Review of "DNA Methylation analysis reveals distinct methylation signatures in pediatric germ cell tumors" by Amatruda et al.

Pediatric germ cell tumors are a heterogeneous group of tumors with a spectrum of clinical behavior. Survival of children with pure germinomas and teratomas treated with current approaches is excellent; however, the behavior of mixed germ cell tumors including teratoma with immature features may be unpredictable.

In this study, the authors apply global array-based methylation profiling to examine 51 pediatric germ cell tumors which include germinomas, Teratomas and Yolk Sac Tumors. Using an informatics based approach they define 7 molecular sub-classes of GCT with distinct methylation signatures that largely segregate with tumor histology. They identified 233 CpGs with increased enriched methylation in YSTs and show through pathway analyses that these map to loci with functions in ESC differentiation and WNT signaling. They demonstrate using comparative methylation analyses of a small sub-group of tumors with matched normal tissue, that target loci identified in YSTs were tumor-specific changes. They used a similar approach to examine methylation signatures in mature and immature teratomas. Although they did not find a correlation of methylation signature with teratoma histology, they identified a subset of differentially methylated CpGs that were significantly hypomethylated in mature versus immature teratoma. Notably, a number of pathways were related to stem cell biology.

General comments:
Although similar analyses have been reported (references 5 and 9), this study has the potential to corroborate and extend prior findings that may enable more precise distinction of GCT sub-types and predicate clinical behavior specifically of tumors classified as mature and immature teratomas. However, although the authors set out to investigate the molecular differences between mature and immature teratoma, they have not exploited their data to fully answer this important question which has the potential to make their study novel.

Major revisions:
The manuscript in its present form needs more detailed and robust data analyses
as well as corroboration of methylation studies with gene or protein expression data.

Specific comments:
1. Figure 1 – it’s not clear whether an unsupervised or supervised cluster analyses was performed on the tumour cohort. If the analyses is unsupervised then, a dendrogram in figure A would have been very helpful to assess the specific relationship of each tumour to the tumour molecular sub-groups. The authors indicate that they identified 8 sub-classes of GCT (only seven are shown in the Fig 1). They report an interesting sub-set of genes that define the YSTs, however the data could have been significantly strengthened by correlating significance of differential methylation with gene or protein expression data.

2. The authors observed discrepancies in molecular and histologic classification of the teratomas. This potentially very interesting finding is not fully explored. Was tumour histology re-reviewed after the molecular analyses. The authors performed comparative analyses of methylation signatures of the mature versus immature teratomas – its not clear whether these were performed on histologic sub-groups or molecular sub-groups and may explain the lack of significance in their comparative analyses. Finally, it is very interesting that Sox2 – a pluripotency gene is differentially enriched in pathways analyses of immature versus mature teratomas. Corroboration of these findings at the gene or protein expression level may potentially provide a diagnostic marker to distinguish or identify mature teratomas with foci of immature cells that may not be readily detected by routine morphology.

Minor Essential revisions:
The manuscript could also be significantly enhanced by careful editing. Several discrepancies in numbers of tumours or sub-groups analysed are noted.

The authors indicate that 5 CpG Loci with significant differences by tumor histology (q value <2.2E-16 and fold change >2.50 ) were selected for validation by pyrosequencing. Three of the loci selected appeared in table 2, which lists the top 23 loci with the most significant Q values and fold changes. The remaining loci are not among the top 23 listed in Table 2. Table 2 should include all loci with >2.50 fold change if they are referring to these values.

In the 4th paragraph, the authors are unclear as to the genes mentioned, whether they were found in the study or only in IPA. What is the significance of these genes are there promoters regions hyper or hypomethylated? If no analysis is being performed on this gene lists, perhaps it should be moved to the discussion section of this paper.

Level of interest: An article whose findings are important to those with closely related research interests
**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.