Oracle Health Sciences
Translational Research Center:
A Translational Medicine Platform
to Address the Big Data Challenge
Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Abstract .................................................................................................................................1
Introduction ..........................................................................................................................2
Oracle Health Sciences Translational Research Center Features ..........................5
  Intuitive Patient Stratification Interface with Both Clinical and Omics Criteria ....5
  Cross-Vendor and Cross-Modality Omics Data Repository and its Biological Context – Simple yet Comprehensive .................................................................6
  Variant Impact Assessment Accelerates Finding of Actionable Variants ..........8
  Embracing Community Standards and Efforts .......................................................8
  Data Lineage Across Multi-Level Result Files – Supporting Traceability ..........9
  Consistent, Accurate, Complete, and Standardized Clinical Data .......................10
  Scalability to Solve Big Data Problems in Translational Research..................10
  Multiple Secure Deployment Options to Meet Customer Needs ..................11
Architecture Components ..........................................................................................11
  Oracle Healthcare Data Warehouse Foundation .............................................12
  Oracle Health Sciences Omics Data Bank ..........................................................12
  Oracle Healthcare Analytics Data Integration ...............................................13
  Oracle Health Sciences Cohort Explorer ............................................................13
Conclusions .......................................................................................................................14
Abstract

As organizations move to advance translational research to achieve personalized medicine, researchers and clinicians must manage informatics, however, there is a shortage of fully-integrated informatics solutions that integrate, store, and analyze clinical and omics data from diverse sources – generated in-house as well as public consortiums. Many researchers and clinicians must rely on bioinformaticians to perform mundane data management tasks in order to validate a simple hypothesis. Oracle Health Sciences Translational Research Center provides a complete and scalable informatics solution, with centralized data storage and analysis across genetic information areas (genomics, transcriptomics, and proteomics), vendor platforms, biological data types, and clinical data sources. It also offers flexible deployment options – on-premise, HIPAA-certified Software as a Service (SaaS), and a hybrid of these two. Through an intuitive user interface, Oracle Health Sciences Translational Research Center enables researchers to stratify patients and allows clinicians to evaluate treatment responses for similar patients in a self-sufficient manner, ultimately shortening the biomarker development cycle and accelerating the adoption of personalized medicine.
Introduction

Rapid growth of genomics data, increased patient expectations of quality of care, and financial pressures to improve operational efficiencies have created fundamental challenges and opportunities for healthcare and life sciences organizations to develop and implement biomarker therapies. One of the major promises of personalized medicine is an increase in drug efficacy. In 2008 and 2009, success rates for new drugs in Phase II clinical trials from 16 pharmaceutical companies decreased 10%. Of these failures, 51% were due to insufficient efficacy and 19% due to preclinical safety. Although many organizations, including healthcare providers, insurers, pharmaceutical companies, and diagnostic laboratories, realize the potential of personalized medicine, they still lack adequate implementation of an informatics solution.

The main obstacles to adoption of the new paradigm include:

- Lack of seamlessly integrated, productized, and scalable technology infrastructure to support data management, analysis, and reporting
- Vendor- and modality-specific omics data are difficult to interrelate, and isolated data in clusters (that is, silo omics) are insufficient to provide biological insights
- Difficulty in jointly analyzing public domain data and data generated in-house
- Disparate clinical data sources requiring heavy post processing
- Lack of flexible and HIPPA-compliant deployment models

As the cost of DNA sequencing continues to decrease and technological advances make sequencing both easier and more robust, the bottleneck in translational research has shifted from data gathering to data analysis. Many researchers are finding insurmountable analytical hurdles in testing simple hypotheses without the ad hoc assistance of a bioinformatician. For bioinformaticians, poor infrastructure design has made it difficult to build tools to enable researchers to be self-sufficient in their standard analyses. This has created an environment where bioinformaticians are required to perform mundane tasks, diverting their time and focus away from challenges and innovations. This ad hoc analysis paradigm often leads to scattered data and analysis files across multiple storage devices. As a result, it is difficult to reproduce results and transfer knowledge to external collaborators in accordance with regulatory requirements. In the past, some institutions have contracted consultants to build customized solutions, which is very time-intensive. The customized solution also becomes a maintenance challenge due to the high cost of long-term contracts for external expertise and a disconnect in knowledge between contractors and in-house resources.

