

“生物与信息融合 (BT 与 IT 融合)” 重点专项 2021 年度项目申报指南

为落实“十四五”期间国家科技创新有关部署安排，国家重点研发计划启动实施“生物与信息融合 (BT 与 IT 融合)”重点专项。根据本重点专项实施方案的部署，现发布 2021 年度项目申报指南。

本重点专项总体目标是：聚焦未来生命科学、医药健康产业和经济社会发展等重大需求，通过加强生物技术与信息技术跨界融合研究，兼顾科学创新和技术图谱，引领新经济模式发展。突破信息大数据、生物大数据的获取、管理、分析、挖掘、调控和知识发现等底层支撑技术，提升数据整合与转化利用能力；构建 DNA 存储、类脑智能与人机交互、生物知识图谱、可编程细胞智能、智慧医疗等交叉融合技术，推进大数据驱动的生命科学知识发现及转化应用；催生一批面向生命健康的颠覆性新技术，形成一批新工具、新技术、新标准与新产品，解决医疗大数据、医疗人工智能原创性理论基础薄弱、重大产品和系统缺失等难点问题。

2021 年度指南部署坚持“优化学科布局和研发布局，推进学

科交叉融合,完善共性基础技术供给体系”的要求,围绕基于 DNA 原理的信息存储系统开发、面向生命—非生命融合的智能生物系统构建与开发、BT 与 IT 融合技术的健康医学场景应用示范等 3 个任务,按照基础前沿技术、共性关键技术、示范应用,拟启动 16 个方向,拟安排国拨经费概算 6.7 亿元。其中,围绕 DNA 信息存储等技术方向,拟部署 5 个青年科学家项目,每个项目 500 万元,拟安排国拨经费概算 0.25 亿元。

项目统一按指南二级标题(如 1.1)的研究方向申报。每个项目拟支持数为 1~2 项,实施周期不超过 3 年。申报项目的研究内容必须涵盖二级标题下指南所列的全部研究内容和考核指标。基础研究类项目下设课题数不超过 4 个,项目参与单位总数不超过 6 家,共性关键技术类和示范应用类项目下设课题数不超过 5 个,项目参与单位总数不超过 10 家。项目设 1 名负责人,每个课题设 1 名负责人。

青年科学家项目(项目名称后有标注)不再下设课题,项目参与单位总数不超过 3 家。项目设 1 名项目负责人,青年科学家项目负责人年龄要求,男性应为 1983 年 1 月 1 日以后出生,女性应为 1981 年 1 月 1 日以后出生。原则上团队其他参与人员年龄要求同上。常规项目下设青年科学家课题的,青年科学家课题负责人及参与人员年龄要求,与青年科学家项目一致。

指南中“拟支持数为 1~2 项”是指：在同一研究方向下，当出现申报项目评审结果前两位评价相近、技术路线明显不同的情况时，可同时支持这 2 个项目。2 个项目将采取分两个阶段支持的方式。第一阶段完成后将对 2 个项目执行情况进行评估，根据评估结果确定后续支持方式。

本专项研究涉及人类遗传资源采集、保藏、利用、对外提供等，须遵照《中华人民共和国人类遗传资源管理条例》相关规定执行。涉及人体研究需按照规定通过伦理审查并签署知情同意书。涉及实验动物和动物实验，要遵守国家实验动物管理的法律、法规、技术标准及有关规定，使用合格实验动物，在合格设施内进行动物实验，保证实验过程合法，实验结果真实、有效，并通过实验动物福利和伦理审查。

本专项 2021 年度项目申报指南如下。

1. 基于 DNA 原理的信息存储系统开发

1.1 DNA 分子信息存储的高加密性编码与信息安全体系研究 (青年科学家项目)

研究内容：针对大数据时代海量数据高效存储需求，开发基于 DNA 存储技术特征的高加密性存储编码方法，建立基于数据可靠有效恢复的 DNA 存储特有的新型信息安全体系。开发 DNA 存储个性化编码，实现信源信道联合编码；研究 DNA 存储错误

