The Evolution of Cancer Suppression
Solutions to Peto’s Paradox Revealed by Genomic Analyses of Elephants and Whales

Marc Tollis, Ph.D.
Lisa Abegglen
Aleah Caulin
Andrea Cabrera
Richard E. Green
Nader Pourmand
Jooke Robbins
Per Palsson
Joshua Schiffman
Carlo Maley

Body Size and Macroevolution

- Cope’s Rule

Benson et al. (2014) PLOS Biology

Body Size and Macroevolution

Edward Drinker Cope


Body Size and Macroevolution

Body Size Evolution in Populations

- Fitness consequences → directional selection

Body Size and Cancer within Human Populations

- Caloric intake?
- Growth factors?
- More cells?

Kabat et al. (2013) Adult Stature and Risk of Cancer at Different Anatomic Sites in a Cohort of Postmenopausal Women.


No relationship between adult body mass and cancer incidence in mammals

"A man has 1000 times as many cells as a mouse... and we usually live at least 30 times as long as mice... However... the probabilities of carcinoma induction in mice and in men are not vastly different."

--Sir Richard Peto, enunciating Peto's Paradox

Life histories and Cancer Suppression as an Adaptation

Mouse = fast
High risk of predation
High reproductive output
Short lifespan

Elephant = slow
Low risk of predation
Fewer offspring
Long lifespan

Increased apoptotic response in elephants despite same levels of DNA damage

African Elephant Genome Contains an Extraordinary Expansion in TP53 Copies

- ~20 copies, 40 alleles
- Retrogenes, lack introns, flanked by MIRs
- Two major lineages
A apoptotic response is related to number of functional TP53 alleles

Abegglen et al. (2015) *JAMA*

Coevolution of body size and TP53 copy number in elephant lineage

Sulak et al. (2016) *eLife*

TP53 Retrogene expression

Abegglen et al. (2015) *JAMA*

Coevolution of body size and TP53 copy number in elephant lineage

Figure from Ho (2014) *TREE*

Bowhead Whale
*Balaena mysticetus*

Keane et al. (2015) *Cell Reports*

- Lives over 200 years
- 2nd largest whale
  - ~76,000 kg (~83 tons)
- Accelerated evolution in many genes associated with aging and cancer risk:
  - Suppressor of cytokine signaling (SOCS2)
  - Aprataxin (APTX)
  - Noggin (NOG)
  - Leptin (LEP)
  - Forkhead box o3 (FOXO3)
  - Complementation group 3 (ERCC3)
  - Fibroblast growth factor receptor 1 (FGFR1)
**Humpback Whale Input Assembly Sequence Data**

<table>
<thead>
<tr>
<th>Library</th>
<th>estNumReads</th>
<th>Q20 Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>180bp paired end</td>
<td>379,140,000</td>
<td>8.9</td>
</tr>
<tr>
<td>180bp paired end</td>
<td>418,300,000</td>
<td>9.6</td>
</tr>
<tr>
<td>180bp paired end</td>
<td>413,880,000</td>
<td>9.5</td>
</tr>
<tr>
<td>300bp paired end</td>
<td>25,820,000</td>
<td>0.8</td>
</tr>
<tr>
<td>500bp paired end</td>
<td>112,400,000</td>
<td>3.4</td>
</tr>
<tr>
<td>600bp paired end</td>
<td>395,500,000</td>
<td>9.1</td>
</tr>
<tr>
<td>2kb mate pair</td>
<td>348,080,000</td>
<td>4.2</td>
</tr>
<tr>
<td>10kb mate pair</td>
<td>279,000,000</td>
<td>6.6</td>
</tr>
</tbody>
</table>

47mer depth of 32.7X (using paired end libraries only)

**Humpback Assembly Stats**

<table>
<thead>
<tr>
<th></th>
<th>Input Assembly [Meraculous]</th>
<th>Devial Assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Length</td>
<td>2.3 Gb</td>
<td>2.3 Gb</td>
</tr>
<tr>
<td>No. scaffolds</td>
<td>24,319</td>
<td>2,558</td>
</tr>
<tr>
<td>Contig N50</td>
<td>12.4 kb</td>
<td>12.3 kb</td>
</tr>
<tr>
<td>Scaffold N50</td>
<td>198 kb</td>
<td>9,100 kb</td>
</tr>
<tr>
<td>No. scaffolds at N50</td>
<td>3,214</td>
<td>79</td>
</tr>
<tr>
<td>Scaffold N90</td>
<td>53 kb</td>
<td>2,400 kb</td>
</tr>
<tr>
<td>No. scaffolds at N90</td>
<td>11,081</td>
<td>273</td>
</tr>
<tr>
<td>Percent genome in gaps</td>
<td>5.4%</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

