RESPONSE LETTER

Reviewer #1 (Josh's review)

-- Ref 1.0 Clarifying terminology–--

Reviewer Comment	This is a very important and novel angle on interpreting the PCAWG dataset. I expect it will be read with much interest. I do find the terminology to be very confusing to follow. The definitions of passengers and drivers get adorned and blurred. I appreciate that it would seem very tricky to find the correct term to describe "impactful passengers," which itself seems tautological. There seems to be a spectrum between drivers and passengers and, if we believe the results here, another 3rd class in the middle of the two. I found myself wondering what the difference really is between a weak driver and an impactful passenger…? The authors might play with it a bit more until they find terminology that sounds a bit more sensical. I admit that I also do not have a clear idea on what terms should be used.
Author Response	We thank the reviewer for pointing out issues related to our terminology. We would like to point out that in the paper, as much as possible, we've tried to stay with accepted terminology. All of the terminology that we use in the paper has been previously published in literature. We are not inventing new terminology, but rather, trying to use what people have already developed. In particular, terms such as weak drive or mini driver, deleterious passengers have been used already in the literature. We're simply trying to meld them together, here. We're very open to other terminology suggestions, and we would be happy to change the terminology in the paper if people in the steering committee think this makes sense. In particular, mutations with weak effects on fitness are assumed to have a negligible impact on tumor growth and are termed here as "nominal passengers" (i.e. all mutations other than drivers in a cancer). In contrast, an "impactful nominal passengers" here refers to a subset of "nominal passengers", which have high predicted molecular impact scores and thus might play a role in tumor growth. We suggest that, through aggregated effects, such mutations can play weak driver roles and thus be subject to weak positive selection (or negative selection in the case of deleterious passengers). In this manuscript, we considered weak drivers as a subset of impactful passengers.
Excerpt From Revised Manuscript	

Deleted: Majority of

Deleted: used in our manuscript is borrowed from prior

Deleted: cite these references. For the remaining terminologies, we define them upfront and use these consistently throughout the text to avoid any confusion. We also provide detailed definition of all these terms in our supplement section

Deleted: Mutations

Deleted: . ... [1]

-- Ref 1.1 – Background model--

Reviewer Comment	The major sticking point of this work is the definition of the background model that determines their noncancerous set. The authors state that these are created by shuffling the mutations while preserving various features such as overall burden and mutation signatures. While this is convincing, one wonders if there are other possible confounders since the genome is so architected. Could the authors at least address the role different chromatin states might (or might not?) influence the null model? It could be that if they accounted for these effects that the entire signal would drop away. But even then, perhaps the authors could argue that the architecture itself somehow influences the accumulation of such passengers?
Author Response	We thank the reviewer for identifying potential limitations in the background model. In the previous version of this work, we applied a nonparametric null model, which preserves the signature and local burden. We believe this to implicitly reflects chromatin states as well as many other covariates whose influence clusters locally. This null model has been applied across the PCAWG project. However, based on suggestions by all reviewers, we now also apply a null model that considers additional covariates *explicitly* such as chromatin openness, replication timing, and GC content. We observe a slight increase in the variance explained by our model with these explicit covariates, suggesting the signal is robust to the effects mentioned. Our updated randomized model clearly show that our result is robust to the randomized model and we don't lose any signal due to co-variate correction. However, we think that we should stay with the original PCAWG randomization model to report our final result. We think that the key strength of the PCAWG project is having a consistent presentation of signatures, background models, drivers, and mutation calls across a very large data set. We think deviating from this will seemingly hurt the PCAWG presentation in the wider community.
Excerpt From Revised Manuscript	

Deleted: [[SK2MG: We still need to figure out if we are going to stick with Inigo's simulation or the moat-sim simulation, which has better signal for additive variance analysis.]]