Most informatics infrastructure today focuses on a single omics technology vendor (for example, Illumina, Life Technologies, and others) or in a single omics data modality (for example, genome versus transcriptome). While useful, this approach provides a fragmented picture of the underlying biological processes. Overwhelming evidence has shown that an integrated approach is the key to identifying the root cause of a disease based on gene structure, expression, and regulation across multiple omics modalities.²–³ For instance, an integrated analysis of cross-modality glioblastoma (GBM) data, including DNA copy number, gene expression, and DNA methylation aberrations, helped dissect genome-wide regulatory mechanisms for further investigation into the identification of candidate biomarkers for GBM tumors and potential therapeutic targets.⁶ A holistic view of cross-vendor and cross-modality data is thus critical for the development and delivery of targeted medicine.

The research community has devoted significant resources to make large-scale genomics data available to the public. For example, The Cancer Genome Atlas project (TCGA) and the International Cancer Genome Consortium (ICGC) offer comprehensive sets of complementary resources for researchers to understand the complete repertoire of mechanisms contributing to tumor initiation and progression. Similarly, the 1000 Genomes project provides a detailed catalogue of human genetic variation. These efforts have provided a great opportunity for researchers to combine public data with the data generated in their own laboratories; to simulate the testing of new types of hypotheses, increase statistical power of tests, and sub-sequentially generate new biological insights. Given these public data, researchers and clinicians can theoretically map their own patients' molecular profiles against the public data to predict disease progression and treatment responses. However, this requires daunting manual steps or complex programming skills. For example, when verifying whether a particular mutation occurs in TCGA GBM patients, a researcher needs to pass through several files on the TCGA website to identify the relevant ones, download them to the local storage, write script (using bioinformatician's help), and finally perform the analysis to make a prediction. As a result, most public data are grossly under-utilized. Furthermore, this approach also leads to redundant copies of the same public data in an institution as researchers in different labs independently manage their own copy – multiplying data storage and labor costs.

In the clinical environment, many translational research projects disproportionately focus on molecular profiling technologies compared to the clinical information collection systems, which are commonly done through a project-specific case report form. Without heavy and consistent post-processing, clinical data reuse and comparison across projects is impossible. As translational research moves toward personalized medicine, the sources of clinical data will continue to expand beyond case report forms, to include electronic health records, state or national registries, and records from various departments such as accounting, pharmacy, and pathology. From an informatics perspective, translational research tools need to support the integration and standardization of clinical data from disparate sources with the help of consistent terminology and units of measure. Once data is standardized, it will not only benefit translational research, but also the analysis of institutional-wide metrics such as operating resource and physician performance metrics.

Healthcare and life sciences organizations have divergent informatics needs from on-premise infrastructure to integrated and externally-managed solutions such as software as a service (SaaS). Given the sensitive nature of clinical records and increased regulation of omics data, the hosting system must be compliant with the Health Insurance Portability and Accountability Act (HIPAA). For example, the system must be able to encrypt sensitive data and audit each user’s actions on Protected Health Information (PHI). Yet, there are limited deployment options available in HIPAA-certified systems. Standardized data and analysis will likely shift deployment needs from on-premise to SaaS, or to a hybrid model, where only a relatively small set of crucial data and analyses are managed in-house. The remaining data and analysis are managed in the cloud, with the continuous synchronization between local copies and the SaaS solution.

Given the above barriers in informatics, a complete informatics solution for translational research must include the following features:

- A productized solution with quick time-to-value and long-term support
- Scalability for data management, analysis, and reporting
- An easy-to-use interface with seamless integration between public domain data and in-house data, as well as an integration between public algorithms and in-house algorithms
- Support standardization at many levels, from clinical data, to vendor technologies, and omics modalities
- A flexible, regulatory-compliant informatics infrastructure deployment model
Oracle Health Sciences Translational Research Center addresses all of these requirements. It standardizes data sets and brings them together on a common storage, leveraging the economies of scale and keeping track of lineage information. Oracle Health Sciences Translational Research Center also captures biological relationships among data in the repository, allowing it to adequately represent data across vendor platforms and omics modalities. Since Oracle Health Sciences Translational Research Center is agnostic to data-generating vendor platforms and can capture relationships among cross-modality data, both public data and in-house data can be analyzed on a common platform to achieve biological insights. Hosting a large amount of data from diverse sources, Oracle Health Sciences Translational Research Center has been optimized to handle petabyte size data by combining software and hardware features to support extreme performance.

