Genomic and Epigenomic Profiles from South East Asia

Patrick Tan, MD PhD
Professor, Duke-NUS Medical School
Deputy Executive Director, Biomedical Research Council, Agency for Science, Technology and Research (A*STAR)

Weekly Precision Medicine Forum
Duke University, Oct 2017
South East Asia is 650 Million People and >10 Different Countries
Outline of Presentation

1) **Cancer Patterns** and Global Health
   - Lifestyle Factors and Molecular Oncology

2) **Clinical Implementation** of Precision Medicine
   - Intercepting Early Disease

3) **Beyond** Human Diversity (new!)
   - Exposures and Nutrition
Asian Cancers – A Vast Unmet Clinical Need

• Cancers caused 14 million new cases and 8.2 million cancer related deaths in 2012

• New cases expected to rise by 70% over the next 2 decades.

• >60% of new cases occur in Asia, Africa, and Central and South America, accounting for 70% of the world’s cancer deaths.

• Top cancer killers are lung, liver and gastric cancers

Source: WHO
Asian Cancers and Group I Carcinogens

Gastric Cancer

Nasopharyngeal CA Lymphomas

Bile Duct Cancer (Cholangiocarcinoma)

Urinary Tract Cancer

Helicobacter pylori

Epstein-Barr Virus

O. viverrini (Liver Fluke)

Aristolochia Plants (eg Birthwort)

Zang et al., 2012 Nature Genetics
Qamra et al., 2017 Cancer Discovery
Koo et al., 2012 Cancer Discovery
Ong et al., 2012 Nat Genet
Chan-on et al 2013 Nat Genet
Poon et al., 2013 Sci Trans Med
Yao et al., 2017 Cancer Discovery
Cholangiocarcinoma (CCA): A Subtype of Liver Cancer Arising from Bile Ducts
CCA is Highly Prevalent in NE Thailand (Est Population = 20 million)

Age-Standardized Incidence Rate (ASR)

<1.5 in Western Countries
94.8 and 39.4 in Khon Kaen Province

~12-14,000 new cases diagnosed each year

Caused by infection by the liver fluke *Opisthorchis viverrini* (OV)

World J Gastroenterol. 2005 Jun 14;11(22):3392-7
Opisthorchis Viverrini (OV) – Life Cycle and Social Habits

Figure 1. Preparation of a Meal of Koi-Pla Using Uncooked Cyprinoid Fishes

Studying Fluke-Positive CCA
(with Khon Kaen University, NE Thailand)
• 740 Mb draft *O. viverrini* genome (16,000 genes)
• Genomic expansions of cholesterol transporters (*NPC1/2*)
• Disruption of bile homeostasis
• OV secreted factors with homology to human growth factors

*Young et al., Nature Communications* (2014)
A Global Map of Cholangiocarcinoma Genomic Variation (489 CCAs, 10 Countries)

- **71** CCAs WGS (Fluke-Positive and Fluke-Negative)
  - Thailand
  - Singapore
  - Romania
  - Italy
  - Japan

- **418** Validation CCAs
  - Singapore, Romania, Japan, Thailand, Taiwan, Brazil, France, South Korea, China
  - Exome/Targeted Sequencing (~400 genes)
Validation and Discovery of CCA Driver Genes
Striking Convergence of FGFR-related Alterations

**FGFR2/FGFR3 Gene Fusions**

**FGFR2 3’UTR Loss**

**FGFR2 Activating Mutations**

- **S252W**
- **Y376C**
- **N550K/S**
Heterogeneity in DNA Methylation
Distinct Epigenetic Subtypes and Fluke Status
CCA Methylation Subtypes Have Distinct Mutational Signatures and Clonality Patterns
Model: Causes of Cancer Influence Tumor Somatic Landscapes
Summary

• Largest genomic and epigenomic survey of cholangiocarcinoma, from different etiologies

• Discovery of new CCA driver genes (RASA1, MAP2K4, SF3B1)

• FGFR 3’ UTR deletion as a mechanism for FGFR upregulation

• Existence of distinct DNA Hypermethylation Subtypes (CpG Islands vs CpG Shores)

• Non-coding promoter mutations target H3K27me3 sites in CCA

Jusukul et al., 2017 Cancer Discovery
Outline of Presentation

1) Cancer Patterns and Global Health
   - Lifestyle Factors and Molecular Oncology

2) Clinical Implementation of Precision Medicine
   - Intercepting Early Disease

3) Beyond Human Diversity (new!)
   - Exposures and Nutrition
Can We Intercept Early Disease Using Detailed Phenotyping?

