SUPPLEMENTAL MATERIAL.

University of Washington Center for Mendelian Genomics (UW CMG):

Michael J. Bamshad[1,2], Suzanne M. Leal[3], and Deborah A. Nickerson[1]

Peter Anderson[1], Marcus Annable[1], Elizabeth E. Blue[1], Kati J. Buckingham[1], Imen Chakchouk[3], Jennifer Chin[1], Jessica X Chong[1], Rodolfo Cornejo Jr.[1], Colleen P. Davis[1], Christopher Frazar[1], Martha Horike-Pyne[1], Gail P. Jarvik[1], Eric Johanson[1], Ashley N. Kang1, Tom Kolar[1], Stephanie A. Krauter[1], Colby T. Marvin[1], Sean McGee[1], Daniel J. McGoldrick[1], Karynne Patterson[1], Sam W. Phillips[1], Jessica Pijoan[1], Matthew A. Richardson[1], Peggy D. Robertson[1], Isabelle Schrauwen[3], Krystal Slattery[1], Kathryn M. Shively[1], Joshua D. Smith[1], Monica Tackett[1], Alice E. Tattersall[1], Marc Wegener[1], Jeffrey M. Weiss[1], Marsha M. Wheeler[1], Qian Yi[1], and Di Zhang[3]

[1]University of Washington

[2]Seattle Children's Hospital

[3]Baylor College of Medicine

Methods.

Subject enrollment.

All study participants were consented according to protocols approved by the Oregon Health & Science University, the University of Miami, or the Ohio State University. A diagnosis of DCM was made by echocardiography using standard diagnostic criteria for idiopathic dilated cardiomyopathy (IDC): namely left ventricular enlargement (LVE) in the presence of systolic dysfunction (left ventricular ejection fraction, LVEF <50%) after exclusion of other clinically relevant causes such as ischemic, infectious, or toxic cardiomyopathy. DCM severity was graded on a previously described ordinal classification scale[1]: 0 (no cardiovascular disease or evidence of DCM), I (LVE without systolic function or systolic dysfunction without LVE), II (asymptomatic DCM), III (symptomatic DCM with or without medical therapy), and IV (DCM

requiring advanced interventions including mechanical circulatory support, cardiac transplantation, or resulting in DCM-related death).

Exome Sequencing and Variant Adjudication.

Genomic DNA was extracted from peripheral blood samples following standard protocols[2-6]. Exome sequencing was completed at the University of Washington Genome Sciences Center using the NimbleGen V2 in-solution capture (Roche NimbleGen Inc., Madison, WI) and Illumina HiSeq platform (Illumina, San Diega, CA, USA) as part of a broader effort to characterize peripheral blood DNAs from 412 individuals in our DCM cohort. Reads were aligned to the hs37d5 version of the hg19/GRCh37 human reference sequence using BWA (Burrows-Wheeler Aligner). After upload to the GEM.app/GENESIS platform (The Genesis Project Foundation, Miami, FL, USA)[7], duplicate removal was performed using Picard MarkDuplicates, and indel realignment and base quality recalibration were performed using GATK IndelRealigner and BaseRecalibrator, respectively. Genotypes were called jointly for all samples using GATK UnifiedGenotyper, v.2.6.5. Using the GENESIS 2.0 web interface,[7] nonsynonymous or splice-altering variants affecting Consensus Coding Sequence (CCDS) transcripts were filtered with the following thresholds: ExAC MAF <0.001, gnomAD MAF <0.001, heterozygous alleles <6, homozygous alleles <6, Depth >8, Genotype Quality Filter >50, Quality Filter >35. Missense variants affecting the *TTN* gene were excluded from further analysis. Following previous rationale[8], we considered any variant with a pooled MAF <0.05% in gnomAD and/or ExAC to be of potential pathogenic significance. Variants in known or suspected DCM genes meeting above filtering criteria (Table S3) were sequence verified by Sanger methods (Figure S4) and adjudicated using the DCM Consortium and DCM Precision Medicine approach to DCM variant interpretation[9], which have been refined from current American College of Medical Genetics (ACMG)/ClinGen guidelines[10]. All primers used in this study are listed in Table S4.

Plasmid Generation.

To generate *SOS1* expression constructs, a wild-type *SOS1* pBABE-puro expression vector (pBABE-*SOS1*-WT, gifted by Dr. Benjamin Neel, NYUMC, New York, NY, USA) was used as a template for Quikchange II XL site-directed mutagenesis (Qiagen, Germantown, MD, USA)[11]. Variant constructs

were generated for the six *SOS1* variants identified in our cohort (Table 1), as well as for two pathogenic NS-associated variants with known gain of function effects on RAS signaling [p.(Met269Arg) and p.(Glu846Lys)][11, 12] to be used as positive controls for functional impact on RAS signaling. An additional *v*ariant [p.(Leu791Ile)], previously reported in two patients with NS[13] and detected in two unrelated families in our cohort (not shown), was included as a likely negative control. This variant was considered to be likely benign based on frequent occurrence in databases of normal genetic variation (ExAC, gnomAD), low conservation, and minimal predicted impact on protein structure (Table S5).

Immunoblots.

Relative expression levels of RAS signaling components were determined by western blot in HEK293T cells transfected with *SOS1* expression constructs. One day prior to transfection, 5-6 X 10^5 cells HEK293T cells (ATCC, Old Town Manassas, VA, USA, #CRL-3216, passage range: 7-12) were seeded onto poly-L-lysine treated 6-well plates. For each transfection, 3μg of plasmid DNA was complexed at a 1:2 ratio with jetPRIME (Polyplus, New York, NY, USA) or a 1:3 ratio with FuGENE HD (Promega, Madison WI, USA) reagent and incubated for 24 hours. Cells were subsequently serum-starved for 16 hours, stimulated for 15 minutes with 20ng epidermal growth factor (EGF) (MilliporeSigma, Burlington, MA), quenched in ice-cold DPBS, and lysed in 200μl RIPA buffer [50mM Tris, pH 8.0; 150mM NaCl; 0.5% Sodium deoxycholate; 0.1% SDS; 1.0% NP-40; 1X Halt protease and phosphatase inhibitor cocktail (Fisher, Hampton, NH, USA)]. 10μg of each protein sample was loaded onto 8% tris-glycine SDS-PAGE gels. Electrophoresis was carried out at 100V and proteins were transferred to Immobilon-P PVDF membranes (MilliporeSigma, Burlington, MA) overnight at 30V. Membranes were trimmed to size and incubated overnight using primary antibodies (Cell Signaling, Danvers, MA, USA; Abcam, Cambridge, UK) titrated to optimal concentrations: α-actinin (Cell Signaling, #6487, 1:10,000); Akt (Cell Signaling, #4691; 1:5,000), pAkt (Cell Signaling, #8200; 1:5,000), β-tubulin (Abcam, #ab6046; 1:20,000), ERK (Cell Signaling, #4695; 1:5,000), pERK (Cell Signaling, #8201; 1:5,000), JNK (Cell Signaling, #9252; 1:1,000), pJNK (Cell Signaling, #4668, 1:1,000, SOS1 (Abcam, #ab140621; 1:10,000). Membranes were washed 3 times for 15 minutes each in TBS with 0.1% Tween (TBST), incubated for 1 hour at room temperature with HRP-linked Anti-rabbit IgG (Cell Signaling, #7074; 1:2,000), and washed an additional 3 times in TBST. Membranes were developed for 5

minutes at room temperature in Clarity Western ECL substrate (Biorad, Hercules, CA, USA) and imaged on an Odyssey Fc Imaging System (Licor, Lincoln, NE, USA) running Image Studio, v.5.2 software.

Structural Analyses.

Computational modeling was completed using the human SOS1 protein (PDB: 3KSY)[14]. Missing regions of the core structure (residues 177-194, 405-417, and 747-749) were modeled based on sequence homology and inserted into the protein structure. This structure was then energy minimized in YASARA, v.17.1.28 (YASARA Biosciences GmBH, Vienna, Austria). The complete model was allowed to equilibrate for ~20-40ns until stable, as previously described[15]. Individual mutations were introduced via the 'swap' command, and the resulting models were allowed to equilibrate for ~25-40ns in explicit solvent at 310K, 150mM NaCl in duplicate, as previously described[16]. Specific interactions were visually confirmed and then monitored in 25ps intervals using YASARA. A .pdb snapshot of each mutant was generated every 5ns and resulting models were visualized using Pymol, v.2.0.2 (Schrodinger Inc., New York, NY, USA).

Statistical Analysis of Immunoblot Data.

The design structure of all experiments was blocked, with all lysates in each block derived from HEK293T cells aliquoted from the same flask. Within each block, null/untransfected (UT), empty vector (EV), wild-type *SOS1* (WT) and up to 5 variant *SOS1* plasmid transfection conditions were applied independently to fixed well positions on paired (EGF stimulated, EGF unstimulated) sets of six-well plates. Because transfection treatments and EGF stimulation were independently applied to wells containing cells from the same flask, each well constituted a single experimental unit. Total protein isolated from each experimental unit was quantified by Bradford Assay, and 10μg of lysate was loaded into separate wells of one or two 8% tris-glycine SDS-PAGE gels. Each gel contained a distinct experimental unit from the same block (see Table S6) in each lane, and lysates from the UT, EV, and WT transfection conditions were measured on every gel run for a given block.