Oracle Health Sciences Translational Research Center was developed in collaboration with strategic development partners of repute, including: Erasmus Medical Center, Inova Health System, Merck & Co., Moffitt Cancer Center, Oregon Health and Sciences University, and Roche Ltd. Its productized approach ensures that product features support best practices in translational research while accelerating time-to-value.

**Oracle Health Sciences Translational Research Center Features**

**Intuitive Patient Stratification Interface with Both Clinical and Omics Criteria**

Finding subjects with certain characteristics is a fundamental step in biomarker discovery and clinical decision support. Oracle Health Sciences Translational Research Center provides an easy-to-use interface that eliminates the need for programming or scripting skills. With a few clicks, users can specify both clinical and omics criteria of a desired patient group.

![Figure 1. Oracle Health Sciences Translational Research Center patient stratification use case](image-url)
Figure 1 illustrates a use case where the user wants a group of male patients with non-small-cell-lung carcinoma (NSCLC) with a mutation in the EGFR gene, but no copy number gain of the MET gene. There are four criteria specified in the interface:

- Gender criterion in the Patient Information category
- Diagnostic criterion in the Clinical Data category
- Mutation criterion in the Genomics Data category
- Copy number variant criterion in the Genomics Data category

Once all the criteria are specified, the user can obtain the number of patients who meet the criteria along with detailed information of selected patients in either a tabular or an event timeline view.

Cross-Vendor and Cross-Modality Omics Data Repository and its Biological Context – Simple yet Comprehensive

A critical factor in the success of an integrated approach to translational research is the ability to interrelate omics data. The relational tables within Oracle Health Sciences Translational Research Center is designed to follow the Central Dogma of Molecular Biology, following the series of biological events in the cell, from DNA sequencing, to splicing, to RNA transcription, to translation, and finally to protein.

There are two sets of relational tables modeling omics data in Oracle Health Sciences Translational Research Center: the reference and the result tables. The reference tables store the reference genome builds and associated annotations, while the result tables store specimen data generated by laboratories and data from public data repositories (for example, TCGA). Selected entities in the reference tables are shown in Figure 2 (For illustration purposes, the exact table names in the product are slightly different from those shown in the figure. Subset of reference tables is not included in the pictorial illustration for simplicity purposes.).

- DNA_STRAND table stores each species' genome build.
- GENE and GENE_COMPONENT tables store genes and their sub-components, as well as additional annotations.
- PROTEIN and PROTEIN_COMPONENT tables store information about proteins, their sub-components, and annotations.
- As shown in Figure 2, Oracle Health Sciences Translational Research Center captures the fact that only subsets of gene components translate into protein, accounting for alternative splicing events.
- PATHWAY table models pathway memberships.
- VARIANT table stores the differences compared to a reference genome, including single nucleotide polymorphisms (SNP), insertions, and deletions.
Figure 2. Graphical illustration of the core reference tables in the Oracle Health Sciences Translational Research Center solution (The 3D protein image is downloaded from Protein Data Bank, and it is only for illustration purposes.)

The generality of the reference tables can accommodate data from many reference sources such as Ensembl, NCBI, HUGO, and others. To facilitate adoption, a set of data adapters are included as part of the product for reference source data (details are provided in Section 3.2). Although many existing genomics data processing solutions are restricted to store known variants, this is insufficient for users of the Next Generation Sequencing (NGS) technology, who commonly discover new polymorphic sites. Thus, Oracle Health Sciences Translational Research Center solution models both known and novel variants by combining variant attributes in the VARIANT table and coordinate information in the DNA_STRAND table.

These reference tables not only provide biological context, but also serve as an anchor for results data acquired from different platforms. For example, each sample’s sequence variant data are linked to the VARIANT table to access the biological context. It is also cross-referenced to dbSNP, the Catalogue of Somatic Mutation In Cancer (COSMIC), and other public annotation databases. Each sample’s gene expression data is linked to the GENE table, and each sample’s copy number variant data is linked to DNA_STRAND table. With this anchor design, it is straightforward to find patients with the EGFR mutation, but not MET gene copy number gain. This biology-focused nature of the Oracle Health Sciences Translational Research Center simplifies extensions to new types of omics modalities, adapting to continuous innovation in molecular technology.
Variant Impact Assessment Accelerates Finding of Actionable Variants

Exome and targeted sequencing projects routinely report hundreds of thousands of variants, while whole-genome projects report millions of variants. Since researchers often start with nonsynonymous changes, that is, which alter amino acid sequence, Oracle Health Sciences Translational Research Center offers two sets of impact assessment for each variant, both capturing multiple transcripts per gene.

