

医疗和生命科学大数据 引领个性化医疗

魏 智, Frank.wei@intel.com
首席架构师, 医疗和生命科学部
Intel 亚太研发有限公司

www.intel.com/healthcare/bigdata
www.intel.com/healthcare/optimizecode



网络化的医疗服务

从单一资源向基于跨组织和IT系统的支撑团队变迁



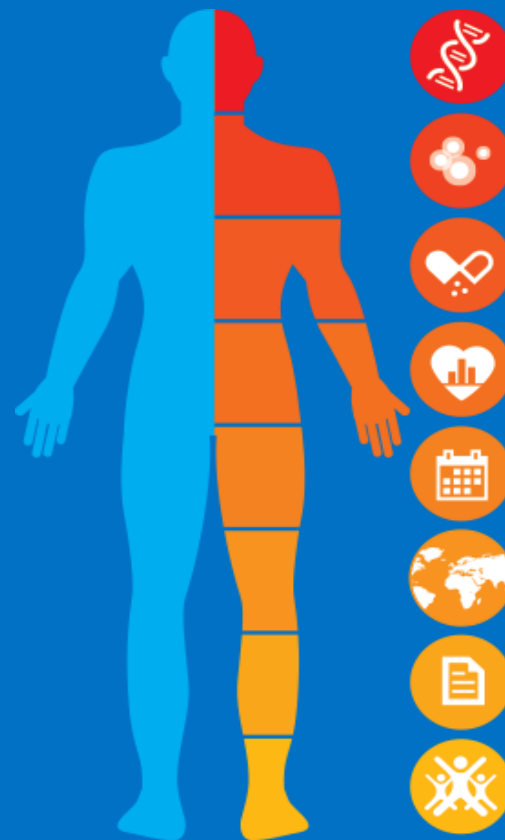
无所不在的医疗服务

移动、家庭和社区医疗补充院内资源



个性化的医疗服务

从基于人群的一般治疗，发展到针对个体的治疗



Intel推动全球医疗和生命科学发展发展的基本理念

什么是精准医疗？

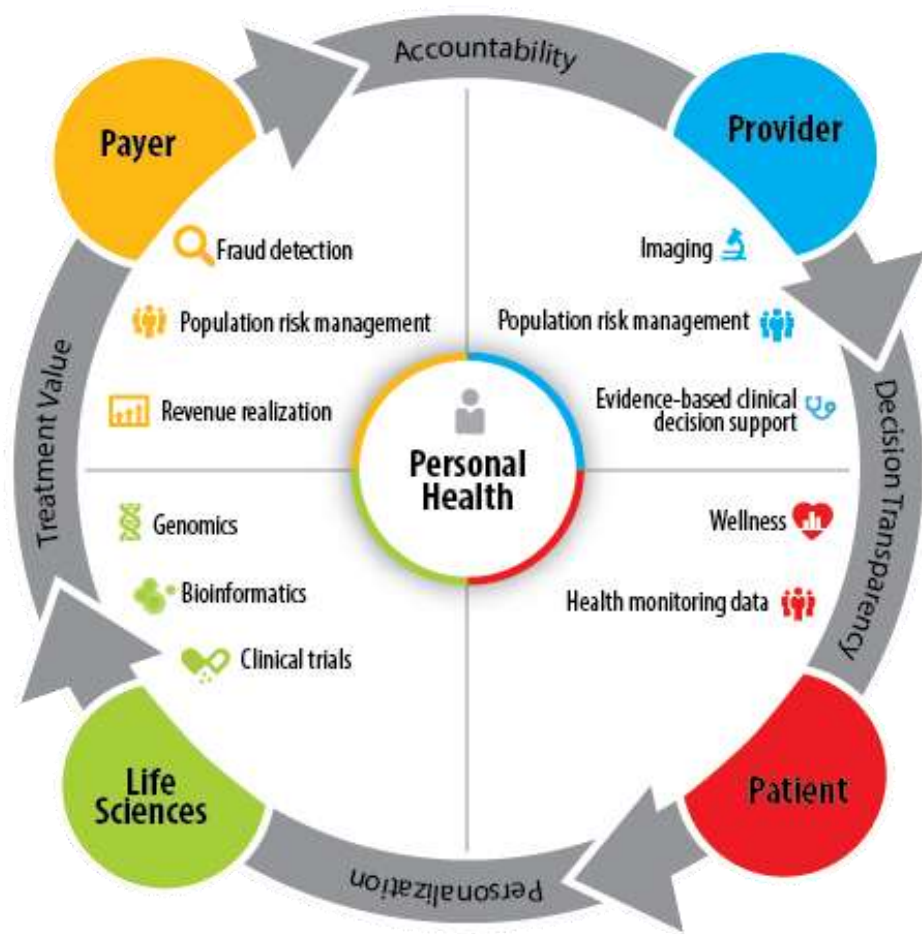
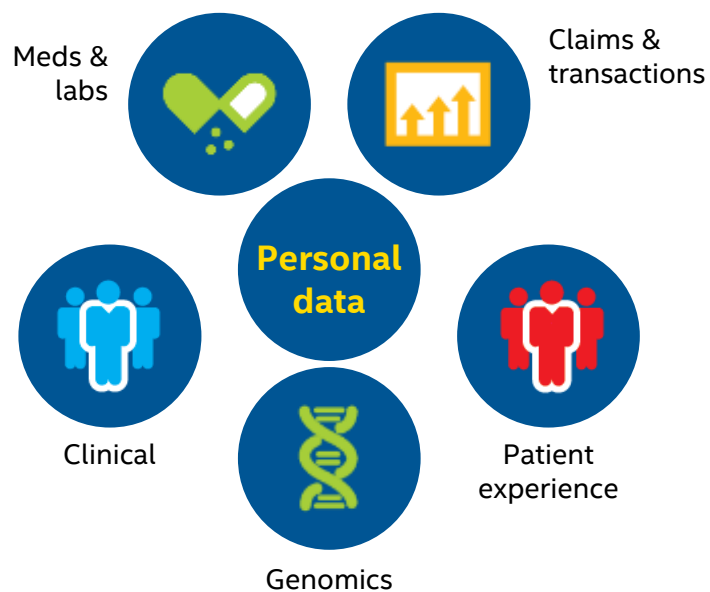
一种将每个个体在基因、环境、和生活方式上的差异用于疾病预防、诊断和治疗的新方法

