

华大智造AgriHigh低深度全基因组测序组合产品赋能农业分子育种

华大智造基于多年技术积累开发了AgriHigh低深度全基因组测序组合产品 (AgriHigh Low-pass WGS整体解决方案)。基于该方案，本应用简报对384例猪耳样本开展低深度全基因组测序分析，所得结果与固相芯片的准确率或高深度WGS的一致性高度一致。该方案为动植物大规模育种提供了一种新选择，可实现超低成本、快速、自动化、高通量的基因分型全流程。

推荐应用: 农业基因组学

推荐机型: MGISEQ-2000RS, DNBSEQ-T7RS (测序仪)

MGISP-NE384RS, AlphaTool, MGISP-960RS, MGISP-100RS(自动化仪器)

- 大规模、高通量、低成本的低深度全基因组整体解决方案

AgriHigh Low-pass WGS 整体解决方案提供流程极简(3步建库)、通量灵活(96~1536样本/天)、自动化友好、分析高效(1536样本/24 h)的全流程解决方案,可获取百万级SNP和InDel信息,为畜禽遗传改良和分子育种提供足够遗传信息。

- 测序数据产出质量高

DNBSEQ测序技术具有高准确性,低重复序列率以及低标签跳跃等重要特性。



背景介绍

种质资源是人类赖以生存与农业发展的物质基础，也是作物、畜禽遗传改良和相关基础研究的物质基础。育种的本质是培育优良生物的生物学技术。纵观世界农业育种史，生物育种经历了传统育种(杂交育种)、诱变育种、倍性育种(单倍体育种、多倍体育种)、细胞工程和染色体工程育种、分子标记辅助育种、全基因组育种和基因编辑育种等¹。

基因组编辑技术为功能基因的挖掘和新品种的培育提供了契机，全基因组选择分子育种为功能基因组的应用提供了广阔平台。全基因组选择技术(Genomic Selection, GS)是基于覆盖全基因组的高密度分子标记的基因型和大量表型数据，估计每个与表型相关的标记的育种值，然后计算出全基因组预估育种值(GEBVs)，对所有与表型变异相关的位点予以检测，实现多个微效数量性状基因座(QTLs)贡献的精准解析，真正实现所有位点的高通量选择，通过在全基因组水平上聚合优良基因型，改良重要农艺性状²。

全基因组选择不但可以缩短育种年限，而且可以大幅提升育种效率。基因组选择技术在奶牛育种中已广泛应用，在猪、绵羊及肉牛育种中的应用正在稳步开展。随着育种目标的不断变化，更多地强调绿色、健康、环境友好型的动物生产，全基因组选择技术育种无疑能在种业创新中发挥更重要的作用。

在过去十年中，高通量基因分型技术(包括单核苷酸多态性 (SNP) 芯片³、简并基因组 (RRS)^{4,5}和全基因组测序(WGS)⁶)的快速发展显著增强了基因组预测在植物⁷和动物⁸中的应用。虽然芯片或简并基因组的低密度SNP数据可满足大部分应用场景，通过WGS实现更高SNP密度理论上可以提高基因组预测的准确性^{9,10}。尽管如此，使用WGS对数千个体进行测序的成本很高，限制了其在各种物种基因组预测中的广泛使用。低深度全基因组测序 (Low-pass WGS) 与数据填充相结合的方法应运而生¹¹。该方法不依赖于有限数量基因位点的高深度基因型测定，从数百万个随机采样的低置信度变异中收集信息，然后使用参考群库将这些信息用于填充可能的基因型。该参考群库由大量高深度WGS数据集组成，这些数据集代表了种群中的潜在单倍型。

华大智造DNBSEQ测序平台以高准确度和灵敏度、超低重复率、低标签跳跃率等优势越来越受到全球客户的广泛青睐。在农业科研与应用领域，该测序平台已广泛用于物种基因组解析、与性状相关的信号通路分析、等。基于DNBSEQ测序平台的Low-pass WGS技术已成功应用于猪、牛、鱼等物种的遗传分析与优良性状培育^{12,13}。基于上述情况，华大智造推出AgriHigh low-pass WGS的整体解决方案。该方案专为动植物大规模育种场景设计，以超低成本实现快速、自动化、高通量的基因分型全流程。种质资源是畜禽作物育种的基础，育种过程中样本的保存与处理尤为重要。针对生物育种大样本量的保存和使用需求，基于华大智造自身生物样本库，自动化设备、测序设备、信息处理的实力，华大智造开发了一整套从样本自动保存、自动建库、测序到分析的整套方案——AgriHigh Low-pass WGS Package。该方案通过灵活组合不同型号自动化设备与测序仪可实现年通量1.25万到40万，满足客户多样化需求。