Based on prior research showing the appropriateness of a lognormal multiplicative error model in the analysis of immunoblot data,[17] we posited such a model relating loading-control normalized densitometry measurements for a given protein (pERK, ERK, pAKT, AKT, pJNK, JNK, SOS1) to the

unobserved underlying abundance. For a given experimental unit, we assumed that the unobserved target protein abundance in 10μg of lysate, X_{ijn}, was determined according to the lognormal model:

$$X_{ijn} = \exp\left(\mu_{ij} + b_{k_{ijn}} + e_{w_{ijn}(k_{ijn})}\right) \quad (1)$$

where $i = 1\ (\mathrm{WT}), \ldots, I$ indexes variant transfection condition, $j = 1\ (\mathrm{No\ EGF}), 2$ indexes EGF stimulation condition, and $n = 1, \ldots, N_{ij}$ indexes the experimental units that received treatment combination ij (see Table S6 for N_{ij} in various experiments). The indexes $k_{ijn} \in \{1, \ldots, K\}$ and $w_{ijn} \in \left\{1, \ldots, W_{k_{ijn}}\right\}$ indicate the block and the well within block k_{ijn} containing experimental unit n for treatment ij (see Table S6 for K and $W_{k_{ijn}}$ in various experiments). In this model, μ_{ij} represents the expected log protein abundance for experimental condition ij, $b_{k_{ijn}} \sim N(0, \sigma_b^2)$ is a random effect reflecting the impact of block-level factors on target protein abundance in all experimental units from block k_{ijn}, and $e_{w_{ijn}(k_{ijn})} \sim N(0, \sigma_e^2)$ is a random effect reflecting the impact of unit-specific factors on target protein abundance for experimental unit ijn, which is also uniquely identified by its location in well w_{ijn} within block k_{ijn}. The notation $\sim$ is used throughout to denote independent realizations from the distribution on the right for each unique subscript of the term on the left; random effects are assumed throughout to be independent of one another unless otherwise noted.

If protein abundance for the target protein is within the linear range of detection for all conditions, we can incorporate an additional multiplicative error term[17] into a standard model for the background-corrected intensity measurements, Y_{ijnm}, on a particular gel[18] to obtain:

$$Y_{ijnm} = a_{g_{ijnm}} X_{ijn} \exp\left(\varepsilon_{ijnm}\right) \quad (2)$$

where $m = 1, \ldots, M_{ijn}$ indexes the measurements of experimental unit n of treatment combination ij and $g_{ijnm} \in \{1, \ldots, G\}$ indexes the gel on which this measurement was performed (see Table S6 for M_{ijn} in various experiments). In this model, $a_{g_{ijnm}}$ is the gel-specific constant of proportionality for all measurements of the target protein on gel g_{ijnm}, and $\varepsilon_{ijnm} \sim N(0, \sigma_\varepsilon^2)$ is a random error specific to that measurement.

Variation in $a_{g_{ijnm}}$ among gels is the justification for analyzing within-gel ratios of intensity measurements because $a_{g_{ijnm}}$ cancels out.[18] However, under the error model above, these ratios have a

lognormal rather than a normal distribution, which implies that fitting linear mixed models to untransformed ratios does not yield maximum likelihood estimates of relevant model parameters. While linear mixed models fit to log-transformed ratios could yield maximum likelihood estimates for certain experimental designs, some of the experiments analyzed here measured the same experimental units on multiple gels. For example, in the assessment of ERK and pERK expression, the total number of experimental units in certain blocks exceeded the available lanes on the gel. Thus, measurements needed to be made on two gels, and the same protein lysates for the UT, EV, and WT transfection conditions were measured on both gels along with different sets of variant transfection conditions from the same block (see Table S6). This design structure induces a complex dependence structure between ratios on different gels from the same block that is difficult to model. A simpler alternative approach is to model the gel-to-gel variation in the proportionality constant using a multiplicative error model:

$$a_{g_{ijnm}} = \exp\left(\alpha + d_{g_{ijnm}}\right) \tag{3}$$

In this model, α is the expected log proportionality constant for the target protein across gels, and $d_{g_{ijnm}} \sim N(0, \sigma_d^2)$ is a random effect reflecting gel-to-gel variation in this proportionality constant. Combining (1), (2), and (3) and taking the natural logarithm yields:

$$\ln Y_{ijnm} = \mu_{ij} + \alpha + b_{k_{ijn}} + d_{g_{ijnm}(k_{ijn})} + e_{w_{ijn}(k_{ijn})} + \varepsilon_{ijnm} \tag{4}$$

where now $g_{ijnm} \in \left\{1, \ldots, G_{k_{ijn}}\right\}$ indexes the gel within the block k_{ijn} on which measurement m of experimental unit n of treatment combination ij was performed. The nesting of the gel indexes within each block reflects that each gel contained only experimental units from that block.

Background-corrected intensity measurements of the target protein are typically divided by lane normalization factors based on the ratio of the background-corrected intensity measurement of the loading control protein in the same lane to the maximum across lanes on the same gel. The unobserved abundance of an ideal loading control protein in 10μg of lysate, X_{ijn}^z, should be independent of experimental conditions and have essentially no biological variability across blocks or experimental units. Thus, (1) can be replaced by $X_{ijn}^z \equiv \exp(\gamma^z)$ for an ideal loading control protein. Letting Z_{ijnm} denote the background-corrected loading control intensity measurement from the same lane as Y_{ijnm}, combining the modified version of (1) with (2) and (3), and taking the natural logarithm yields:

$$\ln Z_{ijnm} = \gamma^z + \alpha^z + d^z_{g_{ijnm}(k_{ijn})} + \varepsilon^z_{ijnm} \tag{5}$$

where α^z is the expected log proportionality constant for the loading control protein across gels, $d^z_{g_{ijnm}(k_{ijn})} \sim N\left(0, \sigma^2_{d^z}\right)$ is a random effect reflecting gel-to-gel variation in this proportionality constant, and $\varepsilon^z_{ijnm} \sim N\left(0, \sigma^2_{\varepsilon^z}\right)$ is a random error specific to this measurement. Under this model, variability in Z_{ijnm} reflects gel-to-gel variation in the proportionality constant as well as lane-specific variation due to technical factors, including loading and transfer, rather than biological variability in the loading control protein abundance. Letting $Z^*_{g_{ijnm}(k_{ijn})}$ be the maximum loading control intensity measurement obtained across all lanes on gel $g_{ijnm}\left(k_{ijn}\right)$ and using (4) and (5), we can model the loading-control normalized intensity measurement $\tilde{Y}_{ijnm}$ as:

$$\begin{aligned} \ln \tilde{Y}_{ijnm} &= \ln Y_{ijnm} + \ln Z^*_{g_{ijnm}(k_{ijn})} - \ln Z_{ijnm} \\ &= \mu_{ij} + \alpha + b_{k_{ijn}} + d_{g_{ijnm}(k_{ijn})} + e_{w_{ijn}(k_{ijn})} + u_{g_{ijnm}(k_{ijn})} + \varepsilon_{ijnm} - \varepsilon^z_{ijnm} \\ &= \mu_{ij} + \alpha + b_{k_{ijn}} + d_{g_{ijnm}(k_{ijn})} + e_{w_{ijn}(k_{ijn})} + \delta_{ijnm} \end{aligned} \tag{6}$$

where $u_{g_{ijnm}(k_{ijn})} = \max\left(\boldsymbol{\varepsilon}^z_{g_{ijnm}(k_{ijn})}\right)$, $\boldsymbol{\varepsilon}^z_{g_{ijnm}(k_{ijn})}$ denotes the vector of ε^z_{ijnm} from experimental units in the same block k_{ijn} measured on the same gel $g_{ijnm}\left(k_{ijn}\right)$, and $\delta_{ijnm} = u_{g_{ijnm}(k_{ijn})} + \varepsilon_{ijnm} - \varepsilon^z_{ijnm}$. Note that the term $\gamma^z + \alpha^z + d^z_{g_{ijnm}(k_{ijn})}$ from (5) cancels out when this normalization is used because it is the same for all loading control measurements on the same gel. This cancellation also permits the use of different loading controls on different gels in the same experiment by assuming that $\sigma^2_{\varepsilon^z}$ is the same for all loading controls. Because loading controls are specifically chosen to reflect variation due to technical factors that should have similar multiplicative effects on the intensity measurements for all proteins in a lane, this assumption should be reasonable.

