Investigations into SCAMP5, a candidate lupus risk gene expressed in plasmacytoid dendritic cells

Mustafa H Ghanem, Andrew J Shih, Himanshu Vashistha, Latanya N Coke, Wentian Li, Sun Jung Kim, Kim R Simpfendorfer, Peter K Gregersen

ABSTRACT
Objective We have investigated the molecular function of SCAMP5, a candidate lupus risk gene for SLE exclusively expressed in plasmacytoid dendritic cells (pDCs) among peripheral leukocytes.
Methods We tested the independence of the association in SCAMP5 with SLE by performing conditional analyses. We profiled the expression pattern of SCAMP5 among circulating leukocytes at the transcript and protein levels. Using lentiviral vectors, we localised the subcellular distribution of SCAMP5 alongside the interferon secretory pathway. We analysed pDCs for the expression of SCAMP5 and interferon production capacity by SCAMP5 genotype. Finally, we examined pDC-specific SCAMP5 isoforms by total RNAseq analysis and examined for genotype-associated quantitative differences therein.
Results A conditional analysis revealed evidence of an independent genetic association of SCAMP5 with SLE. Among circulating leukocytes, SCAMP5 is uniquely expressed in pDCs at the transcript and protein levels, with main presence in the Golgi apparatus and minor presence at the cell periphery. In live cells, SCAMP5 displayed dynamic Golgi–cell surface trafficking and localised with the interferon secretory pathway. SCAMP5 did not differ in expression levels in pDCs between genotyped donors; however, a transient interferon secretory defect was noted in pDCs from donors carrying the risk genotype.
Conclusions SCAMP5 constitutes a novel SLE risk gene on the basis of genomic data and expression in a cell type widely implicated in SLE pathogenesis. While we could not find evidence of quantitative expression differences in SCAMP5 between genotyped donors, SCAMP5 remains an attractive gene to explore given its highly restricted expression pattern and colocalisation with interferon secretion.

INTRODUCTION
Plasmacytoid dendritic cells (pDCs) are rare innate immune cells which express the pattern recognition receptors toll-like receptors 7 and 9 (TLR7 and TLR9), which together with their constitutive expression of the type I interferon (IFN-I) inducer interferon regulatory factor 7 (IRF7) are poised to rapidly respond to viral infection.1 Secretion of massive amounts of IFN-α, particularly IFN-α, is characteristic of pDCs, and downstream signalling in response to IFN drives induction of an antiviral programme in host cells, an ‘IFN signature’ that is constitutively activated in the blood of many patients with SLE.2 A role of pDCs in the pathogenesis of SLE has also been suggested by their presence in cutaneous lupus lesions and in lupus nephritis.3 4 Furthermore, antibody-mediated depletion of pDCs has demonstrated therapeutic efficacy in human trials.5 From the genomic standpoint, several SLE risk genes are enriched for expression in pDCs, including TLR7, IRF7, IRF8 and DNASE1L3.6 7

What is already known about this subject?
► SLE is associated with overactivity of the type I interferon pathway, and plasmacytoid dendritic cells (pDCs) are major producers of type I interferon in this disease.

What does this study add?
► A polymorphism on chromosome 15q24 associates with SLE by virtue of a novel candidate SLE risk gene, SCAMP5.
► We confirmed that among circulating leukocytes, SCAMP5 is uniquely expressed in pDCs at the transcript and protein levels, with main presence in the Golgi apparatus and minor presence at the cell periphery.
► We detected a transient interferon secretory defect in pDCs from donors homozygous for the SLE risk genotype, providing a potential link between a functional consequence of the genetic risk variant and the SLE risk association.

How might this impact on clinical practice or future developments?
► Molecular dissection of heritable traits in type I interferon secretion may unveil novel opportunities for drug development.

Key messages
pDC production of IFN-\(\alpha\) is potentially a major driver of the IFN signature. IFN-\(\alpha\) signalling has been shown to enhance antigen presentation, promote B cell differentiation and suppress regulatory T cells—processes central to the development of SLE.8–10

In their recent meta-analysis of genomewide association studies (GWAS) of SLE, Bentham et al11 reported an association at single-nucleotide polymorphism (SNP) rs2289583, with an OR of 1.19 and a \(p\) value of \(6\times10^{-15}\). rs2289583 occurs in intron 6 of \(\text{SCAMP5}\), a member of the secretory carrier membrane protein (SCAMP) family. The SCAMPs are tetraspan membrane proteins with cytoplasmic N-termini and C-termini and both luminal and extraluminal peptide loops connecting the transmembrane domains along the surface of intracellular trafficking vesicles (online supplemental figure S1).12 SCAMPs 1–3 are ubiquitously expressed, whereas SCAMPs 4–5, lacking the characteristic N-terminal NPF repeats of the other SCAMPs, display a more limited expression pattern, largely restricted to the brain.13 This expression profile of \(\text{SCAMP5}\) in the central nervous system appears to be conserved among mammalian species.

