Watershed Informatics[data is inevitable]

From: Casey Caruso
To: TLC Investment Committee
Date: May 2nd, 2020
Re: Watershed Informatics
Note: This is a public-facing document with some information redacted. Design inspired by the incredible Ari Melenciano.

For more info: visit tlccollective.space

--> Deal Summary:

Watershed is raising a $550k pre-seed round at a $3m pre-money valuation.

Watershed was founded in August 2019 and is a Boston-based software development platform for biologists. The Watershed team deliberately filled the round with ~10 angels and 2 micro-funds, including Dorm Room Fund. This round should provide ~15 months of financing and enable the team to (1) create an alpha prototype and (2) secure commercial relationships. By EOY 2020, seven months from now, the goal is to have three organizations using Watershed and 15 total users.

Prior to this current fundraise, they have been bootstrapped. Casey was introduced to Watershed through a prior Bessemer intern and current MIT engineering student.

--> Team:

Jonathan Wang is the co-founder and CEO. Jonathan studied Computer Science and Electrical Engineering (M.S., M.Eng) at MIT. Prior to Watershed Jonathan built a petabyte-scale machine learning research platform for generating high-frequency trading strategies as the co-founder of Domeyard.

Mark Kalinich, PhD (part-time) is the co-founder and CSO. Mark studied Biochemical Engineering at MIT and is currently a Biophysics PhD student at Harvard University, focusing on genomic technologies for cancer detection. He is focusing on product, biz dev, and sales and is devoting 40+ hours a week while taking two medical school courses.

Alvin Wang (part-time) is the lead front-end developer. Alvin studied Information Systems and Human Computer Interaction at Carnegie Mellon University. Before Watershed Alvin was the lead developer for Google Hangouts Chat and is a co-creator of Materialize, a popular front-end framework. He is working 25 hours currently but planning on bringing him on as work scales up.

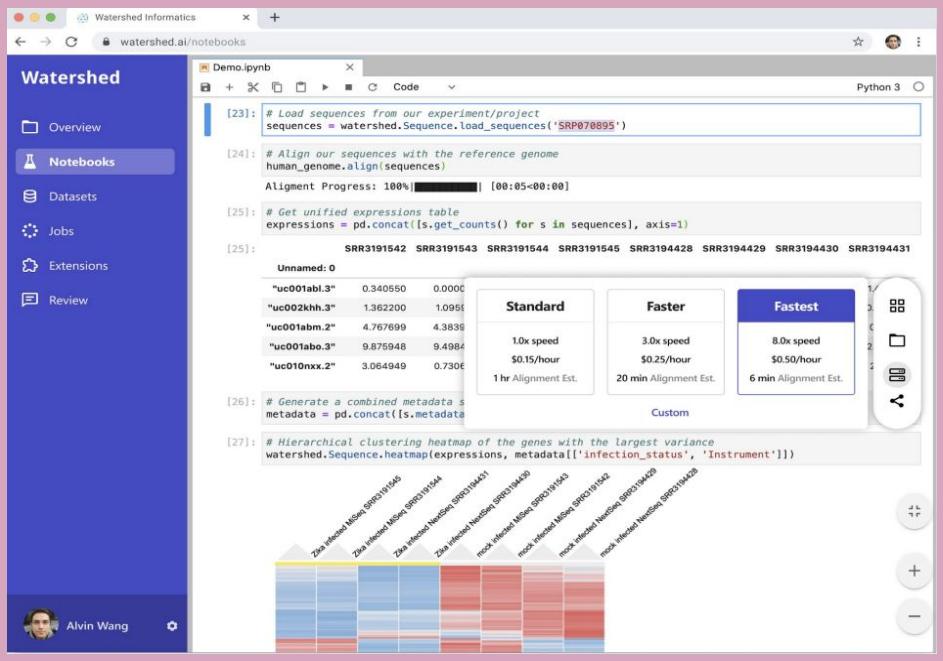
Michael Nackoul (part-time) is the VP of Business Operations. Michael studied Mechanical Engineering at MIT and received his M.S. in MechE and MBA from

University of Pittsburgh. Before Watershed Michael was a senior product manager at Microsoft's computing division. He will remain a part-time hire.

