An open approach to solving the challenges and complexity of the genomic workflow

Cancer is complex. Each patient is unique. Medical information is growing explosively, but not readily usable in clinical settings due to healthcare data silos. The current approach to precision medicine as a multi-disciplinary endeavor requires sophisticated technologies and high levels of expertise from bioinformaticians, pathologists, radiologists, and oncologists. Philips Genomics solutions are designed to solve this challenge; it is a secure cloud-based software solution that enables end-to-end clinical workflows by integrating genomic data, along with disease histology and patient phenotype, across a healthcare system to help provide a holistic view of the patient. It protects patient data, streamlines collaboration, supplies a comprehensive molecular picture, and delivers actionable reports for a better path towards precision care.

The Philips Genomics solutions are SaaS applications built on Philips HealthSuite Digital Platform. It supports a genome analytics platform on which bioinformatics pipelines are configured for NGS or other molecular test data processing with embedded quality control. It provides an intuitive web-based user interface to expose automated clinical workflows for clinicians: to select relevant genomic aberrations; prioritize evidence-driven therapies; match patients with biomarker-specific clinical trials; and generate an actionable clinical report. It has HIPAA certification from Coalfire Systems, Inc. and enables comprehensive patient data security and privacy.

Highlights
- **End-to-end clinical workflows**
- **Open pipeline architecture**
- **Multi-tenant cloud base SaaS built on Philips HealthSuite Digital Platform**

Abbreviation	Definition
SaaS	Software as a Service
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
HSDP	HealthSuite Digital Platform
REST API	Representational State Transfer Application Programming Interface
EMR	Electronic Medical Record
SNV	Single Nucleotide Variant
NGS	Next-generation sequencing
FASTQ	It is a text-based file format for storing both a nucleotide sequence and its corresponding quality score. It has become the standard for storing the output of high-throughput sequencing instruments.
BAM	Binary Alignment Map. It is a compressed format to store sequence alignment data.
VCF	Variant Call Format. It specifies the format of a text file for storing gene sequence variations.
PHI	Protected Health Information
LIS	Laboratory Information System. A software system that records, manages, and stores test data for clinical laboratories.
LIMS	Laboratory Information Management System. A software system that tracks and manages samples and associated data.

IntelliSpace Precision Medicine Genomics Overview

Philips IntelliSpace Precision Medicine Genomics processes the raw genomic data from the sequencer, or vcf file from a commercial genomics laboratory, bringing the comprehensive molecular patient data alongside disease histology and patient phenotype for a complete biomarker-informed diagnostic and therapeutic picture. Our solution can ingest commercial and internal NGS test reports to deliver a personalized therapy plan with literature references for each biomarker in the context of the disease sub type. Through exclusive partnerships with MD Anderson and automated access to the Nof1 and Jackson Laboratory clinical knowledge bases, clinicians can easily identify and prioritize therapies and potential clinical trials.

Because of Philips robust data model, we are able to provide a structured discrete data set for exported advanced analytics. All of this feeds into our molecular tumor board with the goal of creating a standardized, holistic genomic view of the patient.

