Supplemental Information

An activating germline \textit{IDH1} variant associated with a tumor entity characterized by unilateral and bilateral chondrosarcoma of the mastoid

Patrick R. Blackburn, Jodi M. Carter, Devin Oglesbee, Jennifer J. Westendorf, Brian A. Neff, Damian Stichel, David W. Tsen, Ralitza H. Gavrilova, Pieter Wesseling, Andreas von Deimling, Thomas R. Caulfield, Eric W. Klee, Stefan Pusch, and Carrie Y. Inwards
Supplemental Material

Table S1: Table of all variants meeting strict filtering criteria.

Table S2: Read mapping statistics and sample metrics.

Supplemental Methods

Whole exome sequencing

Ingenuity Variant Analysis version 5.3.20180727 was used for the analysis with the following content versions: CADD (v1.3), Allele Frequency Community (2018-05-25), EVS (ESP6500SI-V2), JASPAR (2013-11), Ingenuity Knowledge Base Snapshot Timestamp (2018-06-23 17:40:02.0), Vista Enhancer (2012-07), Clinical Trials (Rohan 180623.001), PolyPhen-2 (v2.2.2), 1000 Genome Frequency (phase3v5b), ExAC (0.3.1), iva (Jun 26 11:55 iva-1.0.476233.jar), PhyloP (2009-11), DbSNP (151), TargetScan (6.2), CentoMD (4.2), Ingenuity Knowledge Base (Rohan 180623.001), OMIM (May 26, 2017), gnomAD (2.0.1), BSIFT (2016-02-23), TCGA (2013-09-05), Clinvar (2018-04-06), DGV (2016-05-15), COSMIC (v84), HGMD (2018.1), SIFT4G (2016-02-23).

Variant annotation and analysis

Each variant in coding regions was functionally annotated by CAVA, as well as ClinVar, dbNSFP, OMIM, and the Human Gene Annotation Database to predict biological effects. Each variant was annotated with allele frequency from the Exome Aggregation Consortium (ExAC). Variants of significant interest were visually inspected using IGV, and then biologically validated using standard molecular biology techniques. The annotated variant call format file was loaded into the QIAGEN’s Ingenuity® Variant
Analysis™ software (www.qiagen.com/ingenuity) (QIAGEN, Redwood City) and causal variants were identified. Filter Description: Starting with 156,698 variants spanning 27,059 genes, variants were: kept with call quality at least 20.0 in cases or at least 20.0 in controls AND outside top 5.0% most exonically variable 100base windows in healthy public genomes (1000 genomes); excluded that are observed with an allele frequency greater than or equal to 0.01% of the general population ExAC Frequency (non-TCGA) OR greater than or equal to 0.01% of the general population gnomAD Frequency OR Filter variants unless established Pathogenic common variant; kept (up to 20 bases into intron) that are experimentally observed to be associated with a phenotype: Pathogenic, Possibly Pathogenic OR Disease-associated according to HGMD OR Frameshift, in-frame indel, or stop codon change OR Missense OR predicted to disrupt splicing by MaxEntScan OR within 4 bases into intron; kept which are associated with gain of function OR compound_heterozygous OR homozygous OR haploinsufficient OR heterozygous_alt OR hemizygous OR heterozygous_amb OR heterozygous AND occur in at least 3 of the case samples at the variant level in the Case samples AND not which are associated with gain of function OR compound_heterozygous OR homozygous OR haploinsufficient OR heterozygous_alt OR hemizygous OR heterozygous_amb OR heterozygous AND occur in at least 1 of the control samples at the variant level in the Control Samples.

Model generation and molecular dynamics simulations

The model refinement was built within Schrodinger for molecular checks (Schrödinger, LLC, New York, NY)\textsuperscript{9,10}, YASARA for simulated annealing and packing\textsuperscript{10-16}, and finally
imported into NAMD2 for simulations using generated PSF. The structures were adjusted with simulated annealing with explicit solvent, and small equilibration simulations using YASARA’s refinement protocol and prior to NAMD2 simulations with Amber force field. The final models were subjected to energy optimization with PR conjugate gradient with an R-dependent dielectric prior to heating protocols and equilibration runs with the NAMD2 engine. Each model was exported to the following formats: Maestro (MAE), NAMD (PDB and DCD) (Supplemental Material). Model manipulation was done with Maestro (Macromodel, version 9.8, Schrödinger, LLC, New York, NY, 2010), or Visual Molecular Dynamics (VMD).