Outside of the brain, we noted a strong signal for \(\text{SCAMP5}\) expression in pDCs.14 Given the multiplicity of evidence for the role played by pDCs in SLE, and the association of this risk locus with the disease, we sought to determine whether \(\text{SCAMP5}\) contributed to SLE by modulating a key function of pDC biology: IFN-\(\alpha\) secretion.

METHODS
See online supplemental material.

RESULTS
\(\text{SCAMP5}\) associates independently with SLE on chr15q24

A recent SLE GWAS meta-analysis reported on susceptibility loci concentrated in genes involved in innate and adaptive immune function.11 This study coupled analyses of their respective discovery and replication cohorts with a previous SLE GWAS from our group.15 Among the associations was the polymorphism rs2289583. CSK was annotated as the likely gene explaining the association signal, citing prior work from our group establishing a link between \(\text{CSK}\) and SLE in this genomic region.16 However, inspection of SNP associations across this region reveals that the lead SNP from this analysis lies within the \(\text{SCAMP5}\) gene, ~155 kb telomeric to \(\text{CSK}\). This raised the question as to whether \(\text{CSK}\) drives the association signal at rs2289583. We examined the primary data in this region, with individual-level data on 4036 SLE cases and 6059 controls from the main GWAS of the aforementioned study.11 As shown in figure 1A (left panel), two SNPs associated with SLE risk achieve genome-wide significance in the region: rs2289583 in the last intron of \(\text{SCAMP5}\) (OR=1.2, \(p=3.4\times10^{-6}\)) and rs1378940 in the first intron of \(\text{CSK}\) (OR=1.17, \(p=2.8\times10^{-6}\)). Our group previously identified B cell phenotypes associated with a lupus risk variant in the \(\text{CSK}\) gene (rs34933034).16 The rs34933034 SNP was not genotyped in the Bentham et al data set. In fact, this variant in \(\text{CSK}\) is poorly imputed by other variants in the region—the lead SNP in \(\text{CSK}\) in this analysis (rs1378940) exhibits an \(r^2=0.34\) with rs34933034. Moreover, the signal at rs2289583 is unlikely to reflect the effects of our previously identified association at \(\text{CSK}\), as the linkage disequilibrium (LD) between these two SNPs is also quite modest (\(r^2=0.2\) between rs34933034 and rs2289583).

Therefore, we carried out formal conditional analysis to seek evidence for \(\text{SCAMP5}\) as a secondary or alternative driver of the SLE risk association in the \(\text{CSK}/\text{SCAMP5}\) region. On conditioning on the associated SNP within \(\text{SCAMP5}\), the signal at rs1378940 within \(\text{CSK}\) is substantially reduced (figure 1A, middle panel). In contrast, the association signal in \(\text{SCAMP5}\) (rs2289583 OR=1.2, \(p=3.4\times10^{-6}\)) was only partially reduced following conditioning on the lead risk-association SNP in \(\text{CSK}\) (\(p_{\text{cond}}=1.66\times10^{-3}\)) (figure 1A, right panel), indicating that the signal at \(\text{SCAMP5}\) is independent of the association signals observed within \(\text{CSK}\) itself.

\(\text{SCAMP5}\) is uniquely expressed in pDCs among peripheral blood leukocytes

