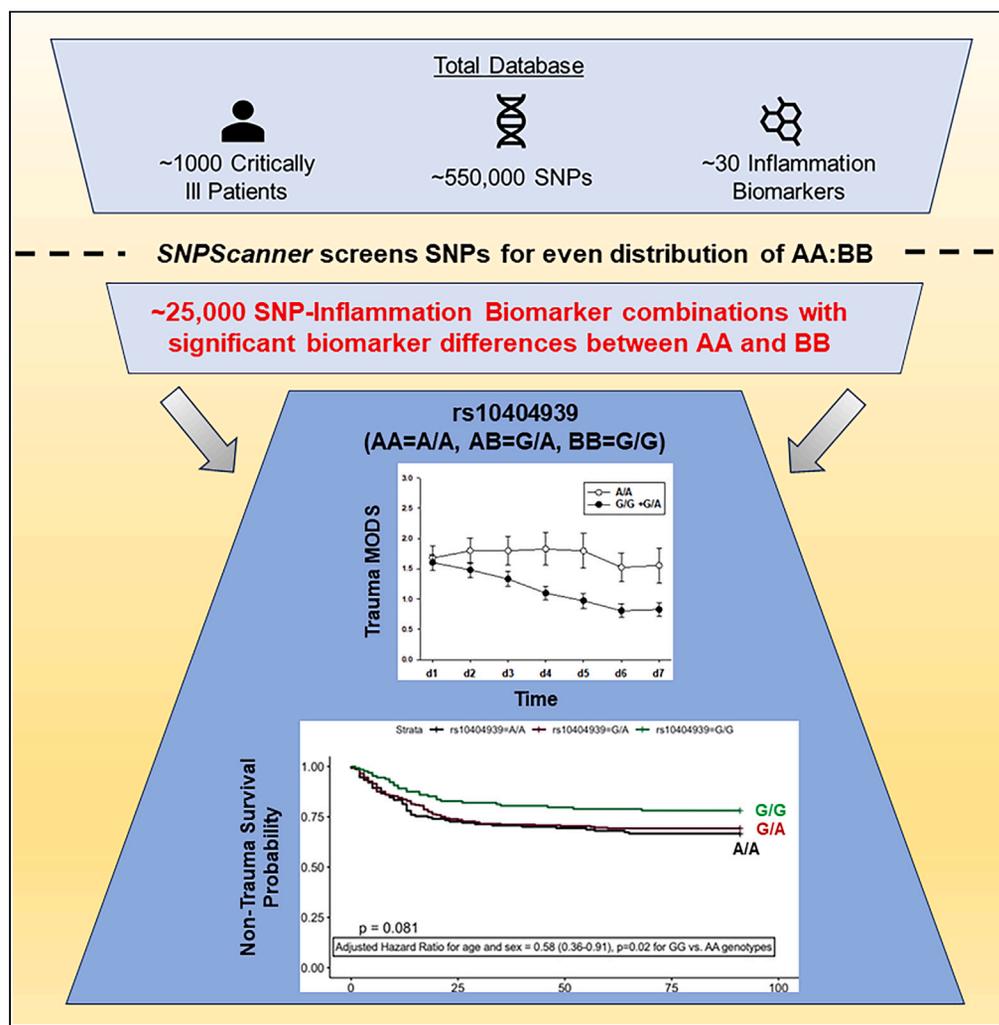**Article**

# A common single nucleotide polymorphism is associated with inflammation and critical illness outcomes

Fayten El-Dehaibi, Ruben Zamora, Josiah Radder, ..., Yingze Zhang, Georgios D. Kitsios, Yoram Vodovotz

vodovotzy@upmc.edu

**Highlights**

Dysregulated inflammation contributes adversely to critical illness outcomes

Common SNPs may associate with altered inflammation in critical illness

>500,000 SNPs were assessed algorithmically in critically ill trauma and non-trauma patients

rs10404939 in *LYPD4* was significantly associated with inflammation and adverse outcomes

## Article

# A common single nucleotide polymorphism is associated with inflammation and critical illness outcomes

Fayten El-Dehaibi,[1,6] Ruben Zamora,[1,6] Josiah Radder,[2,6] Jinling Yin,[1,6] Ashti M. Shah,[3] Rami A. Namas,[1] Michelle Situ,[1] Yanwu Zhao,[2] William Bain,[2] Alison Morris,[2] Bryan J. McVerry,[2] Derek A. Barclay,[1] Timothy R. Billiar,[1,7] Yingze Zhang,[2,7] Georgios D. Kitsios,[2,7] and Yoram Vodovotz[1,4,5,7,8,*]

## SUMMARY

**Acute inflammation is heterogeneous in critical illness and predictive of outcome. We hypothesized that genetic variability in novel, yet common, gene variants contributes to this heterogeneity and could stratify patient outcomes. We searched algorithmically for significant differences in systemic inflammatory mediators associated with any of 551,839 SNPs in one derivation (n = 380 patients with blunt trauma) and two validation (n = 75 trauma and n = 537 non-trauma patients) cohorts. This analysis identified rs10404939 in the *LYPD4* gene. Trauma patients homozygous for the A allele (rs10404939AA; 27%) had different trajectories of systemic inflammation along with persistently elevated multiple organ dysfunction (MOD) indices vs. patients homozygous for the G allele (rs10404939GG; 26%). rs10404939AA homozygotes in the trauma validation cohort had elevated MOD indices, and non-trauma patients displayed more complex inflammatory networks and worse 90-day survival compared to rs10404939GG homozygotes. Thus, rs10404939 emerged as a common, broadly prognostic SNP in critical illness.**

## INTRODUCTION

Critically ill patients admitted to the intensive care unit (ICU) have heterogeneous clinical course and outcomes despite seemingly similar clinical diagnoses, demographics, and protocolized care regimens.[1] We and others have hypothesized that genetic variation may account for such clinical heterogeneity,[2] having identified distinct dynamic networks of systemic inflammation in critically ill patients following severe traumatic injury associated of genetic polymorphisms.[3–6] Such associations between single nucleotide polymorphisms (SNPs), trajectories of inflammatory response, and clinical outcome suggest that genetic variation may play a role in determining the host response to acute insults, and ultimately predicting the outcome of critical illness.

Studies investigating gene polymorphisms associated with critical illness have centered primarily around candidate genes,[7] with broad-scope genome-wide association studies (GWASs) being relatively scarce. Several studies have focused on identifying novel gene variants linked to specific aspects of organ dysfunction in critical illness, such as acute respiratory distress syndrome (ARDS)[8,9] and acute kidney injury.[10] In the context of the COVID-19 pandemic, multiple GWAS were conducted to elucidate the associations of gene variants with severe COVID-19, with a particular focus on SNPs associated with the inflammatory response.[11–13] Furthermore, investigations have honed in on individual candidate genes and their corresponding protein products, aiming to uncover their relevance in the context of GWAS of critical illness. For instance, the *fas* gene has been studied in that regard.[14] Despite such efforts, more extensive genetic association studies are warranted to provide a deeper understanding of the genetic basis of host response during critical illness.

Here, we reasoned that while many genetically defined diseases involve the genetic variation of low prevalence and high impact on the phenotype,[15–22] in the context of traumatic injury we might identify a small number of SNPs present at high frequency in critically ill patients with dysregulated systemic inflammation. We further hypothesized that such genetic variants may also predispose to differential host responses and outcomes in non-traumatic acute insults, such as sepsis or acute lung injury among medical ICU patients. Therefore, we conducted a discovery analysis following the filtering of homozygous genotypes present at the same frequency in a cohort of 380 trauma patients,
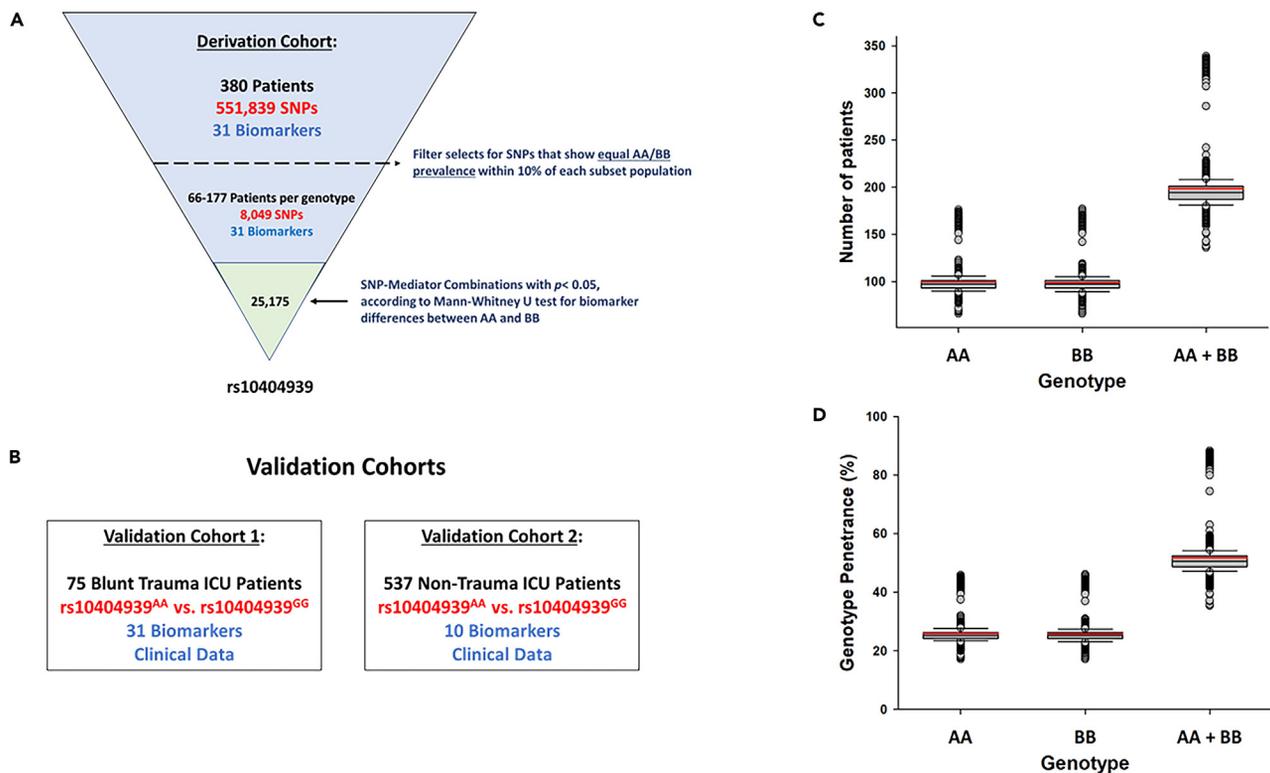
**Figure 1. *SNPScanner* generation and patient segregation**

(A) Data (SNPs, inflammation biomarkers) from a derivation cohort of 380 critically ill blunt trauma patients were assessed by the *SNPScanner* algorithm to find common SNPs associated with statistically significant (p < 0.05) differences in systemic inflammatory mediators between AA and BB genotypes; only one SNP (rs10404939) met the ultimate selection criteria and was associated with adverse clinical outcomes (see text).

(B) The association of rs10404939 genotype with adverse clinical outcomes and distinct systemic inflammatory mediator profiles was validated in an independent, 75-patient cohort of critically ill blunt trauma patients as well as in a 537-patient cohort of non-trauma critically ill patients.

(C and D) (C) In the derivation cohort of blunt trauma patients, the *SNPScanner* algorithm defined patient subgroups that ranged from 66 to 177 per genotype, and thus the prevalence (D) of SNPs detected by *SNPScanner* ranged from 17.2 to 45.1% in the derivation trauma patient cohort.

which revealed SNP rs10404939 as a prevalent predictor of systemic inflammatory responses and clinical outcomes. These findings were validated in two independent cohorts of trauma patients and medical ICU non-trauma patients.

## RESULTS

### Initial delineation of common, single nucleotide polymorphism-based trauma patient sub-cohorts with distinct profiles of circulating biomarkers

Our hypothesis centers around the existence of a limited number of SNPs that possess two key characteristics: being common (i.e., present at high frequencies) within the population and serving as reliable predictors for a specific phenotype of dysregulated inflammatory responses in critically ill patients. Operationally, we defined these as SNPs present with equal distributions (within 10%) of homozygous (nominally AA) vs. alternative homozygous (nominally BB) genotypes associated with systemic inflammatory responses, clinical severity indices (such as the severity/duration of Multiple Organ Dysfunction Syndrome (MODS)), and adverse clinical outcomes, such as length of stay (LOS) in the ICU, or requirement for mechanical ventilation. To search for SNPs meeting our criteria, we generated an algorithm (*SNPScanner*; Data S1; Figure 1A), which first scans for SNPs that show equal AA/BB prevalence within 10% of each subset population. In the next step, all the time courses of systemic inflammatory biomarkers in each SNP-based trauma patient subgroup were analyzed for statistical significance (p < 0.05) using the Mann-Whitney U Test.