-- Ref 1.2 – Comments on the setup of the paper--

Reviewer Comment	Abstract and Introduction are very well written and intriguing. Pitched with the right amount of background, motivation, controversy, and reservation.
Author Response	We thank the reviewer for this positive comment.
Excerpt From Revised Manuscript	

-- Ref 1.3 –Clarifying random effects model parameters--

Reviewer Comment	I believe the phenotype variable y[j] records if the sample is a sample from PCAWG (y[j]=1) versus a randomly generated sample (y[j]=0). The authors should make this just a little more explicit.
Author Response	As per suggestion, this is now explicitly defined in the updated text.
Excerpt From Revised Manuscript	

-- Ref 1.4 – Rationale for using random effects model --

Reviewer Comment	Page 3. Not clear why the particular model used was chosen. Is this standard from the GWAS community or is it the idea of the authors? Either provide a citation or refer the reader to the appropriate part of the supplement that gives justifications for the form chosen.
Author Response	This particular model is commonly used in complex trait analysis – a recent approach (PMC3232052) which used it to explain variability in human height has been adapted for many studies. The first supplemental note referred to on page 3 addresses suitability of this for somatic mutation in cancer. The model is appropriate for detecting the cumulative effect of variants, which may be individually weak.
Excerpt From Revised Manuscript	

-- Ref 1.5 – Related to LoF spectrum –

Reviewer Comment	I don't quite follow the argument and interpretations under the mutational signatures section. Are there certain signatures that will lead to stop codons? Which ones are these and can they be noted in Fig 3?
Author Response	Yes, reviewer's interpretation is correct. For LoFs, we exclusively look at mutational spectrum in the context of "nominal passenger" mutations altering stop codons. In Figure 3a, mutational spectrum is plotted for mutations leading to LOF events.
Excerpt From Revised Manuscript	

Deleted: ... [2]

Deleted: on tumorigenesis

Deleted:

Deleted: signature analysis --

Deleted: Or are the authors looking at associations of any mutational signature? I lose sight here of how their nominal passengers have been used in this analysis if at all.

Deleted: In the mutational signature section, we analyze the role of mutational signatures to the differential burdening of genomic elements by nominal passengers. We perform this analysis for the coding LoF mutations, as well as for non-coding mutations leading to TF motif break events. ... [3]

Deleted: ... [4]

-- Ref 1.6 – Related to signature analysis –

Reviewer Comment	Or are the authors looking at associations of any mutational signature? I lose sight here of how their nominal passengers have been used in this analysis if at all.
Author Response	In addition to LoF spectrum, we also compare the difference in mutational signature for "nominal passengers" with low and high impact scores for various PCAWG cancer cohorts (Figure 3c & 3D). Categorization of nominal passengers into low and high impact score groups was done based on functional impact score threshold.
Excerpt From Revised Manuscript	

-- Ref 1.7 – Clarifying comparative statements --

Reviewer Comment	There are quite a few comparative statements where one of the classes being compared is implied and this leads to an ambiguity and lack of clarity in following the logical arguments of the manuscript in many places. For example, statements like "As expected, we observe lower mutational heterogeneity among high impact nominal passenger SNVs," are hard to decipher because its not clear what is lower? To random? To low-impact nominal passengers? To drivers? What?
Author Response	The reviewer makes a good point. In the updated version, we explicitly state these comparisons. In the particular example the reviewer mentions; the intended comparison is to low-impact nominal passengers rather than to random set.
Excerpt From Revised Manuscript	

Reviewer #2 (Peter's comments)

-- Ref 2.0 Overall comment–--

Reviewer Comment	The paper is considerably improved from earlier versions we have seen. In particular, I like the general concept of estimating the size of the set of unobserved driver mutations using random effects models. This is potentially very powerful, but requires rigorous attention to detail in its specification.
Author Response	We thank the reviewer for the encouraging comments and highlighting the importance of this work.
Excerpt From Revised Manuscript	

-- Ref 2.1 – Clarifying the model--

Reviewer Comment	I didn't have the Supp Notes in reviewing this, but if I understand correctly, the model fitted is basically a GLMM? That is, basically a random effects logistic regression (in which case the formula y_ij = … should have the logit function specified)?
Author Response	The Model is a GLMM (General linear mixed model) with random effects, where the random effects (SNV effect sizes) are modeled as Gaussian distributed, and are not estimated directly but integrated across when estimating the variance explained by the SNVs. As in previous analyses which use this model for complex trait analysis (PMC3232052), we do not use a logit linking function as would be used in logistic regression, but use the linear outputs directly to predict the binary phenotype. The additive variance thus estimated is referred to as using the 'observed scale'. We also quote results on the 'liability' scale, which uses a probit link function.
Excerpt From Revised Manuscript	

