan 2D Balance 0.5 m system).[23] Each participant was asked to stand barefoot in a comfortable self-chosen stance facing the positive anterior–posterior (AP)direction on a plate with the dominant leg, which is described as the preferred leg for kicking a football,[24] as motionless as possible. The other leg was fixed 90° at hip and knee joint flexion. Both arms hung relaxed at the sides. Two conditions of standing were tested randomly: one when participants were asked to perform single-leg standing for 22 s with eyes open while looking straight ahead at a dot on the wall 2.5 m away[25]; another one was when they performed single-leg standing for 12 s with eyes closed.[26] The trial was unavailable and repeated if the participants moved the supported leg or if the non-weight leg touched the supporting surface during the testing duration. Three successful trials of each SLS with eyes open and closed were tested after two familiarized test procedures. The time interval for breaks was 1 min between two trials. All measurement procedures were performed under the supervision of a technician.

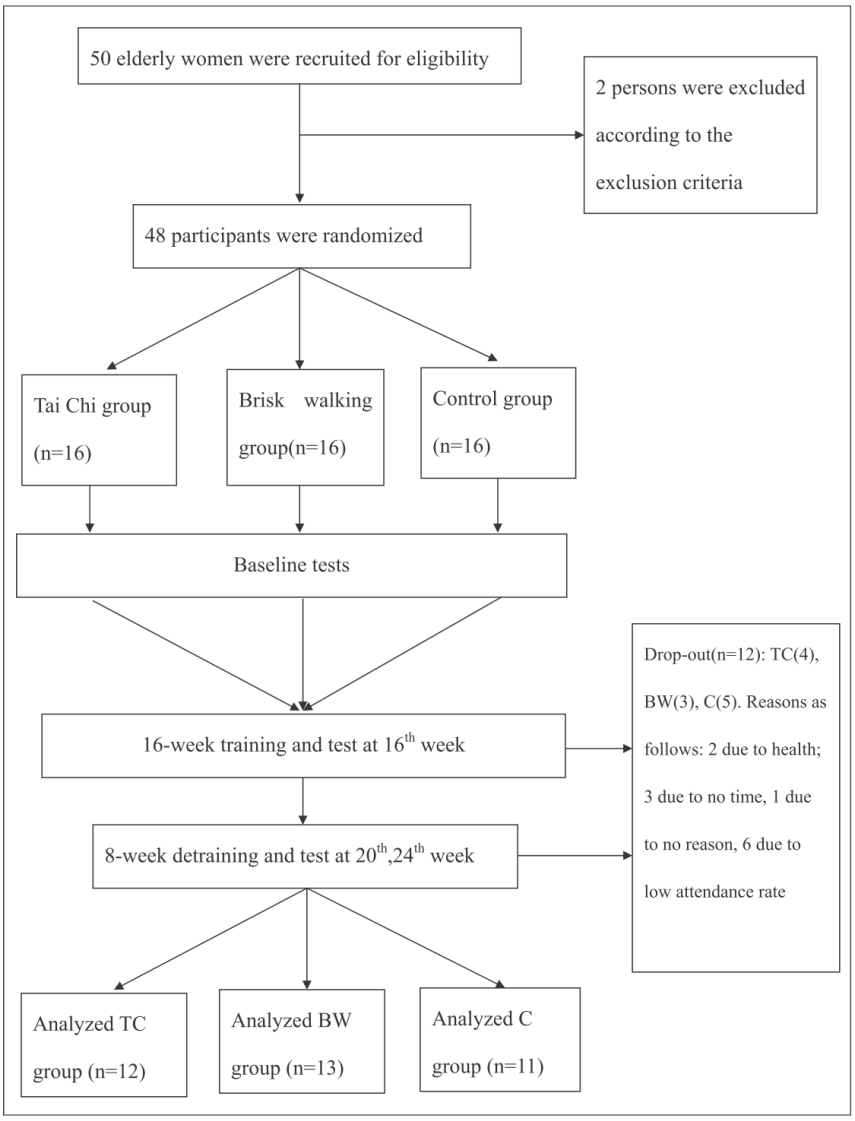


Figure 1. Flow diagram for randomized controlled trial.
TC, Tai Chi; BW, brisk walking; C, control.

The data were sampled at 17 Hz and low-pass filtered with a cut-off frequency of 6 Hz (Butterworth).[23] Each trial data of the first and the last 1 s were not considered for stability. Three trial data with the same visual condition were averaged for analysis. All output variables were calculated based on the mathematical formula in the previous study.[27,28] The maximal displacement of the COP in the AP direction (D-ap), the maximal displacement of the COP in the medial–lateral (ML) direction (D-ml), the total length of the COP trajectories (Lng), and the 95% confidence ellipse area of the COP movements (area), the mean AP total excursion velocities (V-ap), and the mean ML total excursion velocities (V-ml) were calculated.

Statistical analysis

The SPSS 17.0 was used for data analysis. All variables were presented as mean ± standard deviation. The variables of referring to four times were named $week_0$, $week_{16}$, $week_{20}$, and $week_{24}$ in this study. One-way ANOVA was employed to compare the differences of the demographic and baseline variables among the three groups. Two-way repeated ANOVA was used to determine the main effects of groups, time durations, and their interaction on the measurements. If any significant main and interaction effects were found, the Boneferroni method was conducted for post-hoc comparisons. The significant level was set at 0.05.

Results

Baseline characteristics of the participants

A total of 50 participants were screened for eligibility; 48 were qualified and were divided into three groups; and 36 participants (12, 13, and 11 in the TC, BW, and control groups) completed the whole 24-week study. Twelve participants dropped out because of health issues (2C), low attendance rate (3 B W, 3 TC), and no time (1 TC, 2C), and for no reason at all (1C) (Figure 1). The characteristics of the participants were showed in Table 1.

Postural control ability results during the 16-week training period

Tables 2 and 3 show no significant between-groups difference across all the variables at $week_0$ among the three groups. During the training periods, after the 16-week interventions, the participants in the TC and BW groups have significant better within-group performance than during $week_0$ and better between-groups performance compared with the control group in all the variables with the two visual conditions. No significant difference between pre- and post-exercise in the control group during training periods existed.

Postural control ability results during the 8-week detraining period

Table 2 shows that during detraining periods, the significant within-group difference compared with $week_0$and the significant between-groups difference compared with the control group were found across all the variables with the eyes open condition at the last three tests in the TC group. The significant within- and between-group differences were found in time, area, V-ap, and V-ml with eyes open at the last three tests in the BW group. However, the gains decreased significantly at $week_{24}$ in D-ap, D-ml, and V-ml compared with $week_{16}$ during detraining periods in the BW group. No significant difference was found in the control group.

