MEDomics
Towards Self-Cognizant Hospitals in the Treatment of Cancer

Olivier Morin, PhD
Chief of Physics, Interim
University California San Francisco, CA, USA

Joint AAPM | COMP Annual Meeting
July 15th, 2020

www.medomics.ai
Lab Efforts
Main Collaborators

UCSF RadOnc:
• Catherine Park, MD
• Jorge Barrios, PhD
• Taman Upadhaya, PhD
• Steve Braunstein, MD PhD
• Sue Yom, MD
• David Raleigh, MD PhD
• Joe Hsu, MD
• Gilmer Valdes, PhD
• Jean Nakamura, MD
• Jason Chan, MD
• Penny Sneed, MD
• Lijun Ma, PhD
• Benjamin Ziemer, PhD

UCSF Radiology/surgery:
• Javier Villanueva-Meyer MD
• Spencer Behr, MD
• Janine Luppo, PhD
• Antonio Carlos Westphalen, MD
• Michael McDermott, MD

McGill/Sherbrooke:
• Martin Vallières, PhD
• Jan Seuntjens, PhD

Dresden, Germany
• Alexander Zwanenburg, PhD
• Steffen Lock, PhD

D-Lab/Maastricht U:
• Philippe Lambin, MD
• Simon Keek, PhD student
• Henry Woodruff, PhD
• Abdalla Ibrahim, MD PhD
• Avishek Chatterjee, PhD

USF (data science):
• Yannet Interian, PhD
• Jeremy Howard, PhD
Point-of-Care Opportunities

- Many statistical models have been developed.
- Few have been integrated in the clinic.
- My lab is studying various point-of-care interventions using informatics.
Oncology

Today
- Increasingly digital, not always accessible, not centralized
- Single shot research
- Transactional
- Advances driven mainly by clinical trials
- Complex, for all
- Frustrating, for patients

Future
- Fully digital and accessible
- Technologies employed for data centralization and governance
- Advances driven by clinical trials but also influenced by real-world data
- Complexity will be increasingly hidden
- Quality will be assessed in real-time
- Participatory, data ownership?

Where do we start?
Self-Cognizant Hospitals

having knowledge and being aware of itself and its goals

• Have clear rules on the data needed for each medical intervention and decision.

• Missing data will be identified and collected.

• Data quality will be assessed and corrected.

• Data will be synthesized (MEDomics).

• Hospital value/cost and performance (patient quality of life) will be measured and compared to regional/national/international trends.

MEDomics Animation: https://youtu.be/2030Pdgm3_4
Meningioma example

Presentation: Demographic
Imaging: Radiographic Radiomic Volume
Surgery/Radiation: Therapy
Pathology: Pathology Grade

Prognostic models

Local failure
Overall survival

306 patients with comprehensive clinical, radiological, radiomics, molecular and outcome data
# Radiomics Baseline

## Model Analysis Module

<table>
<thead>
<tr>
<th>Model</th>
<th>Meningioma</th>
<th>Overall Survival</th>
<th>Random Forest</th>
<th>Test</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>0.56</td>
<td>0.59</td>
<td>0.56</td>
<td>0.58</td>
<td>0.43</td>
</tr>
<tr>
<td>Sphericity</td>
<td>0.66</td>
<td>0.77</td>
<td>0.60</td>
<td>0.81</td>
<td>NaN</td>
</tr>
<tr>
<td>Volume_Sphericity</td>
<td>0.63</td>
<td>0.78</td>
<td>0.57</td>
<td>0.72</td>
<td>0.43</td>
</tr>
<tr>
<td>Radiomics basics</td>
<td>0.68</td>
<td>0.58</td>
<td>0.70</td>
<td>0.63</td>
<td>0.66</td>
</tr>
<tr>
<td>Radiomics_full</td>
<td>0.77</td>
<td>0.82</td>
<td>0.76</td>
<td>0.84</td>
<td>0.69</td>
</tr>
</tbody>
</table>

**Test Sets:**
- test_random || All, n=74 (3:1)
- test_random || PMH, n=20 (19:1)
- test_random || UCSF, n=54 (2:1)
- test_PMH, n=62 (15:1)
- test_UCSF, n=163 (2:1)
We need data!

1000 object classes
1.2 M train
100k test
We need diverse data with labels!
Data Flow

EHR SYSTEM (ApeX/Epic)

User request entries

Server 1
Server 2
...
Server N

Backup every 24 hrs

CLARITY

16 custom reports per day

UCSF Registry

PACS

MOSAIQ

MIM

De-identification (PHILiter)

MEDomics

PostgreSQL

MIM

Updated every 24 hrs

Find immunotherapy drug hits

```sql
select qipm.medication.mrn, qipm.medication.medication_startdate, qipm.medication.medication_name, qipm.medication.medication_route from qipm.medication
join qipm.medication on qipm.medication.mrn = qipm.medication.mrn
where qipm.medication.medication_name ilike '%Pepilimumab%'
or qipm.medication.medication_name ilike '%Nivolumab%'
or qipm.medication.medication_name ilike '%Pembrolizumab%'
or qipm.medication.medication_name ilike '%Atezolizumab%'
or qipm.medication.medication_name ilike '%Avelumab%'
or qipm.medication.medication_name ilike '%Durvalumab%'
```

