以海量数据计算揭示人类疾病发生机制及相关分子标志物

杨力, 魏刚, 唐鲲, Christine Nardini, 韩敬东*

中国科学院-马普学会计算生物学伙伴研究所, 计算生物学重点实验室, 上海 200031

* 联系人, E-mail: jdhan@picb.ac.cn

收稿日期: 2012-08-07; 接受日期: 2012-08-30

国家自然科学基金(批准号: 31271390, 31271354, 31171277, 31070748, 30890033, 31210103916, 91019019)资助项目

摘要 高通量芯片和深度测序技术为在全基因组水平上绘制高分辨率的基因组变异、RNA 转录、转录因子结合及组蛋白修饰图谱等研究提供了前所未有的机遇。这些技术彻底改变了以往有关转录组学、调控网络以及表观遗传调控的研究方法, 产生了海量的多水平组学数据, 并开启了高效数据整合研究的先河。然而, 如何有效地整合这些数据仍然是一个巨大的挑战。本文总结了高通量组学数据的产生对相关领域研究的主要影响及其与人类疾病的关系, 并介绍了多种用于数据整合分析的生物信息学方法。最后, 以炎症疾病为例进行说明。

关键词

组学
数据整合
生物信息学
系统生物学
分子标志物

十几年前, 人类基因组草图及其序列的公布^[1,2]向人们展示了基因组水平的发现将革命性地改变人类疾病研究及其临床治疗应用。随着不同个体基因组测序的完成, 基于计算分析的比较基因组学已成为在全基因组水平上研究基因功能的一个强大工具。同时研究表明, 个体之间的差异远远超过预期^[3]。鉴别遗传性疾病相关的基因组变异(包括单核苷酸多态性以及 DNA 片段的插入和删除事件)已被广泛应用于检测个体基因变异与特定性状表现之间的相关性^[4]。这些研究先采用微阵列芯片技术, 接着采用外显子测序技术, 而现在普遍流行的是全基因组测序技术^[5,6]。如今, 得益于测序成本的极大降低(1999 年与 2009 年的测序费用相差 14000 倍^[7])以及计算分析能力的飞速发展^[8], 多个位点的基因组变异已成功关联到不同的人类疾病, 包括多种类型的癌症^[6]。

尽管大多数研究专注于基因突变(包括常见和罕见的)与人类疾病的相关性, 但是其 DNA 变异的机制却不是很清楚, 而且仅依靠基因组信息来解释复杂疾病是远远不够的^[9]。因此, 更多基于表观基因组学、转录组学以及人体微生物组学的证据为人类疾病的相关研究带来了新的曙光。

1 表观基因组学与人类疾病

表观遗传学是一门研究在基因组序列不发生变化的情况下, 基因活性和表达水平发生可遗传变化的学科。表观遗传信息主要包括 DNA 甲基化、组蛋白修饰、非编码 RNA 和染色质高级结构, 这些信息在调节基因组功能, 如基因转录和基因组 DNA 复制等过程中发挥着重要作用^[10]。越来越多的证据表明,

引用格式: 杨力, 魏刚, 唐鲲, 等. 以海量数据计算揭示人类疾病发生机制及相关分子标志物. 中国科学: 生命科学, 2013, 43: 72-79, doi: 10.1360/052011-720
英文版见: Yang L, Wei G, Tang K, et al. Understanding human diseases with high throughput quantitative measurement and analysis of molecular signatures. Sci China Life Sci, 2013, 56, in press, doi: 10.1007/s11427-013-4445-9