We first deployed the *SNPScanner* algorithm on data from a cohort of critically ill blunt trauma patients since the time of onset of the acute insult is better defined as compared to critical illness from other causes.[23] In this blunt trauma derivation cohort (n = 380; Figure 1A), algorithmic analysis of a Human Core Exome-24 v1.1 BeadChip interrogating 551,839 SNPs resulted in groups of patients that ranged from 66 to 177 (Figure 1C) (i.e., SNPs present in 17–45% of the 380 derivation cohort patients) [Figure 1D]). Table 1 summarizes the numbers of SNPs with 0–14 statistically significant (p < 0.05 by Mann-Whitney U-test) trajectories of systemic inflammatory mediators when comparing each SNP-based trauma patient subgroup. At this level of statistical stringency, over 6,000 SNPs were clustered in groups containing the

**Table 1. Results of *SNPScanner* analysis**

| Frequency of Significant Cytokines (p < 0.05) | Number of SNPs | Percentage of Candidate SNPs | Percentage of Total SNPs |
|---|---|---|---|
| 0 | 567 | 7.0% | 0.1027% |
| 1 | 1348 | 16.7% | 0.2443% |
| 2 | 1677 | 20.8% | 0.3039% |
| 3 | 1519 | 18.9% | 0.2753% |
| 4 | 1091 | 13.6% | 0.1977% |
| 5 | 780 | 9.7% | 0.1413% |
| 6 | 449 | 5.6% | 0.0814% |
| 7 | 284 | 3.5% | 0.0515% |
| 8 | 163 | 2.0% | 0.0295% |
| 9 | 94 | 1.1% | 0.0170% |
| 10 | 45 | 0.6% | 0.0082% |
| 11 | 16 | 0.2% | 0.0029% |
| 12 | 15 | 0.2% | 0.0027% |
| 14 | 1 | 0.01% | 0.0002% |

SNPs with equal (±10%) distribution in the study cohort of 380 blunt trauma survivors were ranked by number of systemic inflammatory biomarkers that exhibited statistically significant differences (p < 0.05) in biomarker levels between the AA and BB genotypes.

following numbers of significantly altered inflammation biomarkers: 0 (567 SNPs); 1–4 (~4500 SNPs); 5–8 (~1,700 SNPs); 9–12 (~170 SNPs). Only one SNP group contained 14 significantly altered inflammation biomarkers; the SNPs in this final group corresponded to ~0.01–21% of SNPs meeting algorithmic criteria and ~0.0002–0.30% of total SNPs in the Human Core Exome-24 v1.1 BeadChip (Table 1).

As a control, we generated multiple, random groups of 15 SNPs from the group of 567 SNPs exhibiting 0 significantly different inflammation biomarkers by the initial Mann-Whitney U-test in the trauma derivation cohort (Table 1 and Data S2). We then carried out a permutation analysis for the presence of significant clinical outcome differences in total hospital length of stay (LOS), ICU LOS, days on mechanical ventilation, and MODS between the AA and BB genotype subgroup for a given SNP (Data S2), hypothesizing that the majority of the 567 SNPs would be associated with zero significant differences as a function of SNP genotype. In support of this hypothesis, 345 (61%) SNP-based trauma patient subgroups exhibited zero significant clinical outcomes, and a further 169 (30%) SNP-based trauma patient subgroups exhibited only one significant clinical outcome (Data S2). In essence, 91% of the SNPs predicted to have no association with clinical outcome also lacked any association with altered systemic inflammation as a function of SNP genotype. Of the remaining 9% of SNPs in the group of 567, 32 (6%) SNP-based groups exhibited two significant clinical outcomes, 13 (2%) SNP-based groups exhibited three significant clinical outcomes, and four (1%) SNP-based groups exhibited four significant clinical outcomes (Data S2).

We examined the results of this permutation test further in a more stringent analysis of 15 SNPs selected randomly from the 567 SNP-based patient subgroup with 0 significant differences in systemic inflammatory biomarker trajectories, wherein we tested for significance in clinical outcomes by two-Way ANOVA; of these, only 6 SNPs were associated with both significantly different systemic inflammatory biomarker levels and clinical outcomes (Figure 2B).

## Stringent statistical analysis of derivation cohort SNPScanner output identifies a single SNP (rs10404939)

The data from 15 SNPs initially identified algorithmically as having the greatest differences in inflammatory biomarker levels were next analyzed in a more stringent fashion for SNP-defined patient subgroup/time interactions using two-Way ANOVA, with statistical significance set a p < 0.05. rs2204925, the top-scoring SNP based on the initial *SNPScanner* analysis, was associated with 14 significantly different biomarkers (out of 31 total biomarkers analyzed; p < 0.05, initial analysis using the Mann–Whitney U [MWU] test). Subsequent testing with rs2204925 by two-Way ANOVA revealed 5 significantly different biomarkers (IL-15, IL-17A, IP-10, IL-23 and $NO_2^-/NO_3^-$, p < 0.05) but did not show significant differences in ICU LOS, total hospital LOS, number of days on mechanical ventilation, and average Marshall MODS score over the first 7 days in the ICU.

However, rs10404939, in the second-highest category of SNPs identified by the *SNPScanner* algorithm and initially associated with 12 significantly different biomarkers, was also significantly associated with 22 biomarker levels out of 31 assessed by two-Way ANOVA (Figure 2A; Figure S1). The two other leading SNPs identified algorithmically (rs7915065 and rs589804) were associated with 9/31 and 8/31 biomarkers by two-Way ANOVA. As a control, 15 SNPs selected randomly from the SNP-based patient subgroup with 0 significant differences in systemic inflammatory biomarker trajectories were also tested by two-Way ANOVA; of these, only 6 SNPs were associated with both significantly different systemic inflammatory biomarker levels and clinical outcomes (Figure 2B).

Taken together, these analyses pointed to rs10404939 as the leading candidate for an SNP present at high frequency in critically ill patients (Figure 2A). Based on the above results and following the validation of the Human Core Exome-24 v1.1 BeadChip results by DNA sequencing,
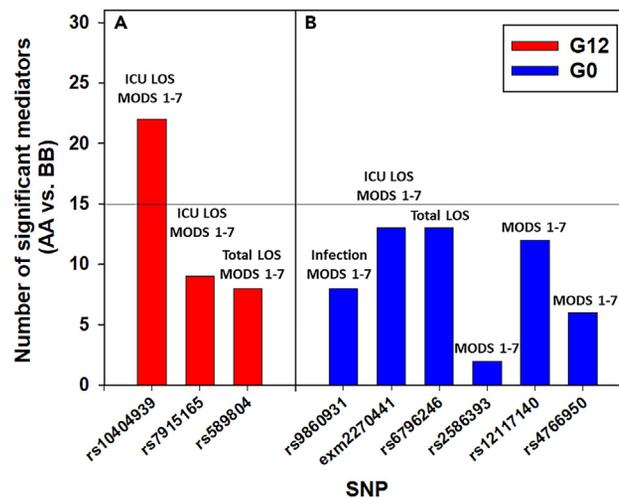
**Figure 2. Number of significant mediators by two-Way ANOVA (AA vs. BB) in two groups of SNPs with 12 (G12) or none (G0) significant mediators**
The 15 SNPs with the greatest differences in differentially distributed biomarkers inflammatory mediator counts by Mann-Whitney were analyzed in a more stringent fashion for subgroup/time interactions using two-Way ANOVA (p < 0.05). Subsequently, SNPs were divided into two groups based on the number of significant mediators (12 [G12, Panel A] or none [G0, Panel B]) in AA vs. BB patients and a threshold of 50% (15/30) of Two-Way ANOVA-defined differences in circulating inflammatory mediators was designated as the cutoff for defining an SNP as being associated with significant impact on systemic inflammation as described in Results. Figure shows the total number of significant inflammatory mediators (AA vs. BB) by two-Way ANOVA (p < 0.05).

all subsequent analyses were focused on this SNP, an intronic, G/A/C single nucleotide variant in the *LYPD4* gene on chromosome 19. LYPD4 is a sperm acrosomal membrane protein associated with sperm motility[24,25] with no known association with inflammatory or immune processes. Since the C allele was not observed in our cohorts, we refer to rs10404939 alleles as G/A.

## rs10404939 genotypes predict clinical outcomes in the derivation cohort of blunt trauma patients

We examined for clinical characteristic and outcome differences by rs10404939 genotypes according to a dominant (A/A + A/G vs. G/G), recessive (A/A vs. A/G + G/G), or additive (A/A vs. G/A vs. G/G) genetic model for the A allele.[26,27] Trauma patients in the derivation cohort stratified by rs10404939 genotypes (26.6% rs10404939[AA], 27.1% rs10404939[GG] and 46.1% rs10404939[GA]) had similar distribution of demographic characteristics , including age (A/A: 49.2 ± 20.3 years, G/G: 50 ± 18.8 years, and G/A: 49 ± 18.7 years) and sex (A/A: 29 females/ 72 males, G/G: 37 females/66 males, and G/A: 53 females/122 males), as well as similar injury severity, assessed by the composite Injury Severity Score (Figure 3A) or the components of the Abbreviated Injury Scale (AIS) (Figure 3B). Despite these similar demographic and injury characteristics, rs10404939[AA] homozygotes experienced significantly elevated MODS trajectories (Figure 3C; p < 0.001) compared to rs10404939[GG] and rs10404939[GA] patients. Notably, rs10404939[GG] and rs10404939[GA] patients had similar MODS trajectories, suggestive of a recessive model for rs10404939 (p < 0.001; Figure 3D). When the data were re-analyzed based on such a recessive model (A/A vs. G/A + G/G), the mean age, ISS, and AIS components of the comparison groups remained similar (p > 0.05, data not shown). The higher illness severity in rs10404939[AA] homozygotes was also supported by numerically (albeit not statistically significant) higher ICU length of stay (p = 0.075; Figure 3E) and duration of mechanical ventilation (p = 0.071; Figure 3F) compared to rs10404939[GG] and rs10404939[GA] patients combined. As described previously, death was a rare outcome (4.4%) in the parent 493-patient cohort from which the 380-patient derivation cohort of trauma patients (all survivors) for the present study was derived,[3] and therefore rs10404939 genotype-specific survival curves were not assessed.

## Validation of distinct clinical outcomes as a function of rs10404939 genotype in a separate cohort of blunt trauma patients

We next sought to validate and generalize the findings from the derivation cohort using an independent, more contemporary, cohort of 75 blunt trauma patients. The distribution of genotypes in the validation cohort (24.0% rs10404939[AA], 41.3% rs10404939[GG] and 34.7% rs10404939[GA]) was significantly different compared to the derivation cohort with higher frequency of the rs10404939[GG] in the validation cohort (Chi-square p = 0.04). Patients in the validation cohort were significantly younger and more seriously injured (higher ISS) as compared to the derivation cohort both overall (p < 0.001; see STAR Methods for details) and also when stratified by rs10404939 genotypes (Figure S2). Compared to the derivation cohort, rs10404939 genotype-based subgroups in the trauma validation cohort had similar injury characteristics, though there was a non-significant trend toward differences in the head injury component of the ISS as function of rs10404939 genotype (Figures 4A and 4B). The significant differences in injury severity compared to the derivation cohort were reflected in the generally longer hospital LOS (23.2 ± 21.3 days), ICU LOS (9.4 ± 12.9 days), and MODS trajectories in the validation vs. derivation cohort (Figure 4C). Despite significant differences in genotype distribution, age, and injury severity compared to the derivation cohort, trauma patients in the validation
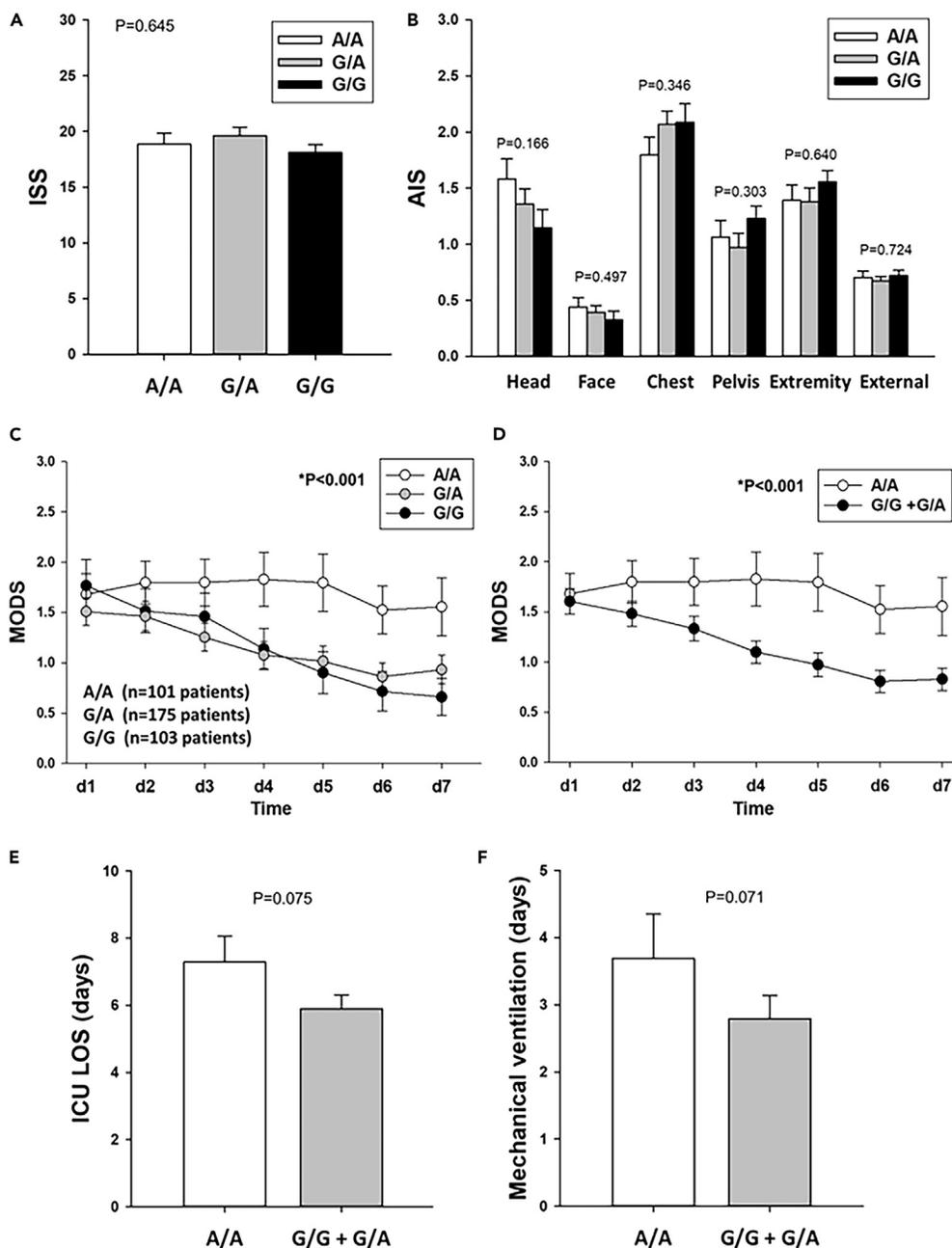
**Figure 3. Comparison of clinical outcomes in patients with rs10404939 genotype in the derivation cohort**