特征，开发 DNA 存储专用纠错编码体系；研究多类型 DNA 存储介质选择及长效远距离传输等的数据安全相关理论与方法；挖掘 DNA 存储的生化特性，开发依靠数学方法的 DNA 存储的数据加密关键技术、防恶意复制的生物方法关键技术、防篡改的类区块链方法关键技术。

考核指标：开发高效可靠 DNA 存储编解码方法，可支持低冗余条件下（合成冗余率 $\leq 20\%$ ）数据有效恢复（恢复率 $\geq 90\%$ ）所需平均拷贝的分子数达到 10。开发高容错编解码和纠错方法，较现有国际先进水平容错率提升一个数量级，达到 1%读写容错率；开发不少于 3 种 DNA 存储专用介质及适配存储条件下的数据加密技术体系，实现 $K \geq 128$ 的密钥强度（ 2^K ）；开发不少于 2 种防复制、防篡改方法，实现单碱基水平的改动可识别精度。形成 3 项以上行业或团体等技术标准。

1.2 基于多类型生物分子的新一代超高密度信息存储技术研发（青年科学家项目、基础研究类）

研究内容：研究基于多类型生物分子的新型 DNA 信息存储技术，实现对现有 DNA 分子编码效率极限的突破；研究包括但不限于镜像核酸（L-DNA）、D 型非天然核酸或带有修饰基团的天然核酸等的信息写入/读取成套技术；研究新型 DNA 存储的编解码算法与系统设计新方法；研究非天然碱基在 DNA 存储应用

中的错误特性并开展相应纠错方法研究；开发制备适合镜像核酸和 D 型非天然核酸的扩增酶，实现信息的无损复制与扩增；开发非天然核酸的高效生物、化学合成方法及适配高通量测序技术。

考核指标：在信息与生物分子的编解码理论和分子信息寻址、加密等方面形成理论基础，开发不少于 2 种基于新型存储介质分子的存储技术体系，实现实际净信息密度 ≥ 3.8 bits/nt 的原理验证，数据保真度 $\geq 99.9\%$ 的突破；开发不少于 3 种适配新型存储介质分子的转码算法与纠错方法；建立生产级（kg 级）低成本适配碱基单体的制备工艺与技术流程；研制不少于 2 种适配核酸扩增酶，实现 > 200 nt 的高效扩增；完成适配高通量测序技术原理验证，验证通量大于 10M 碱基。

1.3 基于大规模可寻址可控催化原理的 DNA 合成新技术研发（青年科学家项目、基础研究类）

研究内容：针对高通量 DNA 合成的需求，设计独立寻址、可控的微电极阵列，开发亚微米尺度的 DNA 可控合成单元，通过集成电路设计理念实现超大规模反应矩阵；研究光化学电极、酶化学电极或热电极的制备，通过光触、电触、热触酶化学或化学反应控制 DNA 实时合成过程；结合快速微流体等控制技术，实现局部可控的间歇合成方式，将碱基或者短链碱基片段逐个加入到 DNA 链中，实现高通量的“数据写入”过程。

考核指标：建立至少 1 种基于大规模寻址改变物理量控制生物酶活性或化学催化活性的创新合成理论模型；搭建可实现寻址控制功能的配套芯片；实现兼容 CMOS 或 MEMS 工艺和可控合成反应的电极或光极材料、加工工艺和合成方法；DNA 可控合成阵列可达 1000 万单元每平方厘米；开发可用光、电或热控制活性的合成酶；实现单元间的合成的交叉干扰几率 < 3%；实现可控合成长度 > 40nt。申请核心发明专利 3~5 件。

1.4 基于晶体管场效应的单分子测序关键技术研发（青年科学家项目）

研究内容：针对 DNA 数据存储所需的高速“读出”的需求，依靠半导体先进制程的发展缩小晶体管等微纳器件的物理尺寸，赋能高灵敏单分子传感和基因测序的新机遇，开发基于可规模化制造的晶体管或晶体管纳米孔的新一代超高通量单分子测序技术。研究固态晶体管或晶体管纳米孔用于液相单分子基因测序的器件原理和工作机制，开发预测单分子测序性能数值模型；攻克 CMOS 工艺兼容的测序芯片的单元架构、关键制造工艺、测试系统等核心技术难题，实现原型器件；开发适配高速、并行化测试数据的智能算法和分析技术。

