“Le grand et le petit”: splicing factors SF3B1 and SUGP1 and their cancer mutations leading to aberrant acceptor usage

Tatiana Popova Institute Curie
CGC webinar 22.09.2021
DRUM team: DNA repair and Uveal Melanoma
Collaborative project with Seven Bridges CGC

- Project: Comprehensive detection and analysis of mutant-\textit{SF3B1} splice pattern

- Erik Lehnert

- Acknowledgements. The Seven Bridges Cancer Genomics Cloud has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, contract no. HHSN261201400008C and ID/IQ agreement no. 17 × 146 under contract no. HHSN261201500003I.
OUTLINE

• «Le grand» SF3B1 splicing factor and its cancer mutations
• Discovery of SUGP1 as a genocopy of SF3B1
• «Le petit» SUGP1 a new player in acceptor selection
• Sequence Bloom Tree (SBT) in deciphering aberrant splicing
RNA splicing is an important step in eukaryotic gene expression.
Splicing is orchestrated by ~150 proteins and require specific sequence organization

Most splicing factors mutated in cancer belongs to splicing initiation complex

Adapted from: Anczuków & Krainer DOI:10.1261/rna.057919.116
« Le grand » SF3B1 in splicing

SF3B1 is major subunit of U2 complex, which play multiple roles

Cretu C et al Mol Cell 2016
SF3B1 is abundant and plays a role in each splicing complex

<table>
<thead>
<tr>
<th>Spot</th>
<th>Protein</th>
<th>GI number</th>
<th>U1 snRNP</th>
<th>17S U2 snRNP</th>
<th>tri-SnRNP</th>
<th>&quot;A&quot; complex</th>
<th>B complex</th>
<th>B^ts complex</th>
<th>C complex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>kD</td>
<td>PAFs</td>
<td>Phospho</td>
<td>PAFs</td>
<td>Phospho</td>
<td>PAFs</td>
</tr>
<tr>
<td>Sm proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B*</td>
<td>1</td>
<td>24.6</td>
<td>gi</td>
<td>4507125</td>
<td>254</td>
<td>233</td>
<td>270</td>
<td>221</td>
<td>308</td>
</tr>
<tr>
<td>D1</td>
<td>2</td>
<td>13.3</td>
<td>gi</td>
<td>5902102</td>
<td>145</td>
<td>123</td>
<td>225</td>
<td>264</td>
<td>214</td>
</tr>
<tr>
<td>D2</td>
<td>3</td>
<td>13.5</td>
<td>gi</td>
<td>29294624</td>
<td>129</td>
<td>98</td>
<td>205</td>
<td>279</td>
<td>356</td>
</tr>
<tr>
<td>D3**</td>
<td>4</td>
<td>13.9</td>
<td>gi</td>
<td>4759160</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>10.8</td>
<td>gi</td>
<td>4507129</td>
<td>201</td>
<td>147</td>
<td>295</td>
<td>213</td>
<td>434</td>
</tr>
<tr>
<td>G</td>
<td>7</td>
<td>8.5</td>
<td>gi</td>
<td>4507133</td>
<td>192</td>
<td>132</td>
<td>284</td>
<td>212</td>
<td>374</td>
</tr>
<tr>
<td>U1 snRNP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1-70K</td>
<td>9</td>
<td>31.3</td>
<td>gi</td>
<td>4759156</td>
<td>186</td>
<td>205</td>
<td>202</td>
<td>418</td>
<td>336</td>
</tr>
<tr>
<td>U1-A</td>
<td>10</td>
<td>17.4</td>
<td>gi</td>
<td>4507127</td>
<td>97</td>
<td>122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1-C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17S U2 snRNP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U2A'</td>
<td>11</td>
<td>28.4</td>
<td>gi</td>
<td>50593002</td>
<td>215</td>
<td>169</td>
<td>216</td>
<td>181</td>
<td>115</td>
</tr>
<tr>
<td>U2A''</td>
<td>12</td>
<td>25.4</td>
<td>gi</td>
<td>4507123</td>
<td>114</td>
<td>222</td>
<td>251</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>SF3b120</td>
<td>13</td>
<td>88.9</td>
<td>gi</td>
<td>5032067</td>
<td>186</td>
<td>175</td>
<td>159</td>
<td>128</td>
<td>33</td>
</tr>
<tr>
<td>SF3a66</td>
<td>14</td>
<td>49.3</td>
<td>gi</td>
<td>2361376</td>
<td>117</td>
<td>100</td>
<td>134</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>SF3b60</td>
<td>15</td>
<td>58.8</td>
<td>gi</td>
<td>5033167</td>
<td>221</td>
<td>267</td>
<td>240</td>
<td>204</td>
<td>71</td>
</tr>
<tr>
<td>SF3b155</td>
<td>16</td>
<td>145.8</td>
<td>gi</td>
<td>5412117</td>
<td>120</td>
<td>109</td>
<td>133</td>
<td>79</td>
<td>29</td>
</tr>
<tr>
<td>SF3b130</td>
<td>18</td>
<td>135.5</td>
<td>gi</td>
<td>5411212</td>
<td>148</td>
<td>128</td>
<td>131</td>
<td>128</td>
<td>44</td>
</tr>
<tr>
<td>SF3b49</td>
<td>19</td>
<td>44.4</td>
<td>gi</td>
<td>5032069</td>
<td>112</td>
<td>158</td>
<td>173</td>
<td>173</td>
<td>52</td>
</tr>
<tr>
<td>SF3b14a/p14</td>
<td>20</td>
<td>14.6</td>
<td>gi</td>
<td>7706326</td>
<td>76</td>
<td>105</td>
<td>162</td>
<td>198</td>
<td>50</td>
</tr>
<tr>
<td>SF3b14b</td>
<td>21</td>
<td>12.4</td>
<td>gi</td>
<td>14249398</td>
<td>81</td>
<td>145</td>
<td>175</td>
<td>203</td>
<td>63</td>
</tr>
<tr>
<td>SF3b10</td>
<td>22</td>
<td>10.1</td>
<td>gi</td>
<td>13775200</td>
<td>104</td>
<td>40</td>
<td>93</td>
<td>42</td>
<td>67</td>
</tr>
<tr>
<td>17S U2 related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tat SF1</td>
<td>23</td>
<td>85.7</td>
<td>gi</td>
<td>21361437</td>
<td>78</td>
<td>18</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPRPS</td>
<td>24</td>
<td>117.4</td>
<td>gi</td>
<td>41327773</td>
<td>75</td>
<td>103</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPF45</td>
<td>25</td>
<td>45</td>
<td>gi</td>
<td>14249670</td>
<td>102</td>
<td>126</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUF60</td>
<td>26</td>
<td>59.9</td>
<td>gi</td>
<td>17298690</td>
<td>126</td>
<td>56</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPFF30</td>
<td>27</td>
<td>26.7</td>
<td>gi</td>
<td>5032113</td>
<td>56</td>
<td>69</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U2AF65</td>
<td>28</td>
<td>53.5</td>
<td>gi</td>
<td>6005926</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U2AF35</td>
<td>29</td>
<td>27.9</td>
<td>gi</td>
<td>5803207</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Agafonov DE et al Mol and Cell Biology 2011
«Le grand» SF3B1 is the most frequently mutated splicing factor in cancer