We now establish the properties of δ_{ijnm}. Random measurement errors for the target and loading control proteins, ε_{ijnm} and ε^z_{ijnm}, are independent of the random effects to the left of $u_{g_{ijnm}(k_{ijn})}$ in (6), as is $u_{g_{ijnm}(k_{ijn})}$ as a function of ε^z_{ijnm}. The term δ_{ijnm} is therefore independent of the other random effects in (6). It is also independent of the same term for measurements on another gel due to the independence of the ε^z_{ijnm} across measurements. For a gel with l lanes, the term $u_{g_{ijnm}(k_{ijn})}$ is the maximum of l independent $N\left(0, \sigma^2_{\varepsilon^z}\right)$ random variables, which has expected value $v(l)\sigma_{\varepsilon^z}$ and variance $\sigma^2_v(l\,)\sigma^2_{\varepsilon^z}$ that are both functions

of l.[19, 20] While the exact distribution of $u_{g_{ijnm}(k_{ijn})}$ is not normal, it is known to diverge slowly from normality as the number of random variables increases and has fairly modest skewness and excess kurtosis for 12 or fewer random variables.[19, 20] In calculating the variance of δ_{ijnm}, we also need to consider that ε_{ijnm} and ε^z_{ijnm} for the same measurement are positively correlated because both are affected by the same lane-specific technical factors. Under the above assumptions, ε^z_{ijnm}, and by extension ε_{ijnm}, are also positively correlated with $u_{g_{ijnm}(k_{ijn})}$, which is a function of ε^z_{ijnm} for the current measurement. In fact, because the ε^z_{ijnm} are independent and identically distributed normal random variables, $\mathrm{Cov}\left(\varepsilon^z_{ijnm}, u_{g_{ijnm}(k_{ijn})} \middle| l\right) = l^{-1}\sigma^2_{\varepsilon^z}$.[21] Moreover, we have:

$$\mathrm{Cov}\left(\varepsilon_{ijnm}, u_{g_{ijnm}(k_{ijn})} \middle| l\right) = \mathrm{E}\left[\varepsilon_{ijnm} u_{g_{ijnm}(k_{ijn})} \middle| l\right] = l^{-1}\mathrm{E}\left[\varepsilon_{ijnm}\varepsilon^z_{ijnm}\right] = l^{-1}\sigma_{\varepsilon,\varepsilon^z}$$

where $\sigma_{\varepsilon,\varepsilon^z}$ is the covariance between ε_{ijnm} and ε^z_{ijnm} within the same measurement. The first equality follows from the fact that ε_{ijnm} has mean zero, and the second follows from the facts that $\mathrm{E}\left[\varepsilon_{ijnm}\varepsilon^z_{i'j'n'm'}\right] = 0$ unless $i'j'n'm' = ijnm$ and the maximum occurs at measurement $ijnm$ with probability l^{-1}. Thus, for a gel with l lanes, we have:

$$\begin{aligned}
\mathrm{E}\left[\delta_{ijnm} \middle| l\right] &= \upsilon(l)\sigma_{\varepsilon^z} \\
\mathrm{Var}\left(\delta_{ijnm} \middle| l\right) &= \sigma^2_{\upsilon}(l\,)\sigma^2_{\varepsilon^z} + 2l^{-1}\sigma_{\varepsilon,\varepsilon^z} - 2l^{-1}\sigma^2_{\varepsilon^z} + \sigma^2_{\varepsilon} - 2\sigma_{\varepsilon,\varepsilon^z} + \sigma^2_{\varepsilon^z} \\
&= \sigma^2_r(l) + \sigma^2_{\varepsilon'} \\
\mathrm{Cov}\left(\delta_{ijnm}, \delta_{i'j'n'm'} \middle| l\right) &= \sigma^2_{\upsilon}(l\,)\sigma^2_{\varepsilon^z} + 2l^{-1}\sigma_{\varepsilon,\varepsilon^z} - 2l^{-1}\sigma^2_{\varepsilon^z} \\
&= \left(\sigma^2_{\upsilon}(l\,) - 2l^{-1} + 2l^{-1}\rho\frac{\sigma_{\varepsilon}}{\sigma_{\varepsilon^z}}\right)\sigma^2_{\varepsilon^z} \\
&= \sigma^2_r(l)
\end{aligned} \tag{7}$$

where $i'j'n'm'$ is the index of a different measurement on gel $g_{ijnm}(k_{ijn})$ and ρ is the correlation coefficient between ε_{ijnm} and ε^z_{ijnm}. Using tabled values of $\sigma^2_{\upsilon}(l\,)$,[20] we found that $\sigma^2_r(8) = \left(0.123 + 0.250\rho\frac{\sigma_{\varepsilon}}{\sigma_{\varepsilon^z}}\right)\sigma^2_{\varepsilon^z}$, $\sigma^2_r(10) = \left(0.144 + 0.200\rho\frac{\sigma_{\varepsilon}}{\sigma_{\varepsilon^z}}\right)\sigma^2_{\varepsilon^z}$, and $\sigma^2_r(12) = \left(0.157 + 0.167\rho\frac{\sigma_{\varepsilon}}{\sigma_{\varepsilon^z}}\right)\sigma^2_{\varepsilon^z}$. If $\sigma^2_{\varepsilon^z}$ and σ^2_{ε} are similar in magnitude, which is a reasonable assumption since they reflect technical factors that should have similar multiplicative effects on measurements of all proteins in a lane, or the correlation between ε_{ijnm} and ε^z_{ijnm} is low to moderate, the scaling factor in brackets should be well approximated by a constant λ for the range of l that occurred in our experiments (Table S6). Based on this observation, we used the approximation

$\sigma_r^2(l) \approx \lambda\sigma_{\varepsilon^z}^2 = \sigma_r^2$ when l varied within experiments. Note that $\sigma_r^2(l) \equiv \sigma_r^2$ and no approximation is involved when l is the same for all gels in an experiment.

The same mean and covariance structure for δ_{ijnm} can also be obtained under a simplified hierarchical model:

$$\delta_{ijnm} = \upsilon\sigma_{\varepsilon^z} + \left(\upsilon\left(l_{g_{ijnm}(k_{ijn})}\right) - \upsilon\right)\sigma_{\varepsilon^z} + r_{g_{ijnm}(k_{ijn})} + \varepsilon'_{ijnm} \tag{8}$$

where υ is the expectation of $\upsilon\left(l_{g_{ijnm}(k_{ijn})}\right)$ over the distribution of $l_{g_{ijnm}(k_{ijn})}$ in the experiment, $r_{g_{ijnm}(k_{ijn})} \sim N(0, \sigma_r^2)$, and $\varepsilon'_{ijnm} \sim N\left(0, \sigma_{\varepsilon'}^2\right)$. In experiments with a variable number of lanes run per gel, $\left(\upsilon\left(l_{g_{ijnm}(k_{ijn})}\right) - \upsilon\right)\sigma_{\varepsilon^z}$ contributes to the gel-specific component of deviations from $\upsilon\sigma_{\varepsilon^z}$ for a given measurement. We therefore modeled the combined gel-level deviation $\left(\upsilon\left(l_{g_{ijnm}(k_{ijn})}\right) - \upsilon\right)\sigma_{\varepsilon^z} + r_{g_{ijnm}(k_{ijn})}$ with a single gel-specific random effect, which allowed us to accommodate variable or fixed numbers of lanes per gel in the same model. In particular, we assumed:

$$\delta_{ijnm} = \upsilon\sigma_{\varepsilon^z} + d'_{g_{ijnm}(k_{ijn})} + \varepsilon'_{ijnm} \tag{9}$$

where $d'_{g_{ijnm}(k_{ijn})} \sim N\left(0, \sigma_{d'}^2\right)$. Note that, when l is the same for all gels in an experiment, the second term in (8) is identically zero for all gels, so $d'_{g_{ijnm}(k_{ijn})} \equiv r_{g_{ijnm}(k_{ijn})}$.

Combining (9) with (6) and collapsing the two gel-specific terms that are not separately identifiable yields:

$$\ln \tilde{Y}_{ijnm} = \mu^*_{ij} + b_{k_{ijn}} + d^*_{g_{ijnm}(k_{ijn})} + e_{w_{ijn}(k_{ijn})} + \varepsilon'_{ijnm} \tag{10}$$

where $\mu^*_{ij} = \mu_{ij} + \alpha + \upsilon\sigma_{\varepsilon^z}$ and $d^*_{g_{ijnm}(k_{ijn})} = d_{g_{ijnm}(k_{ijn})} + d'_{g_{ijnm}(k_{ijn})} \sim N\left(0, \sigma_{d^*}^2 = \sigma_d^2 + \sigma_{d'}^2\right)$. Equation (10) defines a linear mixed model for the log-transformed loading control-normalized intensity measurements with mean μ^*_{ij} and variance components at the block (σ_b^2), gel within block ($\sigma_{d^*}^2$), experimental unit within block (σ_e^2), and measurement ($\sigma_{\varepsilon'}^2$) levels. The resulting model is similar to ones previously proposed for unnormalized measurements.[17] For designs in which all experimental units from a particular block are measured exactly once on a single gel, $d^*_{g_{ijn1}(k_{ijn})}$ and ε'_{ijn1} are indistinguishable from $b_{k_{ijn}}$ and $e_{w_{ijn}(k_{ijn})}$, respectively, and the model in (10) simplifies to:

$$\ln \tilde{Y}_{ijn} = \mu^*_{ij} + b^*_{k_{ijn}} + e^*_{ijn} \tag{11}$$

where $b^*_{k_{ijn}} = b_{k_{ijn}} + d^*_{g_{ijn1}(k_{ijn})} \sim N\left(0, \sigma^2_{b^*} = \sigma^2_b + \sigma^2_{d^*}\right)$ and $e^*_{ijn} = e_{w_{ijn}(k_{ijn})} + \varepsilon'_{ijn1} \sim N\left(0, \sigma^2_{e^*} = \sigma^2_e + \sigma^2_{\varepsilon'}\right)$. Here, the variance components are at the block ($\sigma^2_{b^*}$) and experimental unit ($\sigma^2_{e^*}$) levels.

