



Внедрение и валидация информационных систем на базе программно-аппаратного обеспечения OnSiteSeq.

Методология ALCOA+

История изменений

Версия	Дата изменения	Описание изменения	Автор изменения
0001	01.04.2026	Создание первой версии инструкции	Горбенко Р.

Введение

ALCOA+ — это набор принципов целостности данных (Data Integrity), изначально разработанный для фармацевтической и медицинской отраслей. Он гарантирует, что данные надежны, достоверны и могут быть проверены на любом этапе их жизненного цикла.

1. А – Attributable (Атрибутируемость)

А – Attributable (Атрибутируемость)

Кто провел анализ и какой именно алгоритм выдал результат?

- **Авторизация на устройстве:** При работе через сенсорный экран медицинский работник должен авторизоваться), чтобы в логах было зафиксировано, кто именно запустил процессы «Запуск амплификации» и «Запуск секвенирования». Кто именно выполнил “Исследование”, генерацию и печать отчёта.
- **Версионирование пайплайнов/инструментов/моделей:** В биоинформатических отчетах OnSiteSeq заложена практика указания версий как для собственных пайплайнов/инструментов/моделей (например, onsiteseq_hiv v1.0.0, HIV-1-M-Env-Rus v1.0.0) так и инструментах в зависимостях.



Методология ALCOA+

Это строго соответствует принципу атрибутируемости, в том числе и алгоритмов машинного обучения.

2. L – Legible (Читаемость)

Понятны ли данные врачу и аудитору спустя годы?

- **Двойная отчетность:** В OnSiteSeq отчеты разделены на «Клинический отчет» для врача-клинициста и «Биоинформатический отчет» (QC Report) для специалистов.
- **Физические и цифровые копии:** Наличие встроенного термопринтера обеспечивает мгновенную читаемость на месте. Цифровые PDF-копии отчетов сохраняются на SSD M2 диске с понятной номенклатурой файлов (например, включающей в себя метку времени и ID образца).

. C – Contemporaneous (Своевременность)

Фиксируются ли данные в момент их появления?

- **Таймстемпы:** Все этапы (от base-calling-а на GPU процессоре Nvidia до вывода отчета на экран) логируются в реальном времени.
- **Синхронизация времени:** Поскольку комплекс OnSiteSeq Edge автономен и может работать "в поле", важно, что внутренние часы комплекса не могут быть изменены пользователем вручную. Для синхронизации точного времени используется коммуникационный блок Wi-Fi/4G. Синхронизация отрабатывает сразу же, как OnSiteSeq Edge появляется в сети.

4. O – Original (Оригинальность)

Сохраняются ли первичные «сырые» данные?

- **Хранение сигналов:** В OnSiteSeq Edge до того как ML-модель Dorado (base-caller) переведет электрические сигналы с нанопоры в нуклеотиды, система сохраняет первичные сырые данные (форматы POD5/FAST5) на SSD-диск.
- **Неизменяемость:** Исходные риды (k-меры) никогда не перезаписываются в процессе сборки генома и анализа.

5. A – Accurate (Точность)

Являются ли данные достоверными и есть ли контроль качества?



OnSiteSeq.io

Методология ALCOA+

- **Метрики QS:** Биоинформатический отчет содержит важнейшие метрики: % покрытия генома (например, 99.2347%), среднюю глубину (Depth), слепые зоны. Это прямое доказательство точности.
- **Уверенность ИИ:** Вывод в клинический отчет уверенности нейросети в процентах (например, уверенность 99.98% для субтипа В или Hazard Score для *M. tuberculosis*) позволяет врачу оценить надежность предсказания.

Плюс (+)

- **Complete (Полнота):** Snakemake GUI должен обеспечивать сохранение логов даже прерванных или неудачных запусков (например, статус "FAIL", если покрытие < 80%).
- **Consistent (Последовательность):** Жёсткий workflow: пайплайн анализа не может стартовать, пока аппаратно не завершится этап амплификации и секвенирования.