表观遗传修饰对于正常的细胞发育极为重要,而异常的表观遗传调控则与许多类型的疾病有关^[11]。例如,表观遗传机制在癌症发展中扮演着重要的角色。一方面,在多种类型癌症中均发现存在大量 DNA 甲基化和组蛋白修饰的改变,并伴随着表观遗传调控相关基因上多位点的基因突变,这些结果显示表观遗传修饰的改变有助于癌症的发展^[12];另一方面,表观遗传机制为治疗干预提供了新思路,如使用组蛋白脱乙酰基酶抑制剂(HDACi)治疗 T-细胞淋巴瘤等癌症^[13]。除了癌症外,表观遗传失控还会导致许多免疫系统失调和神经紊乱,如系统性红斑狼疮(systemic lupus erythematosus, SLE)、类风湿性关节炎(rheumatoid arthritis, RA)、I 型糖尿病、Rett 综合征(rettsyndrome, RS)和阿尔茨海默病(Alzheimer disease, AD)^[14-18]。尽管对这些疾病的表观遗传机制已有很多认识,但仍需全面整体地了解表观遗传的过程从而安全有效地对疾病进行诊断和治疗。

了解表观遗传机制的第一步是在正常发育和疾病模型中鉴定与该表型相关的表观遗传学变化。随着新一代测序技术(next generation sequencing, NGS)的发展,可以以更低的成本获得全基因组的表观遗传学图谱。这些图谱展示的 DNA 甲基化、组蛋白修饰及其他染色质特征,可以大大提高人们对表观遗传学与发育和疾病之间关系的认识。2007 年,通过染色质免疫共沉淀及高通量测序技术(chromatin immunoprecipitation sequencing, ChIP-Seq),第一个人类 CD4 细胞表观基因组图谱绘制成功。这张图谱包括 20 种组蛋白修饰,组蛋白变体 H2A.Z, RNA 聚合酶 II 以及绝缘子结合蛋白 CTCF^[19]。2008 年,美国国立卫生研究院启动表观基因组学路线图计划(Roadmap Epigenomics Mapping Consortium),旨在建立用于生物医学研究的表观基因组学公共数据库。在此项目中,研究人员将在不同细胞类型中产生多种表观遗传修饰的数据。例如, DNA 甲基化可以通过亚硫酸氢盐测序法(bisulfite sequencing, BS-Seq)检测,即在测序前用亚硫酸氢盐处理基因组 DNA^[20]; ChIP-Seq 技术可以检测组蛋白修饰;而分析 DNA 酶 I 超敏感位点可以检测染色质亲和性^[21]。

虽然这些新颖的全基因组方法可以产生不同水平的表观基因组数据,但如何将这些数据和转录组学相整合,以揭示特定细胞或疾病类型相关的表观遗传标志,或是鉴定正常发育和病理过程中关键的

表观遗传调控因子,成为人们面临的新的挑战。此外,高通量测序技术产生的海量数据也需要发展新的计算工具进行分析。例如,研究人员已开发出贝叶斯网络模型来推断不同的组蛋白修饰及转录调控间的相互作用^[22]。此外,利用比较基因组学可从 DNA 和组蛋白修饰数据中进一步揭示基因组的调节功能^[23]。

