

Oracle® Health Sciences Omics Data Bank

Release Content Document

Release 2.5

E41335-01

June 2013

1 Introduction

1.1 Purpose of Document

The Release Content Document (RCD) communicates information about Oracle Health Sciences Omics Data Bank (ODB) v2.5 release.

1.2 Release Overview

ODB v2.5 release includes significant extensions in omics data types and clinically-related annotations.

2 Product Overview

ODB is a flexible, cross-platform omics data management system that delivers meaningfully biological contents and the extreme performance required by today's translational research studies. Moreover, it is fully integrated with the Oracle Health Sciences Cohort Explorer (OHSCE), bringing both phenotypic and genotypic data to researchers and clinicians.

The data model supports reference omics information and annotation, as well as result data from each individual sample. ODB 2.5 is packaged with a set of data loaders for reference and result data, thereby reducing your adoption barrier. It also includes Oracle SecureFile as an optional file storage mechanism.

ODB 2.5 includes the following components with indication of new components added in this release:

- Reference and annotation tables for biological context
 - Reference genome build, genome annotation, and variant annotation from Ensembl
 - Protein information from Swiss-Prot and UniProt
 - Gene information from HUGO
 - Pathway information from Pathwaycommons
 - SIFT and Polyphone functional annotations

- Multi-version reference support (new)
- Selected clinical annotations from BIOBASE's GenomeTrax database, including HGMD and Drug Targets annotations (new)
- Result tables for individual sample data across multiple omics datatypes
 - Simple variation including SNV, small insertion, and small deletion
 - Gene expression including both single-channel and two-channel microarray data
 - Copy number variation
 - RNA sequencing data
 - Full genotype support including both reference and mutant allele (new)
 - Structural variant including chromosomal re-arrangement, large insertion, and large deletion (new)
 - Data for experiment and analysis details (new)

Data Loaders

- Reference data loaders for
 - Ensembl (for example, genome definition, sequence variant annotations from dbSNP and COSMIC, and so on)
 - Swiss-Prot and UniProt
 - HUGO
 - Pathwaycommons
 - SIFT and Polyphen functional annotations
 - HGMD and Drug Targets from BIOBASE's GenomeTrax (new)
- Result data loaders for
 - VCF v4.1 including both simple sequence variant and large-size structural variant with improved flexibility and usability (new)
 - GVCF for variant and monomorphic reference genotype (new)
 - Tab-delimited format for both single-channel and two-channel microarray expression data
 - Mutation annotation format
 - Copy number variant format defined in The Cancer Genome Atlas (TCGA) project
 - RNA-seq format defined in TCGA project

SecureFile (optional)

- Oracle SecureFile as an option for all data files that should be kept in a secure environment

3 Features

- **Cross-Platform Omics Data Model:** ODB supports omics data management and analysis regardless of the scientific approach (for example, next generation sequencing, gene expression, and so on) and technology platform (for example, Illumina, Life Technologies, CompleteGenomics, and so on).
- **Multiple Species:** ODB can model reference DNA, gene, and protein information from multiple species.
- **Cross Reference Omics Information:** ODB can handle cross reference information for DNA source, gene, gene component, variant, protein, and protein component.
- **Support of Multiple Versions of Reference Genomes:** ODB can organize your result data based on the version of reference genome the data was aligned with. This design improves data management efficiently and reduces interpretation error.
- **Integration with Clinical Data:** ODB is fully compatible with the Cohort Data Mart (part of OHSCE v2.0.2.1) to let end users integrate both clinical and omics information in their cohort selection.
- **Monomorphic Reference Genotype vs. No-call Region:** ODB differentiates monomorphic reference genotype from no-call regions to improve genotype characterization.
- **Annotation Management:** Bringing annotation data to the results, ODB efficiently manages ever improving omics annotations.
- **Scalability:** Result tables are partitioned by study or gene based on your data distribution and usage patterns.
- **Data Loaders:** Easy-to-use command line data loaders with comprehensive instructions.

[Table 1](#) lists the terms used in this document.

Table 1 Terminology

Term	Definition
COSMIC	Catalogue of Somatic Mutations in Cancer
DNA	Deoxyribonucleic Acid
VCF	Genome Variant Call Format
HGMD	Human Gene Mutation Database
HUGO	HUGO Gene Nomenclature Committee
MAF	Mutation Annotation Format
ODB	Omics Data Bank
OHSCE	Oracle Health Sciences Cohort Explorer
RNA	Ribonucleic Acid
SNV	Single-nucleotide Variation
TRC	Translation Research Center
VCF	Variant Call Format

4 Disclaimer Regarding Third Party Data

Oracle makes no express or implied warranty, including but not limited to warranties regarding the accuracy, completeness, merchantability, or fitness for a particular purpose, with respect to third party data loaded into this application or the results of

any functions of the application using such data. It may be used for information purposes only, and no medical, clinical or other health related decisions may be based upon such results. You are solely responsible for your use of the third party data, including your right to use the data for your purposes.

5 Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at

<http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

Oracle Health Sciences Omics Data Bank Release Content Document, Release 2.5
E41335-01

Copyright © 2013, Oracle and/or its affiliates. All rights reserved.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

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.

This software or hardware and documentation may provide access to or information on content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.