Success
2799 rows

<table>
<thead>
<tr>
<th>mrn</th>
<th>medication_startdate</th>
<th>medication_name</th>
<th>medication_route</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-03-14</td>
<td>2018-03-14</td>
<td>NIVOLUMAB 100 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2018-02-21</td>
<td>2018-02-21</td>
<td>PEMBROLIZUMAB 50 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2017-12-27</td>
<td>2017-12-27</td>
<td>PEMBROLIZUMAB 50 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2017-09-27</td>
<td>2017-09-27</td>
<td>NIVOLUMAB 100 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2017-09-19</td>
<td>2017-09-19</td>
<td>PEMBROLIZUMAB 50 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2017-09-27</td>
<td>2017-09-27</td>
<td>NIVOLUMAB 100 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2018-01-26</td>
<td>2018-01-26</td>
<td>NIVOLUMAB 100 ML IVPB</td>
<td>Intravenous</td>
</tr>
<tr>
<td>2017-09-19</td>
<td>2017-09-19</td>
<td>PEMBROLIZUMAB 50 ML IVPB</td>
<td>Intravenous</td>
</tr>
</tbody>
</table>
### Careful Patient Selection

#### Table A

<table>
<thead>
<tr>
<th>Feature</th>
<th>Total # of days with notes</th>
<th>% of patients with notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at Dx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical Margin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erasmus stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3N3M stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional node status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laterality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targeted drug</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Completeness of features [%]

- Age at Dx: 100%
- Race: 100%
- Surgical Margin: 98.5%
- Smoking status: 100%
- Depression: 99%
- Hypertension: 100%
- Overall survival: 100%
- Vital status: 100%
- Living distance: 100%
- Mental status: 100%
- Living region: 100%
- Erasmus stage: 100%
- P3N3M stage: 100%
- Regional node status: 100%
- Tumor size: 100%
- Laterality: 100%
- Targeted drug: 100%

---

#### Diagram B

The diagram illustrates the completeness of various features in the dataset, with percentages ranging from 100% for all features.

#### Diagram C

The C diagram shows a scatter plot with patients and drug targeting, highlighting the correlation between patient selection and targeted drug treatment.
Lung Cancer

Staging

- Stage 0-2, n=981
- Stage 3, n=408
- Stage 4, n=1070

Surviving Fraction

Smoking Status

Framingham Risk

Immunotherapy

Time in Years
MEDomics: preliminary results

Breast Cancer

Lung Cancer
Exploring Medical Notes

left breast
right breast
lymph node

cell carcinoma
squamous cell

RADIATION ONCOLOGY

Physical Exam

Department Radiation Oncology

Example of radiology extract:

Example of pathology extract:

UNIVERSITY OF CALIFORNIA SAN FRANCISCO DEPARTMENT OF PATHOLOGY 1600 DIVISADERO STREET Room: R-200, Box: 1785 SAN FRANCISCO, CA 94115-1785 TEL: (415) 885-7301 FAX: (415) 353-7676 CYTOPATHOLOGY REPORT Patient Name: XXXXXXXXX J. Med. Rec.#: XXXXXX DOB: XXXXXX (Age: 32) Sex: X Accession #: XXXXXXX Visit #: XXXXXX Service Date: 4/18/2018 Received: 4/19/2018 Location: SU8 Client: Parnassus Physician(s): XXXXXXXX ((415) 885-7788) &x20; Source of Specimen: Cervical, ThinPrep FINAL CYTOLOGIC INTERPRETATION/RESULT: A: Cervical, ThinPrep ATYPICAL SQUAMOUS CELLS OF UNDETERMINED SIGNIFICANCE (ASC-US). SPECIMEN ADEQUACY: Satisfactory for evaluation. Transformation zone components are present. COMMENTS: The specimen will be sent for high risk HPV testing with HPV16/18 genotyping. Please refer to the separate report for results. Clinical History Date of Last Menstrual Period: 4/7/2018 1 Thin Prep Pap. Co-test with reflex 16/18 genotyping in women 30 years and older, if Pap negative and high risk HPV is positive History Abnormal Pap?: No Treatment: Not Applicable Submitting Diagnosis: ICD-10-CM: Z12.4: Encounter for screening for malignant neoplasm of cervix Number of slides: 1 &x20; The Pap test is a screening test to aid in the detection of anogenital cancers and their precursors. Both false-negative and false-positive results have been experienced. The Pap should not be used as the sole means to detect anogenital cancers; regular periodic testing and follow-up of unexplained clinical signs and symptoms is suggested. &x20; XXXXXXXX/Pathologist Electronically signed out on 4/23/2018 17:2
Need Feature Extraction -> Pattern Recognition -> Deep learning
Training Pipeline and Modelling

10-fold Stratified Sampling

TF-IDF Vectorization

Random Over Sampling

Logistic Regression

Train Physician Notes

Test Physician Notes

fit

predict

Stop Words

Frequent Words

Rare Words

Occurrence (High)

Occurrence (Low)

Value (Low)

Value (High)

TF-IDF

0.50

1.00

1.50

2.00

2.50

3.00
Breast: Predicting 5 yrs Survival, Train 401/Test 102

- Trastuzumab
- Skeleton
- Tickening
- Herceptin
- Atypical
- Triple
- Enhancement
- SUV
- Progression
- Four

Glioma: Predicting 14 months Survival, Train 378/Test 107

- Anaplastic
- Residual
- Astrocytoma
- Avastin
- Multifocal
- Basal
- Worsening
- Ganglia
- Wildtype
- 1p19q

Number of Days of Medical Text Included in the Model after Diagnosis
In Summary

• The needs for medical informatics are diverse (administrative, QI, research).

• **Data governance and aggregation** should be planned carefully.

• We have created a MEDomics framework:
  
  • To explore current and new hypotheses from real-world data.
  
  • To develop novel algorithms and clinical tools.
  
  • To take a first step towards self-cognizant and responsive EHR/OIS.