The above models were fit via restricted maximum likelihood (REML) using PROC MIXED available in SAS/STAT 14.1 software, Version 9.4 (TS1M3) of the SAS System for 64-bit Windows (SAS Institute Inc., Cary, NC, USA). The primary targets of inference were $\theta_{i1j} = \exp\left(\mu^*_{ij} - \mu^*_{1j}\right) = \exp\left(\mu_{ij} - \mu_{1j}\right)$, which can be interpreted as the ratio of the median target protein abundance in 10 ug of lysate in variant condition $i > 1$ to that in WT ($i = 1$) within a given EGF stimulation condition j. Under the assumed model, this can also be interpreted as the ratio of marginal means because:

$$\theta_{i1j} = \exp\left(\mu_{ij} - \mu_{1j}\right) = \frac{\exp\left(\mu_{ij} + \frac{\sigma^2_b + \sigma^2_e}{2}\right)}{\exp\left(\mu_{1j} + \frac{\sigma^2_b + \sigma^2_e}{2}\right)} = \frac{E\left[X_{ijn}\right]}{E\left[X_{1jn}\right]}$$

Consistent point estimates of the contrasts $\mu_{ij} - \mu_{1j}$ were obtained by REML,[22, 23] and standard errors and degrees of freedom were determined by the Kenward-Roger (KR) adjustment to reduce small sample bias.[24] Pointwise 95% confidence intervals for each $\mu_{ij} - \mu_{1j}$ were calculated using the KR-adjusted standard errors and the *t* distribution with the KR-adjusted degrees of freedom. Point estimates and these confidence intervals were then transformed to consistent point estimates of and pointwise 95% confidence intervals for θ_{i1j} by exponentiation. Two-sided p-values under the null hypothesis that $\mu_{ij} - \mu_{1j} = 0$ (equivalently $\theta_{i1j} = 1$) were calculated by comparing the t statistic using the KR-adjusted standard error to the *t* distribution with the KR-adjusted degrees of freedom.

Our scientific hypotheses translated to independent statements about each θ_{i1j} for each protein rather than to joint statements about vectors of these parameters. In such inferential settings, simultaneous confidence intervals and multiplicity-adjusted p-values are unwarranted because the per-comparison confidence levels and p-values provide the relevant quantification of statistical uncertainty for each statement and, therefore, scientific hypothesis.[25-28] We, therefore, based our inferences on pointwise confidence intervals and unadjusted p-values for each θ_{i1j} for each protein.

Diagnostics were reviewed to evaluate overall model fit. For all models, the fitting algorithm converged with all variance components greater than zero, suggesting that the model was not overspecified. Cholesky residuals[29, 30] are equivalent to the residuals from a standard linear regression

using the design matrix and outcome variable vector scaled by the inverse Cholesky decomposition of the estimated covariance matrix.[31] If the vector of log-transformed outcome measurements has the multivariate normal distribution with mean vector and covariance matrix implied by (10) or (11) and the estimated covariance matrix used in the Cholesky decomposition is close to the true value, then, conditional on the realized value of this estimated covariance matrix, these residuals will behave like residuals from a standard regression model and be approximately normally distributed with mean zero and covariance matrix given by $\mathbf{I} - \mathbf{H}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{H}$ is the hat matrix for the transformed design matrix. This covariance matrix will approach $\mathbf{I}$ as the sample size grows large.[32, 33] We therefore examined quantile-quantile plots of Cholesky residuals against a fitted normal distribution with mean and variance estimated from the residuals to evaluate the overall tenability of the normality assumption and assumed model. These did not show any systematic deviation from the fitted normal distribution aside from a handful of outliers discussed further below.

Outliers were identified by an internally studentized marginal or conditional residual greater than $\sqrt{\chi_1^2(1-0.05/n)}$, where n is the sample size.[30, 34, 35] Single-case deletion influence diagnostics with a maximum of 5 refitting iterations[30] were used to identify influential observations having restricted likelihood distance greater than $\chi^2_{p+q}(0.5)$,[30, 32] Cook's D for μ_{ij} greater than $F_{p,n-p}(0.5)$,[30, 32] or Cook's D for covariance parameters greater than $\chi^2_q(0.5)$,[36] where p denotes the number of fixed-effect parameters and q denotes the number of covariance parameters in the appropriate model for the experiment and outcome (see Table S6). We identified a total of 6 outlying observations, 3 of which were influential, and a single influential observation that was a second measurement on the same experimental unit for which the other measurement was an outlier (Table S7). Importantly, outliers identified in the ERK experiments were not also influential based on the diagnostic measures described above. For the AKT and JNK experiments, the 3 outliers had minimal influence on the estimated μ_{ij} for SOS1, pAKT, and pJNK, although they did reduce the estimated precision of the μ_{ij} through their influence on covariance parameter estimates. Thus, 95% confidence intervals might be narrower in analyses excluding these observations. Nonetheless, because all 7 observations were verified to be correct and there were no technical issues that would warrant their exclusion, all were retained in the final analysis.

Figure S1. Impact of DCM-associated *SOS1* variants on AKT signaling. HEK293T cells expressing variant or wild-type (WT) *SOS1* were serum starved for 16 hours and alternatively stimulated for 15 minutes with epidermal growth factor (EGF) (15) or left unstimulated (0). (A). Representative immunoblots for each *SOS1* variant are shown. Total protein isolates were probed using antibodies specific to phosphorylated AKT (pAKT, Cell Signaling #8200, 1:5,000), total AKT (AKT, Cell Signaling #4691, 1:5,000) and SOS1 (Abcam ab140621, 1:10,000). Levels of β-tubulin were used to control for loading (Abcam, ab6046, 1:20,000). Note that pAKT and total AKT levels were determined using separate blots generated from equal quantities of the same protein lysate. For figure clarity, SOS1 and β-tubulin are shown for only the pAKT blot. Full blots can be found in Figure S3. (B) Differences in AKT activation between each variant and WT *SOS1* are shown as estimated ratios of median protein expression (variant/WT) at baseline (no-EGF) and 15 min EGF stimulation. Bars represent point estimates of this ratio from a linear mixed model fit to data from three independent experimental replicates for each variant, and error bars represent pointwise 95% confidence intervals for this ratio. *, **, and *** denote two-sided $p \leq 0.05$, 0.01, and 0.001, respectively, for the null hypothesis that this ratio was 1. Relative to WT *SOS1*, observed differences in pAKT levels were small for all tested *SOS1* variants and did not rule out the possibility of no difference in any case. Abbreviations: UT = untransfected, Vector/EV = vector only, WT = wild-type.