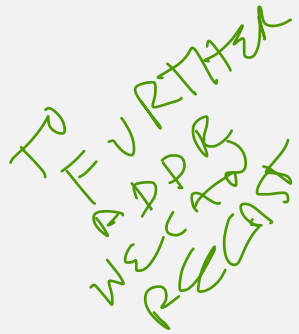
Deleted: . ... [5]

-- Ref 2.2 – Clarifying the model--

Reviewer Comment	It is difficult to assess without the Supp Methods, but how well have the authors tested / controlled for over-fitting? That is, the statement of 64.5% variance explained with all mutations versus 52.5% with drivers only could be due to the vastly larger numbers of data-points in the former analysis – even small mis-specifications could accumulate to lead to apparent better predictive power. Separate test-retest cohorts are essential here.
Author Response	The issue of overfitting on the SNV effect sizes does not arise in the random effects model, since the model is not trained as a predictive model. Instead, these parameters are integrated across in order to estimate the single hyper-parameter which controls their variance. In its simplest form, the model thus uses only one degree of freedom to model the SNV effects regardless of the number of SNVs, and hence does not suffer from overfitting as more SNVs are added. We verified this directly through simulations, which show that adding Poisson-distributed SNVs to the model lowers the estimated additive variance. While not fit as a predictive model, it is possible to derive a point estimate of the SNV effect sizes using the 'best linear unbiased predictor' in the random effects framework. Based on the reviewer's suggestion, we constructed multiple training/hold-out partitions of each cohort's data, and analyzed the predictive accuracy of the blup parameters for predicting the cancer phenotype on the hold-out partition after fitting on the training partition. The generalization accuracy consistently falls between

Deleted: .

Deleted: 1



Deleted: In addition, we also performed sensitivity analysis to cross-check the issue of over-fitting in our random effects model. We computed the additive variance of two random samples. If this model were overfitting the data, one would get non-zero additive variance for such a double random dataset. In contrast, we observe 0% additive variance in all cancer cohorts, suggesting no overfitting in our analysis.

	60-75% for all cohorts using our pooled model (with chance at 50%, and halved statistical power due to partitioning the data), with the generalization error negatively correlated with the additive variance as measured on the training partition (r=0.53, p=1.5e-6, where r is Pearson's correlation). This confirms empirically that the additive variance is an appropriate statistical estimator for differences in aggregate predictive potential. Additionally, we confirmed that the additive variance estimated on training and hold-out partitions was highly correlated (r=0.73, p=2.5e-14).
Excerpt From Revised Manuscript	

-- Ref 2.3 – Improving background model --

Deleted: 2

Reviewer Comment	I remain concerned about the generation of the null model samples. Any factor that influences true passenger mutation distribution that is not accounted for in the null model redistribution will have the potential to get picked up by the additive model as containing predictive power, but not for the 'functional impact' reasons, but rather for 'uncorrected information in mutation signatures' reasons. In particular, the following factors could well play a role and should be included in the null model redistribution: a) Replication timing (especially likely to have an effect) b) Intergenic versus intragenic; Gene expression c) Chromatin openness d) Replication and transcription strand (less concerned about this) e) Nucleotide context beyond trinucleotide (especially problematic for POLE hypermutators and also the UV light signature).
Author Response	We update our null model to include some of suggested covariates. We also use penta-nucleotide context for the melanoma and liver cancer cohorts.
Excerpt From Revised Manuscript	

