Research progresses of imaging genomics in Alzheimer disease

LYU Yuting^{1,2}, YAO Xufeng^{1,2,3*}, BU Xixi^{1,2}, XU Song^{1,2}, HUANG Gang^{2,3}

(1. School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200082, China; 2. College of Medical Imaging, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China; 3. Shanghai Key Laboratory of Molecular Imaging, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China)

[Abstract] Alzheimer disease (AD) is a progressive neurodegenerative disease, with phenotype changes closely related to the influence of genotype variation. Imaging genomics can vastly and non-invasively reveal the physiopathological process of AD by associating multimodality phenotypes with genotypes. Current status and progresses of imaging genomics in AD were reviewed in this paper focused on imaging phenotypes, genomics characteristics and phenotype-genotype association.

[Keywords] Alzheimer disease; imaging genomics; phenotype; genotype

DOI:10.13929/j.issn.1003-3289.2020.08.029

阿尔茨海默病影像基因组学研究进展

吕玉婷^{1,2}, 姚旭峰^{1,2,3*}, 卜溪溪^{1,2}, 徐松^{1,2}, 黄钢^{2,3}

(1. 上海理工大学医疗器械与食品学院, 上海 200082; 2. 上海健康医学院医学影像学院, 上海 201318; 3. 上海市分子影像学重点实验室, 上海健康医学院, 上海 201318)

[摘要] 阿尔茨海默病(AD)是一种进行性神经系统变性疾病,其表型变化与遗传变异因素密切相关。影像基因组学通过将多模态影像表型与基因数据关联,以高通量和非侵入式等特点揭示 AD 的生理病理过程。本文从 AD 相关影像学表型、基因组学特征及影像-基因关联各方面对影像基因组学 AD 研究现状及进展进行综述。

[关键词] 阿尔茨海默病; 影像基因组学; 表型; 基因型

[中图分类号] R745.1; R445 **[文献标识码]** A **[文章编号]** 1003-3289(2020)08-1243-04

阿尔茨海默病(Alzheimer disease, AD)是一种病因未明的进行性神经系统退行性疾病,也是最常见的痴呆类型,占全部痴呆的 50%~70%,病因尚未明确,且无可有效治愈药物或改变其病理进程的手段^[1]。为诊断及预测 AD,有学者^[2]基于影像组学方法,将影像学特征与临床信息等相结合,并建立模型进行个性化评估;分子生物学研究^[3]也证实多个易感基因与 AD

密切相关。本文从 AD 相关影像学表型、基因组学特征及影像-基因关联各方面对影像基因组学 AD 研究现状及进展进行综述。

1 影像基因组学简介

影像基因组学对影像学表型与基因组数据进行关联分析,剖析其间的关联性,以影像学表型作为遗传变异和诊断之间的中间表型来探究 AD 的发病机制,兼

[基金项目] 国家自然科学基金重点项目(81830052)、国家自然科学基金面上项目(61971275)、分子影像学重点实验室建设项目(18DZ2260400)。

[第一作者] 吕玉婷(1996—),女,河南洛阳人,在读硕士。研究方向:医学图像处理与分析。E-mail: 1807200675@qq.com

[通信作者] 姚旭峰,上海理工大学医疗器械与食品学院,200082;上海健康医学院医学影像学院,201318;上海市分子影像学重点实验室,上海健康医学院,201318。E-mail: yao6636329@hotmail.com

[收稿日期] 2019-11-20 **[修回日期]** 2020-05-12

具可早期预测疾病及多模态、高通量、非侵入式获取病变表型特征的双重优势。将多模态脑影像学表型与基因数据结合分析,通过影像学表型检测遗传变异或评估已确定的遗传风险对大脑的影响,比单纯基因研究更为有效。

