Dear Dr. Villanueva and the Committee Members of the manuscript meeting of *BMJ*,

My colleagues and I are pleased to receive your mail dated May 25 informing us that you recognize the potential importance and relevance to general medical readers of our manuscript entitled "HLA-B*58:01 genotyping to prevent allopurinol-induced severe cutaneous adverse reactions: national prospective study" (Manuscript ID BMJ.2015.025516) and that you would consider a revised manuscript for publication in the BMJ. We were pleased to carefully examine and fully address your as well as reviewers' thorough and thoughtful comments.

To begin, we would like to emphasize three critical issues related to several comments raised by you and the Reviewers.

(A) A prospective study is essential to verify whether the use of genetic testing to prevent adverse drug reactions in proper clinical setting is of clinical feasibility and reality. Our group first reported the association between the HLA-B*58:01 allele and skin adverse reactions in 2005 (Hung *et al.*, Proc Natl Acad Sci USA.2005;102:4134-9). This finding was subsequently confirmed by several independent studies. However, all these studies were performed retrospectively. **The purpose of the prospective study described in this manuscript was not to explore this association but to assess the clinical utility of HLA-B*58:01 screening in preventing allopurinol-induced severe cutaneous adverse reactions (SCARs), as well as to demonstrate the feasibility of incorporating personalized medicine to the clinical setting.**

(B) As clearly indicated in our manuscript, we were unable to perform a direct comparison in our study because of ethical considerations. The ethical consideration in question is not only whether it is ethical to administer allopurinol to someone who has tested positive for the HLA subtype of concern. More importantly, the relevant ethical consideration is also "*whether it is ethical to administer allopurinol to somebody who has not been tested if testing is available",* a consideration pointed by one of your editors which we greatly appreciate it.

We used historical incidence in the present study, in which the incidence of allopurinol-induced adverse drug reaction (ADR) was calculated based on information from an internationally recognized and highly reliable database, the National Health Insurance Research Database (NHIRD) of Taiwan. Diagnosis of allopurinol-induced SCARs was confirmed after crosschecking with the NHIRD and patients' medical history to confirm the diagnosis of SCAR and to ensure that the SCARs were not due to other medications. The approach of using historical

incidence and the NHIRD has been used in many studies, including a study published by our group in 2011 that successfully demonstrated screening of HLA-B*15:02 to prevent carbamazepine-induced toxic effects (Chen *et al*., N Engl J Med. 2011;364:1126-33.). More importantly, **for a nationwide prospective study to recruit patients from different participating hospitals, the use of highly valid national statistical data from the NHIRD is more reliable and much more comprehensive than performing a direct comparison in individual hospitals.**

(C) The issue of our two-month follow-up period was raised. The two-month follow-up period chosen with care at the beginning of this study was based on evidence from our previous study published by our group in PNAS in 2005 (Hung *et al.*, Proc Natl Acad Sci USA 2005;102:4134-9), which was one of the largest studies to report the history of allopurinol-induced severe cutaneous adverse reactions (SCARs). In this study, all allopurinol-related SCAR patients showed onset of SCARs within the first 2 months of allopurinol use. This finding has been consistently reported in many studies, suggesting that the typical latency to SCAR onset is 2 to 6 weeks (http://livertox.nih.gov/Allopurinol.htm).

Below, we respond point-by-point to the critical issues and comments raised by you in the manuscript meeting as well as by the reviewers.

We hope that the article will now be suitable for publication in the BMJ. We look forward to hearing from you.

Chen-Yang Shen, Ph.D.
Institute of Biomedical Sciences, Academia Sinica
Taipei, 11529, Taiwan
e-mail: bmcys@ibms.sinica.edu.tw

Our responses to comments raised by the committee:

1. *The authors only followed participants for two months. Could they please provide stronger evidence and data supporting that adverse events are likely to occur within 2 months? (the current phrase is that "in general SCARs onset occurs within 2 months" which is not particularly convincing that all cases would have been detected). It would seem a weakness of the study not to have followed-up individuals for a longer time period, particularly seeing that the control group data are based on 3-monthly prescriptions.*

Response: This is a very critical issue related to the validity of our results and we thank you for the opportunity to address this issue in more detail. We are aware that there are very rare cases that SCAR occur more than 2 months of allopurinol exposure, however, the 2-month follow-up period carefully determined at the beginning of this study was based on evidence from a previous study published by our group in PNAS in 2005 (Hung *et al.*, Proc Natl Acad Sci USA 2005;102:4134-9), which was one of the largest studies to report the history of allopurinol-induced severe cutaneous adverse reactions (SCARs). In this study, all allopurinol-related SCAR patients showed SCAR onset within the first 2 months of allopurinol use. Included below is Table 1 from our 2005 PNAS paper, in which we clearly demonstrate that all 51 patients developed SCARs within 2 months and show that the longest exposure was 56 days. This finding has been consistently reported in other studies, which suggested that "the typical latency to onset is 2 to 6 weeks" (http://livertox.nih.gov/Allopurinol.htm).

