

证券代码：000710

证券简称：贝瑞基因

成都市贝瑞和康基因技术股份有限公司
投资者关系活动记录表

编号：2025-001

投资者关系活动类别	<input type="checkbox"/> 特定对象调研 <input type="checkbox"/> 分析师会议 <input type="checkbox"/> 媒体采访 <input type="checkbox"/> 业绩说明会 <input type="checkbox"/> 新闻发布会 <input type="checkbox"/> 路演活动 <input type="checkbox"/> 现场参观 <input checked="" type="checkbox"/> 其他（电话交流会）
参与单位名称	国盛证券、信达澳亚基金、国寿安保基金、彤源投资、天治基金、太平基金、深圳市鹏城基石投资管理有限公司、玖鹏资产、青骊投资、北信瑞丰、亚太财产保险有限公司、国华兴益资管、德睿恒丰、中银基金、同泰基金、正圆投研、光大保德信、青榕资产、汇添富基金、平安养老、德邦基金、华安基金、易方达、兴全基金、民生加银基金、平安资产、华宝基金、建信养老、建信基金、东吴基金、东北证券、国泰基金、鹏华基金管理有限公司、摩根基金、贝莱德、碧云银霞、申万菱信、敦和资产、金科控股、景领投资、银石投资、鹏扬基金、南土资产、长江资管、峰岚资产、中欧基金、富国基金管理有限公司、上海冰河资产管理有限公司、中信建投、健顺投资、光大保德信基金、先锋基金、中信证券、华夏基金、达诚基金、海通证券资管、誉辉资本、嘉实基金、新活力资本、永赢基金、凯石基金、资衡（上海）商业管理有限公司、深圳金泊投资管理有限公司、兴业证券、长江养老、富安达基金、泓澄投资、兴证全球基金、兴业基金、Mighty Divine、财通资管、途灵资管、中银基金、中金资管、红土创新、华泰柏瑞基金、鸿商投研、博道基金、鑫元基金、中华联合保险、中

	海基金管理有限公司、工银瑞信、理成资产、万家基金、野村资管、前海联合基金管理有限公司、华创资管、金鹰基金、上海榕湖投资管理有限公司
时间	2025 年 2 月 11 日下午 14:00
会议地点	线上电话会议
上市公司接待人员姓名	董事长-高扬、智能数据部负责人-李岳珍、董事会秘书-许菲
投资者关系活动主要内容介绍	<p>董事长高扬博士介绍整体情况：</p> <p>贝瑞基因长期从事基因测序行业，聚焦生育健康、遗传病检测、科技服务等领域，业务中产生的每个样本诊断数据都是海量数据，从几百 M 到几十 G 数据，公司长期在处理数据方面有大量积累。从 2015 年开始，行业内数据处理从算法时代进入以大模型为代表的智能化时代。不久前，美国的 ARK 研究团队发布了名为《Big Ideas 2025》的报告，引起了市场的广泛关注，公司也是十分关注全球的市场变化和同行业公司的发展，目前基因检测行业在智能化分析和 AI 大模型的应用在逐步进入实际应用阶段。比如美国公司 Tempus 主要通过测序及检测业务积累数据，通过人工智能的接入实现医疗数据业务发展。</p> <p>国内方面，医疗行业发布了大量规范性政策，就检测方向而言，样本出院受限，医院目前都在寻求本地化解决方案，人工智能算法为公司本地化解决方案落地、扩大公司产品、服务、销售打下了良好基础。同时，国家卫生健康委、国家中医药局、国家疾控局三部门发布《卫生健康行业人工智能应用场景参考指引》，明确提出医院各种“人工智能+”应用场景，《“健康中国 2030”规划纲要》也提出要积极推进数字健康战略，公司努力在符合纲要的情况之下，将贝瑞的智能化产品加速推向医院，提升医院数字化转型速度，加快数字技术与医疗卫生健康深度融合。</p>