Table 3 shows that during detraining periods, the significant within-group difference compared with $week_0$ and the significant between-groups difference compared with the control group were found across all the variables with the eyes closed condition, except for D-ml in the last three tests of the TC group. Significant within-group difference was found in D-ml at $week_{24}$ compared with $week_{16}$ in the TC group. With the eyes closed condition, significant within-group difference compared with $week_0$ and significant between-group difference compared with the control group in D-ml at $week_{16}$ and $week_{20}$ were found during the detraining periods in the TC group. With the eyes closed condition, significant within-group difference compared with $week_0$ and significant between-group difference compared with the control group in Lng, D-ap, V-ap, and V-ml at $week_{16}$, $week_{20}$ and $week_{24}$ were found during the detraining periods in the BW group. The gains decreased significantly at $week_{24}$ in time, area, and D-ml compared with $week_{16}$ during the detraining periods in the BW group. No significant difference was found in the control group.

Discussion

The first hypothesis was demonstrated in the present study. Our results showed that after the 16-week interventions, the postural control ability with two visual conditions in SLS improved in the TC and BW groups, which concurred with previous studies.[29,30] The variables assessing postural control, including the medial–lateral, total, and anterior–posterior path lengths of the COP trajectories, significantly improved after the TC[7] and BW exercise interventions.[17]

Regular physical activities,[31] especially moderate-intensity exercises,[32] could be helpful in improving postural control ability. Both exercises were moderate-intensity exercises with approximately 55% of maximal oxygen intake.[21,33] Several studies corroborated that TC was an effective practice to improve postural control.[7,8] A study by Zhou[7] examined the effects of 24 weeks of TC on the postural control of the older adults. The results validated that positive improvements were found in time, paths, and velocity of the COP in the TC group. In the current study, positive results were found after 16 weeks. Although the direct comparisons between two studies were infeasible because of different exercise frequencies, participants, and sample sizes, our findings still partly support that TC could improve postural control.

These positive effects of TC on postural control were related to various factors. Postural control ability is the integrated result from the center neural, peripheral nervous, and musculoskeletal

Table 1
The baseline characteristics of the participants.

N	Tai Chi group	Brisk walking group	Control group	F value	P value
	12	13	11		
Age (years)	64.12 ± 3.21	63.26 ± 2.20	65.36 ± 4.31	0.712	0.498
Weight (kg)	62.81 ± 8.37	62.00 ± 7.49	62.63 ± 7.21	0.004	0.996
Height (cm)	157.56 ± 5.45	158.50 ± 4.40	156.45 ± 4.43	2.607	0.089
BMI(kg/m^2)	25.12 ± 3.19	24.69 ± 2.97	26.21 ± 3.82	0.617	0.546

Table 2
Comparisons of study variables with eyes open in single leg stance among three groups.

	TC group (N = 12)	BW group (N = 13)	C group (N = 11)	time		group		time × group	
				P value	η^2_p	P value	η^2_p	P value	η^2_p
Time (second)				<0.001	0.337	0.014	0.229	0.011	0.152
$Week_0$	32.73 ± 16.69	40.80 ± 15.13	38.37 ± 13.73						
$Week_{16}$	55.61 ± 10.20[a,c]	55.03 ± 12.80[a,c]	39.94 ± 8.05						
$Week_{20}$	54.90 ± 10.95[a,c]	56.89 ± 8.78[a,c]	41.18 ± 11.07						
$Week_{24}$	54.20 ± 19.98[a,c]	56.90 ± 28.55[a,c]	39.18 ± 8.56						
Lng (mm)				<0.001	0.346	0.026	0.199	0.074	0.108
$Week_0$	422.57 ± 105.56	363.43 ± 104.7	427.93 ± 152.3						
$Week_{16}$	217.89 ± 45.82[a,c]	245.42 ± 80.48[a,c]	375.9 ± 69.78						
$Week_{20}$	256.27 ± 84.94[a,c]	260.73 ± 83.73[a,c]	357.29 ± 78.42						
$Week_{24}$	319.34 ± 82.98[a,c]	276.83 ± 90.25[a]	367.22 ± 84.32						
Area (cm^2)				<0.001	0.517	<0.001	0.570	<0.001	0.307
$Week_0$	1.45 ± 0.53	1.32 ± 0.61	1.38 ± 0.71						
$Week_{16}$	0.40 ± 0.11[a,c]	0.34 ± 0.25[a,c]	1.29 ± 0.20						
$Week_{20}$	0.44 ± 0.18[a,c]	0.35 ± 0.24[a,c]	1.32 ± 0.14						
$Week_{24}$	0.38 ± 0.13[a,c]	0.63 ± 0.38[a,c]	1.23 ± 0.40						
D-ap (mm)				<0.001	0.413	0.004	0.280	0.026	0.132
$Week_0$	30.07 ± 4.82	28.47 ± 7.96	25.95 ± 6.71						
$Week_{16}$	16.52 ± 3.62[a,c]	15.60 ± 5.64[a,c]	25.61 ± 6.51						
$Week_{20}$	16.46 ± 3.44[a,c]	15.44 ± 4.45[a,c]	27.59 ± 6.08						
$Week_{24}$	16.80 ± 3.49[a,c]	21.47 ± 6.58[b]	25.46 ± 7.22						
D-ml (mm)				<0.001	0.528	0.003	0.293	<0.001	0.417
$Week_0$	33.62 ± 11.49	31.97 ± 9.88	33.59 ± 7.32						
$Week_{16}$	18.61 ± 6.49[a,c]	16.46 ± 6.08[a,c]	29.26 ± 3.96						
$Week_{20}$	19.48 ± 8.22[a,c]	16.71 ± 5.00[a,c]	28.84 ± 8.30						
$Week_{24}$	22.85 ± 7.80[a,c]	24.30 ± 9.37[b]	28.18 ± 5.17						
V-ap (mm/s)				<0.001	0.418	0.002	0.302	0.027	0.130
$Week_0$	7.61 ± 2.61	7.26 ± 2.24	7.63 ± 1.66						
$Week_{16}$	4.22 ± 1.47[a,c]	3.74 ± 1.38[a,c]	6.64 ± 0.90						
$Week_{20}$	4.43 ± 1.86[a,c]	3.79 ± 1.13[a,c]	6.55 ± 1.88						
$Week_{24}$	5.19 ± 1.77[a,c]	4.52 ± 2.58[a,c]	6.40 ± 1.17						
V-ml (mm/s)				<0.001	0.524	0.003	0.301	<0.001	0.413
$Week_0$	6.49 ± 1.18	6.21 ± 1.73	5.66 ± 1.46						
$Week_{16}$	3.56 ± 0.83[a,c]	3.41 ± 1.22[a,c]	5.58 ± 1.41						
$Week_{20}$	3.54 ± 0.79[a,c]	3.36 ± 0.97[a,c]	6.02 ± 1.32						
$Week_{24}$	3.62 ± 0.81[a,c]	4.68 ± 1.43[b]	5.55 ± 1.57						

Abbreviations: Time, the single-leg stance time; Lng, the total length of the COP trajectories; Area, the 95% confidence ellipse area of the COP movements; D-ap, the maximal displacement of the COP in the anterior–posterior direction; D-ml, the maximal displacement of the COP in the medial–lateral direction; V-ap, the mean AP total excursion velocities; V-ml, the mean ML total excursion velocities.
[a] Denotes significant difference compared with the $week_0$ value within each group.
[b] Denotes significant difference compared with the $week_{16}$ value within each group.
[c] Denotes significant difference compared with the control group.

systems. Physical function declines with aging; however, regular TC could reshape brain structures, such as thickened cortex in the precentral gyrus and insula sulcus in the right hemisphere[34] and improve lower limb strength,[14] ankle and knee joint proprioception,[13] and neuromuscular reaction ability.[15] The abovementioned factors could help to improve balance and postural control after the intervention in the older adults. Moreover, BW is a popular moderate-intensity exercise, which could also enhance physical function similar to TC.[35] Our results were consistent with those of previous studies,[16,18] which indicated that after a 12-week BW, the participants significantly performed well in the postural control test.