(A–F) Injury characteristics and clinical outcomes of derivation cohort trauma patients in the rs10404939[AA] (n = 101 patients), rs10404939[GA] (n = 175 patients), and rs10404939[GG] (n = 103 patients) subgroups were compared using Kruskal-Wallis ANOVA on Ranks (ISS, AIS), Mann-Whitney Rank-Sum test (ICU LOS, Mechanical Ventilation) and two-Way ANOVA (MODS) as described in *Materials and Methods*. Results represent mean ± SEM. No statistically significant differences were found in the composite ISS (A) and the components of the AIS (B) between the three patient subgroups. Trauma patients in the rs10404939[AA] subgroup experienced significantly elevated MODS as compared to rs10404939[GA] and rs10404939[GG] patients, both when comparing each genotype separately (C) and when rs10404939[GA] and rs10404939[GG] patients were combined (D). In addition, the rs10404939[AA] subgroup had longer stay in the ICU as compared to rs10404939[GA] and rs10404939[GG] patients, both when comparing each genotype separately (E) and when rs10404939[GA] and rs10404939[GG] patients were combined (F).

cohort with the rs10404939[AA] genotype also exhibited worse MODS trajectories as compared to patients with either the rs10404939[GG] or rs10404939[GA] genotype (Figure 4C), in line with the derivation cohort results. These results further supported the hypothesis that this genotype follows a recessive model in trauma patients (p < 0.001; Figure 4D). There was no significant difference in total hospital LOS, ICU LOS, and duration of mechanical ventilation by genotype (Figure S3).
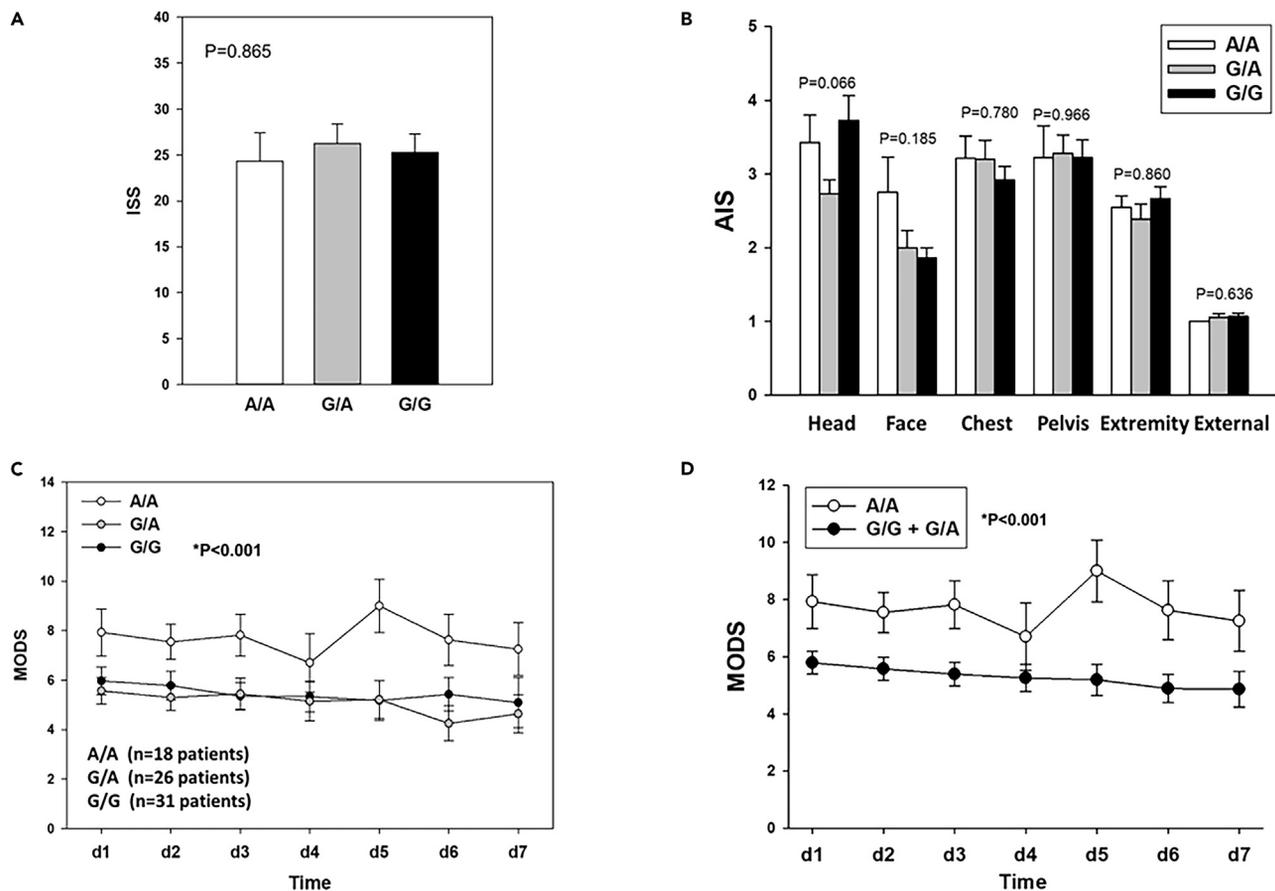
**Figure 4. Comparison of clinical outcomes in patients with rs10404939 genotype in the validation cohort**

(A–D) Injury characteristics and clinical outcomes of validation cohort trauma patients in the rs10404939[AA] (n = 18 patients), rs10404939[GA] (n = 26 patients), and rs10404939[GG] (n = 31 patients) were compared using One-Way ANOVA (ISS, Panel A), Kruskal-Wallis ANOVA on Ranks (AIS, Panel B), and two-Way ANOVA (MODS, Panels C and D) as described in *Materials and Methods*. Results represent mean ± SEM. Trauma patients with the rs10404939[AA] genotype exhibited worse MODS trajectories (C) as compared to rs10404939[GG] and rs10404939[GA] patients despite having similar injury characteristics (ISS [Panel a] and AIS [Panel B]). Similarly, trauma patients with the rs10404939[AA] genotype exhibited worse MODS trajectories as compared to rs10404939[GG] and rs10404939[GA] patients combined (D).

### Distinct dynamic networks of inflammation biomarkers as a function of rs10404939 genotype in the derivation trauma cohort

Since the initial goal of this study was to discover common SNPs associated with dysregulated systemic inflammatory programs, we next examined the temporal evolution of inflammation networks in the rs10404939[AA] (Figure 5A), rs10404939[GA] (Figure 5B), and rs10404939[GG] (Figure 5C) genotypes in the derivation trauma patient cohort. To accomplish this, we used the DyNA computational algorithm[28] at a stringency of 0.95 (i.e., p = 0.05).[29–31] This analysis showed distinct dynamic network patterns associated with the different rs10404939 genotypes. These findings supported the hypothesis that rs10404939 is associated with broad alterations in inflammatory programs in the context of trauma-induced critical illness.

Quantification of dynamic network complexity (Figure 5D) suggested that the greatest degree of inflammatory orchestration occurred early following hospital admission in rs10404939[AA] patients. In contrast, the other two rs10404939 genotypes exhibited much lower complexity, suggesting a more subdued systemic inflammatory response in these patients. Notably, the lowest inflammatory network complexity was observed in rs10404939[GA] patients (Figure 5D). Furthermore, we found differences in the connectivity of individual inflammatory biomarkers in rs10404939[AA] vs. rs10404939[GG] patient subgroups (Figure 5E), suggesting an impact of rs10404939 on a broad array of pathways. Specifically, relative to rs10404939[GG], rs10404939[AA] patients exhibited an inflammatory program involving elevated chemokines (eotaxin, MIP-1β, MIG/CXCL9), Th1 (IFN-γ), Th2 (IL-4, IL-13), Treg (IL-2, sIL-2Rα), and NK cell (IL-15) pathways. Network analysis also suggested that rs10404939[AA] patients have downregulated Th17 (IL-17A, IL-23) coupled to elevated pathways that serve to down-regulate IL-17A and that are tissue-protective, such as IL-17E/IL-25, IL-22, and IL-33. Due to the smaller size of the trauma validation cohort and the potential for spurious associations, DyNA was not carried out on data from this cohort.
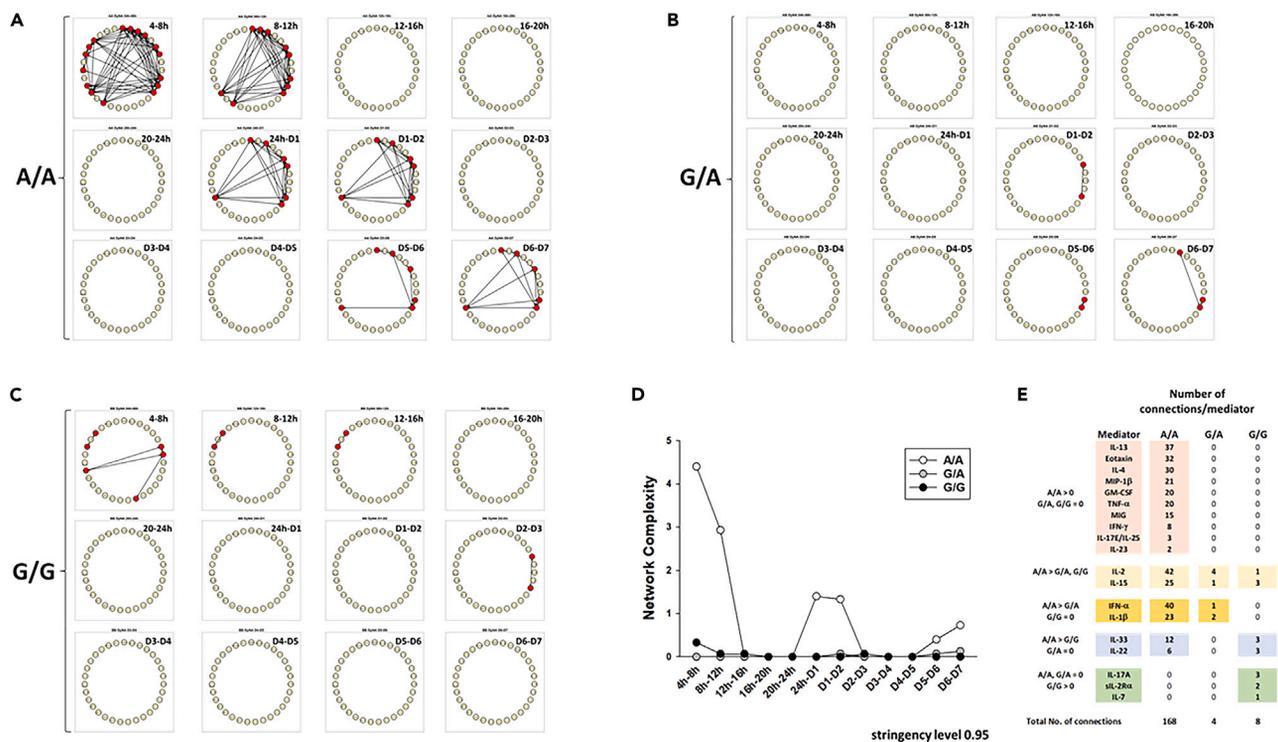
**Figure 5. Dynamic Network Analysis (DyNA) of inflammatory biomarkers in patients with rs10404939 genotype in the derivation cohort**

(A–E) Circulating inflammatory biomarkers in samples from the trauma patient derivation cohort were measured using Luminex, and DyNA (stringency level 0.95) was performed during different time-intervals as described in *Materials and Methods*. Panels A-C show the individual inflammatory networks in rs10404939[AA], rs10404939[GA] and rs10404939[GG] patients, respectively. Closed red circles represent biomarkers with at least one connection to another biomarker, while open yellow circles represent biomarkers that had no connections to other biomarkers. Panel D shows the network complexity of each network shown in Panels A-C for rs10404939[AA], rs10404939[GA] and rs10404939[GG] patients, respectively, calculated as described in *Materials and Methods*. Panel E shows a comparison of the total number of connections/biomarker for the three patient genotypes grouped as indicated.