Table 1. Clinical characteristics of patients with allopurinol-induced SCAR

Patient no.	Age/sex	Phenotype	Dose, mg/d/ duration of drug exposure, days	Exanthema %/blister or detachment,† %	Mucous membrane erosions, number:site	Systemic manifestations
1	80/F	SJS	100/33	31–50/2	(+) 2: Oral, eyes	None
2	47/F	HSS	100/23	31–50/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
3	60/F	SJS	100/14	31–50/5	(+) 2: Oral, eyes	Fever, worsening RF, leukocytosis, atypical lymphocytosis‡
4	25/M	SJS	200/24	51–70/5	(+) 1: Oral	Fever, LFI, leukocytosis, eosinophilia
5	50/F	HSS	300/33	51–70/0	(+) 1: Oral	Fever, LFI, eosinophilia
6	74/F	HSS	100/22	31–50/0	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia, atypical lymphocytosis
7	78/M	HSS	NA/21	71–90/0	(−)	Worsening RF, eosinophilia
8	37/F	SJS/TEN	100/38	51–70/10	(+) 3: Oral, eyes, genital	Fever, LFI, leukocytosis, eosinophilia
9	18/F	HSS	200/21	51–70/0	(−)	LFI, worsening RF, leukocytosis, eosinophilia
10	63/F	SJS	200/37	51–70/5	(+) 2: Oral, genital	LFI, worsening RF, eosinophilia
11	55/F	HSS	100/35	51–70/0	(−)	Fever, LFI, leukocytosis
12	70/F	HSS	NA/56	31–50/0	(+) 2: Oral, eyes	Fever, LFI, leukocytosis, eosinophilia
13	52/F	SJS	100/35	51–70/5	(+) 3: Oral, eyes, genital	Fever, LFI, eosinophilia
14	29/F	SJS	NA/28	51–70/5	(+) 2: Oral, eyes	None
15	73/M	SJS	200/30	51–70/1	(+) 3: Oral, eyes, genital	Fever
16	78/M	SJS	100/1§	51–70/1	(+) 1: Oral	Worsening RF, eosinophilia
17	70/M	SJS/TEN	NA/27	51–70/10	(+) 2: Oral, eyes	Fever
18	69/F	HSS	100/53	31–50/0	(+) 1: Oral	LFI, worsening RF, eosinophilia
19	91/F	TEN	100/26	71–90/30	(+) 2: Oral, eyes	Atypical lymphocytosis
20‖	60/M	SJS/TEN	200/21	51–70/15	(+) 2: Oral, eyes	Fever, LFI, worsening RF, eosinophilia
21	77/F	SJS	NA/14	51–70/5	(+) 2: Oral, genital	Worsening RF
22	62/F	HSS	100/38	51–70/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
23	41/M	HSS	200/32	51–70/0	(+) 1: Oral	LFI, leukocytosis
24	72/F	HSS	100/45	>90/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
25	51/M	SJS	200/2§	51–70/5	(+) 3: Oral, eyes, genital	None
26	35/M	HSS	NA/30	31–50/0	(−)	Fever, leukocytosis, eosinophilia
27	65/F	HSS	300/49	51–70/0	(−)	Leukocytosis, eosinophilia
28‖	54/F	SJS	100/14	51–70/3	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia
29	51/M	SJS	NA/21	51–70/5	(+) 1: Oral	Fever, LFI, atypical lymphocytosis
30	52/F	HSS	NA/45	51–70/0	(+) 1: Oral	Fever, LFI, leukocytosis, severe myositis
31	50/M	HSS	50/15	31–50/0	(−)	LFI, worsening RF
32	80/M	SJS	NA/12	31–50/5	(+) 2: Oral, genital	None
33	80/M	HSS	100/21	31–50/0	(−)	LFI, worsening RF, leukocytosis, eosinophilia
34	70/M	HSS	50/18	31–50/0	(−)	LFI, leukocytosis, eosinophilia, atypical lymphocytosis
35	52/M	HSS	100/7	71–90/0	(−)	Leukocytosis, eosinophilia
36	63/M	HSS	100/52	31–50/0	(−)	LFI, eosinophilia
37	67/F	HSS	100/21	51–70/0	(−)	Fever, worsening RF, eosinophilia
38	55/M	HSS	100/14	51–70/0	(−)	Fever, worsening RF, eosinophilia
39‖	66/F	TEN	100/23	71–90/30	(+) 3: Oral, eyes, genital	Leukopenia
40	62/M	HSS	300/30	71–90/0	(−)	LFI, worsening RF, eosinophilia
41	73/F	HSS	100/39	71–90/0	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia
42	84/M	HSS	100/30	51–70/0	(−)	Worsening RF, eosinophilia
43	24/M	HSS	NA/28	51–70/0	(−)	Fever, worsening RF, eosinophilia
44‖	78/F	HSS	100/30	71–90/0	(+) 1: Oral	Fever, worsening RF, eosinophilia, leukocytosis
45	69/F	HSS	100/26	51–70/0	(−)	Worsening RF, leukocytosis, eosinophilia
46	73/M	SJS/TEN	100/1§	51–70/10	(+) 3: Oral, eyes, genital	LFI, worsening RF
47‖	73/F	HSS	100/30	71–90/0	(+) 1: Oral	Fever, LFI, worsening RF, leukocytosis, eosinophilia
48	70/F	SJS/TEN	200/12	51–70/15	(+) 2: Oral, eyes	Fever, leukocytosis
49	70/M	HSS	100/20	51–70/0	(+) 3: Oral, eyes, genital	LFI, worsening RF, eosinophilia, atypical lymphocytosis
50	71/M	TEN	100/26	71–90/30	(+) 1: Oral	Fever, LFI, worsening RF
51	71/M	HSS	100/7	51–70/0	(+) 1: Oral	Fever, LFI, worsening RF, eosinophilia

See *Methods* for the diagnostic criteria. NA, data not available; RF, renal function; LFI, liver function impairment; M, male; F, female.
†Exanthema, erythematous or purpuric macules/papules; blister or detachment, extent of blisters or epidermal detachment; both exanthema and blister/detachment were expressed as % of total body surface area.
‡Second attack on reexposure.
§Cases deceased.

We are continuing to follow the study participants. All study participants have now been followed for at least 9 months and none have developed SCARs. We have also followed the approach suggested by the reviewer to identify possible SCARs in these prospective cohort members by linkage with the National Health Insurance Research Database (NHIRD) dataset based on unique identification numbers of individual Taiwanese patients, but no SCARs were identified using this method. Moreover, as one editor mentioned, a sensitivity analysis could be performed to assess the consequences of using a longer period of presentation of SCAR's. To date (June 2nd, 2015), based on our genetic screening protocol, we have not identified cases of SCARs in the study participants; thus, we suggest that

the incidence of SCARs found in the 2-month follow-up was equal to the incidence of SCARs found in the 9-month follow-up.
In response to this critical comment, we have included some details regarding this data in the revised manuscript (lines 14-17 on page 10) (lines 12-19 on page 17).

2. *The power calculation is written as if the comparison is being made between a known fixed value of 0.3% and a hypothesised value of 0.03% to be estimated within the sample. Given that the comparison is known and estimated with high precision this seems reasonable. However, the computed sample size does not agree with this (my sample size calculator using a normal approximation suggests that the stated sample size of 2169 would have 85.8% power (not the 99% stated). Could the authors fully justify and explain their calculation.*
Response: The original power calculation was performed under two-sample one-side test to determine whether the incidence of new approach (with HLA-B*58:01 genotyping screening) was lower than the incidence of previous approach (without HLA-B*58:01 genotyping screening).
We have followed your suggestions and use one-sample binomial test. The power of the two-side one-sample binomial test is about 86%, assuming p0 = 0.003 and p1 = 0.0003, with normal distribution approximation. We have revised manuscript in response to this critical comment (lines 12-14 on page 13).

3. *The statistical method the authors state has been used is a Fisher's Exact test. This is a two group test and requires data from both the screening cohort and the comparison group. It is not clear what data has been used for the comparison group in the analysis presented in the paper. However, in line with point 2) the comparison between the observed result in the collected cohort and the historical figure of 0.3% might be better based on a one sample Binomial test (to work out the probability that the see whether the observed incidence rate differs from the fixed value of 0.3%). My estimate of the associated P-value from this test is 0.0003 – thus similar to that reported by the authors.*
Response: We have followed your suggestions and used a one-sample binomial test. The p-value is equal to 0.0026, assuming X ~ B (p0 = 0.003, n = 2169) with a normal distribution approximation (lines 5-6 on page 6; lines 16-17 on page 16) (**Table 3**).

4. *It is important that the authors report the 95% confidence interval for the observed event rate in the cohort – (0% to 0.13%) as this gives an upper limit on the observed event rate.*

Response: The 95% confidence interval was added (0%, 0.17%) (lines 5-6 on page 6; lines 16-17 on page 16).

5. *The statistical methods here are not quite right – both for the power calculation and the comparison, but it makes no difference to the conclusions which would have been drawn. This is a "one sample" study which compares the observed rate to a fixed known value (0.3%), and an appropriate "one sample" sample size calculation should have been done (they have overestimated power in their calculation) and comparison of the observed to a fixed value using a binomial test (which gives nearly the same P-value).*
   Response: We appreciate your suggestions and have re-performed all calculations using the one-sample settings. As you have mentioned, all results continued to support the original conclusions.

6. *The economic analysis does not provide enough detail to be helpful.*
   Response: We agree with this comment and concur that a pharmacoeconomic analysis should be presented in more detail. We are working on such an analysis, which we plan to publish in an independent article. We thought that the readers of BMJ would benefit from an acknowledgement of the importance of particular consideration of pharmacoeconomic issues in the context of the practice of genetic screening to prevent adverse drug reactions.
   In response to this comment, which was also raised by one reviewer and one editor suggested removal of economic data, we now only briefly mention the importance of pharmacoeconomic analysis in the revised manuscript (lines 9-11 on Page 19).

7. *One editor wondered about the large number or authors….He also wondered whether results applied to people who are not of Han Chinese descent. Does this limit the generalizability of the findings? In the US, most physicians will identify patients of Chinese descent but he doesn't think they would differentiate the different ethnicities. Would a physician starting allopurinol ask for this information? Is genomic screening before prescribing allopurinol the standard of care? From the paper, he did not get the sense that it is, even in Taiwan. As such, he felt that doing a RCT is not necessarily unethical and would be better than using historical control data. He felt the economic data should be removed. Finally, he wondered whether the paper is useful, as the association between the presence of HLA-B*58:01 and skin reactions is well known. He felt that without a suitable control group, we cannot draw firm inferences about the clinical impact*

of screening.