Having established that \(\text{SCAMP5}\) may represent a novel SLE risk gene, we sought to probe the functional biology underlying this association. \(\text{SCAMP5}\) is reported to be expressed uniquely by pDCs among peripheral blood leukocytes in humans (BioGPS17 and Human Protein Atlas14; online supplemental figure S2). To confirm this, we sorted leukocyte subsets from four healthy donors and examined \(\text{SCAMP5}\) expression by reverse transcription (RT)–PCR. \(\text{SCAMP5}\) transcript was detected in pDCs from all donors and was absent in their B, T and NK cells, as well as monocytes and conventional dendritic cells (cDCs) (figure 1B, left panel). To determine whether \(\text{SCAMP5}\) is expressed at the protein level in pDCs, we resolved lysates from several leucocyte subsets and detected \(\text{SCAMP5}\) protein using the polyclonal antibody (pAb) against \(\text{SCAMP5}\) employed by the Human Protein Atlas. We found that \(\text{SCAMP5}\) is indeed expressed at the protein level only in pDCs (figure 1B, middle panel). To agnostically determine whether any other (unsorted) peripheral blood leucocyte subset expressed \(\text{SCAMP5}\), RBC-depleted whole blood was stained intracellularly for \(\text{SCAMP5}\) using the same pAb. Only pDCs stained for \(\text{SCAMP5}\), confirming the unique expression of this protein in the circulation (figure 1B, right panel).

To localise expression of \(\text{SCAMP5}\) within pDCs, fixed and permeabilised pDCs were stained for \(\text{SCAMP5}\) and visualised by immunofluorescence. Colocalisation was evident with the Golgi marker golgin-97, with additional evidence of staining for \(\text{SCAMP5}\) at the cell periphery (figure 1C).

The novelty of \(\text{SCAMP5}\) as an SLE risk gene, coupled with the striking expression pattern restricted to pDCs, a cell type highly invoked in SLE pathogenesis, prompted us to further explore a role of \(\text{SCAMP5}\) in modulating pDC function.
SCAMP5 traffics between the Golgi stack and cell surface and colocalises with clathrin and the IFN-I export pathway

The SCAMP protein family remains understudied. To develop tools to enable a more complete understanding of SCAMP5 in particular, several SCAMP5 constructs were generated and cloned into lentiviral vectors. To localise and track SCAMP5 in live cells, a SCAMP5-mKate2 fusion protein was designed. SCAMP5-mKate2 was transduced...
into human embryonic kidney (HEK) cells. SCAMP5-mKate2 localised mainly to the Golgi body, with puncta observed at the cell periphery (figure 2). This distribution matched the SCAMP5 localisation patterns observed by immunofluorescence on endogenously expressed SCAMP5 in primary pDCs (figure 1C). Intracellular vesicles bearing SCAMP5 were observed to traffic in both anterograde and retrograde directions between the cell surface and the Golgi (lower right images representing the magnified region indicated by the white box, white arrows trace a unique vesicle across acquisition intervals in the retrograde (upper panels) and anterograde (lower panels) trajectories). See online supplemental materials online supplemental files for SCAMP5-mKate2 construct design. Correlate with online supplemental videos 1–4. 63× magnification with oil immersion. NucBlue delineates the nucleus. BF, bright-field; HEK, human embryonic kidney.

To probe for a possible interaction between SCAMP5 and the IFN-α trafficking pathway, the fluorescent protein mVenus was generated as a soluble expression construct, designated ‘cytosolic/c-mVenus’ or with the signal peptide from human IFN-α at the N-terminus (designated ‘IFNsp-mVenus’). Both c-mVenus and IFNsp-mVenus were cloned downstream of a doxycycline-inducible promoter. Expression of c-mVenus in HEK cells revealed a fluorescence pattern that was diffusely present in the cytoplasm, as expected, including in lamellipodia. IFNsp-mVenus displayed a more restricted fluorescence pattern, highlighting the endoplasmic reticulum (ER) throughout the cell (figure 3A and online supplemental videos 5; 6). Both c-mVenus and IFNsp-mVenus were induced in a doxycycline dose-responsive manner. IFNsp-mVenus was detected in the culture media and cell lysates, whereas expression of c-mVenus was only detectable in cell lysates (figure 3B), demonstrating that IFNsp-mVenus was actively secreted. Coexpression of IFNsp-mVenus with SCAMP5-mKate2 demonstrates abundant colocalisation of SCAMP5 and IFNsp-mVenus within the Golgi (figure 3C and online supplemental video 7). Moreover, IFNsp-mVenus, SCAMP5 and the clathrin light chain could be observed within the same vesicle (figure 3D). As clathrin is involved in both endocytic and exocytic pathways, we hypothesised that SCAMP5 is an integral component of the trafficking machinery in cells in which it is expressed and is positioned to participate in IFN-I secretion mechanisms.