IntelliSpace Precision Medicine Genomics workflow features

Genomics clinical reporting requires a combination of bioinformatics expertise, variant annotation, and interpretation, capturing clinical evidence associated with variants as well as the clinical trials, to produce a high-quality report. IntelliSpace Precision Medicine Genomics has all these steps fully automated into a seamless workflow for the stakeholders: molecular pathologist, oncologist, geneticist, curator, a bioinformaticist, as shown in Figure 1. With the full integration and automation, our goal is to produce an efficient workflow and minimize the time that molecular lab personnel must spend on clinical reporting while staying entirely in control of the reporting process and while leveraging high quality interpretation resources both external and in-house. While each of the modules described in this section could be covered separately by different tools, it's important to keep in mind that the full end-to-end integration in IntelliSpace Precision Medicine Genomics enables provenance of the data transformations and that the system keeps track of the meta-data for all the versions of the tools and the knowledge bases.
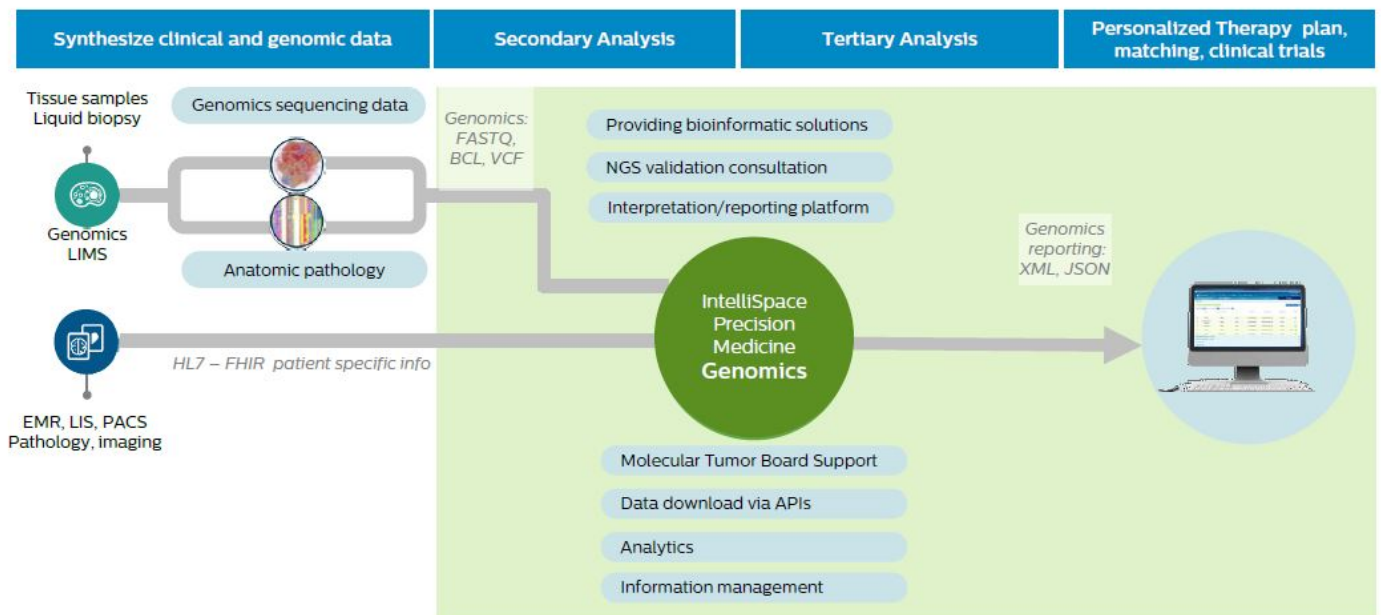