2 AD 影像基因组学研究现状

2.1 相关影像学表型的研究现状 目前 AD 常用影像学检查方法有 MRI 和 PET。MRI 主要包括结构 MRI (structural MRI, sMRI)、功能 MRI (functional MRI, fMRI) 和弥散张量成像 (diffusion tensor imaging, DTI)。sMRI 能直接反映脑区关键部位的形态学改变,清楚显示大脑灰白质体积萎缩程度; fMRI 特指血氧水平依赖 (blood oxygenation level dependent, BOLD)-fMRI,通过检验脑细胞在血液流动时的磁场变化完成脑部功能成像;DTI 通过检测水分子扩散运动的各向异性来显示脑白质纤维束的走向和完整性,是目前唯一可精确显示活体脑白质纤维束的成像方法。PET 则依据标记物分为 ¹⁸F 标记的氟代脱氧葡萄糖 (¹⁸F-fluorodeoxyglucose, ¹⁸F-FDG)PET 和淀粉样 PET,分别用于评估大脑糖代谢和淀粉样蛋白变化,显示早期 AD 患者脑部代谢。见表 1。

国内外影像基因组学 AD 研究所选取的影像学表型特征大多来自 MRI。研究^[4]表明,海马体积萎缩是鉴别 AD 临床前阶段的关键影像学表型。海马区及内嗅皮层等结构体积明显变小的遗传性强,故海马体积和皮层厚度是较为适用的 AD 影像学特征参数, LORENZI 等^[5]、CACCIAGLIA 等^[6]及 LI 等^[7]对影像学表型与基因进行关联分析时,所选取大脑皮层厚度、灰质体积和右侧海马体积参数均来自 sMRI。由于常规脑影像学特征难以表征 AD 全部特点,根据既往研究结果,可结合高维影像组学特征对 AD 影像学特征与易感基因进行关联分析,如 SCELSI 等^[8]对 MRI 所示双侧海马体积和 PET 所见皮层蛋白样沉积负荷进行其与基因的关联研究,ZHANG 等^[9]自 sMRI 和 fMRI 中提取海马体积和部分脑功能特性进

行相关性分析。随着脑网络研究的进展,有学者^[10]开始利用 DTI 构建大脑结构网络,并对网络特征进行脑影像学表型相关研究。

2.2 基因组学方法研究现状 AD 基因研究方法现已从经典的基于连接和候选基因的关联研究发展到全基因组关联研究 (genome-wide association study, GWAS) 和下一代测序 (next generation sequencing, NGS)。鉴定新易感基因有助于深入理解 AD 发病机制。GWAS 是目前最常用的 AD 影像基因组学基因分析方法,以单核苷酸多态性 (single nucleotide polymorphism, SNP) 为标记单位,在全基因组水平进行比对分析,挖掘与疾病表型存在强相关性基因的变异位点,可鉴定与 AD 表型或病理变化具有强相关的特定 SNP 组合,于基因层面反映 AD 的病理特征。DERRK 等^[11]以 GWAS 分析来自 50 个群组的 30 717 个个体的皮层体积和颅内体积,确定了影响核壳和尾状核体积的 5 个新的遗传变异,并证实 3 个新位点对海马体积和颅内体积存在影响。NHO 等^[12]通过整合遗传学信息与影像学表型,利用全基因组测序对 PESN1 基因与负载定量生物标记物进行关联分析,证明 PESN1 变异与关键脑区萎缩显著相关。此外,基因分析方法还包括主成分分析 (principal component analysis, PCA)、多因子降维 (multifactor dimensionality reduction, MDR) 和独立成分相关算法 (independent component correlation algorithm, ICA)^[13],但均未考虑之前的研究结果,且需对大量数据集进行多重校正,造成很多微小风险位点被遗漏。基因本体 (gene ontology, GO)、京都基因和基因组数据库 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 富集分析可解决上述问题,故常用于分析具有各异性表达基因的通路。ZENG 等^[14]分析 AD 相关基因通路富集分析和蛋白质-蛋白质相互作用 (protein-protein interaction, PPI),确定了 6 个潜在影响微小核糖核酸 (micro RNA, miRNA) 结合的先导 SNP (rs10119、rs1048699、rs148763909、rs610932、rs6857 和 rs714948) 和 2 个代理 SNP (rs12539172 和