考核指标：建立基于晶体管或晶体管纳米孔为核心传感器件的单分子测序的完整工程化理论模型和工作机制；实现预测物理

器件在微纳流体中的性能数值仿真技术，开发一套单分子传感和测序的仿真模拟器；实现检测单元的核心器件，最优可达单分子灵敏度，单个单元的最快数据采样频率高于 1MHz；验证 DNA 碱基识别原理，单种碱基的识别准确性不低于 90%；单分子 DNA 的最长读长优于 10kb；开发核心器件的 CMOS 和生物应用兼容的可规模化的制造工艺，可实现阵列化独立检测（如 4×4 ）；申请核心发明专利 2~3 件。

2. 面向生命—非生命融合的智能生物系统构建与开发

2.1 蛋白质结构折叠的精准预测与设计应用（青年科学家项目、基础研究类）

研究内容：利用蛋白质氨基酸序列、结构性质、主侧链、相互作用、功能等多种大数据信息，运用生物物理及人工智能原理开发有效的深度学习解析算法，重点解决非同源蛋白质、未知功能新蛋白、膜蛋白等的结构精准预测问题。研究蛋白质序列、结构与功能的内在关系，建立蛋白结构精准预测技术体系，并取得大规模测试。利用典型的生物表达体系，结合高分辨率结构生物学技术，与合成生物学、分子生物学方法等测试蛋白质结构，揭示蛋白功能。

考核指标：建立非同源蛋白、未知功能新蛋白与膜蛋白的精准预测方法 2~3 个；参与最新国际蛋白质结构预测评估实验，预

测方法整体排名进入前 3，并在最新评估实验测试数据集上的平均准确度 GDT_TS 分值达到 90；建立发展 5~10 种核心深度学习精准演算法，申报 3~5 件新方法核心发明专利；设计 1~2 个具有新结构的蛋白质，完成生物功能应用验证。

2.2 原子尺度生物系统基础元件的人工智能理性设计技术 (青年科学家项目、基础研究类)

研究内容：围绕原子尺度精度上设计和制备的复杂生物系统的基础模块解析、理性设计、组合优化与制备所面临的瓶颈性、基础性问题，发现并结合自然界中基于进化的针对复杂蛋白元件以及原子精度的类酶人工生命体等生物系统底层逻辑的理性设计的研究和认知，完善基于代谢通路的发现、组合和迁移的数据库、知识库体系，发现大数据基础上的通路以及基础元件设计原则、公理体系并实现设计工具的智能迭代和优化；研究建立高效、灵活、理性的基于进化选择、大数据和知识融合、面向人工合成生物系统智能构建的高性能计算分析技术体系，建立有机和有机—无机复合生命元件的创造、设计和优化的理论基础，研究大数据和知识指导下的基本元件系统自我进化的模型与算法，开发基于第一性原理的新型统计模型和基于小样本稀疏监督下的“设计—构建—测试—学习”闭环系统，实现自动化的分子生成与化合物逆向合成，研究建立涵盖智能设计、智能控制和智能风险评估三

个方面理论和技术的人工生物系统构建的基础性支撑平台。

考核指标：开发基于第一性原理的 3~5 种新型统计关联的精准预测模型和人工智能学习模型，为提高定向进化速度、提高虚拟筛选精度提供理论依据和算法基础；针对离子通道等大尺寸、复杂蛋白元件，以及原子精度的类酶人工生命体，明确各自体系底层的关键特征，开发基于小样本的、稀疏监督下的“设计—构建—测试—学习”闭环系统，提高虚拟筛选的精度，指导原子尺度精度上设计和制备合成生物系统；基于生命功能模块构成的公理体系和设计原则，设计、合成和构建 10~30 种具有基础性和框架性的有机、无机体系生命元件，建立基本原则发现和理论实现、大数据、智能分析、高效理性设计与高性能计算一体化的全链条生物系统合成的技术支撑体系，形成验证示范。

