Clinical Research Seminar Series
February 24, 2017

Alberta’s Tomorrow Project
Presented by: Drs. Paula Robson & Sambasivarao Damarju

Alberta’s Tomorrow Project (ATP) is Alberta’s largest longitudinal health research study, with 55,000 participants enrolled to date. Dr. Paula Robson (Scientific Director, ATP) will introduce the ATP and how it can support researchers in the areas of cancer and chronic disease etiology. Dr. Sambasivarao Damarju (Professor, FoMD) will present highlights from his Cancer Genomics projects and how ATP data contributed to novel discoveries.

Questions?
Dr. Jill Byrne (Director, Clinical Research; FoMD)
jill.byrne@ualberta.ca

Alberta’s Tomorrow Project: Leveraging Provincial Data & Sample Accessibility

Inspiring Research For A Healthier Tomorrow

To provide:

- An overview of Alberta’s Tomorrow Project (ATP):
  - Phase I (2000 to 2008)
  - Phase II (Canadian Partnership for Tomorrow Project; 2008 to 2015)
- Information on how to access data and biological samples
- Examples of past and current projects using ATP data and biological samples

Objectives
**Alberta’s Tomorrow Project – prospective cohort**

- **35-69y, no cancer**
- **Cancer? Chronic disease?**

**Data/Samples**

What differentiates those who develop disease from those who do not?

**Health & Lifestyle data**
**Biological samples**
**Consent for linkage and active follow-up**

---

**ATP – Phase I**

- **‘Original’ Tomorrow Project (2000-2008)**
- **Started with a grant from New Initiatives Program, Alberta Cancer Board (PI: Heather Bryant, MD PhD)**

---

**ATP – Phase I - Original Proposal**

Objectives – to determine feasibility of:

- establishing cohort of Albertans to create a platform for health research
- collecting follow-up questionnaires from cohort participants
- linkage with administrative data to explore health services utilization
- collecting blood samples for banking

---

**Phase I - Recruitment**

- Random digit dialing (RDD by health region)
- Selected eligible individual from within household
  - 35-69 years old
  - able to complete questionnaires in English
  - no prior history of cancer, other than non-melanoma skin
  - if more than one eligible person, select most recent birthday

- Are you willing to receive an enrollment package in the mail?
Phase I – Enrollment package

- Health & Lifestyle Questionnaire (HLQ)
- Tape measure (210cm long)
- Study information booklet and consent form (included request for Personal Health Number and consent for linkage)

Phase I – Diet and physical activity assessment

Phase I – Questionnaires completed

- 31,072 people (39% men) enrolled by returning HLQ and consent in the mail
- 98% consented to linkage with administrative databases and provided a viable Personal Health Number (PHN)
- 26,843 completed diet questionnaire (86.4%)
- 26,769 completed physical activity questionnaire (86.2%)

In 2008, 20,707 people completed follow-up questionnaire (some updated information and new information)

Phase I – who joined ATP?

<table>
<thead>
<tr>
<th>Category</th>
<th>Men (n=12,116)</th>
<th>Women (n=18,956)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-54y</td>
<td>67.5</td>
<td>67.9</td>
</tr>
<tr>
<td>55-69y</td>
<td>32.5</td>
<td>32.1</td>
</tr>
<tr>
<td>Less than high school</td>
<td>11.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Completed high school</td>
<td>14.9</td>
<td>20.8</td>
</tr>
<tr>
<td>Some post-secondary</td>
<td>18.7</td>
<td>22.3</td>
</tr>
<tr>
<td>Completed post-secondary</td>
<td>55.5</td>
<td>47.9</td>
</tr>
<tr>
<td>Underweight</td>
<td>0.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Normal</td>
<td>23.0</td>
<td>39.2</td>
</tr>
<tr>
<td>Overweight</td>
<td>48.4</td>
<td>33.2</td>
</tr>
<tr>
<td>Obese</td>
<td>28.4</td>
<td>26.0</td>
</tr>
<tr>
<td>Daily smoker</td>
<td>16.0</td>
<td>15.2</td>
</tr>
<tr>
<td>Occasional smoker</td>
<td>3.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Former smoker</td>
<td>38.9</td>
<td>35.0</td>
</tr>
<tr>
<td>Never smoker</td>
<td>41.6</td>
<td>45.9</td>
</tr>
</tbody>
</table>
ATP – Phase II
Canadian Partnership for Tomorrow Project (CPTP)
2008-2015

Phase II - CPTP

- Canadian Partnership Against Cancer – catalyzed CPTP by bringing together five existing/emerging cohorts in 2008
- Aimed to enroll ~300,000 participants from 8 provinces to CPTP
- Alberta committed to ~40,000 participants for CPTP

Phase II – re-consenting participants to CPTP protocol

- Similar elements to ATP Phase I
  - Questionnaire on health and lifestyle
  - Cancer and chronic disease focus
  - Long-term follow up of participants
  - Linkage with administrative health databases
  - Consent to allow access by ‘qualified’ researchers to repositories of data and biological samples

Phase II – re-consenting participants to CPTP protocol

BUT BIG DIFFERENCES TOO...