MDS was completed on each model for conformational sampling, using methods previously described in the literature. Briefly, each IDH1 molecule was minimized with relaxed restraints using either Steepest Descent or Conjugate Gradient PR, and equilibrated in solvent with physiological salt conditions, as shown in the literature. Following equilibration each system was allowed to run MD calculations between ~0.2 microseconds in length. The protocol for refinement includes the following steps: (1) Simulated annealing with explicit water molecules and ions, (2) Energy minimization, (3) MDS for 25 ns to relax to the force field (both AMBER03). Production run simulations were conducted using NVE and NPT ensembles for comparison. Berendsen protocols were applied in the heating phase as described.

Charmm27, AMBER02, and OPLS2005 force fields were tested with the current release of NAnoscale Molecular Dynamics 2 engine, so that each force field can be examined for
comparison of generated stabilized, refined structure. The IDH1 wildtype with hydrogens consists of 3,381 atoms and the IDH1 R100Q mutant with hydrogens consists of 3,388. In all cases, we neutralized with counter-ions, and then created a solvent with 150 mM Na+ Cl- to recreate physiological strength. Essential K+ ions from the Qd central axis were included. TIP3P water molecules were added around the structure at a depth of >18 Å from the edge of the molecule depending upon the side. Our protocol has been previously described in the literature. Solvated protein simulations consist of a box from 4.83 x 10^4 to 5.31 x 10^4 atoms including protein, counter-ions, solvent ions, and solvent waters. Simulations were carried out using the particle mesh Ewald technique with repeating boundary conditions with a 9 Å nonbonded cut-off, using SHAKE with a 2-fs timestep. Pre-equilibration was started with 5,000 steps of minimization followed by 1000 ps of heating under MD, with the atomic positions of protein fixed. Then, two cycles of minimization (5000 steps each) and heating (2000 ps) were carried out with restraints of 10 and 5 kcal/(mol·Å^2) applied to all protein atoms. Next, 5000-steps of minimization were performed with solute restraints reduced by 1 kcal/(mol·Å^2). Then, 1000 ps of unrestrained MD were carried out, and the system was slowly heated from 1 to 310 K. The production MD runs were carried out with constant pressure boundary conditions (relaxation time of 1.0 ps). A constant temperature of 300 K was maintained using the Berendsen weak-coupling algorithm with a time constant of 1.0 ps. SHAKE constraints were applied to all hydrogens to eliminate X-H vibrations, which yielded a longer simulation time step (2 fs). Our methods for equilibration and production run protocols are in the literature. Equilibration was determined from a flattening of RMSD over time after an interval of >25ns, whereupon long-timescale MDS was allowed.
to start (>18 microseconds). Translational and rotational center-of-mass motions were initially removed. Periodically, simulations were interrupted to have the center-of-mass removed again by a subtraction of velocities to account for the “flying ice-cube” effect. Following the simulation, the individual frames were superposed back to the origin, to remove rotation and translation effects.

*Biochemical testing*

Organic acids (including 2-hydroxyglutarate, 2-oxoglutarate, and citrate) were prepared by stable isotope dilution for capillary gas chromatography-mass spectrometry (GC-MS) analysis as oximated, silyl- derivatives. Briefly, 100 µL of each plasma sample was diluted 1:1 into isotopic internal standard mix (containing D⁴-citrate, D⁴-lactate, ¹³C₄-2-oxoglutarate) and followed by oximation with pentafluorobenzyl hydroxylamine. Organic oximes were extracted with ethyl acetate and derivatized with N,O₂-bis-(trimethylsilyl) trifluoroacetamide and 1% trimethylchlorosilane (BSTFA+TMCS) to form silyl-derivatives and 1 µL was injected into an Agilent 6890/5973 GC-MS using a splitless/split combination (50:1 split mode). The GC-MS system was operated under the control of Agilent MSD Chemstation software running on a Microsoft Windows 7 Professional workstation. Quantification of analytes was performed by comparing measured peak areas of quantifying ions on the GC-MS against unique quantifying ions of known concentrations of isotope-labeled internal standards. For urine organic acids, 0.25 mg of creatinine urine equivalent (not in excess of 3 mL of urine) with a known concentration of pentadecanoic acid (PDA, internal standard) was acidified and extracted with ethyl acetate. After evaporation, dry residue was silylated with N,O₂-bis-
(trimethylsilyl) trifluoroacetamide and 1% trimethylchlorosilane (BSTFA+TMCS) and analyzed by a GC/MS method as previously described.\textsuperscript{27} Six point calibration curves were run for each quantified analyte each time a new PDA internal standard was prepared and renewed at least every six months. Plasma free fatty acids (i.e. phytanic acid) were quantified by stable isotope dilution for GC/MS analysis as previously described.\textsuperscript{28}