2.3 自动化细胞设计流水线开发及应用（青年科学家项目、基础研究类）

研究内容：利用自动化软件设计、深度学习算法和高通量测试平台，全面提高生物设计过程的速度与精度，开辟以智能工程细胞为基础的高效率、低污染的材料和医药产业。在基础研究方面，研究细胞调控线路计算设计原理、细胞调控线路的稳定性和鲁棒性、细胞调控线路的跨物种适配原则、合成多细胞系统的动态自组织理论，形成可定量预测的人工细胞调控线路的设计算法

理论体系。在关键技术方面，开发高时空精确度、微型高通量的细胞测量与操作平台，开发用于大规模细胞调控线路装配的自动化工程细胞改造平台，并开发实验设计、数据收集和管理机器学习工具，形成全整合的软件平台。在产业辐射方面，开发数字模拟系统，将细胞计算功能与产业应用需求对接。

考核指标：以功能性细胞和细胞群体为设计目标，在人工细胞调控线路的定量预测理论上产生重要突破。在工程菌、病毒载体、HEK293T 和 CHO 哺乳细胞系等 3~5 种模式生物系统中，建立 4~6 种功能性单细胞或多细胞中人工调控线路的可定量预测模型。建立 1 个标准化表征的细胞设计元件库；建立 1~2 套基于模型和基于深度学习的智能设计算法，以及 1~2 套定量测试的标准体系。建成 1 套自动化的大片段细胞调控线路 DNA 组装与转化平台，实现功能密集型工程细胞的高效改造与测试。利用以上理论和工具，对定制化合成生物细胞系实现“设计—装配—测试”的快速迭代，服务于科研、绿色生产、新型智能药物研发等领域。

2.4 高精度非侵入神经电生理编解码计算芯片体系

研究内容：针对现有国产非侵入式神经电生理信号采集计算系统研发使用设备需大量依赖进口，及现有神经电生理信号计算、处理系统性能不足、功耗大、集成度低等重大现实问题，构建完全自主国产高精度非侵入神经电生理编解码计算芯片体系。研究低

功耗仪表放大器设计技术、高精度模数转换器技术，开发高精度神经电生理信号采集放大模拟芯片，获取极微弱信号；研究信号处理加速模块，实现芯片级信号时\频\空\非线性等特征设计解析，支持 SVM\LDA\DNN 等分类模型加速器，形成非侵入神经电生理信号编解码计算芯片；研究芯片级精识别、快通讯、高指令等范式，形成高效基础算法库和接口标准；开发信号采集、计算系统原型。

考核指标：非侵入神经生理信号采集模拟芯片的任务指标 ADC 分辨率 ≥ 24 bit，ADC 采样率 ≥ 16 KHz，ADC 信噪比 ≥ 120 dB，ADC 动态范围 ≥ 125 dB，单通道功耗 ≤ 5 mW，共模抑制比 ≥ 120 dB，输入噪声 $\leq 0.8 \mu$ Vpp，输入阻抗：1G Ω ；基于自主模拟芯片的非侵入神经生理信号采集系统，无线传输速率 ≥ 216 KB/s，重量（不含电池） ≤ 120 g；智能编解码计算芯片要求采用 55nm 以下工艺；架构安全高性能 RISC-V 处理器，主频不低于 300 MHz，支持实时操作系统；内嵌 FPGA 模块，查找表数量不低于 60KLUTs，典型工作频率不低于 200 MHz，BRAM 不低于 3 Mbits，DSP 不低于 150 个，支持千兆以太网接口、ADC DAC 接口、音视频接口；支持微伏以下精度微弱生理解码、信息传输速率不低于 350bit/min、超百指令集的交互范式。在国产自主高精度非侵入神经生理信号采集芯片和编解码计算芯片方面取得突

破，申请 10 项以上核心发明专利；在 5 个以上的场景上实现示范应用。

有关说明：由企业牵头，其他经费（包括地方财政经费、单位出资及社会渠道资金等）与中央财政经费比例不低于 3:1。

2.5 新一代高相容性生物植入电极设计与应用（青年科学家项目）

研究内容：针对各类全侵入型、半侵入型等生物植入电极传感材料与生物组织界面的生物相容性和检测灵敏度问题，通过原子工程设计开发人工酶改性的生物植入电极。通过人工智能技术研究人工酶结构与活性的关系，设计开发新的原子精度人工酶，对植入电极进行表面改性，提高植入电极与生物组织接触界面的生物相容性，提高生物植入电极的灵敏度与精确度。开展改性后植入电极多通道信号采集与分析研究，促进其临床应用。

