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 have updated the text to define various terminologies upfront and use these consistently throughout the text to avoid any confusion. 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 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	

-- 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 an increase in the variance explained by our model with these explicit covariates, suggesting the signal is robust to the effects mentioned.
Excerpt From Revised Manuscript	

-- 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 ("With a number of caveats…"). In short, in the case of GWAS related studies, the model predicts the 'narrow-sense' (additive) heritability. Additive heritability can be justified for many germ-line traits since recombination limits the amount of epistasis that can be maintained for polygenic traits. The model does not directly model heritability for clonal evolution, since in this case 'broad-sense' heritability acquires greater importance, which includes non-additive effects. Further, we analyze a balanced observed/null sample with a binary phenotype, rather than individual subclones and their associated fitness within a tumor, as would be required to estimate clonal fitness heritability. However, the model can still serve as an indicator of the first-order (additive) information contained in the SNVs about tumorigenesis, which in many cases is substantial. We have updated the manuscript to make the above clearer, explicitly stating that the additive variance we calculate does not directly measure heritability as in complex trait analysis, but that the form of the model is nevertheless appropriate to measure the cumulative effects of variants on tumorigenesis (with the provisos above regarding the null model).
Excerpt From Revised Manuscript	

-- Ref 1.5 – Related to signature analysis --

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? 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 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, by closely inspecting their underlying mutation spectrum. For LoFs, we look at mutational spectrum in the context of mutations altering stop codons. In Figure 3a, mutational spectrum is plotted for mutations leading to LOF events. In contrast, for TF motif breaking events, we look at the entire mutation spectrum. We highlight this specifically for the renal cell carcinoma cohort. 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 their functional impact score.
Excerpt From Revised Manuscript	

-- Ref 1.6 – 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)? 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 Model is a GLMM with random effects. We don't use a logit linking function: prior GWAS analyses have applied either a direct linear model (observed scale) or a probit model (liability scale). Both of these give similar results (we will quote both), though estimates of variance explained are slightly higher for probit model, and interpretation of 'liability' is unclear in our model. Regarding the issue of overfitting, the random effects model adapted from previous GWAS studies does not directly estimate the effect sizes of individual SNVs, but rather estimates the variance of a normal distribution which acts as a common prior for these effect sizes (which is a hyper-parameter of the GLMM). 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. Further, in the updated manuscript we apply a nested random effects model, which shows that the nominal passengers (coding and non-coding) are capturing strictly non-redundant information that is not contained in the drivers. This remains significant when the drivers are directly included in the model (~15% extra). 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.
Excerpt From Revised Manuscript	

-- Ref 2.2 – Improving background model --

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 2.3 – 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 updated our downstream analysis based on the new null model, which corrects for the covariates suggested by the reviewer. In addition, our SNV-level estimates of weak drivers and deleterious passengers are conservative, and sample size appears to have little effect.
Excerpt From Revised Manuscript	

-- Ref 2.4 – 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 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.
Excerpt From Revised Manuscript	

-- Ref 2.5 – 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	As per reviewer's suggestion, we are now more careful not to suggest that missing heritability applies to the same extent in cancer. Also, as above, we update the text to 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. 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. 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.
Excerpt From Revised Manuscript	

-- Ref 2.6 – 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. 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	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.
Excerpt From Revised Manuscript	

-- 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 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. In addition, we also mention these limitations explicitly in our discussion section.
Excerpt From Revised Manuscript	

-- 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	"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.
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. <talk about JW's result>
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?).
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 --

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 --

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	The singletons do not affect the additive variance, since they do not contribute to genetic relatedness, and so lead to a model with an identical likelihood. For efficiency, we do not include them in the model, although the results are identical if they are included. 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.
Excerpt From Revised Manuscript	

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

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	As the reviewer points out for certain cohorts it's very likely that signature plays an important role. In this work, we perform an unbiased holistic analysis to characterize passenger variants. One could justify our observations, as driven either by signatures or weak selection. In order to establish role of weak selection, we compare original observations to randomized set. In the updated random model, we take into account of many signature effects. In contrast, some of our observations are mechanistic in nature. For example, Figure 1d shows correlation between number of high impact nominal passengers and total mutation in burden for various samples in a particular cancer cohort. In certain cancer cohorts, we observed strong negative correlations, which were statistically significant. As the reviewer points out for certain cohorts it's very likely that signature plays an important role. That's why we look at signature differences between high and low impact nominal passengers. In the updated manuscript, we clarify this point more explicitly.
Excerpt From Revised Manuscript	

-- 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	The goal of this analysis was to highlight the observation that both germline and somatic large deletions prefer to engulf genomic elements rather than partially break it. This is very interesting and provides mechanistic insight into how different categories of SVs burden various genomic elements. As the reviewer points out, we use a uniform randomization scheme to perform this analysis.
Excerpt From Revised Manuscript	

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

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 --

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 --

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.