PCR 产物的高通量测序方法及优化

李桂澜[1]，　胥传来[2]

（1. 北京大学 生命科学学院，北京 100871；2. 食品科学与技术国家重点实验室，江南大学，江苏 无锡 214122）

摘要：PCR 产物的高通量测序被广泛应用于功能基因筛选、肿瘤相关基因的突变和甲基化检测等。在高通量测序技术中，建库和上机测序实验一直是决定最终 DNA 数据质量的关键。作者优化了 PCR 产物样品的 DNA 文库制备条件和体系，设计了适合 PCR 产物特点上机测序方法，将已建库的 PCR 样品中混入一定量的基因组标准品后再上机测序，由此保证了测序数据的高质量和低冗余度。为低多样性 DNA 样品高通量测序技术方法运用提供方法上的参考。

关键词：PCR 产物；高通量测序技术；样品文库制备；上机测序

中图分类号：Q 523.8　　**文献标志码**：A　　**文章编号**：1673—1689(2016)12—1317—06

Methods and Optimization of High Throughput Sequencing Technologies on PCR Products

LI Guilan[1]，　*XU Chuanlai*[2]

（1. School of Life Sciences，Peking University，Beijing 100871，China；2. State Key Laboratory of Food Science & Technology，Jiangnan University，Wuxi 214122，China）

Abstract：The development of the new generation sequencing technology extends the application field of gene sequencing. High throughput sequencing of PCR products has been widely used in functional gene screening，mutation and methylation detection of tumor related genes，etc. In the high throughput sequencing technology，database and computer sequencing experiments are the key factor to decide the quality of DNA data. Based on Illumina sequencing instruments，the sample preparation method and sequencing running for PCR products were optimized，the results make us clearly understand the primary features of high-throughput sequencing techniques，and provide us important references of sequencing methods for low diversity DNA samples to address biological questions of interest.

Keywords：PCR products，high-throughput sequencing，sample library preparation，sequencing runs

DNA 测序是常用的分子生物学研究技术，通过测序分析能提供最真实可靠的基因序列信息。从 1977 年第一代传统 DNA 双脱氧链末端终止的 sanger 测序法问世以来，DNA 测序技术经历了快速发展[1-2]。现在基于边合成边测序（sequencing by synthesis）为主的第二代高通量测序技术逐渐被广

收稿日期：2015-02-03

基金项目：国家“十二五”科技支撑计划项目(2012BAC01B071)。

作者简介：李桂澜(1980—)，女，重庆人，理学博士，工程师，主要从事 DNA 测序技术的应用及生物分子相互作用方面的研究。

E-mail：liglan@pku.edu.cn

泛运用于基础研究、临床医学诊断和肠道微生物群与营养及代谢研究领域[2-5]。二代测序以单个读长较短、通量大为特点,实现了对单个物种的转录组和基因组细致全貌分析,也被称为深度测序或高通量测序[6]。二代测序研究平台从最初 Roche 公司的 454 焦磷酸测序和 Illumina 公司早先的 Solexa 测序技术发展到目前广泛使用的 Hiseq、Miseq 系列, 和 Life Technoloiges 的 Ion PGM 和 Ion Proton 的测序,测序技术无论从速度和通量上得到了前所未有的发展[7]。目前以 Illumina 为技术平台的 Hiseq X 5 系统、Hiseq3000/4000 最新的测序仪器也相继发布,使二代测序样品建库更加高效,读长(reads)更长,通量更高。

PCR 产物测序通过设计恰当引物探针,运用各种 PCR 技术将待测序的目的基因片段扩增产物进行测序[8-10]。PCR 产物的高通量测序技术被辅助运用于细胞文库中功能基因的筛选,肿瘤相关基因突变及甲基化检测和微生物菌群的营养及代谢相关探索研究[11-12]。以 Illumina 为平台的高通量测序,对片段大小在 150~350 bp 的 PCR 混合物样品, 与基因组 DNA 和微量的 ChIP-seq (Chromatin Immuno Precipitation,ChIP)样品的文库制备方法不同[13-14]。虽然目前建库技术整体一直在被优化,测序通量和测序长度也在不断的进行技术革新使之达到更高的测序要求, 但 Illumina 的边合成边测序技术核心仍无法改变, 也决定了 Illumina 二代测序的技术局限性。所以根据 PCR 产物样品自身的特点,文库制备和上机测序实验都需要进行特别的设计和处理。作者主要以 Illumina 系列的二代测序平台, 在普通的 ChIP 建库方法,基于片段已知的 DNA 样品通过两步末端修饰反应后加 DNA 通用接头的实验原理,设计建立了一种适合 PCR 混合物样品性质特点的文库制备及上机操作实验,以确保高效低成本的建库技术及上机测序中得到最佳的测序数据质量。作者以 Illumina 为代表的高通量测序技术特点及具体实验方法,为不同的研究领域中选择不同测序方法提供参考。

