Large Language Models to understand biomedical text

Yuan Luo, PhD, FIAHSI, FAMIA
Chief AI Officer
Northwestern University Clinical and Translational Sciences Institute
Institute for Augmented Intelligence in Medicine
Associate Professor
Department of Preventive Medicine
Northwestern University
yuan.luo@northwestern.edu
@yuanhypnosluo

9/18/2023
Large Language Models – oversimplified review

Roadmap

• Communication
  – ChatGPT in medical and scientific writing

• Resource
  – BERT in building knowledge graphs

• Practice
  – Clinical-Longformer and Clinical-BigBird

• Surveillance
  – ChatGPT in pharmacovigilance
Study design

Prompt: ‘Please write a scientific abstract for the article [title] in the style of [journal] at [link]’

**JAMA**
- Importance
- Objective
- Design, Setting, and Participants
- Interventions
- Main Outcomes and Measures
- Results
- Conclusions and Relevance

**The BMJ (variable headers depending on study type)**
- Objective
- Design
- Data source
- Setting
- Participants
- Data extraction and synthesis
- Main outcome measures
- Results
- Conclusions

**The NEJM**
- Background
- Methods
- Results
- Conclusions

**Nature Medicine** - headerless paragraph abstract
- Findings
- Interpretation
Generated abstracts have a similar patient cohort size as original abstracts
Many generated abstracts can be detected using an AI output detector.
Generated abstracts are original and do not plagiarize from other written work
Reviewers use criteria different than the AI output detector for flagging abstracts

Reviewers were able to correctly identify 68% of generated abstracts as being generated by ChatGPT, but incorrectly identified 14% of original abstracts as being generated.

Roadmap

• Communication
  – ChatGPT in medical and scientific writing

• Resource
  – BERT in building knowledge graphs

• Practice
  – Clinical-Longformer and Clinical-BigBird

• Surveillance
  – ChatGPT in pharmacovigilance
Fine-tuning KG-BERT for predicting triples, links, and relations

Predicting the plausibility of a triple or a link

Triplet Label $y \in \{0, 1\}$

$$\mathcal{L} = -\sum_{\tau \in \mathbb{D}_+ \cup \mathbb{D}_-} (y_\tau \log(s_{\tau 0}) + (1 - y_\tau) \log(s_{\tau 1}))$$