- Variant impact: For each sequence variant at the coding region of a gene (for example, missense, nonsense, frameshift and others), Oracle Health Sciences Translational Research Center computes the type of effect it has on the protein sequence. This computation includes both known variants and novel variants reported in an experiment.

- Effect of amino acid substitutions on protein function: Oracle Health Sciences Translational Research Center predicts amino acid changes as a result of a sequence variant that affect protein function using the methods of Sorting Intolerant From Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen).<sup>7, 8</sup>

These two approaches allow users to easily narrow millions of sequence variants down to a manageable set of candidate variants to conduct further analysis.

Embracing Community Standards and Efforts

The bioinformatics community has developed standards and formats for omics data from various modalities. Specifically, in the area of NGS, one to two data formats dominate at each level of data analysis. Oracle Health Sciences Translational Research Center includes multiple data adapters from common genomics data formats (details are in Section 3.2). Oracle Health Sciences Translational Research Center also supports the integration of the Visquick library for a circular genomics viewer<sup>9</sup> and of the Integrated Genomics Viewer (IGV)<sup>10</sup>. As shown in Figure 3, users can specify a particular specimen and its available molecular data to be shown in the circular genomics viewer. If multiple omics modalities are selected, the viewer shows each one in a concentric circle. In the IGV, users can specify specimen, genomics data modality, and genome region of interest at the control panel, and export a file in one of the IGV-supported formats.

---


<sup>9</sup> Visquick: Javascript project providing a library of configurable SVG and Canvas-based visual tools 2012, Institute of Systems Biology.

Figure 3. Oracle Health Sciences Translational Research Center’s integration with the Visquick library provides a cross-modality circular genomics viewer for a selected specimen.

Data Lineage Across Multi-Level Result Files – Supporting Traceability

In our solution, lineage is maintained among the entities in the underlying data and their corresponding raw data. For example, Oracle Health Sciences Translational Research Center stores the file name and directory of aligned reads and then links this entry to the sequence variant results that are derived from these aligned reads. If a user finds an interesting sequence variant, a variant with a particularly low quality score, or an SNP that prompts doubt about the accuracy of the raw data, users can trace it back to the corresponding sequence alignment and raw reads. Oracle Health Sciences Translational Research Center also lets users export the alignment files for further examination in a genome browser. This feature is particularly important for oncologists in the clinical decision support setting due to the extremely heterogeneous nature of tumor samples. Tracing back to the raw data helps clinicians confidently differentiate mutations from noise in the data.
Consistent, Accurate, Complete, and Standardized Clinical Data

Clinical data are often inconsistent and incomplete, especially when gathered for different purposes and sourced from multiple and diverse systems. Thus, any data integration approach must provide separation of data from non-validated, uncertain, or poor quality sources from consistent, accurate, complete, and standardized data. Data management by Oracle Health Sciences Translational Research Center is achieved through source-system agnostic data structures that facilitate the ease of loading and managing of semantics, relationships, and business rules from multiple source systems. The data and associated metadata are then standardized and loaded into a comprehensive warehouse available to downstream applications. Oracle Health Sciences Translational Research Center can also simultaneously refer to multiple terminologies fitting into diverse use cases. All of these features are critical to support queries and analyze information from multiple, disparate, heterogeneous systems – demographic, diagnostic, laboratory, phenotypic, life style, claims, and more – to enable comprehensive, focused clinical research.

Scalability to Solve Big Data Problems in Translational Research

When targeting therapies, large amounts of information must be analyzed and distilled down to drive decisions. Oracle Health Sciences Translational Research Center is capable of integrating millions of patient records and a hundred thousand whole genome sequences. To manage petabytes of data records, Oracle Health Sciences Translational Research Center was developed with expertise in jointly optimizing software and hardware features. Leveraging Oracle Exadata technologies including smart scan, storage index, and flash cache, and our proposed partition method, Oracle Health Sciences Translational Research Center ensures sub-second response times for many queries in scenarios with 100,000 whole genome sequence data. It is highly scalable, which ensures the research platform can scale up and continue to deliver results as the needs for genotype-phenotype analyses expand.