Development of Disease

- Baseline Risk
- Earliest Molecular Detection
- Earliest Clinical Detection
- Typical Current Intervention

Initiating Events

Quantify Baseline Risk
Monitor Progression
Health Enhancement

Refine Risk Prediction
Monitor
1° Prevention
Define Disease

Personalize Therapy
Disease Management

Time

Cost

1/reversibility

Disease Burden

Courtesy Ralph Snyderman
SingHealth Duke-NUS Academic Medical Center
Potential Epicenter for Precision Medicine

14,000 Eye Surgeries/yr

1,700 Beds

9,000 Inpatient Admissions/yr

>125,000 Clinic Attendees/yr

>20 Subspecialties

>175,000 Pediatric Emergency Visits/yr
A Global Missing Gap: Asian Genomic Data


NEJM, 2016
SPECTRA – A Genome/Phenome Encyclopedia for Asian Patient Normality

Collaboration with NHC Biobank, SingHEART and IDA
(PIs : Stuart Cook and Yeo Khung Keong)

Volunteer

National Heart Centre Singapore
SingHealth

Consented for Research, Incidental Findings, and Long Term Follow Up

Lifestyle Factors and ECG

Imaging Studies (MRI, Calcium)

Activity/Sleep Monitoring (Wearables)

Whole Genome Sequencing (30x)

Serum Metabolomics (Targeted)
Recruitment of volunteers

Inclusion criteria
- Aged ≥ 16 years and ≤ 90 years

Exclusion criteria
- Heart attack/ coronary heart disease with stenosis/ percutaneous coronary intervention (PCI)
- Stroke
- Pregnancy
- Diabetes
- Cardiac pacemaker, brain aneurysm or clips
- Medication
- First degree family member with a cardiac condition
- Family members volunteered for this study

Criteria for biobank

- Aged ≥ 16 years and ≤ 90 years
- Heart attack/ coronary heart disease with stenosis/ percutaneous coronary intervention (PCI)
- Stroke
- Pregnancy
- Diabetes
- Cardiac pacemaker, brain aneurysm or clips
- Medication
- First degree family member with a cardiac condition
- Family members volunteered for this study
Current Progress of SPECTRA (August 2017)

- Lifestyle Habits (1840)
- Cardiac MRI (1600)
- Cardiac NGS (800)
- WGS (750)
- Lipidomics (550)
- Activity (520)

- SingHEART
- PRISM
- Biobank
Consent process includes:

- Access to medical records
- Collection of family history

Choice to:
- Receive significant findings from medical screen
- Receive genetic findings
- Be contacted annually to review health status
Family history collection - MeTree

- Online family history collection tool developed by Duke University
MeTree Risk Results (Singapore)
Online vs Static Data Collection

<table>
<thead>
<tr>
<th>Risk Recommendation</th>
<th>Online Collection (N=99)</th>
<th>Static Collection (N=372)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine Risk</td>
<td>78 (78%)</td>
<td>152 (93%)</td>
</tr>
<tr>
<td>Cancer genetic counselling</td>
<td>6 (6%)</td>
<td>6 (1.6%)</td>
</tr>
<tr>
<td>Thrombosis genetic counselling</td>
<td>2 (2%)</td>
<td></td>
</tr>
<tr>
<td>Cardiac genetic counselling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertain increased cancer risk</td>
<td>2 (2%)</td>
<td>1 (0.3%)</td>
</tr>
<tr>
<td>CRC - Early/more frequent screening</td>
<td>6 (6%)</td>
<td>10 (2.7%)</td>
</tr>
<tr>
<td>Breast MRI</td>
<td>2 (2%)</td>
<td></td>
</tr>
<tr>
<td>Chemoprophylaxis for breast cancer</td>
<td>2 (2%)</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer screening</td>
<td>4 (4%)</td>
<td>2 (0.5%)</td>
</tr>
<tr>
<td>Cardiac screening</td>
<td>1 (1%)</td>
<td></td>
</tr>
<tr>
<td>Liver genetic testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes screening</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Ryanne Wu
### Comparisons against USA Data