Predicting relations

Relation Label $y \in \{1, \ldots, R\}$

$$\mathcal{L}' = -\sum_{\tau \in \mathbb{D}_+} \sum_{i=1}^{R} y'_{\tau i} \log(s'_{\tau i})$$
### Summary statistics of datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Ent</th>
<th># Rel</th>
<th># Train</th>
<th># Dev</th>
<th># Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN11</td>
<td>38,696</td>
<td>11</td>
<td>112,581</td>
<td>2,609</td>
<td>10,544</td>
</tr>
<tr>
<td>FB13</td>
<td>75,043</td>
<td>13</td>
<td>316,232</td>
<td>5,908</td>
<td>23,733</td>
</tr>
<tr>
<td>WN18RR</td>
<td>40,943</td>
<td>11</td>
<td>86,835</td>
<td>3,034</td>
<td>3,134</td>
</tr>
<tr>
<td>FB15K</td>
<td>14,951</td>
<td>1,345</td>
<td>483,142</td>
<td>50,000</td>
<td>59,071</td>
</tr>
<tr>
<td>FB15k-237</td>
<td>14,541</td>
<td>237</td>
<td>272,115</td>
<td>17,535</td>
<td>20,466</td>
</tr>
<tr>
<td>UMLs</td>
<td>135</td>
<td>46</td>
<td>5,216</td>
<td>652</td>
<td>661</td>
</tr>
</tbody>
</table>
## Triple classification accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>WN11</th>
<th>FB13</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTN (Socher et al. 2013)</td>
<td>86.2</td>
<td>90.0</td>
<td>88.1</td>
</tr>
<tr>
<td>TransE (Wang et al. 2014b)</td>
<td>75.9</td>
<td>81.5</td>
<td>78.7</td>
</tr>
<tr>
<td>TransH (Wang et al. 2014b)</td>
<td>78.8</td>
<td>83.3</td>
<td>81.1</td>
</tr>
<tr>
<td>TransR (Lin et al. 2015b)</td>
<td>85.9</td>
<td>82.5</td>
<td>84.2</td>
</tr>
<tr>
<td>TransD (Ji et al. 2015)</td>
<td>86.4</td>
<td>89.1</td>
<td>87.8</td>
</tr>
<tr>
<td>TEKE (Wang and Li 2016)</td>
<td>86.1</td>
<td>84.2</td>
<td>85.2</td>
</tr>
<tr>
<td>TransG (Xiao, Huang, and Zhu 2016)</td>
<td>87.4</td>
<td>87.3</td>
<td>87.4</td>
</tr>
<tr>
<td>TranSparse-S (Ji et al. 2016)</td>
<td>86.4</td>
<td>88.2</td>
<td>87.3</td>
</tr>
<tr>
<td>DistMult (Zhang et al. 2018)</td>
<td>87.1</td>
<td>86.2</td>
<td>86.7</td>
</tr>
<tr>
<td>DistMult-HRS (Zhang et al. 2018)</td>
<td>88.9</td>
<td>89.0</td>
<td>89.0</td>
</tr>
<tr>
<td>AATE (An et al. 2018)</td>
<td>88.0</td>
<td>87.2</td>
<td>87.6</td>
</tr>
<tr>
<td>ConvKB (Nguyen et al. 2018a)</td>
<td>87.6</td>
<td>88.8</td>
<td>88.2</td>
</tr>
<tr>
<td>DOLORES (Wang, Kulkarni, and Wang 2018)</td>
<td>87.5</td>
<td>89.3</td>
<td>88.4</td>
</tr>
<tr>
<td>KG-BERT(a)</td>
<td><strong>93.5</strong></td>
<td><strong>90.4</strong></td>
<td><strong>91.9</strong></td>
</tr>
</tbody>
</table>
Test accuracy of triple classification by varying training data proportions

WN11

FB13
## Link prediction results

<table>
<thead>
<tr>
<th>Method</th>
<th>WN18RR</th>
<th></th>
<th>FB15k-237</th>
<th></th>
<th>UMLS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MR</td>
<td>Hits@10</td>
<td>MR</td>
<td>Hits@10</td>
<td>MR</td>
<td>Hits@10</td>
</tr>
<tr>
<td>TransE (our results)</td>
<td>2365</td>
<td>50.5</td>
<td>223</td>
<td>47.4</td>
<td>1.84</td>
<td>98.9</td>
</tr>
<tr>
<td>TransH (our results)</td>
<td>2524</td>
<td>50.3</td>
<td>255</td>
<td>48.6</td>
<td>1.80</td>
<td>99.5</td>
</tr>
<tr>
<td>TransR (our results)</td>
<td>3166</td>
<td>50.7</td>
<td>237</td>
<td>51.1</td>
<td>1.81</td>
<td>99.4</td>
</tr>
<tr>
<td>TransD (our results)</td>
<td>2768</td>
<td>50.7</td>
<td>246</td>
<td>48.4</td>
<td>1.71</td>
<td>99.3</td>
</tr>
<tr>
<td>DistMult (our results)</td>
<td>3704</td>
<td>47.7</td>
<td>411</td>
<td>41.9</td>
<td>5.52</td>
<td>84.6</td>
</tr>
<tr>
<td>ComplEx (our results)</td>
<td>3921</td>
<td>48.3</td>
<td>508</td>
<td>43.4</td>
<td>2.59</td>
<td>96.7</td>
</tr>
<tr>
<td>ConvE (Dettmers et al. 2018)</td>
<td>5277</td>
<td>48</td>
<td>246</td>
<td>49.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ConvKB (Nguyen et al. 2018a)</td>
<td>2554</td>
<td>52.5</td>
<td>257</td>
<td>51.7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>R-GCN (Schlichtkrull et al. 2018)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>41.7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>KBGAN (Cai and Wang 2018)</td>
<td>–</td>
<td>48.1</td>
<td>–</td>
<td>45.8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RotatE (Sun et al. 2019)</td>
<td>3340</td>
<td><strong>57.1</strong></td>
<td>177</td>
<td><strong>53.3</strong></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>KG-BERT(a)</td>
<td><strong>97</strong></td>
<td>52.4</td>
<td><strong>153</strong></td>
<td>42.0</td>
<td><strong>1.47</strong></td>
<td>99.0</td>
</tr>
</tbody>
</table>
## Relation prediction results