Data loading speed from external system to Oracle Health Sciences Translational Research Center was benchmarked with the 69 publicly available data sets from the Complete Genomics Inc. website\(^\text{11}\) and the approximately one thousand samples in the 1000 Genomes project\(^\text{12}\). It is able to process hundreds of whole-genome samples per day. This robust performance lets Oracle Health Sciences Translational Research Center handle a large amount of initial data load and accommodate the ever-increasing speed at which DNA sequencing data is obtained. All performance details will be published in the upcoming Oracle Health Sciences Translational Research Center performance benchmarking whitepaper.

\(^{11}\) Public Genome Data Repository, 2012, Complete Genomics: Mountain View, CA.

Multiple Secure Deployment Options to Meet Customer Needs

Oracle Health Sciences Translational Research Center can be deployed in a variety of modes depending on organizational preference. In addition to an on-premise option, the solution can also be deployed as SaaS in a private cloud, or as a hybrid option. Data transparency is assured in the cloud environment. In particular, our cloud service customers benefit from the standard formats we support, and at any point in time customers can export their data stored in Oracle Health Sciences Translational Research Center into the standard formats.

Translational research informatics solutions are often required to handle highly sensitive data, such as patient records, genomics data, and data potentially classified as PHI. Oracle Health Sciences Translational Research Center provides a SaaS environment certified by HIPAA regulations with independent auditing from third-party agencies. The data centers hosting the SaaS environment are also trusted by the world’s largest pharmaceutical companies to manage their clinical trials.

Architecture Components

Oracle Health Sciences Translational Research Center is composed of four fully integrated components: Oracle Healthcare Data Warehouse Foundation, Oracle Healthcare Analytics Data Integration, Oracle Health Sciences Omics Data Bank, and Oracle Health Sciences Cohort Explorer. Oracle Healthcare Data Warehouse Foundation and Oracle Health Sciences Omics Data Bank are the back-end data schemata storing clinical and omics data, respectively. Oracle Healthcare Analytics Data Integration is the back-end clinical data integrator source system. Oracle Health Sciences Cohort Explorer is a Web-based application accessing both clinical and omics data stored in the back-end components.

![Figure 4. Oracle Health Sciences Translational Research Center architecture overview.](image-url)
Oracle Healthcare Data Warehouse Foundation

Oracle Healthcare Data Warehouse Foundation is a comprehensive set of logical and physical data models for deploying pre-built business intelligence, analytics, data mining, and performance management applications from Oracle and its partners. While supporting both standard terminologies and proprietary data dictionaries, the solution includes an expanding list of more than 1,000 entities and 12,000 attributes spanning the clinical, financial, operational, and research domains. It stores data at the most granular level, thereby maximizing the potential for reuse by multiple analytics applications.

Oracle Healthcare Data Warehouse Foundation supports data standardization and cross-validation check, an important step when dealing with clinical data originating from diverse sources. For example, it can find and eliminate duplicate data while ensuring correct data attribute survivorship to create a “single source of truth.” With such a capability, the master data in the warehouse are always up-to-date, consistent and complete. Furthermore, all the transformations during the data standardization process are logged, providing a complete data lineage record with permanent traceability.

Oracle Healthcare Data Warehouse Foundation is a main building block in Oracle Health Sciences Translational Research Center, supplying standardized and high quality clinical data critical for accurate disease classification, and thereby enabling high confidence biomarker discovery and clinical decision support activities.

Oracle Health Sciences Omics Data Bank

Oracle Health Sciences Omics Data Bank is a critical component of the Oracle Health Sciences Translational Research Center solution, as a cross-platform and cross-omics modality data model that delivers the extreme performance required for querying vast omics data sets typical of today’s translational research studies. The solution design was based on the Central Dogma of Molecular Biology. In the latest release, the Oracle Health Sciences Omics Data Bank schema models provide multiple data modalities, including sequence variant, copy number variant, and gene expression through both microarray and sequencing technology (that is, RNA-seq). Oracle Health Sciences Omics Data Bank can also store data from multiple biological species.
Oracle Health Sciences Omics Data Bank includes out-of-the-box adapters for loading customer-generated molecular profiling data, as well as public domain data from sources such as Ensembl, Pathway Commons, TCGA, and the 1000 Genomes Project to further speed time to value. For data generated in-house, it provides adapters to Variant Call Format (VCF)\textsuperscript{13}, Mutation Annotation Format (MAF)\textsuperscript{14}, tab-delimited format for expression data, RNA-seq expression-exon data format\textsuperscript{15}, and MasterVar format from Complete Genomics Inc.\textsuperscript{16}. Oracle Health Sciences Omics Data Bank is scalable to projects with 100,000 whole genome sequences, exceeding the five-year ambitions of most large organizations.