AgriHigh Low-pass WGS方案全流程概述

样本储存

样本收集阶段往往需要持续较长时间，故需要合适的存储条件以保证样本的完整性。华大智造提供多种自动化样本库产品，如MGICLab-RG系列(适用于4°C~-20°C的血液样本临时存储)、MGICLab-LT系列(提供-80°C环境，适合长时间保存血液、组织、核酸等样本)和MGICLab-LN Pro系列(用于长期存储珍贵样本，提供稳定的液氮低温条件)，这些产品能够满足数十万甚至数百万样本的存储需求，为科研工作提供了可靠的样本管理解决方案。

核酸提取

华大智造的AgriHigh Low-pass WGS方案基于不同的样本通量需求，为客户提供基础版和标准版两种不同策略。

核酸提取阶段如手工操作可使用华大智造MGIEasy基因组DNA提取试剂盒(磁珠法)完成样本DNA的提取工作。

针对基础版，推荐使用MGIEasy基因组DNA提取试剂盒(磁珠法)在自动化仪器MGISP-960上进行样本的自动化提取，动物组织样本投入量范围为5mg~20mg。针对标准版，推荐使用MGIEasy基因组DNA提取预装试剂盒(MGISP-NE384)在自动化仪器MGISP-NE384上进行样本的自动化提取，样本投入量范围为5mg~20mg。相关自动化脚本已优化内嵌，无需客户验证。具体操作可参考相关说明书。

文库制备

文库制备推荐投入量为200ng gDNA，使用MGIEasy Large-scale PCR-FREE 全基因组低深度酶切文库制备试剂套装(板式)进行文库构建。文库构建包括打断末端修复，接头连接，等体积混样，具体操作可参考相关说明书。基础版和标准版方案均使用MGISP-960完成文库的构建。相关自动化脚本已优化内嵌，无需客户验证。

测序

文库通过一步法制备成DNA纳米球 (DNA nanoballs, DNBs)后，等质量混合后加载到MGISEQ-2000或DNBSEQ-T7测序仪上完成双端150bp(PE150)测序。

该方案以猪为例，每个样本推荐数据量为3GB (1~1.5×)。MGISEQ-2000(PE150)一张芯片可以上机96个样本，DNBSEQ-T7(PE150)一张芯片可以上机测序384个样本。

生信分析

本研究使用BOLT Low-pass WGS Software(BOLT_Low-pass)进行生信分析。BOLT_Low-pass是基于低深度测序(0.1×~10×)进行基因型填充(Imputation)的软件，该基因型填充步骤使用华大智造自研的高效软件。从测序下机的原始数据开始，数据质控、比对、排序等步骤使用华大智造的MegaBOLT生信分析加速器完成，极大提升了这些步骤的分析速度。

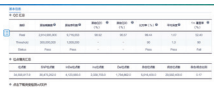
样本储存	核酸提取	文库制备	测序	数据分析
 MGICLab-RG系列 (4°C~ -20°C)	 MGIEasy基因组DNA提取试剂 盒(磁珠法)	 MGIEasy Large-scale PCR-FREE 全基因组低深度酶 切文库制备试剂套装(板式)	 DNBSEQ高通量测序试剂套装	 BOLT Low-pass WGS 软件数据包
AgriHigh Low-pass WGS 基础版				
 MGICLab-LT系列 (-80°C)	 MGISP-960 MGISP-100	 MGISP-960	 MGISEQ-2000RS (PE150)	 MegaBOLT生信分析加速器 (工作站式服务器)
AgriHigh Low-pass WGS 标准版				
 MGICLab-LN Pro系列 (液氮低温)	 MGISP-NE384 MGISP-960 AlphaTool	 MGISP-960	 DNBSEQ-T7RS (PE150) MGISP-100	 ZBOLT Pro生信分析加速器 (机架式服务器)

图1. AgriHigh Low-pass WGS 方案全流程概述。

研究概述

为验证Low-pass WGS基因分型在畜禽作物育种中应用的可行性，本研究选取大白猪和梅山猪两个品种的猪耳样本，获取1~1.5×WGS高精度SNP结果，后分别与50K固相芯片和50×WGS方法进行比较。综合分析表明Low-pass WGS和固相芯片、高深度WGS结果极为相近，分型一致性分别为97.8±1.1%和98.0±0.5%。本研究表明：Low-pass WGS可作为一种可用于大规模畜禽基因组研究和分子育种高效的方法。