- Consider attending a ‘study centre’
- Have a suite of measurements taken
- Provide samples of blood and urine for banking
- Consent to having biosamples and coded questionnaire data stored outside Alberta
Phase II Study centre visit – physical measures
- Height (standing and sitting)
- Weight
- Waist circumference
- Hip circumference
- Grip strength
- Body composition (bioelectrical impedance)
- Blood pressure
- Resting heart rate

Phase II Study centre visit – biological samples
- Non-fasting venous blood
  - SST tube (~20ml)
  - EDTA tube (~30ml)
- Separated into serum, plasma, red cells and buffy coat
- Spot urine sample
- Aim for samples to be processed and frozen at -80°C within 2 hours of collection

Phase II by the numbers

<table>
<thead>
<tr>
<th>Study Location</th>
<th>n</th>
<th>Men %</th>
<th>Women %</th>
<th>Blood n</th>
<th>Physical measures n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic PATH</td>
<td>31,173</td>
<td>30</td>
<td>70</td>
<td>23,897</td>
<td>11,271</td>
</tr>
<tr>
<td>CARTaGENE</td>
<td>42,472</td>
<td>45</td>
<td>55</td>
<td>29,874</td>
<td>19,748</td>
</tr>
<tr>
<td>Ontario Health Study</td>
<td>161,491</td>
<td>40</td>
<td>60</td>
<td>35,561</td>
<td>13,395</td>
</tr>
<tr>
<td>Alberta’s Tomorrow Project</td>
<td>39,124</td>
<td>35</td>
<td>65</td>
<td>29,193</td>
<td>30,511</td>
</tr>
<tr>
<td>BC Generations Project</td>
<td>29,172</td>
<td>31</td>
<td>69</td>
<td>26,562</td>
<td>16,029</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>303,432</strong></td>
<td><strong>145,087</strong></td>
<td><strong>90,954</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Goal of ATP and CPTP

➢ To make data and biosamples available to researchers
  ▪ Do NOT have to be affiliated with ATP or CPTP
  ▪ Have to be a bona fide researcher
  ▪ Need a research protocol and evidence of ethics approval for that protocol
  ▪ Need to submit a formal application (which may be peer reviewed)
  ▪ Need to sign a research agreement (and a materials transfer agreement for biosamples)

➢ Will also provide letters of feasibility to support grant applications

Where to find out more … ATP

➢ For Alberta’s Tomorrow Project – start with myatp.ca

➢ Send an email to atp.research@ahs.ca – a team member will get back to you to discuss feasibility of your idea (i.e. does ATP have the data you need?)

➢ If ATP has what you need, we will provide letters of feasibility/support for grant applications

➢ If you already have resources to do the work, we will support you through the application process

What sort of research have people done with ATP resources?

Working with researchers at the School of Public Health

Project Team
Dr Jeffrey Johnson (PI), Dr Dean Eurich, Dr Paula Robson, Dr Ming Ye (post doctoral fellow)

Project title
Patterns of healthcare utilization among a population-based cohort in Alberta, 2000-2015: Linking Alberta’s Tomorrow Project data to Alberta Health Administrative Databases

What is ATP doing to link with administrative health databases?

Working with researchers at the School of Public Health

Project Team
Dr Jeffrey Johnson (PI), Dr Dean Eurich, Dr Paula Robson, Dr Ming Ye (post doctoral fellow)

Project title
Patterns of healthcare utilization among a population-based cohort in Alberta, 2000-2015: Linking Alberta’s Tomorrow Project data to Alberta Health Administrative Databases
Linking datasets

Alberta Cancer Registry
- PHN: 123456789
- Name/Sex/DOB

ATP Cohort Data
- PHN: 123456789
- Study ID: 00000001
- Name/Sex/DOB

Alberta Health Datasets
- PHN: 123456789
- PIN: AHS-123456789
- Name/Sex/DOB

Causes of Death
- Age at Diagnosis
- Diagnosis (ICDO-3)
- Treatment
- Primary Site of Cancer