Advisors include Frank Moss (previous director of MIT Media Lab), Daniel Haber (Director of MGH Cancer Center, Harvard Professor) and Michael Lawrence (Bioinformatics leader w/ 70k+ citations and assistant professor at Harvard).

--> Product:

Watershed is a software platform that enables biologists to explore and analyze genomics data. The product consists of a Python API, cloud notebook and scalable compute cluster. These three core offerings come together to reduce the effort required to analyze sequencing data and eliminate the challenge of setting up a computational environment.



The current DIY genomics data analysis workflow requires knowledge in many domains including compute resources, python, bash, scripting and alignment APIs. Through our diligence calls we have confirmed that these are skill sets that the average biologist does not have which makes data processing a major bottleneck.

"Coming into the lab, I didn't know computer science fundamentals like how to set up a vm or write a bash script. I had to learn this all on the job. Honestly the first two years of grad school were spent just learning these skills so I could do what I needed to." - Computational Biologist PhD, Researcher at Google

Watershed provides biologists with basic python skills with a one-stop-shop to perform data science from alignment to visualization.

In terms of product priorities, the team is currently laser focused on enabling two use cases for beta users: bulk and single-cell RNA sequencing. For these two use cases there will be different onboarding flows and default notebooks. The team plans to maintain extremely tight feedback loops with beta users.

While Watershed is starting with the bioinformatics space, the long-term vision is to scale this platform to "omics" in general and eventually all scientific disciplines.

--> Diligence Conversations:

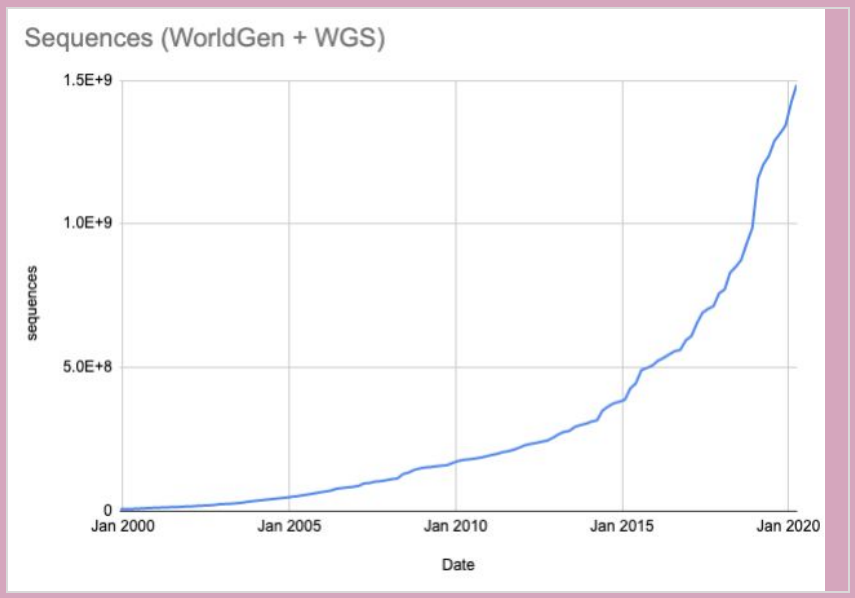
Our nine conversations with potential users and industry experts came back generally positive. One item of note that became crystal clear was the problem: data analysis for biologists is hard right now. Casey spoke with the healthcare team at Bessemer who agreed the problem space is real and hasn't been cracked.

Terri spoke with a computational biologist at Calico Life, founders of another SaaS product working in a similar space, and other industry experts who all agreed that the problem space is real. These conversations did, however, highlight that a software solution might not be the end-all be-all (is it better to just hire specialists? Do SaaS products introduce too much abstraction?), but noted that perhaps the right product hasn't hit the market yet.

Lauren spoke with a biophysics PhD candidate at UCSF who agreed that this problem hasn't been solved but had a neutral outlook on Watershed, as he felt there were many people in "-omics" who had strong computational skills and many open source toolkits that facilitate data analysis. He mentioned that Watershed's success would depend on finding a niche that isn't already served and which has an audience with little to no programming skills.

--> Market:

The amount of sequencing data has been growing exponentially over the past thirty years. The number of sequences published by the leading public repositories for DNA/RNA, WorldGen and WGS, can be seen in the adjacent chart.