表 1 不同成像方式下影像学表型提取及相关变化

成像方式	影像学表型提取或研究方法	AD 患者相关表型变化
sMRI	大脑皮层灰质体积、厚度、表面积和密度、白质高信号、海马体积、脑脊液、脑室等	灰质、白质、海马体积缩小 脑脊液和脑室所占颅内体积百分比增大
fMRI	脑功能区血流量改变和各个脑区之间功能连接等	与认知功能相关的关键脑区血流量降低
DTI	基于体素分析、纤维束追踪、基于纤维束追踪的空间统计分析等	脑白质微小萎缩,纤维束走向改变
PET	大脑相关区域糖代谢、血流灌注信息、横断面和纵向淀粉样蛋白负荷等	双侧颞叶和顶叶等代谢降低、淀粉样蛋白负荷随时间异常增加

rs2847655),表明 SNP 通过影响 miRNA 的结合而作用于 AD。朱士胜等^[15]利用从数据库中得到的差异表达基因(differentially expressed genes, DEG)进行 GO 富集分析和 KEGG 通路富集分析,筛选出在 AD 发展过程中起重要作用的关键基因。考虑到此两种方法可能存在富集假阳性和一定主观性,实际分析时可加入加权基因共表达网络构建(weighted gene co-expression network analysis, WGCNA)^[16],以鉴定共表达的基因模块,分析基因模块与样本特征间的相关性,利于挖掘潜在标记基因,因其可信度高于 GO 和 KEGG 富集分析而应用广泛。

2.3 影像-基因关联分析方法研究进展 关联分析方法是当前用影像基因组学 AD 研究重点。早期单变量研究需要对候选基因变异和候选脑影像学特征进行科学假设,且必须同时检验大量假设,并对结果加以多重校正,过程十分繁复,故应用较少。STEIN 等^[17]首先以多变量关联分析观察受试者整个大脑 31 622 个体素中的 448 293 个 SNP,并以集群节点来执行计算;HIBAR 等^[18]也在基因水平进行类似分析;二者均应用 PCA 将多维遗传数据归一化,得到特征 SNP,最后回归到单变量线性分析。WANG 等^[19]提出群稀疏多任务回归模型,对多模态的影像学表型与候选基因进行关联分析,不仅毋须将多变量影像数据归一化,免于存在大量独立性假设的问题,还具有更强的抗噪声能力。ZHU 等^[20]开发的贝叶斯简化秩模型利用贝叶斯框架中的函数来最大化影像学表型与基因数据的关联,既能降低单变量回归中独立性假设的数量,又具有更高的检测遗传变异能力。GE 等^[21]在此基础上提出一种基于核机的多维变量集交互效应检测方法,可同时检测多个遗传变异之间的相互影响。STINGO 等^[22]利用分层混合模型对整个大脑功能连接网络和

全基因进行关联分析;HUANG 等^[23]开发的快速全基因组关联分析(fast genome-wide association study, FGWAS)框架在 GWAS 基础上结合大脑功能特性,显著提高了对重要基因变异的检测能力。目前大量多元数据关联研究均对已有统计学分析模型进行扩展与改进,以简化结果的多重验证和解释。有学者^[24]建立新的快速多任务稀疏正则相关分析,在相关系数和正则权值模式方面均具有更好的性能,且运行速度快,易于实现。茹峰等^[25]提出一种新的基于典型相关分析(canonical correlation analysis, CCA)的高阶统计量结构化稀疏算法,以探索大量基因与脑影像学数据之间的关联性,便于提取两类数据中高度相关的特征,并保持其相互独立。

近来 AD 研究多转向纵向影像基因组学,以观察大脑结构动态变化与基因之间的相关性,但总体而言仍处于起步阶段(表 2)。SZEFER 等^[26]分析 75 845 个 SNP 与 56 个大脑区域结构 MRI 变化率的关系,以引导程序增强的稀疏正则相关分析验证了 1 649 个 SNP 对促进 AD 结构缺失具有较大可能性,并首次用逆概率加权解释了 ADNI 研究中存在的偏置采样设计和低关联信号的高维数据。LU 等^[27]将 ZHU 等^[20]的贝叶斯低秩模型扩展为纵向设置,将时变影像学数据直接与遗传标记联系起来进行剖析。WANG 等^[28]将机器学习方法扩展到纵向影像基因组学研究中,通过输入 3 123 个 SNP 位点,基于体素分析法(voxel-based morphometry, VBM)和 Freesurfer 两种成像标记物自动检测其间的关系,发现 APOE-rs429358 与大脑皮层体积、MS4A4E-rs670139 与脑白质体积、MS4A6A-rs66798 与灰质密度之间存在相关性,有望为未来将人工智能融入纵向影像基因组学研究提供参考依据。

