

MolecularMatch Conditions 2018 Update

In 2013, MolecularMatch integrated SNOMED-CT as our technology's primary condition (disease) dataset. For months and years after, oncologists and scientists at MM added a layer of curation so that it fits with modern cancer diagnoses. This included editing and adding new conditions and synonyms, and creating genetic composite conditions. Since then, SNOMED has improved their oncology terms. Also, several open-source disease datasets have been launched; further improving how cancer diagnoses are described and coded.

Matching patients to clinical trials, targeted drugs, and assertion evidence guidelines depends on an accurate diagnoses and consistent interpretation between healthcare systems. Therefore in 2018, we upgraded our condition system, accomplishing the goals of:

1. Allowing cohesive code-based searching to improve EHR/Payer/LIMS integrations.
2. Updating to the most modern data ontologies.
3. Expanding our abilities outside-of-cancer.
4. Data validation and duplicates merging with pathologist oversight.

MolecularMatch NLP entity extraction and search engines run on ontologies. These include our conditions and findings, our in-house developed [molecular ontology](#) and our global clinical trial aggregation. As our experience with data aggregation has grown, we've developed a protocol that allows for easy integration of big data to enhance our search engines. This high-level protocol is:

1. Acquire datasets in their native form on frequent intervals.
2. Identification strategy and merging to allow diverse inputs and avoid duplicates.
3. Incorporate into MMPOWER for broad use.

Condition Data Incorporated into MM

Condition terms are now searchable by prefix_code (e.g. `SNOMEDID_255035007`).

Search Prefix	Dataset Name	Records Count
---------------	--------------	---------------

SNOMEDID	Systematized Nomenclature of Medicine -- Clinical Terms	313,940
ICD10	10th rev. of the International Statistical Classification of Diseases	94,127
DOID	DiseaseOntology -- open source medical vocabulary	12,498
ONCOTREE	OncoTree -- curated cancer type ontology by Memorial Sloan Kettering Cancer Center	627

Additional datasets possibly worth integrating are: **ICD9** and **LOINC**. DiseaseOntology contains cross mapping of terms to MeSH, ICD, NCI's thesaurus, SNOMED and OMIM.

All datasets can be easily updated on new releases. The current SNOMED-CT version is [v20170901 U.S. edition](#).

Dataset Mapping

We always design a feedback loop between clinical expertise and software automation. We do this by building easy to use, internal data management tools for training from M.D. and Ph.D. curators. This condition data upgrade required more than **100** M.D. pathologist curation hours.

Mapping efforts:

- ICD10 → SNOMED -- official map [file](#) implemented.
- OncoTree → SNOMED -- MolecularMatch pathologist manually mapped.
- DOID → SNOMED -- MolecularMatch pathologist manually mapped.

Screenshot of internal MM tools curators use to adjust mappings.

Data Management

Conditions

Clinical Findings

Anatomies

Stages

OncoTree

SNOMED

ICD10

DiseaseOntology

Processes

Invalid Only

OncoTree Set is **INVALID**

Search...

ID (OncoTree Code) ↑	Condition Name
<input type="checkbox"/> AA	Aggressive Angiomyxoma
<input type="checkbox"/> AASTR	Anaplastic Astrocytoma
<input type="checkbox"/> ACA	Adrenocortical Adenoma
<input type="checkbox"/> ACBC	Adenoid Cystic Breast Cancer
<input type="checkbox"/> ACC	Adrenocortical Carcinoma
<input type="checkbox"/> ACCC	Acinic Cell Carcinoma
<input type="checkbox"/> ACN	Acinar Cell Carcinoma, NOS
<input type="checkbox"/> ACPG	Craniopharyngioma, Adamantinomatous Type
<input type="checkbox"/> ACPP	Atypical Choroid Plexus Papilloma
<input type="checkbox"/> ACRM	Acral Melanoma
<input type="checkbox"/> ACYC	Adenoid Cystic Carcinoma
<input type="checkbox"/> ADMA	Adamantinoma
<input type="checkbox"/> ADNOS	Adenocarcinoma, NOS
<input type="checkbox"/> ADPA	Aggressive Digital Papillary Adenocarcinoma
<input type="checkbox"/> AECA	Sweat Gland Carcinoma/Apocrine Eccrine Carcinoma
<input type="checkbox"/> AFX	Atypical Fibroxanthoma