<http://www.nih.gov/precisionmedicine/>

在精准医疗方面Intel的愿景

目前的问题:

存在很多不同数据、数据流



未来的方向:

集成的临床指导信息

带来更好的诊断

- 以病人为中心
- 更健康的人群
- 基于价值的治疗, 降低成本的高效治疗

大量数据带来的挑战



数据

- 收集和管理分布在不同的医疗和研究机构
- 健康信息的安全性担忧
- 大部分没有进行应有的监管和定量分析
- 原始数据不能移动
- 缺乏标准
- 分析需要密集计算
- 在诊断和EMR（电子病历）之间缺乏便捷链接



数据分析

- 分析需要几天到几周完成
- 软件多元化、差异化
- 人力和架构结构成本难以获得
- 支持临床诊断的分析需要多个机构的数据和经验积累的汇总
- 公共数据不易获得，数据分析需要内部IT系统的大运算量支持

我们的使命

提供创新的计算技术、方案
改进医疗服务的成本、质量，
并连通全球数十亿人



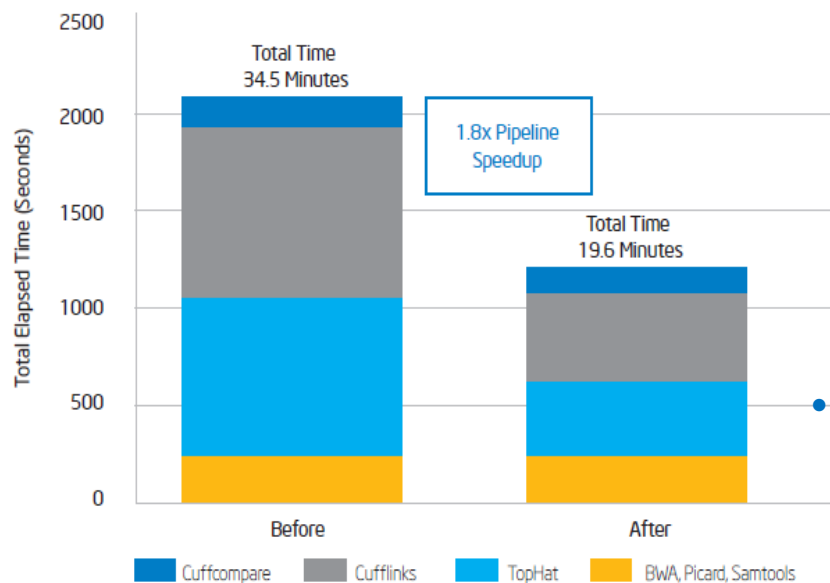
用于基因组序列的一体机方案

TGen* RNA-Seq



RNA-Seq 从7天改进到4小时

- **问题：**检测以前需要7天时间. 拖延了非常紧急的临床治疗，例如儿科癌症
- **新方案：**戴尔使用了基于Intel处理器的基因组数据分析平台
 - 均衡HPC计算、组网和存储的单机架方案
 - Intel® Xeon® E5 v2 处理器基础上的9万亿次运算能力
 - Intel® Enterprise Edition for Lustre*
 - Intel® Cluster Studio XE
- **新方案的改进效果：**
 - RNA-Seq pipeline 从 7 天 减少到四个小时
- **带来的价值：**快速的测序结果，使治疗过程中 可以多次测序、观察患者反应, 调整方案, 改进效果