表观基因组结合计算分析方法还揭示了表观遗传修饰在许多人类疾病(如衰老^[24-26])与 T 细胞成熟进程中具有关键的调控作用^[27,28]。

2 转录组学与人类疾病

转录组(所有 RNA 转录本)在不同个体、组织、细胞类型和发育阶段都有很大的差异。近年来,随着高分辨率微阵列芯片技术和新一代高通量测序技术在转录组学研究中的应用,使得在单核苷酸分辨率水平上的转录组解析成为可能。而进一步的关联研究发现,人类疾病发生过程中转录组 RNA 水平也发生了显著的变化。例如,全转录组学分析证实了正常人和自闭症患者间大脑的转录组表达差异,并表明转录和剪接失调导致的神经功能障碍是引起自闭的原因^[29]。而全面的人脑转录组分析不仅在转录组水平揭示了人脑神经发育的分子基础,更为深入进行相关神经性疾病发生的分子机制研究提供了新的思路^[30]。此外,对特殊富集方法获得的转录组分析研究也发现了来源于内含子区域的新型长非编码 RNA(lncRNA)^[31]。而一类双末端含有小核仁 RNA(snoRNA)的、来源于内含子区域的新型长非编码 RNA,特异定位于人 PWS 综合征(Prader-Willi syndrome, 小胖威利症)紧密关联的 15 号染色体区域,且高度表达^[32],提示这种长非编码 RNA 与人类疾病的相关性,其机制有待进一步研究阐明。虽然有待于进一步的临床研究证实,但已有初步证据表明微生物组在糖尿病、肥胖和肝功能中都发挥一定作用^[33]。当然,把各种组学数据集整合在一起分析,将为研究相关疾病机制提供更广阔的视野^[34]。最新研究展示了对一个个体病例长达 14 个月的追踪研究,收集了包括基因组、转录组、蛋白组、代谢组和自体抗体等多种水平的组学数据图谱。这些数据在组学水平显示了健康和病理条件下的人体动态变化,并揭示了两种状态下的 DNA 变异和转录后调控机制的存在^[35]。

3 数字定量表型组学与人类疾病

新型的测序技术扩大了基因组学的应用领域, 而表型数据采集方法同样面临着巨大的转变. 在后基因组时代, 大规模高通量的表型分析正变得越来越重要. 表型组是指某一生物, 包括细胞、组织、器官和机体在内的全部表型性状特征^[36]. 表型组学是利用综合基因组、转录组、蛋白组等多种组学系统研究某一生物或细胞在各种不同环境条件下的所有表型.

表型组学研究的核心是研究形态学和表面性状特征. 有机体的形态和外观来源于不同层面的一系列生物学反应, 这些协调反应为许多领域, 如进化、发育和人类疾病等研究提供了丰富的信息. 在模式生物的大规模遗传修饰研究中, 形态学变化是特定突变的主要的病理学证据^[37-39]. 人面部软组织是一个复杂的几何表面, 包括眼睛、鼻子、耳朵、嘴巴等许多重要的器官. 颅面结构的形态学发生是最精巧的发育过程, 它涉及神经回路闭合、面部中线成型、神经嵴的发生和迁移、面部原基和鳃弓外长、成型和分化等^[40]. 这些过程均受 SHH, FGF 和 BMP 等信号转导通路的调控^[41]. 鉴于其重要的生物学功能和结构的复杂性, 人脸可以多方面展示一个人的健康状况, 从遗传缺陷到常见疾病, 从衰老到心理健康. 罕见的遗传性疾病可以导致特定异常的面部特征, 如唐氏综合征(Down syndrome)^[42]、鲁宾斯坦-泰必氏综合征(Rubinstein-Taybi syndrome)^[43]、索托斯综合征(Sotos syndrome)^[44]和努南综合征(N Noonan syndrome)^[45]. 更为普遍的是, 常见的基因突变具有导致腭裂、兔唇等面部畸形的高风险性^[46-49]. 有趣的是, 引起唇裂的某些高风险的等位基因与患者父母和正常人的面部形态学变异相关^[50], 暗示了面部检测在疾病风险评估中的潜在用途. 事实上, 中医长期通过面诊来判断常规的疾病(面部识别系统用于中医的临床诊断). 除疾病研究外, 人面部特征还被广泛应用于其他医疗领域, 如法医学^[51]、心理学^[52,53]和老龄化^[54,55].

无论是从人脸部捕捉表型数据还是提取特定的生理信息, 都需要开发新技术和新算法. 快速人脸识别算法已在安全监控中应用很长时间^[56,57]. 最近, 有研究从人脸的视频数据中获取微弱的信号来测量心脏脉搏^[58]. 高分辨率的三维图像获取技术, 如 3dMDface[®]系统(www.3dmd.com), 可在表型组水平

