A. Research Support Available

The PI has support available for Dr. Berg from two sources, as follows:

Source	Title	Dates	Amount
R01 GM115889 (role: PI)	The Population Genetics of Disease Risk and Other Quantitative Traits	7/2015-8/2020	$198,000 direct costs per year
Start-up funds, Columbia University	N/A	No time limit	$497,000

B. Sponsor's/Co-Sponsor's Previous Fellows/Trainees

The PI has supervised six pre-doctoral trainees, five for their M.Sc. and one for a Ph.D., and one postdoc trainee, not including those currently in his group. All of these trainees completed their training at the Hebrew University in Israel, before the PI moved to Columbia 3.5 years ago. Unlike in US institutions, completing a M.Sc. is the standard in Israel, and students generally do their postdocs in the US and Europe. Two of the students received Israel's most prestigious fellowships (Klore and Horowitz). Two have continued to pursue an academic route, while the rest have taken enviable positions in industry. Below are five representative examples:

Name	Position in the lab	Dates	Current title and position
Dr. Eyal Elyashiv	Ph.D. student	10/2009-12/2015	Bioinformatics Scientist, Ancestry.com
Dr. Samuel Sattath	Postdoc	10/2006-9/2013	Chairman and Founder of Mobeeart.com
Arbel Harpak	M.Sc. student	10/2011-10/2013	Ph.D. student with Jonathan Pritchard at Stanford
Dr. Oren Kolodny	M.Sc. student	10/2008-9/2010	Postdoc with Marc Feldman at Stanford
Etam Benger	M.Sc. student	10/2007-9/2009	Head of research at Taykey Inc.

C. Training Plan, Environment, Research Facilities

Environment. The PI is an Associate Professor in the Departments of Biological Sciences and of Systems Biology at Columbia University. His group currently consists of four graduate students and one postdoc from diverse backgrounds, including Mathematics, Statistics, Physics and Biology, making for a stimulating intellectual environment. Much of the work in the lab is related to the topic of the proposed research (but not redundant with it), including work on the maintenance of genetic variation and response to directional selection in continuous quantitative traits and work aimed at quantifying positive and negative selection in the human genome (see the biosketch and below).

Beyond the immediate group, there is a strong community of researchers working on related questions. The Sella lab space is contiguous to that of Molly Przeworski, a population geneticist who studies mutation, recombination and selection in humans and other taxa, and who currently has eight people in her group. In addition, Joe Pickrell (New York Genome Center/Columbia), a human geneticist whose work includes novel statistical analyses of GWAS data and whose lab members spend a day a week in the Sella lab. The three groups share a weekly journal club and lab meeting attended by all labs (see: http://przeworski.c2b2.columbia.edu/index.php/journal-club/). Journal club typically focus on one or two papers, where the discussion is led by one of the participants (in rotation). About a third of the lab meetings are given by outside speakers, who typically stay around to meet with students, providing wonderful opportunities for informal interactions with both junior and senior researchers. In the rest of the meetings, students and postdocs present ongoing work, with considerable time devoted to discussion. Occasionally, lab meetings are also used to practice talks for conferences and seminars, giving Dr. Berg the opportunity to receive extensive feedback from faculty and peers. In addition, the Sella group benefits from close ties with Itzik Pe'er's group in Computer Science, who also works on questions in human population genetics and runs a monthly seminar series in Statistical Genetics, as well as from interactions with computational biologist Harmen Bussmaker's, whose lab is across the hall. Lab members also regularly attend seminars at the New York Genome Center, in the department of Systems Biology in the medical school, and in the Biological Sciences department. Thus, Dr. Berg benefits from a rich and stellar intellectual environment.

Facilities. Dr. Berg has all the practical resources needed for the proposed project. The Sella lab is housed within a newly renovated computational space consisting of the office of the PI, three student offices, one of which Dr. Berg shares with one more student, as well as a lounge area, a kitchen area, and a conference room shared with Dr. Przeworski's group. In addition to being provided with a powerful laptop, and a printer in the lounge area, he has high-speed access to the powerful computer cluster of the Center for Computational Biology and Bioinformatics (http://www.c2b2.columbia.edu/index.php?q=node/6), on which the Sella lab has dedicated space.

Training plan. Dr. Berg already has a strong modeling and statistics toolbox, as well as more general background in population genetics that is unusual in both breadth and depth. His postdoctoral training will be used to further develop and hone his modeling skills, deepen his knowledge in human population genetics, as well as other, more general skills required for a successful academic career. Examples of the means by which this will be achieved include:

- He will hold weekly meetings with me to discuss ongoing research and the preparation of manuscripts, as well as issues related to career development (e.g., applying for jobs and funding).
- The communal area in the lab, alongside a weekly cookie hour, present ample opportunities for spontaneous scientific conversations with the PI and other lab members on an almost daily basis.
- Dr. Berg will attend at least two scientific meeting a year to present his work. In addition to holding practice talks and receiving feedback in lab meetings, I will provide extensive feedback throughout the stages of preparing these talks.
- Dr. Berg will present the results of his research in lab meetings twice a year, where he will receive extensive feedback from members of the Sella, Przeworski

and Pickrell labs.
- Dr. Berg will lead the discussion in our journal club 3-4 times a year.
- Dr. Berg have been supervising an undergrad research project. While I am involved and provide support when needed, it is quite clear that he is a natural and effective mentor.

D. Number of Fellows/Trainees to be Supervised During the Fellowship

During the 2017-2018 academic year, Dr. Sella anticipates training one postdoc in addition to Dr. Berg, and five graduate students.