Research Activities

Next steps for ATP/CPTP

- Continue building relations with academic research community to promote use of the data/biosamples, and help shape future follow-up

Alberta's Tomorrow Project thanks:
- Alberta Health and the Alberta Cancer Prevention Legacy Fund
- Alberta Cancer Foundation
- Canadian Partnership Against Cancer
- Alberta Health Services (host organization)
- ATP staff
- ATP participants
GWAS for Breast Cancer Susceptibility & Machine Learning Predictive Models

Dr. Sambasivarao Damaraju
Professor
Laboratory Medicine and Pathology
University of Alberta

Damaraju lab Projects

- Genetic predisposition to
  - Sporadic breast cancer
  - Disease prognosis (recurrence)
  - Adverse drug reactions
  - Cancer cachexia
  - Non-coding RNAs and Gene regulation
  - Drug resistance
  - Machine learning models

- Approach
  - Genome wide association studies (GWAS)
  - Breast cancer cases from Alberta (CBCF TB)
  - Healthy controls from Alberta (Tomorrow Project)

Damaraju lab Progress 2006-

- First GWAS for sporadic breast cancer (2011)
- First GWAS for germline prognostic markers (2013)
- First for non-coding RNAs/NGS for prognostic markers (2015, 2016)
- First for cancer cachexia
  - GWAS (2017)
  - Non-coding RNAs (2017)
Human Genome Sequencing
Implications & Milestones

• 2001 first draft of human genome
• 2003 next refined draft of human genome
• 2006 first commercial high density polymorphism arrays
• 2007 first GWAS for familial breast cancer (international consortia)

Damaraju lab-Humble
Beginnings……..

• 2003-2004
  – Collaboration with Tomorrow Project
    • Drs Damaraju, Mackey, Robson, Bryant
• Pilot project
  – 800+ Controls from Tomorrow Project
    • First specimens banked
  – 600+ Breast cancer cases
• Developed SOPs for sample banking
  – Started Provincial Tumor Banking Project
  • Damaraju- Founding Scientific Director
  • Mackey- Founding Medical Director
  • Cass- Project Lead/coordinator

Genetic predisposition to BCa

![Diagram of Genetic predisposition to BCa](image)

Etiology of BCa

![Diagram of Etiology of BCa](image)
Genetic Association Studies

- Case-control studies

- Take advantage of thousands of known SNPs in the human genome

- Compare the frequencies of SNPs/CNVs (or other genetic variations) between cases and controls and seek statistical significance

GWAS- Multi-stage Association Study Design

Heritability of Breast Cancer and Genetic Variants Identified

Damaraju lab findings:
Breast Cancer Susceptibility
Accessing Tomorrow Project Samples

- Go to web site
  - http://myatp.ca/for-researchers/access-guidelines-and-procedures
- Be clear about your study design
  - Clearly explain the exclusion and inclusion criteria and access the right samples and data
- Give sufficient time (cannot be rushed)
- For grant applications, give at least a couple of months time to understand the data complexity and what you exactly need
- Costs to access are nominal (charges for application process and small fee to access specimens from freezers)
  - Costs to access DNA or serum are very low compared to other biobanks
- Access only when you are ready to process/conduct experiments as this resource (samples) cannot be replenished.
- Try collaborate with researchers who have common goals and have prior experience in accessing/handling data and samples
- Tomorrow project will not do statistical analysis for you on the data you have generated to optimize the resources

Affymetrix SNP 6.0 genotyping

www.affymetrix.com

First GWAS paper from Canada on BC Susceptibility
Focus: Late age at onset breast cancer and no family history (Sporadic cases)

Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility

Sporadic Breast Cancer Studies Insights from GWAS and Candidate SNPs

Identification of a Breast Cancer Susceptibility Locus at 4q31.22 Using a Genome-Wide Association Study Paradigm

Yadav Sapkota\*1, Yutaka Yasa\*, Raymond La\*, Malinee Sridharan\*, Paula J. Robson\*, Carol E. Cass\*1, John R. Mackay\*2, Sanmugasundaram Damara\*2

IDNP: ACC0205 | Ready available online

Assessing SNP-SNP interactions among DNA Repair, Modification and Metabolism Related Pathway Genes in Breast Cancer Susceptibility

\*1: University of Alberta | \*2: University of British Columbia | \*Correspondence: Yadav Sapkota, yadav.sapkota@ualberta.ca, Sanmugasundaram Damara, sanmugasundaram.damara@ualberta.ca

www.affymetrix.com
Damaraju lab Breast Cancer GWAS Findings.....