上收集完整的面部形态和结构数据. 然而, 收集的数据仅有一小部分应用于研究, 通常是采用一系列标记点和/或是它们之间的距离和角度等信息^[59-61]. 这些不足很大程度上是由于病理表型通常是整个面部细微的曲率差别和色素沉着变化的总体表型, 因此难以简单地将面部特征和特定疾病相关联. 最近, 研究人员利用三维密度表面定位模型(3D dense surface registration, 3D-DSR)来分析三维面部图像, 将成千上万的三维像素点比对到来自不同性别、年龄、民族以及疾病或对照组的面部样本上^[62-64]. 接着用类似于分析基因组数据的方法定量分析这些面部表型组数据. 并在此基础上加入精确的 3D 面部标记点, 使得 3D-DSR 能达到完全自动化, 且具有高通量和更好的稳定性^[65]. 3D-DSR 方法已被应用于在不同民族、性别和疾病人群中计算平均面部模型^[66-68], 以及模拟随年龄衰老的面部变化情况^[69]. 最重要的是, 许多 3D-DSR 技术可以提取许多遗传性疾病的面部标志特征, 显示出其在疾病自动判别中的潜在应用价值^[70]. 类似的方法在自闭症和癫痫等复杂疾病中也进行了相关测试^[71,72]. 但是, 由于大多数研究来源于欧洲人群, 且不同种群的脸部形状和遗传背景极其不同, 其结果不能直接用于其他种群. 最新研究发现, 在中国汉族人群中, 一个非综合性唇裂(nonsyndromic cleft lip, NSCL)关联位点同样可预测正常人面部形态学变化(数据尚未发表). 长远来看, 从面部提取的细微特征不仅可以诊断罕见的遗传性疾病, 还可以监测个体的健康状况, 如常见疾病、皮肤衰老和压力等. 鉴于面部成像技术的非侵入性, 此技术将大大推进个体化医疗的发展.

4 高通量异质数据的整合分析

不论是在分子水平上的芯片和深度测序, 还是在表型水平上的数字化定量图像分析, 均产生了大量的高通量数据, 这就迫切需要开发新的方法将其有效地整合. 从简单的关联网络到复杂的贝叶斯网络, 都已被应用于这些数据的整合分析^[73], 研究人员也不断开发出针对不同类型、不同层次的数据整合的算法. 例如, 信息流优化的算法已被用于从蛋白质-蛋白质相互作用网络中寻找连接遗传学筛选获得的靶标基因及同一生物过程中表达变化的基因之间的信号通路^[74], 寻找连接参与高原反应的 microRNA 和

mRNA 的信号通路^[75], 或通过全基因组关联分析 (genome-wide association study, GWAS) 找出与病理相关的差异表达基因所在的信号通路^[76]. 这些创新性的数据集成方法, 不仅可以鉴定调控表型的关键基因和通路, 还可以为解析复杂表型和疾病发生等提供新的研究思路(图 1).

5 案例分析: 在免疫性疾病研究中的应用

如前所述, 高通量测序技术在生物学研究领域的创新应用给转化医学研究带来了革命性的变化, 一个重要影响体现在自身免疫性疾病的研究中, 这种病在人群中的发病率惊人. 据 AARDA(American Autoimmune Related Diseases Association)统计, 自身

免疫性疾病的发病率是癌症的 2.5 倍, 略超过心脏病的发病率. 事实上, 测序技术的发展使得这些组学研究不仅对患者有利, 也有利于社会经济和生产力的发展. 因为这些疾病通常在患者年富力强时发生, 且对于这些疾病的治疗和预防会占用很多公共资源.

类风湿性关节炎常被用来作为疾病模型进行研究. RA 是一种致残性疾病, 病因不明, 攻击关节的滑膜组织, 造成关节功能障碍或失去活动能力. 全世界有 1%, 中国约 4 百万人患有此病, 且女性发病率较高(约 70%)^[77].