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Rank</th>
<th>Hits@1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TransE (Lin et al. 2015a)</td>
<td>2.5</td>
<td>84.3</td>
</tr>
<tr>
<td>TransR (Xie, Liu, and Sun 2016)</td>
<td>2.1</td>
<td>91.6</td>
</tr>
<tr>
<td>DKRL (CNN) (Xie et al. 2016)</td>
<td>2.5</td>
<td>89.0</td>
</tr>
<tr>
<td>DKRL (CNN) + TransE (Xie et al. 2016)</td>
<td>2.0</td>
<td>90.8</td>
</tr>
<tr>
<td>DKRL (CBOW) (Xie et al. 2016)</td>
<td>2.5</td>
<td>82.7</td>
</tr>
<tr>
<td>TKRL (RHE) (Xie, Liu, and Sun 2016)</td>
<td>1.7</td>
<td>92.8</td>
</tr>
<tr>
<td>TKRL (RHE) (Xie, Liu, and Sun 2016)</td>
<td>1.8</td>
<td>92.5</td>
</tr>
<tr>
<td>PTransE (ADD, len-2 path) (Lin et al. 2015a)</td>
<td><strong>1.2</strong></td>
<td>93.6</td>
</tr>
<tr>
<td>PTransE (RNN, len-2 path) (Lin et al. 2015a)</td>
<td>1.4</td>
<td>93.2</td>
</tr>
<tr>
<td>PTransE (ADD, len-3 path) (Lin et al. 2015a)</td>
<td>1.4</td>
<td>94.0</td>
</tr>
<tr>
<td>SSP (Xiao et al. 2017)</td>
<td><strong>1.2</strong></td>
<td>–</td>
</tr>
<tr>
<td>ProjE (pointwise) (Shi and Weninger 2017)</td>
<td>1.3</td>
<td>95.6</td>
</tr>
<tr>
<td>ProjE (listwise) (Shi and Weninger 2017)</td>
<td><strong>1.2</strong></td>
<td>95.7</td>
</tr>
<tr>
<td>ProjE (wlistwise) (Shi and Weninger 2017)</td>
<td><strong>1.2</strong></td>
<td>95.6</td>
</tr>
<tr>
<td>KG-BERT (b)</td>
<td><strong>1.2</strong></td>
<td><strong>96.0</strong></td>
</tr>
</tbody>
</table>

Roadmap

• Communication
  – ChatGPT in medical and scientific writing

• Resource
  – BERT in building knowledge graphs

• Practice
  – Clinical-Longformer and Clinical-BigBird

• Surveillance
  – ChatGPT in pharmacovigilance
Background and motivation

• Transformer-based models, such as BERT, ClinicalBERT, and BioBERT, are designed to handle text inputs that are up to a maximum length of 512 tokens
• This limit poses a challenge for clinical texts, such as pathology reports, which tend to be much longer
• The Longformer and BigBird models extend the maximum input length from 512 tokens to 4,096 tokens by implementing sparse attention mechanisms. Both models have achieved great success in the general domain
• However, the adaptability of both models to the clinical contexts remain unclear
• To create two clinical knowledge-enriched language models, Clinical-Longformer and Clinical-BigBird, through pre-training on large-scale clinical notes
• To compare the effectiveness of Clinical-Longformer and Clinical-BigBird with short text models in improving the performance of various downstream clinical NLP tasks
# Large Language Models

**Long-sequence Models**

- **len(tokens) < 4096**
  - Longformer
  - BigBird

**Short-sequence Models**

- **len(tokens) < 512**
  - BERT

## Pre-trained Transformers

- Longformer
- BigBird

## Knowledge-enriching Corpora

- MIMIC III

## Knowledge-enriched Transformers

- Clinical-Longformer
- Clinical-BigBird

## Downstream Clinical NLP Tasks

- **Natural Language Inference**
  - medNLI

- **Named Entity Recognition**
  - i2b2 2006
  - i2b2 2010
  - i2b2 2012
  - i2b2 2014