1 材料与方法

1.1 试剂与材料

1.1.1 试剂 ChIP-Seq Sample Prep Master Mix 试剂盒: 美国 New England Biolabs 公司;NEBNext Multiplex Oligos;Agencourt AMPure XP 的 DNA 纯化磁珠: 美国 BECKMAN COULTER 公司;Brilliant SYBR Green QPCR 试剂盒:美国 Agilent 公司;上机双端测序试剂 Miseq Reagent Kit V3 (2×150 cycles): 美国 Illumina 公司;100 bp plus DNA Marker 相对分子质量标准品:北京全式金生物技术公司;100% 乙醇、Tris、Tween20 和 NaOH: 美国 Sigma 公司。1.5 mL 低 DNA 吸附(LoBind)微量离心管:美国 Eppendorf。

1.1.2 DNA 分子 New England Biolabs 公司提供:接头分子(Adaptor):5'-pGATCGGAAGAGCACACGTCTGAACTCCAGTC/ideoxyU/ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3';上游引物:5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3';下游引物:5'-CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3';下游引物中的下划线加粗的 6 个碱基为第 10 号标签(index)DNA 分子。Takara 合成荧光定量 PCR 引物:上游引物 P1:5'-AATGATACGGCGACCACCGA-3';下游引物 P2:5'-CAAGCAGAAGACGGCATACGA-3'。

1.2 仪器

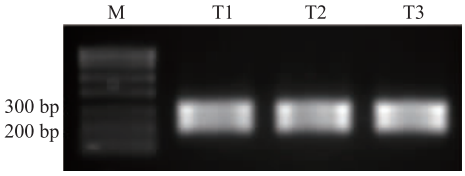
Miseq 第二代高通量测序仪: 美国 Illumina 公司;美国 Bio-Rad 核酸电泳仪、凝胶成像仪和 PCR 核酸扩增仪; 电子天平和 pH 计: 瑞士 METTLER TOLEDO 公司;NanoDrop2000 超微量分光光度计:美国 Thermo Scientific 公司;Qubit 2.0 荧光定量仪:美国 Life Technologies 公司;Mx300P 荧光定量 PCR 仪:美国安捷伦公司;Fragment Analyzer 全自动毛细管电泳仪及其配套试剂:美国 Advanced Analytical。ThermoMixer 恒温混匀仪:美国 Eppendorf。

1.3 实验方法

1.3.1 测序样品文库制备 切胶纯化回收后的 PCR 产物样品 T1、T2 和 T3, 琼脂糖凝胶电泳图中条带是分布在 200~300 bp 宽峰, 见图 1。用 NanoDrop2000 测定质量浓度分别为 24.2、25.2、26.2 ng/μL;A_{260}/A_{280} 的 OD 值为 1.68、1.59 和 1.76。

建库的初始样品用 Qubit 2.0 荧光定量仪对 PCR 产物样品准确定量后各取 2 μL (DNA 总量在 10~50 ng)于 1.5 mL 的 Lobind 离心管,准备第一步的平末端修复反应:1 μL 末端修复反应聚合酶和 5

μL 的 10× 末端修复反应缓冲液,用 RNase free H_2O 补足至反应终体积 50 μL,恒温混匀仪中 30 ℃反应 20 min。结束后加入 80 uL 的 AMPure XP 磁珠纯化反应后样品,用 44 uL RNase free H_2O 洗脱。第二步直接向样品溶液加入 1 μL Klenow 片段酶和 5 μL 的 10×反应缓冲液(含 0.2 mmol/L dATP),在 37 ℃反应 30 min,完成末端加"A"修饰反应。同样用 80 μL 磁珠纯化,19 μL 的 RNase free H_2O 洗脱。