考核指标：应用原子工程开发人工酶构建 10 种以上的新型生物植入电极。通过人工智能算法指导团簇实验，在原子掺杂、合成方法以及处理方法组成的实验参数空间中进行优化设计，探索 10 万种合成参数空间，至少找出 50 种可将电极性能提升 10~100 倍以上的人工酶。结合量子力学方法预测 1 万种以上可能的人工酶结构，进一步获得 20 种以上的原子水平人工酶。开发 10 种以上团簇级的人工酶改性的生物植入电极，人工酶均匀度 >

90%，表面粒径 2~3nm，抗氧化反应速率不低于 20 μ M/s，类氧化还原酶活性的最大反应速率不低于 1mM/min，催化的转换频率（TOF）值不低于 320 /min。通过人工酶降低炎症反应提高电极材料的生物相容性，疤痕组织在电极材料表面生长的抑制率>80%，实现电极催化稳定性和植入电极的检测时间>1 年。神经电生理信号检测在全频带灵敏度提升 10 倍以上，实现 1 μ V 以下的微弱信号检测，电荷存储量达到 50mC。开展人工酶生物植入电极在神经系统疾病如癫痫中的临床应用，实现 100 例以上病人的电生理信号安全可靠采集 1 年以上，建立发病预测、监测模型 2 种以上，为临床诊断和治疗效果检测提供新的技术方案，预测准确度>99%。

2.6 组织工程类脑智能复合体设计与开发(青年科学家项目)

研究内容：建立诱导性多能干细胞增殖分化体系，2D 实现干细胞的定向分化和神经元突起的精确定位，3D 构建具有神经生物学功能的脑类器官；研发模拟脑组织弹性模量的柔性电极，与人工脑组织构建神经—电子复合体，并开发适用于生物类脑信息分析的神经形态器件与芯片；建立体外训练调控体系，实现神经—电子复合体的双向交互，在信息和实体层面构建可植入的神经—电子智能复合体。将这种复合体植入体内，在结构上形成植入物靶向神经环路的定向、定点、可控地整合，在功能上实现本体神

经—植入神经—神经电极信号传导回路的建立。研究其长期植入的可能性、生物相容性以及神经信号空间与时间分辨能力。通过对植入物的构建、植入方式的优化、植入后的控制和数据分析算法等一系列关键技术的突破，最终实现生物胞体介导的新一代生命体—非生命体混合智能体的融合和交互。

考核指标：构建血管化脑类器官，可通过血液给脑类器官供氧；脑类器官体外培养时间不少于 90 天，直径不小于 1 cm。研发不少于 3 种柔性电极；柔性电极在拉伸至 100% 时，电学性质衰减至 95% 以内；开发基于神经形态器件的芯片可用于生物类脑信号处理；建立 2~3 种人工脑学习训练的刺激信号；人工脑可识别的图像和声音信号不少于 5 种；在 3 种任务模态下，实现人工脑对外部电子系统的控制；完成 2D 和 3D 可植入的神经—电子复合体，通道数不少于 60 个。完成中枢神经系统植入测试，体内有效观察时间不少于 30 天，有效通道数目保持在 60% 以上。在自体神经—植入复合智能体之间实现至少 3 种不同模态的信息双向交互传输。

3. BT 与 IT 融合技术的健康医学场景应用示范

3.1 数字化细胞参照系研究、建设与示范应用（青年科学家项目）

研究内容：利用结构复杂、关系多维、蕴含丰富功能规律的

多组学大数据,发展各类数据深度解析与关联分析算法,研究人体各系统、器官、组织及各主要发育阶段和重要疾病状态下各类细胞的内在规律、相互关系、稳态失衡和命运决定,发展能适应细胞多重结构解析,空间、时间和多种内在功能关系表征的信息表示体系,建立包含细胞分类的基因组学、表观基因组学、转录组学和重要细胞表型信息的人体细胞全息图谱,建立具有更完备信息框架、包含更全面信息的人类细胞体系全信息基础参照系,实现对人体复杂系统全方位的细胞分子特性观测,发现人体几十万亿个细胞具有复杂多样的多重分子特征、稳态维持和动态规律,并开展医学示范应用。