- **Question Answering**
  - emrQA Relation
  - emrQA Medication
  - emrQA Heart Disease

- **Text Classification**
  - MIMIC AKI
  - openI

---

**Pre-Training Architecture** | **Average Precision** | **Area Under ROC**
--- | --- | ---
Clinical Longformer | 0.88 | 0.93
PubMedBERT | 0.79 | 0.86
LongformerBASE | 0.79 | 0.85
Bio+DischargeSummaryBERT | 0.78 | 0.85
SapBERT | 0.77 | 0.85
BERTBASE | 0.76 | 0.82

---

@JonWCunningham

Clinical Longformer by @yikuani18 was the best pre-training for our task.

Lesson: clinical pre-training and long attention window are key for medical NLP.

---

Jon Cunningham
@JonWCunningham

Thanks Faraz! Let's do it. We are grateful to the Clinical Longformer team. Starting to use this model partway through our process really improved the test set performance compared to clinical models without the longformer window.
### Results

#### Long-text Models

<table>
<thead>
<tr>
<th>Model</th>
<th>acc @ medNLI</th>
<th>F1 @ i2b2 2006</th>
<th>F1 @ i2b2 2010</th>
<th>F1 @ i2b2 2012</th>
<th>F1 @ i2b2 2014</th>
<th>F1 @ emrQA Med</th>
<th>F1 @ emrQA Rela</th>
<th>F1 @ emrQA CVD</th>
<th>acc @ openI</th>
<th>AUC @ MIMIC AKI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical-Longformer</td>
<td>0.842</td>
<td>0.974</td>
<td>0.887</td>
<td>0.800</td>
<td>0.961</td>
<td>0.716</td>
<td>0.948</td>
<td>0.734</td>
<td>0.977</td>
<td>0.762</td>
</tr>
<tr>
<td>Clinical-BigBird</td>
<td>0.827</td>
<td>0.967</td>
<td>0.872</td>
<td>0.787</td>
<td>0.952</td>
<td>0.715</td>
<td>0.944</td>
<td>0.711</td>
<td>0.972</td>
<td>0.755</td>
</tr>
</tbody>
</table>

#### Short-text Models

<table>
<thead>
<tr>
<th>Model</th>
<th>acc @ medNLI</th>
<th>F1 @ i2b2 2006</th>
<th>F1 @ i2b2 2010</th>
<th>F1 @ i2b2 2012</th>
<th>F1 @ i2b2 2014</th>
<th>F1 @ emrQA Med</th>
<th>F1 @ emrQA Rela</th>
<th>F1 @ emrQA CVD</th>
<th>acc @ openI</th>
<th>AUC @ MIMIC AKI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT</td>
<td>0.776</td>
<td>0.939</td>
<td>0.835</td>
<td>0.759</td>
<td>0.928</td>
<td>0.675</td>
<td>0.924</td>
<td>0.698</td>
<td>0.952</td>
<td>0.514</td>
</tr>
<tr>
<td>BioBERT</td>
<td>0.808</td>
<td>0.948</td>
<td>0.865</td>
<td>0.789</td>
<td>0.93</td>
<td>0.700</td>
<td>0.926</td>
<td>0.702</td>
<td>0.954</td>
<td>0.534</td>
</tr>
<tr>
<td>ClinicalBERT</td>
<td>0.812</td>
<td>0.951</td>
<td>0.861</td>
<td>0.773</td>
<td>0.929</td>
<td>0.698</td>
<td>0.929</td>
<td>0.711</td>
<td>0.967</td>
<td>0.738</td>
</tr>
</tbody>
</table>
Results

• We released the pre-trained models and codebase on HuggingFace and GitHub
  – https://huggingface.co/yikuan8/Clinical-Longformer
  – https://huggingface.co/yikuan8/Clinical-BigBird
  – https://github.com/luoyuanlab/Clinical-Longformer