Figure 1. Example of IntelliSpace Precision Medicine Genomics for Pathology Workflow
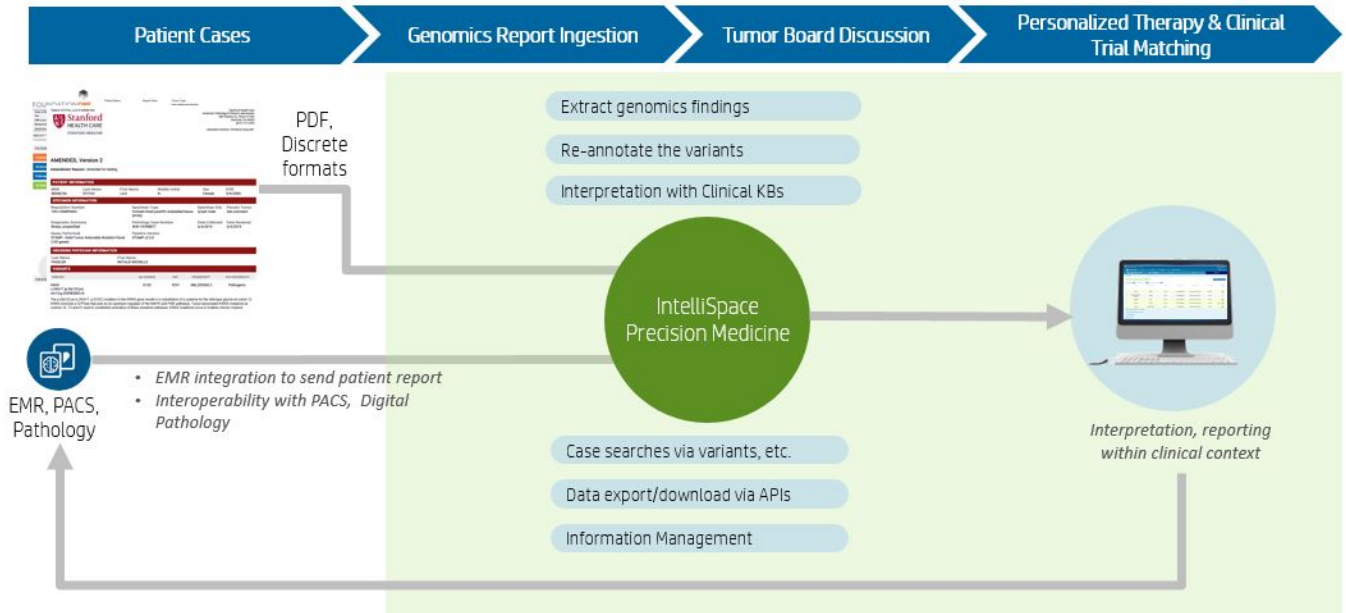
Figure 2. Example of IntelliSpace Precision Medicine Genomics for Oncology Workflow

Philips Intellispace Precision Medicine Genomics feature descriptions

Decode

Overview:

Bioinformatics pipeline processing is only the first and computationally most demanding step in the overall NGS process. As the field evolves, the tools and pipelines require highly specialized expertise and change, so there is the need to continuously improve upon existing computational workflows in a manner which also enables traceability and full transparency. We establish a way of working with the lab so that new versions of the pipelines are validated before entering production.

Details:

Our scalable 'genomic processor' platform provides the ability to define and design bioinformatics workflows. It also allows them to monitor the current processes that are being executed and alerts them to any errors that may result from the substandard input. While there are default pipelines and workflows predesigned in the system, the platform enables users to customize and create new pipelines by mix-and-match of tools. All the modifications are versioned and dated, so a pipeline upgrade is traceable and visible to the pathologist/geneticist who signs out the case. The platform hosts a broad diversity of publicly available genomics tools associated with pipelines and modules (C, C++, Java, Perl, Ruby, Python, and R) asynchronously in parallel and at scale:

- Provision and host the variety of jobs and their ecosystems dynamically on demand,
- Schedule and run these jobs as part of predefined workflows asynchronously,
- Pass parameters and results that are output between jobs,
- Check the status and progress of jobs, as well as cancel and retry specific jobs,
- Isolate jobs and associated data for each particular tenant

The platform has the versatility to handle various genomic data, including raw FASTQ, BCL formats, as well as VCF and other formats for SNV, CNV, fusions and gene expression for cancer and other conditions. It offers reproducibility, reliability, scalability, security, and traceability, built into medical device QMS standards.

Annotate

Overview:

This module provides annotation of identified genomics aberrations (SNV, CNV, insertions, deletions, fusions) according to functional, biological, and disease indication. It supports VCF format as well as customer defined file formats. The annotation sources include UCSC, dbNSFP, COSMIC, COSMIC Fusions, ClinVar, Ensembl VEP annotations, N-of-One therapy availability, and population frequencies such as 1000Genome and ExAC, and customer's annotation sources if desired. It presents the effect of the aberrations on genes, transcripts, protein sequence, as well as regulatory regions. Annotation sources are automatically updated whenever there is a new release. The annotation source version is stored as well for reproducibility. For aberrations that fail predefined QC criteria, a warning sign will be displayed along with the annotations to remind the clinicians during the review process. The module also provides the capability to initiate a curation request when the annotation for an aberration is not deemed to be up-to-date.