M:100 bp plus DNA ladder

图 1 二代测序的 PCR 产物样品 T1–T3

Fig. 1 PCR products T1–T3 for next-generation sequencing

将 DNA 接头分子稀释到 5 μmol/L 后取 1 μL 到样品溶液中准备连接反应,加 6 μL 的 5× 快速连接缓冲液和 4 μL 的快速 DNA 连接酶,小心混匀,室温 25 ℃静置孵育 15 min。结束后加入试剂盒的 USER 酶 2 μL,在 37 ℃反应 15 min 完成接头开环剪切。最后用与样品等倍体积 30 μL 的磁珠纯化两遍,检测浓度。取 10 μL(总量在 5~10 ng)样品进行 PCR 反应,循环数减少至 15 次。最后用与样品等体积的磁珠纯化两遍,30 μL 的 ddH_2O 洗脱,检测浓度。

1.3.2 建库后样品质量检控 将建库后样品稀释至 0.5~2 ng/μL,分别用 Fragment Analyzer 全自动毛细管电泳仪进行片段分析;用 Mx300Ps 荧光定量 PCR 仪对样品进行荧光定量检测。选用的标准品为已准确定量并上机测序后的建库样品。用 10 mmol/L Tris–HCl pH 8.0,0.05% Tween 20 缓冲液按 10 倍梯度配制成 0.002~20 pmol/L 五个不同浓度以获得定量标准曲线。建库样品再次稀释 10^4 倍到 0.1~2 pmol/L 后取 2 μL,2× Brilliant SYBR Green Master Mix 5 μL,10 μmol/L 引物 P1/P2 Mix 0.4 μL,最后 ddH_2O 补足为 10 μL 的反应体系,阴性对照组 NTC (No Template Control) 和所有样品及标准品做三组重复。

$$[\text{Final Conc (nmol/L)}]=[\text{QPCR}]\times 340\times 10^4\times 10^{-6}/\text{Fragm}$$

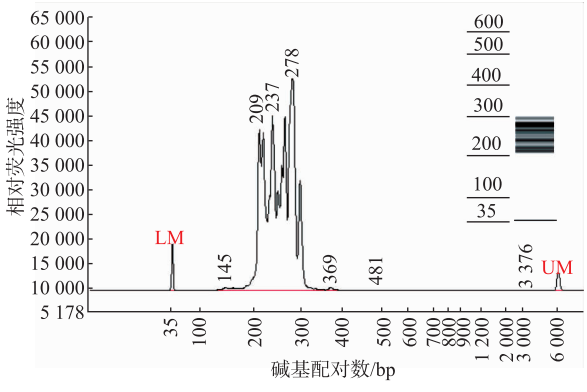
式中,[QPCR] 代表仪器根据标准曲线检测出来的平均样品浓度 (fM);340 bp 代表选用的 DNA 标准品的片段大小;10^4 代表样品稀释倍数;Fragm 代表建库后样品的平均片段大小。

1.3.3 Illumina Miseq 上机测序 对已建库样品 T2 准备上机测序,将 T2 样品与 Phix 基因组标准样品等摩尔混合,用 10 mmol/L Tris–HCl pH 8.0 缓冲液稀释至终浓度 2 nmol/L,取 10 μL 的上机混合样品,加入 10 μL 2 mol/L 的 NaOH 充分混匀后室温变性 5 min。用双端测序试剂盒里的杂交缓冲液将样品稀释到 18 pmol/L,取 600 μL 上机测序。Illumina 数据收集和分析软件:Miseq Control Software;Real Time Analysis (RTA);Offline Basecaller (OLB);CASAVA 软件用于进一步数据分析。

2 结果与分析

2.1 建库后样品检测

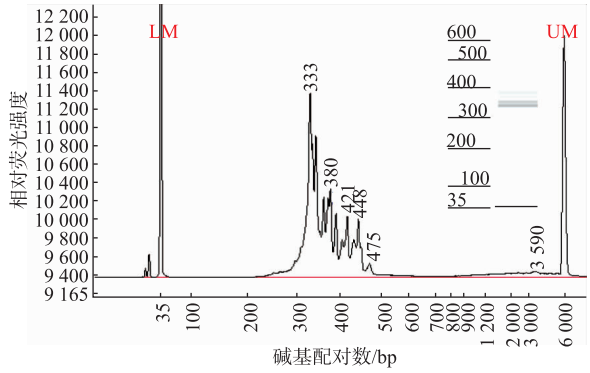
PCR 产物样品 T1,T2 和 T3 为同一样本的 3 个平行生物组重复。建库结束后检测终质量浓度为:22.2、15.3、16.1 ng/μL 各 30 μL 体系。随机选取样品 T2 的建库及上机测序结果分析。用 Fragment Analyzer 分析初始样品和建库后样品片段分布,见图 2–3。样品建库后片段的出峰位置出现移动,增加约 120 bp,满足接头序列成功连接后的片段分布情况。根据标准品所得的荧光定量标准曲线其相关系数为 Rsq:99.9%,计算得到稀释后的 3 个建库样品浓度 12.46、6.78、7.81 nmol/L,满足上机测序要求。T2 样品准备上机测序。