-- Ref 2.4 – Background model related issues--

Reviewer Comment	Many of the downstream analyses depend critically on the accuracy of these models – even small inaccuracies in inferences can lead to quite large numbers of, for example estimated weak drivers or negatively selected mutations (Figure 5) when multiplying up by the total number of mutations and/or number of samples. This is particularly concerning for the melanoma and liver cancer findings in Figure 5. The authors will need to be scrupulous in reassuring the reader that these calculations are valid.
Author Response	We would like to clarify that while there has been a tremendous amount of discussion about background models from all the reviewer, the random model has been utilized in a very small fraction of our paper. In particular, only three of the 20 sub-figures in the paper involve explicit background models. Majority of our analyses simply involve looking at the overall functional impact of mutations, looking at the contrast in the overall impact of early versus late mutations, and underlying mutational signature. There's no model at all involved in this discussion. We can easily downplay or even remove a lot of discussion of background model in the paper without significantly affecting it. We think that, for reasons we don't fully understand, a lot of the discussion has become very preoccupied with the background model, which we think is tangential to the functional impact aspects of the paper. For few instances, where we do apply randomized model, we have updated the null model, which now corrects for covariates suggested by all reviewers. In addition, we apply additional filters (used in the main driver paper) on our putative list weak drivers and deleterious passengers. This provide a very conservative estimate on the number of weak drivers and deleterious passengers in Figure 5.
Excerpt From Revised Manuscript	

Deleted: 3

Deleted: our downstream analysis based on the new

Deleted: the

Deleted: the reviewer

Deleted: .

Deleted: Excerpt From ... [6]

Moved (insertion) [1]

-- Ref 2.5 – Terminological inconsistency --

Reviewer Comment	The Introduction and much of the paper suffers from terminological difficulties – for example, the phrase "…and large numbers of passengers (with weak or neutral fitness effects) is analogous to…" [Page 2] is not accurate. By definition, passengers cannot have weak fitness effects – this would make them drivers. It is, however, perfectly reasonable to ask whether mini-drivers exist, and what they might look like if they do.
Author Response	We address this issue in section Ref 1.0 of this response document.
Excerpt From	

Deleted:

Deleted: 4

Deleted: We agree with the reviewer that there are challenges in defining terminology. We explicitly use the term "nominal passenger" throughout the text. Nominal passengers correspond to all non-driver variants. These potentially include weak drivers, neutral passenger and deleterious passenger. In Figure5, we explained this in detail.

Moved (insertion) [2]

Revised Manuscript	

-- Ref 2.6 – Missing heritability relevance --

Reviewer Comment	I didn't find the GWAS analogy in the Introduction especially helpful. Whatever the truth about the driver – passenger – deleterious mutation landscape in cancer, it is clearly vastly different to the complex diseases studied by GWAS. Also, whether there is the somatic equivalent of "missing heritability" is moot – it is not at all clear to me that we are missing anywhere near the same proportion in cancers as in GWAS.
Author Response	In our updated text, we emphasize that additive variance in our model does not directly measure heritability as in the GWAS case, but rather the combined effects of SNVs on tumorigenesis. The GWAS analogy can be justified due to lack of power issues in PCAWG as well as prior complex trait studies. In complex trait analysis, it was strongly felt that there must exist many SNPs of low-penetrance or low effect because of the missing heritability problem. This motivated the development of the random effects model to measure the aggregate effect of variants in a statistically rigorous way. The driver discovery exercise in PCAWG suggests that we don't find many non-coding drivers in PCAWG cohorts, also due to limited sample size. Thus, we argue that if there is a long tail of low effect mutations in cancer, the random effects model is a powerful way to find their aggregated effect, where the additive variance measures the collective predictive power of the variants against an appropriate null model.
Excerpt From Revised Manuscript	

-- Ref 2.7 – Suggestion on modifying introduction --

Reviewer Comment	Instead, in the Introduction, I would sharpen the discussion about the methodological constraints of current approaches to driver-passenger dichotomisation and whether there might be fundamental limits to recurrence-based methods (especially given the long tail of cancer genes). Make the point that such limits might mean that an undefined proportion of rare drivers might be being mis-annotated as passengers, and with current sample sizes, we are likely only to be able to detect a global signal of their existence (but this is hugely challenging), rather than identify each one specifically.
Author Response	We thank the reviewer for this suggestion. We include some of these points in our updated text.
Excerpt From Revised Manuscript	

Moved up [1]: Excerpt From ... [7]

Deleted: 5

Moved (insertion) [3]

Deleted: The Driver discovery exercise in PCAWG suggest that we don't find many non-coding drivers in PCAWG cohorts as well as on pan-cancer level. This can be potentially attributed to the lack of power for identifying non-coding driver due to limited sample size. In this context, our additive variance analysis provides an alternative approach to address this issue. ... [8]

Deleted: If there are variants of low effect in cancer, the random effects model is a powerful way to find their aggregated, where the additive variance here measures the collective predictive power of the variants against an appropriate null model.