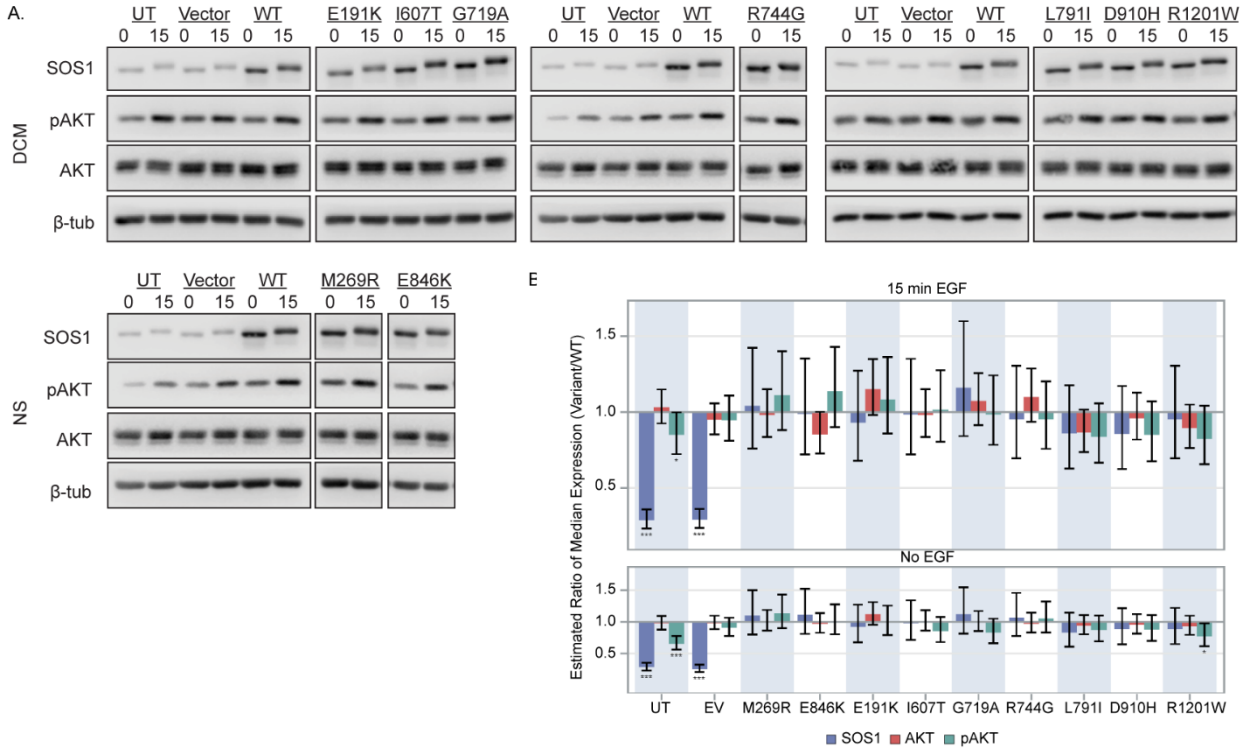

Figure S2. Impact of DCM-associated *SOS1* variants on JNK activation. HEK293T cells expressing variant or wild-type (WT) *SOS1* were serum starved for 16 hours and alternatively stimulated for 15 minutes with epidermal growth factor (EGF) (15) or left unstimulated (0). (A). Representative immunoblots for each *SOS1* variant are shown. Total protein isolates were probed using antibodies specific to phosphorylated JNK (pJNK, Cell Signaling #4668, 1:1,000), total JNK (JNK, Cell Signaling #9252, 1:1,000), and SOS1 (Abcam ab140621, 1:10,000). Levels of α-actinin were used to control for loading (Cell Signaling # 6487, 1:10,000). Note that pJNK and total JNK levels were determined using separate blots generated from equal quantities of the same protein lysate. For figure clarity, SOS1 and α-actinin are shown for only the pJNK blot. Full blots can be found in Figure S3. (B) Differences in JNK activation between each variant and WT *SOS1* are shown as estimated ratios of median protein expression (variant/WT) at baseline (no-EGF) and 15 min EGF stimulation. Bars represent point estimates of this ratio from a linear mixed model fit to data from three independent experimental replicates for each variant, and error bars represent pointwise 95% confidence intervals for this ratio. End caps are not displayed on error bars for SOS1 for UT and EV at baseline to avoid overplotting for these narrow 95% confidence intervals. *, **, and *** denote two-sided p ≤ 0.05, 0.01, and 0.001, respectively, for the null hypothesis that this ratio was 1. Relative to WT *SOS1*, pJNK levels were lower for most variants, although the data ruled out the possibility of no difference only for D910H. Abbreviations: UT = untransfected, Vector/EV = vector only, WT = wild-type.

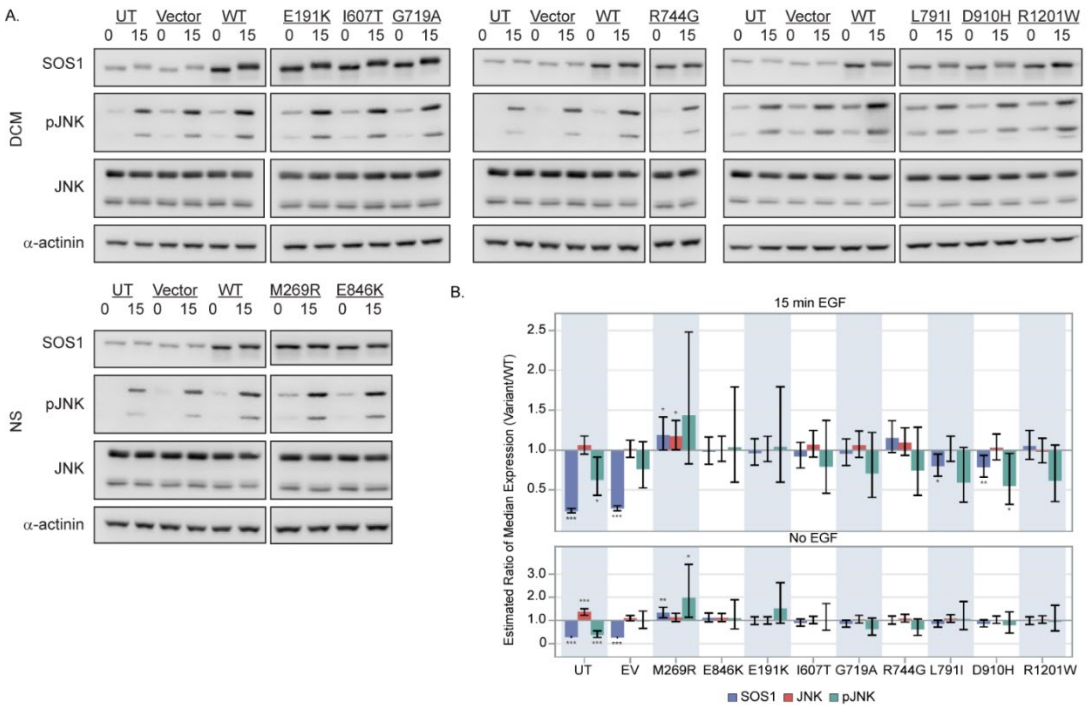

Figure S3. Full immunoblots for (A) ERK (B) AKT and (C) JNK activation experiments depicted in Figure 3, Figure S1, and Figure S2.

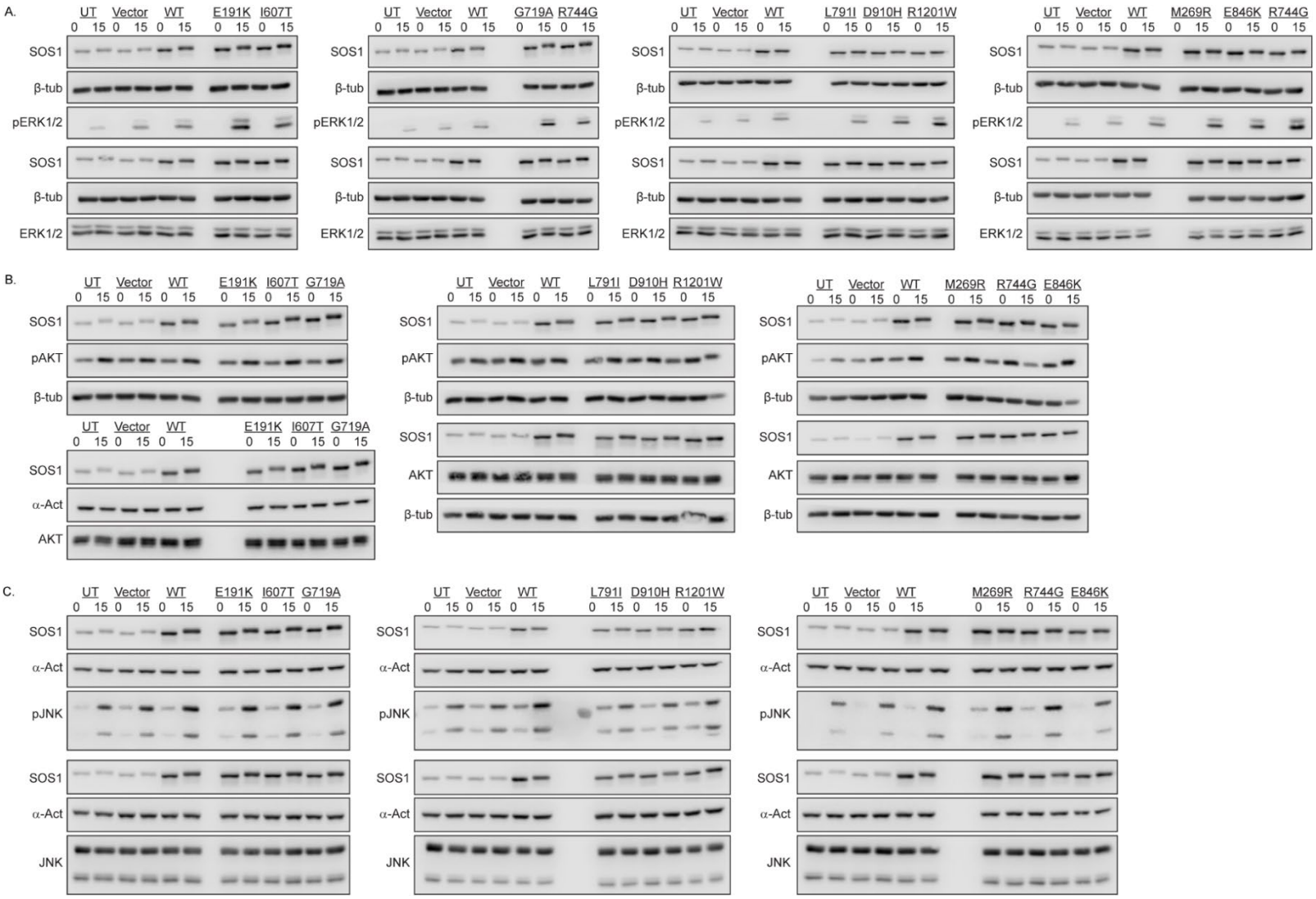

Figure S4. Sanger sequence confirmation of DCM-associated variants identified in Pedigrees A-E.

