Advanced development of an open-source platform for web-based integrative digital image analysis in cancer

David Gutman, MD, PhD
Assistant Professor
Department of Neurology
Emory University
U24 PI

Lee AD Cooper, PhD
Assistant Professor
Department of Biom. Inf.
Emory University
Georgia Institute of Technology
U24 PI

NIH
Informatics Technology for Cancer Research
U24CA194362
The (C)DSA: Developing a platform and infrastructure
Expertise in building open-source communities

Software process and project management. Design for long-term maintainability and extensibility.

Jonathan Beezley
User Interfaces

Zack Mullen
Workflow / DB

Dave Manthey
Visualization

Brian Helba
Project Manager

Deepak Chitajalu
HistomicsTK

Jonathan Beezley
User Interfaces

Zack Mullen
Workflow / DB

Dave Manthey
Visualization

http://www.kitware.com/
**Project Goals:**
Infrastructure + algorithms for the management, analysis and integration of digital pathology data

**Development philosophy:**
Installable, scalable, maintainable, extensible

**Open-source community development:**

https://github.com/DigitalSlideArchive

Project Started May 1st, 2016
Supporting Cancer Research

**Digital Slide Archive**
TCGA PanCancer Pathology Review (Alex Lazar)
TCGA Analysis Working Groups (All)
Lymphoma Epidemiology of Outcomes Cohort Study (Flowers)
ISIC Melanoma Working Group
Emory Winship Cancer Institute Biobank
Emory Molecular Pathology

**HistomicsTK**
TCGA PanCancer Heterogeneity & Evolution (Lazar, Getz)
TCGA Sarcoma AWG (Lazar)
Digital Slide Archive
Visualize Human and Algorithm Generated Results
Core Technologies

- Girder (Kitware Platform) for user management/permissions
- Python Based Backend
- MongoDB for Primary Database
- OpenSeadragon Image Viewer
- **Evaluating Docker, CWL and Workflow Tools**
(Not) Dealing with PHI
HistomicsTK
HistomicsTK = Infrastructure + Algorithms for Whole Slide Image Analysis

Infrastructure:
- User Interfaces
- APIs
- Extensibility
- Execution and resource mapping (Girder)
- Provenance and results management (Girder)
- Visualization
HistomicsTK = Infrastructure + Algorithms for Whole Slide Image Analysis

Infrastructure:

- User Interfaces
- APIs
- Extensibility
- Execution and resource mapping (Girder)
- Provenance and results management (Girder)
- Visualization
Step 1. Define Algorithm Interface

Using Slicer Execution Model XML Spec

Step 2. Write Algorithm Source

Python or C/C++

```
import histomicstk as htk
import ctk_cli

def run( args ):
    imInput = htk.read_image( args.inputImage )
    ...

if __name__ == "__main__":
    # cmd-line argument parsing and help with ctk_cli
    run( ctk_cli.CLIArgumentParser().parse_args() )
```

Step 3. List Algorithms in a JSON

```
{
    "NucleiDetection" : {"type": "python"},
    "CellClassification" : {"type": "c++"}
}
```

Step 4. Write Dockerfile For Containerization

(Generates REST end-points)

```
# Specify root docker image
FROM: dsarchive/histomicstk:v0.1.3

# Install system prerequisites
RUN apt-get update && \
    apt-get install -y wget git python
...

# Copy source code into docker image
COPY .
...

# Install dependencies
RUN pip install -r requirements.txt
...

# Use entry-point provided by HistomicsTK
ENTRYPOINT ["python", "cli_list_entrypoint.py"]
```

**GitHub Auto-build w/ upload to DockerHub**
Performs Adaptive Color Deconvolution

Uses sparse non-negative matrix factorization to adaptively deconvolve a given RGB image into intensity images representing distinct stains.

Jonathan Beezley
## HistomicsTK Algorithms

### Transformations and Color Normalization
- ColorConvolution
- ColorDeconvolution
- ComplementStainMatrix
- OpticalDensityFwd
- OpticalDensityInv
- ReinhardNorm
- ReinhardSample
- RudermanLABFwd
- RudermanLABInv
- SparseColorDeconvolution

### Segmentation
- ChanVese
- DregEdge
- GaussianVoting
- GradientFlow
- MaxClustering
- MergeSinks
- SimpleMask

### Filtering
- Del2
- EstimateVariance
- GaussianGradient
- GradientDiffusion
- cLoG
gLoG

### Utilities
- CondenseLabel
- ConvertSchedule
- EmbedBounds
- FilterLabel
- GraphColorSequential
- MergeColinear
- RegionAdjacencyLayer
- Sample
- ShuffleLabel
- SubmitTorque
- TilingSchedule

---

> git clone https://github.com/DigitalSlideArchive/HistomicsTK
Video

https://youtu.be/bWv-XrTE5Qc
Acknowledgement

David Gutman (co-PI)
Michael Nalisnik
Sanghoon Lee
Jun Kong
Pooya Mobadersany
Mohammed Khalilia

Deepak Chittajallu
Jonathan Beezley
Dave Manthey
Brian Helba
Zach Mullen
Charles Law

Clinical Collaborators:
Alexander Lazar
(MD Anderson)
Daniel Brat (Emory)
Christopher Flowers (Emory)
Brian Pollack (Emory)

Funding:
NCI U24CA194362
NLM K22LM011576