研究结果

AgriHigh Low-pass WGS方案可高质量完成猪耳样本自动化核酸提取和建库

本研究使用DNBSEQ-T7RS方案对384个有芯片结果的猪耳样本(190例大白猪，194例梅山猪)进行Low-pass WGS分析。MGISP-NE384、AlphaTool和MGISP-960自动化设备在一天内完成384个样本的提取和建库，DNBSEQ-T7在24小时内完成测序，ZBLOT Pro在6小时内完成分析，整体周转时间(Turn around time, TAT)少于3天。

提取步骤通过MGISP-NE384 完成384个猪耳样本的DNA提取，产量均高于2μg，A260/A280介于1.7~2.0之间，符合实验要求(表1，图2，随机选取8个样本的提取信息和电泳结果展示)。

样本编号	提取浓度 (ng/μL)	260/280	260/230
S1	258.3	1.87	2.15
S2	235.2	1.89	2.35
S3	241.4	1.88	2.40
S4	263	1.90	2.30
S5	250.1	1.89	2.38
S6	218.5	1.87	2.24
S7	292.4	1.88	2.22
S8	218.9	1.87	2.24

表1.随机挑选的8个猪耳DNA样本的浓度以及纯度信息。

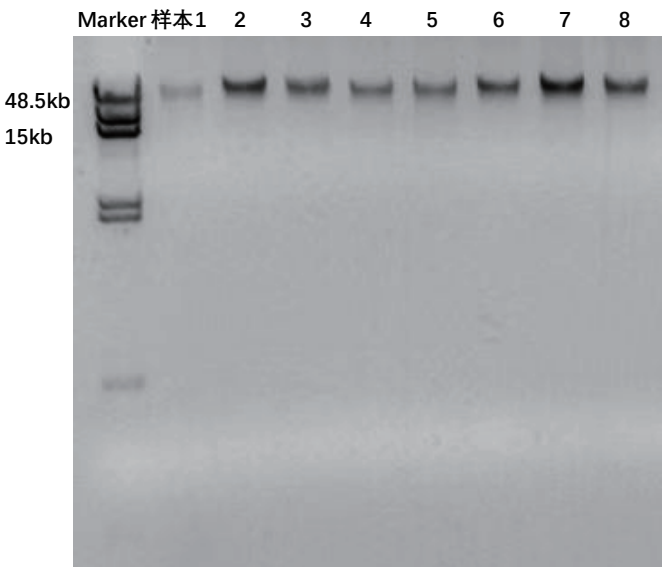


图2. 随机挑选的8个猪耳DNA样本的电泳结果展示。

在建库过程中，取200 ng基因组DNA，进行打断未修、接头连接和样本混合。48样本混合为一组，最终得到8个混合产物并进行后续产物纯化和DNB制备(表2)，8个DNB等质量混合进行后续的上机测序。

建库 起始量	连接产物等 体积pooling 48 pool 1	纯化产物 产量(ng)	OSV4 DNB 投入量 (ng)	DNB浓度 OSV4(ng/μL)
200ng	1	115	24	18
	2	105	24	16
	3	109	24	20
	4	92	24	18
	5	114	24	20
	6	103	24	18
	7	89	24	21
	8	98	24	18

表2. 文库构建过程中关键步骤的信息展示。

AgriHigh Low-pass WGS方案可获得高质量的测序数据

使用DNBSEQ-T7测序仪完成文库的测序工作，测序数据质量高，总体Q30为96.49%，具体每个循环的测序质量如图3A所示。下机数据共5634M reads，每个样本平均数据为14.7 M reads，比率为99.6±0.3%，测序深度为1.5±1×，基因组覆盖度为60.8±16.9%，1×测序深度数据条件下基因组覆盖度为53.7±2.5%(图3B)。通过模拟不同测序深度下全基因组的覆盖百分比可知，覆盖度随测序深度的增加而逐渐增加。