Details:

The annotation of an aberration is grouped into the following sections for review.

- Therapy-specific annotations: for example, details of the availability of therapy options in various phases from N-of-One. It may include approved therapies by FDA, experimental therapies, and resistance to any treatments. This information is sourced from the N-of-One aberration database.
- Genomic: chromosomal position, strand definition, aberration location (exon, 3'UTR, 5'UTR, etc.), transcript ID, reference and alternate allele/sequence. This information is based on UCSC gene annotation.
- Functional: aberration impact prediction scores, aberration type, and population frequency scores from ExAC and 1000Genomes.
- Clinical: COSMIC and ClinVar links to the respective databases.

Philips Intellispace Precision Medicine Genomics feature descriptions (cont.)

Review and filter

Overview:

This module provides the end user (e.g., clinician, pathologist, lab technologist, or bioinformaticist) comprehensive and flexible aberration filtering options, as the list of aberrations from a panel based NGS genomic test can be very long. The list is automatically filtered using the default filter configuration that can include tumor allele frequency, tumor read depth, ExAC allele frequency, and the availability for interpretation by N-of-One. The filter setting can be quickly cleared, modified, added, or removed. Commonly used filter options can be saved for fast recall in successive tests. The availability of filters depends on the configuration of IntelliSpace Precision Medicine Genomics at the customer site. Filtering parameters are organized in the following groups.

Quality:

Parameter	Explanation	Application
Normal allele frequency	Expected normal allele frequency	Mutation detection with tumor-normal sample
Tumor read depth	Required reads covering the mutation position	Mutation detection
Tumor allele frequency	Expected tumor allele frequency	Mutation detection
Fusion – DNA reads	Total read supporting the fusion	Fusion detection
RNA-Seq fpkm	Gene expression level in FPKM	RNA-Seq gene quantification

QC/BioInfo Filter	Additional QC parameters such as low depth of coverage, somatic caller, etc. It is an optional filter and depends on the file generated by the institute.	Mutation detection

Genomic Position:

Parameter	Explanation	Application
Chromosome	Chromosome where the aberration is found	Mutation detection
Location	Location of the aberrationposition	Mutation detection
Aberration type	Deletion, Insertion, Noncoding, Nonsynonymous, Stop gain, Stop loss, Splice site variant, synonymous	Mutation detection
Gene	HGNC gene nomenclature	Mutation detection
Fusion – gene 5'	5' Gene partner name	Fusion detection
Fusion – gene 3'	3' Gene partner name	Fusion detection
Fusion–Chromosome 5'	5' Gene chromosome	Fusion detection
Fusion–Chromosome 3'	3' Gene chromosome	Fusion detection

1000Gp3 allele frequency	1000 Genome phase 3 allele frequency	Mutation detection
dbSNP presence	Aberration reported in dbSNP	Mutation detection
COSMIC presence	Aberration reported in COSMIC	Mutation detection
ClinVar presence	Aberration reported in ClinVar	Mutation detection
ExAc allele frequency	Population frequency in ExAc database	Mutation detection
COSMIC match summary	Aberration reported in COSMIC matched on the basis of either amino acid change, genomic coordinate, nucleotide change for a gene	Mutation detection

Prediction:

Parameter	Explanation	Application

FATHMM	Predicting the functional effects of protein missense mutations using hidden Markov models (HMMs)	Mutation detection
SIFT	Uses sequence homology to predict whether a substitution affects protein function	Mutation detection
PolyPhen 2 HDIV	Uses eight sequence-based and three structure-based predictive features which were selected automatically by an iterative greedy algorithm. It uses HumDiv database	Mutation detection
PolyPhen 2 HVAR	Same as above, except it uses the HumVar database.	Mutation detection

Clinical: Parameter	Explanation	Application
ClinVar clinical significance	Filters on significance levels noted in ClinVar database	Mutation detection
Has approved therapies	Has FDA approved therapies for the gene for any disease condition based on N-of-One database	Mutation & fusion detection
Has experimental therapies	Has experimental therapies for the gene for any disease condition based on N-of-One database.	Mutation & fusion detection
Infers resistance to therapies	The gene is known to confer resistance to therapies based on N-of-One database	Mutation & fusion detection
Available for interpretation by N-of-One	The gene aberration is available for interpretation by N-of-One	Mutation & fusion detection

Philips Intellispace Precision Medicine Genomics feature descriptions (cont.)

