Systematic Analysis of Genetic Variation of Duchenne Muscular Dystrophy and Implication for Cancer

Hubert Chen
West Windsor-Plainsboro South High School, NJ
Mentor: Pingzhang Wang, Betty Wang

3/2/2020
Duchenne muscular dystrophy (DMD)
disease overview

- DMD is a rare, severe, progressive genetic disorder causing disability and premature death.

- Mutations in DMD gene, encoding dystrophin protein, lead to DMD.

- DMD primarily affects boys. The prevalence is approximately 1 in 3500 to 5000 male births worldwide.

- DMD symptom onset usually between ages 3 and 5 years.

- Phenotypic variations in DMD may also occur in patients with same primary mutation due to secondary genetic modifiers.

Reference source: CureDuchenne
Duchenne muscular dystrophy (DMD) gene overview

- **DMD**: One of the largest known human gene, spanning 2.4 Mb genomic sequence.
- **DMD gene consists**: 79 exons encoding a 14,000 bp messenger RNA transcript.
- **DMD patients**: Protein translation is stopped prematurely.
  - Frame-shifting mutations (e.g. deletion of exons 47–50, Figure 2-A, top panel)
    - Lead to inclusion of aberrant amino acids
    - Generally premature truncation of translation.
  - Alternatively, a point mutation (nonsense mutation)
    - Can change an amino acid codon into a stop codon (Figure 2-B, bottom panel)

Figure 2: Schematic depiction of dystrophin transcripts in healthy and DMD

Dystrophin protein and dystroglycan complex overview

- Full length of dystrophin protein consists of 3,685 amino acids with 427 kDa.

- It consists of four major functional domains: actin-binding Nterminal domain (encoded by exon 1-8), central rod domain (encoded by exon 8-61), cysteine-rich domain (encoded by exon 62-69) and C-terminal domain (encoded by exon 69-79).

- Cysteine-rich domain together with C-terminal domain interact with different proteins including β-dystroglycan, syntrophin and dystrobrevin to make up dystrophin-glycoprotein complex.

Reference source: ResearchGate.net

Figure 3: Dystrophin and dystrophin-associated glycoprotein complex
Research objectives and workflows

- Carry out a systematic analysis of the DMD genetic variants via dbSNP database
- Explore protein-protein interactions for genetic modifiers identified in DMD patients
- Investigate potential relationships of genetic alternations in the DMD gene with cancer

Figure 4: Research workflows
Research materials and methods

- Extract DMD genetic variants via dbSNP Database with variant call format (VCF)
- Functional annotation with wANNOVAR: Variant prioritization (Figure 5)
- Retrieve the longest transcript ENST0000035703 (Figure 6)

- Focus on variants in exonic (coding) region can alter the protein function

Figure 5: Genomic catalog in DMD gene

Figure 6: R coding for the longest transcript
Research materials and methods

- Protein–protein interactions (PPI) map for genetic modifiers identified in DMD patients was constructed using STRING v11. Subsequently analyzed using Cytoscape 3.8.1 plugin Network Analyzer.

- Genetic alternations in the DMD gene with cancer was examined by using cBioPortal.
  
  - Data from 25 published TCGA cancer studies and 4 pediatric cancer studies that included a minimum of 100 samples. One study that reported 43 rhabdomyosarcomas cases has also been included.
  
  - Total 11927 patients (age from ~ 3 years to 90 years; ~ 48% male and ~ 46% female; ~ 60% White, ~7% black or Africa America and ~ 5% Asian).
  
  - Kaplan-Meier curves were stratified by genotype and comparisons were tested using the Log-rank test.
Variants type and frequency of amino acid change in the DMD gene

Table 1: Examples of DMD gene mutation

<table>
<thead>
<tr>
<th>Ref</th>
<th>Alt</th>
<th>Ref</th>
<th>Alt</th>
<th>Ref</th>
<th>Alt</th>
<th>Ref</th>
<th>Alt</th>
<th>Ref</th>
<th>Alt</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>A</td>
<td>-</td>
<td>GGT</td>
<td>AAAC</td>
<td>CTT</td>
<td>-</td>
<td>G</td>
<td>AAT</td>
</tr>
<tr>
<td>AAAGACTTC</td>
<td>-</td>
<td>AAAC</td>
<td>TCCAAG</td>
<td>CC</td>
<td>ACTGAT</td>
<td>-</td>
<td>TGA</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>AACGGGACT</td>
<td>-</td>
<td>ACCATGTGAG</td>
<td>TT</td>
<td>A</td>
<td>AGG</td>
<td>-</td>
<td>ATC</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>AACTGTCT</td>
<td>-</td>
<td>AGAC</td>
<td>GGT</td>
<td>-</td>
<td>ACA</td>
<td>AGA</td>
<td>-</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>AG</td>
<td>-</td>
<td>AT</td>
<td>AGA</td>
<td>-</td>
<td>GGACGA</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>ATAA</td>
<td>-</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>T</td>
</tr>
</tbody>
</table>

** Ref: Original nucleotide(s) present before mutation  
Alt: Alternative nucleotide(s) present after mutation

Figure 7: Frequency of amino acid change in DMD gene

- Insertion, deletion, substitution that cause frameshift changes in protein coding sequence.
- The largest category: nonsynonymous, follow by synonymous and stop gain.
Distribution of SNPs by exonic region

- SNPs distributed across almost all exons. Exon 79 is the longest with 2703 bp in length. Exon 78 is the shortest with 32 bp.
- **Normalized** exon length, then Exon 19 has most density of pathogenic SNP distribution.