Interestingly, no significant between-group difference was found in all variables at $week_{16}$ between the TC and BW groups. To the authors' knowledge, a 4-week TC[36] and a 12-week BW[16,18] could significantly improve postural control. Perhaps, TC was more efficient than BW in improving postural control. In our study, 16 weeks could be sufficiently long to improve postural control for the two intervention exercises. However, only two tests were performed before and after the intervention in the present study, and no more data to support our speculation came to light.

The second hypothesis was proven as follows: according to our data, during 8-week detraining periods, all variables with two visual conditions, except for D-ml with eyes closed in a single stance, indicated no significant decline in the TC group. The results confirmed that the maintenance of intervention gains for 8 weeks was good in the TC group. Our results were consistent with the findings from the study by Li et al.[8] The results corroborated that the gains of postural stability were maintained during the 12-week detraining periods.[8] The underlying mechanism can be attributed to the following factors in the present study. First, as aforementioned, regular exercise could have positive effects on the central nervous, musculoskeletal, and peripheral nervous systems to improve balance control. Once the plastic structure and physical function changes were established, adequate time to return to the original condition after the post-intervention would be necessary. In the present study, some decreasing trends emerged, but no significant differences came to light. The 8-week detraining may be insufficient to observe significant changes. This standpoint was supported by Miles and Eighmy's study,[37] which showed that experimental monkeys who wore telescopic, fixed-field, and dove prism spectacles for one week experienced vestibule-ocular reflex changes. However, after gaining adaptive reflex, the monkeys needed many days to readapt and readjust the vestibule-ocular reflex once the spectacles were off.[37] Moreover, another study[13] also proved that the improved proprioception at the ankle joint in the TC groups did not significantly decrease after the 8-week intervention was stopped. Finally, intervention exercises could

Table 3
Comparisons of study variables with eyes closed in single leg stance among three groups.

	TC group (N = 12)	BW group (N = 13)	C group (N = 11)	time		group		time × group	
				P value	η^2_p	P value	η^2_p	P value	η^2_p
Time (second)				<0.001	0.576	0.005	0.271	<0.001	0.373
$Week_0$	16.78 ± 7.10	15.63 ± 8.30	18.26 ± 7.63						
$Week_{16}$	39.95 ± 11.67[a,c]	31.68 ± 12.4[a,c]	19.11 ± 8.19						
$Week_{20}$	36.77 ± 13.47[a,c]	31.45 ± 11.44[a,c]	19.89 ± 6.93						
$Week_{24}$	35.53 ± 12.09[a,c]	26.76 ± 11.61[a,b]	20.41 ± 7.55						
Lng (mm)				<0.001	0.669	<0.001	0.575	<0.001	0.511
$Week_0$	519.51 ± 105.54	582.20 ± 85.69	545.79 ± 98.21						
$Week_{16}$	194.70 ± 83.76[a,c]	227.52 ± 113.06[a,c]	558.55 ± 90.29						
$Week_{20}$	231.19 ± 85.79[a,c]	241.95 ± 97.73[a,c]	553.50 ± 8870						
$Week_{24}$	278.14 ± 95.06[a,c]	278.43 ± 105.95[a,c]	518.98 ± 104.9						
Area (cm^2)				<0.001	0.362	0.002	0.304	0.072	0.108
$Week_0$	2.50 ± 1.15	2.28 ± 0.86	2.46 ± 0.91						
$Week_{16}$	1.14 ± 0.97[a,c]	1.25 ± 0.94[a,c]	2.17 ± 0.71						
$Week_{20}$	1.40 ± 0.81[a,c]	1.39 ± 0.81[a,c]	2.05 ± 1.00						
$Week_{24}$	1.58 ± 0.71[a,c]	1.69 ± 0.93[b]	2.10 ± 1.10						
D-ap (mm)				<0.001	0.401	<0.001	0.377	<0.001	0.232
$Week_0$	41.15 ± 6.12	40.07 ± 7.55	38.96 ± 9.50						
$Week_{16}$	25.88 ± 8.95[a,c]	21.54 ± 7.38[a,c]	37.98 ± 7.02						
$Week_{20}$	29.17 ± 8.47[c]	24.69 ± 8.15[a,c]	38.40 ± 7.14						
$Week_{24}$	29.63 ± 10.19[c]	25.70 ± 9.06[a,c]	39.43 ± 9.65						
D-ml (mm)				<0.001	0.418	<0.001	0.401	<0.001	0.289
$Week_0$	39.33 ± 8.85	41.75 ± 6.24	38.96 ± 9.50						
$Week_{16}$	25.40 ± 9.83[a,c]	30.52 ± 14.98[a,c]	36.16 ± 8.72						
$Week_{20}$	28.96 ± 10.47[a,c]	35.66 ± 17.25[a,c]	33.86 ± 5.98						
$Week_{24}$	36.17 ± 10.22[b]	36.36 ± 17.68[b]	35.80 ± 10.12						
V-ap (mm/s)				<0.001	0.405	0.005	0.269	0.045	0.119
$Week_0$	13.92 ± 2.68	12.69 ± 2.11	12.62 ± 3.45						
$Week_{16}$	7.93 ± 2.72[a,c]	7.80 ± 2.84[a,c]	10.31 ± 3.07						
$Week_{20}$	9.53 ± 2.33[a,c]	8.91 ± 3.10[a,c]	10.13 ± 3.32						
$Week_{24}$	9.39 ± 2.11[a,c]	9.39 ± 2.25[a,c]	10.54 ± 3.61						
V-ml (mm/s)				<0.001	0.376	<0.001	0.358	<0.001	0.213
$Week_0$	7.48 ± 1.11	7.28 ± 1.61	7.08 ± 1.72						
$Week_{16}$	4.71 ± 1.62[a,c]	3.91 ± 0.92[a,c]	6.91 ± 1.27						
$Week_{20}$	5.30 ± 1.53[a,c]	4.48 ± 1.36[a,c]	6.98 ± 1.29						
$Week_{24}$	5.38 ± 1.85[a,c]	4.67 ± 1.41[a,c]	7.17 ± 1.75						

Abbreviations: Time, the single-leg stance time; Lng, the total length of the COP trajectories; Area, the 95% confidence ellipse area of the COP movements; D-ap, the maximal displacement of the COP in the anterior–posterior direction; D-ml, the maximal displacement of the COP in the medial–lateral direction; V-ap, the mean AP total excursion velocities; V-ml, the mean ML total excursion velocities.
[a] Denotes significant difference compared with the $week_0$ value within each group.
[b] Denotes significant difference compared with the $week_{16}$ value within each group.
[c] Denotes significant difference compared with the control group.

improve balance control and decrease the incidence rate of falls.[8] The older adults could reduce the fear of falling and increase difficult physical activities in daily life. Conversely, these physical activities possibly further delayed the reduction of balance control.