## rs10404939 genotypes predict clinical outcomes in a large, inclusive cohort of critically ill, uninjured patients

We next sought to determine whether rs10404939 could predict outcomes in critical illness unrelated to trauma. Accordingly, we genotyped rs10404939 in 537 critically ill patients from the Pittsburgh Acute Lung Injury Registry and Biospecimen Repository (ALIR)[32–34] cohort with acute respiratory failure from different etiologies (including acute respiratory distress syndrome [ARDS], pneumonia, extra-pulmonary sepsis, congestive heart failure or encephalopathy requiring airway protection) and no major traumatic injury (Table 2). The distribution of genotypes (28.1% rs10404939[AA], 23.9% rs10404939[GG], and 48.0% rs10404939[GA]) in ALIR was not significantly different from the trauma derivation cohort (n = 380, Chi-square p = 0.52) but differed from the trauma validation cohort (n = 75, Chi-square p = 0.005).

We found no significant differences in age, sex, body-mass index (BMI), history of chronic obstructive pulmonary disease (COPD) or diabetes by genotypes in ALIR subjects (Table 2), but detected significant differences by race, with higher prevalence of the rs10404939[AA] genotype in whites compared to blacks (p < 0.01). We found no differences in etiology of acute respiratory failure (e.g., incidence of ARDS), type/burden of (non-traumatic) lung injury risk factors (direct vs. indirect; lung injury prediction scores), or baseline organ dysfunction severity by sequential organ failure assessment (SOFA) scores. However, rs10404939[AA] homozygotes and rs10404939[GA] heterozygotes had worse 90-day survival by Kaplan-Meier curves compared to rs10404939[GG] homozygotes (Figure 6A), suggestive of a dominant model for the A allele in non-trauma patients. The dominant model was independently prognostic of survival in a Cox-proportional hazards model adjusted for age, sex, and race (adjusted HR [95% confidence interval]: 1.40 [1.04–1.88], p = 0.02). Similarly, A allele carrier homozygotes had longer time to liberation from mechanical ventilation compared to rs10404939[GG] homozygotes (Figure 6B), an effect that was also independently prognostic in a Cox-proportional hazards model adjusted for age, sex, and race (adjusted HR 0.84 [0.71–0.99], p = 0.04). Thus, in contrast with our findings in the trauma cohorts, rs10404939 appeared to follow a dominant model in non-trauma patients for the hard clinical end-points of survival and ventilatory liberation.

Despite the significantly different distribution of rs10404939 genotypes by race, we found similar prognostic effects for the dominant model of rs10404939 on survival and ventilatory liberation in sensitivity analyses of blacks and whites examined separately (Figure S4). Furthermore, we found no differences in prognostication by rs10404939 in sensitivity analyses stratified by direct (e.g., pneumonia, aspiration, or inhalation injury) vs. indirect (e.g., extra-pulmonary sepsis, pancreatitis, or massive transfusion) lung injury risk factors. These findings suggest a generalizable effect of rs10404939 on critical illness outcome, independent of race and etiology of acute respiratory failure.

**Table 2. Clinical characteristics of three patient cohorts (trauma-derivation, trauma-validation and ALIR-non-trauma) used in the present study**

| Derivation trauma cohort | rs10404939[AA] | rs10404939[GA] | rs10404939[GG] | p |
|---|---|---|---|---|
| n[a] | 101 | 175 | 103 | |
| Age (yrs.), mean (SD) | 49.2 (20.5) | 49.0 (18.7) | 50.0 (18.8) | 0.911 |
| Male, n (%) | 72 (71.3) | 122 (69.7) | 66 (64.1) | 0.492 |
| White, n (%) | 100 (99.01) | 169 (96.6) | 98 (95.15) | 0.279 |
| Black, n (%) | 0 (0.0) | 6 (3.4) | 4 (3.88) | 0.151 |
| ISS, avg (SD) | 18.85 (10.0) | 19.6 (10.3) | 18.3 (10.0) | 0.645 |
| COPD, n (%) | 5 (4.95) | 10 (5.7) | 2 (1.9) | 0.329 |
| Diabetes Mellitus, n (%) | 11 (10.9) | 20 (11.4) | 15 (14.6) | 0.671 |
| Validation trauma cohort | rs10404939[AA] | rs10404939[GA] | rs10404939[GG] | p |
| n | 18 | 26 | 31 | |
| Age (yrs.), mean (SD) | 36.7 (0.5) | 33.2 (10.3) | 36.1 (10.7) | 0.465 |
| Male, n (%) | 11 (61.1) | 20 (76.9) | 21 (67.7) | 0.518 |
| White, n (%) | 13 (72.2) | 18 (69.2) | 14 (45.2) | 0.087 |
| Black, n (%) | 0 (0.0) | 4 (15.4) | 14 (45.2) | **0.002** |
| ISS, avg (SD) | 24.3 (13.1) | 26.2 (10.6) | 25.3 (11.25) | 0.865 |
| [b]COPD, n (%) | 6 (46.15) | 0 (0.0) | 4 (18.2) | **0.014** |
| [b]Diabetes Mellitus, n (%) | 1 (7.7) | 3 (23.1) | 3 (13.6) | 0.531 |
| Validation ALIR cohort | rs10404939[AA] | rs10404939[GA] | rs10404939[GG] | P |
| n | 151 | 258 | 128 | |
| Age, years (mean (SD)) | 56.6 (16.3) | 55.3 (15.1) | 56.5 (16.3) | 0.65 |
| Male, n (%) | 89 (59.3) | 140 (54.5) | 59 (46.1) | 0.08 |
| White, n (%) | 145 (96.7) | 244 (94.9) | 96 (75.0) | **<0.001** |
| Black, n (%) | 4 (2.7) | 11 (4.3) | 31 (24.2) | **<0.001** |
| COPD, n (%) | 33 (22.0) | 47 (18.3) | 36 (28.1) | 0.09 |
| Diabetes Mellitus, n (%) | 59 (39.3) | 81 (31.5) | 40 (31.2) | 0.22 |

[a]One patient with incomplete/missing data was not included.
[b]Represents available data from only 13 (**rs10404939[AA]** and **rs10404939[GA]**) and 22 patients (**rs10404939[GG]**).

We then examined for associations of rs10404939 genotypes with validated, prognostic plasma biomarkers of host response.[33,35] Due to the diverse etiologies of acute respiratory failure among ALIR participants, we first examined for systematic differences in biomarker levels by type of lung injury risk factors and found systematically higher levels of IL-8, sTNFR1, Ang2 and procalcitonin in patients with indirect vs. direct lung injury, whereas patients with no lung injury risk factors (intubated for airway protection or congestive heart failure) had lower levels for all 10 biomarkers examined (Figure S5). Therefore, we examined for rs10404939-biomarker associations in analysis stratified by direct vs. indirect lung injury. For subjects with direct lung injury (n = 229), rs10404939[GG] homozygotes had higher baseline levels of sST2 and Pentraxin-3 (p < 0.05) compared to A allele carriers (p < 0.05, Figure S6). Similarly, direct lung injury patients who survived in the ICU through the very late sampling interval (days 11–14), rs10404939[GG] homozygotes at that time point had higher levels of sST2, sRAGE, procalcitonin, and sTNFR1, indicating persistently higher inflammation in this group (Figure S6A). Conversely, for subjects with indirect lung injury (n = 121; Figure S6B) or no lung injury risk factor (n = 168; Figure S6C), there were no differences in serial plasma biomarker levels by rs10404939 genotypes.

The higher baseline levels of several inflammatory mediators in non-trauma rs10404939[GG] patients as compared to A allele carriers contrasted our findings in trauma patients. To further examine this difference, inflammation networks were assessed by DyNA in non-trauma patients stratified by rs10404939 genotype (Figures 6C–6E). This analysis demonstrated more complex networks at all time intervals in rs10404939[AA] patients as compared those of G allele carriers. Distinct from that of trauma patients, inflammatory complexity in non-trauma patients peaked at later time points. However, in agreement with findings in the trauma validation cohort, inflammatory network complexity was lowest in non-trauma rs10404939[GA] patients. Key mediators inferred to participate in the networks of non-trauma patients included sTNFR1, sRAGE, and IL-6.

### Linkage disequilibrium analysis suggests possible alternative, inflammation-related genes highly associated with rs10404939

Genetic variants identified in population-wide genetic studies are often not themselves causative but rather are in linkage disequilibrium (LD) with causal variants. To explore this possibility, we utilized LD data generated from a European population in the webtool *LDlink* from the
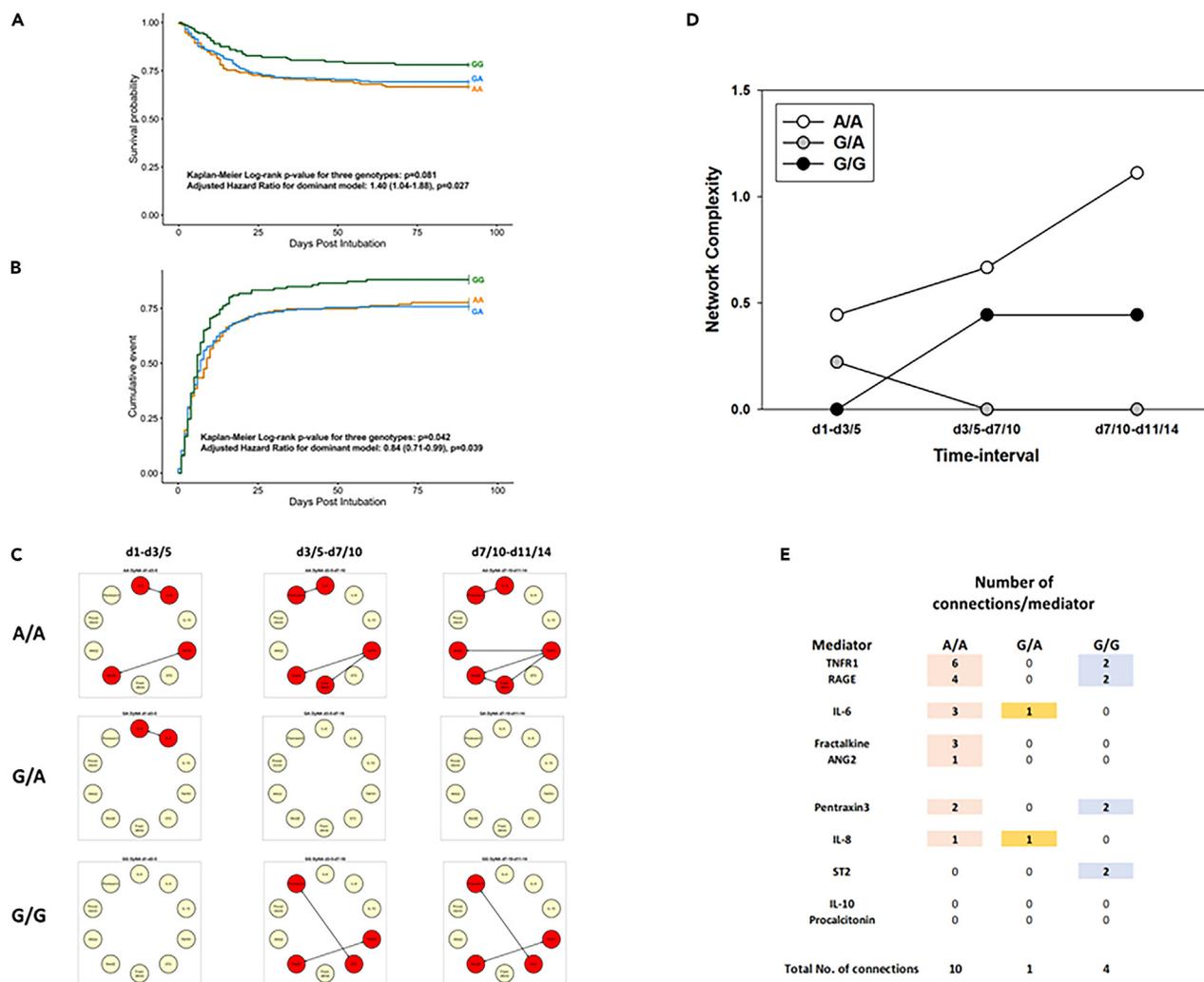
**Figure 6. Improved 90-day survival and liberation from mechanical ventilation for non-trauma, critically ill rs10404939GG patients and Dynamic Network Analysis (DyNA) of inflammatory biomarkers**

(A–E) Homozygotes of the G allele of the rs10404939 SNP had improved 90-day survival compared to heterozygotes or homozygotes of the A allele in Kaplan-Meier curve analyses (A). Assuming a dominant model for the A allele, A allele carriers had higher hazards of death and lower probability for liberation from mechanical ventilation compared to rs10404939[GG] homozygotes in Cox proportional hazards models adjusted for age, sex, and race (B). Circulating inflammatory biomarkers in samples from non-trauma, critically ill rs10404939 patients were measured using Luminex and DyNA (stringency level 0.95) was performed during different time-intervals as described in *Materials and Methods*. Panel C shows the individual inflammatory networks in rs10404939[AA], rs10404939[GA] and rs10404939[GG] patients, respectively. Closed red circles represent biomarkers with at least one connection to another biomarker, while open yellow circles represent biomarkers that had no connections to other biomarkers. Panel D shows the network complexity of each network shown in Panel A for rs10404939[AA], rs10404939[GA] and rs10404939[GG] patients, respectively, and calculated and described in *Materials and Methods*. Panel E shows a comparison of the total number of connections/biomarker for the three patient genotypes grouped as indicated.