Moved up [3]: In our updated text, we emphasize that additive variance in our model does not directly measure heritability as in the GWAS case, but rather the combined effects of SNVs on tumorigenesis.

Moved up [2]: Excerpt From ... [9]

Deleted: 6

-- Ref 2.8 – Suggestion on modifying introduction --

Reviewer Comment	Furthermore, make the point that selection acting on somatic cells is dynamic, and that just because something is a passenger now, doesn't mean that it couldn't become a driver when treatment is given or the clone spreads to another organ – therefore of value to study how many of them have functional effects, even if these functional effects are not of selective consequence currently
Author Response	We thank the reviewer for this suggestion. We include some of these points in our updated text.
Excerpt From Revised Manuscript	

Reviewer #3 (Gaddy's review)

-- Ref 3.0 Issue with background model–--

Reviewer Comment	In their paper "Passenger mutations in 2500 cancer genomes: Overall molecular functional impact and consequences", Kumar et al. present a very intriguing analysis of putative passenger mutations and their potential functional impact in cancer. Overall, the paper presents a very interesting analysis with important implications to our understanding of positive and negative selection in cancer. My main concern about this paper remains the uncertainty in our background models. The use of the additive effect model is very interesting. However, as opposed to GWAS studies where one uses cases and controls, here the authors use a simulated dataset of non-cancer "neutral" cases using a mutation randomization scheme.
Author Response	Please see section "Ref 2.3" for the detailed response.
Excerpt From Revised Manuscript	

Deleted: 6

Deleted: . ... [10]

Deleted: We agree with the reviewer's comment regarding potential limitations in the background model. As noted earlier, we update our background model to address some common issues brought up by all reviewers.

-- Ref 3.1

– Some limitations in current background models--

Reviewer Comment	My concern is that we (the community) are far from understanding the background model and even our best randomization scheme does not truly reflect the underlying processes which can vary at the single-base level (as opposed to the 10-100kb resolution of the randomization scheme). As we are seeing in the PCAWG drivers paper, even after controlling for many effects, there are still significant hits that are false positives due to many effects, such as (i) ignoring the lack of repair due to DNA binding proteins (most prominent in melanoma); (ii) inaccurate modeling of AID and APOBEC processes; (iii) Not taking into account the amount of DNA available for mutagenesis during tumor development (e.g. if a chromosomal region is lost early on, the rate of mutations in that region will be lower since there is only one copy to mutate); (iv) our ability to detect mutations as a function of sequence coverage (ie. GC-content affects coverage and our ability to detect mutations). Promoters typically have high GC-content, hence we will detect fewer mutations in them (whereas the simulated datasets will not have the same decrease in density); (v) ignoring local DNA structure such as palindromes and other sequence motifs; (vi) ignoring the association between signatures and timing; and (vii) ignoring different repair processes that operate different at various scales and in different genomic regions (introns, exons, intragenic regions, early vs. late replication timing). The uncertainty and inaccuracy of the background model can lead to incorrectly reaching the conclusion of positive selection or negative selection (depending on the genomic regions and mutational processes that operate in the specific tumor type).
Author Response	Thanks to these comments, our updated background model now incorporates many of these features explicitly. We present results from both models in the supplement and main text. By comparing the results from the two models, the reader can appreciate how further, yet undiscovered covariates might continue to confound the results. Furthermore, we apply some of the suggested filters on our putative list of weak drivers and deleterious passengers to provide a more conservative estimate. In addition, we also mention these limitations explicitly in our discussion section.
Excerpt From Revised Manuscript	

Moved down [4]: Sensitivity analysis of additive variance model---
Calculate additive variance for double randomized sample set, i.e., use two iteration of random sample to calculate additive variance for different cancer cohort. ... [11]

Deleted: -- Ref 3.2

Moved down [5]: Sensitivity analysis of background model---
Calculate additive variance with different randomized sample set. Vary the length of local window to generate multiple random set and compute the additive variance. ... [12]

Deleted: -- Ref 3.3

Moved down [6]: Excerpt From ... [13]