经 GWAS 证实, RA 是一种复杂的遗传性疾病. 最近基于欧洲人的元分析^[78]发现了 7 个新的易感基因, 其中一个同样存在于亚洲人群^[79]. 这些结果更好地描绘了 RA 在欧洲和亚洲人群中的分布, 同时也表

图 1

明 RA 是一种极其复杂的疾病。

目前, 仍面临着一系列挑战, 包括如何更好地定义/细化和解释用于分析现有高通量测序方法所产生的数据(DNA 测序数据、RNA 测序数据以及元基因组数据等)的算法, 以及在系统层面上开发整合这些数据的方法。事实上, 当这些方法被应用于类风湿性关节炎研究时, 不同组学数据的整合使得人们能够更深入地了解 AIDS-CIDs 与元基因组变异之间的关系。同时, 可以研究这些变化与其他分子之间的相互作用, 判断现有的治疗方法对这些变化的影响以及考量是否可以改进现有方法以平衡常规治疗的副作用。

研究分析元基因组数据时发现^[80], 相对于单组学数据分析的结果, 多组学数据整合分析可挖掘更多的信息。应用多元统计方法(如因子分析)对来自不同组学的数据进行整合分析, 如 mRNA 和 miRNA 数据, 已鉴定出一些采用单组学数据分析时无法发现的特征^[81,82]。

其他方法依赖于网络化呈现数据的可行性, 这

比一系列的分子能提供更多的信息。最近研究报道, 构建了一个包含所有可用的转录和转录后修饰等组学信息的类风湿性关节炎图谱(该文发表时还主要依靠微阵列方法检测基因表达数据)^[83]。随后, 应用拓扑分析方法将该图谱与已知的信号通路信息整合并得到了关联图。有意思的是, 基于这种方法可以更好地解释 R406 在 RA 治疗过程中的作用以及提供新的处方禁忌, 尤其是治疗由糖尿病引起的类风湿性关节炎患者。

6 结语

综上所述, 不同层面的组学研究发现了疾病进程中不同的分子标志物和调控机制。结合高通量测序技术和计算方法的各自优势, 可绘制出一幅全面的基于序列的分子标志图谱, 从而进一步用系统生物学方法来理解人类疾病的发生机制, 为个性化诊断和个体化药物的临床应用提供理论基础。

参考文献

- 1 Lander E S, Linton L M, Bruce B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860–921
- 2 Venter J C, Adams M D, Myers E W, et al. The sequence of the human genome. *Science*, 2001, 291: 1304–1351
- 3 Pennisi E. Breakthrough of the year. Human genetic variation. *Science*, 2007, 318: 1842–1843
- 4 Nielsen R. Genomics: in search of rare human variants. *Nature*, 2010, 467: 1050–1051
- 5 Nik-Zainal S, Alexandrov L B, Wedge D C, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*, 2012, 149: 979–993
- 6 Pasaniuc B, Rohland N, McLaren P J, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet*, 2012, 44: 631–635
- 7 Collins F. Has the revolution arrived? *Nature*, 2010, 464: 674–675
- 8 Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet*, 2007, 3: e114
- 9 Ashley E A, Butte A J, Wheeler M T, et al. Clinical assessment incorporating a personal genome. *Lancet*, 2010, 375: 1525–1535
- 10 Bird A. Perceptions of epigenetics. *Nature*, 2007, 447: 396–398
- 11 Greer E L, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*, 2012, 13: 343–357
- 12 Chi P, Allis C D, Wang G G. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer*, 2010, 10: 457–469
- 13 Mann B S, Johnson J R, Cohen M H, et al. FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist*, 2007, 12: 1247–1252
- 14 Sekigawa I, Kawasaki M, Ogasawara H, et al. DNA methylation: its contribution to systemic lupus erythematosus. *Clin Exp Med*, 2006, 6: 99–106
- 15 Maciejewska-Rodrigues H, Karouzakis E, Strietholt S, et al. Epigenetics and rheumatoid arthritis: the role of SENP1 in the regulation of MMP-1 expression. *J Autoimmun*, 2010, 35: 15–22
- 16 Miao F, Smith D D, Zhang L, et al. Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes. *Diabetes*, 2008, 57: 3189–3198
- 17 Amir R E, Van den Veyver I B, Wan M, et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, 1999, 23: 185–188