In the BW group, a significant difference was found at $week_{24}$ compared to $week_{16}$. Postural control ability improvements were fully maintained for 4 weeks and partly for 8 weeks in the BW group. The differences on the maintenance of intervention gains during the detraining periods between TC and BW could be caused by different movement characteristics. Tai Chi referred to body–mind movements that required upper extremities to move in coordination with squatting leg movements and eyes to follow the hands. These characteristics may improve coordination of eyes, upper body, and lower extremities and be helpful to enhance postural control ability.[38] Moreover, participants concentrated their attention on slow movements of TC in practicing, which could improve cognitive function.[39] However, compared with TC, BW was a subconscious movement needing less coordination and concentration from participants.

It is noteworthy that in the present study, during the detraining periods, D-ml on two visual conditions and V-ml without vision at $week_{24}$ significantly increased compared with $week_{16}$ in the BW group. This result validated that the balance control maintenance effectiveness of BW in the ML direction was poor. The lack of balance in the ML direction could lead to falls, which was an important indicator of the risk of falling.[17] The poor maintenance effectiveness with eyes closed in single leg stance could be related to BW movement characters and visual condition. Walking movements, including the ankle/knee joint flexion and extension, repeatedly occurred in the sagittal plane. This special uniaxial movement character may be helpful for improving the musculoskeletal system function in the AP direction but not in the ML direction. In addition, the visual information input system was important for postural control. The balance control sway without visual feedback could increase by 20%–70% and rely on joint proprioceptive and vestibular feedback in the older adults.[40] However, a study affirmed that the ankle joint proprioception in the ML direction did not significantly improve during the 16-week BW.[13] The abovementioned factors may lead to poor postural control ability maintenance in the BW group. Therefore, the author recommends that the older adults could take TC to control balance in the ML direction.

This study has three limitations. First, only female participants were recruited; hence, the effects of the two exercises on the balance control with SLS in older men were not detected. Second, 8 weeks was not sufficiently long to measured significant differences during detraining; thus, further study should prolong the detraining periods. Thirdly, only 36 participants completed the entire 24-

week study, so the findings of this study should be interpreted with caution. Further studies with large sample sizes should be required to determine the detaining effects of TCC and BW intervention on balance in elderly.

Conclusion

The 24-form TC and BW significantly improved postural control ability with SLS after the 16-week training in older women. During the 8-week detraining, the gains of intervention were fully maintained in the TC group and partly maintained in the BW group.

Conflicts of interest

None.

Acknowledgments

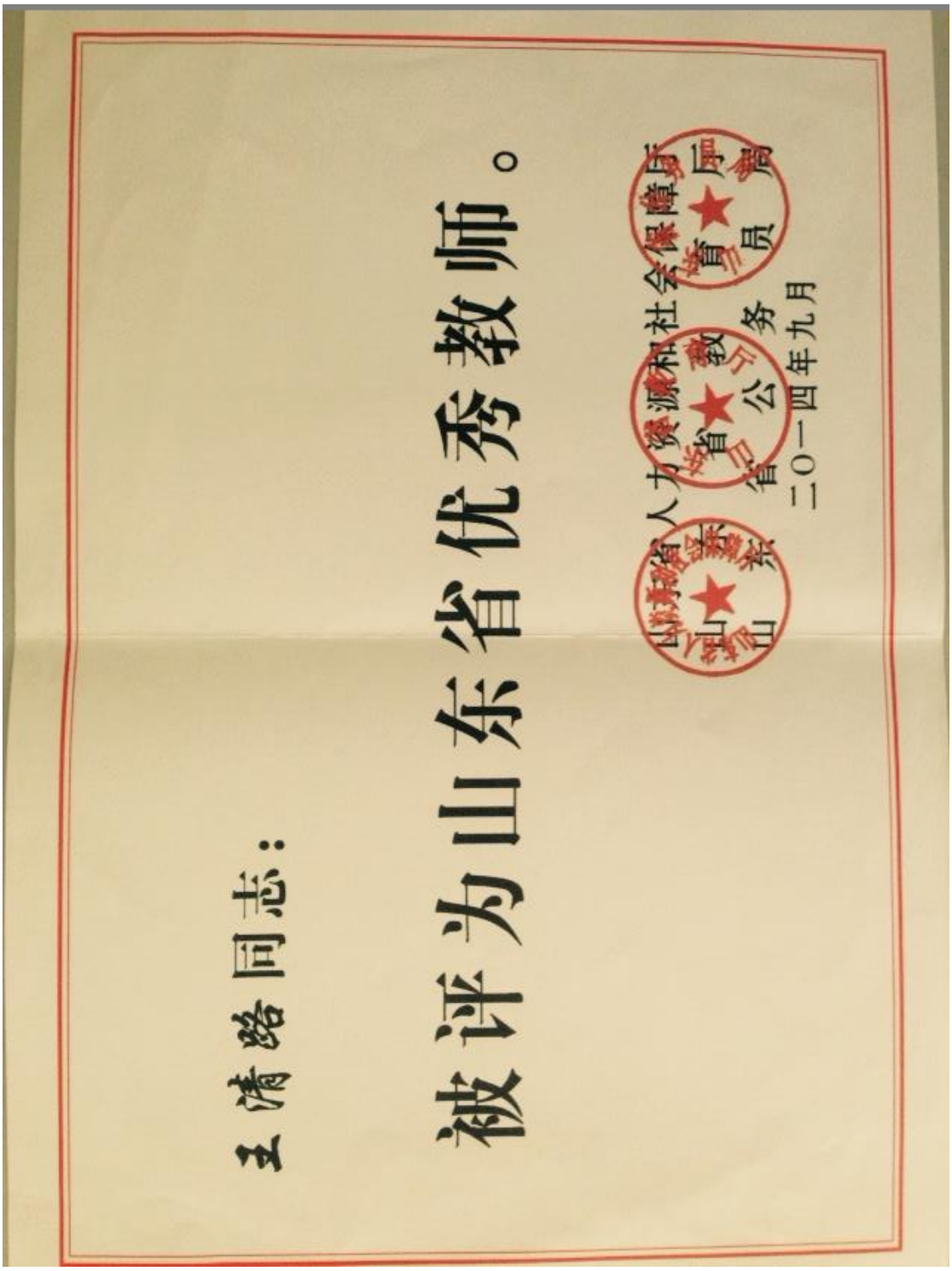
This work was supported by the National Natural Science Fund of China (No.31700815; 11572202).