-- Ref 3.2 – Consistent use of terminology--

Reviewer Comment	The authors define "nominal passengers" but it is unclear exactly how these are defined. It is critical for the rest of the paper. I would remove any event in a very inclusive list of cancer genes and not only the ones that we detected as drivers using the PCAWG dataset. I think the term "potential passengers" is better. It is crucial that the term is used throughout the manuscript. There are several places where they are referred to as just "passengers". This is confusing since true passengers cannot be weak drivers, but potential/nominal passengers can.
Author Response	Please see section "Ref 1.0" for the detailed response.
Excerpt From Revised Manuscript	

-- Ref 3.3 – Additive variance in coding region--

Reviewer Comment	If the authors focus only on coding regions, are the results of the random effects model consistent with significance analyses such as MutSig and dN/dS ?
Author Response	We now perform this comparison directly using the new nested model results. We compare our list of weak driver genes with the gene set curated by the driver group. We specifically look for overlap between our weak driver genes and driver discovery gene set, which didn't satisfy the statistical significance criterion during the driver discovery process. We find good overlap between these two-list validating our approach to certain extent. Presumably, some of these genes are weak drivers and failed the statistical significance threshold due to limited cohort size and thus insufficient power.
Excerpt From Revised Manuscript	

-- Ref 3.4 – Impact score and additive variance model --

Reviewer Comment	It is not exactly clear how the authors use the FunSeq score to derive the z_ij. How would the results change if only evolutionary conservation is used? (Even evolutionary conservation has its own problems since some of the same processes that affect cancer shape evolution)What are the underlying features that contribute to the peaks in the impact score (conservation? specific chromatin marks? alteration of TF binding sites?).

Deleted: 4

Deleted: "Nominal passengers" include all mutations that are not identified as drivers in PCAWG, or previously known to be cancer driving events. ... [14]

Moved (insertion) [6]

Moved down [7]: Excerpt From ... [15]

Deleted: 5

Deleted: We find good overlap between these two-list validating our approach to certain extent.

Deleted: Excerpt From ... [16]

Moved (insertion) [7]

Deleted: ... [17]

Deleted: 6

Author Response	We have attempted to explain this more clearly in the manuscript. The FunSeq threshold is treated as an additional optimization parameter. We also compared our results with a model optimized by conservation score (e.g. GERP) and found no significant difference in total additive variance explained. The different peaks in the impact score distribution can be attributed to combinations of features (conservation, chromatin marks, and TF motif alterations).
Excerpt From Revised Manuscript	

-- Ref 3.5 – SV impact score definition --

Deleted: 7

Reviewer Comment	It is unclear how the impact score for SVs was defined.
Author Response	We have provided a detailed description in the supplemental methods section.
Excerpt From Revised Manuscript	

-- Ref 3.6 – Singletons and additive effect model --

Deleted: 8

Reviewer Comment	How many of the variants are singletons (ie. occur only in one sample across the cancer and "neutral" cases)? How do they affect the model and its predictive power?
Author Response	We don't use singletons in our model to calculate additive variance. Singletons cannot increase the additive variance, since they do not contribute to genetic relatedness. As noted in response 2.1, we do not train a predictive model, so the issue of overfitting does not apply to the SNV effects. We verified directly through simulations that adding singletons decreases the estimated additive variance.
Excerpt From Revised Manuscript	

Deleted:

Deleted: . Hence, we remove singletons for all additive variance calculations in the paper. We note that, for the SNV-level model, singletons are SNVs occurring in only one sample, while in the gene-level model, they are genes which have non-zero mutation burden only in one sample.

-- Ref 3.7 – Signature and impact score distribution --

Deleted: 9

Reviewer Comment	Not clear what is the contribution of different mutational signatures to the 3 peaks on functional impact.
Author Response	We address this question in Figure3d, which shows the signature differences between nominal passengers with high impact score (mutations present in second and third peak) and low impact score (mutations present in the first peak). As noted, these differences vary between different cancer types and we highlight some of these differences in updated text more explicitly.
Excerpt From Revised Manuscript	