Response: Our point-by-point responses to the many critical issues raised by this editor are included below.

(A) This is a nation-wide prospective study that would not be possible without a group effort. Importantly, when we started this study, all authors, along with many other participating doctors not listed in the manuscript, discussed the justification of the use of genetic screening to prevent adverse reactions, the feasibility of this study, the study design, and finally determined the protocols. Because of the large number of collaborators who participated in conducting this study, we were forced to limit authorship only to those key participants who put forth tremendous effort towards the study design, patient recruitment, and preparation of the manuscript.

(B) The comment questioning whether the findings of this study can be applied to people who are not of Han Chinese descent is important. This concern can be properly addressed by examining findings of studies independently conducted in other populations along with our recent experimental findings. As mentioned in our manuscript (lines 15-19 on page 21) (lines 1-11 on page 22), we suggest that the application of our findings may not be limited to individuals of Han Chinese descent, because different independent groups have replicated and confirmed the association between HLA-B*58:01 and allopurinol-induced severe cutaneous adverse reactions (SCARs), originally reported in Han Chinese by us in 2005 (Hung *et al.*, Proc Natl Acad Sci USA.2005;102:4134-9), and later reported in multiple populations and races (e.g. Thai, Japanese, Korean, and European). In addition, based on our recent mechanistic study (Lin *et al.*, J Allergy Clin Immunol 2015;135(4):1063-5.e5), HLA-B*58:01 can **directly** present allopurinol metabolites (e.g. oxypurinol) to cytotoxic T cells without antigen processing, while allopurinol/oxypurinol-specific T cell-mediated cytotoxicity is restricted to carriers of HLA-B*58:01, suggesting that HLA-B*58:01 is the key determinant of susceptibility to allopurinol-induced SCARs and that HLA-B*58:01 carriers are susceptible to SCARs when taking allopurinol, regardless of their ethnic group. We have mentioned this point in the revised manuscript (lines 7-10 on page 8). Therefore, in the manuscript, we emphasize that HLA-B*58:01 screening could be beneficial for preventing allopurinol-induced SCARs in countries where HLA-B*58:01 is relatively prevalent (not restricted only to Han Chinese) (lines 15-19 on page 21) (lines 1-11 on page 22). Even in countries where the allele frequency of HLA-B*58:01 is relatively low (~ 1%), HLA-B*58:01 screening may be

beneficial in a subset of high-risk patients (e.g. chronic renal failure). For countries or populations, in which the prevalence is ill-defined, further studies are suggested to estimate the prevalence for possible application of this screening.

Therefore, because HLA-B*58:01 is a relatively common allele in multiple populations, we suggest that genetic screening prior to prescription of allopurinol could be a standard of care. Indeed, this is the primary reason why we hope to publish our manuscript in BMJ; our study would provide essential justification for this life-saving practice in clinics, which would reach an exceptionally wide audience due to the popularity of BMJ in the medical community, raising awareness of this life-saving issue.

(C) As mentioned above, we agree that a pharmacoeconomic analysis should be presented in more detail and plan to do so in an independent article. Originally, we thought that it might be helpful to the readers of BMJ to acknowledge the importance of particular consideration of pharmacoeconomic issue in the practise of genetic screening to prevent adverse drug reactions. In response to this comment, we have removed the economic data and only briefly mention the importance of pharmacoeconomic consideration in the revised manuscript (lines 9-11 on page 19).

(D) Regarding "*whether the paper is useful, as the association between the presence of HLA-B*58:01 and skin reactions is well known"*

The association between the presence of HLA-B*58:01 and skin reactions was first reported by us in 2005 (Hung *et al.*, Proc Natl Acad Sci USA.2005;102:4134-9) and has been further confirmed by several independent studies. However, all these prior studies were retrospective. The purpose of this prospective study was not to explore this association. Instead, this prospective study was conducted to demonstrate the clinical feasibility of the use of HLA-B*58:01 screening to prevent allopurinol-induced severe cutaneous adverse reactions (SCARs) and to demonstrate that the practice of personalized/precision medicine can be a clinical reality. As mentioned in the manuscript (lines 17-19 on page 20) (lines 1-13 on page 21), there are two other reports regarding involvement of HLA-B in adverse drug reactions (ADR): HLA-B*15:02 in carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis (SJS/TEN), and HLA-B*57:01 in abacavir-induced drug hypersensitivity, both required a prospective study to convince the medical community the clinical utilization of genetic screening (Mallal *et al.*, N Engl J Med. 2008;358:568-79) (Chen *et al*., N Engl J Med

2011;364:1126-33.). Similarly, we believe that our large-scale nation-wide prospective study would have a significant impact in clinical settings by minimizing the incidence of allopurinol-induced SCARs.

(E) Regarding "*without a suitable control group, we cannot draw firm inferences about the clinical impact of screening"*

As we mentioned in the manuscript (lines 7-9 on page 9), because of ethical consideration, we can not perform a randomized controlled trial. We used historical incidence in the present study (line 9 on page 9), in which the incidence of allopurinol-induced adverse drug reaction (ADR) was calculated based on information from an internationally recognised and highly reliable database, the National Health Insurance Research Database (NHIRD) of Taiwan. Thus, the reliability of the NHIRD is critical in assessing the validity of our estimation. To address this issue, it should be noted that the NHIRD was established in Taiwan when the National Health Insurance (NHI) system was launched in 1995. The NHI is a mandatory single-payer social health insurance system administered by the government that provides universal high-quality health care for almost all people in Taiwan, with an enrolment rate of 99.5% (in 2008). Furthermore, 95% of all health care facilities in the country have been contracted by the NHI system. Therefore, the NHIRD is an exceptionally accurate database regarding the medical status of the Taiwanese population, because information from the vast majority of such patients is included in the database. The Taiwanese NHI system has been recognized as one of the best health insurance systems in the world and has an international reputation for high-quality patient diagnosis and care. The NHIRD collects registration data, including data from the registry of medical services and registry of drug prescriptions. The Bureau of National Health Insurance is required to review reimbursement claims and to screen the type, volume, quality, and appropriateness of medical services provided under NHI program, ensuring the quality and accuracy of the NHIRD. More than 400 papers based on the NHIRD have been published in international peer-reviewed journals (for a detailed list, please see http://w3.nhri.org.tw/nhird/en/Research.html.)

Therefore, we respectfully disagree with the comment that "*without a suitable control group, we cannot draw firm inferences about the clinical impact of screening."* The approach of using historical incidence and the NHIRD has been utilized in many studies, including a study from our group that successfully demonstrated screening of HLA-B*15:02 to prevent carbamazepine-induced toxic effects in 2011 (Chen *et al.*, N Engl J Med.

2011;364:1126-33.). For a nationwide prospective study to recruit patients from different participating hospitals, the use of highly valid national statistical data from the NHIRD is more reliable and much more comprehensive than performing a direct comparison of individual hospitals. Furthermore, the clinical impact of screening can be assessed more effectively at a national scale. In response to this critical comment, we have included some relevant information in the revised manuscript (lines 8-17 on page 12).

8. *Another editor felt that the study was useful and that this was an interesting clinical development. However, he felt that the study is methodologically not strong but still convincing.*
   Response: We greatly appreciate the positive comment raised by the reviewer.