SF3B1 mutations in the TCGA

SF3B1 mutations target HEAT domains

Cretu C et al Mol Cell 2016
Hotspot mutations in SF3B1 lead to aberrant acceptor usage

Alsafadi S et al Nat commut 2016
3’splice site aberrations are the main effect of SF3B1 cancer mutations

Alexandre Houy

Uveal melanoma series

Position of alternative 3’SS

SF3B1mut  SF3B1 wt

Alsafadi S et al Nat commut 2016
Mutated SF3B1 shifts the complex to alternative branch site

Shift to BPS’ leads to alternative 3’ splice site usage

To get affected the splice junction should have alternative BPS and corresponding alternative 3’ acceptor within 10-30nts.

Adapted from: Anczuków & Krainer 2016 DOI:10.1261/rna.057919.116
The starting point of the project: SF3B1-like

ToDo: Screen public databases for SF3B1-like cases and look for mutations
**Collaboration with Cancer Genome Cloud:**

**Sensitive screen for SF3B1-like pattern**

*Erik Lehnert*

<table>
<thead>
<tr>
<th>Sequence Bloom Tree</th>
<th>Index file for each RNAseq Fastq</th>
<th>Query: Aberrant junction</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBT</td>
<td>ACC, BLCA, BRCA, OV, STAD, UM.</td>
<td>TCTGCGGGGAGAGATGGAGGCCTCCGc, GCATCCTGTGGGCGGCACCGCTTGAGC, ...</td>
<td>yes, no, no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTTCA GTGTACGAGAATTTACTG</td>
<td></td>
</tr>
</tbody>
</table>

Data indexing structure that allow **FAST** screen for sequence occurrence

N of occurrence in each case: SBT score
Aberrant junctions from SF3B1mut from several tumor types

Michele Cornella

Alsafadi et al

Darman et al

~1000 junctions not present in transcript variants were tested by SBT
SBT scores, SF3B1mut and aberrant junctions

- SBT score is the number of aberrant junctions found
- High SBT score well corresponded to SF3B1mut
- High SBT score pointed SF3B1mut even in low tumor content samples
- ~10 cases with NO SF3B1mut!!!

