



# BASEPAIR SAAS PLATFORM ORCHESTRATES NKARTA'S USE OF ITS OWN AWS RESOURCES TO DEMOCRATIZE GENOMIC DATA ANALYSIS

Learn how Basepair, a next-generation bioinformatics company, is reinventing bioinformatics on AWS through its federated self-service NGS analysis and visualization solution.

**50%**

reduction in compute costs

**50%**

improvement in bioinformatics response times for R&D projects

**25%**

time saved in routine analysis of wide variety of NGS data types

**3-5**

full-time engineer jobs saved versus do-it-yourself solution

**Facilitated**

scientists' ability to analyze & interactively visualize NGS data

## OVERVIEW

Basepair sought to address a critical bottleneck in drug research and development (R&D): how to accelerate making sense of the complex raw data that results from the sequencing of a genome. This process, called next-generation sequencing (NGS), typically requires bioinformaticians to analyze raw data to provide meaningful results for research scientists who are looking for potential treatments. "Our ability as an industry to generate NGS data has far surpassed our ability to make sense of it," says Simon Valentine, Basepair's chief commercial officer. "That is a software problem."

Basepair recognized an opportunity to create a cloud-native solution that would democratize not just access to but also analysis and interpretation of genomic data while keeping an organization's data secure within its own account, without the need to install inside a virtual private cloud. To address demanding storage and high-performance computing requirements, the company built the architecture for its [NGS Analysis and Visualization Software](#) on Amazon Web Services (AWS). Using AWS, Basepair developed a point-and-click solution that helps bench scientists do their own NGS analysis, cutting compute costs by 50 percent and improving support for R&D. "We're freeing bioinformaticians' time to work on more advanced data-mining tasks and accelerating time to scientific insights for researchers, streamlining collaboration between the teams," says Amit Sinha, Basepair's founder.

## OPPORTUNITY | USING AMAZON EC2 TO SIMPLIFY NGS ANALYSIS FOR BASEPAIR

As a research scientist at Harvard Medical School, Sinha saw firsthand how scientists sometimes had to wait months for bioinformaticians to complete their analyses of NGS data. “We had more samples year over year at a lower price, generating bigger files and taking on more collaborators, but the analysis was the bottleneck,” says Sinha. “The vision was to create software so that all scientists, from bioinformaticians to bench biologists, could do their own analyses and get results within minutes.”

Basepair helps its customers to securely deploy, scale, and run bioinformatics tools and pipelines in a cost-effective way that is HIPAA and General Data Protection Regulation compliant. The solution uses [Amazon Elastic Compute Cloud](#) (Amazon EC2), which provides secure and resizable compute capacity for virtually any workload. It efficiently processes huge datasets, where the size of a single sample ranges to several gigabytes. “Traditional commercial data sharing and analysis approaches involve the movement of sensitive data into centralized bioinformatics products, increasing security risk and cloud costs,” says Sinha. “It made sense to build a scalable cloud-native solution, and AWS was the obvious choice. Compute and storage capabilities no longer max out.”

Basepair’s intuitive interface lets bench scientists, who might lack programming expertise, run complex analyses simply. “They get results within minutes and can play with their own data: run quality control, slice and dice it, and look at it in different ways,” says Sinha. “AWS was leaps and bounds ahead of any other cloud providers in terms of resources to build on. We can continue to innovate and build features that our customers will use. If there were no AWS, there would be no Basepair.”

*“Traditional commercial data sharing and analysis approaches involve the movement of sensitive data into centralized bioinformatics products, increasing security risk and cloud costs.”*

**Amit Sinha**  
Founder, Basepair



## SOLUTION | CUTTING COMPUTE COSTS BY 50% WHILE ACCELERATING TIME TO MARKET FOR DRUG THERAPIES

Founded in 2015, the biotech company [Nkarta](#)—Nk from the abbreviation for natural killer (NK) cells—was looking for a user-friendly, scalable bioinformatics solution to accelerate the development of NK cell therapies for cancer patients. “As a startup, it did not have enough bandwidth to adequately manage all its compute resources,” says Valentine. “This meant that it wasn’t efficiently spinning down or optimizing instances.”