9. *Another editor was supportive and felt that most of the concerns raised in the reviews seem addressable. He added that the time period for selecting the historical controls seemed odd and felt that it should be more recent. Moreover, he considered that the limitations should mention the 2 month follow-up and that a sensitvity analysis could be done to assess the consequences of a longer period of presentation of SCAR's.*
   He added that the health economic analysis is rather summary and should be removed, unless it is properly done. It would be a stand-alone paper in its own right. The findings of an economic review would support whether this should be routine practice, so the authors should not claim this should be routine practice.
   Response: We greatly appreciate the positive comment raised by the reviewer.
   (A) The comment regarding the time period of the historical controls is interesting. The time period was chosen based on an important study by our group in PNAS in 2005. In this study (Hung *et al.*, Proc Natl Acad Sci USA 2005;102:4134-9), we provided evidence to support the causal relationship between allopurinol-induced adverse reactions and harbouring HLA-B*58:01, which strongly suggests that HLA-B*58:01 can serve as a marker that could be used to prevent allopurinol-induced SCARs. After our 2005 PNAS paper, some physicians in Taiwan began to genotype the HLA-B region before initiating allopurinol treatment. While this was a beneficial measure, it may have caused a confounding effect on our analysis. Therefore, to obtain a suitable control group, we adopted the most recent, non-confounded data (i.e. the 2001-2004 data) from the National Health Insurance Research Database (NHIRD). In response to this comment, we

have mentioned the material discussed above in the revised manuscript (lines 3-7 on page 13).

(B) As the editor mentions, a sensitivity analysis could be performed to assess the consequences of using a longer period of presentation of SCARs. All study participants were followed-up for at least 9 months and the follow-up is still ongoing. To date (June 2nd, 2015), based on our genetic screening protocol, we have found no cases of SCARs in our participants; therefore, we suggest that the incidence of SCARs at the 2-month follow-up was the same as the incidence of SCARs at the 9-month follow-up. In response to this critical comment, we have included some of these details in the revised manuscript (lines 14-17 on page 10) (lines 12-19 on page 17). Following the reviewer's suggestion, we attempted to identify possible SCARs in this prospective cohort by screening the NHIRD using the identification numbers of individual Taiwanese patients; no new SCAR cases were identified through this approach.

(C) As mentioned above, we agree that a more detailed pharmacoeconomic analysis should be presented and plan to do so in an independent article. We thought that it might be helpful for the readers of BMJ to acknowledge the importance of particular consideration of pharmacoeconomics in the practise of genetic screening to prevent adverse drug reactions. In response to this comment, we have removed economic data and only briefly mention the importance of pharmacoeconomic considerations in the revised manuscript (lines 9-11 on page 19).

10. *Finally, another editor did not have a strong feeling about this paper. Nevertheless, she acknowledged that the other papers in this area were well cited, which was one point in favour. She considered that this is a question about standard of care. She felt that the question is not whether it's ethical to give allopurinol to someone who has tested positive for the HLA subtype of concern. The question is whether it's ethical to give it to somebody who hasn't been tested if testing is available.*

Response: We appreciate and agree the comment that "*the question is whether it's ethical to give it to somebody who hasn't been tested if testing is available"*. Because severe cutaneous adverse reactions (SCARs) are life-threatening events, they should be prevented when the underlying mechanisms are well understood and prevention measures are available, as in the case of the use of HLA-B*58:01 allele genotyping to prevent allopurinol-induced SCARs. Indeed, the purpose of our prospective study is to provide the evidence that screening can make such

prevention a clinical reality.

Responses to the reviewers' comments

Reviewer 1.

We greatly appreciate the positive comments made by the reviewer**.**

The following points were raised by the reviewer:

1. *In the definition of SCR, DRESS was a qualifying presentation. How likely is it that some of the patients with systemic adverse reactions while on allopurinol, reported in Table 2, might in fact have had DRESS? The authors ought to describe what quality control procedures they had in place to make sure that no case is misclassified and not reported? For Stevens-Johnson syndrome, or toxic epidermal necrolysis this is more unlikely given the severity of the phenotype.*
   Response: We appreciate and agree with the reviewer's comment. As mentioned by the reviewer, DRESS is more complicated than SJS or TEN because of its diverse presentation. To qualify presentation in a standard way, diagnosis of DRESS is based on criteria established by the RegiSCAR Group Diagnosis (Roujeau *et al.,* Toxicology 2005;209:123-129; Cacoub *et al.,* 2011;124:588-597; Kardaun *et al.,* Br J Dermatol 2013;169:1071-1080). Based on this definition, DRESS diagnosis is primarily based on two important characteristics: multi-systemic involvement and frequent eosinophilia. In this prospective study, we did not identify cases with multi-systemic involvement or frequent eosinophilia among our study participants.

2. *For estimating the number of cases of SCR avoided the authors used rather old 2001-2004 data. Why were more recent data not used to ensure that any misclassification biases are minimised or at least consistent for both their active cohort and the historical cohorts?*
   Response: The comment regarding the time period chosen for our historical controls is interesting and merits explanation. The control time period was chosen based on an important study published by our group in PNAS in 2005. In this study (Hung *et al.,* Proc Natl Acad Sci USA 2005;102:4134-9), we provided evidence to support the causal relationship between allopurinol-induced adverse reactions and harbouring HLA-B*58:01, which strongly suggested that HLA-B*58:01 could serve as a marker to prevent allopurinol-induced SCARs. After this publication, some physicians in Taiwan began to genotype HLA-B region before starting allopurinol treatment. This measure was certainly beneficial,

but inclusion of data from such patients could confound our analysis. Therefore, to obtain a suitable control group, we adopted the most recent, non-confounded data, which was that collected prior to our 2005 publication. We have mentioned the rationale for selecting the control time period in the revised manuscript (lines 3-7 on page 13).

3. *In their database data-trawl, they identified patients who had been on at least 3 months of allopurinol. Yet their follow-up is only of 2-month duration in their prospective study. Given that the time lag between exposure and SCR is rather ill-defined and may be as long as more than 12 weeks (Ramasamy et al. 2013) how confident are they that their allopurinol-treated cohort may not yet develop SCR?*

   Response: This is a very critical issue related to the validity of our results. We thank you for the opportunity to address this issue in more detail. We are aware that there are very rare cases that SCAR occur more than 2 months of allopurinol exposure, however, the 2-month follow-up had been carefully chosen at the beginning of the study was based on evidence reported in a study published by our group in PNAS in 2005 (Hung *et al.*, Proc Natl Acad Sci USA 2005;102:4134-9). Our 2005 PNAS study was one of the largest studies conducted on the history of allopurinol-induced severe cutaneous adverse reactions (SCARs). In our 2005 PNAS study, all allopurinol-related SCAR patients had SCAR onset within the first 2 months of allopurinol use.

   Table 1 from our 2005 PNAS paper is shown on the next page to clearly demonstrate that all patients developed SCARs within 2 months and show that the longest exposure was 56 days. This finding has been consistently reported in other studies, which suggested that “the typical latency to onset is 2 to 6 weeks” (http://livertox.nih.gov/Allopurinol.htm).