- 18 Tsankova N, Renthal W, Kumar A, et al. Epigenetic regulation in psychiatric disorders. *Nat Rev Neurosci*, 2007, 8: 355–367
- 19 Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, 129: 823–837
- 20 Eckhardt F, Lewin J, Cortese R, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*, 2006, 38: 1378–1385
- 21 Hesselberth J R, Chen X, Zhang Z, et al. Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat Methods*, 2009, 6: 283–289
- 22 Yu H, Zhu S, Zhou B, et al. Inferring causal relationships among different histone modifications and gene expression. *Genome Res*, 2008, 18: 1314–1324
- 23 Xiao S, Xie D, Cao X, et al. Comparative epigenomic annotation of regulatory DNA. *Cell*, 2012, 149: 1381–1392
- 24 Jin C, Li J, Green C D, et al. Histone demethylase UTX-1 regulates *C. elegans* life span by targeting the insulin/IGF-1 signaling pathway. *Cell Metab*, 2011, 14: 161–172
- 25 Xue H, Xian B, Dong D, et al. A modular network model of aging. *Mol Syst Biol*, 2007, 3: 147
- 26 Zhou B, Yang L, Li S, et al. Midlife gene expressions identify modulators of aging through dietary interventions. *Proc Natl Acad Sci USA*, 2012, 109: E1201–E1209
- 27 Wei G, Wei L, Zhu J, et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity*, 2009, 30: 155–167
- 28 Wei G, Abraham B J, Yaqi R, et al. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity*, 2011, 35: 299–311
- 29 Voineagu I, Wang X, Johnston P, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 2011, 474: 380–384
- 30 Kang H J, Kawasawa Y I, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature*, 2011, 478: 483–489
- 31 Yang L, Duff M O, Graveley B R, et al. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol*, 2011, 12: R16
- 32 Yin Q F, Yang L, Zhang Y, et al. Long noncoding RNAs with snoRNA ends. *Mol Cell*, 2012, 48: 219–230
- 33 Hvistendahl M. My microbiome and me. *Science*, 2012, 336: 1248–1250
- 34 Snyder M, Du J, Gerstein M. Personal genome sequencing: current approaches and challenges. *Genes Dev*, 2010, 24: 423–431
- 35 Chen R, Mias G I, Li-Pook-Than J, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, 2012, 148: 1293–1307
- 36 Davis B D. The isolation of biochemically deficient mutants of bacteria by means of penicillin. *Proc Natl Acad Sci USA*, 1949, 35: 1–10
- 37 Bedell M A, Jenkins N A, Copeland N G. Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. *Genes Dev*, 1997, 11: 1–10
- 38 Bedell M A, Largaespada D A, Jenkins N A, et al. Mouse models of human disease. Part II: recent progress and future directions. *Genes Dev*, 1997, 11: 11–43
- 39 Hardouin S N, Nagy A. Mouse models for human disease. *Clin Genet*, 2000, 57: 237–244
- 40 Francis-West P H, Robson L, Evans D J. Craniofacial development: the tissue and molecular interactions that control development of the head. *Adv Anat Embryol Cell Biol*, 2003, 169: III–VI, 1–138
- 41 Wilkie A O, Morriss-Kay G M. Genetics of craniofacial development and malformation. *Nat Rev Genet*, 2001, 2: 458–468
- 42 Allanson J E, O'Hara P, Farkas L G, et al. Anthropometric craniofacial pattern profiles in Down syndrome. *Am J Med Genet*, 1993, 47: 748–752
- 43 Allanson J E, Hennekam R C. Rubinstein-Taybi syndrome: objective evaluation of craniofacial structure. *Am J Med Genet*, 1997, 71: 414–419
- 44 Allanson J E, Cole T R. Sotos syndrome: evolution of facial phenotype subjective and objective assessment. *Am J Med Genet*, 1996, 65: 13–20
- 45 Allanson J E, Hall J G, Hughes H E, et al. Noonan syndrome: the changing phenotype. *Am J Med Genet*, 1985, 21: 507–514
- 46 Beaty T H, Murray J C, Marazita M L, et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet*, 2010, 42: 525–529
- 47 Birnbaum S, Ludwig K U, Reutter H, et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet*, 2009, 41: 473–477
- 48 Mangold E, Ludwig K U, Birnbaum S, et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet*, 2010, 42: 24–26
- 49 Rahimov F, Marazita M L, Visel A, et al. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet*, 2008, 40: 1341–1347
- 50 Boehringer S, van der Lijn F, Liu F, et al. Genetic determination of human facial morphology: links between cleft-lips and normal variation.