National Cancer Institute. Using the *LDproxy* tool, we identified a large number of SNPs in high LD within a 220-kb region surrounding the rs10404939 on chromosome 19 (Figure 7A). A total of 408 unique variants with assigned reference SNP cluster ID (rs) and 11 without assigned rs ID were identified to be in LD with rs10404939. This analysis implicated multiple inflammation-related genes within this region including *CEACAM6, CEACAM3, DMRTC2, RPS19, CD79A, and RABAC1* that may be related to the biological pathway captured by rs10404939 (Figure 7A). Notably, none of the 408 variants other than rs10404939 (Table S1) were identified in the original *SNPScanner* discovery analysis depicted in Table 1.

## Genotype-specific expression of genes associated with rs10404939

To determine where rs10404939 is also associated with quantitative expression genes in different tissues, we analyzed genes identified as its expression quantitative trait loci (eQTL) and splicing quantitative trait loci (sQTL) using the GTEx portal. Based on the patient populations in
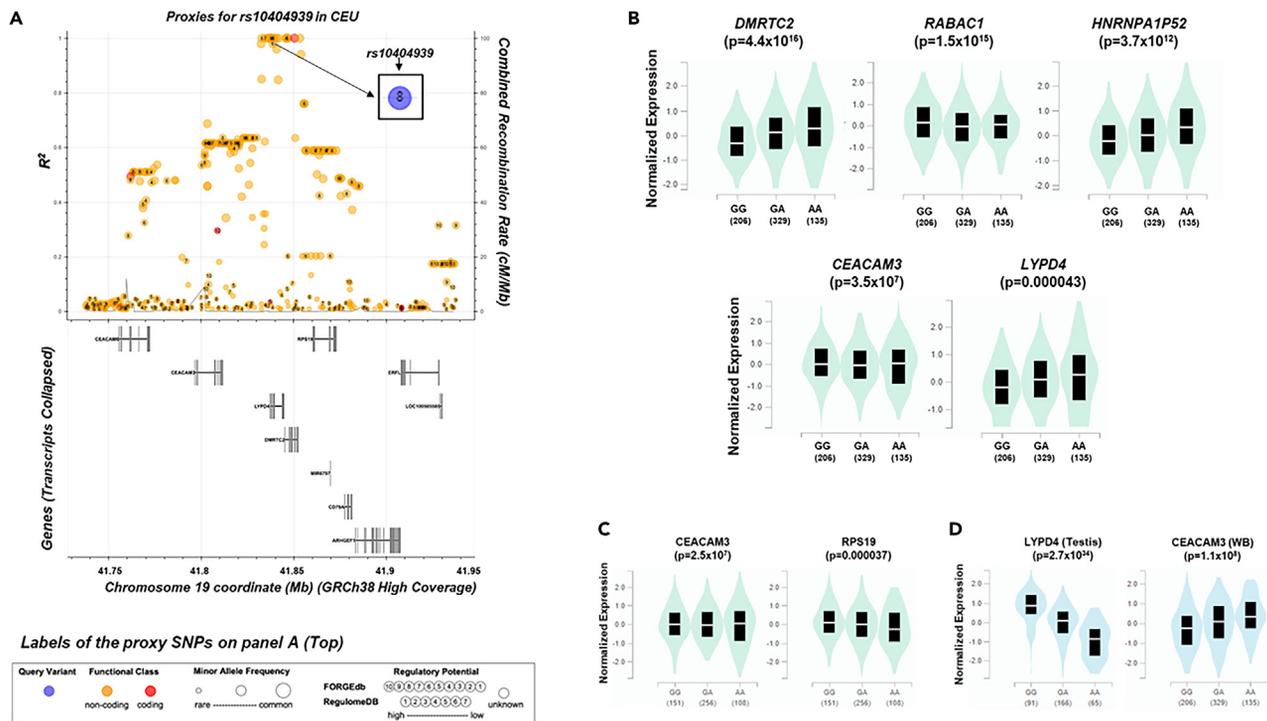
**Figure 7. Genes with SNPs in linkage disequilibrium with rs10404939 and quantitative tissue specific expression of genes with rs10404939**

(A) The proxy SNPs within an approximately 220 kb region surrounding rs10404939 are shown. The query was based on an American population with European ancestry (CEU) with a 100,000 bp scanning window using the LDproxy of the LDlink webtool (https://analysistools.cancer.gov/LDlink/).[36] The location of rs10404939 is shown in the inside panel and the SNP labeling is shown in the right bottom. Genomic coordinates based on GRCh38 high coverage genome on chromosome 19 are shown. The genes with high LD SNPs to rs10404939 are shown on the bottom of the panel (A).

(B and C) The gene structure of each gene is shown in its corresponding chromosome 19 coordinates. Tissue specific quantitative expression genes by the rs10404939 in blood (B) and lung (C) tissue.

(D) Tissue specific quantitative expression of specific splicing variants by the rs10404939 for the *LYPD4* in testis and *CEACAM3* in whole blood (D). Expression data were queried using the publicly available data from the Genotype-Tissue Expression (GTEx) Project portal (https://www.gtexportal.org/home/snp/rs10404939).

which we observed the prognostic effects of rs10404939 (trauma patients and non-trauma patients with acute respiratory failure), we focused our QTL analyses on whole blood (Figure 7B) and lung tissues (Figure 7C). In whole blood, *DMRTC2, HNRNPA1P52,* and *LYPD4* showed dose-dependent increased expression for the A allele, whereas rs10404939[GG] homozygotes had higher expression levels for the *RABAC1* gene. For the *CEACAM3* gene, higher expression was observed for the rs10404939[AA] homozygotes in both whole blood and lung tissues. Lower expression levels of the rs10404939[AA] homozygotes were expressed for *RPS19* gene in lung tissues. Splicing QTLs were observed in the testis for the *LYPD4* gene (exon 4 inclusion variants) and the *CEACAM3* (exon 4 skipping variants) gene in whole blood (Figures 7D; Figure S7). Overall, these results illustrate that the rs10404939 SNP may be in LD with genetic variation influencing the expression of different genes in several tissues, but the underlying genetic model was inconsistent by gene or tissue examined.

## DISCUSSION

The present study describes a novel genomic analysis workflow that identified a single SNP (rs10404939) that is both common and predictive of systemic inflammatory response and adverse clinical outcomes in critically ill patients. We first analyzed a derivation cohort of 380 blunt trauma patients that led to the identification of rs10404939, and then we validated associations in two separate cohorts. Remarkably, we found that despite differences in the insult and etiology of critical illness and MODS, rs10404939 genotypes were associated with illness severity and predicted clinical outcomes. Despite the discrepancies observed in the underlying genetic model regarding the prediction of outcome in our cohorts and differential gene expression levels in publicly available data, the rs10404939 SNP proved to be a robust and generalizable predictor of inflammatory responses and clinical outcomes in critical illness.

The SNP rs10404939 is an intronic variant in the *LYPD4* gene, and it is associated with genotype-specific expression in whole blood and the expression levels of specific splicing variants. The *LYPD4* gene encodes the protein LY6/PLAUR Domain-Containing 4 (LYPD4) that affects protein metabolism and posttranslational modification synthesis of glycosylphosphatidylinositol (GPI)-anchored proteins.[37] GPI-anchored proteins are lipid anchors for many cell-surface proteins, which are typically associated with membrane microdomains (rafts) enriched in

sphingolipids and cholesterol. GPI-anchored proteins often exist on the cell surface as transient homodimers and transduce proliferation or motility signals.[37] We hypothesize that LYPD4 may play a role in cell proliferation and repair of dysfunctional tissues in the context of multiple organ dysfunction. An alternative hypothesis is that, as a GPI-anchored protein associated with lipid rafts, LYPD4 is involved with lipid metabolism. We have shown recently that there is a substantial reprogramming of lipid, and especially sphingolipid, metabolism in the context of critical illness following blunt trauma.[38–40]

Another possibility is that rs10404939 is in linkage disequilibrium (LD) with a variant in a gene other than *LYPD4* that represents the causal association underlying the observed one reported here. Among the candidate genes identified as containing variants in LD with rs10404939, *CEACAM3* and *CEACAM6* are particularly intriguing. These genes encode the carcinoembryonic antigen-related cell adhesion molecule 3 and 6, respectively, which are proteins that play a major role in the opsonin-independent identification and phagocytosis of bacteria. In addition to this critical role in the innate immune system, these genes have been demonstrated to be rapidly evolving compared to many other genes as they interact with pathogen evolution directly, but it is reasonable to assume that these mutations have not evolved under the survival pressure of critical illness.[41]

These studies point to some important issues deserving of further study. The different MODS trajectories in the derivation and validation cohorts, while supporting the finding of adverse clinical outcomes in rs10404939[AA] vs. rs10404939[GG] patients, suggest that injury severity and/ or age may modify the impact of rs10404939. Intriguingly, the rs10404939 heterozygotes behaved differently between the trauma and acute lung injury patients although the underlying mechanism is unknown. It is possible that the mechanism of acute illness (traumatic vs. non-traumatic insult) as well as the timing of presentation in the ICU (close to acute insult for trauma patients and variable time course of non-traumatic illness in patients with acute respiratory failure) may influence the biological behavior of heterozygotes and the impact of differential levels of inflammatory mediators. In this regard, dynamic network analysis suggests that, although the broad inflammatory programs exhibit distinct dynamics between trauma and non-trauma critically ill patients, the inflammatory network complexity is notably higher in rs10404939[AA] vs. rs10404939[GG] patients. Interestingly, the finding of very sparse network connectivity in both trauma and non-trauma rs10404939[GA] patients may indicate a subphenotype of inappropriately low innate immune response in critical illness, which may lead to poor outcome.[42] The *in silico* analyses of eQTL and sQTL also suggested inconsistent genetic models based on tissue and gene examined, and also necessitate additional mechanistic studies to determine if *lypd4* or one or more of the other implicated genes is involved causally in the observed dysregulated inflammation and adverse clinical outcomes. Therefore, further study is necessary to define the mechanism and inheritance model underlying the prognostic effect of this SNP on critical illness outcomes.

We note that we attempted to address the potential for artifact and to test the null hypothesis that systemic inflammation does not relate to clinical outcomes by examining clinical outcomes in *SNPScanner*-identified, SNP-based patient subgroups that were predicted to have zero statistically significant differences in circulating inflammation biomarkers. While the majority of the 567 SNP-based patient subgroups indeed were associated with no clinical outcome differences as a function of SNP genotype, ~1% of SNP-based patient subgroups did have an association with either three or four of the four key clinical outcomes assessed. The majority of the SNPs were not associated with any known gene; the few genes that have been studied were in associated with cancers of various tissues (*lrrc4c*[43] and *linc01317*[44]), xenophagy of bacteria and viruses (*sacm1l*[45]), Ca$^{2+}$ channels on endothelial cells (*trpc4*[46]), as well as neuronal dysfunction (*lrrc4c*[47,48] and *slc20a2*[49]). Further study is needed to determine if and how these genes (or SNPs in linkage disequilibrium with these particular SNPs) might be associated with organ dysfunction.

## Conclusions

In conclusion, we report novel associations for a common SNP with systemic inflammatory responses and critical illness outcomes. Our findings transcend distinct clinical groups of critically ill patients and raise new hypotheses for the potential of a genetically preserved, generalized mechanism of host response to acute insults. Furthermore, we suggest that the algorithmic approach utilized to find common SNPs associated with adverse clinical phenotypes might be broadly applicable to similar, large-scale systematic screens in other contexts.