All cases to the right from red line were tested directly using splice index
SUGP1 (SF4) alterations lead to SF3B1-like splice aberration

Principal Component Analysis on splice index

Only SUGP1 mutations with LOH give SF3B1-like phenotype ($p<10^{-8}$)
SUGP1 (SF4) alterations lead to SF3B1-like splice aberration

To conclude 7 cases are explained by alteration of SUGP1
HEK293T SUGP1\textsuperscript{KD} and SF3B1\textsuperscript{K700E} overexpression

Samar Alsafadi

Aberrant junctions with top difference (MAX - Control)

Phenotype was validated in experimental model
SUGP1 alterations mimic effect of SF3B1 mutations

- LUAD cohort from the TCGA: 5 SUGP1alt + 6 SF3B1mut +400 Controls
- ~100,000 junctions
- High inter-tumor heterogeneity
- Plenty of significant differences just by chance + possibly different histo types
- Supervised Principal Component analysis

1. Wilcoxon or t-test comparison:

   ![Histograms showing p-values](image)

   - More than by chance
   - Not more than by chance
   - More than by chance
2. Supervised Principal component analysis

Principal component analysis of differentially expressed junctions instead of multiple testing correction

Aberrant 3’ ss
Both are separated from Control
SUGP1 is not separated from SF3B1

Aberrant 5’ ss
Both are not separated from Control
SUGP1 is separated from SF3B1

Bair et al (Thibshirani) JASA 2006
3. No difference SUGP1alt vs SF3B1mut in 3’ss aberration in LUAD

No difference from Controls in 5’ss and exon usage in LUAD
«Le petit» SUGP1 a new player in acceptor selection

- SUGP1 is barely visible only in A complex

<table>
<thead>
<tr>
<th>Spot protein</th>
<th>gi number</th>
<th>U1 snRNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>kD</td>
<td>PAFs</td>
</tr>
<tr>
<td>TRAP3 (TRAP150)</td>
<td>g</td>
<td>107234419</td>
</tr>
<tr>
<td>CCR1</td>
<td>g</td>
<td>46852388</td>
</tr>
<tr>
<td>RBM5/LUCA15</td>
<td>g</td>
<td>5032031</td>
</tr>
<tr>
<td>S164 (ISAP94)</td>
<td>g</td>
<td>4050087</td>
</tr>
<tr>
<td>RBM10</td>
<td>g</td>
<td>12644371</td>
</tr>
<tr>
<td>SF1</td>
<td>g</td>
<td>42544130</td>
</tr>
</tbody>
</table>

Abundant first or only present in A complex

- Function of SUGP1 was not established

- Mutated SF3B1 shifts the complex to alternative branch site


- How this concept is changed with discovery of SUGP1?

Agafonov DE et al Mol and Cell Biology 2011
« Le petit » SUGP1 in splicing

Disease-Causing Mutations in SF3B1 Alter Splicing by Disrupting Interaction with SUGP1

Zhang et al

- 4 missense mut around G-patch domain
- 1 deleterious with alternative start
- 2 extremely reduced (but!) expressed

We see hypomorphic mutations, seems like cell needs some functional SUGP1
Le grand et le petit: the model

- SF3B1mut always heterozygous
- SUGP1alt are homozygous
- SUGP1alt do the correct job in the majority of exons!
- Or SUGP1 is only needed in the processing of “sensitive” junctions?

Exact role of SUGP1 in splicing complex stabilization is still not clear!
SBT score for estimation of the strength of SF3B1 mutation

**SBT score is not very much affected by these biases:**
- SBT score show nice correlation to Million reads per mutated allele (~0.8).
- No obvious effect of the tumor content (the size of mutation ID is proportional to the tumor content)
- Tissue bias is also moderate (blue and red represent different tissues)

**SF3B1mut effect is biased by:**
- RNA-seq parameters
- Tissue-specific gene expression
- Abundance of mutated allele
- Tumor content, etc.

Million reads per mutated allele = N junction reads x VAF
Using the linear model for SF3B1 mutation effect estimation

Model (green) on R625 and K700 mutations (blue and red) 
\[ p < 10^{-11} \]

Deviation: % from the model value

Conclusions and Follow-up

- Thanks to collaboration with CGC sensitive screening of the TCGA helped revealing *SUGP1* as a “cancer” genocopy of *SF3B1*

- One case of AML with clear SF3B1-like splice aberration is not deciphered yet

- Studying splicing defects in cancer have a fundamental and translational values. First, it clarifies the gene function and second it have a therapeutic implications for immuno- and targeted therapy
Thank you!!!


O Delattre & U830 & Bioinfo court

Seven Bridges Cancer Genomics Cloud

A Zinovyev, M. Kondratova & Sysbio