- Eur J Hum Genet, 2011, 19: 1192–1197
- 51 Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet*, 2011, 12: 179–192
- 52 Little A C, Jones B C, DeBruine L M. Facial attractiveness: evolutionary based research. *Philos Trans R Soc Lond B Biol Sci*, 2011, 366: 1638–1659
- 53 Meyer-Marcotty P, Alpers G W, Gerdes A B, et al. Impact of facial asymmetry in visual perception: a 3-dimensional data analysis. *Am J Orthod Dentofacial Orthop*, 2010, 137: 168 e1–e8; discussion 168–169
- 54 Albert A M, Ricanek K Jr, Patterson E. A review of the literature on the aging adult skull and face: implications for forensic science research and applications. *Forensic Sci Int*, 2007, 172: 1–9
- 55 Fu Y, Guo G, Huang T S. Age synthesis and estimation via faces: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2010, 32: 1955–1976
- 56 Bowyer K W, Chang K, Flynn P. A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition. *Comput Vis Image Und*, 2006, 101: 1–15
- 57 Burton A M, Wilson S, Cowan M, et al. Face recognition in poor-quality video: evidence from security surveillance. *Psychol Sci*, 1999, 10: 243–248
- 58 Poh M Z, McDuff D J, Picard R W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt Express*, 2010, 18: 10762–10774
- 59 Boehringer S, van der Lijin F, Liu F, et al. Genetic determination of human facial morphology: links between cleft-lips and normal variation. *Eur J Hum Genet*, 2011, 19: 1192–1197
- 60 Farkas L G, Katic M J, Forrest C R. International anthropometric study of facial morphology in various ethnic groups/races. *J Craniofac Surg*, 2005, 16: 615
- 61 Weinberg S M, Naidoo S D, Bardi K M, et al. Face shape of unaffected parents with cleft affected offspring: combining three—dimensional surface imaging and geometric morphometrics. *Orthod Craniofac Res*, 2009, 12: 271–281
- 62 Hammond P, Hutton T J, Allanson J E, et al. 3D analysis of facial morphology. *Am J Med Genet A*, 2004, 126: 339–348
- 63 Maal T J, van Loon B, Plooi J M, et al. Registration of 3-dimensional facial photographs for clinical use. *J Oral Maxillofac Surg*, 2010, 68: 2391–2401
- 64 Wan J, Shen L, Fang S, et al. A framework for 3D analysis of facial morphology in fetal alcohol syndrome. *MIAR'10 Proceedings of the 5th international conference on Medical imaging and augmented reality Pages*, 2010, 118–127
- 65 Guo J, Mei X, Tang K. Automatic landmark annotation and dense correspondence registration for 3D human face images, *arXiv:1212.4920*
- 66 Bhuiyan Z A, Klein M, Hammond P, et al. Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. *J Med Genet*, 2006, 43: 568–575
- 67 Hammond P, Hannes F, Suttie M, et al. Fine-grained facial phenotype-genotype analysis in Wolf-Hirschhorn syndrome. *Eur J Hum Genet*, 2012, 20: 33–40
- 68 Hammond P, Hutton T J, Allanson J, et al. 3D dense surface models identify the most discriminating facial features in dysmorphic syndromes. *54th Annual Meeting of the American Society for Human Genetics, Toronto, Canada*, 2004, 1
- 69 Hutton T J, Buxton B F, Hammond P, et al. Estimating average growth trajectories in shape-space using kernel smoothing. *Medical Imaging, IEEE Transactions on*, 2003, 22: 747–753
- 70 Hammond P, Hutton T J, Allanson J E, et al. Discriminating power of localized three-dimensional facial morphology. *Am J Hum Genet*, 2005, 77: 999–1010
- 71 Hammond P, Forster-Gibson C, Chudley A E, et al. Face-brain asymmetry in autism spectrum disorders. *Mol psychiatry*, 2008, 13: 614–623
- 72 Kasperavičiūtė D, Catarino C B, Chinthapalli K, et al. Uncovering genomic causes of co-morbidity in epilepsy: gene-driven phenotypic characterization of rare microdeletions. *PLoS ONE*, 2011, 6: e23182
- 73 Han J D. Understanding biological functions through molecular networks. *Cell Res*, 2008, 18: 224–237
- 74 Yeger-Lotem E, Riva L, Su L J, et al. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat Genet*, 2009, 41: 316–323
- 75 Chen F, Zhang W, Liang Y, et al. Transcriptome and network changes in climbers at extreme altitudes. *PLoS One*, 2012, 7: e31645
- 76 Huang J, Liu Y, Zhang W, et al. eResponseNet: a package prioritizing candidate disease genes through cellular pathways. *Bioinformatics*, 2011, 27: 2319–2320
- 77 Scher J U, Abramson S B. The microbiome and rheumatoid arthritis. *Nat Rev Rheumatol*, 2011, 7: 569–578
- 78 Stahl E A, Raychaudhuri S, Remmers E F, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*, 2010, 42: 508–514