## Limitations of the study

There are multiple limitations to this study. There were notable differences in design, eligibility criteria, sampling times, and assayed biomarkers between the three cohorts included in our analysis. Our cohorts had limited representation of racial minorities and reported associations mostly reflect effects on white populations. Finally, our results are hypothesis-generating, and meticulous basic and translational work is needed to define the causal genetic variant(s) and the involved biological pathways by which rs10404939 is associated with critical illness.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

- METHOD DETAILS
  - Adherence to reporting guidelines
  - Trauma patient cohorts
  - Acute lung injury cohort
  - Genomic analyses
  - Assays of inflammation biomarkers
  - Algorithm for determining high-prevalence SNPs
  - Assessment of clinical differences in control SNPs
  - Statistical and computational analyses
  - Power calculations
  - Dynamic network analysis
  - Genetic association testing
  - Study approval
  - Role of funding source

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108333.

## AUTHOR CONTRIBUTIONS

FE (data analysis and interpretation, writing), RZ (data analysis and interpretation, writing), JR (data analysis and interpretation, writing), JY (genomic and inflammatory biomarker analysis), AMS (writing), RAN (trauma patient data collection), MS (trauma patient data collection), YZ-hao (non-trauma patient data collection), WB (non-trauma patient data collection), AM (non-trauma patient data collection oversight and funding, writing), BJM (data interpretation, writing), DAB (analysis of inflammatory biomarkers), TRB (trauma patient data collection oversight and funding, writing), YZhang (non-trauma patient genomic data collection and analysis, writing), GDK (data analysis and interpretation, writing, project leader), YV (data analysis and interpretation, writing, project leader).

## DECLARATION OF INTERESTS

The authors declare no competing interests associated with this article.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

## REFERENCES

1. Marra, A., Ely, E.W., Pandharipande, P.P., and Patel, M.B. (2017). The ABCDEF bundle in critical care. Crit. Care Clin. 33, 225–243. https://doi.org/10.1016/j.ccc.2016.12.005.

2. Hildebrand, F., Mommsen, P., Frink, M., van Griensven, M., and Krettek, C. (2011). Genetic predisposition for development of complications in multiple trauma patients. Shock 35, 440–448. https://doi.org/10.1097/SHK.0b013e31820e2152.

3. Schimunek, L., Namas, R.A., Yin, J., Liu, D., Barclay, D., El-Dehaibi, F., Abboud, A., Lindberg, H., Zamora, R., Billiar, T.R., and Vodovotz, Y. (2018). An enrichment strategy yields seven novel single nucleotide polymorphisms associated with mortality and altered TH17 responses following blunt trauma. Shock 49, 259–268. https://doi.org/10.1097/shk.0000000000000987.

4. Schimunek, L., Namas, R.A., Yin, J., Barclay, D., Liu, D., el-Dehaibi, F., Abboud, A., Cohen, M., Zamora, R., Billiar, T.R., and Vodovotz, Y. (2019). MPPED2 polymorphism is associated with altered systemic inflammation and adverse trauma outcomes. Front. Genet. 10, 1115. https://doi.org/10.3389/fgene.2019.01115.

5. Lamparello, A.J., Namas, R.A., Schimunek, L., Cohen, M., El-Dehaibi, F., Yin, J., Barclay, D., Zamora, R., Billiar, T.R., and Vodovotz, Y. (2020). An aging-related single-nucleotide polymorphism is associated with altered clinical outcomes and distinct inflammatory profiles in aged blunt trauma patients. Shock 53, 146–155. https://doi.org/10.1097/shk.0000000000001411.

6. Brown, D., Namas, R.A., Almahmoud, K., Zaaqoq, A., Sarkar, J., Barclay, D.A., Yin, J., Ghuma, A., Abboud, A., Constantine, G., et al. (2015). Trauma in silico: individual-specific mathematical models and virtual clinical populations. Sci. Transl. Med. 7, 285ra61. 285ra261.

7. Abraham, E. (2013). It's all in the genes: Moving toward precision medicine in critical illness. Crit. Care Med. 41, 1363–1364. https://doi.org/10.1097/Ccm.0b013e31827c02dd.

8. Christie, J.D., Wurfel, M.M., Feng, R., O'Keefe, G.E., Bradfield, J., Ware, L.B., Christiani, D.C., Calfee, C.S., Cohen, M.J., Matthay, M., et al. (2012). Genome wide association identifies PPFIA1 as a candidate gene for acute lung injury risk following major trauma. PLoS One 7, e28268. https://doi.org/10.1371/journal.pone.0028268.

9. Lynn, H., Sun, X., Casanova, N., Gonzales-Garay, M., Bime, C., and Garcia, J.G.N. (2019). Genomic and genetic approaches to deciphering acute respiratory distress syndrome risk and mortality. Antioxid. Redox Signal. 31, 1027–1052. https://doi.org/10.1089/ars.2018.7701.

10. Zhao, B., Lu, Q., Cheng, Y., Belcher, J.M., Siew, E.D., Leaf, D.E., Body, S.C., Fox, A.A., Waikar, S.S., Collard, C.D., et al. (2017). A Genome-wide association study to identify single-nucleotide polymorphisms for acute kidney injury. Am. J. Respir. Crit. Care Med. 195, 482–490. https://doi.org/10.1164/rccm.201603-0518OC.

11. Niemi, M.E.K., Karjalainen, J., Liao, R.G., Neale, B.M., Daly, M., Ganna, A., Pathak, G.A., Andrews, S.J., Kanai, M., Veerapen, K., et al. (2021). Mapping the human genetic architecture of COVID-19. Nature 600, 472–477. https://doi.org/10.1038/s41586-021-03767-x.

12. Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C.A., Walker, S., Russell, C.D., Malinauskas, T., Wu, Y., Millar, J., et al. (2022). Whole-genome sequencing reveals host factors underlying critical COVID-19. Nature 607, 97–103. https://doi.org/10.1038/s41586-022-04576-6.

13. Pairo-Castineira, E., Rawlik, K., Bretherick, A.D., Qi, T., Wu, Y., Nassiri, I., McConkey, G.A., Zechner, M., Klaric, L., Griffiths, F., et al. (2023). GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. Nature 617, 764–768. https://doi.org/10.1038/s41586-023-06034-3.

14. Mikacenic, C., Bhatraju, P., Robinson-Cohen, C., Kosamo, S., Fohner, A.E., Dmyterko, V., Long, S.A., Cerosaletti, K., Calfee, C.S., Matthay, M.A., et al. (2022). Single nucleotide variant in FAS associates with organ failure and soluble Fas cell surface death receptor in critical illness. Crit. Care Med. 50, e284–e293. https://doi.org/10.1097/ccm.0000000000005333.

15. Filippini, S.E., and Vega, A. (2013). Breast cancer genes: beyond BRCA1 and BRCA2. Front. Biosci. 18, 1358–1372. https://doi.org/10.2741/4185.

16. Zavala, V.A., Serrano-Gomez, S.J., Dutil, J., and Fejerman, L. (2019). Genetic epidemiology of breast cancer in Latin America. Genes 10, 153. https://doi.org/10.3390/genes10020153.

17. Mahdavi, M., Nassiri, M., Kooshyar, M.M., Vakili-Azghandi, M., Avan, A., Sandry, R., Pillai, S., Lam, A.K.Y., and Gopalan, V. (2019). Hereditary breast cancer; Genetic penetrance and current status with BRCA. J. Cell. Physiol. 234, 5741–5750. https://doi.org/10.1002/jcp.27464.

18. Read, J., Wadt, K.A.W., and Hayward, N.K. (2016). Melanoma genetics. J. Med. Genet. 53, 1–14. https://doi.org/10.1136/jmedgenet-2015-103150.

19. Rosenberg, A.Z., and Kopp, J.B. (2017). Focal Segmental Glomerulosclerosis. Clin. J. Am. Soc. Nephrol. 12, 502–517. https://doi.org/10.2215/cjn.05960616.

20. Nikolac Perkovic, M., and Pivac, N. (2019). Genetic markers of Alzheimer's disease. Adv. Exp. Med. Biol. 1192, 27–52. https://doi.org/10.1007/978-981-32-9721-0_3.

21. Brenner, D., and Weishaupt, J.H. (2019). Update on amyotrophic lateral sclerosis genetics. Curr. Opin. Neurol. 32, 735–739. https://doi.org/10.1097/wco.0000000000000737.

22. Blauwendraat, C., Nalls, M.A., and Singleton, A.B. (2020). The genetic architecture of Parkinson's disease. Lancet Neurol. 19, 170–178. https://doi.org/10.1016/s1474-4422(19)30287-x.

23. Lamparello, A.J., Namas, R.A., Constantine, G., McKinley, T.O., Elster, E., Vodovotz, Y., and Billiar, T.R. (2019). A conceptual time window-based model for the early stratification of trauma patients. J. Intern. Med. 286, 2–15. https://doi.org/10.1111/joim.12874.

24. Fujihara, Y., Noda, T., Kobayashi, K., Oji, A., Kobayashi, S., Matsumura, T., Larasati, T., Oura, S., Kojima-Kita, K., Yu, Z., et al. (2019). Identification of multiple male reproductive tract-specific proteins that regulate sperm migration through the oviduct in mice. Proc. Natl. Acad. Sci. USA 116, 18498–18506. https://doi.org/10.1073/pnas.1908736116.

25. Wang, D., Cheng, L., Xia, W., Liu, X., Guo, Y., Yang, X., Guo, X., and Xu, E.Y. (2020). LYPD4, mouse homolog of a human acrosome protein, is essential for sperm fertilizing ability and male fertility. Biol. Reprod. 102, 1033–1044. https://doi.org/10.1093/biolre/ioaa018.

26. Liu, H.M., Zheng, J.P., Yang, D., Liu, Z.F., Li, Z., Hu, Z.Z., and Li, Z.N. (2021). Recessive/dominant model: Alternative choice in case-control-based genome-wide association studies. PLoS One 16, e0254947. https://doi.org/10.1371/journal.pone.0254947.

27. Zhao, F., Song, M., Wang, Y., and Wang, W. (2016). Genetic model. J. Cell Mol. Med. 20, 765. https://doi.org/10.1111/jcmm.12751.

28. Mi, Q., Constantine, G., Ziraldo, C., Solovyev, A., Torres, A., Namas, R., Bentley, T., Billiar, T.R., Zamora, R., Puyana, J.C., and Vodovotz, Y. (2011). A dynamic view of trauma/hemorrhage-induced inflammation in mice: Principal drivers and networks. PLoS One 6, e19424.

29. Ziraldo, C., Vodovotz, Y., Namas, R.A., Almahmoud, K., Tapias, V., Mi, Q., Barclay, D., Jefferson, B.S., Chen, G., Billiar, T.R., and Zamora, R. (2013). Central role for MCP-1/CCL2 in injury-induced inflammation revealed by in vitro, in silico, and clinical studies. PLoS One 8, e79804. PONE-D-13-32473 [pii]. https://doi.org/10.1371/journal.pone.0079804.

30. Abboud, A., Namas, R.A., Ramadan, M., Mi, Q., Almahmoud, K., Abdul-Malak, O., Azhar, N., Zaaqoq, A., Namas, R., Barclay, D.A., et al. (2016). Computational analysis supports an early, type 17 cell-associated divergence of blunt trauma survival and mortality. Crit. Care Med. 44, e1074–e1081.

31. Zamora, R., Korff, S., Mi, Q., Barclay, D., Schimunek, L., Zucca, R., Arsiwalla, X.D., Simmons, R.L., Verschure, P., Billiar, T.R., and Vodovotz, Y. (2018). A computational analysis of dynamic, multi-organ inflammatory crosstalk induced by endotoxin in mice. PLoS Comput. Biol. 14, e1006582.

32. Kitsios, G.D., Kotok, D., Yang, H., Finkelman, M.A., Zhang, Y., Britton, N., Li, X., Levochkina, M.S., Wagner, A.K., Schaefer, C., et al. (2021). Plasma 1,3-β-d-glucan levels predict adverse clinical outcomes in critical illness. JCI Insight 6, e141277. https://doi.org/10.1172/jci.insight.141277.

33. Drohan, C.M., Nouraie, S.M., Bain, W., Shah, F.A., Evankovich, J., Zhang, Y., Morris, A., McVerry, B.J., and Kitsios, G.D. (2021). Biomarker-based classification of patients with acute respiratory failure into inflammatory subphenotypes: A single-center exploratory study. Crit. Care Explor. 3, e0518. https://doi.org/10.1097/cce.0000000000000518.

34. Kitsios, G.D., Yang, H., Yang, L., Qin, S., Fitch, A., Wang, X.H., Fair, K., Evankovich, J., Bain, W., Shah, F., et al. (2020). Respiratory tract dysbiosis is associated with worse outcomes in mechanically ventilated patients. Am. J. Respir. Crit. Care Med. 202, 1666–1677. https://doi.org/10.1164/rccm.201912-2441OC.

35. Kitsios, G.D., Yang, L., Manatakis, D.V., Nouraie, M., Evankovich, J., Bain, W., Dunlap, D.G., Shah, F., Barbash, I.J., Rapport, S.F., et al. (2019). Host-response subphenotypes offer prognostic enrichment in patients with or at risk for Acute Respiratory Distress Syndrome. Crit. Care Med. 47, 1724–1734. https://doi.org/10.1097/ccm.0000000000004018.

36. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. 12, 996–1006. https://doi.org/10.1101/gr.229102.

37. Kinoshita, T., and Fujita, M. (2016). Thematic Review Series: Glycosylphosphatidylinositol (GPI) anchors: Biochemistry and cell biology biosynthesis of GPI-anchored proteins: special emphasis on GPI lipid remodeling. J. Lipid Res. 57, 6–24. https://doi.org/10.1194/jlr.R063313.

38. Cyr, A., Zhong, Y., Reis, S.E., Namas, R.A., Amoscato, A., Zuckerbraun, B., Sperry, J., Zamora, R., Vodovotz, Y., and Billiar, T.R. (2021). Analysis of the plasma metabolome after trauma, novel circulating sphingolipid signatures, and in-hospital outcomes. J. Am. Coll. Surg. 232, 276–287.e1. https://doi.org/10.1016/j.jamcollsurg.2020.12.022.

39. Wu, J., Vodovotz, Y., Abdelhamid, S., Guyette, F.X., Yaffe, M.B., Gruen, D.S., Cyr, A., Okonkwo, D.O., Kar, U.K., Krishnamoorthi, N., et al. (2021). Multi-omic analysis in injured humans: Patterns align with outcomes and treatment responses. Cell Rep. Med. 2, 100478. https://doi.org/10.1016/j.xcrm.2021.100478.

40. Wu, J., Cyr, A., Gruen, D.S., Lovelace, T.C., Benos, P.V., Das, J., Kar, U.K., Chen, T., Guyette, F.X., Yazer, M.H., et al. (2022). Lipidomic signatures align with inflammatory patterns and outcomes in critical illness. Nat. Commun. 13, 6789. https://doi.org/10.1038/s41467-022-34420-4.

41. Adrian, J., Bonsignore, P., Hammer, S., Frickey, T., and Hauck, C.R. (2019). Adaptation to host-specific bacterial pathogens drives rapid evolution of a human innate immune receptor. Curr. Biol. 29, 616–630.e5. https://doi.org/10.1016/j.cub.2019.01.058.

42. Schimunek, L., Lindberg, H., Cohen, M., Namas, R.A., Mi, Q., Yin, J., Barclay, D., El-Dehaibi, F., Abboud, A., Zamora, R., et al. (2020). Computational derivation of core, dynamic human blunt trauma inflammatory endotypes. Front. Immunol. 11, 589304. https://doi.org/10.3389/fimmu.2020.589304.

43. Yang, X., Lei, P., Huang, L., Tang, X., Wei, B., and Wei, H. (2021). Prognostic value of

LRRC4C in colon and gastric cancers correlates with tumour microenvironment immunity. Int. J. Biol. Sci. 17, 1413–1427. https://doi.org/10.7150/ijbs.58876.

44. Chen, B., Dong, D., Yao, Q., Zou, Y., and Hu, W. (2021). A novel prognostic cancer-related lncRNA signature in papillary renal cell carcinoma. Cancer Cell Int. 21, 545. https://doi.org/10.1186/s12935-021-02247-6.

45. Carey, K.L., Liu, K., and Xavier, R.J. (2022). Innate host defense mechanisms SAC bacteria by regulating phosphoinositide kinases and phosphatases. Autophagy 18, 452–454. https://doi.org/10.1080/15548627.2021.2002102.

46. Plant, T.D., and Schaefer, M. (2003). TRPC4 and TRPC5: receptor-operated Ca2+-permeable nonselective cation channels. Cell Calcium 33, 441–450. https://doi.org/10.1016/s0143-4160(03)00055-1.

47. Maussion, G., Cruceanu, C., Rosenfeld, J.A., Bell, S.C., Jollant, F., Szatkiewicz, J., Collins, R.L., Hanscom, C., Kolobova, I., de Champfleur, N.M., et al. (2017). Implication of LRRC4C and DPP6 in neurodevelopmental disorders. Am. J. Med. Genet. 173, 395–406. https://doi.org/10.1002/ajmg.a.38021.

48. Choi, Y., Park, H., Jung, H., Kweon, H., Kim, S., Lee, S.Y., Han, H., Cho, Y., Kim, S., Sim, W.S., et al. (2019). NGL-1/LRRC4C deletion moderately suppresses hippocampal excitatory synapse development and function in an input-independent manner. Front. Mol. Neurosci. 12, 119. https://doi.org/10.3389/fnmol.2019.00119.

49. Lemos, R.R., Ramos, E.M., Legati, A., Nicolas, G., Jenkinson, E.M., Livingston, J.H., Crow, Y.J., Campion, D., Coppola, G., and Oliveira, J.R.M. (2015). Update and mutational analysis of SLC20A2: A major cause of primary familial brain calcification. Hum. Mutat. 36, 489–495. https://doi.org/10.1002/humu.22778.

50. Little, J., Higgins, J.P.T., Ioannidis, J.P.A., Moher, D., Gagnon, F., von Elm, E., Khoury, M.J., Cohen, B., Davey-Smith, G., Grimshaw, J., et al. (2009). STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med. 6, e22. https://doi.org/10.1371/journal.pmed.1000022.

51. BAKER, S.P., O'NEILL, B., Haddon, W., Jr., and LONG, W.B. (1974). The injury severity score: A method for describing patients with multiple injuries and evaluating emergency care. J. Trauma 14, 187–196.

52. Gennarelli, T.A., and Wodzin, E. (2006). AIS 2005: a contemporary injury scale. Injury 37, 1083–1091. https://doi.org/10.1016/j.injury.2006.07.009.

53. Marshall, J.C., Cook, D.J., Christou, N.V., Bernard, G.R., Sprung, C.L., and Sibbald, W.J. (1995). Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit. Care Med. 23, 1638–1652.

54. Zhang, Y., Tedrow, J., Nouraie, M., Li, X., Chandra, D., Bon, J., Kass, D.J., Fuhrman, C.R., Leader, J.K., Duncan, S.R., et al. (2021). Elevated plasma level of Pentraxin 3 is associated with emphysema and mortality in smokers. Thorax 76, 335–342. https://doi.org/10.1136/thoraxjnl-2020-215356.

55. Mann, H.B., and Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18, 50–60. https://doi.org/10.1214/aoms/1177730491.

56. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

57. Abboud, A., Namas, R.A., Ramadan, M., Mi, Q., Almahmoud, K., Abdul-Malak, O., Azhar, N., Zaaqoq, A., Namas, R., Barclay, D.A., et al. (2016). Computational Analysis Supports an Early, Type 17 Cell-Associated Divergence of Blunt Trauma Survival and Mortality. Crit. Care Med. 44, e1074–e1081. https://doi.org/10.1097/ccm.0000000000001951.

58. Zamora, R., Korff, S., Mi, Q., Barclay, D., Schimunek, L., Zucca, R., Arsiwalla, X.D., Simmons, R.L., Verschure, P., Billiar, T.R., and Vodovotz, Y. (2018). A computational analysis of dynamic, multi-organ inflammatory crosstalk induced by endotoxin in mice. PLoS Comput. Biol. 14, e1006582. https://doi.org/10.1371/journal.pcbi.1006582.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Critical commercial assays** | | |
| Human Cytokine/Chemokine MILLIPLEX™ Panel kit | Millipore Corporation, Billerica, MA | Cat# HCYTA-60K |
| Illumina® arrays | Illumina, San Diego, CA | WG-331-1102 |
| QIAamp® DNA Blood Kit | QIAGEN, Valencia, CA | Cat#51185 |
| **Deposited data** | | |
| Genotype-Tissue Expression (GTEx) Project | National Human Genome Research Institute | https://www.gtexportal.org/home/ |
| LDlink | National Cancer Institiute | https://analysistools.cancer.gov/LDlink/ |
| **Oligonucleotides** | | |
| forward primer sequence 5′-CCAGAGGCCCCTCACAAAC-3′ | This paper | N/A |
| reverse primer sequence 5′-GGCAGGTGAGAGGAGAGGAA-3′ | This paper | N/A |
| **Software and algorithms** | | |
| SNPScanner | This Paper | N/A |
| Python version 3.7 | Python Software Foundation | https://www.python.org/ |
| SciPy | https://scipy.org/ | https://doi.org/10.1038/s41592-019-0686-2 |
| Mann-Whitney U Test | Mann & Whitney, 1947[22] | https://doi.org/10.1214/aoms/1177730491 |
| Illumina GenomeStudio | Illumina | https://www.illumina.com/techniques/microarrays/array-dataanalysisexperimentaldesign/genomestudio.html |
| Dynamic Network Analysis (DyNA) | MATLAB Software, Ziraldo et al. (2013)[24] | https://doi.org/10.1371/journal.pone.0079804 |
| SigmaPlot | Systat Software | https://systatsoftware.com/sigmaplot/ |
| **Other** | | |
| Luminex™ 100 IS Analyzer | Luminex, Austin, TX | |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Yoram Vodovotz (vodovotzy@upmc.edu).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

All data reported in this paper will be shared by the lead contact upon reasonable request.

This SNP Scanner Pseudocode has been submitted together with the manuscript. The SNP Scanner code in full is available at https://github.com/fayteneldehaibi/SNPScanner.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon reasonable request.

## METHOD DETAILS

### Adherence to reporting guidelines

The STrengthening the REporting of Genetic Association Studies (STREGA)[50] reporting guidelines were utilized in this study.

### Trauma patient cohorts

Initial studies (Figure 1A) were carried out using a derivation cohort of 380 blunt trauma survivors enrolled by screening after presentation to the emergency department of the UPMC Presbyterian hospital (a Level 1 trauma center) between May 5, 2004, to May 2, 2012. A separate validation cohort of 75 blunt trauma patients (Figure 1B) was enrolled at the same institution from July 6, 2019, to July 19, 2021. Patients eligible for enrollment in the study were at least 18 years of age, admitted to the (ICU) after being resuscitated, and, per treating physician, were expected to live more than 24h. Exclusion criteria were isolated head injury, pregnancy, and penetrating trauma. Informed consent was obtained from each patient or next of kin as per the University of Pittsburgh Institutional Review Board (IRB; Protocol No. MOD08010232-19/PRO08010232) and in accordance with the Declaration of Helsinki. Blood samples for DNA and inflammatory biomarker analysis were obtained upon admission to the trauma bay, at first 24 h of hospitalization, and daily thereafter for 7 days.

Clinical data, including injury severity score (ISS), abbreviated injury scale (AIS) score, Marshall Multiple Organ Dysfunction (MOD) score, ICU LOS, hospital LOS, and days on mechanical ventilation were collected from the hospital inpatient electronic and trauma registry database. ISS[51] and AIS[52] were calculated for each patient by a single trauma surgeon after attending radiology evaluations were finalized. The ISS is based on an anatomical scoring system that provides an overall score for patients with multiple injuries. Each injury is assigned an AIS score, allocated to one of six body regions: head, face, chest, abdomen, extremities (including pelvis), and external. As an index of organ dysfunction, the MOD score[53] ranging from 0 to 24) was calculated. In brief, six variables were obtained from the electronic trauma data registry including (a) the respiratory system ($PO_2/FIO_2$ ratio); (b) the renal system (serum creatinine concentration); (c) the hepatic system (serum bilirubin concentration); (d) the hematologic system (platelet count); (e) the central nervous system (Glasgow Coma Scale); and (f) the cardiovascular system-the pressure-adjusted heart rate (PAR).

The overall characteristics of the patients in the derivation cohort (n = 380) were as follows (mean $\pm$ SD): (ISS) = 19.0 $\pm$ 10.2 [min: 1; max: 54]; mean age: 49.3 $\pm$ 19.2 years (min: 18y; max: 90y); 119 women, 261 men; 96.83% whites, 2.64% blacks and 0.53% others. The characteristics of the validation cohort (n = 75) were ISS: 25.4 $\pm$ 11.4 [min: 1; max: 54]; mean age: 35.2 $\pm$ 10.6 years (min: 19y; max: 55y); 23 women, 52 men; 60% whites, 24% blacks and 16% declined/not specified race.

### Acute lung injury cohort

From October 2011 – January 2018, we enrolled mechanically-ventilated patients from the Medical ICU at the University of Pittsburgh Medical Center to the Pittsburgh Acute Lung Injury Registry and Biospecimen Repository (ALIR) study[32–34] (Figure 1B). Eligible patients were 18 years or older with acute respiratory failure (of any cause) requiring mechanical ventilation via endotracheal intubation. Exclusion criteria were the inability to obtain informed consent, presence of tracheostomy, or mechanical ventilation for >72 h. The University of Pittsburgh Institutional Review Board approved the study (STUDY19050099) and written informed consent was provided by all participants or their surrogates in accordance with the Declaration of Helsinki.

Baseline clinical data (including demographics, comorbidities, physiologic and laboratory variables) were collected prospectively at the time of enrollment and baseline severity of illness was quantified with modified sequential organ failure assessment (SOFA) scores (i.e., excluding the neurologic component). We analyzed data for 537 patients (288 men, 249 women), 90.6% whites, 8.6% blacks, with mean age 56.4 $\pm$ 15.9 years, who were followed prospectively for 90-day survival and ventilator-free days at day 28 from enrollment (VFD). Blood samples for genomic and biomarker assay analyses were collected upon enrollment (within 48 h of intubation) and then at three follow-up time intervals (middle: days 3–6; late: days 7–10; very late: days 11–14) for patients who remained in the ICU.