E. Applicant's Qualifications and Potential for a Research Career

I first met Jeremy at the SMBE conference in Chicago in the summer of 2013 and have been following his work with great interest ever since. Jeremy completed his Ph.D. at UC Davis in one of the top population genetics programs in the country, where he worked with Graham Coop, who is one of the world's leading theoretical population geneticists. Before detailing his outstanding research achievements, I'd like to note that Jeremy's population genetics background is quite exceptional in both breadth and depth. Moreover, he is one of the very best population geneticists in his cohort and I have little doubt that he will have a stellar scientific career. I was therefore overjoyed when he joined my lab in November 2016.

When Jeremy just entered graduate school circa 2011, our view of the genetic basis of adaptation in humans was shifting. After a decade in which the study of adaptation was predicated on the classic sweep model, in which a beneficial mutation arises and is rapidly driven to fix in the population, multiple lines of evidence converged in suggesting that this mode of adaptation was rare in recent human evolution. The field then shifted its attention to plausible alternatives, most notably to adaptation from standing genetic variation (i.e., that was segregating in the population before it became beneficial). In particular, there were good reasons to believe that adaptation was often polygenic, i.e., involving small changes to allele frequencies at numerous loci affecting quantitative traits.

Jeremy's graduate work played a key role in developing the theory and statistical methodology to study these modes of adaptation. In one project, he used coalescent theory to characterize the footprints that adaptation from standing variation should leave in patterns of genetic variation (Berg & Coop *Genetics* 2015). Beyond being extremely timely, and clearly articulating when we should expect to see footprints of this mode of adaptation in population genetic data, this work also has the distinct quality of good theory–of taking a difficult question and providing a simple and intuitive answer. In a second project (Berg & Coop *PLoS Genetics* 2014), Jeremy developed statistical methods to detect adaptive changes to quantitative traits between human populations. This problem is challenging because adaptive changes to quantitative traits are expected to involve minute changes to allele frequency at many loci affecting the trait. To overcome this problem, Jeremy combined information from genome-wide association studies (GWAS) and genomic polymorphism data from multiple human populations in order to calculate the breeding value (the phenotypic value predicted based on the genotype) for traits in different populations. He then developed a statistical framework to test whether the observed differences between populations significantly exceed the expectation in the absence of selection. While the basic idea was also presented in a

couple of other studies, Jeremy's work stood out in being the most comprehensive and best grounded in population genetics theory. While this work has been incredibly influential, it still suffered from limited statistical power, leading it to detect polygenic selection in only a couple of traits. Jeremy's third project (the manuscript is to be submitted shortly) addresses this problem by extending the statistical methodology to incorporate GWAS results not only from loci that show significant associations but rather from all loci included in the study. Doing so is a statistical and population genetics feat, but it is worth the effort: with this extension, Jeremy uncovers evidence for polygenic adaptations in many traits, across many human populations. I have little doubt that this work will have an important and lasting impact in the field.

For his postdoctoral research, Jeremy proposes to address a complementary side of the population genetics of quantitative traits: to understand the processes that shape the genetic basis of complex (polygenic) disease susceptibility. While we know that the risk of developing most common diseases is highly heritable and polygenic, we have a poor understanding of the population genetics processes that account for their prevalence and shape their genetic architecture (e.g., of the frequencies and effect sizes of alleles that underlie risk). To address these important questions, Jeremy proposes to model how salient genetic factors and population genetics processes affect disease prevalence and architecture. In contrast to the little previous theory that addressed this question, Jeremy's point of departure is to build an explicit model relating an individual's genotype with his risk of developing the disease. His preliminary results already indicate that using such a generative modeling approach leads to results that contradict assumptions that are widely held in the field, e.g., that variants that affect diseases with a greater fitness cost would tend to be under stronger purifying selection. Importantly, the modeling he proposes will generate quantitative predictions that he will later test against the results from human GWAS. This research is expected to substantially advance our understanding of the population genetic processes that underlie complex disease. Beyond the obvious basic importance of such an understanding, it is also likely to inform us about questions of practical import, such as the design of future mapping studies and building models for phenotypic predictions (e.g., in personalized medicine).

As noted, Jeremy has an extremely strong background in population and evolutionary genetics in general and in doing population genetics theory and statistical analyses in particular. His postdoctoral training will focus on further developing his modeling and statistical tools and on deepening and broadening his knowledge of human population genetics and evolutionary quantitative genetics. My lab and the closely affiliated Przeworski and Pickrell labs (see above) provide an excellent environment in these regards. Specifically, my lab is focused on understanding the population genetics processes that shape the genetic architecture of continuous quantitative traits (e.g., traits like height, body mass index, age of menarche, etc.). The continuous quantitative traits we have studied are clearly distinct from the complex diseases Jeremy proposes to focus on (e.g., they tend to be under stabilizing rather than directional selection and are described well by additive models rather than by nonlinear functions of an underlying additive liability). Nonetheless, their study shares many common features and I therefore expect that some of the analytical and statistical machinery that we have developed will prove to be extremely useful to Jeremy. In addition, research in both the Przeworski and Pickrell labs addresses complementary aspect of the study of quantitative traits. More generally, the three labs cover a wide range of expertise in human population genetics. Taken together, the combination of the three labs is probably one of the best

environments in the country to perform the proposed research, and to learn more about theoretical and human population genetics in general.

The proposed research would consolidate Jeremy's position as one of the top experts in the population genetics of quantitative traits, an extremely strong launching point for an independent academic career. Not only has the evolution of quantitative traits been one of the central questions in population genetics pretty much since its conception, but the progress ushered by GWAS has also brought related questions to the forefront of human genetics. Moreover, the new wealth of empirical data about quantitative traits and complex diseases promises considerable progress in studying these questions in years to come, and the scope of this topic guarantees that important work is not going to be exhausted any time soon. I have little doubt that Jeremy will be a major contributor to this progress.