SNomed ID 416712009 is not in Condition table

ID: CCOV

Condition Name: Clear Cell Ovarian Cancer

MetaMainType: Ovarian Cancer

SNOMED IDs

+

-

SNOMED IDs	Custom
416712009	true

Parents from OncoTree

ID	Name
OVARY	OVARY
OVT	Ovarian Epithelial Tumor

«

<

Page 1 of 32

>

»

↺

Displaying 1 - 20 of 632

Revert

Save

Custom Conditions and Genetic Composites

Only a handful of genetic composite conditions exist in the public datasets, like "EGFR positive NSCLC". At MM, we've created many more composites, which are necessary in our NLP engine for coalescing meaning from medical documents.

Screenshot of internal MM tools for creating composite conditions.

Data Management

Conditions

Clinical Findings

Anatomies

Stages

OncoTree

SNOMED

ICD10

DiseaseOntology

Processes

+

-

Suppression

All Fields

Exact Match

Hide Suppressed

Composites

Search...

Name ↑	Composite	Custom	Suppress
<input type="checkbox"/> BCL-ABL1			
<input type="checkbox"/> B lymphoblastic leukaemia/lymphoma with TEL-AML1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> B lymphoblastic leukaemia/lymphoma with t(1;19)(q23;p13.3)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> B lymphoblastic leukaemia/lymphoma with t(5;14)(q31;q32)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> B lymphoblastic leukaemia/lymphoma with t(v;11q23); MLL rearranged	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> BRAF mutated colorectal cancer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> BRAF mutated melanoma	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Carney's Triad	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Childhood myelodysplastic syndrome	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Classical Hodgkin lymphoma type PTL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> EGFR mutated GBM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/> ER+ HER2- PIK3CA Breast cancer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> FLT3 AML	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> GIST - NF1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> GIST KIT exon 11	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> GIST KIT exon 13	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

ID: d470c5e0-4154-43c4-a842-625464745efa

Name: ER+ HER2- PIK3CA Breast cancer

Alias: ER+ HER2- PIK3CA Breast cancer

Clear Alias

Composite: ☒

Associated Tags (these tags get created in addition to this domain tag when this rec...)

Term	Custom	Suppress
PIK3CA	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Neoplasm of breast	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ERBB2 Loss	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ESR1	<input checked="" type="checkbox"/>	<input type="checkbox"/>

«

<

Page 1 of 3

>

»

↺

Displaying 1 - 25 of 72

Revert

Save

Lets Use It! (**warning: not available on production yet**)

These codes can now be used in the whole MM ecosystem. From the application search, to API queries, to EHR integration.

App Search

- http://app.molecularmatch.com/search/ICD10_D45
- http://app.molecularmatch.com/search/SNOMEDID_255035007
- http://app.molecularmatch.com/search/DOID_3950
- http://app.molecularmatch.com/search/ONCOTREE_AML

API Queries

See specs on api.molecularmatch.com

```
#####
# Trials Search -- save this as a .sh file
#####

# to run include apiKey as first argument
# $ chmod 777 file.sh
# $ ./file.sh apiKey

curl -X POST 'https://api.molecularmatch.com/v2/search/trials' \
--data "apiKey=$1" \
--data-urlencode 'filters=[ \
    {"facet":"ICD10","term":"ICD10_D45"}, \
    {"facet":"STATUS","term":"Enrolling"}, \
    {"facet":"TRIALTYPE","term":"Interventional"} \
  ]'

curl 'https://api.molecularmatch.com/v2/search/trials' \
--data "apiKey=$1" \
--data-urlencode 'filters=[ {"facet" : "SNOMEDID", "term" : "SNOMEDID_254626006"}]'
```

```
#####
# Condition Search -- used to normalize conditions
# If you have a condition name, or code,
# you can search our conditions table to find the best match.
# This can then be used as a term in subsequent trials and drugs searches.
#####

curl -X POST 'https://api.molecularmatch.com/v2/search/conditions' \
--data "apiKey=$1" \
--data-urlencode 'filters=[{"facet":"PHRASE","term":"Lung cancer"}]'
```