Figure 8: Exon 78 and Exon 79 length example of DMD gene ENST00000357033

<table>
<thead>
<tr>
<th>Exon</th>
<th>Start</th>
<th>End</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>31144759</td>
<td>3114790</td>
<td>32</td>
</tr>
<tr>
<td>79</td>
<td>31137345</td>
<td>31140047</td>
<td>2703</td>
</tr>
</tbody>
</table>

Figure 9: ClinVar classification VS Exon frequency for DMD Transcript ENST00000357033
Distribution of SNPs by ACMG-AMP classifications

- Nonsense mutation (i.e. stopgain) or frameshift mutation likely lead to more pathogenic.
- Observed a few cases with synonymous mutation (2%) also associated with pathogenicity.
- Interestingly, some pathogenetic variants were also observed in healthy individuals.

Figure 10-a: Partial examples of DMD gene variants by ACMG-AMP classifications

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>Ref</th>
<th>Alt</th>
<th>ExonicFunc.ensGene</th>
<th>Exon</th>
<th>1000G AFR</th>
<th>1000G AMR</th>
<th>1000G EAS</th>
<th>1000G EUR</th>
<th>1000G SAS</th>
<th>COSMIC_DIS</th>
<th>ClinVar_SIG</th>
<th>ClinVar_DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>31196876</td>
<td>T</td>
<td>-</td>
<td>frameshift deletion</td>
<td>exon70</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>large intestine</td>
<td>Pathogenic</td>
<td>Duchenne muscular dystrophy</td>
</tr>
<tr>
<td>X</td>
<td>33229415</td>
<td>TT</td>
<td>A</td>
<td>frameshift substitution</td>
<td>exon1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>Pathogenic</td>
<td>Duchenne muscular dystrophy</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>32867917</td>
<td>C</td>
<td>G</td>
<td>startloss</td>
<td>exon3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>Pathogenic</td>
<td>Duchenne muscular dystrophy</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>31196906</td>
<td>TCT</td>
<td>-</td>
<td>nonframeshift deletion</td>
<td>exon70</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>Pathogenic</td>
<td>Duchenne muscular dystrophy</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>31514988</td>
<td>G</td>
<td>A</td>
<td>stopgain</td>
<td>exon57</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>Pathogenic</td>
<td>Dilated cardiomyopathy 3B</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>32305668</td>
<td>T</td>
<td>A</td>
<td>stopgain</td>
<td>exon43</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>Pathogenic</td>
<td>Dilated cardiomyopathy 3B</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10-b: SNPs by ACMG-AMP Classification Example in DMD

<table>
<thead>
<tr>
<th></th>
<th>Stopgain</th>
<th>Frameshift deletion</th>
<th>Nonsynonymous SNV</th>
<th>Frameshift Insertion</th>
<th>Synonymous SNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathogenic</td>
<td>151</td>
<td>57</td>
<td>26</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Likely Pathogenic</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Benign</td>
<td>0</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Likely Benign</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Uncertain Significance</td>
<td>1</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Interaction network resulting from the genetic modifiers identified in DMD patients

- Only interactions with confidence score over 0.9 were mapped to network.

- SPP1 interact with DMD through ITGB1, which has highest node degree and BC values in the network.

- Among the “seed” genetic modifiers, THBS1 has higher network topological parameters, followed by SPP1, ACTN3 and LTBP4. The network enrichment p-value was < 3.09e-08.

Reference source: STRING Consortium 2000
DMD genetic alterations using cBioPortal data

Figure 12: Frequency of GMD genetic alternations in different types of tumors

- The majority of genetic alterations corresponded to mutations, deep deletions.
- Low frequency of gene amplifications.
- DMD alterations were not found in samples (rhabdomyosarcomas)

Reference source: cBioPortal for Cancer Genomics
Cancer Patients with DMD alterations have poorer overall survival

- Other tumor types e.g. ovarian carcinomas showed similar trend, but not statistically significant.
- The majority of tumor specimens had lower DMD expression compared to the normal adjacent tissue.
- The relationship between DMD genetic status and prognosis may be tumor-type specific.
Conclusions

- To our knowledge, this is the first data mining study with a systematic analysis of all exon variants, especially SNPs, in the one of the largest known human gene. 

- This study examined total 3,627 exonic SNPs in the DMD gene. Nonsynonymous account for nearly 64% of all mutations. Exon 19 appeared to have most density of pathogenic SNP distribution. Nonsense mutation (i.e. stopgain) or frameshift mutation likely lead to more pathogenic.

- According to 1000 Genomes project, genetic variants (i.e., nonsynonymous mutation) associated with relatively higher alteration frequency in African. Similar frequency distributions were observed among America, Europe, East and South Asia.

- Protein network analysis highlighted non-random interconnectivity between the genetic modifiers identified in DMD patients, and potentially shed light on new genetic modifiers by their functional coupling to these known genes.

- This study result also suggest DMD gene may serve as a diagnostic and therapeutic target for certain types of cancer.