• Conducted 4 Stage study
  – 5014 breast cancer cases (Provincial Tumor bank)
  – 5000 controls (Tomorrow Project)
• Sporadic breast cancer variants identified
  – All associations adjusted for BMI
• Two SNPs from Chr. 4 and 5 showed associations (Bonferroni significant)
  • Allelic association p-value- 3.33E-09 (OR 1.3)
  • Allelic association p-value- 1.38E-09 (OR 1.3)

• When cases were stratified for menopausal status, Chr. 4 and 5 SNPs showed-
  – Allelic Association with Premenopausal
    p-values 1.82E-10 and (OR- 1.42) and 1.25E-11 (OR- 1.46)
    • Highest ORs for BC GWAS variants in literature
  – Postmenopausal- allelic association
    p-values 1.07E-04 (OR - 1.21) and 1.0E-04 (OR-1.21)
    • Not Bonferroni significant

Important Findings.....

• Validated in international cohorts
  – Chinese ancestry (~5000 samples)
  – African Ancestry (~2000 samples)
  – Caucasian Postmenopausal cohorts (~2500 samples)
• Chr. 4 and Chr. 5, SNPs were not associated with BC risk
• Damaraju lab findings from Alberta (Canada) are unique non-consortia work
• Unique in GWAS literature and show highest risk for pre-menopausal women

Important Findings.....

Machine Learning models/tools developed using TP cohorts

1. ML models using SNP data
   - For breast cancer prediction
   - For genetic ancestry prediction
2. How accurately can we predict who will get breast cancer using health, lifestyle and physical activity questionnaire data?
Definitions

- Predictive model
  - encompasses a variety of statistical techniques from modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future, or otherwise unknown, events (Wikipedia)
- To know that a certain population are at higher risk for a disease is different from which individual is going to be affected

Risk models

- http://www.yourdiseaserisk.wustl.edu
  - different diseases; cancer, bronchitis, emphysema, diabetes, heart disease, stroke, osteoporosis
- Gail model for breast cancer risk
- Risk estimates do not allow one to say precisely which woman will develop breast cancer. In fact, some women who develop breast cancer may have lower estimated risks than some women who do not develop breast cancer.
Data (2003-present)

- 1386 individuals
  - 810 healthy individuals with no cancer
  - 576 individuals with breast cancer (and no previous cancer)

- Features
  - 378 Questionnaire Data
    - Health and Lifestyle Questionnaire
    - Diet History Questionnaire
    - Physical Activity Questionnaire

- The baseline of all the data is 58.44%

Goal/Question: Who will get breast cancer?

Well studied risk factors for Bca: Who might be at risk in populations

- Age
- Ethnicity
- Family history (with or without BRCA gene mutations)
  - Other high penetrant gene mutations
- BMI (several proxy indices-obesity, HWC, etc)
- Physical activity
- Diet

Machine learning approach (Collaboration with Dr. Greiner)

- Data split into training (90%) and validation sets (10%)
- 10 fold cross validation on Training data
- Feature selection in each fold – did this for different # of features
- Take the highest accuracy based on # of features and use the Validation set (10%) to report the classifier accuracy

Top 8 Selected Features

1. Age
2. Caucasian
3. Type of Food/protein (Oz_lean_meat_equivalent_from_eggs)
4. Body measurement (BMI)
5. Social: About how many close friends and close relatives do you have, that is, people you feel at ease with and can talk to about what is on your mind? Close friends and relatives
6. Exercise: Average weekly metabolic output from recreation and leisure activities
   \[ \sum_{\text{all recreational activities}} \text{MET value} \times \text{hours/week} \]
7. Family/Friends: How often is each of the following kinds of support available to you? Someone to help you with daily chores if you were sick
8. Env exposure: In the past year, were you exposed to second hand smoke in public places (bars, restaurants, shopping malls, arenas, bingo halls, bowling alleys)?
Acknowledgements

- Team
  - Badar Sehrawat
  - Ashok Narasimhan
  - Preetha Krishnan
  - Mihir Arora
  - Malinee Sridharan
  - Yadav Sapkota
  - Mohsen Hajiloo
  - Jennifer Dufour
- The CBCF and the Tomorrow Project team members
- Funding to Damaraju lab
  - AGF/ACB
  - CBCF
  - CIHR

Collaborators

- Dr. John Mackey
- Dr. Yutaka Yasui
- Dr. Russ Greiner
- Dr. Paula J Robson
- Dr. Sunita Ghosh
- Dr. Carol Cass
- Dr. Raymond Lai
- Dr. Liang Li
- Dr. David Wishart
- Dr. Olga Kovalchuk
- Dr. Vickie Baracos
- Dr. Anil Joy

Thank You