Roadmap

• Communication
  – ChatGPT in medical and scientific writing

• Resource
  – BERT in building knowledge graphs

• Practice
  – Clinical-Longformer and Clinical-BigBird

• Surveillance
  – ChatGPT in pharmacovigilance
ChatGPT-based Adverse Drug Reaction (ADR) detection

• To detect a rare ADR from clinical trials, 10,000 participants is needed\(^1\)
• Only 1-10% of ADRs are reported to the FDA Adverse Event Reporting System\(^2\)

1. Lee et al., 2021
2. Meyboom et al., 1999
ChatGPT in pharmacovigilance

Drug abuse identification through social media data analysis with LLMs

GPT3 was previously shown to expand the lexicon of colloquial drug synonyms from social media posts. ChatGPT can recognize drug abuse risks from tweets

<table>
<thead>
<tr>
<th>Tweet</th>
<th>Risk Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ever since my Acid trips like whenever I get super high I just start—lightly hallucinating and it’s tbh creepy”</td>
<td>High</td>
</tr>
<tr>
<td>“drove like 10 miles on these icy ass roads all to get some weed if imma—be locked up in my house for awhile imma need some weed”</td>
<td>High</td>
</tr>
<tr>
<td>“Just watched Fear and Loathing in Las Vegas for the first time—and I think I should have been on acid to fully understand it”</td>
<td>High</td>
</tr>
<tr>
<td>“Smoking a blunt at home so much better than going to the woods in—Brooksville and puking on yourself Bc you drank too much reball”</td>
<td>High</td>
</tr>
<tr>
<td>“today I was asked if I do heroin because I went to Lancaster????”</td>
<td>High</td>
</tr>
<tr>
<td>“Morgan told me my Bitmoji looks like a heroin addict?”</td>
<td>High</td>
</tr>
</tbody>
</table>

Table adapted from Hu et al., 2019

1. Carpenter et al., 2023
ADR ranking and signal detection capabilities of ChatGPT

1. Rugo et al., 2020
2. PIQRAY® (alpelisib) tablets, for oral use. 2019
3. Agency EM. Assessment report - Piqray. 2020
Using ChatGPT as a starting point for scientific evidence retrieval

1. Jain et al., 2018
2. Hopkins et al., 2018
Potential of ChatGPT as a pharmacovigilance knowledge database

Choice of query language matters

- Adverse effects?
- Furosemide
- Lasix (brand name)
- \( A_j \)
- \( A_j \)
- Diuretic to treat fluid retention
  - electrolyte imbalance
  - hypotension
  - ototoxicity
  - ...
- Chinese herbal compound
- Antibiotic: furazolidone

- BCRP inhibitors
- P-glycoprotein inhibitors
- Strong CYP3A inhibitors
- Acid-reducing agent
- CYP2C9 substrates
- Strong CYP3A inducers

Selective coverage in response

Drug-drug interaction?

Listed on drug label but not included by ChatGPT

Included by ChatGPT but weak interaction

Included by ChatGPT but no clear evidence
ChatGPT in pharmacovigilance

ADR case review and text summarization capabilities of ChatGPT

Tasks done by ChatGPT for summarization:
- Named entity recognition
- Relation extraction

Cost of manual case review

Possible ADRs mentioned in the notes?

1. Eye discharge after starting Dilantin postoperatively for seizure prophylaxis
2. Oral sores, rash...might be associated with Dilantin
3. Diffuse erythema and pustules...after Decadron was tapered and discontinued
4. Positive ketones...could indicate an adverse effect of Lipitor or Tylenol with codeine

H Wang, J Ding, Y Luo. Future of ChatGPT in Pharmacovigilance. Drug Safety, 2023 accepted
Future Directions

Streamlined, interoperable and scalable biomedical NLP

R&D, medical/regulatory
- Cross Functioning Datamart
  - Pharmacovigilance
  - Real-world drug usage trends
  - Clinical trial eligibility criteria
  - Literature mining, knowledge graphs

Cloud processing
- NLP as a service
- Reducing technical barrier
- Enabling in-depth information access