Table 1. Clinical characteristics of patients with allopurinol-induced SCAR

Patient no.	Age/sex	Phenotype	Dose, mg/d/ duration of drug exposure, days	Exanthema %/blister or detachment,† %	Mucous membrane erosions, number:site	Systemic manifestations
1	80/F	SJS	100/33	31–50/2	(+) 2: Oral, eyes	None
2	47/F	HSS	100/23	31–50/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
3	60/F	SJS	100/14	31–50/5	(+) 2: Oral, eyes	Fever, worsening RF, leukocytosis, atypical lymphocytosis‡
4	25/M	SJS	200/24	51–70/5	(+) 1: Oral	Fever, LFI, leukocytosis, eosinophilia
5	50/F	HSS	300/33	51–70/0	(+) 1: Oral	Fever, LFI, eosinophilia
6	74/F	HSS	100/22	31–50/0	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia, atypical lymphocytosis
7	78/M	HSS	NA/21	71–90/0	(−)	Worsening RF, eosinophilia
8	37/F	SJS/TEN	100/38	51–70/10	(+) 3: Oral, eyes, genital	Fever, LFI, leukocytosis, eosinophilia
9	18/F	HSS	200/21	51–70/0	(−)	LFI, worsening RF, leukocytosis, eosinophilia
10	63/F	SJS	200/37	51–70/5	(+) 2: Oral, genital	LFI, worsening RF, eosinophilia
11	55/F	HSS	100/35	51–70/0	(−)	Fever, LFI, leukocytosis
12	70/F	HSS	NA/56	31–50/0	(+) 2: Oral, eyes	Fever, LFI, leukocytosis, eosinophilia
13	52/F	SJS	100/35	51–70/5	(+) 3: Oral, eyes, genital	Fever, LFI, eosinophilia
14	29/F	SJS	NA/28	51–70/5	(+) 2: Oral, eyes	None
15	73/M	SJS	200/30	51–70/1	(+) 3: Oral, eyes, genital	Fever
16	78/M	SJS	100/1§	51–70/1	(+) 1: Oral	Worsening RF, eosinophilia
17	70/M	SJS/TEN	NA/27	51–70/10	(+) 2: Oral, eyes	Fever
18	69/F	HSS	100/53	31–50/0	(+) 1: Oral	LFI, worsening RF, eosinophilia
19	91/F	TEN	100/26	71–90/30	(+) 2: Oral, eyes	Atypical lymphocytosis
20‖	60/M	SJS/TEN	200/21	51–70/15	(+) 2: Oral, eyes	Fever, LFI, worsening RF, eosinophilia
21	77/F	SJS	NA/14	51–70/5	(+) 2: Oral, genital	Worsening RF
22	62/F	HSS	100/38	51–70/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
23	41/M	HSS	200/32	51–70/0	(+) 1: Oral	LFI, leukocytosis
24	72/F	HSS	100/45	>90/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
25	51/M	SJS	200/2§	51–70/5	(+) 3: Oral, eyes, genital	None
26	35/M	HSS	NA/30	31–50/0	(−)	Fever, leukocytosis, eosinophilia
27	65/F	HSS	300/49	51–70/0	(−)	Leukocytosis, eosinophilia
28‖	54/F	SJS	100/14	51–70/3	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia
29	51/M	SJS	NA/21	51–70/5	(+) 1: Oral	Fever, LFI, atypical lymphocytosis
30	52/F	HSS	NA/45	51–70/0	(+) 1: Oral	Fever, LFI, leukocytosis, severe myositis
31	50/M	HSS	50/15	31–50/0	(−)	LFI, worsening RF
32	80/M	SJS	NA/12	31–50/5	(+) 2: Oral, genital	None
33	80/M	HSS	100/21	31–50/0	(−)	LFI, worsening RF, leukocytosis, eosinophilia
34	70/M	HSS	50/18	31–50/0	(−)	LFI, leukocytosis, eosinophilia, atypical lymphocytosis
35	52/M	HSS	100/7	71–90/0	(−)	Leukocytosis, eosinophilia
36	63/M	HSS	100/52	31–50/0	(−)	LFI, eosinophilia
37	67/F	HSS	100/21	51–70/0	(−)	Fever, worsening RF, eosinophilia
38	55/M	HSS	100/14	51–70/0	(−)	Fever, worsening RF, eosinophilia
39‖	66/F	TEN	100/23	71–90/30	(+) 3: Oral, eyes, genital	Leukopenia
40	62/M	HSS	300/30	71–90/0	(−)	LFI, worsening RF, eosinophilia
41	73/F	HSS	100/39	71–90/0	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia
42	84/M	HSS	100/30	51–70/0	(−)	Worsening RF, eosinophilia
43	24/M	HSS	NA/28	51–70/0	(−)	Fever, worsening RF, eosinophilia
44‖	78/F	HSS	100/30	71–90/0	(+) 1: Oral	Fever, worsening RF, eosinophilia, leukocytosis
45	69/F	HSS	100/26	51–70/0	(−)	Worsening RF, leukocytosis, eosinophilia
46	73/M	SJS/TEN	100/1§	51–70/10	(+) 3: Oral, eyes, genital	LFI, worsening RF
47‖	73/F	HSS	100/30	71–90/0	(+) 1: Oral	Fever, LFI, worsening RF, leukocytosis, eosinophilia
48	70/F	SJS/TEN	200/12	51–70/15	(+) 2: Oral, eyes	Fever, leukocytosis
49	70/M	HSS	100/20	51–70/0	(+) 3: Oral, eyes, genital	LFI, worsening RF, eosinophilia, atypical lymphocytosis
50	71/M	TEN	100/26	71–90/30	(+) 1: Oral	Fever, LFI, worsening RF
51	71/M	HSS	100/7	51–70/0	(+) 1: Oral	Fever, LFI, worsening RF, eosinophilia

See *Methods* for the diagnostic criteria. NA, data not available; RF, renal function; LFI, liver function impairment; M, male; F, female.
†Exanthema, erythematous or purpuric macules/papules; blister or detachment, extent of blisters or epidermal detachment; both exanthema and blister/detachment were expressed as % of total body surface area.
§Second attack on reexposure.
‖Cases deceased.

We are continuing to follow the study participants. All patients have been followed for at least 9 months and none have developed SCARs. We have also performed the approach suggested by the reviewer to identify possible SCARs in these prospective cohort members by linkage with the National Health Insurance Research Database (NHIRD) dataset based on the unique identification numbers of individual Taiwanese patients and did not identify SCARs. Moreover, as one editor mentions, a sensitivity analysis could be performed to assess the consequences of using a longer period of presentation of SCARs. To date (June 2nd, 2015), based on our genetic screening protocol, we have found no cases of SCARs in our participants; thus, we suggest that the incidence of SCARs at the

2-month follow-up is the same as the incidence of SCARs at the 9-month follow-up.
In response to this critical comment, we have included some relevant information in the revised manuscript (lines 14-17 on page 10) (lines 12-19 on page 17).

4. *In their 'cost-effectiveness' study the authors assume equi-effectiveness (beneficial and adverse effects). Comparative data suggests that febuxostat may well have more adverse cardiovascular and hepatic effects profiles (see product label) although the available data are rather sparse. How should trade-offs be made between these adverse effects with febuxostat and potential SCAR with allopurinol? Given this uncertainty, the authors should be more reserved in their claims.*
Response: We agree with this comment and concur that a pharmacoeconomic analysis should be presented in more detail. We plan to provide such an analysis in an independent article. We thought that it might be helpful for the readers of BMJ to acknowledge the importance of particular consideration of pharmacoeconomic issues in the practise of genetic screening to prevent adverse drug reactions.
In response to this comment, which was also raised by the editors, we removed the section in question from the manuscript and only briefly mention the importance of pharmacoeconomic analysis in the revised manuscript (lines 9-11 on page 19).

5. *By their own estimates the positive predictive value is only 2%. Given this, the point raised under (iv) is particularly important.*
Response: We appreciate the reviewer's comment. Please see the response above, we remove this part from the manuscript, and only briefly mention the importance of pharmacoeconomic analysis in the revised manuscript (lines 9-11 on page 19).