对于“AI+医疗”方面，结合公司主营业务，公司希望以测序为基础，提供更多创新的检测产品。贝瑞基因拥有 15 年的基因检测基础，基因检测覆盖全生命周期，积累了大量遗传相关的人类基因组数据，包含实验、科研、临床数据等方面，对公司智能化算法、软件开发及大模型应用奠定了良好的数据基础，结合人工智能算法的应用，公司将更高质高效地将数据资源推向医疗临床使用。

在数据资源的积累与建设方面，我司已取得显著成效。目前，基因检测数据中心的数据存储量已突破 PB 级，海量数据资源中，经过专业开发流程，超亿条数据已具备直接用于公司内部科研、分析及应用的条件，为我们的智能化算法开发提供了坚实的数据支撑。同时，公共数据中心数据库与文献数据中心数据库的建设同样成果丰硕，两者合计数据量亦超亿条，进一步丰富了我们的数据维度。此外，公司构建的医疗知识图谱规模宏大，关系数量超数亿条，节点数接近亿级。所有这些数据，均已严格按照数据处理规范，完成了清洗、脱敏、治理、开发以及分类等一系列精细化处理流程，确保了数据的高质量与高可用性，为公司在医疗大数据领域的深入探索与创新应用奠定了坚实基础。

智能化工具方面，公司拥有多个自研算法，以及以开源大模型为基座的应用大模型，包括：遗传病预测算法、文献智能检索算法、疾病解读大模型、健康管理大模型等。

公司目前拥有两大智能化平台：WESisi 平台助力医院搭建遗传病本地化综合管理，CNVisi 平台智能化解读报告系统助力临床高效解读。

智能化工具能极大提高临床数据解读效率，公司市场优势较强，目前常规为 2000 多家医院和科研机构提供咨询服务，涵盖公司所有临床基因检测产品。公司有信心在未来全新的 AI 应

用落地后，在政策支持背景下获得大量医疗医疗机构的接受和认可。

对 AI 的应用方面，首先从公司内部出发，智能算法对公司降本增效有极大帮助，从每个解读人员从每天个位数的报告解读效率，到目前已经可以达到每天解读几十份报告，极大提高临床报告产出能力，有利于未来更复杂、更高难度的基因检测产品在商业上的推广；其次，大模型极大增加了检测分析能力，有利于消除不同临床机构对于部分疾病的差异表达造成的理解误差，提高基因检测整体服务质量；最后，使用智能化算法极大提高了公司研发团队的研发效率，可快速、定制化地开发应用于特定临床场景的检测产品。

对于 B 端客户，公司计划逐步打造 AI 诊断、AI 慢病管理以及 AI 医疗服务几大板块的智能 Agent；对于 C 端客户，公司也将推出适合其使用的平台，希望能为国内围产期家庭提供家庭医生服务。

公司目前已接入 Deepseek、千问、llama、智谱 AI、BiomedGPT 和 LucaOne 等多种开源模型，通过大模型的应用提升公司运营效能、降低运营成本，为公司临床客户提供更优质的服务。截至目前，贝瑞已与多家大型医疗机构达成战略规划，计划为其开展以分子诊断遗传病咨询为基础、逐渐向其他领域发展的 AI 智能大数据建设，预计年内逐渐落地。

智能数据部负责人李岳珍博士介绍公司 AI 智能化进展：

在数智化转型的征程中，贝瑞基因从数据、软件、工具三个维度协同发力，全面推进公司产品的智能化升级。

在数据层面，众所周知，基因检测行业是一个高度专业化且知识密集领域。在实际的测序工作中，从接到案例开始，我们就需要从患者表型、可能得疾病、相关的基因、已报到的致病变异等多个维度查询专业的知识库。对于那些未

被数据库收录的信息，我们还需从海量文献中寻找可能的支持证据。而这些信息，不仅要在准确性上有所保证，其实时性和全面性同样至关重要。鉴于此，公司依托大数据算法，构建了一套能够自动获取信息的离线数仓。并在底层设计中就进行了完整的数仓规划，同时建立了相应的数据标准和数据指标。我们希望通过深入理解原始数据以及对使用场景的调研，能够在初期就将模型设计得足够完善。随后，将这些模型推进到各自的实际开发空间中，从而让我们的数据开发工程师能够更加高效地完成数据开发工作。为了确保数据的安全性，我们还进行了严格的空间隔离和数据落盘加密等措施。