Quantitative expression analysis on SCAMP5 in pDCs from genotyped donors and assays of pDC IFN-I output

We next sought to investigate the effect SCAMP5 genotypes may have on SCAMP5 expression in pDCs and to determine a potential impact on pDC function. To this end, we obtained fresh blood from homozygous carriers of the risk haplotype and non-risk haplotype tagged by rs2289583 from the Genotype and Phenotype Registry, a resource previously used...
in genotype-phenotype investigations in autoimmune disease (online supplemental tables S1 and S2). 19–21

For expression quantitative trait locus (eQTL) analysis, quantitative RT-PCR for SCAMP5 and the neighboring genes, CSK, COX5A, RPP25 and PPCDC, was performed. No significant differences in mRNA levels were seen for SCAMP5, COX5A, RPP25 or PPCDC (figure 4A and online supplemental figure S3).

Nominal evidence of an eQTL effect was observed for CSK, with elevated CSK transcript in the risk haplotype when normalised to 18s rRNA or IRF7 (online supplemental figure S3). To explore whether the risk haplotype polymorphisms might exert their effect at the level of translation, SCAMP5 protein was measured in pDC lysates. Median SCAMP5 protein levels were increased in the risk haplotype when normalised to

---

**Figure 3** SCAMP5 co-traffics with the IFN-α secretory pathway. Soluble c-mVenus and IFNsp-mVenus were transduced into HEK cells under the regulation of a Tet-On system. (A) Induction using doxycycline revealed that c-mVenus was diffusely expressed in the cytoplasm, whereas IFNsp-mVenus exhibited a lacy pattern characteristic of the ER. Insets represent brightfield images. (B) c-mVenus and IFNsp-mVenus were expressed in a doxycycline dose-responsive manner as determined by western blotting using an anti-GFP antibody. c-mVenus was restricted to intracellular expression, whereas IFNsp-mVenus was detected in cell lysates as well as the culture media. (C) Coexpression of IFNsp-mVenus and SCAMP5-mKate2 reveals colocalisation of SCAMP5 with the IFN-α secretory pathway. (D) SCAMP5 colocalises with IFNsp-mVenus and clathrin in the Golgi apparatus as well as in the vesicles on the cell surface. Magnification of a single vesicle (white box) reveals SCAMP5 and clathrin coating the surface of the vesicle, with IFNsp-mVenus occupying the lumen. (A and C) 63× with oil immersion; (D) 100× with oil immersion. NucBlue delineates the nucleus. See online supplemental material for fluorescent protein designs. Correlate with online supplemental videos 5–7. BF, brightfield; ER, endoplasmic reticulum; GFP, green-fluorescent protein; HEK, human embryonic kidney; IFN, interferon.
As some risk variants only manifest endophenotypic effects upon cell stimulation (response eQTL, reQTL), we next assessed for a phenotypic difference between genotypes at the functional level. Culture media from pDCs stimulated with influenza H1N1, imiquimod or ODN-2216 for 4 and 24 hours were assayed for pan-subtype IFN-α. When comparing total results from donors for either haplotype, IFN output trended towards lower values for pDCs from the risk group (figure 5A). This difference reached statistical significance only for influenza virus stimulation at the earlier time point.

To determine whether the expression of SCAMP5 in pDCs was influenced by stimulation with influenza virus in a genotype-dependent manner, potentially explaining the kinetic differences in IFN-α production seen above, we assayed for SCAMP5 expression and observed that SCAMP5 is downregulated on stimulation with influenza virus by 4 hours, with no significant differences by genotype (figure 5B). Moreover, there were no consistent differences in IFN-α transcript induction in this condition between the genotypes (figure 5B).

Taken together, these findings suggest, but do not prove, that the SCAMP5 risk haplotype may be related to a delay in IFN protein production or secretion in the early period after viral infection.

Identification of novel SCAMP5 splice variants in pDCs

To address the possibility that the SCAMP5 risk haplotype influences pDC-specific SCAMP5 isoform expression not detected by our PCR and protein assays, we recalled an
DISCUSSION

We have investigated a novel candidate SLE risk gene, SCAMP5, located on chromosome 13q24. An intronic variant in CSK, ~155 kb distant from SCAMP5, was previously reported to be associated with SLE by our group and others.11 16 However, as shown in figure 1, a more detailed analysis of this region in a larger data set has suggested the presence of an independent genetic effect located within the SCAMP5 gene.