Nkarta adopted Basepair’s NGS Analysis and Visualization Software and off-loaded routine DevOps tasks, such as compute optimization, job handling, and access permissions. Instead of a multiday training course, Nkarta bench scientists needed 1–2 hours of training to interpret results using Basepair’s built-in interactive visualizations and reports. “We have freed about 25 percent of the time for our small team to focus on more valuable data-mining tasks,” says Sombeet Sahu, Nkarta’s associate director of bioinformatics. “What’s more, bench scientists now come to my team with an informed question rather than a request to work on small, repetitive tasks.”

Nkarta has increased operational resilience, cutting internal response time by 50 percent and accelerating time to market for therapies. Additionally, Nkarta saves the equivalent of three to five full-time engineers who otherwise would have to configure, connect, and maintain the technology behind the pipeline. “The resources can be better allocated for scientific innovation and other core business areas, like exploring the power of artificial intelligence and machine learning for R&D,” says Valentine.

Basepair has helped Nkarta cut compute costs in half by automatically selecting the right instances to improve performance. The platform uses its own [Amazon EC2 Spot Instances](#), which take advantage of unused Amazon EC2 capacity on AWS. For storage, Basepair uses Nkarta’s instances of [Amazon Simple Storage Service](#) (Amazon S3), an object storage service offering industry-leading scalability, data availability, security, and performance. Basepair estimates that the automated data archival and dynamic retrieval capabilities could cut storage costs by up to 80 percent.

For added data privacy and security, Nkarta connects to Basepair’s compute and storage resources from its own AWS account. “Organizations can keep all the data in their own environments and use Basepair to automate the management of the building blocks,” says Valentine. “It opens new horizons to use NGS for discovery and development of therapies. We accelerate time to scientific insight for researchers and diagnostic insights for patients while eliminating the need for data movement.”

## OUTCOME | PROVIDING AN AUTOMATED SOLUTION TO DEMOCRATIZE THE USE OF GENOMIC DATA

Basepair has started to incorporate [AWS HealthOmics](#), which helps healthcare and life science organizations build at scale to store, query, and analyze genomic, transcriptomic, and other omics data. Basepair plans to use Amazon Omics for even more cost-effective storage of genomic data and to help customers perform cohort-level clinical-genomics-based queries across patient datasets.

Basepair became an AWS Partner in May 2023 after completing the [AWS Foundational Technical Review](#), which helps organizations identify and remediate risks in software or solutions. Basepair's NGS Analysis and Visualization Software is available in [AWS Marketplace](#), a curated digital catalog that simplifies procurement, provisioning, and governance of third-party software. "We want to democratize the use of NGS by helping more people analyze the data without replacing bioinformaticians' valuable work," says Valentine. "When we combine improved cross-team collaboration with effective use of the power of AWS, we will definitely accelerate the discovery process."

## **ABOUT BASEPAIR**

Founded in 2017 by a Harvard Medical School researcher, Basepair is a market leading point & click bioinformatics company focused on unlocking the full potential of next generation sequencing (NGS) technology. By empowering more scientists to understand the meaning and value of genomic data while keeping it secure in-place, we help accelerate the scientific and diagnostic breakthroughs that bring lifesaving medicines and treatments to the patients that need them.

## **AWS SERVICES USED**

### **Amazon S3**

Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance.

[Learn more »](#)

### **Amazon EC2**

Amazon Elastic Compute Cloud (Amazon EC2) offers the broadest and deepest compute platform, to help you best match the needs of your workload.

[Learn more »](#)

### **AWS HealthOmics**

AWS HealthOmics helps healthcare and life science organizations build at-scale to store, query, and analyze genomic, transcriptomic, and other omics data.

[Learn more »](#)

### **AWS Foundational Technical Review**

The AWS Foundational Technical Review (FTR) enables you to identify and remediate risks in your software or solutions.

[Learn more »](#)



Basepair. NGS insight.



+1 (347) 428-9399



[www.twitter.com/basepairtech](https://www.twitter.com/basepairtech)



New York, NY



[www.linkedin.com/company/basepair](https://www.linkedin.com/company/basepair)



[sales@basepairtech.com](mailto:sales@basepairtech.com)



[www.basepairtech.com](https://www.basepairtech.com)