6. *The authors make inadequately unsupported generalisations about potential use in other populations for which only very limited data are available. I think that they should refrain from this particularly given the lower and/or ill-defined prevalence of the influential allele in those populations.*
Response: We agree with this comment and incorporated the limitation into the revised manuscript (lines 9-11 on page 22).
The issue regarding whether the findings of this study can be applied to people who are not of Han Chinese descent is important. This can be properly addressed by examining findings of studies independently conducted in other populations along with our recent experimental findings. As mentioned in our manuscript (lines 15-19 on page 21) (lines 1-2 on page 22), we suggest that the application of our findings may not be limited to individuals of Han Chinese descent,

because different independent groups have replicated and confirmed the association between HLA-B*58:01 and allopurinol-induced severe cutaneous adverse reactions (SCARs), originally reported in Han Chinese by us in 2005 (Hung et al., Proc Natl Acad Sci USA.2005;102:4134-9), and later in multiple populations and races (e.g. Thai, Japanese, Korean, and European). In addition, based on our recent mechanistic study (Lin et al., J Allergy Clin Immunol 2015;135(4):1063-5.e5), HLA-B*58:01 can directly present allopurinol metabolites (e.g. oxypurinol) to cytotoxic T cells without antigen processing, while allopurinol/oxypurinol-specific T cell-mediated cytotoxicity is restricted to carriers of HLA-B*58:01, suggesting that HLA-B*58:01 is the key determinant of susceptibility to allopurinol-induced SCARs and that HLA-B*58:01 carriers are susceptible to SCARs when taking allopurinol, regardless of their ethnic group. We have mentioned this point in the revised manuscript (lines 7-10 on page 8). Therefore, in the manuscript, we emphasize that HLA-B*58:01 screening could be beneficial for preventing allopurinol-induced SCARs in countries where HLA-B*58:01 is relatively prevalent (not restricted only to Han Chinese). In addition, even in countries where the allele frequency of HLA-B*58:01 is relatively low (~ 1%), HLA-B*58:01 screening may be beneficial in a subset of high-risk patients (e.g. chronic renal failure). For countries and populations, in which the prevalence is ill-defined, further investigations are suggested to estimate the prevalence for possible application of this screening. We have added this information in the revised manuscript (lines 2-11 on page 22).

Therefore, because HLA-B*58:01 is a relatively common allele in multiple populations, we suggest that genetic screening prior to prescription of allopurinol could be a standard of care.

Reviewer 2.

We greatly appreciate the very positive comments made by the reviewer.

The following points were raised by the reviewer:

1. *The entire basis of the conclusion is based on the comparison to the historical incidence of allopurinol-induced SCARs. The authors use data from 2001-2004; however, there is no clear rationale why this time period was selected. A wider selection of time periods, or at least some more recent data are needed to strengthen their findings.*
   Response: The question regarding the time period we selected for our historical controls merits explanation. The time period was selected based on an important study by our group in PNAS in 2005. In this study (Hung *et al.*, Proc Natl Acad Sci USA 2005;102:4134-9), we provided convincing evidence to support the causal relationship between allopurinol-induced adverse reactions and harbouring HLA-B*58:01, which strongly suggested that HLA-B*58:01 could serve as a marker to allow prevention of allopurinol-induced SCARs. After this publication, some physicians in Taiwan began to genotype the HLA-B region before administering allopurinol to patients. This measure was certainly beneficial, but inclusion of data from such patients may lead to confounding effects in our study. Therefore, to obtain a suitable control group, we adopted the most recent, non-confounded data prior to our 2005 publication. We have mentioned this rationale in the revised manuscript (lines 3-7 on page 13).

2. *The authors include p=0.0026 in the abstract and throughout the manuscript as the primary p value; however, it is not one of the p-values in Table 3 and does not seem to be an average value from their time period. This requires clarification.*
   Response: We appreciate the reviewer's suggestion. We combined the data from 2001 to 2004 together, and did the analysis, resulting in a p value of 0.0026. Based on the suggestions from the reviewer and editor, we analysed our data using the two-side one-sample binomial test (**Table 3**).

3. *The estimated incidence of allopurinol-induced SCARs was calculated by SCARs cases divided by annual number of new allopurinol users. The SCARs number was based on ICD-9 code; were any of these cases validated or confirmed manually (or otherwise) to be true allopurinol-induced SCARs and not another drug induced SCARs? Any confirmation data is needed to support these numbers, or if*

they are not necessary, an explanation as to why not is warranted.

Response: In our study design for the identification of allopurinol-induced SCARs from the National Health Insurance Research Database (NHIRD), we purposely excluded patients who were also administered other medications that may cause SCARs such as carbamazepine and phenytoin.

The estimated incidence of allopurinol-induced SCARs was calculated by dividing the number of SCAR cases by the annual number of new allopurinol users as defined by the ICD-9-CM 695.1 code which covers all SCARs. This information was obtained from the Taiwan NHIRD, which was established at the same time as the National Health Insurance (NHI) system in 1995. The NHI is a mandatory single-payer social health insurance system run by the government that procures high-quality healthcare with national coverage in Taiwan. The population enrolment rate was 99.5% in 2008. Furthermore, 95% of all healthcare facilities in the country are contracted through the NHI system. The NHIRD is an internationally recognized and exceptionally accurate database of the medical status of the Taiwanese population; clinical information of most patients is included in the database. The NHIRD collects registration data including those from the registry of medical services and drug prescriptions. The Bureau of NHI reviews reimbursement claims and screens the type, volume, quality, and suitability of medical services provided by the NHI program, in order to guarantee NHIRD data quality and accuracy. **For serious diseases such as SCARs, they are routinely reviewed, and diagnoses are validated and confirmed manually by the National Health Insurance Bureau. In addition, when we performed this study, to estimate the incidence of allopurinol-induced SCARs, diagnosis of allopurinol-induced SCARs was confirmed after crosschecking with the NHIRD and patients' medical history to confirm the diagnosis of SCAR and to ensure that the SCARs were not due to other medications.**

4. *The most common medication taken for the positive subjects was benzbromarone. Can the authors include more discussion about the efficacy of this alternative and why these results do not suggest just using this alternative for all gout patients instead of genotype-directed selection? Is this driven by cost differences or efficacy or both?*

Response: Thank you for the comments. Benzbromarone is a uricosuric agent that has been used to control hyperuricemia. It is effective in lowing serum uric acid levels, especially in patients with urate under-excretion. However, benzbromarone has a risk of severe hepatotoxicity as well as acute renal colic, and it has been withdrawn from the market or not available in many countries, including US and

some European countries (Lee et al., Drug Saf. 2008;31:643-65). These are the major reasons why allopurinol (a xanthine oxidase inhibitor) ultimately replaced benzbromazone and other uricosuric medications as the treatment of choice for hyperuricemia and the reasons why benzbromarone is not used in all gouty patients. Instead we proposed in this study the genotype-directed selection of patients on allopurinol for efficacy and safety considerations. In response to this comment, we have added this information in the revised manuscript (lines 11-19 on page 18) (lines 1-2 on page 19).

5. *Citing and commenting on the CPIC guideline for HLA-B/allopurinol in the Introduction is suggested (PMID: 23232549).*
   Response: We have cited the paper as suggested by the reviewer in the Discussion section (lines 1-2 on page 18; Reference No.32).

6. *Introduction, page 7, line 32-35: could be refined to remove the term 'is rather lethal'.*
   Response: We agree with this comment, and have revised the manuscript as suggested by the reviewer (lines 6-7 on page 7).

7. *Results, page 14, lines 38-39: suggest using 'counseled' instead of 'given advice'.*
   Response: We agree with this comment and have revised the manuscript as suggested by the reviewer (lines 13-14 on page 14).

8. *Discussion, page 18, first sentence: suggest rewording, 'prior to' instead of 'before', 'subsequently' instead of 'then', and 'would likely' instead of 'could'.*
   Response: We agree with this comment and have revised the manuscript as suggested by the reviewer (lines 4-5 on page 17).

9. *Discussion, second paragraph: suggest 'Our results support HLA...' instead of 'Our results suggest the merit of HLA...'.*
   Response: We agree with this comment and have revised the manuscript as suggested by the reviewer (line 1 on page 18).

10. *Discussion, page 20, line 51: suggest removing the word 'culprit'.*
    Response: We agree with this comment and have revised the manuscript as suggested by the reviewer (line 8 on page 20).