-- Ref 3.8 –Potential role of signature driving some observation--

Reviewer Comment	It is concerning that many of the results are in tumor types with specific signatures that we are probably not modeling accurately, such as UV, lymphomas (canonical and non-canonical AID), lung cancer, liver cancer, and esophageal cancer.
Author Response	We would like to emphasize that the main aim of our work is to characterize "nominal passenger" landscape in PCAWG. For this purpose, we perform many empirical analyses, which doesn't require any background model or signature modeling. As the reviewer points out, for certain cohorts it's very likely that signature plays an important role in influencing our observations. We think this is perfectly fine as this provide mechanistic understanding to some of our observations. Moreover, this is one key reason why we perform analyses to characterize mutation spectrum and signature differences among "nominal passengers". In the updated manuscript, we clarify this point more explicitly.
Excerpt From Revised Manuscript	

Deleted: 10

Deleted: specifically look at their overall burden among different genomic elements and their predicted molecular functional impact.

Deleted: look at

Deleted: . ... [18]

-- Ref 3.9 – comparison of germline and somatic SVs --

Reviewer Comment	I am not sure that the comparison to germline for SV is meaningful – there are very different patterns of SVs in germline and somatic. It is also unclear what is the random model (is it uniform?)
Author Response	As the reviewer points out, we use a uniform randomization scheme to perform this analysis. Comparison between germline and somatic SV provides mechanistic insight into how different categories of SVs burden various genomic elements. Based on reviewer's suggestion, we further downplay this analysis.
Excerpt From Revised Manuscript	

Deleted: 11

Deleted: The goal of

Deleted: was to highlight the observation that both

Deleted: large deletions prefer to engulf genomic elements rather than partially break it. This is very interesting and

Deleted: As the reviewer points out

Deleted: use a uniform randomization scheme to perform

-- Ref 3.10 – Essential gene and LOF analysis --

Deleted: 12

Reviewer Comment	What is the list of essential genes that is used? Keep in mind that essential genes can be expressed at higher levels and therefore subject to different DNA damage and repair mechanisms (such as transcription-coupled repair and transcription-coupled damage (in liver cancer)).
Author Response	The essential gene list was based on previous publication "Essential genes - CRISPR knockouts in four cancer cell lines". The reviewer suggests a good point, which we explicitly mention in the updated draft.
Excerpt From Revised Manuscript	

-- Ref 3.11 – Consistency in PCAWG wide terms --

Deleted: 13

Reviewer Comment	We need to be consistent across PCAWG papers. We typically use SCNA (somatic copy number alteration) rather than CNV.
Author Response	The reviewer makes a good point here. We update the text accordingly to keep the nomenclature consistent with other PCAWG papers.
Excerpt From Revised Manuscript	

-- Ref 3.12 – TFBS related analysis --

Deleted: 14

Reviewer Comment	I am skeptical about the analysis of hits in different binding sites of a TF. There is no correction for signatures. I believe this is mostly mechanistic, e.g. CTCF binding sites in liver cancer.
Author Response	We concur with reviewer that TF related analysis is mechanistic and not trying to highlight any role of selection. This analysis is aimed towards highlighting the differential burdening of various TFs and the corresponding gene regulatory network. We have modified the main text to clarify this point more explicitly.

Response to extra comments from Gaddy based on phone call

-- Ref 3.13 Sensitivity analysis of additive variance model---

Moved (insertion) [4]

Reviewer Comment	Calculate additive variance for double randomized sample set, i.e., use two iteration of random sample to calculate additive variance for different cancer cohort.
Author Response	As per suggestion, we performed sensitivity analysis to cross-check the issue of over-fitting in our random effects model. We computed the additive variance of two random samples. Across cancer types, we observe ~0% additive variance suggesting no overfitting in our analysis.
Excerpt From Revised Manuscript	

-- Ref 3.14 Sensitivity analysis of background model---

Moved (insertion) [5]

Reviewer Comment	Calculate additive variance with different randomized sample set. Vary the length of local window to generate multiple random set and compute the additive variance.
Author Response	As per suggestion, we performed sensitivity analysis to cross-check the influence of background model on additive variance. We generated two distinct randomizations set within a local window length of 50kb and 100kb. Overall, there is a slight variation in the total additive variance calculated based on these distinct randomizations set. This can be attributed to small effect due to various genomic co-variates. We also note that in our updated analysis, we perform additive variance on a separate background model, which corrects for various covariates suggested by all the reviewers.
Excerpt From Revised Manuscript	

Page 2: [1] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

Page 3: [2] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

Page 3: [3] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

In the mutational signature section, we analyze the role of mutational signatures to the differential burdening of genomic elements by nominal passengers. We perform this analysis for the coding LoF mutations, as well as for non-coding mutations leading to TF motif break events.