References

1. Gill T, Taylor AW, Pengelly A. A population-based survey of factors relating to the prevalence of falls in older people. *Gerontology*. 2005;51:340–345.
2. Konak HE, Kibar S, Ergin ES. The effect of single-task and dual-task balance exercise programs on balance performance in adults with osteoporosis: a randomized controlled preliminary trial. *Osteoporos Int*. 2016:1–8.
3. Davis JC, Robertson MC, Ashe MC, Liu-Ambrose T, Khan KM, Marra CA. International comparison of cost of falls in older adults living in the community: a systematic review. *Osteoporos Int*. 2010;21:1295–1306.
4. Enderlin C, Rooker J, Ball S, et al. Summary of factors contributing to falls in older adults and nursing implications. *Geriatr Nurs*. 2015;36:397–406.
5. Novak AC, Komisar V, Maki BE, Fernie GR. Age-related differences in dynamic balance control during stair descent and effect of varying step geometry. *Appl Ergon*. 2016;52:275–284.
6. Zhang C, Mao D, Riskowski JL, Song Q. Strategies of stepping over obstacles: the effects of long-term exercise in older adults. *Gait Posture*. 2011;34:191–196.
7. Zhou J, Chang S, Yan C, et al. Effects of 24 weeks of tai chi exercise on postural control among elderly women. *Res Sports Med Int J*. 2015;23:302–314.
8. Li F, Harmer P, Fitzgerald K, et al. Tai chi and postural stability in patients with Parkinson's disease. *N Engl J Med*. 2012;366:511–519.
9. Hong Y, Li JX, Robinson PD. Balance control, flexibility, and cardiorespiratory fitness among older Tai Chi practitioners. *Br J Sports Med*. 2000;34:29.
10. Tsang WW, Huichan CW. Comparison of muscle torque, balance, and confidence in older tai chi and healthy adults. *Med Sci Sports Exerc*. 2005;37:280–289.
11. Tsang WW, Huichan CW. Effects of tai chi on joint proprioception and stability limits in elderly subjects. *Med Sci Sports Exerc*. 2003;35:1962–1971.
12. Lu X, Siu KC, Fu SN, Huichan CW, Tsang WW. Effects of Tai Chi training on postural control and cognitive performance while dual tasking - a randomized clinical trial. *J Compl Integr Med*. 2016;13:181–187.
13. Zhang C, Sun W, Yu B, Song Q, Mao D. Effects of exercise on ankle proprioception in adult women during 16 Weeks of training and eight weeks of detraining. *Res Sports Med*. 2014;23:1–12.
14. Xu DQ, Hong Y, Li JX. Tai Chi exercise and muscle strength and endurance in older people. In: Hong Y, ed. *Tai Chi Chuan. State of the Art International Research. Basel, Switzerland: Karger*. 2008:20–29.
15. Sun W, Zhang C, Song Q, et al. Effect of 1-year regular Tai Chi on neuromuscular reaction in elderly women: a randomized controlled study. *Res Sports Med Int J*. 2016;24:1–12.
16. Okubo Y, Osuka Y, Jung S, et al. Walking can be more effective than balance training in fall prevention among community-dwelling older adults. *Geriatr Gerontol Int*. 2015;16:1–8.
17. Paillard T, Lafont C, Costes-Salon MC, Rivière D, Dupui P. Effects of brisk walking on static and dynamic balance, locomotion, body composition, and aerobic capacity in ageing healthy active men. *Int J Sports Med*. 2004;25:539–546.
18. Gába A, Cuberek R, Svoboda Z, et al. The effect of brisk walking on postural stability, bone mineral density, body weight and composition in women over 50 years with a sedentary occupation: a randomized controlled trial. *BMC Wom Health*. 2016;16:1–10.
19. Lan C, Chen S, Lai J. Relative exercise intensity of tai chi Chuan is similar in different ages and gender. *Am J Chin Med*. 2004;32:151–160.
20. G.Hopkins W. Estimating sample size for magnitude-based inferences. *Sport Sci*. 2006;10:63–70.
21. Murtagh EM, Boreham CA, Murphy MH. Speed and exercise intensity of recreational walkers. *Prev Med*. 2002;35:397–400.
22. Franchignoni F, Tesio L, Martino MT, Ricupero C. Reliability of four simple, quantitative tests of balance and mobility in healthy elderly females. *Aging Clin Exp Res*. 1998;10:26–31.
23. Hijmans JM, Zijlstra W, Geertzen JH, Hof AL, Postema K. Foot and ankle compression improves joint position sense but not bipedal stance in older people. *Gait Posture*. 2009;29:322–325.
24. Gribble PA, Tucker WS, White PA. Time-of-day influences on static and dynamic postural control. *J Athl Train*. 2006;42:35–41.
25. Nilsson G, Ageberg E, Ekdahl C, Eneroth M. Balance in single-limb stance after surgically treated ankle fractures: a 14-month follow-up. *BMC Muscoskel Disord*. 2006;7:1–8.
26. Alsubiheen A, Jerrold P, Noha D, Everett L, Edward B. Effect of tai chi exercise Combined with mental imagery theory in improving balance in a diabetic and elderly population. *Med Sci Monit :Int Med J Exp Clinic Res*. 2015;21:3054–3061.
27. Prieto TE, Myklebust JB, Hoffmann RG, Lovett EG. Measures of postural steadiness: differences between healthy young and elderly adults. *IEEE Trans Biomed Eng*. 1996;43:956–966.
28. Howcroft J, Lemaire ED, Kofman J, Mcilroy WE. Elderly fall risk prediction using static posturography. *PLoS One*. 2017;12. e0172398.
29. Won SH, Kim JC, Oh DW. Effects of a novel walking training program with postural correction and visual feedback on walking function in patients with post-strokehemiparesis. *J Phys Ther Sci*. 2015;27:2581–2583.
30. Maciaszek J, Osinski W. Effect of Tai Chi on body balance: randomized controlled trial in elderly men with dizziness. *Am J Chin Med*. 2012;40:245–253.
31. Perrin P, Gauchard G, Perrot C, Jeandel C. Effect of physical activity and sporting activities on balance control in elderly people. *Br J Sports Med*. 1999;33:121–126.
32. Sherrington C, Tiedemann A, Fairhall N, Close JC, Lord SR. Exercise to prevent falls in older adults: an updated meta-analysis and best practice recommendations. *New S. Wales Public Health Bull*. 2011;22:78–83.
33. Li J, Hong Y, Chan KM. Tai chi: physiological characteristics and beneficial effects on health. *Br J Sports Med*. 2001;35:148–156.
34. Wei GX, Dong HM, Yang Z, Luo J, Zuo XN. Tai Chi Chuan optimizes the functional organization of the intrinsic human brain architecture in older adults. *Front Aging Neurosci*. 2015;6, 54–54.
35. Raffin J, Barthelemy JC, Terrat P, et al. Effects of brisk walking on autonomic nervous system reactivation in nursing home residents. Additional effects of transcutaneous vagus nerve stimulation. *Ann Phys Rehabil Med*. 2016;59S:e57–e58.
36. Tsang WW, Hui-Chan CW. Effect of 4- and 8-wk intensive Tai Chi Training on balance control in the elderly. *Med Sci Sports Exerc*. 2004;36:648–657.
37. Miles FA, Eighmy BB. Long-term adaptive changes in primate vestibuloocular reflex. I. Behavioral observations. *J Neurophysiol*. 1980;43:1406–1425.
38. Pei YC, Chou SW, Lin PS, Lin YC, Hsu TH, Wong AM. Eye-hand coordination of elderly people who practice Tai Chi Chuan. *J Formos Med Assoc*. 2008;107:103–110.
39. Sungkarat S, Boripuntakul S, Kanokwan Watcharasaksilp, Lord SR. Effects of tai chi on cognition and fall risk in older adults with mild cognitive impairment: a randomized controlled trial. *J Am Geriatr Soc*. 2017;65:721–727.
40. Lord SR, Dayhew J. Visual risk factors for falls in older people. *J Am Geriatr Soc*. 2001;49:508–515.

