

Design von Primern für viral-zelluläre-Übergänge basierend auf NGS-Shortreads

Exposé zur Bachelorarbeit

Daniel Hartmann

Humboldt Universität zu Berlin

daniel.hartmann@informatik.hu-berlin.de

Betreuer: Professor Dr. Ulf Leser, Dr. Karsten R. Heidkte

21. März 2017

1 Einführung

In Zusammenarbeit zwischen der Frauenklinik der Universität Jena und der ATLAS Biolabs läuft derzeit die Validierung eines Next-Generation-Sequencing-Verfahrens sowie eine begleitende klinische Studie an Patientinnen mit einer vorliegenden HPV16/HPV18-Infektion, zur Untersuchung von Übergangsstellen des HPV Virus in die menschliche DNA.

2 Motivation

Zu Grunde liegt die Annahme, dass der Bruchpunkt des Insekrats bei jeder Infektion mit dem Virus einzigartig ist. Geht aus einer Infektion ein Tumor hervor, so wird dieser einen einheitlichen, beschreibenden Bruchpunkt aufweisen. Ziel ist es die Bruchpunkte aller vorliegenden HPV Insekrats in Gewebeprouben des Tumors und Zellabstrichen der Patientinnen detektieren zu können. Somit wäre es möglich Rezidive von Neuinfektionen zu unterscheiden, da Neuinfektionen nicht den bekannten Bruchpunkt des Tumors aufweisen werden.

Als Grundlage dienen Daten, welche mithilfe von Illumina / Nextera Sequenzierung erhoben und anschließend durch den am Deutschen Krebsforschungszentrum, DKFZ, entwickelten TEN16-Algorithmus prozessiert wurden. Das Ten16 Verfahren mappt die Daten auf ein Referenzgenom und filtert mögliche Kandidaten für Insekratsstellen heraus. Die Validierung dieser Bruchpunkte erfolgt anhand einer PCR, indem für die virale und die zelluläre Seite ein Primer bestimmt wird.

3 Ziele der Arbeit

Im Rahmen der Arbeit soll ein Verfahren entwickelt werden, welches für die gegebenen Daten Primer im Bereich der Bruchpunkte bestimmt, damit in der Nachsorge gezielt diese Bereiche in der Patienten DNA auf Inerate geprüft werden können.

4 Lösungsweg

Hierzu werden zunächst die zu Grunde liegenden Verfahren erläutern und die Studie im Detail erklärt. Das Verfahren selbst wird grob in drei Schritte unterteilt sein.

1. Auswahl der Reads: Unter den vorliegenden Daten stehen verschiedene Quellen zur Verfügung, welche Reads zum Design der Primer liefern können. (Rohdaten, verschiedene Filterstufen der Short-Reads des TEN16-Algorithmus). Hier gilt es herauszufinden welche Short-Reads den Bruchpunkt unterstützen bzw. beinhalten.
2. Design der Primer: Anhand der gewählten Reads werden Sequenz-Templates bestimmt, welche dann an das Tool Primer3 übergeben werden, um Primer im gewünschten Bereich zu wählen. Hierzu müssen diverse Parameter berücksichtigt werden, damit im Ergebnis Abstand, Ausrichtung und ähnliche Eigenschaften der Primer wie gewünscht vorliegen.
3. Aufbereitung der Ergebnisse: Zur weiteren Nutzung und Verarbeitung der Daten, wird ein Tool entwickelt, das die Ergebnisse visualisiert und mit Primern, welche bereits mittels PCR und anschließender Sanger-Sequenzierung validiert wurden, darstellt. Außerdem wird eine Test-Suite implementiert, um die entwickelten Funktionalitäten zu Prüfen und Fehler detektieren zu können

5 Verwandte Arbeiten

Bisherige Untersuchungen von Zervixkarzinomen haben ergeben, dass in mehr als 90 % aller Fälle eine Integration des Virus in die Humane DNA Vorliegt. In [3] wird beschrieben, dass bei HPV16 induzierten Karzinomen ca. 80% eine Integration aufweisen. Bei HPV18 sind es sogar 100%. In [4] wird dargelegt, dass diese Information genutzt werden kann um chirurgische Eingriffe bei frühen Stadien einer Infektion zu reduzieren. Bisher werden alle "high-grade-Läsionen per Conziation entfernt. Erkennt man Läsionen in denen das Virus frei vorliegt und nicht integriert ist, geht die Infektion in den meisten Fällen selbstständig zurück. Hier reicht eine regelmäßige Untersuchung der Wunde.

Auf Algorithmischer Ebene wurde von Yunxin Chen et al. die Sowftware VirusSeq entwickelt, welche Integrate bekannter Viren in Humaner DNA detektiert [5]. Hierzu wurde aus bekannten Virus-Sequenzen und der Humangenomsequenz hg19 ein Referenzgenom erstellt. Auf dieses Referenzgenom werden NGS paired-end Reads gemappt. Dann filtert VirusSeq die Read Pairs heraus, welche Junctions zwischen Humaner und Viraler DNA zeigen. Tambunan et al haben in [6] Algorithmische Verfahren genutzt, um Epitope verschiedener HPV-Arten zu finden. Anhand dieser Erkenntnisse soll die Entwicklung von Impfstoffen auf Basis der Epitope vorangetrieben werden.

6 Referenzen

- [1] XU, Bo, et al. Multiplex identification of human papillomavirus 16 DNA integration sites in cervical carcinomas. *PloS one*, 2013, 8. Jg., Nr. 6, S. e66693.
- [2] Li, Heng, et al. The sequence alignment/map format and SAMtools.” *Bioinformatics* 25.16 (2009): 2078-2079.
- [3] PETT, M.; COLEMAN, N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?. *The Journal of pathology*, 2007, 212. Jg., Nr. 4, S. 356-367.
- [4] LUFT, Frank, et al. Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. *International journal of cancer*, 2001, 92. Jg., Nr. 1, S. 9-17.
- [5] CHEN, Yunxin, et al. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics*, 2013, 29. Jg., Nr. 2, S. 266-267.
- [6] TAMBUNAN, Usman Sumo Friend; PARIKESIT, Arli Aditya. HPV bioinformatics: in silico detection, drug design and prevention agent development. INTECH Open Access Publisher, 2012.