SCAMP5 is a member of the SCAMP gene family and is specifically expressed in neuronal cells of the central nervous system. Interestingly, SCAMP5 is present in the neurohypophysis, derived from neuroectoderm, and is absent in the adenohypophysis, derived from Rathke’s pouch, implying tight ontogenic control over SCAMP5 expression.13 In cells of the immune system, we found that SCAMP5 is uniquely and highly expressed in pDCs, a rare cell type with the capacity to produce large amounts of IFN-I. Given the importance of IFN-I in SLE, and the large body of literature attributing pDCs as a major source of IFN-I in SLE, we pursued further analysis of SCAMP5 function in pDCs.

SCAMP5 is highly conserved. Rare copy number variations in SCAMP5 can lead to severe neurological phenotypes in humans and mice, while coding-variant variants in SCAMP5 are extremely rare.23-26 Therefore, we expected that a pDC-specific regulatory variant acting in pDCs underlies the association with SLE. In neuronal cells, the SCAMP family of proteins has been implicated in various roles in intracellular vesicular trafficking, including cargo sorting, endocytosis and exocytosis.27 SCAMP5 was shown to be necessary for the sustained release of neurotransmitter-preloaded synaptic vesicles in response to tonic stimulation. In the absence of SCAMP5, membrane recycling pathways were disrupted, leading to the depletion of the synaptic vesicle reserve pool and neurons refractory to further stimulation.28 Elsewhere, SCAMP5 has been reported to facilitate the export of signal-peptide containing cytokines, whereas cytokines exported via Golgi-independent pathways were unaffected by SCAMP5 expression.29 Taken together, a highly specialised role for SCAMP5 in protein trafficking/membrane recycling seems likely.

pDCs are noteworthy for their capacity to secrete tremendous amounts of protein, including all members of the IFN-I family (with the exception of IFN-α). This undoubtedly requires efficient mechanisms for protein packaging and subsequent export. IFN-I is secreted along the classical pathway: a signal sequence directs nascent polypeptide chains to the ER, where the protein is translated directly into the ER lumen. Proteins destined for export are then transported to the Golgi apparatus, where sequential protein processing and sorting take place en route to the cell surface via protein-loaded vesicles.30 Using IFNsp-mVenus as a traceable surrogate for IFN-I secretion, we have observed that the signal peptide from IFN-α indeed directs protein synthesis to the ER, with later presence in the Golgi. We have also observed that SCAMP5 localises predominantly to the Golgi, with puncta present on the cell surface which traffic to the Golgi in a bidirectional fashion. With these observations, we hypothesised that SCAMP5 licenses the hypersecretory capacity of pDCs by continuous replenishment of Golgi membrane lost to exocytosis on the cell surface. While it stands to reason that any protein destined for secretion via the Golgi may be affected by this process, we focused on IFN-α secondary to its association with SLE and its relative abundance versus other pDC-derived cytokines.

To test our hypotheses, we isolated pDCs to high purity from female donors of childbearing age homozygous for risk or non-risk haplotypes about rs2289583. pDCs from these donors were used in expression-level and functional assays. SCAMP5 expression was similar between the genotypes in unperturbed pDCs, ruling out a baseline eQTL. At the protein level, SCAMP5 had a median higher expression in the risk group when normalised to a Golgi membrane protein loading control; however, the magnitude of this difference and the limited number of samples precluded statistical significance. The possibility that SCAMP5 is differentially regulated between the genotypes upon stimulation was not supported by our reQTL analysis. Finally, whether a novel SCAMP5 isoform exists in pDCs and explains the association with SLE was explored using RNAseq. Novel putative isoforms of SCAMP5 in pDCs were detected; however, they did not differ in expression levels between the genotypes.

The expression of neighbouring genes to either side of SCAMP5 was likewise similar, with the exception of CSK, which exhibits a minor (<1.2-fold) yet statistically significant increase in expression in the risk group. This may reflect the small residual LD between rs2289583 and the previously identified independent CSK risk haplotype, which has been associated with increased CSK expression in B cells. However, we note that CSK expression in pDCs is low in comparison with B cells. Thus, to the extent that we can identify pDC phenotypes regulated by rs2289583, SCAMP5 remained a leading candidate gene to explore.

Functionally, pDCs from risk donors produce less IFN-α on average compared with non-risk, when stimulated with imiquimod, ODN-2216 or influenza virus at 4 or 24 hours of stimulation. These differences were more pronounced with imiquimod and influenza virus—both TLR7 agonists—and at the earlier time point. These data run counter to the expectation that SLE risk variants in
the IFN pathway increase IFN production. With the exception of SLE variants in STAT4 which increase sensitivity to signalling through the IFN-α/β receptor, resulting in increased serum IFN activity,31 most IFN-related SLE risk variants indeed confer a greater predilection for IFN production.32 Statistically significant differences in these studies usually involve dozens or hundreds of representatives of each genotype, and IFN presence is often inferred through biological activity assays rather than direct IFN measurements as performed here.