### Genomic analyses

For the blunt trauma patient cohorts, DNA was prepared from whole blood samples and analyzed using Illumina® arrays as described previously.[3–5] Whole blood samples were collected into EDTA-treated tubes and DNA was extracted using the QIAamp® DNA Blood Midi Kit (QIAGEN, Valencia, CA) as per manufacturer's specifications. Single nucleotide polymorphism genotyping was performed with 200 ng of genomic DNA input using the Human Core Exome-24 v1.1 BeadChip (Illumina, San Diego, CA) following the manufacturer's Infinium® HTS Assay protocol. This array interrogated 551,839 SNPs, including both SNPs previously implicated in genome-wide association studies as well as additional coding and non-coding SNPs spaced along the human genome. Briefly, DNA was denatured in 0.1N NaOH and neutralized prior to isothermal amplification. Amplified DNA was fragmented and then hybridized to locus-specific 50mers that make up the array for 16–24 h with rocking at 48°C. After removal of unbound or non-specifically annealed DNA, single base extension of the 50mer oligonucleotides was performed with labeled nucleotides, which were scanned using an Illumina iScan with autoloader 2.x. Data analysis was performed using Illumina Genome Studio 2.0.

To validate the sequence of rs10404939 in samples from blunt trauma patients, PCR primers were designed to amplify 216 base pairs around the SNP base. The forward primer sequence was 5'-CCAGAGGCCCCTCACAAAC-3', and the reverse primers sequence was 5'-GGCAGGTGAGAGGAGAGGAA-3'. DNA concentration was quantified using NanoDrop (ThermoFisher Scientific, Pittsburgh, PA). Cleaned PCR samples were sequenced using Sanger sequencing with the reverse primer at University of Pittsburgh Genomics Research

Core and Psomagen (Rockville, MD). Sequence alignments were performed using SnapGene (GSL Biotech LLC, San Diego, CA) and genotypes were confirmed by visual inspection of the sequencing chromatogram.

For the ALIR cohort, genomic DNA was isolated from whole blood using Puregene DNA isolation kit (Qiagen). A custom Taqman SNP assay was designed using the Custom TaqMan® Assay Design Tool (ThermoFisher Scientific) for genotyping rs10404939. Since the genomic region flanking rs10404939 is extremely polymorphic (https://www.ensembl.org), we manually interrogated the allele frequencies of 17 known SNPs in the 84 base pair sequences flanking rs10404939. All these SNPs are extremely rare with 2 SNPs lacking population frequency data and 15 SNPs with <0.01% population frequencies. The ancestral alleles for these 17 known SNPs were used for primer and probe design. The assay was verified using 24 DNA samples with known genotypes determined by Sanger Sequencing. Genotyping was performed as described[54] using QuantStudio™ 5 System (ThermoFisher Scientific).

### Assays of inflammation biomarkers

To assay inflammation biomarkers in blunt trauma patients (Figure 1A), whole blood samples were drawn in EDTA-treated tubes, which were kept on ice and centrifuged to obtain plasma, and then stored at −80°C until assays were performed. The Luminex™ 100 IS analyzer (Luminex, Austin, TX) and Human Cytokine/Chemokine MILLIPLEX™ Panel kit (Millipore Corporation, Billerica, MA) were used to measure plasma levels of Eotaxin (CCL11), interleukin (IL)-1β, IL-1 receptor antagonist (IL-1RA), IL-2, soluble IL-2 receptor-α (sIL-2Rα), IL-4, IL-5, IL-6, IL-7, IL-8 (CCL8), IL-10, IL-13, IL-15, IL-17A, interferon (IFN)-α, IFN-γ, IFN-γ inducible protein (IP)-10 (CXCL10), monokine induced by gamma interferon (MIG; CXCL9), macrophage inflammatory protein (MIP)-1α (CCL3), MIP-1β (CCL4), monocyte chemotactic protein (MCP)-1 (CCL2), granulocyte-macrophage colony stimulating factor (GM-CSF) and tumor necrosis factor alpha (TNF-α). Human Th17 MILLIPLEX™ Panel kit (Millipore Corporation, Billerica, MA) was used to measure IL-9, IL-21, IL-22, IL-23, IL-17E/25, and IL-33. $NO_2^-/NO_3$ levels were measured by a Griess Reagent colorimetric assay (Cayman Chemical, Ann Arbor, MI). IL-1 receptor-like 1/suppression of tumorigenicity-2 (ST2) was measured by a sandwich ELISA assay (R&D Systems, Minneapolis, MN). All cytokine/chemokine concentrations are given in pg/ml; $NO_2^-/NO_3^-$ concentrations are in µM. Experimental data are shown as mean $\pm$ SEM.

For the ALIR cohort (Figure 1B), plasma samples were collected and processed as described previously,[33] and with a custom made Luminex™ panel (R&D Systems) we measured 10 host response biomarkers belonging to different biologic pathway categories: a. innate immune responses (IL-6, IL-8, IL-10, soluble tumor necrosis factor receptor 1 [sTNFR1], ST-2, fractalkine), b. epithelial injury (receptor of advanced glycation end-products [RAGE]), c. endothelial injury (Angiopoietin-2) and d. host-response to bacterial infections (procalcitonin and pentraxin-3).

### Algorithm for determining high-prevalence SNPs

To determine which SNPs from the CoreExome array in our trauma patient cohort were both prevalent and potentially associated with significant differences in plasma biomarkers, we created a Python (Python Software Foundation, Python Language Reference, version 3.7. Available at http://www.python.org)-based algorithm (*SNPScanner* [pseudocode in Data S1]). The algorithm proceeds as follows. Let $x$ denote a singular SNP listed in an input data set of biomarker levels and SNP genotypes. Only its two mutually exclusive homozygous genotypes—$x^{AA}$ and $x^{BB}$—were considered for this study, with their respective populations denoted as $p_x^{AA}$ and $p_x^{BB}$ and their population sizes as $n_x^{AA}$ and $n_x^{BB}$. SNPScanner located each of the biomarkers, rows containing SNPs, and the placement of AA and BB in all of the SNP rows. All SNPs were screened for a minimum of 20 patients required for each genotype and a population distribution $0.9 \times n_x^{AA} \leq n_x^{BB} \leq 1.1 \times n_x^{AA}$ (allowing a maximum deviation of $\pm 0.10 \times$ genotype population size). For each biomarker associated with each SNP that met the population distribution criteria, a two-sided Mann-Whitney U[55] (MWU) function from SciPy[56] was used to calculate the MWU score and p value. If the corresponding p value was less than 0.05, the SNP-biomarker permutation was considered statistically significant; this relatively high p value was used given the additional levels of stringency inherent in the subsequent aspects of the analysis. To process this dataset of 380 patients, 551,839 SNPs, and serial levels of 30 biomarkers, *SNPScanner* required an average of 96.2 s $\pm$ 1.7 s per SNP and resulted in a total runtime of 774,073.16 s (8d:23h:01min:13.16s) on an Intel Xeon®-class workstation.

### Assessment of clinical differences in control SNPs

After *SNPScanner* had been performed on the trauma cohort, the evenly-distributed SNPs which had resulted in zero statistically significant plasma biomarkers were examined for genotype-correlated differences in clinical post-traumatic outcomes. We created a Python-based code to pull the names of the 567 SNPs with zero significant biomarkers from the output file of the previous *SNPScanner* run (Data S2). Only the two homozygous genotypes were considered, and the SNPs' inclusion in the output signified that *SNPScanner* had already identified them as having the same population distribution as before. The Python random module (https://docs.python.org/3/library/random.html) was used to select a random sample of fifteen of the 0-biomarker SNPs. From there, each SNP's homozygous population was pulled and the patients' ICULOS, total LOS, ventilation days, and MODS were collated by parameter. The MODS values, from MOD1 to MOD7 were combined by ascending order of day into one Series. Once the data was organized, a MWU was performed between the AA and BB populations on each parameter to determine a p value. This process was iterated 200 times. If a SNP had a p value less than 0.05 for a parameter, the parameter was considered significant for that SNP. The number of SNPs with at least one significant parameter in each batch of fifteen was also recorded.

## Statistical and computational analyses

Student's $t$ test, Mann-Whitney Rank Sum Test as well as Krukal-Wallis Analysis of Variance (ANOVA) on Ranks/One-Way and Two-Way ANOVA were conducted to analyze the differences in ISS, AIS, ICU LOS, MODS, and time-dependent changes in plasma biomarkers in A/A, G/A, and G/G patients using SigmaPlot (Systat Software, San Jose, CA) as appropriate.

## Power calculations

To examine the available statistical power offered by our feasibility datasets to detect statistically significant results for the main endpoints of our genetic association analyses, we conducted the following power calculations:

For the trauma cohorts, we considered the MODS difference between genotypes as the primary endpoint for comparisons and performed separate power calculations given the observed differences in MODS between genotypes in the derivation and validation cohort.

For the derivation cohort which included less severely injured patients, we considered a range of differences in means of 0.05–1.2 (based on distribution of observed scores in Figure 3D). For the validation trauma cohort, which included more severely injured patients, we considered a range of differences in means between 1.0-2.0. We conducted power calculations with the *pwr* package in R and the pwr.t2n.test function for the range of the effect sizes, the given sample sizes for each of the genotypes, and an alpha level of 0.05 for the two-way ANOVA comparisons. Results are shown in Table S2 and validate that our cohorts had sufficient sample size that afforded statistical power to detect the observed differences between genotypes for MODS.

For the validation non-trauma cohort and the primary endpoint of differences in survival as detected by a Cox-proportional hazards model (observed hazards ratio [HR] of 1.4), we conducted a power analysis for a range of HRs from 1.2-1.8. We used the *powerSurvEpi* package in R. Results are shown in Table S3 and demonstrate that our sample size afforded a 56% power to detect a significant HR of 1.4, with greater than 80% power to detect a significant HR $\geq$ 1.6.

## Dynamic network analysis

Dynamic Network Analysis (DyNA) was carried out to define, in a granular fashion, the central inflammatory network nodes as a function of both time and patient subgroup. In the trauma cohort, DyNA networks were created over twelve consecutive time periods (4-8h, 8-12h, 12-16h, 16-20h, 20-24h, 24h-D1, D1-D2, D2-D3, D3-D4, D4-D5, D5-D6, D6-D7) using MATLAB® software as described previously.[29,57,58] In the ALIR cohort, DyNA networks were created over three time-intervals (baseline [d1] - middle [d3-6], middle-late [d7-10], and late-very late [d11-14]) as described above. Connections, defined as the number of inflammatory biomarkers that were positively or negatively correlated across time intervals, were created if the Pearson correlation coefficient between any two nodes (biomarkers) at the same time-interval was greater or equal to a threshold ranging from an absolute value of 0.7 (a correlation value commonly used to characterize trajectories that move in parallel either up or down) to 0.95, as appropriate. The network complexity for each time-interval was calculated using the following formula: Sum $(N_1 + N_2 + … + Nn)/(n - 1)$, where N represents the number of connections for each biomarker and n is the total number of biomarker analyzed.

## Genetic association testing

Genetic associations for clinical/biomarker data and genotypes of interest were tested for the recessive, additive and dominant genetic models. We report resultant p values without correction for multiple testing. We conducted analyses in all subjects with available biospecimens in the discovery and validation cohorts. We did not perform any post-hoc power calculations given that the samples sizes were predetermined based on feasibility. We conducted survival analysis by constructing Kaplan-Meier curves by genotype (and the corresponding models of inheritance) and by building Cox proportional hazards models adjusted for the effects of age, sex, and race.

*Linkage disequilibrium analysis.* Publicly available linkage disequilibrium (LD) data from the LDlink webtool (https://analysistools.cancer.gov/LDlink/)[36] in the 220-kilobase pair (kb) region surrounding rs10404939 were used to identify variants in LD with rs10404939. Candidate genes were identified as those genes containing single nucleotide variants that are in measurable LD as measured by $R^2$ with rs10404939 using 100,000 base pair scanning windows in the LDproxy tool based on an American population with European ancestry (CEU).

*Quantitative trait locus analysis.* The gene expression potential affected by rs10404939 was queried using the publicly available data from the Genotype-Tissue Expression (GTEx) Project portal (https://www.gtexportal.org/home/snp/rs10404939). We focused our analysis on the expression of quantitative trait loci (eQTL) on whole blood (670 subjects) and lung (514 subjects) tissues. Additionally, known splicing quantitative trait loci (sQTL) were also queried in testis because rs10404939 is located in the *LYPD4* gene encoding a sperm-related protein (322 subjects), as well as whole blood (670 subjects). Only known genes were included in the results.

## Study approval

As stated above, the University of Pittsburgh Institutional Review Board approved the study and written informed consent was provided by all participants or their surrogates in accordance with the Declaration of Helsinki.

## Role of funding source

No funding sources had any role in the writing of the manuscript or the decision to submit it for publication. The authors were not precluded from accessing data in the study, and they accept responsibility to submit for publication.