<table>
<thead>
<tr>
<th></th>
<th>US population</th>
<th>Singapore population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (SD)</strong></td>
<td>59 (12)</td>
<td>47</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>59%</td>
<td>57%</td>
</tr>
<tr>
<td><strong>Racial make-up</strong></td>
<td>82% White</td>
<td>85% Chinese</td>
</tr>
<tr>
<td></td>
<td>13% Black</td>
<td>7.3% Filipino</td>
</tr>
<tr>
<td></td>
<td>5% Other</td>
<td>4.5% Indian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.7% Malay</td>
</tr>
<tr>
<td><strong>Genetic counselling – Cancer</strong></td>
<td>26%</td>
<td>6.1%</td>
</tr>
<tr>
<td><strong>CRC screening</strong></td>
<td>18.6%</td>
<td>6.1%</td>
</tr>
<tr>
<td><strong>Breast MRI/chemoprophylaxis</strong></td>
<td>9.8%</td>
<td>6.3%</td>
</tr>
</tbody>
</table>

Ryanne Wu
Genomic screen

Variants within 117 genes associated with:

- Cancer
- Cardiometabolic
- Respiratory
- Haematology
- Neurology
- Ophthalmology
- Organ specific
- Multi-organ

40 drug response variants (pharmacogenomics)

The type of genetic information returned may indicate:

- A diagnosis of a treatable genetic condition (primary diagnosis)
- An increased risk of developing a genetic condition (secondary finding)
- Information that you are a carrier for a genetic condition (carrier risk)
- Information about how you may respond to medication (pharmacogenomics)
WGS variant filters

Variants present in 117 PRISM genes
Quality score >10, depth filter >30x

Exonic, intronic, non-synonymous, <1% population

• Predicted likely pathogenic/ pathogenic by ClinVar and/ or HGMD, OR
• Novel loss of function mutations (stop gain, stop loss, frameshift, splice site)
Integration of data for visualisation and interpretation

MDT meeting

Participant information
- Family history
- Medical history and clinical screening results
- MeTree
- Cholesterol
- Renal panel
- Fasting glucose
- LFT
- FBC
- MRI
- CT scan
- ECG
- Fitbit
- 24hr ABPM

Genomic data
- WGS variants
- Human variant databases
- ClinVar
- ExAC
- CardioDB
- HbVAR
- BIC/InSIGHT
- Inhouse SEC

Information resources
- Publication access
- PubMed
- HGMD

Multidisciplinary expertise
- Health professionals
  - Clinician specialists
  - Geneticists
  - Genetic counsellors/nurses
  - Laboratory specialists
  - Bioinformaticists
- Health disciplines
  - cardiology
  - oncology
  - ophthalmology
  - immunology
  - diabetes
  - metabolic
  - neurodegenerative
  - musculoskeletal
  - etc.

Publication access
- Human variant databases
  - ClinVar
  - ExAC
  - CardioDB
  - HbVAR
  - BIC/InSIGHT
  - Inhouse SEC
**Case 1: 46y Chinese Male**

**Genomic data:** *BRCA2* c.7377_7380del p.K2459fs
- Frameshift deletion of ACAA occurring exon 14 of 27 causing truncated protein at aa 2467

**Medical history:** none

**Clinical screen:** normal

**Human variant databases:**
- ClinVar: pathogenic by 5 submitters
- Pathogenic in Breast Cancer Information Core (BIC)
- ExAC: .000004067, SEC: not present

**Publications:**
  - Verified in one Chinese female diagnosed by breast cancer
**Variant classification according to ACMG**