Our observations raise several possibilities. The IFN production defect is more pronounced at 4 hours than at 24 hours, suggesting that a kinetic trait in IFN export may be at play. Rapid production of IFN is critical for antiviral defence. As viral prodrugs often precede onset of SLE symptoms, it may be the case that enhanced susceptibility to viral infection through slower IFN secretion may induce SLE in an otherwise susceptible host. Alternatively, the observed IFN secretion defect may represent an incidental finding unrelated to SCAMP5’s role in pDCs. That SCAMP5 colocalises with the IFN-α secretory pathway does not necessarily indicate an interaction between the two. CRISPR knockouts studies in primary pDCs may be necessary for establishing a definitive link between SCAMP5 and IFN secretion. We have successfully used CRISPR to knock out BDCA4 expression in primary human pDCs (online supplemental figure S5); however, knockout of SCAMP5 has proven elusive, possibly due to longer half-life of this protein in these short-lived cells.

Separately, we have investigated human pDC responses to influenza virus by secretome profiling using liquid chromatography tandem mass spectrometry and at the single pDC level using single cell RNAseq (Ghanem MH et al, 2021, manuscript in preparation). There, we observed that SCAMP5 is widely expressed in pDCs immediately ex vivo, but is downregulated upon stimulation, implying a role of SCAMP5 that is distinct from ongoing IFN export. Moreover, we observed evidence for granzyme B release from these cells via a mechanism consistent with degranulation rather than the classic secretory pathway. A role of SCAMP5 in this process would bear more similarity to synaptic vesicle release from neuronal cells where the function of SCAMP5 is more thoroughly described. These possibilities are being addressed by ongoing investigation. This work marks the first description of a role of SCAMP5 in pDC biology and IFN secretion; much remains to be explored.

Acknowledgements We would like to acknowledge the participants of the GaP Registry for their contribution to this work. Gila Klein, Margaret DeFranco, Kristine Elmakili and Doreen Flores coordinated GaP donor recall and sample collections. Professor Timothy Vyse (King’s College London) kindly provided primary SNP data on SLE cases and controls. Christopher Colon, Heriberto Borrero and Barrett Waling provided assistance with flow cytometry. Anthony Liew performed the library construction and sequencing for total RNAseq. We thank Professor David Castle (University of Virginia) for helpful discussions on SCAMP5. Lentiviral constructs were cloned using vectors provided by Addgene.

Contributors MHG, SJK, KRS and PKG conceived and designed the study. MHG, HV and LNC collected the experimental data. MHG and KRS analysed the experimental data. AS performed the RNAseq analysis. WL performed the statistical genetics analysis. MHG, KRS and PKG wrote the manuscript. PKG was the principal investigator on the study and serves as the guarantor of this work. All authors revised and approved the final manuscript.

Funding This work is supported by a career development award to KRS (NIH K01AR071502) and awards to SJK (R01 AR075565 and U19 AI144306), as well as departmental funds provided by the Center for Genomics and Human Genetics at the Feinstein Institutes for Medical Research.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The Genotype and Phenotype (GaP) Registry at the Feinstein Institutes for Medical Research provided de-identified blood from donors consented under an IRB-approved protocol (IRB# 09-081). The Committee for Participant Protection at the Feinstein Institutes for Medical Research approved this study (TAP0307.5.73.). The GaP is a subprotocol of the Tissue Donation Program (TDP) at Northwell Health and a national resource for genotype–phenotype studies (https://www.feinsteininstitute.org/robert-sboa-center-for-genomics-and-human-genetics/gap-registry/).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID ids Mustafa H Ghanem http://orcid.org/0000-0002-8019-5324
Andrew J Shih http://orcid.org/0000-0001-8375-7313
Himanshu Vashishtha http://orcid.org/0000-0003-4653-5011
Latanya N Coke http://orcid.org/0000-0001-9018-0188
Wentian Li http://orcid.org/0000-0003-1155-110X
Sun Jung Kim http://orcid.org/0000-0001-8535-5537
Kim R Simpfendorfer http://orcid.org/0000-0002-1906-2899
Peter K Gregersen http://orcid.org/0000-0003-1613-1518

REFERENCES