**Oracle Healthcare Analytics Data Integration**

Oracle Healthcare Analytics Data Integration provides a rich platform for processing data from diverse source systems and loading of the homogenized, consistent, complete, correct, and standardized analytics-ready data into Oracle Healthcare Data Warehouse Foundation.

Oracle Healthcare Analytics Data Integration includes a set of interface tables that provide abstraction from Oracle Healthcare Data Warehouse Foundation, serves as a system of original record proxy, supports versioning of data, and acts as a persistent placeholder for incoming data from source systems. Foundation validation and integration rules are integral parts of Warehouse Integration, during which Oracle Healthcare Analytics Data Integration can identify exception conditions and provide comprehensive exception logs with traceability to erroneous data and configurable exception messages. Customers can also configure the integration rules to allow for suspension of data to accommodate late arriving data. Furthermore, Oracle Healthcare Analytics Data Integration supports data governance through data profiling, management of data quality, and metadata management. With these features, customers can obtain data lineage and impact analysis, versioning of data, fulfilling auditing requirements.

**Oracle Health Sciences Cohort Explorer**

Oracle Health Sciences Cohort Explorer is a Web-based interface that provides scientists and clinicians with a dashboard to monitor patient population, based on commonly used metrics, with an agile query interface, and a detailed event viewer for each patient’s longitudinal information. With this tool, clinicians and scientists can be self-sufficient and timely in their own analysis.


Oracle Health Sciences Cohort Explorer accesses integrated clinical and omics data from Cohort Data Mart (CDM) and Oracle Health Sciences Omics Data Bank. As a data mart off Oracle Healthcare Data Warehouse Foundation, CDM contains a subset of entities and attributes that are relevant to translational research. Due to the data standardization capability in Oracle Healthcare Data Warehouse Foundation, the clinical data in CDM are always ready for use by analytic tools such as Oracle Health Sciences Cohort Explorer without further post processing.

Oracle Health Sciences Cohort Explorer provides a highly interactive user-friendly interface for identifying and exploring patient sub-populations based on clinical and omics information such as demography, diagnosis, treatment, specimen, mutation, and copy number variant. The selected cohorts can also be saved for future use. In Figure 1, Oracle Health Sciences Translational Research Center patient stratification use case illustrates a patient stratification use case. Researchers and clinicians can further examine selected patients using the longitudinal viewer, and thus notice any inconsistencies in treatment schedules and treatment outcomes.

In addition to patient-focused cohort queries, Oracle Health Sciences Cohort Explorer also supports queries that are gene-centric and includes state of the art molecular data visualizers, which provide a holistic view of the multiple omics modalities stored in Oracle Health Sciences Omics Data Bank. Figure 3 illustrates an example. This integrated view is the perfect foundation for systems biology-based approaches to further the molecular understanding of disease in the discovery of new drug targets.

Conclusions

Oracle Health Sciences Translational Research Center is a scalable informatics solution for translational research. Its back-end data components – Oracle Healthcare Data Warehouse Foundation, Oracle Healthcare Analytics Data Integration, and Oracle Health Sciences Omics Data Bank – seamlessly integrate clinical and omics data from diverse clinical data sources as well as from vendor-specific and modality-specific omics data silos, providing standardized data readily available for the front-end application, Oracle Health Sciences Cohort Explorer. With a few clicks, researchers and clinicians can identify patients with specific clinical and omics characteristics of interest with full traceability support. This self-sufficient environment will significantly shorten the biomarker discovery cycle and transform clinical decision capability, opening the way for the practice of personalized medicine.
Oracle Health Sciences Translational Research Center: A Translational Medicine Platform to Address the Big Data Challenge
June 2012
Authors: Wanmei Ou, PhD
Jonathan Sheldon, PhD
Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Oracle is committed to developing practices and products that help protect the environment.

Copyright © 2012, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0612

Hardware and Software, Engineered to Work Together