Reviewer's

Page 3: [4] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

Furthermore, we also compare the difference in mutational signature for "nominal passengers" with low and high impact scores for various PCAWG cancer cohorts. Categorization of nominal passengers into low and high impact score groups was done based on functional impact score threshold.

Page 5: [5] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

Page 7: [6] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

Excerpt From Revised Manuscript	

Page 8: [7] Moved to page 7 (Move #1)	10-1 to 10-7	10/8/17 10:40:00 PM

Excerpt From Revised Manuscript	

Page 8: [8] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

The Driver discovery exercise in PCAWG suggest that we don't find many non-coding drivers in PCAWG cohorts as well as on pan-cancer level. This can be potentially attributed to the lack of power for identifying non-coding driver due to limited sample size. In this context, our additive variance analysis provides an alternative approach to address this issue.

Page 8: [9] Moved to page 7 (Move #2)	10-1 to 10-7	10/8/17 10:40:00 PM

Excerpt From Revised Manuscript	

Page 9: [10] Deleted	10-1 to 10-7	10/8/17 10:40:00 PM

Page 10: [11] Moved to page 11 (Move #4)	10-1 to 10-7	10/8/17 10:40:00 PM

Sensitivity analysis of additive variance model–--

Reviewer Comment	Calculate additive variance for double randomized sample set, i.e., use two iteration of random sample to calculate additive variance for different cancer cohort.
Author Response	As per suggestion, we performed sensitivity analysis to cross-check the issue of over-fitting in our random effects model. We computed the additive variance of two random samples. Across cancer types, we observe ~0% additive variance suggesting no overfitting in our analysis.
Excerpt From Revised Manuscript	

Page 10: [12] Moved to page 11 (Move #5)	10-1 to 10-7	10/8/17 10:40:00 PM

Sensitivity analysis of background model–--

Reviewer Comment	Calculate additive variance with different randomized sample set. Vary the length of local window to generate multiple random set and compute the additive variance.
Author Response	As per suggestion, we performed sensitivity analysis to cross-check the influence of background model on additive variance. We generated two distinct randomizations set within a local window length of 50kb and 100kb. Overall, there is a slight variation in the total additive variance calculated based on these distinct randomizations set. This can be attributed to small effect due to various genomic co-variates. We also note that in our updated analysis, we perform additive variance on a separate background model, which corrects for various covariates suggested by all the reviewers.
Excerpt From Revised Manuscript	

Page 10: [13] Moved to page 11 (Move #6)	10-1 to 10-7	10/8/17 10:40:00 PM

Excerpt From Revised Manuscript	

Page 11: [14] Deleted **10-1 to 10-7** **10/8/17 10:40:00 PM**

"Nominal passengers" include all mutations that are not identified as drivers in PCAWG, or previously known to be cancer driving events.
We agree with reviewer's comment to consistently use the terminologies across the manuscript. We have updated the manuscript accordingly.

We also would like to point out that some of our terminologies are borrowed from prior literature and we cite them accordingly.

Page 11: [15] Moved to page 11 (Move #7) **10-1 to 10-7** **10/8/17 10:40:00 PM**

Excerpt From Revised Manuscript	

Page 11: [16] Deleted **10-1 to 10-7** **10/8/17 10:40:00 PM**

Excerpt From Revised Manuscript	

Page 11: [17] Deleted **10-1 to 10-7** **10/8/17 10:40:00 PM**

Page 13: [18] Deleted **10-1 to 10-7** **10/8/17 10:40:00 PM**

We would also like to point out that additive variance analysis is a subset of many analyses performed in this work. This analysis only addresses a specific question related to "nominal passengers (i.e. whether their cumulative effect has any role in cancer progression). For this purpose, we use an updated random model, which takes into account of many mutational processes mentioned by the reviewer.