<table>
<thead>
<tr>
<th>Population Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2</td>
<td>Absent in population databases PM2</td>
<td>Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5</td>
<td>Same amino acid change as an established pathogenic variant PS1</td>
<td>Prevalence in affecteds statistically increased over controls PS4</td>
<td>Predicted null variant in a gene where loss of function is a known mechanism of disease PVS1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computational And Predictive Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple lines of computational evidence suggest no impact on gene/gene product BP4</td>
<td>Multiple lines of computational evidence support a deleterious effect on the gene/gene product PP3</td>
<td>Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5</td>
<td>Protein length changing variant PM8</td>
<td>Predicted null variant in a gene where loss of function is a known mechanism of disease PVS1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functional Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-established functional studies show no deleterious effect RS3</td>
<td>Missense in gene with low rate of benign missense variants and path. missenses common PP2</td>
<td>Mutational hot spot or well-studied functional domain without benign variation PM1</td>
<td>Well-established functional studies show a deleterious effect PS3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Segregation Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-segregation with disease RS4</td>
<td>Co-segregation with disease in multiple affected family members PP1</td>
<td>Increased segregation data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>De novo Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>De novo (without paternity &amp; maternity confirmed) PM6</td>
<td>De novo (paternity &amp; maternity confirmed) PS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Allelic Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed in trans with a dominant variant BP2</td>
<td>For recessive disorders, detected in trans with a pathogenic variant PM1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Database</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reputable source w/o shared data = benign BP5</td>
<td>Reputable source w/o pathogenic PP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Data</th>
<th><strong>Strong</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Supporting</strong></th>
<th><strong>Moderate</strong></th>
<th><strong>Strong</strong></th>
<th><strong>Very Strong</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Found in case with an alternate cause BP5</td>
<td>Patient’s phenotype or FH highly specific for gene PP4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PM2 + PVS1 + PP5 = pathogenic
1. BAM files are reviewed to visualise variant
2. Research report is generated

Date: 4 July 2017

<table>
<thead>
<tr>
<th>Name:</th>
<th>Specimen: Blood</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOB:</td>
<td>Lab reference no: PRM0416-0879</td>
</tr>
<tr>
<td>Sex: M</td>
<td>Referring physician: PRISM</td>
</tr>
<tr>
<td>Ethnicity: Chinese</td>
<td>Test: Whole Genome Sequencing</td>
</tr>
<tr>
<td>MRN:</td>
<td></td>
</tr>
</tbody>
</table>

**RESULTS SUMMARY**

**PATHOGENIC VARIANTS IDENTIFIED**

- **MONOGENIC DISEASE RISK**
  - BRCA2 (NM_000059.c.7377_7380del:p.K2459fs)

- **CARRIER RISK**
  - None

- **PHARMACOGENOMIC ASSOCIATIONS**
  - Clinically relevant variants detected. See below for details.

**DETAILS ON VARIANTS OF MEDICAL SIGNIFICANCE**

**A. MONOGENIC DISEASE RISK**
This test identified a heterozygous likely pathogenic variant in BRCA2 associated with an increased risk of developing breast, ovarian and prostate cancer.

**B. CARRIER RISK**
The test did not identify any pathogenic genetic variants associated with being a carrier of a genetic condition.

**C. PHARMACOGENOMIC ASSOCIATIONS**
Clinically relevant variants were detected. Details of variant and associated drug(s) are as follows:
- rs13745274 [CYP2B6 (G/T)] - Efavirenz is likely to be less effective in someone with this genotype. Monitor response and for side effects.
- rs4149056 [SLC01A1 (T/C)] - Modest increase in risk for myopathy even at lower simvastatin doses (40mg daily); if optimal efficacy is not achieved with a lower dose, alternate agents should be considered.
- rs2228001 [APC (T/T)] - Decreased risk for toxicity with cisplatin treatment.
Case 2: 55y Chinese Male

Genomic data: \textit{LDLR} c.G268A p.D90N
- Missense mutation of glycine (hydrophobic, nonpolar, neutral) to alanine (hydrophobic, nonpolar, neutral) in exon 3 of 90/861 aa
- Conserved across species, located in the LDL-receptor class A 2 domain
- Predicted damaging by computational algorithms (ie Polyphen, SIFT)

Medical history: none
Clinical screen: normal

Family history:

Human variant databases:
- Pathogenic/ likely pathogenic by 5 submitters
- ExAC: .00005685, SEC: not present
- CardioDB: VUS

Publications:
- Multiple references with Asian data (Mak 1998, Yang, 2007, Jiang, 2015, Khoo 2000 indicating genotype/ phenotype association and segregation)
- Chang et al. 2003 demonstrated reduced function of LDL receptor
Variant classification according to ACMG