表 2 不同影像-基因关联分析方法的优点

文献序号	关联分析方法	模型优点
[18]	主成分回归、线性回归	对多维数据利用主成分分析得到特征 SNP,并进行单变量线性回归
[19]	群稀疏多任务回归分析	可解决需要大量独立性假设的问题
[20]	贝叶斯简化秩模型	利用贝叶斯框架进行关联分析,可降低单变量回归中独立性假设的数量
[21]	最小二乘法核机的多轨迹分析	结合多个遗传变异的影响,并建模分析其相互作用
[22]	分层混合分析	研究大脑功能与遗传变异的相关性
[23]	功能全基因组关联分析	可显著提高对影响大脑功能的遗传变异的检测能力
[24]	快速多任务稀疏正则相关分析	在相关系数和正则权值模式方面性能较好,且运行速度快,易于实现
[25]	基于 CCA 的高阶统计量结构化稀疏算法	能快速提取出两类数据中高度相关的特征
[26]	引导程序增强的稀疏正则相关分析	研究大脑结构纵向变化与遗传变异的相关性
[27]	纵向扩展的贝叶斯低秩分析	直接剖析时变影像学数据与遗传变异的关联
[28]	时间结构自动学习模型	能自动发现基因型与表型间的相互关系

3 小结与展望

影像基因组学目前已大量用于 AD 研究,通过对多维脑部特征和特异性风险基因进行多方位有效信息的提取和关联分析而实现 AD 诊断、分型、预测及预防,但用于影像学表型及基因分析方面仍存在限制。未来将聚焦于扩充影像学表型和选取易感基因,改进关联分析模型,提高分析的准确性;还可考虑在影像基因组学研究应用深度学习相关算法,通过融合影像学表型与基因而进一步探究其相关性。

[参考文献]