Estimated Chicago physical coverage: 93.2X

**Estimating the Gene Space in the Humpback Genome by Core Eukaryotic Genome Gene Analysis**

<table>
<thead>
<tr>
<th>Species</th>
<th>CEGMA complete</th>
<th>CEGMA partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog genome</td>
<td>98.0%</td>
<td>99.6%</td>
</tr>
<tr>
<td>Macaque genome</td>
<td>96.0%</td>
<td>100%</td>
</tr>
<tr>
<td>Cat genome</td>
<td>58.1%</td>
<td>75.8%</td>
</tr>
<tr>
<td>African elephant genome</td>
<td>46.0%</td>
<td>63.5%</td>
</tr>
<tr>
<td>Humpback whale genome</td>
<td>91.1%</td>
<td>96.8%</td>
</tr>
</tbody>
</table>

CEGMA: Parra et al. (2009) Nucleic Acids Research

**BUSCO results for humpback whale genome**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>BUSCO results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humpback whale genome</td>
<td>C:85% [D:1.5%], F:10%, M:4.9%, n:3023</td>
</tr>
<tr>
<td>Human genome</td>
<td>C:89% [D:1.5%], F:6.0%, M:4.5%, n:3023</td>
</tr>
<tr>
<td>Mouse genome</td>
<td>C:78% [D:3.0%], F:19%, M:2.5%, n:3023</td>
</tr>
<tr>
<td>Green anole lizard genome</td>
<td>C:73% [D:1.4%], F:15%, M:8.9%, n:3023</td>
</tr>
<tr>
<td>Painted turtle genome</td>
<td>C:76% [D:1.4%], F:15%, M:8.9%, n:3023</td>
</tr>
<tr>
<td>D. melanogaster genome</td>
<td>C:98% [D:6.4%], F:0.6%, M:0.3%, n:2675</td>
</tr>
<tr>
<td>C. elegans genome</td>
<td>C:85% [D:6.9%], F:2.8%, M:11%, n:843</td>
</tr>
</tbody>
</table>

Simão et al. (2015) Bioinformatics
Estimating the Number of Repeats in the Humpback Whale Genome

RepeatMasker with a mammal-specific RepBase library
- Total masked: 38.9%
  - SINEs: 6.1%
  - LINEs: 19.5%
  - LTRs: 6.3%

RepeatModeler: de novo ID
- Total masked: 30.2%
  - SINEs: 3.3%
  - LINEs: 19.1%
  - LTRs: 4.2%

Humpback Whale Annotation
- MAKER2 pipeline
  - Skin RNA-seq
  - Protein homology (10 datasets)
  - Ab-initio gene models (SNAP, Augustus)
- Results in 24,140 protein coding genes
  - Includes 90% of vertebrate BUSCOs

UCHL3 was duplicated in the ancestral cetacean

Accelerated Evolution in Humpback Whale Tumor Suppressor Genes
- Pairwise Ka/Ks with human orthologs
- 412 genes have Ka/Ks > 1
- 20 are known human TSGs

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIF</td>
<td>Leukemia inhibitory factor</td>
<td>Upregulates MDM2</td>
</tr>
<tr>
<td>MDM2</td>
<td>Mouse double minute 2 homolog</td>
<td>Negative regulator of TP53</td>
</tr>
<tr>
<td>TP53BP2</td>
<td>Tumor protein P53 binding protein 2</td>
<td>Stimulates apoptosis by binding to TP53</td>
</tr>
</tbody>
</table>

Future for humpback whale
- Single copy orthologs analysis
  - Mary O’Connell (Leeds UK) and Andrew Webb (Rutgers Univ.)
  - Dn/ds pipeline of 30 mammalian genomes
- Finding duplicated regions in humpback whale
  - WSSD and WGA methods
- Demographic analysis (PSMC)

Conclusion
- Evolution has solved the problem of cancer multiple times
- Likely independent mechanisms of cancer suppression in elephants and cetaceans
- Utilizing the shared genetic code to illuminate potential avenues of cancer prevention
Acknowledgements

University of California San Francisco
Aleah Caulin
Arizona State University
Carlo Maley
University of Utah
Josh Schiffman
Lisa Abegglen

Funding sources:
National Institutes of Health
Arizona State University
Hunstman Cancer Institute
University of Utah
Ringling Bros and Barnum and Bailey Circus
University of Groningen

University of Groningen
Per Patapoli
Andrea Cabrera

University of California Santa Cruz
Ed Green
Nader Pourmand
Center for Coastal Studies
Joel Robbins

Dovetail Genomics
PAG XXV workshop organizer:
Aaron Avivi

Sulak et al. (2016) eLife