5 教师个人荣誉（部分）

5.1 山东省优秀教师（王清路）


王清路同志：

被评为山东省优秀教师。

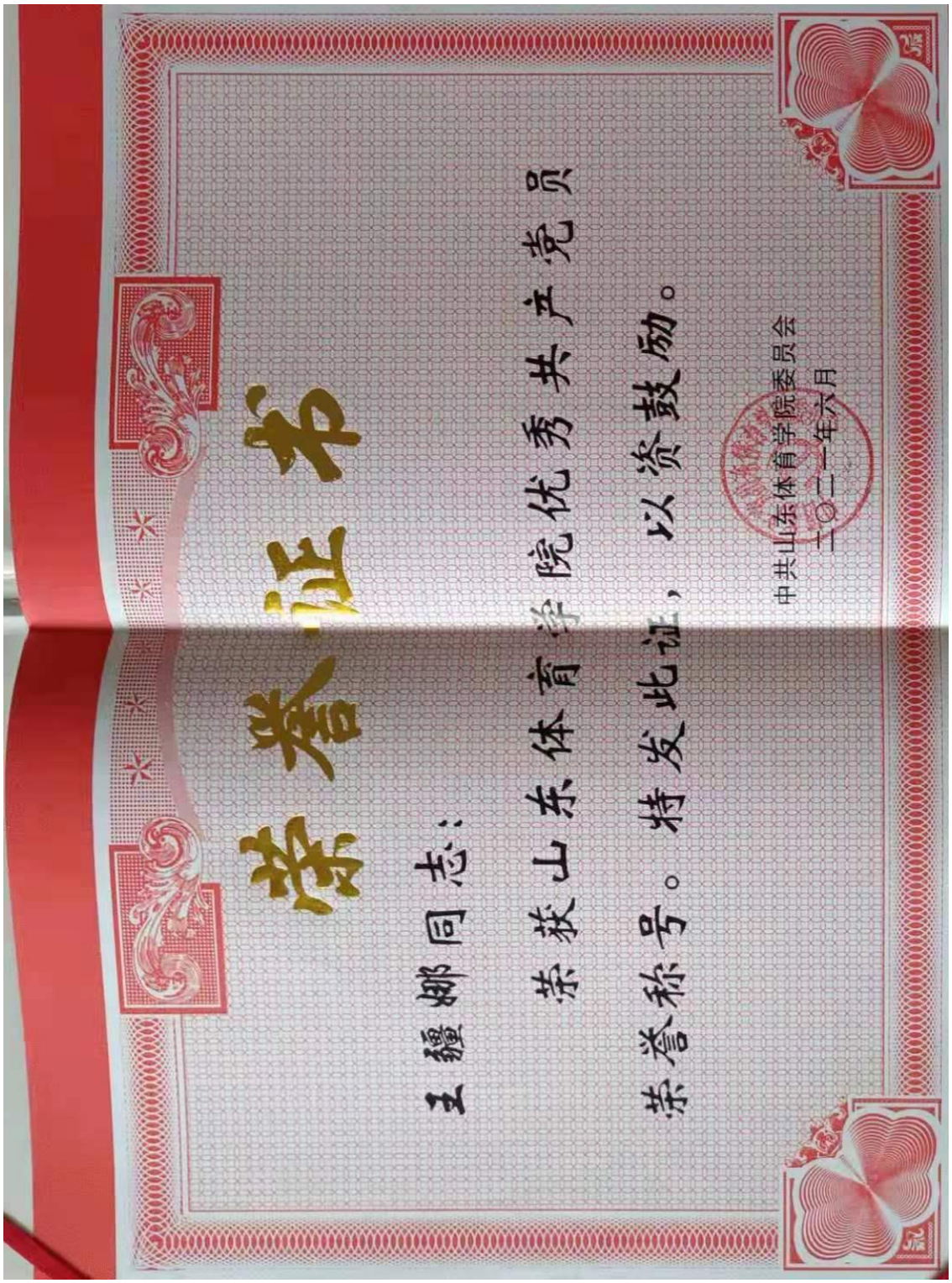
山东省人力资源和社会保障厅 山东省教育厅 山东省公务员局

二〇一四年九月

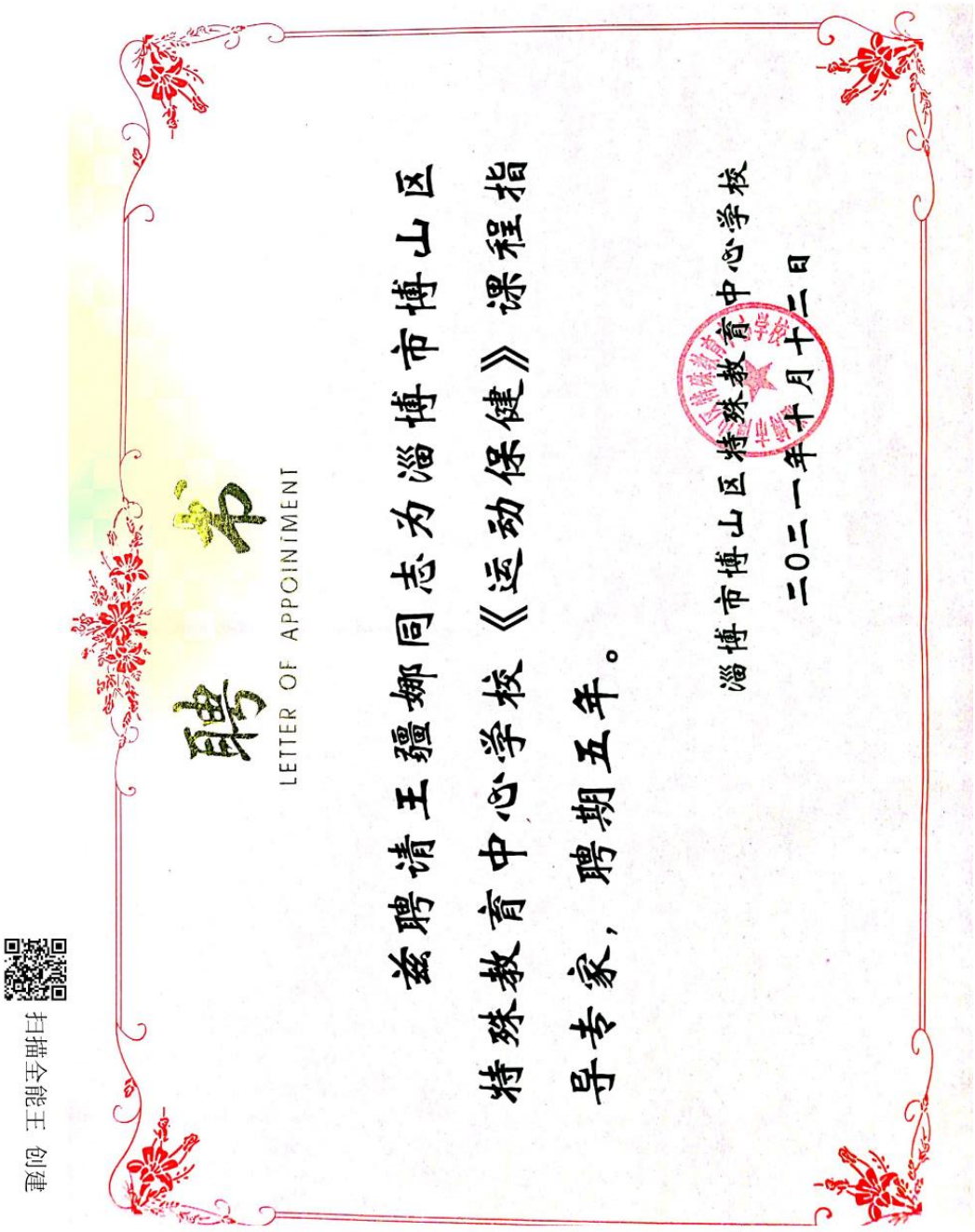
5.2 淄博市有突出贡献中青年专家（王清路）

5.3 泰山学者青年专家（田雪文）

5.4 校级优秀共产党员（王疆娜）



5.5 博山区特殊教育中心学校课程指导专家（王疆娜）



聘书

LETTER OF APPOINTMENT

兹聘请王疆娜同志为淄博市博山区特殊教育中心学校《康复训练》课程指导专家，聘期五年。

淄博市博山区特殊教育中心学校

二〇二一年十月十二日

扫描全能王 创建

6 媒体宣传类

6.1 中国教育报“根植齐鲁沃土 滋养体育人才”报道截图

中国教育報

ZHONGGUO JIAOYU BAO

2018年10月8日 星期一

山东体育学院实施以需求为导向的人才培养模式改革，坚持实践育人、特色发展——

根植齐鲁沃土 滋养体育人才

学习贯彻十九大精神 写好教育奋进之笔

福建建消除大班额工作台账

对存在大班额学校进行销号管理

安徽高校设艺术教育俱乐部

开启公共艺术教育新改革

6.2 中国教育报“全面健身体育教育责无旁贷”报道截图

02 中教评论

用专业化的教师教育培养卓越教师

当教师是在做一个知识农民

出门，我得从网上搜搜看先迈哪个脚！

防控网瘾刻不容缓

人性化是高校大门的开放之道

全民健身体育教育责无旁贷

消除留守儿童厌学情绪要破“心中贼”

教育观察

6.3 我院特殊教育专业学生赴济南市市中区未成年人素质教育基地