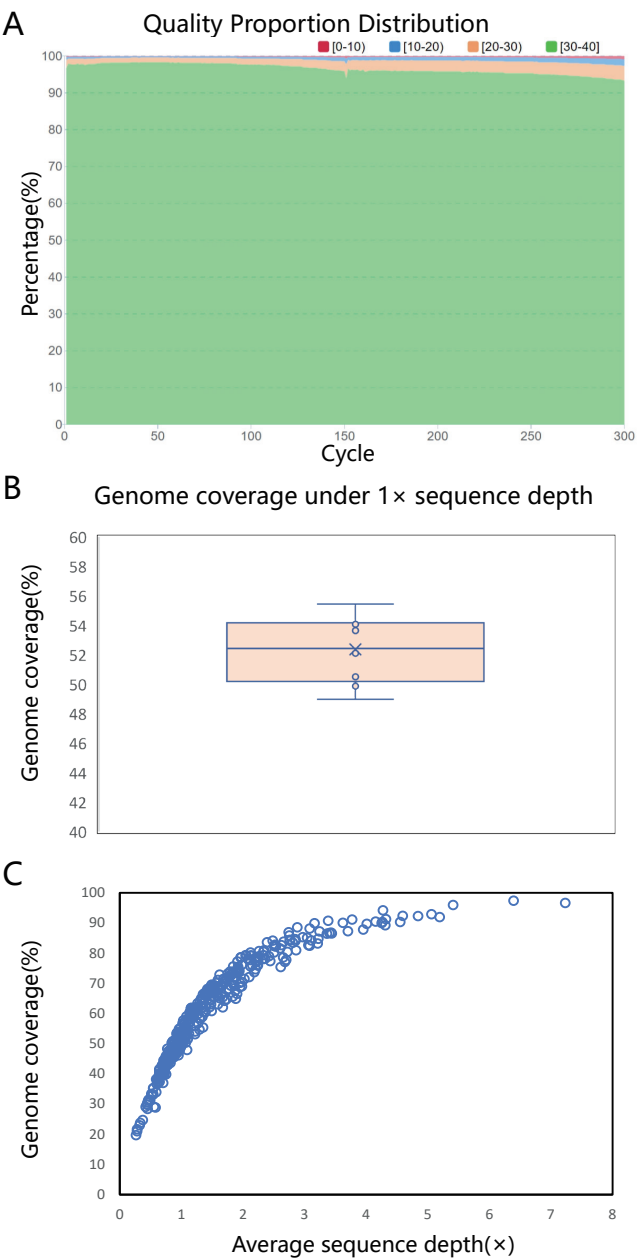


图3. A)测序数据质量图, B)1×平均测序深度下全基因组覆盖百分比; C)不同测序深度下全基因组覆盖百分比。

AgriHigh Low-pass WGS数据与固相芯片、高深度WGS无差别

利用自建的Reference Panel，对测序结果进行填充(Imputation)，Reference Panel覆盖了30,475,252个SNP位点和4,123,565个InDel位点。在所有个体中，实际平均检出的SNP数为8,524,950±1,705,032，实际平均检出的InDel数为1,293,998±175,194，远远超过芯片检测出的SNP数21382±4485个，且低深度全基因组测序覆盖了芯片上~90%以上的SNP位点(图4A)。采用RTG Tools工具对低深度测序(1×)Imputation结果和芯片共有的20,133±4,055个SNP进行准确性验证，

结果显示所有样本中的准确率(Precision)为98.7±1.1%，大白猪种群中的准确率为98.4±0.8%，梅山猪种群中的准确率为97.2±1.0%(图4B)。

同时，我们对其中6个大白个体和2个梅山个体进行了高深度测序(50×)，并使用GATK Genotype Concordance工具将低深度测序(1×)Imputation结果和全基因组高深度测序的共有检出位点进行比较，结果显示低深度数据和高深度数据在大白猪种群中的SNP和InDel一致性(Concordance)分别为98.1±0.5%和97.8±0.6%，在梅山猪种群中的SNP和InDel一致性分别为98.7%和97.5%(图4C)。