<table>
<thead>
<tr>
<th></th>
<th>Strong</th>
<th>Supporting</th>
<th>Supporting</th>
<th>Moderate</th>
<th>Strong</th>
<th>Very Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Data</td>
<td>MAF is too high for disorder BS1/BS1 OR observation in controls inconsistent with disease penetrance BS2</td>
<td></td>
<td></td>
<td>Absent in population databases. PM2</td>
<td>Prevalence in affecteds statistically increased over controls PS4</td>
<td></td>
</tr>
<tr>
<td>Computational And Predictive Data</td>
<td>Multiple lines of computational evidence suggest no impact on gene /gene product BP4</td>
<td>Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3</td>
<td>Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5</td>
<td>Same amino acid change as an established pathogenic variant PS3</td>
<td>Predicted null variant in a gene where LOF is a known mechanism of disease PS5</td>
<td></td>
</tr>
<tr>
<td>Functional Data</td>
<td>Well-established functional studies show no deleterious effect BS3</td>
<td>Missense in gene with low rate of benign missense variants and path. missenses common PP2</td>
<td>Mutational hot spot or well-studied functional domain without benign variation PM1</td>
<td>Well-established functional studies show a deleterious effect PS3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segregation Data</td>
<td>Non-segregation with disease BS4</td>
<td>Co-segregation with disease in multiple affected family members PP1</td>
<td>Increased segregation data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De novo Data</td>
<td></td>
<td></td>
<td></td>
<td>De novo (without paternity &amp; maternity confirmed) PM6</td>
<td>De novo (paternity &amp; maternity confirmed) PS2</td>
<td></td>
</tr>
<tr>
<td>Allelic Data</td>
<td>Observed in trans with a dominant variant BP2</td>
<td>Observed in cis with a pathogenic variant BP2</td>
<td>For recessive disorders, detected in variants with a pathogenic variant PM1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Database</td>
<td>Reputable source w/out shared data = benign BP6</td>
<td>Reputable source = pathogenic PS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Data</td>
<td></td>
<td></td>
<td></td>
<td>Patient’s phenotype or FH highly specific for gene PP4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PM2 + PP3 + PM1 + PP2 + PP5 = Pathogenic
1. BAM files are reviewed to visualise variant
2. Research report is generated

**Clinical Finding**

**Clinical Interpretation**

Potential drug interactions and adverse effects
Clinical significance of *LDLR p.D90N*

- Severely elevated LDL cholesterol

**Males:**
- 50% risk for a fatal or non-fatal coronary event by age 50 years

**Females:**
- 30% risk of fatal or non-fatal coronary event by age 60 years

Cascade testing to family members is available:

➢ *Our participant has 3 brothers, 1 sister and 1 son.*
Returning variants of significance

Genetic counselling appointment #1: return of research results

- Research findings and report
- Clinical significance to their health management and family members
- Validation of the variant offered—new consent and blood sample

Genetic counselling appointment #2: return of validated results

- Clarification of clinical management
- Referral to specialist clinics
- Cascade testing offered to family members
# Pathogenic variants from 450 WGS

## MONOGENIC DISORDERS (n=8)

<table>
<thead>
<tr>
<th>Variant no.</th>
<th>Volunteer</th>
<th>Pathogenic variant</th>
<th>Surveillance</th>
</tr>
</thead>
</table>
| 1.          | 46y Chinese male | **BRCA2** p.K2459fs  
Previously reported* frameshift causing protein truncation | Breast-ovarian cancer, familial, 2  
↑ risk of prostate cancer |
| 2.          | 48y Chinese female | **BRCA2** p.I1446fs  
Novel frameshift causing protein truncation | Breast-ovarian cancer, familial, 2  
↑ risk of breast and ovarian cancer |
| 3.          | 47y Chinese male** | **MSH6** p.S536fs  
Novel frameshift causing protein truncation | Lynch syndrome  
↑ risk of bowel cancer |
| 4.          | 51y Chinese female | **PROS1** p.K196E  
Previously reported missense mutation | Thrombophilia  
↑ risk of deep vein thrombosis |
| 5.          | 78y Chinese male | **MYBPC3** p.R502W  
Previously reported missense mutation | Hypertrophic cardiomyopathy  
- ongoing cardiac screening |
| 6.          | 47y Chinese female | **MYL3** p.A57G  
Previously reported missense mutation | Hypertrophic cardiomyopathy  
- ongoing cardiac screening |
| 7.          | 51y Chinese male | **DSC2** p.S740X  
Novel nonsense mutation causing protein truncation | Arrhythmogenic right ventricular dysplasia (ARVC)  
- ongoing cardiac screening |
| 8.          | 55y Chinese male | **LDLR** p.D90N  
Previously reported missense mutation | Hypercholesterolemia  
↑ risk of coronary heart disease |