Curate

Overview:

The curation module enables scientists and clinicians to apply their expertise and curate genomic aberrations to associate with biological features, as well as clinical significance: diagnostic, prognostic, therapeutic significance and related clinical trials in the context of disease states. Multiple types of genomic variants can be curated, including SNVs and indels, gene fusions and CNVs – both gene-based and region based. Numerous biological annotations are automatically populated, such as gene descriptions from NCBI, and other annotations that are available from our annotation module. Also, the user can access automated Google and PubMed searches. It provides the capability to curate genomic aberrations and their associated clinical significance from literature and in-house expertise. Multiple roles are defined to ensure the content quality. A contributor role can add content, and a curator role approves the content upon successful review.

Details:

The curation process can be initiated in two different workflows as shown in Figure 2: 1) by logging into the Curation module as a contributor or curator to manually create an entry of the aberration; or 2) by selecting and submitting variants for curation during the aberration review step by a pathologist within a clinical workflow.

In the curation workflow once the variant is approved it becomes available within the new version of the knowledgebase. The approved variants subsequently can be used for clinical interpretation. All revisions of variant curation are tracked, which allows the capability to roll back to an earlier version if desired.
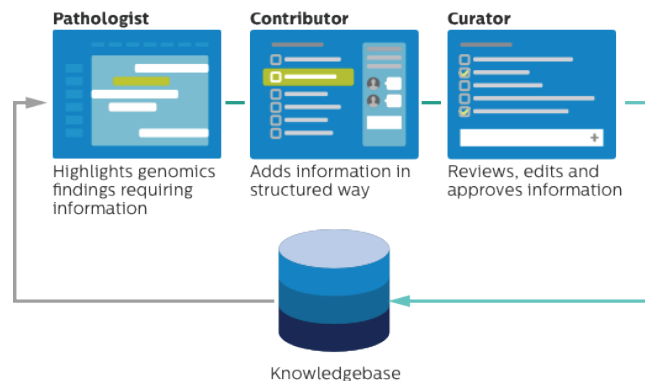

Figure 2. IntelliSpace Precision Medicine Genomics Curation Workflow

Therapy references

Overview:

This module provides automated therapy references based on a curated knowledge base and clinical trial listing inclusion and exclusion criteria through external knowledge bases (e.g., clinicaltrials.gov). It is executed as part of a clinical reporting workflow. After clinical expert review, prioritization, and selection, for each selected genomic aberration, the match microservice sends a query to the therapy and clinical trial knowledge bases, which can be externally available via business partnership (e.g., N-of-One) or internally available via the Curation knowledgebase.

Philips Intellispace Precision Medicine Genomics feature descriptions

Therapy references (continued)

Details:

Before the query is sent out to an external database via a secure and trusted channel, there is a de-identification process which removes all the PHI clinical data. Results are received in a predefined format and our system pools the results and remove the duplicates when necessary to provide an accurate and concise presentation to clinicians for clinical reporting or therapy prioritization. Therapy references and supporting information are listed in the following order:

- FDA-approved therapies that are within the indication schedule and run these jobs as part of predefined workflows asynchronously,
- FDA-approved therapies that are outside indication check the status and progress of jobs, as well as cancel and retry specific jobs,
- Investigational therapies that are within indication
- Resistance and interactions
- Clinical Guideline inclusion
- Treatment approach summary as well as detailed description.