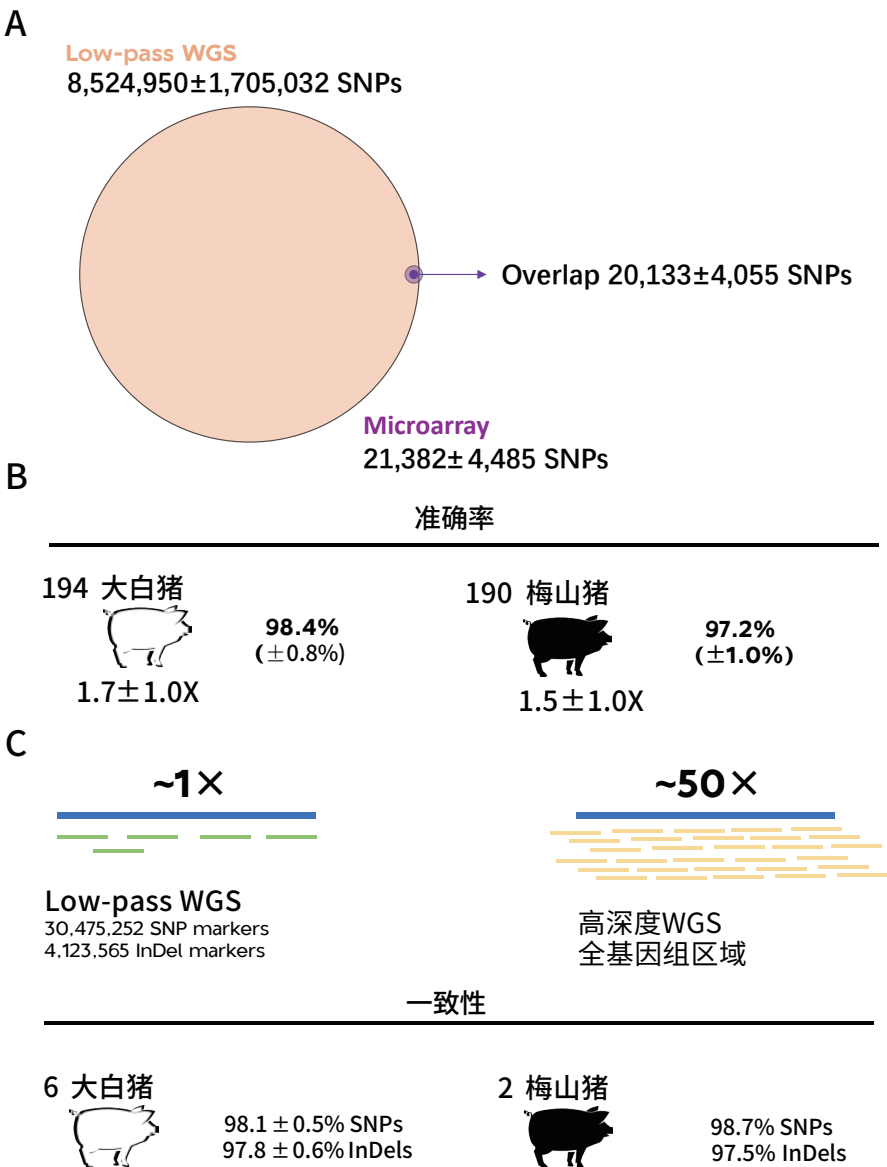


图4. 使用AgriHigh Low-pass WGS方案所获每个个体所检出的平均SNPs数量(A)和芯片方案共有的SNPs数量(A)，以及两种方案的准确率比较(B)；AgriHigh Low-pass WGS方案与高深度测序(50×)方案的SNP和InDel一致性(Concordance)比较(C)。

总结

AgriHigh低深度全基因组测序组合产品，旨在解决全球人口增长、气候变化和环境压力下的农业挑战。通过降低测序成本和技术创新，为农业领域提供从核酸提取到数据分析的全流程组合产品，缩短育种周期，提升育种效率，促进农民、育种者和研究人员培育性能优越的品种，减少资源消耗，促进农业的可持续发展。



MGISEQ-2000 基因测序仪



DNBSEQ-T7 基因测序仪

参考文献

1. Bawa, G., Liu, Z., Yu, X., Tran, L. P. & Sun, X., Introducing single cell stereo-sequencing technology to transform the plant transcriptome landscape. *TRENDS PLANT SCI* 29 249 (2024).
2. Desta, Z. A. & Ortiz, R., Genomic selection: genome-wide prediction in plant improvement. *TRENDS PLANT SCI* 19 592 (2014).
3. LaFramboise, T., Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *NUCLEIC ACIDS RES* 37 4181 (2009).
4. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L., Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE* 7 e32253 (2012).
5. Chen, Q. et al., Genotyping by genome reducing and sequencing for outbred animals. *PLOS ONE* 8 e67500 (2013).
6. Giani, A. M., Gallo, G. R., Gianfranceschi, L. & Formenti, G., Long walk to genomics: History and current approaches to genome sequencing and assembly. *COMPUT STRUCT BIOTEC* 18 9 (2020).
7. Parveen, R. et al., Understanding the genomic selection for crop improvement: current progress and future prospects. *MOL GENET GENOMICS* 298 813 (2023).
8. Georges, M., Charlier, C. & Hayes, B., Harnessing genomic information for livestock improvement. *NAT REV GENET* 20 135 (2019).
9. Meuwissen, T. & Goddard, M., Accurate prediction of genetic values for complex traits by whole-genome resequencing. *GENETICS* 185 623 (2010).
10. Iheshiulor, O. O., Woolliams, J. A., Yu, X., Wellmann, R. & Meuwissen, T. H., Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *GENET SEL EVOL* 48 15 (2016).
11. Martin, A. R. et al., Low-coverage sequencing cost-effectively detects known and novel variation in under-represented populations. *AM J HUM GENET* 108 656 (2021).
12. Yang, R. et al., Accelerated deciphering of the genetic architecture of agricultural economic traits in pigs using a low-coverage whole-genome sequencing strategy. *GIGASCIENCE* 10 (2021).
13. Zhuangbiao Zhang, A. W. H. H., The efficient phasing and imputation pipeline of lowcoverage whole genome sequencing data using a highquality and publicly available reference panel in cattle. *Animal Research and One Health* 1 4 (2023).