On-premise
Cloud environment

Common Data Model Tables
- Persistent and interoperable

Bulk-NLP Pipeline
- Section recognition
- Sentence breaking
- Tokenization
- Regex Pattern Matcher
- Stemming
- Part-of-Speech tagging
- Coreference resolution
- Concept recognition
- Graph representation and graph mining

Example clinical tasks
- Computational phenotyping
- Adverse drug event detection
- Trial eligibility criteria
- Specialist referral

Streamlining

Interoperability

Scalability

Democratize AI/ML literacy, tooling and results to lower access barriers for partners

PMID: 30591037
PMID: 29854260
PMID: 32308812
PMID: 32308871
PMID: 30943974
How much can LLM help?

**Tag part of speech for the sentence**

"Progress notes and pathology report contain rich information on breast cancer local, distant recurrences and contralateral events"

- Progress - Noun
- notes - Noun
- and - Conjunction
- pathology - Noun
- report - Noun
- contain - Verb
- rich - Adjective
- information - Noun
- on - Preposition
- breast - Noun
- cancer - Noun
- local - Adjective
- , - Punctuation
- distant - Adjective
- recurrences - Noun
- and - Conjunction
- contralateral - Adjective
- events - Noun

Overall, the sentence contains 9 nouns, 4 adjectives, 1 verb, 2 conjunctions, 1 preposition, and 1 punctuation mark.

**Find named entities in the sentence**

"Progress notes and pathology report contain rich information on breast cancer local, distant recurrences and contralateral events"

- "Progress notes" (referring to a type of medical record)
- "pathology report" (referring to a medical document containing analysis of tissue samples)
- "breast cancer" (referring to a type of cancer)
- "local" (referring to a type of recurrence of breast cancer)
- "distant" (referring to a type of recurrence of breast cancer)
- "contralateral" (referring to an event affecting the opposite side of the body)

Note that the last three named entities ("local," "distant," and "contralateral") are technically adjectives describing different types of recurrences and events associated with breast cancer, but they can be considered named entities in the context of the sentence as they refer to specific medical concepts.
How much can LLM help?

In this constituency parse, the sentence is composed of a subject noun phrase "Progress notes and pathology report", and a verb phrase "contain rich information on breast cancer local, distant recurrences and contralateral events". The verb phrase is divided into a verb "contain", an object noun phrase "rich information on breast cancer local, distant recurrences and contralateral events", and a prepositional phrase "on breast cancer local, distant recurrences and contralateral events" modifying the object. The object noun phrase is composed of a noun "information", modified by an adjective phrase "rich", and a prepositional phrase "on breast cancer local, distant recurrences and contralateral events". The prepositional phrase is composed of a preposition "on", a noun phrase "breast cancer", and a coordination of three noun phrases "local", "distant recurrences", and "contralateral events". The coordination is joined by a comma and the conjunction "and".
Future Directions

LLM to integrate multi-modal healthcare data

CURRENT LIMITATIONS IN THE ML DATA FUSION PIPELINE

MODEL TRAINING
- Training multiple models
- Weighting of data interaction
- Voting issues for multiple models

MODEL BUILDING
- Complex and time-consuming multi-modal models, creating a barrier to creation
- Unclear which fusion models are superior

MODEL TESTING
- Lack of comparison with single modality
- Lack of comparison with alternative fusion strategies

DATA FLOW THROUGH
- Digitally recorded data/retrospective
- Missing data
- Single site

TRANSLATIONAL SUPPORT
- Lack of FDA approved tools (0%)
- Ease of use for clinical partnerships
- Clinical relevance is unclear

LLMs can help us move from reactive to proactive machine learning

Dialogue LLMs such as ChatGPT fits perfectly in the proactive ML paradigm

Let us work together and bring it to a whole new level

• Collaboration welcome
• yuan.luo@northwestern.edu
• @yuanhypnosluo
• We are hiring, multiple postdoc positions available
• https://labs.feinberg.northwestern.edu/luolab/
• Main funding support acknowledgement
  – U01TR003528
  – U54HL160273
  – R01LM013337
  – R01GM105688
  – R21LM012618
  – UL1TR001422
  – U01HG011169