- 79 Kochi Y, Okada Y, Suzuki A, et al. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat Genet*, 2010, 42: 515–519
- 80 Liu Y, Zhang C, Zhao L, et al. Adapting functional genomic tools to metagenomic analyses: investigating the role of gut bacteria in relation to obesity. *Brief Funct Genomics*, 2010, 9: 355–361
- 81 Fronza R, Tramonti M, Atchley W R, et al. Joint analysis of transcriptional and post-transcriptional brain tumor data: searching for emergent properties of cellular systems. *BMC Bioinformatics*, 2011, 12: 86
- 82 Fronza R, Tramonti M, Atchley W R, et al. Brain cancer prognosis: independent validation of a clinical bioinformatics approach. *J Clin Bioinforma*, 2012, 2: 2
- 83 Wu G, Zhu L, Dent J E, et al. A comprehensive molecular interaction map for rheumatoid arthritis. *PLoS One*, 2010, 5: e10137

杨力, 博士, 中国科学院-马普学会计算生物学伙伴研究所研究员、课题组长, 2012 年入选中国科学院“百人计划”。主要从事 RNA 的表达调控和功能研究, 近年来利用高通量测序和计算生物学相结合的方法在全转录组水平对 RNA 的表达调控进行研究。在全转录组水平揭示了包括 RNA 剪接, RNA 编辑和非编码 RNA 等多个水平的调控作用。近年来在 *Nature*, *Molecular Cell*, *Genome Research*, *Genome Biology* 等国际著名期刊共发表 SCI 文章 20 余篇(www.picb.ac.cn/rnomics)。

韩敬东, 博士, 2004 年入选中国科学院“百人计划”, 担任中国科学院遗传与发育生物学研究所分子系统生物学中心主任和课题组研究员。2005 年获得“国家杰出青年科学基金”资助。2009 年至今, 任中国科学院上海生命科学研究院计算生物学研究所所长。主要研究方向: 生物网络的结构、动态及其功能; 通过表观组学、功能基因组学及整合网络分析, 寻找衰老及干细胞分化相关的关键基因及进一步研究分子及功能作用机制; 探求遗传基因网络中内在的安全稳定机制。