11. *Discussion, Strengths and Limitations of Study section: appears to be written by a*

different person. Acronyms are incorrectly re-defined, SCARs is written out incorrectly, etc. Suggest revising this section for consistency.

Response: We appreciate this comment and have revised the section for consistency as suggested by the reviewer (lines 17-19 on page 20) (lines 1-19 on page 21) (lines 1-15 on page 22).

12. *Table 1: Can p-values be included between positive and negative cases for the clinical characteristics?*

    Response: We appreciate your suggestion and have revised the manuscript as suggested by the reviewer (**Table 1**). All p values of clinical characteristics between positive and negative cases were statistically non-significant.

Reviewer 3.

We greatly appreciate the positive comments made by the Reviewer.

The following points were raised by the reviewer:

1. *One aspect that is worth clarifying is the approaches used to identify individuals with SCARs. In the historical control group this undertaken on the basis of ICD codes recorded in a national health insurance database, whereas in the prospective cohort it was determined via an interview with the subject approximately 2 months after the initial screening. It is implicitly assumed that these two different approaches for identifying SCARs are equivalent. It would be useful to demonstrate that using using the NHIRD data to identify SCARs in the prospective cohort gives the same results as the interview.*

   Response: We appreciate this comment. We used historical incidence in the present study, in which the incidence of allopurinol-induced adverse drug reaction (ADR) was calculated based on information from an internationally recognised and highly reliable database, the National Health Insurance Research Database (NHIRD) of Taiwan. Thus, the reliability of the NHIRD is critical in assessing the validity of our estimation. To address this issue, it should be noted that the NHIRD was established in Taiwan when the National Health Insurance (NHI) system was launched in 1995. The NHI is a mandatory single-payer social health insurance system administered by the government that provides universal high-quality health care for almost all people in Taiwan, with an enrolment rate of 99.5% (in 2008). Furthermore, 95% of all health care facilities in the country have been contracted by the NHI system. Therefore, the NHIRD is an exceptionally accurate database regarding the medical status of the Taiwanese population, because information from the vast majority of such patients is included in the database. The Taiwanese NHI system has been recognized as one of the best health insurance systems in the world and has an international reputation for high-quality patient diagnosis and care. The NHIRD collects registration data, including data from the registry of medical services and registry of drug prescriptions. The Bureau of National Health Insurance is required to review reimbursement claims and to screen the type, volume, quality, and appropriateness of medical services provided under NHI program, ensuring the quality and accuracy of the NHIRD. More than 400 papers based on the NHIRD have been published in international peer-reviewed journals (for a detailed list,

please see http://w3.nhri.org.tw/nhird/en/Research.html.). The approach of using historical incidence and the NHIRD has been utilized in many studies, including a study from our group that successfully demonstrated screening of HLA-B*15:02 to prevent carbamazepine-induced toxic effects in 2011 (Chen *et al.*, N Engl J Med. 2011;364:1126-33.). For a nationwide prospective study to recruit patients from different participating hospitals, the use of highly valid national statistical data from the NHIRD is more reliable and much more comprehensive than performing a direct comparison of individual hospitals.
In the prospective cohort, possible SCAR was determined via an interview with the subject 2 months after the initial screening. We also attempted to identify SCARs using the NHIRD dataset based on the unique identification numbers of individual Taiwanese and no SCARs were identified, which was consistent with the results by interview.
Therefore, in response to Reviewer's comment, we ensure that these two approaches for identifying SCARs are equivalent. We have added this information in the revised manuscript (lines 8-17 on page 12) (lines 12-19 on page 17).

2. *Page 18 - implication for clinical practice discussion. Discussion of cost-effectiveness of screening of the allele is warranted but the approach undertaken is somewhat limited. Results of cost-effectiveness analyses should be reported in the results section and the methods used described in the method section. The approach and results included in the discussion are rather simplistic and it is advised to either undertaken this in a more comprehensive manner or perhaps leave this for another study to focus on exclusively…..*
Response: We agree with this comment and concur that a more detailed pharmacoeconomic analysis should be presented. We plan to publish such an analysis in an independent article. We thought that it might be helpful for the readers of BMJ for us to acknowledge the importance of particular consideration of pharmacoeconomic issues in the practise of genetic screening to prevent adverse drug reactions.
In response to this comment, which was also made by the editors and another reviewer, we have removed the pharmacoeconomic analysis and only briefly mention the importance of pharmacoeconomic consideration in implementing personalized medicine (lines 9-11 on page 19).

3. *page 19 line 15: "which may have a huge impact on reducing the number of patients with allopurinol-induced SCARs in the population" Rather than describing this as a huge impact it may be more informative to the reader to*

*provide some estimates of the impact in the population. For example, one could say (based on numbers in table 3) that screening approximately 110000 new users of allopurinol in Taiwan each year for HLA-B*58:01 may prevent approximately 330 cases of allopurinol-induced SCARs each year.*

Response: We appreciate this comment and have revised the manuscript based on this suggestion (lines 4-7 on page 20).

Reviewer 4.

The following points raised by the reviewer are the same as some of the points raised in the manuscript meeting. Therefore, our responses to these comments are the same as those indicated in our response letter.

1. *The authors only followed participants for two months. Could they please provide stronger evidence and data supporting that adverse events are likely to occur within 2 months? (the current phrase is that "in general SCARs onset occurs within 2 months" which is not particularly convincing that all cases would have been detected). It would seem a weakness of the study not to have followed-up individuals for a longer time period, particularly seeing that the control group data are based on 3-monthly prescriptions.*

   Response: This is a very critical issue related to the validity of our results and we thank you for the opportunity to address this issue in more detail. We are aware that there are very rare cases that SCAR occur more than 2 months of allopurinol exposure, however, the 2-month follow-up period carefully determined at the beginning of this study was based on evidence from a previous study published by our group in PNAS in 2005 (Hung *et al.*, Proc Natl Acad Sci USA 2005;102:4134-9), which was one of the largest studies to report the history of allopurinol-induced severe cutaneous adverse reactions (SCARs). In this study, all allopurinol-related SCAR patients showed SCAR onset within the first 2 months of allopurinol use. Included below is Table 1 from our 2005 PNAS paper, in which we clearly demonstrate that all 51 patients developed SCARs within 2 months and show that the longest exposure was 56 days. This finding has been consistently reported in other studies, which suggested that "the typical latency to onset is 2 to 6 weeks" (http://livertox.nih.gov/Allopurinol.htm).