*reported in ClinVar and literature as pathogenic  **same individual
### Pathogenic variants from 450 WGS

#### RECESSIVE DISORDERS (n=15)

<table>
<thead>
<tr>
<th>Variant no.</th>
<th>Volunteer</th>
<th>Pathogenic variant</th>
<th>Carrier risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CARRIER DISORDERS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1. | 46y Chinese female | *MUTYH* p.W142X  
Previously reported* nonsense mutation | MYH-associated polyposis  
- small risk of bowel cancer for carriers |
| 2. | 64y Indian male | *KCNQ1* p.E492fs  
Previously reported* frameshift causing protein truncation | Long QT syndrome  
- Variant reported in AR state  
- Baseline cardiac screen |
| 3. | 45y Chinese female | *ATP7B* p.A712V  
Previously reported missense mutation | Wilson disease |
| 4. | 31y Chinese female | *ATP7B* p.R778L  
Previously reported missense mutation | Wilson disease |
| 5. | 39y Chinese female | *CPT2* p.W10X  
Novel nonsense mutation causing protein truncation | Carnitine palmitoyltransferase II deficiency |
| 6. | 20yr Chinese male*** | *HBB* p.E27K  
Previously reported missense mutation | Beta thalassemia |
| 7. | 47y Chinese male** | *HBB* c.316-197C>T  
Previously reported cryptic splice site mutation | Beta thalassemia |
| 8. | 63y Chinese female | *HBB* p.W16fs  
Previously reported* frameshift causing protein truncation | Beta thalassemia |

*reported in ClinVar and literature as pathogenic  ** same individual  ***same individual
<table>
<thead>
<tr>
<th>Variant no.</th>
<th>Volunteer</th>
<th>Pathogenic variant</th>
<th>Surveillance</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.</td>
<td>No consent</td>
<td><strong>HBB</strong> p.K18X</td>
<td>Beta thalassemia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Previously reported* nonsense mutation</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>50y Chinese male</td>
<td><strong>LMNA</strong> p.R415C</td>
<td>Hutchinson-Gilford progeria syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Previously reported missense mutation</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>62y Chinese female</td>
<td><strong>RYR1</strong> p.R109W</td>
<td>Malignant hyperthermia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Previously reported missense mutation</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>51y Chinese male</td>
<td><strong>SLC25A13</strong> p.R284fs</td>
<td>Citrin deficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Previously reported* frameshift causing protein truncation</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>40y Chinese female</td>
<td><strong>SLC25A13</strong> p.R284fs</td>
<td>Citrin deficiency</td>
</tr>
<tr>
<td>14.</td>
<td>20y Chinese male***</td>
<td><strong>SLC25A13</strong> p.R284fs</td>
<td>Citrin deficiency</td>
</tr>
<tr>
<td>15.</td>
<td>64y Chinese female</td>
<td><strong>SLC25A13</strong> p.R284fs</td>
<td>Citrin deficiency</td>
</tr>
</tbody>
</table>

*reported in ClinVar and literature as pathogenic **same individual
Citrin Deficiency – Signs and Clinical Symptoms
Caused by mutations in SLC25A13 Mitochondrial Transporter
(Mutations are 20x more prevalent in Asia)