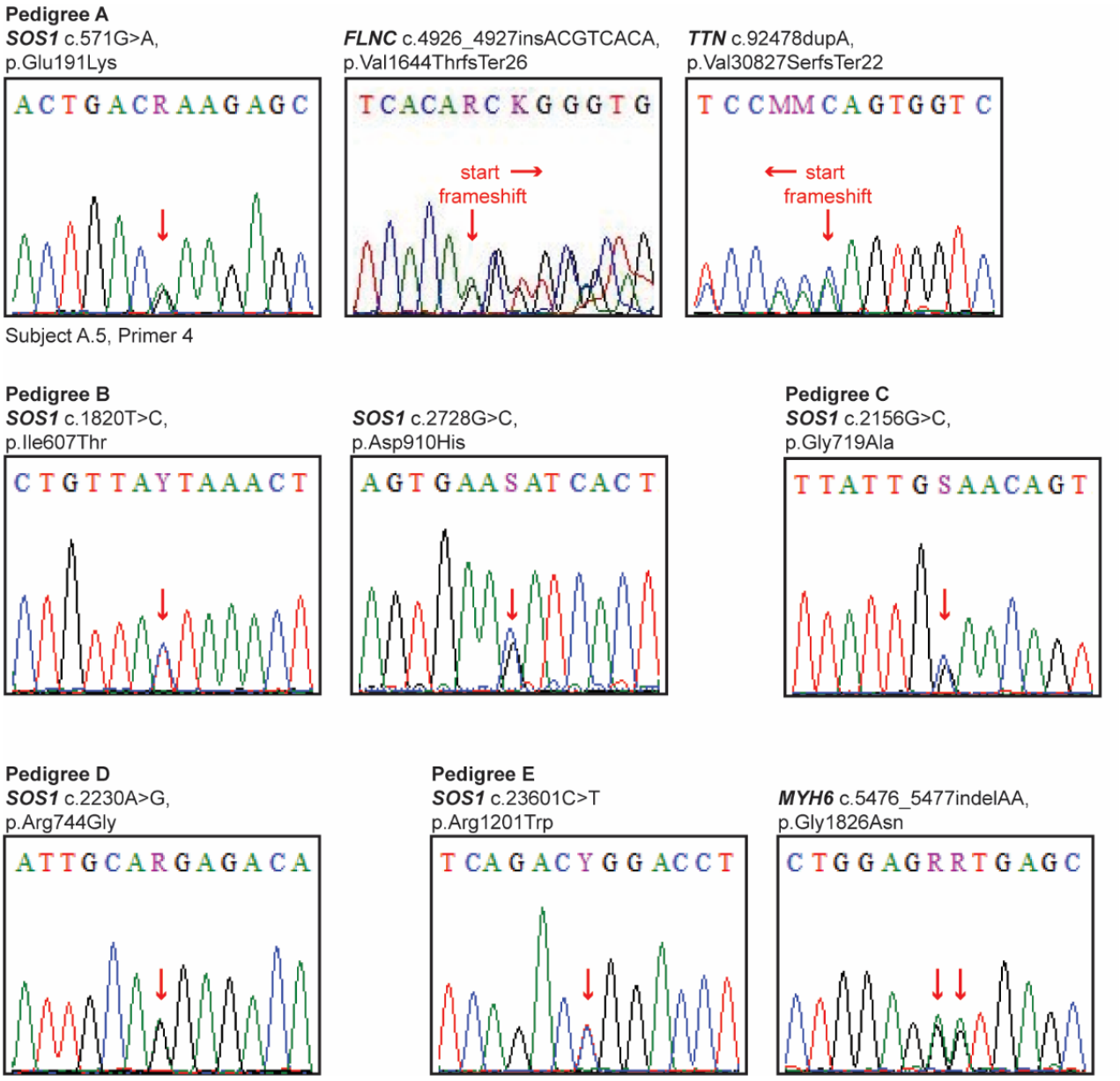

Table S1. Functional impact of reported *SOS1* variants on RAS signaling

Domain	Amino Acid Change	Associated Phenotype	Functional Result	Reference
Histone Fold (HF)	p.Thr158Ala	Cardio-facio-cutaneous Syndrome	Elevated pAKT	[37, 38]
	p.Lys170Glu	Noonan Syndrome	Elevated pERK	[39]
	p.Glu191Lys	Dilated Cardiomyopathy	Elevated pERK	This Study
Dbl Homology (DH)	p.Ile252Thr	Noonan Syndrome	No Change	[39]
	p.Met269Arg	Noonan Syndrome	Elevated pERK	This Study, [11, 12]
	p.Asp309Tyr	Noonan Syndrome	Elevated pERK	[11]
	p.Tyr337Cys	Noonan Syndrome	Elevated pERK	[39]
Pleckstrin Homology (PH)	p. Gly434Arg	Noonan Syndrome	Elevated pERK	[39]
	p.Cys441Tyr	Noonan Syndrome	No Change	[39]
	p.Arg497Gln	Noonan Syndrome	Reduced pERK	[12]
Helical Linker (HL)	p.Ser548Arg	Noonan Syndrome	No Change	[39]
	p.Leu550Pro	Noonan Syndrome	No Change	[39]
	p.Arg552Gly	Noonan Syndrome	Elevated pERK	[11, 40]
Ras Exchanger Motif (REM)	p.Ile607Thr	Dilated Cardiomyopathy	Elevated pERK	This Study
	p.Pro655Leu	Noonan Syndrome	Reduced pERK	[39]
	p.Arg688Ala + p.Leu687Glu	N/A	Reduced pERK	[41]
	p.Gly719Ala	Dilated Cardiomyopathy	Elevated pERK	This Study
	p.Trp729Leu	Noonan Syndrome	Elevated pERK	[40]
	p.Trp729Glu	N/A	Reduced pERK	[41]
	p.Ile733Phe	Noonan Syndrome	Elevated pERK	[39]
	p.Arg744Gly	Dilated Cardiomyopathy	Elevated pERK	This Study
CDC25	p.Leu791Ile	Dilated Cardiomyopathy	No Change	This Study
	p.Glu846Lys	Noonan Syndrome	Elevated pERK	[11]
	p.Pro894Arg	Noonan Syndrome	No Change	[39]
	p.Asp910His	Dilated Cardiomyopathy	Elevated pERK	This Study
	p.Gln977Arg	Noonan Syndrome	Reduced pERK	[39]
Proline-rich (PR)	p.Arg1201Trp	Dilated Cardiomyopathy	Elevated pERK	This Study

Table S2. *SOS1* variants reported in patients with isolated DCM

Domain	Nucleotide Change	Amino Acid Change	Variant Classification	Reference
Dextrin Homology (DH)	c.1135G>T	p.Ala379Ser	VUS	[42]
CDC25	c.2351T>C	p.Ile784Thr	VUS	[42]
CDC25	c.2762C>G	p.Ser921Cys	VUS	[43]
Proline-rich (PR) tail	c.3568A>G	p.Lys1190Glu	VUS	[43]

Table S3. DCM genes included in exome sequencing analysis.

ABCC9	*MYPN*
ACTC1	*NEBL*
ACTN2	*NEXN*
ANKRD1	*PDLIM3*
BAG3	*PKP2*
CRYAB	*PLN*
CSRP3	*PSEN1*
DES	*PSEN2*
DSG2	*RBM20*
DSP	*SCN5A*
EYA4	*SGCD*
FLNC	*TCAP*
ILK	*TNNC1*
LAMA4	*TNNI3*
LDB3	*TNNT2*
LMNA	*TPM1*
MYBPC3	*TTN*
MYH6	*VCL*
MYH7	

Table S4. Primer sequences used in this study.