Table 1. Clinical characteristics of patients with allopurinol-induced SCAR

Patient no.	Age/sex	Phenotype	Dose, mg/d/ duration of drug exposure, days	Exanthema %/blister or detachment,† %	Mucous membrane erosions, number:site	Systemic manifestations
1	80/F	SJS	100/33	31–50/2	(+) 2: Oral, eyes	None
2	47/F	HSS	100/23	31–50/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
3	60/F	SJS	100/14	31–50/5	(+) 2: Oral, eyes	Fever, worsening RF, leukocytosis, atypical lymphocytosis‡
4	25/M	SJS	200/24	51–70/5	(+) 1: Oral	Fever, LFI, leukocytosis, eosinophilia
5	50/F	HSS	300/33	51–70/0	(+) 1: Oral	Fever, LFI, eosinophilia
6	74/F	HSS	100/22	31–50/0	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia, atypical lymphocytosis
7	78/M	HSS	NA/21	71–90/0	(−)	Worsening RF, eosinophilia
8	37/F	SJS/TEN	100/38	51–70/10	(+) 3: Oral, eyes, genital	Fever, LFI, leukocytosis, eosinophilia
9	18/F	HSS	200/21	51–70/0	(−)	LFI, worsening RF, leukocytosis, eosinophilia
10	63/F	SJS	200/37	51–70/5	(+) 2: Oral, genital	LFI, worsening RF, eosinophilia
11	55/F	HSS	100/35	51–70/0	(−)	Fever, LFI, leukocytosis
12	70/F	HSS	NA/56	31–50/0	(+) 2: Oral, eyes	Fever, LFI, leukocytosis, eosinophilia
13	52/F	SJS	100/35	51–70/5	(+) 3: Oral, eyes, genital	Fever, LFI, eosinophilia
14	29/F	SJS	NA/28	51–70/5	(+) 2: Oral, eyes	None
15	73/M	SJS	200/30	51–70/1	(+) 3: Oral, eyes, genital	Fever
16	78/M	SJS	100/1§	51–70/1	(+) 1: Oral	Worsening RF, eosinophilia
17	70/M	SJS/TEN	NA/27	51–70/10	(+) 2: Oral, eyes	Fever
18	69/F	HSS	100/53	31–50/0	(+) 1: Oral	LFI, worsening RF, eosinophilia
19	91/F	TEN	100/26	71–90/30	(+) 2: Oral, eyes	Atypical lymphocytosis
20‖	60/M	SJS/TEN	200/21	51–70/15	(+) 2: Oral, eyes	Fever, LFI, worsening RF, eosinophilia
21	77/F	SJS	NA/14	51–70/5	(+) 2: Oral, genital	Worsening RF
22	62/F	HSS	100/38	51–70/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
23	41/M	HSS	200/32	51–70/0	(+) 1: Oral	LFI, leukocytosis
24	72/F	HSS	100/45	>90/0	(+) 2: Oral, eyes	Fever, LFI, worsening RF
25	51/M	SJS	200/2§	51–70/5	(+) 3: Oral, eyes, genital	None
26	35/M	HSS	NA/30	31–50/0	(−)	Fever, leukocytosis, eosinophilia
27	65/F	HSS	300/49	51–70/0	(−)	Leukocytosis, eosinophilia
28‖	54/F	SJS	100/14	51–70/3	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia
29	51/M	SJS	NA/21	51–70/5	(+) 1: Oral	Fever, LFI, atypical lymphocytosis
30	52/F	HSS	NA/45	51–70/0	(+) 1: Oral	Fever, LFI, leukocytosis, severe myositis
31	50/M	HSS	50/15	31–50/0	(−)	LFI, worsening RF
32	80/M	SJS	NA/12	31–50/5	(+) 2: Oral, genital	None
33	80/M	HSS	100/21	31–50/0	(−)	LFI, worsening RF, leukocytosis, eosinophilia
34	70/M	HSS	50/18	31–50/0	(−)	LFI, leukocytosis, eosinophilia, atypical lymphocytosis
35	52/M	HSS	100/7	71–90/0	(−)	Leukocytosis, eosinophilia
36	63/M	HSS	100/52	31–50/0	(−)	LFI, eosinophilia
37	67/F	HSS	100/21	51–70/0	(−)	Fever, worsening RF, eosinophilia
38	55/M	HSS	100/14	51–70/0	(−)	Fever, worsening RF, eosinophilia
39‖	66/F	TEN	100/23	71–90/30	(+) 3: Oral, eyes, genital	Leukopenia
40	62/M	HSS	300/30	71–90/0	(−)	LFI, worsening RF, eosinophilia
41	73/F	HSS	100/39	71–90/0	(+) 2: Oral, eyes	Fever, worsening RF, eosinophilia
42	84/M	HSS	100/30	51–70/0	(−)	Worsening RF, eosinophilia
43	24/M	HSS	NA/28	51–70/0	(−)	Fever, worsening RF, eosinophilia
44‖	78/F	HSS	100/30	71–90/0	(+) 1: Oral	Fever, worsening RF, eosinophilia, leukocytosis
45	69/F	HSS	100/26	51–70/0	(−)	Worsening RF, leukocytosis, eosinophilia
46	73/M	SJS/TEN	100/1§	51–70/10	(+) 3: Oral, eyes, genital	LFI, worsening RF
47‖	73/F	HSS	100/30	71–90/0	(+) 1: Oral	Fever, LFI, worsening RF, leukocytosis, eosinophilia
48	70/F	SJS/TEN	200/12	51–70/15	(+) 2: Oral, eyes	Fever, leukocytosis
49	70/M	HSS	100/20	51–70/0	(+) 3: Oral, eyes, genital	LFI, worsening RF, eosinophilia, atypical lymphocytosis
50	71/M	TEN	100/26	71–90/30	(+) 1: Oral	Fever, LFI, worsening RF
51	71/M	HSS	100/7	51–70/0	(+) 1: Oral	Fever, LFI, worsening RF, eosinophilia

See *Methods* for the diagnostic criteria. NA, data not available; RF, renal function; LFI, liver function impairment; M, male; F, female.
†Exanthema, erythematous or purpuric macules/papules; blister or detachment, extent of blisters or epidermal detachment; both exanthema and blister/detachment were expressed as % of total body surface area.
§Second attack on reexposure.
‖Cases deceased.

We are continuing to follow the study participants. All study participants have been followed for at least 9 months and none have developed SCARs. We have also followed the approach suggested by the reviewer to identify possible SCARs in these prospective cohort members by linkage with the National Health Insurance Research Database (NHIRD) dataset based on unique identification numbers of individual Taiwanese patients, but no SCARs were identified using this method. Moreover, as one editor mentioned, a sensitivity analysis could be performed to assess the consequences of using a longer period of presentation of SCAR's. To date (June 2nd, 2015), based on our genetic screening protocol, we have not identified cases of SCARs in the study participants; thus, we suggest that

the incidence of SCARs found in the 2-month follow-up was equal to the incidence of SCARs found in the 9-month follow-up.
In response to this critical comment, we have included some details regarding this data in the revised manuscript (lines 14-17 on page 10) (lines 12-19 on page 17).

2. *The power calculation is written as if the comparison is being made between a known fixed value of 0.3% and a hypothesised value of 0.03% to be estimated within the sample. Given that the comparison is known and estimated with high precision this seems reasonable. However, the computed sample size does not agree with this (my sample size calculator using a normal approximation suggests that the stated sample size of 2169 would have 85.8% power (not the 99% stated). Could the authors fully justify and explain their calculation.*
Response: The original power calculation was performed under two-sample one-side test to determine whether the incidence of new approach (with HLA-B*58:01 genotyping screening) was lower than the incidence of previous approach (without HLA-B*58:01 genotyping screening).

   We have followed your suggestions and use one-sample binomial test. The power of the two-side one-sample binomial test is about 86%, assuming p0 = 0.003 and p1 = 0.0003, with normal distribution approximation. We have revised manuscript in response to this critical comment (lines 12-14 on page 13).

3. *The statistical method the authors state has been used is a Fisher's Exact test. This is a two group test and requires data from both the screening cohort and the comparison group. It is not clear what data has been used for the comparison group in the analysis presented in the paper. However, in line with point 2) the comparison between the observed result in the collected cohort and the historical figure of 0.3% might be better based on a one sample Binomial test (to work out the probability that the see whether the observed incidence rate differs from the fixed value of 0.3%). My estimate of the associated P-value from this test is 0.0003 – thus similar to that reported by the authors.*
Response: We have followed your suggestions and used a one-sample binomial test. The p-value is equal to 0.0026, assuming X ~ B (p0 = 0.003, n = 2169) with a normal distribution approximation (lines 5-6 on page 6; lines 16-17 on page 16) (**Table 3**).

4. *It is important that the authors report the 95% confidence interval for the observed event rate in the cohort – (0% to 0.13%) as this gives an upper limit on the observed event rate.*

   Response: The 95% confidence interval is added as (0%, 0.17%) (lines 5-6 on page 6; lines 16-17 on page 16).