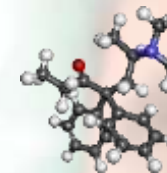


商业价值

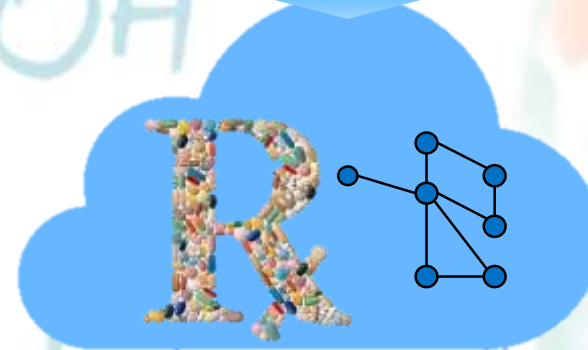
- 通过专注研究，更快、更低成本地发现新药
- 减少患者毒性药物反应
- 通过预测现有复方的特异疗效，减少治疗成本

异构数据集的集合揭示了复方药之间的关系，
获得治疗功效的本质原因

25 个药理学数据集



4千1百万种组合



- 数据准备
- 特征提取
- 网络、图形建模

- 可视化
- 机器学习
- 查询

高级分析

- 医学研究
- 毒性
- 药物相互作用
- 疗效预测



Glucose =
 C_6H_{12}
6 Carbon
12 Hydrogen

综合临床、基因、染色体和家庭健康历史数据的全新 CDS（临床诊断支持系统）的未来应用



- 目标：推广有助于医生、咨询师进行风险评估、帮助遗传咨询专家进行基因测试的临床诊断支持的广泛使用
- 建立一个可扩展的标准化数据基础上的CDS：
 - 家庭健康记录
 - 临床和筛查
 - 基因数据

方案应用:

- 兼容数据收集: 扩展到不同的临床领域和其它健康机构
- 基于标准的应用
- 跨医院数据平台 (EHRs)
- 借助 Intel 基础架构、数据平台等技术
- 灵活的影像、个人数据接入



Intel 架构的定位

大数据和 HPC (高性能计算)



Sequencer



Sequencing Appliance



HPC Pipeline Cluster



Storage and Networking



Research Computing



Precision Medicine



Intel® Xeon® / Intel® Xeon® Phi™ • Rack Scale Architecture • Integrated Fabric • Software Defined Network • Solid-State Drives • Workload Optimization

Hadoop* • Lustre* • Cloudbursting • Virtualization • Encryption • Cloud-based Analytics