- [1] GAUGLLER J, JAMES B, JOHNSON T, et al. 2019 Alzheimer's disease facts and figures[J]. *Alzheimers & Dementia*, 2019, 15(3):321-387.
- [2] 贺江琳,王远军.阿尔茨海默病影像组学关键方法研究进展[J]. *中国医学影像技术*, 2019, 35(10):1569-1573.
- [3] FARRER L A. Expanding the genomic roadmap of Alzheimer's disease[J]. *Lancet Neurol*, 2015, 14(8):783-785.
- [4] 孙梦莎,顾鸣敏.阿尔茨海默病早期诊断研究进展[J]. *中国现代神经疾病杂志*, 2018, 18(3):213-221.
- [5] LORENZI M, GUTMAN B, HIBAR D P, et al. Partial least squares modelling for imaging-genetics in Alzheimer's disease: Plausibility and generalization [C]. *Proc of 13th International Symposium on Biomedical Imaging, Prague: IEEE*, 2016: 838-841.
- [6] CACCIAGLIA R, JOSE L M, CARLES F, et al. Effects of APOE- ϵ 4 allele load on brain morphology in a cohort of middle-aged healthy individuals with enriched genetic risk for Alzheimer's disease[J]. *Alzheimers & Dementia*, 2018, 14(7):902-912.
- [7] LI J Q, WANG H F, ZHU X C, et al. GWAS-linked loci and neuroimaging measures in Alzheimer's disease [J]. *Molecular Neurobiol*, 2016, 54(1):1-8.
- [8] SCELISI M A, KHAN R R, LORENZI M, et al. Genetic study of multimodal imaging Alzheimer's disease progression score implicates novel loci[J]. *Brain*, 2018, 141(7):2167-2180.
- [9] ZHANG X L, YU J T, LI J, et al. Bridging integrator 1 (BIN1) genotype effects on working memory, hippocampal volume and functional connectivity in young healthy individuals [J]. *Neuropsychopharmacology*, 2015, 40(7):1794-1803.
- [10] ELSHEIKH S S, CHIMUSA E R, MULDER N J, et al. Genome-wide association study of brain connectivity changes for Alzheimer's disease[J]. *Sci Rep*, 2020, 10(1):1-16.
- [11] HIBAR D P, STEIN J L, RENTERIA M E, et al. Common genetic variants influence human subcortical brain structures[J]. *Nature*, 2015, 520(7546):224-229.
- [12] NHO K, HORGUSLUOGLU E, KIM S, et al. Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease[J]. *BMC Med Genomics*, 2016, 9(1):12-18.
- [13] 郝小可,李蝉秀,严景文,等.基于统计学习的影像遗传学方法综述[J]. *自动化学报*, 2018, 44(1):13-24.
- [14] HAN Z P, HUANG H, GAO Y, et al. Functional annotation of Alzheimer's disease associated loci revealed by GWASs[J]. *PLoS One*, 2017, 12(6):e0179677.
- [15] 朱士胜,张丽,代亚磊,等.阿尔茨海默病差异表达基因的生物信息学分析[J]. *重庆医科大学学报*, 2019, 44(4):434-440.
- [16] 刘伟,李立,叶桦,等.权重基因共表达网络分析在生物医学中的应用[J]. *生物工程学报*, 2017, 33(11):1791-1801.
- [17] STEIN J L, HUA X, LEE S, et al. Voxelwise genome-wide association study (vGWAS) [J]. *Neuroimage*, 2010, 53(3):1160-1174.
- [18] HIBAR D P, STEIN J L, KOHANNIM O, et al. Voxelwise gene-wide association study (vGeneWAS): Multivariate gene-based association testing in 731 elderly subjects [J]. *Neuroimage*, 2011, 56(4):1875-1891.
- [19] WANG H, NIE F, HUANG H, et al. Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning [J]. *Bioinformatics*, 2012, 28(12):i127-i136.
- [20] ZHU H T, KHONDKER Z, Lu Z H, et al. Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers[J]. *J Am Stat Assoc*, 2014, 109(507):997-990.
- [21] GE T, NICHOS T E, GHOSH D, et al. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application [J]. *Neuroimage*, 2015, 109:505-514.
- [22] STINGO F C, GUINDANI M, VANNUCCI M, et al. An integrative bayesian modeling approach to imaging genetics[J]. *J Am Stat Assoc*, 2013, 108(503):876-891.
- [23] HUANG C, THOMSPON P, WANG Y L, et al. FGWAS: Functional genome wide association analysis [J]. *Neuroimage*, 2017, 159(1):107-121.
- [24] DU L, LIU K F, YAO X H, et al. Fast multi-task SCCA learning with feature selection for multi-modal brain imaging genetics [C]. *Proc of 2018 IEEE International Conference on Bioinformatics and Biomedicine, Madrid:IEEE*, 2018:356-361.
- [25] 茹锋,徐锦,常琪,等.一种用于影像遗传学关联分析的高阶统计量结构化稀疏算法[J]. *计算机科学*, 2019, 46(4):66-72.
- [26] SZEFER E, LU D H, NATHOO F, et al. Multivariate association between single-nucleotide polymorphisms in Alzgene linkage regions and structural changes in the brain: Discovery, refinement and validation[J]. *Stat Appl Genet Mol Biol*, 2017, 16(5-6):367-386.
- [27] LU Z H, KHONDKER Z, IBRAHIM J G, et al. Bayesian longitudinal low-rank regression models for imaging genetic data from longitudinal studies[J]. *Neuroimage*, 2017, 149:305-322.
- [28] WANG X, YAN J, YAO X, et al. Longitudinal genotype-phenotype association study through temporal structure auto-learning predictive model [J]. *J Computat Biol*, 2018, 25(7):809-824.