Coding Sequence Primers

Exon	Primer Sequence (5'-3')
SOS1	
3	F: AAGAGGAAGCGAAGAAACC
5	F: GAGCCCTTTGTCTCCAATTC
5	F: CAGAGAAAACCAAGTCAGGGA
5	R: TCATGCAAATTTCACAACACA
6	R: TGGCCCAGTAACTTTACAC
8	R: ATGGAGACAGTGGTAAACAG
9-10	R: ACCGACATGCAGATTCAC
10	F: GATGACACCAATGAATACAAGCA
10	R: CTAGGCAGCCTCATCTGCTC
10	R: GCCAAGTGACCTCATTTTCTC
10	F: CTTCCTGGTGCTAGCAATG
12	R: TGCGATCAGCTTCTGTTG
12	F: ATATACAGCCTGTGCAACTG
12	R: ACCCAGTGCCGACATACATT
13	F: AGGACCTTTCCCAAACAAGG
13	R: CAGGGTGACTTGAGCCATTT
14	F: AAGATTAGGATTGGGGACCG
14	R: AGAGGGACCAGGGGAGTAAA
15	R: GGGAGGTGAACTCTGAAATG
15	F: TTAGTTGGAAGTGTGTGGAC
16	R: CTCACCACAGCTACTCTTTC
17	F: ATTTGGGCGTTTCTGTTAGC
17	R: GAGTTTGGCTATGCCTCTGC
18	R: TGGTACTGCTGGATCTCTC
19	F: TCCAGCAGTACCAAAATCAGC
22	F: CACTGATGAAGTGCCTGTC
23	R: AGGAGGAGGTGGTGTAAAG
23	F: GCATGTTTGAAAACCCCAAC
23	R: GAACTCAGGAAGAATGGGCA
FLNC	
28	F: ACGTACACTGTGTCCTACC
30	R: GCTGTGTAGTAGATGTCAAAGG
TTN	
339	F: AATGTAGAGAGCCCGTCAACC
339	R: CACATGCCTCTGCATTCACC
MYH6	
36	F: TGTAACACCAAGCCAACTCTGCAG
36	R: ACTCAAGTCAAACTGACTGCAGAGCGTGA

Primers for site-directed mutagenesis of *SOS1* expression constructs.

Nucleotide Change	Amino Acid Change	Primer Sequence (5'-3')*
c.571G>A	p.(Glu191Lys)	F: CTCCTGAGGTGGAAGGCTCTT**T**GTCAGTTAAAGATAATATATT
		R: AATATATTATCTTTAACTGAC**A**AAGAGCCTTCCACCTCAGGAG
c.806T>G	p.(Met269Arg)	F: GGGACTGCCTTCATCTGTC**C**TTTCTACTGTATCTTCTATA
		R: TATAGAAGATACAGTAGAAA**G**GACAGATGAAGGCAGTCCC
c.1820T>C	p.(Ile607Thr)	F: ACGTAAGCCTCTCTATAAGTTTA**G**TAACAGTTCCTGCTTTGATAATT
		R: AATTATCAAAGCAGGAACTGTTA**C**TAAACTTATAGAGAGGCTTACGT
c.2156G>C	p.(Gly719Ala)	F: CATTGCTTTACCTCTTACTGTT**G**CAATAAATTCTTCCATTCGTTGC
		R: GCAACGAATGGAAGAATTTATTG**C**AACAGTAAGAGGTAAAGCAATG
c.2230A>G	p.(Arg744Gly)	F: GACCTGGTCCATTGTCTC**C**TGCAATTTTTTTCCTTTGGATTA
		R: TAATCCAAAGGAAAAAAATTGCA**G**GAGACAATGGACCAGGTC
c.2371C>A	p.(Leu791Ile)	F: TATAGATCTGATTCAA**T**TAAAGTGAGTTGTCGAGCAATTTCTATTGG
		R: CCAATAGAAATTGCTCGACAACTCACTTTA**A**TTGAATCAGATCTATA
c.2536G>A	p.(Glu846Lys)	F: CCACAGCTACTCTTTCTT**T**TAAATTTTCAGTTTCTACAATACATTTCTCAAAC
		R: GTTTGAGAAATGTATTGTAGAAACTGAAAATTTA**A**AAGAAAGAGTAGCTGTGG
c.2728G>C	p.(Asp910His)	F: CAAATATTTCTTATAGTGAT**G**TTCACTCAATTCATGAGCTTCTTCTAAAATTTTCTTC
		R: GAAGAAAATTTTAGAAGAAGCTCATGAATTGAGTGAA**C**ATCACTATAAGAAATATTTG
c.3601C>T	p.(Arg1201Trp)	F: GGTCTGAGATAGAGGTCC**A**GTCTGATATTGAATATCG
		R: CGATATTCAATATCAGAC**T**GGACCTCTATCTCAGACC

*Variant nucleotide is bolded and underlined in each primer sequence.

Table S5. Conservation and pathogenicity predictions for identified variants.

				Conservation				Pathogenicity*									
Pedigree	Gene	Nucleotide Change	Amino Acid Change	Phast Cons (100 way)	phylo P (100 way)	GERP++ (NR/RS)	Mut-assessor	Mut-taster	MetaIr	FATHMM	Polyphen2 (HDIV)	Provean	LRT	SIFT	Metasvm	CADD	REVEL
A	*SOS1*	c.571G>A	p.(Glu191Lys)	1.0000	7.2280	5.88/5.88	Low	+	+	-	+	-	+	-	+	24.9	0.534
	FLNC	c.4926_4927insACGTCACA	p.(Val1643Thrfs*26)	0.9380	5.1490	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	TTN	c.92478dupA	p.(Val30827Serfs*22)	0.9890	1.9180	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
B	*SOS1*	c.1820T>C	p.(Ile607Thr)	0.8670	5.1820	5.78/5.78	Neutral	+	-	-	-	-	-	-	-	16.5	0.057
		c.2728G>C	p.(Asp910His)	1.0000	7.3340	5.39/5.39	Medium	+	-	-	+	+	+	+	-	31	0.394
C	*SOS1*	c.2156G>C	p.(Gly719Ala)	0.9990	2.7340	5.92/5.92	Neutral	+	-	-	-	-	+	-	-	15.9	0.103
D	*SOS1*	c.2230A>G	p.(Arg744Gly)	1.0000	5.5470	5.87/4.65	Neutral	+	-	-	-	-	+	-	-	20.2	0.06
E	*SOS1*	c.3601C>T	p.(Arg1201Trp)	1.0000	8.4900	5.8/5.8	Neutral	+	-	-	+	-	+	+	-	30	0.433
	MYH6	c.5476_5477delinsAA	p.(Gly1826Asn)	0	0.0101575	4.3/1.79	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

*+ denotes damaging - denotes tolerated/neutral/benign

Table S6. Relevant design characteristics of in vitro experiments.

Experiment	Blocks (K)	EUs per Block ($W_{k_{ijn}}$)	EUs per variant-EGF combination (N_{ij})	Target proteins	Gels per Block ($G_{k_{ijn}}$)	Lanes per Gel (l)	Measurements per EU (M_{ijn})	Total N	Model
ERK	12	12 (6 blocks) 14 (4 blocks) 16 (1 block) 18 (1 block)	12 (UT, EV, WT) 5 (variants)	ERK	1 (6 blocks) 2 (6 blocks)	12 (11 gels) 10 (5 gels) 8 (2 gels)	1 or 2 (UT, EV, WT) 1 (other variants)	198	(10)
				pERK	1 (6 blocks) 2 (6 blocks)	12 (10 gels) 10 (7 gels) 8 (1 gels)	1 or 2 (UT, EV, WT) 1 (other variants)	99*	(10)
				SOS1	2 (6 blocks) 4 (6 blocks)	12 (21 gels) 10 (12 gels) 8 (3 gels)	2 or 4 (UT, EV, WT) 2 (other variants)	396	(10)
AKT	9	12	9 (UT, EV, WT) 3 (variants)	AKT	1	12	1	108	(11)
				pAKT	1	12	1	106†	(11)
				SOS1	2	12	2	214‡	(10)
JNK	9	12	9 (UT, EV, WT) 3 (variants)	JNK	1	12	1	108	(11)
				pJNK	1	12	1	108	(11)
				SOS1	2	12	2	216	(10)

EGF = epidermal growth factor; EU = experimental unit; EV = empty vector; UT = untransfected; WT = wild-type.

* pAKT could not be measured under the no EGF condition.

† One UT/No EGF replicate and one UT/15 min EGF replicate were missing pAKT measurements on the same gel.

‡ One G719A/No EGF replicate and one G719A/15 min EGF replicate were missing SOS1 measurements on the same gel.

Table S7. Outlying and/or influential observations

Experiment	Target Protein	Variant/EGF	Block/Well/Gel	Residual Outlier?	Influential?
ERK	SOS1	R744G/15 min EGF	4/14/4	Conditional	-
	pERK	UT/15 min EGF	6/2/1	Conditional	-
		UT/15 min EGF	8/2/1	Conditional	-
AKT	SOS1	EV/No EGF	5/3/1	-	Covariance Parameters
			5/3/2	Conditional Marginal	Covariance Parameters
	pAKT	UT/No EGF	5/1/1	Conditional	Covariance Parameters
JNK	pJNK	UT/No EGF	6/1/1	Conditional	Covariance Parameters

EGF = epidermal growth factor; EV = empty vector; UT = untransfected

Supplemental References.