考核指标:获得 5~10 个人体重要组织器官的包含细胞分类的基因组学、表观基因组学、转录组学和重要细胞表型信息的人体细胞全息图谱和 1 个人类细胞体系全信息基础参照系;建立 3~5 个细胞多重分子特征、稳态维持与命运决定、发育演化及其动态规律解析的计算分析系统和 3~5 个核心深度学习算法;构建 5~10 个涵盖人类细胞分类和功能研究的基础性数据库体系。

3.2 数字化人体表型建模与呈现技术(青年科学家项目)

研究内容:研究分子、细胞、功能和影像等多尺度人体表型信息的可视化技术;针对多尺度人体表型信息,通过信息采集、存储、处理、共享、分析等多个环节,制定功能互联、格式规范、

操作便捷的人体表型信息数据库，并与指南 3.4 中的大型队列间联合研究大数据云平台联动共享；通过对表型数据的系统挖掘，建立表型间的跨尺度关联模型；开展体现分子、细胞、功能、影像等表型变量间关系的动态建模，并建立相应的人体表型数字化呈现系统；在健康医疗领域开展相关系统的验证，辅助临床教学与诊疗等实际应用。

考核指标：同时采集分子、细胞、功能和影像等 4 类表型数据，建立不少于 1 万人的多尺度人体表型队列；建立涵盖以上 4 类表型，数据规模不少于 1PB 的多尺度人体表型数据库；提出表型数据库建设规范与评价规范，形成 6 项以上标准规范，申请或获得不少于 5 项核心发明专利；构建 1 套可面向临床的人体表型数字化呈现系统，可体现分子、细胞、功能、影像等表型之间的动态关系；在 3 家以上综合医院进行该人体表型数字化呈现系统的验证与实际应用。

3.3 融合形态特征和组学信息的智慧病理辅助诊断技术体系

研究内容：以常见恶性肿瘤和重大传染病等重大疾病为主要研究对象，应用 BT 与 IT 融合的最新技术产品，建立或基于已有的多尺度生物样本资源平台，形成包含包括电子病历等规范化、结构化临床诊疗信息、影像特征、病理形态特征和多组织学特征的医疗大数据；基于新型人工智能技术，深入挖掘影像特征、病

理改变、组学信息和临床信息间的准确关联；研发病理与影像、组学、临床信息的交叉和融合技术；建立以病理变化、影像特征和组学信息为基础的人工智能辅助智慧病理整合诊断技术体系。

考核指标：成功构建具有自主知识产权的、融合影像学特征、病理变化和组学特征的智慧病理辅助诊断体系，与目前诊断指南中诊断指标的契合度大于 95%，并能提供个体化诊疗措施建议，提示新的疾病相关组学特征信息；在至少 5 家三甲医院完成对以上诊断技术体系的评估。

3.4 大型队列间联合研究大数据云平台支撑系统研发

研究内容：依托现有、正在建设大型人群队列，应用最新 BT 与 IT 融合新技术，支持不同队列间科研、临床等数据的快速精准获取、质控、共享、数据融合、联合研究等科研活动。针对人群队列数据的专业性、安全性和可共享性，研究不同人群队列之间的通用数据治理、使用及共享的术语和定义、体系框架等基础标准；研究促进队列数据共享的技术和机制、跨队列安全分析技术、共享过程去中心化监管及可溯源技术；研发不同队列数据之间从质控、采集、传输、存储、处理、交换、分享、发布、销毁等环节的共性信息技术；研制并部署涵盖全流程的不同队列间数据交汇分享管理平台。