\textit{p.R100Q IDH1 mutant astrocytoma}

The 41-year-old patient underwent a brain biopsy in 2000 and received a histological diagnosis of low-grade diffuse astrocytoma. In 2004, the patient returned with progressive disease and received 12 cycles of temozolomide (TMZ), which led to a period of stable disease. In 2006, the patient presented again with progressive disease and received another 6 cycles of TMZ, but discontinued treatment because of clinical and radiological progression. In 2006 the patient was admitted to hospital with embolus in lung-artery and died early in 2007. No other evidence of malignancy was mentioned in this patient's clinical history.

References

1. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. \textit{Fly (Austin)} 2012;6:80–92.

2. Münz M, Ruark E, Renwick A, Ramsay E, Clarke M, Mahamdallie S, Cloke V, Seal S, Strydom A, Lunter G, Rahman N. CSN and CAVA: variant annotation
tools for rapid, robust next-generation sequencing analysis in the clinical setting. *Genome Med* 2015;7:1–8.

3. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42:D980–5.

4. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 2011;32:894–9.

5. Tryggvadóttir L, Vidarsdóttir L, Thorgeirsson T, Jonasson JG, Ólafsdóttir EJ, Ólafsdóttir GH, Rafnar T, Thorlacius S, Jonsson E, Eyfjord JE, Tulinius H. Prostate cancer progression and survival in BRCA2 mutation carriers. *J Natl Cancer Inst* 2007;99:929–35.

6. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeysinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;21:577–81.

7. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, Beltran H, Abida W, et al. Integrative clinical genomics of advanced prostate cancer. *Cell* 2015;161:1215–28.

8. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.

9. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology* 2002;320:597–608.
10. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 2009;77 Suppl 9:114–22.

11. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.

12. Krieger E, Dunbrack RL, Hooft RWW, Krieger B. Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods Mol Biol* 2012;819:405–21.

13. Hooft RWW, Sander C, Scharf M, Vriend G. The PDBFINDER database: A summary of PDB, DSSP and HSSP information with added value. *Comput Appl Biosci* 1996;12:525–9.

14. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature* 1996;381:272–2.

15. Qiu J, Elber R. SSALN: An alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. *Proteins* 2006;62:881–91.

16. King RD, Sternberg MJ. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci* 1996;5:2298–310.

17. Laskowski RA, Macarthur MW, Moss DS. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied ...* 1993;
18. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14:33–8–27–8.

19. Caulfield TR, Devkota B, Rollins GC. Examinations of tRNA Range of Motion Using Simulations of Cryo-EM Microscopy and X-Ray Data. *J Biophys* 2011;2011:219515–5.

20. Caulfield T, Devkota B. Motion of transfer RNA from the A/T state into the A-site using docking and simulations. *Proteins* 2012;80:2489–500.

21. Caulfield T, Medina-Franco JL. Molecular dynamics simulations of human DNA methyltransferase 3B with selective inhibitor nanaomycin A. *J Struct Biol* 2011;176:185–91.

22. Caulfield TR. Inter-ring rotation of apolipoprotein A-I protein monomers for the double-belt model using biased molecular dynamics. *J Mol Graph Model* 2011;29:1006–14.

23. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 1983;79:926–35.

24. Réblová K, Lankaš F, Rázga F, Krasovska MV, Koča J, Šponer J. Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers* 2006;82:504–20.

25. Réblová K, Fadrná E, Sarzynska J, Kulinski T, Kulhánek P, Ennifar E, Koča J, Šponer J. Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics. *Biophys J* 2007;93:3932–49.

26. Cheatham TE, Young MA. Molecular dynamics simulation of nucleic acids:
Successes, limitations, and promise. *Biopolymers* 2000;56:232–56.

27. Rinaldo P. Organic Acids. In: Blau N, Duran M, Gibson KM, eds. Laboratory Guide to the Methods in Biochemical Genetics. Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2008. 137–69.

28. Lagerstedt SA, Hinrichs DR, Batt SM, Magera MJ, Rinaldo P, McConnell JP. Quantitative determination of plasma C8-C26 total fatty acids for the biochemical diagnosis of nutritional and metabolic disorders. *Molecular Genetics and Metabolism* 2001;73:38–45.