LM:代表 35 bp 小片段的 lower marker

图 2 高通量测序样品 T2 的 DNA 片段分析

Fig. 2 DNA fragment analysis of sample T2



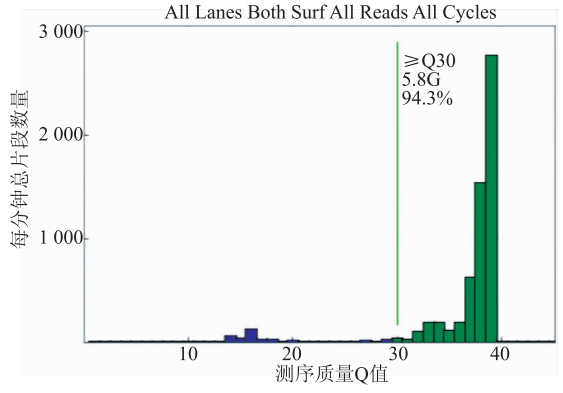
UM:代表 6 000 bp 大片段的 Upper Marker

图 3　测序样品 T2 建库后的 DNA 片段分析

Fig. 3　DNA fragment analysis of final library sample T2

2.2　上机测序结果

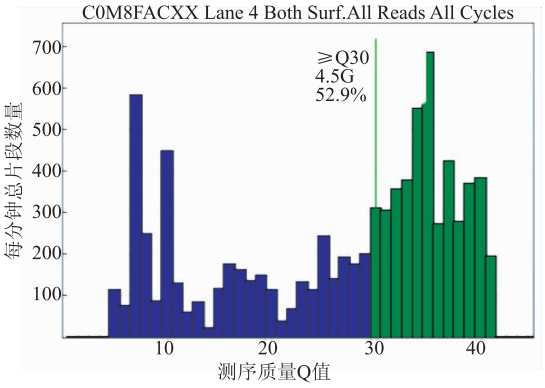
Illumina Miseq 测序中最主要指标参数是数据质量 Q30 值和测序数据量统计结果见图 4–5。其中百分比≥Q30 代表每轮测序碱基的质量不低于 99.9%所占的百分比。统计 T2 样品测序最终的 Q 值分布(横坐标)和得到的数据量(纵坐标,单位是 million)。统计得到全部测序总数据量为 5.8G,数据质量≥Q30 的占 94.3%。最终统计得到的 T2 样品原始测序数据量(Raw Data)为 3.725 G。进一步生物信息过滤处理得有效可用数据(clean data)为3.705 G,数据有效率为 99.46%,Q20 值为94.09%,Q30 值为 92.85%;GC 含量比例为 48.19%。结果满足后续对 T2 测序样品的生物学研究分析的基本要求。



■为测序质量≥Q30 的数据量统计分布情况，■为测序质量<Q30 的数据量统计分布。

图 4　样品 T2 在 Miseq 上机测序得到的全部数据 Q 值(Quality Score)分布统计

Fig. 4　Statistic of the QScore distribution on illumina Miseq



■为测序质量≥Q30 的数据量统计分布情况，■为测序质量<Q30 的数据量统计分布。

图 5　Hiseq2000 的测序芯片（flowcell）中第 4 条测序通道(lane 4)里所有样品都只是 PCR 产物样品测序时的数据 Q30 分布统计

Fig. 5　Statistic of the Qscore distribution without the phix standard mixture on lane 4 at Hiseq 2000

2.3　上机测序实验优化分析

当前二代高通量测序正突飞猛进地不断进行技术革新，但以 Illumina 为平台的边合成边测序的核心技术原理保持不变[2]。在上机测序过程中前 25 个碱基测序循环(cycle)结果质量的统计方法无法改变,依旧决定了整个测序的数据质量。测序中监测参数 Cluster PF(Pass Filter)仍是重要的上机质检标准。当测序的前 25 个碱基低质量的数据有两个以上(即 PF<60%),测序仪器将判定这条读长质检不通过。而 PCR 产物样品 DNA 多样性差,在测序中极易引起两个碱基的识别可靠性低于 60%,从而减损最终测序的 Q30 值和判定合格测序数据总量,见图 6。图 4 为 Illumina Hiseq2000 测序中的第 5 测序通道(Lane 5)标准基因组样品测序监测的%Base 正常分布情况,Cluster PF(%)为 95.23±1.15,可获得的数据量约 34 G；图 7 为同时测序中第 4 通道(Lane 4) 的样品当全为 PCR 产物时%Base 分布情况，Cluster PF(%)为 21.24±14.83。使最终得到的合格数据量和 Q30 的统计值降低，测序通道 Lane 4 里全部测序数据质量≥Q30 仅为 52.9%，最终获得的测序数据量降低到 4.5 G。