预期指标：根据跨队列研究需求，形成 20 项以上国家、行

业或团体标准；构建一套队列间合作研究及数据共享的标准和机制；在研究内容涉及的各个方面研发出先进的新算法、新计算模型和改进方法，形成不少于 50 个具有跨队列研究意义的工作流；在不同队列数据之间的全流程共性信息技术中取得突破，申请 10 项以上核心技术发明专利，申请 30 项以上相应的软件著作权。建立具备支持超过百万规模的跨队列研究平台，该平台需要支持同时 5 个以上万级大型人群队列的联合研究，服务 20 个以上跨人群队列研究相关重大科研项目。

3.5 生物数据深度挖掘与知识融合的智能系统研发与示范应用

研究内容：面向高性能计算环境，构建融合大规模生命组学数据分析系统与工作流的类脑计算环境；研究基于边缘计算、联邦计算和分布式计算的知识访问控制模型，研究具备隐私保护能力的知识存储证明方法，研究实现基于前沿 IT 技术的新一代的知识共享和价值体现模式；研究建立群智协同下的基于规则、统计、深度学习的多源知识图谱构建和标准融合方法、前沿的知识图谱联邦更新机制和多尺度生物知识图谱构建及可视化交互方法；以前沿知识图谱推理和知识融合技术为基础，研究建立精准医学知识图谱推理技术体系，将语义网、图嵌入等前沿知识图谱推理技术应用到智能决策系统当中，面向健康医学场景应用，开发临床

应用的智能诊疗和临床决策支持系统。

考核指标：研发 20~30 项、可部署在云端、具有支撑数据密集型科研的“类脑”计算模式的组学大数据深度解析 workflow；建立 3~5 种基于边缘计算、联邦计算和分布式计算的知识访问控制模型；研发 5~10 项基于规则、统计、深度学习、语义网、图嵌入、平移模型、路径排名算法的多源知识图谱融合推理算法和应用体系，研发 2~3 个可对标 Transformer、Bert 等现有通用技术的自主知识产权的知识图谱网络构建平台，形成相应的自然语言处理技术体系；建立可支持精准医疗临床决策支持的、重要性能指标达到国际同类产品相同水平的集成引擎产品 1 个；研发可部署在大型医院的基于生物学多组学大数据的临床决策支持系统不少于 1 套；建立 5~10 种具备隐私保护能力的知识存储证明方法，实现基于前沿 IT 技术的新一代的知识共享和价值体现模式。

3.6 基于跨尺度多模态生物学大数据的肿瘤智能诊疗共性关键技术研究

研究内容：针对恶性肿瘤生物学大数据的高维度、跨尺度和多模态特点，研究生物学大数据的存储方案、共享体系、融合计算、隐私保护和安全监管技术，发展面向生命组学、影像组学、临床表型组学等的大数据汇聚分析的药物靶点和疾病新型标

标志物发现的“干”“湿”结合的新方法和疾病风险预测新技术；研究整合医学典籍、电子病历等多来源多层次医学知识的智能诊疗知识融合方法，研发构建智能诊疗算法库、通用软件框架及在线分析共享平台。研发以临床需求为导向的面向高维度、跨尺度和多模态临床大数据的集成分析引擎和软件工具，建立面向临床的科研大数据支持系统和操作环境；建立基于生物医学多组学大数据的面向疾病风险预测、疾病诊断、精准治疗、药靶和疾病生物标志物发现、预后监控的智能化诊疗支持系统；推广并验证系统的临床适用性。

考核指标：建立 5~10 种“干”“湿”结合、通用的基于跨尺度、多模态、高维度生物医学大数据融合的肿瘤精准诊疗生物标志物、药物作用靶点发现和个体患病风险预测评估的新算法和新方法；建立 5~10 种基于通用数据标准、诊疗规范的临床大数据与海量医疗知识融合利用的新技术新方法；建立综合病例数 20000 例以上，多组学数据 2000 例以上的肿瘤研究标准化可共享数据集；建立 5~10 种推动临床基础研究向“数据密集型科研新范式”转化的临床大数据新型计算分析技术；开发可整合医院临床真实世界基础大数据，涵盖临床数据质量评估、数据清理和编审、智能搜索、统计关联分析、深度挖掘、知识融合、大数据分析 workflow 等技术的临床的科研大数据支持 HRP 系统和智能操作

环境 1 套。