<table>
<thead>
<tr>
<th>Phenotypes</th>
<th>NICCD</th>
<th>FTTCD</th>
<th>CTLN2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td>Neonatal</td>
<td>Children</td>
<td>Adult</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Neonatal Hepatic Cholestasis</td>
<td>Fatigue</td>
<td>Recurrent hyperammonemia</td>
</tr>
<tr>
<td></td>
<td>Hepatomegaly</td>
<td>Hyperlipidemia</td>
<td>Neuropsychiatric symptoms</td>
</tr>
<tr>
<td></td>
<td>Diffuse fatty liver and parenchymal cell infiltration</td>
<td>Fatty liver</td>
<td>Strong food preference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Food preference</td>
<td>Low BMI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypoglycemia</td>
<td>Hepatoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Growth retardation</td>
<td>Pancreatits</td>
</tr>
<tr>
<td>Severe</td>
<td>Transient and not severe, full recovery by 1 year of age on lactose free milk</td>
<td>Varies, but non-life-threatening</td>
<td>Life-threatening</td>
</tr>
<tr>
<td>Incidence Rate</td>
<td>1 : 17,000 – 1 : 20,000</td>
<td>Unknown, not widely observed or studied outside of China</td>
<td>1 : 100,000 – 1 : 230,000</td>
</tr>
</tbody>
</table>
Registering Clinical Encounters
Outline of Presentation

1) **Cancer Patterns** and Global Health
   - Lifestyle Factors and Molecular Oncology

2) **Clinical Implementation** of Precision Medicine
   - Intercepting Early Disease

3) **Beyond** Human Diversity (new!)
   - Exposures and Nutrition
Cai Jing
Food Nutrition and Consumer Care (FNCC)

Rapidly Emerging Industry Development

A globally unique **Innovation Ecosystem** for the Food, Nutrition and Consumer Care industry has developed in Singapore in the past 5 years, over an established base of commercial and manufacturing activities.
Cracking the Durian Genome

Nature Genetics
(Out next week!)
Durian – The “National Fruit” of Singapore
Durian – Economic Value and Growth

US$800 million
Imported into China (2016)
"the more you eat of it the less you feel inclined to stop... To eat Durians is a new sensation worth a voyage to the East to experience” – Alfred Russel Wallace

“It smelled like you'd buried somebody holding a big wheel of Stilton in his arms, then dug him up a few weeks later.” – Anthony Bourdain
Integration of Sequencing Platforms

PACBIO RS II
- Long read lengths
- Moderate read errors

Illumina HiSeq 2500
- Low read errors
- Short read lengths

Dovetail Chicago + Hi-C
- Highly-contiguous scaffold linking

Long contigs assembly
- Error correction
- Genome assembly technologies
- Assembly & validation
- Chromosome-length scaffolding
D. zibethenus (Durian) Genome Sequencing

**PacBio RSII sequencing**
18.3M reads, 6.2 kb average length, 153X coverage

**FALCON-Unzip assembly**

**Arrow polishing**

**Illumina paired-end sequencing**
748.5M reads, 202X coverage

**Pilón + FSFix polishing**

**PacBio Assembly**

---

**Estimated genome size**
738 Mb

**Number of scaffolds**
677

**Scaffold N50**
22.7 Mb

**Longest scaffold**
36.3 Mb

**Assembly length**
715 Mb

**Assembly % of genome**
96.88%

**Repeat region %**
56.4%

**Predicted gene models**
45,335

**Average coding sequence length**
1700.4 bp

**Average exons per gene**
5.8
Durian – Evolutionary Relationships to T. cacao (Chocolate)
Durian Smell – Related to Expansions in Sulphur Processing
Durian Cultivars Have Distinct Genomic Signatures
Durian Biodiversity in South East Asia

Image sources:
https://www.flickr.com/photos/adaduitokla/12065514655
https://www.rarepalmseeds.com/images/DurDul2.jpg
https://www.yearofthedurian.com/2013/05/the-non-edible-durians.html
Acknowledgements

Apinya Jusukul
Chern Han Yong
Ioana Cutcutache
Jing Quan Lim

Chawalit Pairojkul
Raluca Gorden
Tatsuhiro Shibata

Bin Tean Teh

Steve Rozen

PRISM Team
- Stuart Cook
- Sonia Davila
- Saumya Jamuar
- Weng Khong Lim
- Yasmin Bylstra
- Jing Xian Teo
- Jyn Ling Kuan
- Ryanne Wu
- Jaydutt Digambar
- Khung Keong Yeo
- Siew Ching Kong
- Pei Yi Ho

Kevin Lim
Chern Han Yong
Cedric Ng