我院特殊教育专业学生赴济南市市中区未成年人素质教育基地

发布者：

时间：2019-05-12

2019年5月10日上午，我院特殊教育专业二十名学生赴济南市市中区未成年人素质教育基地开展“牵着蜗牛去散步”公益活动志愿服务。此次活动由济南市泺源学校主办，以“幸福同行，共享和谐——‘小蜗牛’迷你马拉松”为主题，旨在帮助心智障碍者融入社会，增强爱国家、爱家乡的意识。

活动一开始，孩子们和老师表演了精心准备的《非洲鼓》，拉开了本次活动的序幕。本次活动由三个环节组成，分别是：参观素质基地自然环境、体验活动（手工制作馆、捏塑坊、智能搭建）和迷你马拉松。

在“迷你马拉松”活动中，参与活动的每名心智障碍者由一名志愿者和普校小伙伴全程陪跑。我院志愿者主要负责引导、陪同心智障碍者跑完全程，并给其颁发奖品，切实感觉到了融合教育的必要性。

在此次活动中，我院志愿者良好的精神面貌和耐心细致的服务获得了各方的一致好评。通过参与活动，增进了志愿者们对特殊儿童的了解，学习了与特殊儿童相处的技巧，为今后专业课的学习打下了良好的基础。

6.4 我院参与“第二届‘GOSEN 杯’聋健羽毛球邀请赛”体育手语翻译工作

我院参与“第二届‘GOSEN杯’聋健羽毛球邀请赛”体育手语翻译工作

发布者：

时间：2019-12-26

每年的12月3日是国际残疾人日，为了进一步弘扬扶残、助残的传统美德，形成全社会理解、关心、帮助残疾人的良好氛围，进一步唤起社会对聋人群体的关注，传递社会正能量，持续公益活动。山东省羽毛球运动协会、山东金码头体育文化发展有限公司、山东省聋协联合于2019年12月7-8日在济南举办了第二届“GOSEN杯”聋健羽毛球邀请赛，山东体育学院特殊教育专业的十几名默言手语志愿者在张珺老师的带领下参加了本次赛事并担任赛事中的体育手语翻译服务工作。