公司的医疗数据湖涵盖了以下三大核心内容：

第一是公共数据中心。正如我之前提到的，当我们收到到案例时，需要分析其表型并分析可能对应的疾病。在完成检测后，我们将面对一个庞大的数据集。为了能够快速找到可以解释该病例的基因、变异，公司建立了公共数据中心，从整个数据体系出发，将其关系和节点整合形成一个整体的公共数据知识图谱。

第二是文献数据中心。每一种疾病及其相关基因的关联性都源自于一篇篇文献的积累。目前，公司建立的本地文献数据中心已经收录了超过 3500 篇 NCBI 的标题和摘要数据，以及公开的 PDF 和补充材料的原文数据，并会与数据源同步更新，这大大有助于我们高效的使用这些文献。此外我们还通过自研的“医学实体抽取归一化算法”对从标题、摘要中疾病、表型、基因等内容进行了结构化的开发，从而帮助我们快速找到公共知识库中尚未收录的文献。

第三是基因检测数据中心。基于公司 15 年来积累的上千万临床数据，我们对其进行了群体化开发，形成了自有的频率库、报告库以及通过专家解读的变异位点库。在此基础上，我

们对一系列表型进行群体分析，增加了联想功能，显著提高了解读和遗传咨询的效率。同时，我们还设计了实时信息更新提醒机制，以便在最新文献或研究成果发表后，能够及时通知报告解读人员进行相关报告的重分析，进一步提升对于其临床表型的可解释性。

在智能化的道路上，其实贝瑞起步非常早。在外显子（WES）检测在临床推广的初期，就已经开发了 berrylyzer 智能算法，并在 2020 年就已投入真实的临床样本检测中。该算法实现了全流程的智能化，可以从口语化的临床描述直接预测可能的疾病，结合基因检测数据，预测最可能解释该表型的致病性变异。该模型已经经过了几十万来源于不同临床机构的真实世界样本的考验，模型准确率已达 99% 以上。在效率方面，以一个 WES 为例，从临床表型到需要特殊关注的位点分析，可以在一分钟内完成。即使是更复杂的全基因组（WGS）检测样本，也能在十分钟内完成分析。在所需计算资源方面，该算法不依赖 GPU，使用服务器的 CPU 即可完成医院的本地化部署，这将极大地改善医疗机构人员专业化不足的问题，该算法已内置于 WESisi 软件中。该软件聚焦遗传病本地化的痛点，集遗传病数据分析系统、遗传病智能化解读系统、遗传病检测全流程业务管理系统等三大核心系统功能，真正做到了助理院端搭建遗传病本地化综合管理平台。

针对拷贝数变异（CNV）的 CNVisi 智能化解读报告系统也同样实现了“端到端”的输出，大大的降低了如科诺安、贝比安等拷贝数变异检测产品，报告解读的门槛，实现了从临床表型到报告的全流程智能化。该软件内置了 20+ 权威公共数据库及百万中国人 CNV 内部数据，解读结果与“金标准”二分类一致率达 95% 以上。

围绕一线业务人员的痛点和需求，我们还开发了一系列智能化小工具。例如，一整套的医学实体抽取归一化算法，

能够针对大量的报告、文献、公共知识库等信息，快速进行结构化处理，然后针对这些信息进行智能搜索，以及对搜索结果快速解读和总结。这些工具充分利用了我们的海量数据和人工智能大模型，为一线业务人员和临床医生的工作提供了有力支持。

此外，公司还开发了帮助医生和患者更好地理解报告、理解疾病的大模型及配套的对话小助手。使每一个医生都能快速了解某一疾病全方面信息，让罕见病不再“罕见”。该大模型已接入公司的公共数据中心知识图谱，通过 AI 实现智能问答和意图识别，引领医生和患者更好地理解疾病并规划后续治疗方案。