Information about clinical trials, if available, is provided in addition to the therapies, with details of trials that are recruiting patients with specific indication and histological subtype and genomic profile. The trials presented are from the clinicaltrials.gov database, and the geographic context can be specified (e.g., only trials in the United States of America are listed). The trials are matched based on the gene, diagnosis, and age, and they are ordered based on distance from the hospital to the trial location.

Treatment prioritization

Overview:

Therapy reference and clinical trial listings provide many options for an individual patient, and it is up to the oncologists and pathologists to discuss and create a treatment strategy.

Details:

This module provides the option to collaborate on therapy prioritization based on Therapy Match results. It can be configured to include therapy prioritization from multiple oncologists before the final therapy plan is complete. Accompanying each prioritization can be a summary of the rationale behind the choices made by the oncologists or other clinical experts. The virtual molecular tumor boards allow designated users across institutions within a cancer center network to leverage expertise from multiple oncologists and pathologists.

Clinical reporting tool

Overview:

Report creation can be very time-consuming even after the interpretation results are available. This module provides pathologists and oncologists the capability of automatically generating and editing genomics reports after the therapy prioritization using the predefined report template including the organization's logo, header design, and patient-specific information. Also, it can configure the level and depth of information incorporated from therapy references and clinical trial listings.

Details:

The Clinical Reporting module supports reports in both Word™ and PDF formats. It also allows the pathologists and oncologists to modify specific fields and create an addendum report manually. Report files can be downloaded from the IntelliSpace Precision Medicine Genomics web interface, edited locally, and uploaded. Reports can also be shared with users with the proper access privilege within IntelliSpace Precision Medicine Genomics. The final report can be exported in a variety of formats and sent back to the EMR.

IntelliSpace Precision Medicine Genomics deployment

As a cloud-based solution, IntelliSpace Precision Medicine Genomics interfaces with the organization's existing EMR/LIS/LIMS systems for seamless interoperability after the initial integration. It is important to note that this kind of deployment requires no additional investment in computing and storage hardware.

IntelliSpace Precision Medicine Genomics modular architecture facilitates load balancing, disaster recovery, and auto-scaling. It provides the ability to integrate with existing EMR/LIS/LIMS within the organization. Figure 3 depicts the deployment scheme of IntelliSpace Precision Medicine Genomics along with the organization's systems and network. The primary user interface to IntelliSpace Precision Medicine Genomics is through a web browser, and there are two ways to trigger a new case: 1) The web application allows physicians to accession a new case and create a genomic test for a patient, and initiate the corresponding workflow; 2) IntelliSpace Precision Medicine Genomics, in addition, provides a programmatic API interface to the organization's EMR/LIS/LIMS to integrate patient clinical data and genomic tests. There is a one-time integration process required to establish the interfaces as depicted in Figure 3 before going into production. Once the test is requested from the EMR or LIS to IntelliSpace Precision Medicine Genomics, the corresponding test workflow will be automatically initiated and displayed for the clinicians. Secure access management governs both mechanisms to initiate the workflow in IntelliSpace Precision Medicine Genomics.

All function communications are implemented via an HTTPS connection. In addition to the web interface to the end users, IntelliSpace Precision Medicine Genomics also provides a REST API to integrate LIS and LIMS. IntelliSpace Precision Medicine Genomics also provides a File Uploader application on the NGS file server(s). It is responsible for sending NGS files to IntelliSpace Precision Medicine Genomics cloud storage. The File Uploader is installed and scheduled correctly in the local infrastructure. It can periodically scan the local storage server and initiate an upload task whenever it detects new files.