12月7日正逢大雪时节的济南，凛冬悄然来袭，在这一抹寒冷万物皆知的景象里，山东体育学院特殊教育专业的十几名默言手语志愿者就像一抹抹跃动着的红色火焰温暖着整个山东省羽毛球赛场。在山东省乒乓球羽毛球运动中心六楼羽毛球训练馆，举办了一场特别的羽毛球比赛“聋健融合、共享羽乐”聋健羽毛球邀请赛，也是全国唯一首创的“聋健融合”的比赛模式。比赛中，在我院默言手语志愿者的手语翻译沟通协助下，聋人运动员们能够很好的与听人队友融合交流，共商双打战术切磋球技，发挥出了不输于听人的比赛水准，挑球、吊球、杀球样样技术精通，进攻排山倒海，防守滴水不漏，为本次比赛带来了一场视觉盛宴！志愿者们上下翻飞舞动着的手姿、聋听运动员交流会心的笑容与裁判员顺畅肯定的目光共同汇成赛场上一道美丽的爱心融合景象。

上一篇 下一篇

6.5 我院特殊教育志愿者参加“共享一片蓝天 梦想在此腾飞”融合趣味运动会

我院特殊教育志愿者参加“共享一片蓝天 梦想在此腾飞” 融合趣味运动会

发布者：

时间：2019-12-08

2019年12月7日，由济南春晖星儿家长支持中心举办、山东小蝌蚪教育承办、山东体育学院运动与健康学院协办的“共享一片蓝天 梦想在此腾飞”融合趣味运动会在山东体育学院体育场举行。60多个家庭、100多位志愿者，共计240人参加了融合趣味运动会。

本次运动会参与对象为心智障碍家庭、志愿者及相关友好单位，以志愿者与家庭一对一配搭方式共同参与，让身心障碍家庭的家长和孩子们体验一次真正的运动会，通过运动会的可观赏性，参与互动性，来体验亲子合作及家庭与家庭之间团队合作的快乐，使大家在轻松愉快的氛围中释放压力、放松心情。

我院特殊教育教研室主任张东亮带领18特教60余名志愿者参加了融合趣味运动会。18特教志愿者全程陪伴孩子参加各种比赛项目，充分体现出爱心和耐心，受到了孩子喜欢和家长的好评。通过参加融合趣味运动会，志愿者们理论联系实践，参与融合趣味运动会的策划和组织实施，提升了专业实践能力。

上一篇 下一篇

6.6 我院特殊教育专业志愿者参加山东省智协 2020 年 特奥篮球联谊活动

我院特殊教育专业志愿者参加山东省智协2020年 特奥篮球联谊活动

发布者：zgf

时间：2020-10-26

金秋十月，阳光明媚，10月24日，由中国智力残疾人及亲友协会、山东省残疾人联合会指导，山东省智力残疾人及亲友协会主办的“山东省智协2020年特奥篮球联谊活动”在山东省体育中心篮球公园举办。来自济南慧爱残疾人服务中心、济南雨寻心智障碍者小院的运动员以及来自山东大学、山东体育学院的大学生志愿者80余人参加了活动。

山东省残联党组成员、副理事长张文涛，山东省智协主席许淑娟，山东省体育中心体育馆馆长胡学江，山东省体育学院运动与健康学院特殊教育教研室主任张东亮，山东为君健身俱乐部负责人纪卫军等领导参加了活动。

山东体育学院运动与健康学院作为协办单位，18特教学生孙浩文、曾厚隆、邓锋、李渊、董恩和19特教学生张文龙、彭建军、白盈麒等8名同学作为志愿者担任本次特奥篮球联谊活动的裁判工作，顺利地完成了拍球、投篮、亲子传球等项目的裁判任务，得到了省智协主席和家长的赞扬。

6.7 我院 2019 级特教学生参加春晖星儿第四届融合趣味运动会志愿者活动

我院2019级特教学生参加春晖星儿第四届融合趣味运动会志愿者活动

发布者：zgf

时间：2020-11-26

在2020年11月22日下午，山东体育学院运动与健康学院19级特殊教育专业的同学来到济南市历城区特殊教育学校参加春晖星儿融合趣味运动会活动做志愿者工作。

本次运动会是由济南春晖星儿家长支持中心为主办方，山东小蝌蚪为承办方，山东体育学院特殊教育专业为协办方联合举办。经过工作人员及志愿者对各项工作的准备，迎来了我们的趣味运动会的开幕式。首先是各代表队入场，旗帜方队、运动员方队、家长代表方队和志愿者方队。然后介绍到场的领导，由杨艳丽女士致词。接下来由老师带领大家做准备活动，开始进行运动会比赛，每场比赛由一位老师指挥，让示范组做游戏示范，每一位志愿者带一位孩子完成比赛，游戏分为团体赛和个人赛，比赛结束后开始颁奖活动，最后进行的是家长组比赛。活动结束，合影留念。

通过近距离接触自闭症儿童和他们一起参加游戏活动，我们能明显的发现他们虽然能力有限，但在我们志愿者的帮助和鼓励下还是尽力去完成。有一些儿童是属于特别爱动的，不能静下来去认真玩游戏，但他们还是参与，虽然没有遵守规则，但他们在其中也获得游戏的乐趣。一下午的服务，志愿者们虽然身体感到十分的疲惫，但是想想自己所做出的贡献和回报总会感到很值得，在帮忙儿童的时候收获一个个微笑，能够帮忙他们是一种职责更是一种乐趣，或许有的时候会很累，但当我们志愿者和儿童一起参与游戏并获得奖品时，看到儿童那灿烂的微笑，感觉我们辛苦的付出是很值得的。本次活动是我们难得的社会实践活动，是这学期学习的特殊儿童体育游戏和特殊儿童鉴别与评估两门课程的实践机会。

6.8 19 特教志愿者参加山东省第三届听障儿童演讲比赛

19特教志愿者参加山东省第三届听障儿童演讲比赛

发布者：zgf

时间：2021-05-12

听力障碍，使很多孩子生活在无声的世界里。但是，一场特殊的演讲比赛让大家听到了天使的声音。在5月12号上午山东体育学院运动与健康学院的同学代表们就感受到了这美妙的声音。

为充分发挥典型引领作用，全面展示听障儿童乐观向上的精神面貌，以康复的可喜成果献建党一百周年。百周年，省残联、省教育厅在济南举办山东省第三届听障儿童演讲比赛决赛，由山东省听力语言康复中心承办。比赛地点山东教育卫视演播大厅。参赛选手是戴耳蜗或助听器他们通过抢救性语言康复训练，借助人工耳蜗、助听器，越过了重重障碍，走上了演讲比赛的舞台。

比赛严格按照程序进行，体现了公平公正的原则，评委们当场打分、公布成绩，经过一番激烈的角逐。在这个过程中他们的世界开始变得有声有色起来……

在比赛中，他们穿插的节目也是让我们眼前一亮，他们的演讲自然流畅，语言清晰，声情并茂，充满激情，表达了对家人、对社会的感恩之情。赢得在场观众的热烈掌声及评委的一致好评。这是令人震惊的一场比赛也是令人欢呼的一场比赛，也是在祖国100周年之际，送给祖国母亲最好的礼物！省言语康复中心周主任特地代表省残联领导对特殊教育专业的志愿者服务表示感谢！