作者尝试将待测的 PCR 产物样品 T2 中混合了等摩尔的 Phix 基因组标准品后进行 Miseq 测序，监测得到的 Cluster PF(%)为 89.81±0.72。最终目的样品 T2 占整个有效测序数据量的61.77%，大大改

善了 PCR 混合物测序样品的数据质量和有效数据总量。对于多样性差的其它待测序 DNA 样品，比如 RRBS 样品（Reduced Representation Bisulfite Sequencing，简化的表观亚硫酸氢盐测序），由于含有固定的酶切位点序列降低了样品的多样性，但通过混入一定量比例的多样性好的基因组标准品，实验结果显示得到的测序数据质量可提高约 50%。

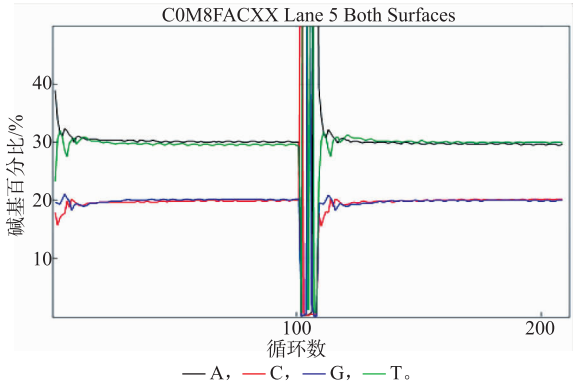


图 6　在 Illumina Hiseq2000 测序仪器中监测的第 5 条测序通道标准基因组样品的碱基分布图

Fig. 6　Base distribution of genomic DNA sequencing on lane 5 of Illumina Hiseq 2000

目前还可以尝试降低待测序样品上样量的方法改善低多样性样品引起的低数据质量的情况。但研究表明至少要降低一半的 PCR 样品量才可能达到略微的效果，也大大折损了测序数据总量，并非经济有效的方案。而选择将多样性好的基因组样品

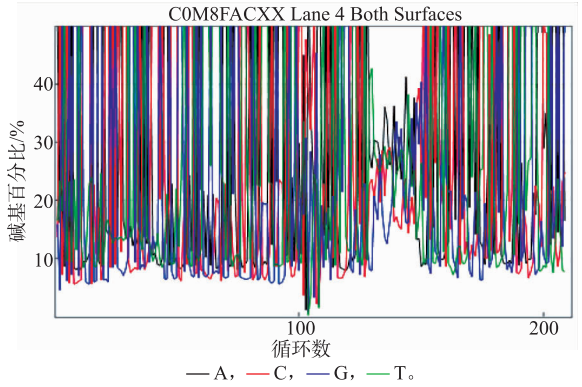


图 7　第 4 条测序通道为全部 PCR 产物样品测序时的碱基分布图

Fig. 7　Base distribution of all PCR samples sequencing on lane 4 of Illumina hiseq 2000

掺入到多样性差的样品中上机测序的方法，实验操作简单，能改善整个测序通道的样品多样性表现，从而提高测序质量，同时也能保证得到的最终有效的测序数据总量损失最少。

3 结　语

作者研究探索了一套对普通 PCR 产物样品的高通量测序建库和上机实验方法，对低多样性样品的高通量测序方法做了探索研究，通过对样品的上机测序数据监测，证实所建立的方法保证样品制备的高效经济和测序质量的稳定可靠。

参考文献：

[1] MARDIS E R. A decade's perspective on DNA sequencing technology[J]. **Nature**, 2011, 470(7333): 198-203.

[2] METZKER M L. Sequencing technologies-the next generation[J]. **Nat Rev Genet**, 2010, 11(1): 31-46.

[3] NIEDRINGHAUS T P, Milanova D, Kerby M B, et al. Landscape of next-generation sequencing technologies[J]. **Anal Chem**, 2011, 83(12): 4327-4341.