1. Cowan JR, Kinnamon DD, Morales A, Salyer L, Nickerson DA and Hershberger RE. Multigenic Disease and Bilineal Inheritance in Dilated Cardiomyopathy Is Illustrated in Nonsegregating LMNA Pedigrees. *Circ Genom Precis Med*. 2018;11:e002038.
2. Hershberger RE, Norton N, Morales A, Li D, Siegfried JD and Gonzalez-Quintana J. Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet*. 2010;3:155-61.
3. Li D, Parks SB, Kushner JD, Nauman D, Burgess D, Ludwigsen S, Partain J, Nixon RR, Allen CN, Irwin RP, et al. Mutations of presenilin genes in dilated cardiomyopathy and heart failure. *Am J Hum Genet*. 2006;79:1030-9.
4. Norton N, Li D, Rampersaud E, Morales A, Martin ER, Zuchner S, Guo S, Gonzalez M, Hedges DJ, Robertson PD, et al. Exome Sequencing and Genome-Wide Linkage Analysis in 17 Families Illustrates the Complex Contribution of TTN Truncating Variants to Dilated Cardiomyopathy. *Circ Cardiovasc Genet*. 2013;6:144-153.
5. Norton N, Li D, Reider MJ, Siegfried JD, Rampersaud E, Zuchner S, Mangos S, Gonzalez Quintana J, Wang L, McGee S, et al. Genome-wide studies of copy number variation and exome sequencing identify rare variants in *BAG3* as a cause of dilated cardiomyopathy. *Am J Hum Genet*. 2011;88:273-282.
6. Parks SB, Kushner JD, Nauman D, Burgess D, Ludwigsen S, Peterson A, Li D, Jakobs P, Litt M, Porter CB, et al. Lamin A/C mutation analysis in a cohort of 324 unrelated patients with idiopathic or familial dilated cardiomyopathy. *Am Heart J*. 2008;156:161-9.
7. Gonzalez M, Falk MJ, Gai X, Postrel R, Schule R and Zuchner S. Innovative Genomic Collaboration Using the GENESIS (GEM.app) Platform. *Hum Mutat*. 2015.
8. Norton N, Robertson PD, Rieder MJ, Zuchner S, Rampersaud E, Martin E, Li D, Nickerson DA and Hershberger RE. Evaluating Pathogenicity of Rare Variants From Dilated Cardiomyopathy in the Exome Era. *Circ Cardiovasc Genet*. 2012;5:167-174.
9. Morales A, Kinnamon DD, Jordan E, Platt J, Vatta M, Dorschner MO, Starkey CA, Mead JO, Ai T, Burke W, et al. Variant Interpretation for Dilated Cardiomyopathy: Refinement of the American College of Medical Genetics and Genomics/ClinGen Guidelines for the DCM Precision Medicine Study. *Circ Genom Precis Med*. 2020;13:e002480.
10. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405-23.
11. Roberts AE, Araki T, Swanson KD, Montgomery KT, Schiripo TA, Joshi VA, Li L, Yassin Y, Tamburino AM, Neel BG, et al. Germline gain-of-function mutations in SOS1 cause Noonan syndrome. *Nat Genet*. 2007;39:70-4.
12. Longoni M, Moncini S, Cisternino M, Morella IM, Ferraiuolo S, Russo S, Mannarino S, Brazzelli V, Coi P, Zippel R, et al. Noonan syndrome associated with both a new Jnk-activating familial SOS1 and a de novo RAF1 mutations. *Am J Med Genet A*. 2010;152A:2176-84.
13. Lepri FR, Scavelli R, Digilio MC, Gnazzo M, Grotta S, Dentici ML, Pisaneschi E, Sirleto P, Capolino R, Baban A, et al. Diagnosis of Noonan syndrome and related disorders using target next generation sequencing. *BMC Med Genet*. 2014;15:14.
14. Gureasko J, Kuchment O, Makino DL, Sondermann H, Bar-Sagi D and Kuriyan J. Role of the histone domain in the autoinhibition and activation of the Ras activator Son of Sevenless. *Proc Natl Acad Sci U S A*. 2010;107:3430-5.

15. Ackermann MA, King B, Lieberman NAP, Bobbili PJ, Rudloff M, Berndsen CE, Wright NT, Hecker PA and Kontrogianni-Konstantopoulos A. Novel obscurins mediate cardiomyocyte adhesion and size via the PI3K/AKT/mTOR signaling pathway. *J Mol Cell Cardiol*. 2017;111:27-39.
16. Rudloff MW, Woosley AN and Wright NT. Biophysical characterization of naturally occurring titin M10 mutations. *Protein Sci*. 2015;24:946-55.
17. Kreutz C, Bartolome Rodriguez MM, Maiwald T, Seidl M, Blum HE, Mohr L and Timmer J. An error model for protein quantification. *Bioinformatics*. 2007;23:2747-53.
18. Degasperi A, Birtwistle MR, Volinsky N, Rauch J, Kolch W and Kholodenko BN. Evaluating strategies to normalise biological replicates of Western blot data. *PLoS One*. 2014;9:e87293.
19. Tippett LHC. On the extreme individuals and the range of samples taken from a normal population. *Biometrika*. 1925;17:364-387.
20. Ruben H. On the Moments of Order Statistics in Samples from Normal Populations. *Biometrika*. 1954;41:200-227.
21. Liu JS. Siegel Formula Via Stein Identities. *Stat Probabil Lett*. 1994;21:247-251.
22. Demidenko E. Mixed models: Theory and applications with R. *Wiley series in probability and statistics 893*. 2013.
23. Richardson AM and Welsh AH. Asymptotic Properties of Restricted Maximum Likelihood (Reml) Estimates for Hierarchical Mixed Linear Models. *Australian Journal of Statistics*. 1994;36:31-43.
24. Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997;53:983-97.
25. Cox DR. A Remark on Multiple Comparison Methods. *Technometrics*. 1965;7:223-224.
26. Mayo DG and Cox DR. Frequentist statistics as a theory of inductive inference. In: J. Rojo, ed. *Optimality* Beachwood, Ohio, USA: Institute of Mathematical Statistics; 2006(Number 49): 77-97.
27. O'Brien PC. The appropriateness of analysis of variance and multiple-comparison procedures. *Biometrics*. 1983;39:787-94.
28. Cook RJ and Farewell VT. Multiplicity considerations in the design and analysis of clinical trials. *J Roy Stat Soc a Sta*. 1996;159:93-110.
29. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O and SAS Institute Inc. *SAS for Mixed Models.* 2 ed. Cary North Carolina: SAS Institute; 2006.
30. SAS Institute Inc. *SAS/STAT®14.1 User's Guide*. Cary, NC: SAS Institute Inc.; 2015.
31. Fitzmaurice GM, Laird NM and Ware JH. *Applied longitudinal analysis.* 2nd ed. Hoboken, N.J.: Wiley; 2011.
32. Cook RD and Weisberg S. *Residuals and Influence in Regression*: New York: Chapman and Hall; 1982.
33. Houseman EA, Ryan LM and Coull BA. Cholesky residuals for assessing normal errors in a linear model with correlated outcomes. *J Am Stat Assoc*. 2004;99:383-394.
34. Zewotir T and Galpin JS. A unified approach on residuals, leverages and outliers in the linear mixed model. *TEST*. 2007;16:58-75.
35. Grégoire TG, Schabenberger O and Barrett JP. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can J For Res*. 1995;25:137-156.
36. Christensen R, Pearson LM and Johnson W. Case-Deletion Diagnostics for Mixed Models. *Technometrics*. 1992;34:38.
37. Tumurkhuu M, Saitoh M, Sato A, Takahashi K, Mimaki M, Takita J, Takeshita K, Hama T, Oka A and Mizuguchi M. Comprehensive genetic analysis of overlapping syndromes of RAS/RAF/MEK/ERK pathway. *Pediatr Int*. 2010;52:557-62.
38. Tumurkhuu M, Saitoh M, Takita J, Mizuno Y and Mizuguchi M. A novel SOS1 mutation in Costello/CFC syndrome affects signaling in both RAS and PI3K pathways. *J Recept Signal Transduct Res*. 2013;33:124-8.

39. Smith MJ, Neel BG and Ikura M. NMR-based functional profiling of RASopathies and oncogenic RAS mutations. *Proc Natl Acad Sci U S A*. 2013;110:4574-9.
40. Tartaglia M, Pennacchio LA, Zhao C, Yadav KK, Fodale V, Sarkozy A, Pandit B, Oishi K, Martinelli S, Schackwitz W, et al. Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. *Nat Genet*. 2007;39:75-9.
41. Sondermann H, Soisson SM, Boykevisch S, Yang SS, Bar-Sagi D and Kuriyan J. Structural analysis of autoinhibition in the Ras activator Son of sevenless. *Cell*. 2004;119:393-405.
42. Aljeaid D, Sanchez AI, Wakefield E, Chadwell SE, Moore N, Prada CE and Zhang W. Prevalence of pathogenic and likely pathogenic variants in the RASopathy genes in patients who have had panel testing for cardiomyopathy. *Am J Med Genet A*. 2019;179:608-614.
43. Ceyhan-Birsoy O, Miatkowski MM, Hynes E, Funke BH and Mason-Suares H. NGS testing for cardiomyopathy: Utility of adding RASopathy-associated genes. *Hum Mutat*. 2018;39:954-958.