Highlights
- **Seamless interoperability with EMR/LIS/LIMS**
- **Load balancing, disaster recovery, and auto-scaling**
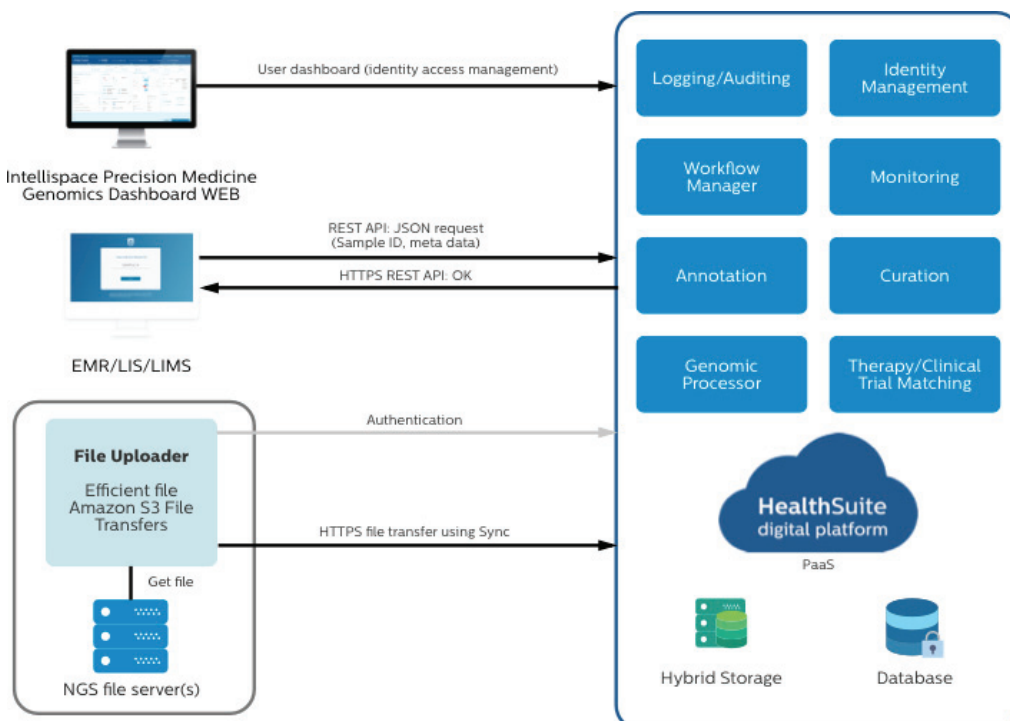- **Distributed software system with micro-service architecture**

Figure 3. Deployment diagram for IntelliSpace Precision Medicine Genomics

Leveraging a cloud-based architecture

IntelliSpace Precision Medicine Genomics is built on Philips HSDP to get advantages of the cloud deployment access and scalability that is essential for genomics. HSDP is an open, cloud-based platform that supports traditional and big data collection, integration, and analysis. It is designed to operate as a multi-tenant and scalable environment. The environment provides several differences from conventional hosting solutions by abstracting critical elements out of the "application" stack and making them available as part of the underlying platform. Client applications are deployed into containers that provide isolation from other applications running on the platform. These containers store application configuration, environment variables, and service credentials in an encrypted database table while also conforming to network traffic rules. The platform provides a mechanism for service provisioning and discovery using Service Brokers developed to integrate with the platform and deploy supporting services with