Intel® 架构推动测序仪、一体机到高性能云计算

处于变化的十字路口

10^{18}

超级计算



开放、互通的云



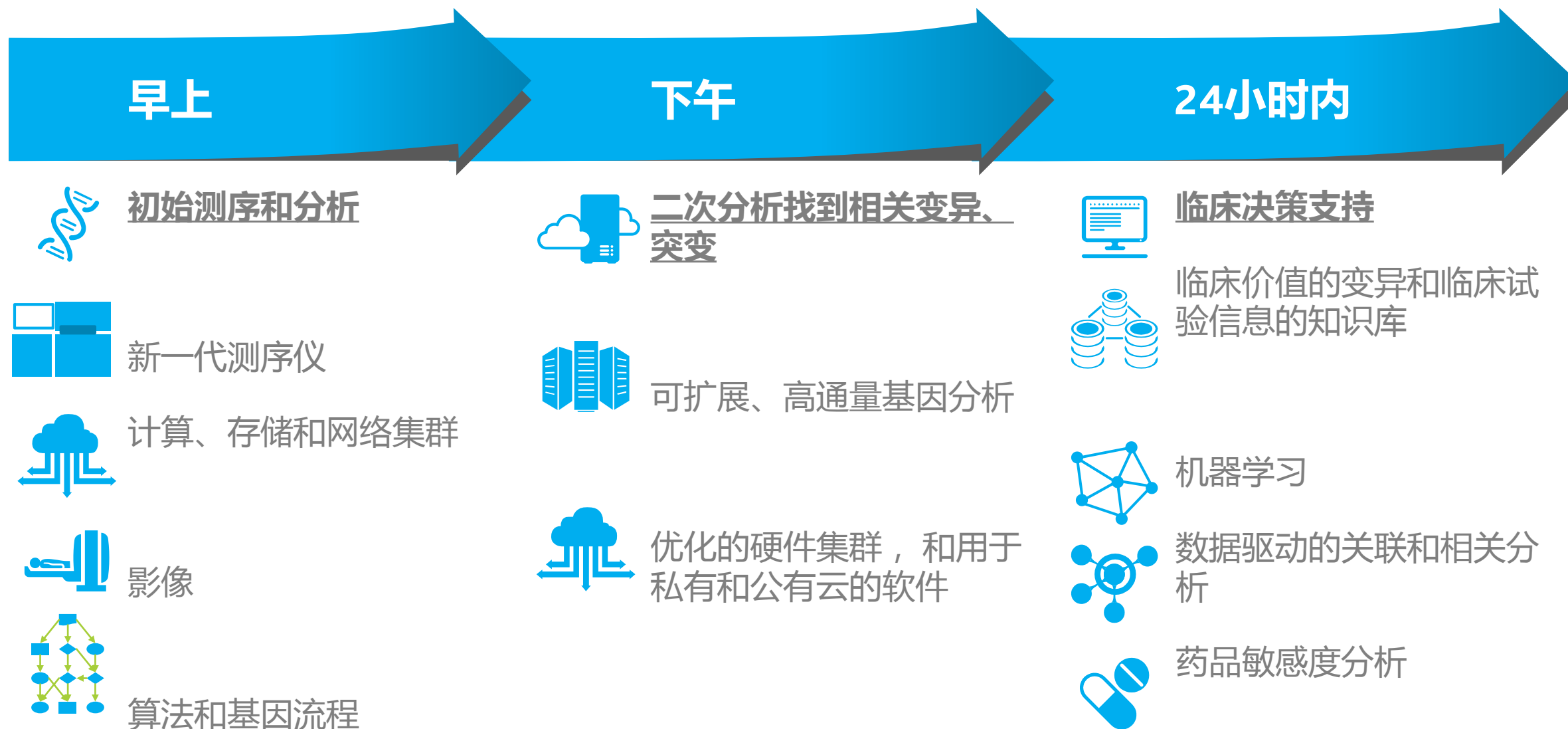
提供代码、培育生态链

精准医疗现在需要多长时间?

活组织检查 -> 测序 -> 分析 -> 诊断 -> 治疗

能不能一天完成(All-in-One-Day)?

2020 ‘一天完成’ – 医院场景背后的技术



支持‘一天完成’的生态链变革



2020年能否实现‘一天实现’？

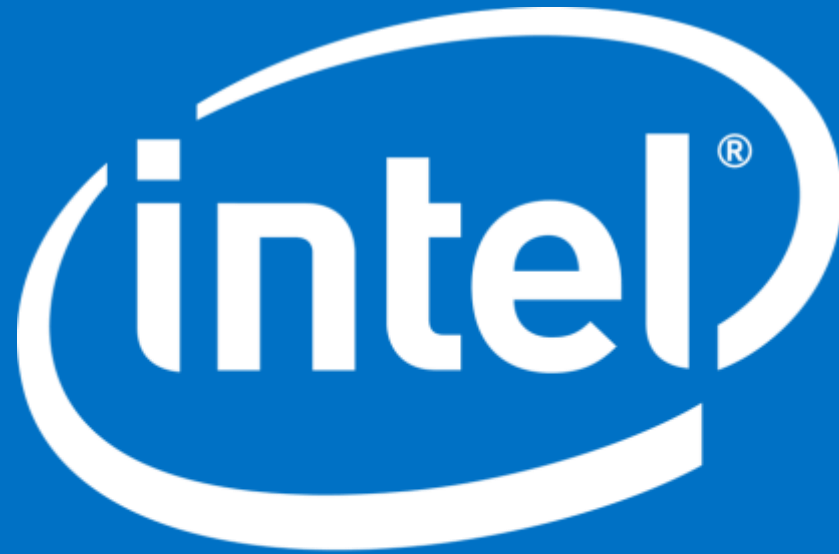
Legal Notices and Disclaimers

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Configurations: Test used a two-socket Intel Xeon processor E5-268W (3.1 GHz). Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Before: Bowtie2 2.0.0-beta7, TopHat 2.0.4, Cufflinks 2.0.2; After: Bowtie2 2.1.0, TopHat 2.0.8b, Cufflinks 2.1.1.

Processor configuration based on current Intel benchmarks: <http://www.intel.com/content/www/us/en/benchmarks/server/xeon-e5-2600-v2/xeon-e5-v2-hpc-life-sciences.html>.

For more information go to <http://www.intel.com/performance>



experience
what's inside™