问答环节

1、贝瑞基因如何进行数据梳理？

答：基因检测领域涉及疾病、表型、基因等多维度知识，数据量庞大且更新迅速，知识体系复杂程度高。在这个背景下，我们在数据梳理面临着诸多挑战，其中数据标准不统一的问题尤为突出。不同业务线的检测数据存在差异，统计口径也不一致，同时，各公共知识库中的命名方式也各具特色。为有效解决这一问题，我们在项目开发的早期阶段便着手建立统一的数据标准，积极推进数据整合工作，以确保数据能够统一且灵活地对接。这一举措不仅为后续的数据处理和分析奠定了坚实的基础，也极大地提高了数据的可用性和可靠性。

在成功建立数据标准并充分理解业务需求之后，我们进一步通过模型带动开发的方式，对海量数据进行系统化的梳理。这一过程涵盖了数据的实体关联、维度退化、聚合汇总等多个方面，旨在确保数据的准确性、完整性和一致性。通过这一步骤，我们能够更加深入地理解和利用数据，为后续的数据分析和应用提供了有力的支持，从而更好地挖掘数据的潜在价值。

为有效避免“烟囱效应”，我们采用了先进的大数据架构来处理数据。具体而言，我们按照数据湖和数据域的原则来开发数据，从原始数据层（ODS）到数据仓库明细层（DWD），再到数据维度层，最终到支持服务的分析数据层（ADS），每一层

都严格遵循大数据架构的设计原则。这一架构不仅使得这套开发标准能够很好地在医院环境中进行落地和复现，还极大地提升了数据的可解释性及数据处理的效率和质量，为数据的广泛应用提供了有力保障。

在数据安全方面，我们对私有数据进行专业的脱敏处理，将单人的数据转化为群体数据。这一过程既满足了公司内部数据挖掘、算法开发的需求，又有效避免了个人信息的泄露风险，确保了数据的安全性和合规性。通过这一措施，我们能够在充分利用数据价值的同时，严格保护用户的隐私权益。

最后，我们还构建了一个完善的数据开发框架，能够帮助客户私有数据建立频率库和位点库。这一框架不仅解决了数据出院的问题，还为客户提供了更加精准和个性化的数据服务，进一步提升了客户满意度和市场竞争力。

2、大模型后续发展和未来应用场景

答：在公司整体规划中，我们精心布局了两大核心板块，致力于推动基因检测领域的创新发展，为医生和用户提供更加精准、全面的医疗服务。

第一大板块，我们聚焦于专业领域，全力向疾病解读的大型模型迈进。这一模型的开发旨在以高效、便捷的方式，助力医生深化对疾病的认知，并能更加快速精准的解读变异的致病性。通过整合海量的基因数据和临床信息，我们致力于打造一个智能化的疾病解读平台，为医生提供有力的决策支持，提升诊断的准确性和效率。

第二大板块，我们在遗传咨询大模型的基础上进行升级，打造健康管理大模型健康。鉴于基因检测覆盖全生命周期，其应用范围不仅局限于检测报告的深度解读，更可延伸至日常生活的疾病知识问答。这一板块的开发将全方位帮助用户，伴随他们健康成长的每一步。通过提供个性化的健康管理方案和疾病预防建议，我们致力于提升用户的健康意识和生活质量，为他们的健康保驾护航。

我们相信，通过这两大板块的协同发展，我们将能够为医生和用户提供更加全面、精准的医疗服务，推动基因检测领域的创新和进步。

3、deepseek 等大模型的接入情况，给贝瑞带来了哪些影响？

答：首先，在变异解读大模型和遗传咨询大模型的开发中，我们已经使用了 llama、qwen、GLM 等开源基座大模型。现正在全力进行与 deepsee 系列模型的平行测试。另一方面 deepseekR1

	提出的蒸馏方案，也给专业化疾病解读等模型的开发、落地带来了希望。使我们在可控的计算资源、和有限的高质量数据集下就能快速进行临床转化和产品升级。
附件清单(如有)	
日期	2025 年 2 月 11 日