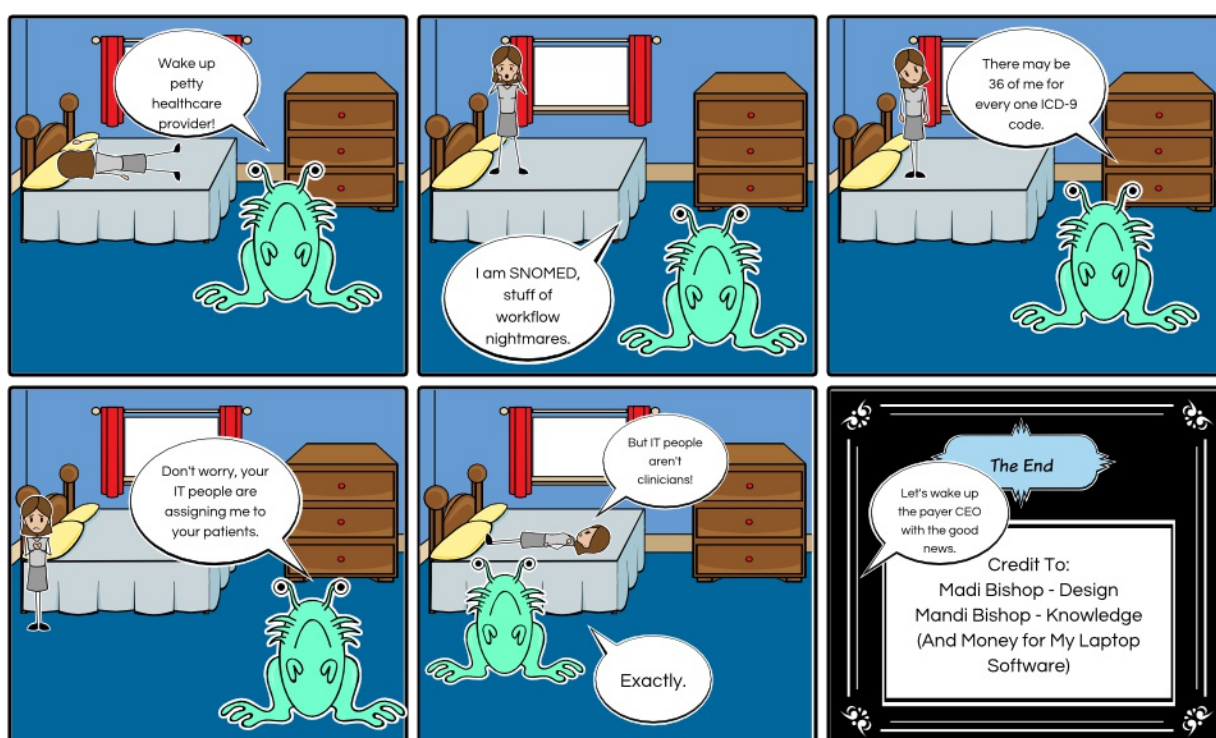
```
curl 'https://api.molecularmatch.com/v2/search/trials' \
--data "apiKey=$1" \
--data-urlencode 'filters=[ {"facet" : "SNOMEDID", "term" : "SNOMEDID_254626006"}]'
```

EHR Integration

Contact Us at info@molecularmatch.com for a demonstration of how to integrate with an EHR, payer system or LIMS.

Mapping Trouble?

No mapping effort is perfect. Work with us to make sure you are getting the results you expect for your patients.



Create your own at StoryboardThat.com

From: <http://healthstandards.com/blog/2014/04/21/snomed-problems/>

published 04/10/2018

MolecularMatch Inc.
2450 Holcombe Blvd.
TMC X+ | TMC Innovation Institute
Houston, TX 77021

713.357.1042
info@molecularmatch.com
www.molecularmatch.com