推荐订购信息

AgriHigh Low-pass WGS 基础版

产品类型	产品名称	产品货号
样本存储	MGICLab-RG低温自动化样本库	定制化产品
	MGICLab-LT -80℃自动化生物样本库	定制化产品
	MGICLab-LN55K Pro自动化液氮存储系统	900-000863-00
自动化仪器	基因测序文库制备仪MGISP-100RS	900-000070-00
	MGISP-960RS 高通盘自动化样本制备系统-订制化配置7	900-000107-00
测序仪	基因测序仪MGISEQ-2000RS	900-000035-00
提取试剂	MGIEasy基因组DNA提取试剂盒(磁珠法)	940-000972-00
建库试剂	MGIEasy Large-scale PCR-FREE	940-002162-00
	全基因组低深度酶切文库制备试剂套装 (96RXN)	
测序试剂	DNBSEQ -步法DNB制备试剂盒V4.0	940-001654-00
	MGISEQ-2000RS 高通量测序试剂套装(FCLPE150)	1000012555
	CPAS 条形码引物3试剂盒	1000020834
分析	MegaBOLT 生信分析加速器(工作站式服务器)	970-000085-00
	BOLT Low-pass WGS软件数据包(96人份)	970-000519-00

AgriHigh Low-pass WGS 标准版

产品类型	产品名称	产品货号
样本存储	MGICLab-RG低温自动化样本库	定制化产品
	MGICLab-LT -80℃自动化生物样本库	定制化产品
	MGICLab-LN55K Pro自动化液氮存储系统	900-000863-00
自动化仪器	全自动核酸提取纯化仪MGISP-NE384RS	900-000356-00
	基因测序文库制备仪MGISP-100RS	900-000070-00
	MGISP-960RS 高通量自动化样本制备系统-订制化配置5	900-000100-00
测序仪	AlphaTool自动化移液机器人	960-001238-00
	全自动样本加载仪MGIDL-T7RS	900-000132-00
	基因测序仪DNBSEQ-T7RS	900-000082-00
	数据分析一体机MGI-ZTRON-LITE	900-000406-00
提取试剂	24孔试剂底座(带试剂槽)	962-000068-00
	MGIEasy基因组DNA提取预装试剂盒	940-000974-00
建库试剂	MGIEasy Large-scale PCR-FREE	940-002049-00
	全基因组低深度酶切文库制备试剂套装 (384RXN)	
测序试剂	DNBSEQ -步法DNB制备试剂盒V4.0	940-001654-00
	DNBSEQ-T7RS 高通量测序试剂套装(FCLPE150) V3.0	940-000268-00
	CPAS 条形码引物3试剂盒	1000020834
分析	ZBOLT Pro 生信分析加速器(机架式服务器)	900-000460-00
	BOLT Low-pass WGS软件数据包(96人份)	970-000519-00

深圳华大智造科技股份有限公司

深圳市盐田区北山工业区综合楼11栋

4000-688-114

www.mgi-tech.com

MGI-service@mgi-tech.com

股票简称：华大智造

股票代码：688114



仅供研究使用

版权声明：本手册版权属于深圳华大智造科技股份有限公司所有,未经本公司书面许可,任何其他个人或组织不得以任何形式将本手册中的各项内容进行复制拷贝、编辑或翻译为其他语言。本手册中所有商标或标识均属于深圳华大智造科技股份有限公司及其提供者所有。

版本：2024年7月版

撰稿：胡豆，杨林，张韶红，梁鑫明

责任编辑：王其伟

审稿：江遥