[4] 王兴春，杨致荣，王敏，等. 高通量测序技术及其应用[J]. 中国生物工程杂志，2012，32(1)：109-114.
WANG Xingchun, YANG Zhirong, WANG Min, et al. High-throughput sequencing technology and its application[J]. **China Biotechnology**, 2012, 32(1): 109-114.(in Chinese)

[5] MCCANN J C, WICKERSHAM T A, Loor J J, et al. High-throughput methods redefine the rumen microbiome and its relationship with nutrition and metabolism[J]. **Bioinformatics and Biology Insights**, 2014, 8: 109-125.

[6] MOROZOVA O, MARRA M A. Applications of next-generation sequencing technologies in functional genomics[J]. **Genomics**, 2008, 92(5): 255-264.

[7] MARDIS E R. Next-generation sequencing platforms[J]. **Annu Rev Anal Chem (Palo Alto Calif)**, 2013(6): 287-303.

[8] YU C, ZHANG Y, YAO S, et al. A PCR based protocol for detecting indel mutations induced by TALENs and CRISPR/Cas9 in Zebrafish[J]. **PLoS One**, 2014, 9(6): e98282.

[9] NELSON M C, MORRISON H G, BENJAMINO J, et al. Analysis, optimization and verification of Illumina-generated 16S rRNA

gene amplicon surveys[J]. **PLoS One**,2014,9(4):e94249.
[10] BOUTIN S,SEVELLEC M,Pavey S A,et al. A fast,highly sensitive double-nested PCR-based method to screen fish immunobiomes[J]. **Mol Ecol Resour**,2012,12(6):1027-1039.
[11] ZHOU Y,ZHU S,CAI C,et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells[J]. **Nature**,2014,509(7501):487-491.
[12] CAPORASO J G,LAUBER C L,WALTERS W A,et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms[J]. **ISME J**,2012,6(8):1621-1624.
[13] AMIR A,ZEISEL A,ZUK O,et al. High-resolution microbial community reconstruction by integrating short reads from multiple 16S rRNA regions[J]. **Nucleic Acids Res**,2013,41(22): e205.
[14] BOWMAN S K,SIMON M D,DEATON A M,et al. Multiplexed Illumina sequencing libraries from picogram quantities of DNA [J]. **BMC Genomics**,2013,14:466.

会　议　信　息

会议名称(中文):第 13 届德国肽研讨会

会议名称(英文):13th German Peptide Symposium
所属学科:生物物理学、生物化学及分子生物学,生物技术与生物工程
开始日期:2017-03-20　　结束日期:2017-03-23
所在国家:德国　　具体地点:纽伦堡
主办单位:德国化工与生物技术学会　　联系人:Petra Hellwig
联系电话:+49 (0)69 7564-167　　传真:+49 (0)69 7564-176
E-MAIL:hellwig@dechema.de　　会议网站:http://dechema.de/en/peptide13_2017.html
会议背景介绍:

We are pleased to announce the 13th German Peptide Symposium at the University of Erlangen-Nürnberg, 20-23 March 2017. The Symposium will cover all areas of peptide science, including peptide chemistry, biochemistry, biophysics, structural biology, immunology, pharmacology, as well as the use of peptides in medicine and material sciences.

You are cordially invited to participate in this event, and to contribute an oral or a poster presentation.

会议名称(中文):2017 年生物特征识别工程与应用国际会议 (ICBEA 2017)

会议名称(英文):2017 International Conference on Biometrics Engineering and Application (ICBEA 2017)
所属学科:生物技术与生物工程,生物信息学,进化与系统生物学,计算机软件,信号与信息处理
开始日期:2017-04-21　　结束日期:2017-04-23
所在国家:中华人民共和国　　所在城市:香港特别行政区　香港
主办单位:香港中文大学　　协办单位:ACM
承办单位:香港中文,香港理工　　会议主席:Prof. David Zhang
组织委员会主席:Prof. Gautam Sethi　　程序委员会主席:Prof. Yasushi Yagi
联系人:Ms. Mandy Yan　　联系电话:+852-3500-0137 (Hong Kong)　　E-MAIL:icbea@cbees.net
会议注册费:400 USD　　会议网站:http://www.icbea.org
会议背景介绍:The objective of the 2017 International Conference on Biometrics Engineering and Application (ICBEA 2017) is to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities on Biometrics Engineering and Application. 2017 International Conference on Biometrics Engineering and Application (ICBEA 2017) will be held in Hong Kong during April 21-23, 2017.