In Jan 2019 there were roughly 770m public sequences and by 2020 there were over 1.15b, yielding a 50% YoY growth rate.

We believe that in previous decades the bottleneck in genomics discovery was the data itself but in this current decade, the tables have turned. The limitation is no longer data generation but rather data analysis.

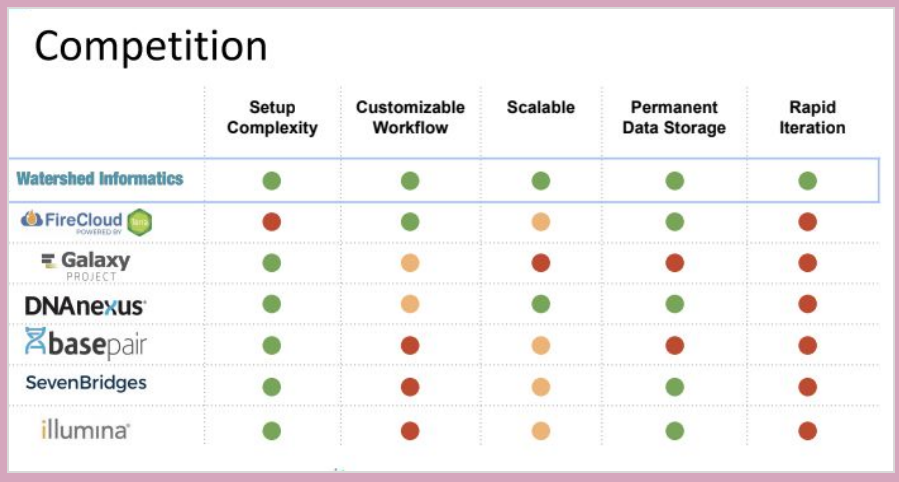
More specifically, within genomics, Watershed is going after three types of users: Academia, biotech and pharma.

+ Academia - ~$100M market as of 2020
+ Pharma - ~$400M market as of 2020
+ Biotech - ~$1.38B market as of 2020

--> Competitive Landscape:

Most competitors, including Seven Bridges Genomics, Illumina's Sequence Hub, and others, offer fixed data workflows. In other words, these are GUI-based solutions where users drag-and-drop components. The problem with this approach is the inherent rigidity. The reality is that scientific discovery is an iterative, dynamic and non-linear process, therefore, we believe fixed workflow tools will not suffice. Also, from an economics standpoint, fixed workflow tools are difficult to scale since every pipeline modification requires an engineering investment.

Unlike GUI-based solutions, **Watershed allows users to write raw code, which provides optimal flexibility and agility, while abstracting the cons of coding such as environment and compute management.**



A perfect analogy is creating a website. On one hand, a user can leverage a website builder like Squarespace. While this choice minimizes the initial setup time, once a user tries to exert some creativity, the limitations are painfully visible. Customization is incredibly difficult and there is only so much a user can do with Squarespace. On the other hand, a user can build a website from scratch. Setup costs are non-negligible, the experience is sub-optimal and the user is often pulled in many directions with having to set up hosting, domains, redirects, along with writing and testing the actual code. The glaring benefit of coding from scratch is the fact that you are given a blank canvas. Now, imagine Watershed being a new option which combines the best of both worlds: optimal flexibility with minimized overhead and room for error.

--> G2M, Business Plan and Pricing:

Watershed plans on targeting KOLs who influence heads of R&D at life sciences companies. The plan is to bring on a senior (potentially C-level) veteran salesperson with experience selling SaaS to biotech/pharma once the product is ready to be sold.

Watershed plans to charge for the underlying compute and storage. B2C customers will experience a metered, pay-as-you service. B2B customers will pay upfront on contract with support SLAs. As this is pre-product, Watershed is still wrapping their arms around how much value this provides and how much their users are willing to spend on it beyond the cost of the underlying data storage and compute.

--> **Risks:**

+ Technical execution is challenging
+ Crowded space with well-funded competitors
+ Founding team doesn’t include sales figure
+ Pharma market is incredibly hard to break into

--> **Recommendation:**

TLC to invest into Watershed!