pre-defined security configurations. HSDP has been designed as an open multi-layer secure platform, optimized to interoperate in health enterprise ecosystems (Figure 4). Central to this system is the notion of microservices. Microservices are considered an extension of service-oriented architectures used to build distributed software systems. These processes communicate with each other over a network using lightweight protocols, with the benefits to enhance the cohesion and decrease coupling of software and to continuously add or drop services and refactor the system.
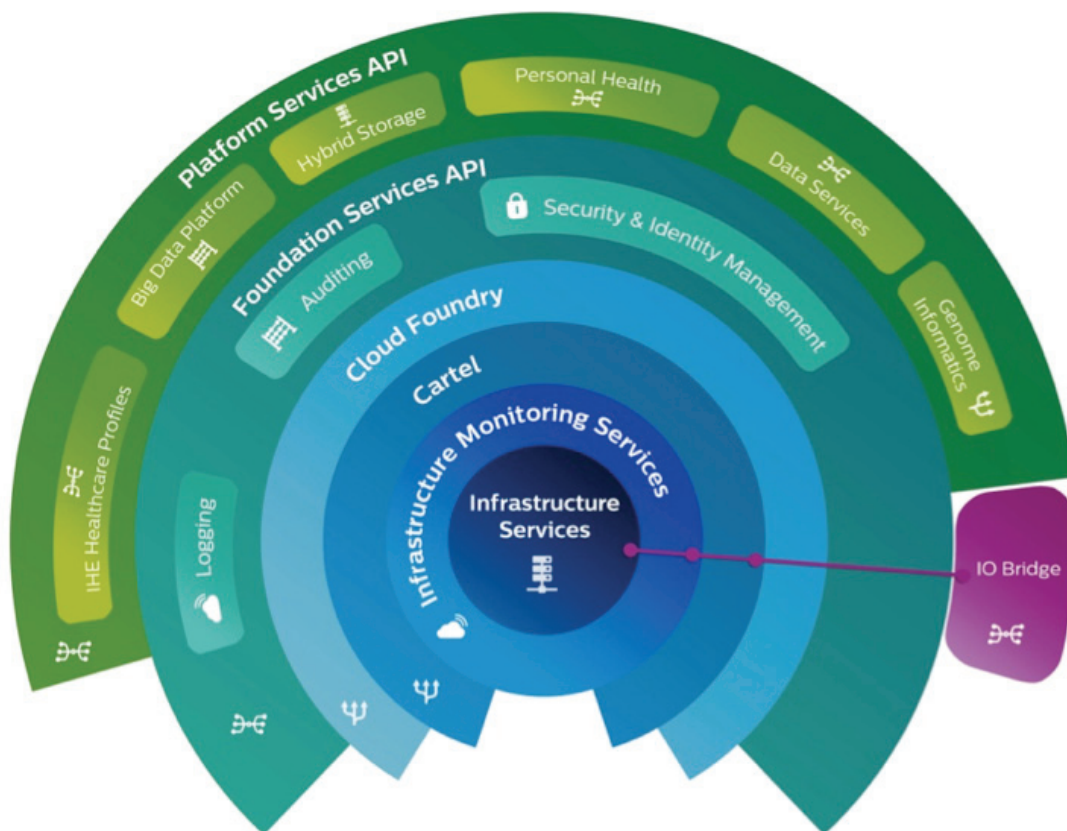

Figure 4. Layered microservices within Philips HealthSuite Digital Platform (HSDP)

Philips Healthsuite Digital Platform

Description

The core of HSDP has the IaaS layer, and its associated monitoring services are designed to host massively scalable processes, services, applications on demand and storing a large amount of data.

A second layer is a PaaS that consists of an elastic runtime self-service application execution component, coupled with an automation engine for application deployment and lifecycle management (Cloud Foundry) and an underlying provisioning API (Cartel) spinning up/down server instances. Containerization ensures that applications and services instances run in isolation without interference from other tenants.

HSDP foundation services API forms the next layer: logging, auditing, identity, and access management are essential cross-cutting services when implementing wellness, healthcare or bio-science services and applications.

The platform service API offers high-level big data and storage services as well as healthcare domain-specific capabilities, personal health, interoperability HL7/IHE integration profiles as well as genome informatics processing services.

As a secure cloud-based software solution, IntelliSpace Precision Medicine Genomics provides interoperability for bridging healthcare informatics gaps. It enables end-to-end clinical workflows by integrating disparate data across a healthcare system for accurate clinical interpretation and reporting.

References

1. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. **The Ensembl Variant Effect Predictor.** *Genome Biology* Jun 6;17(1):122. (2016)
2. Andry, F., Dimitrova, N., Mankovich, A., Agrawal, V., Bder, A. and David, A. **PAPAyA: A Highly Scalable Cloud-based Framework for Genomic Processing**. *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016),* Volume 3: BIOINFORMATICS: 198-206 (2016)
* Trademarks are the property of Koninklijke Philips N.V. (Royal Philips) or their respective owners.