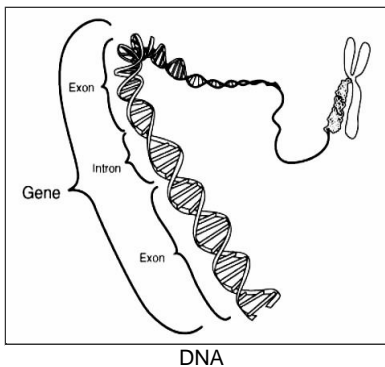


Gene prediction with a combination of generalized hidden Markov model and support vector machine

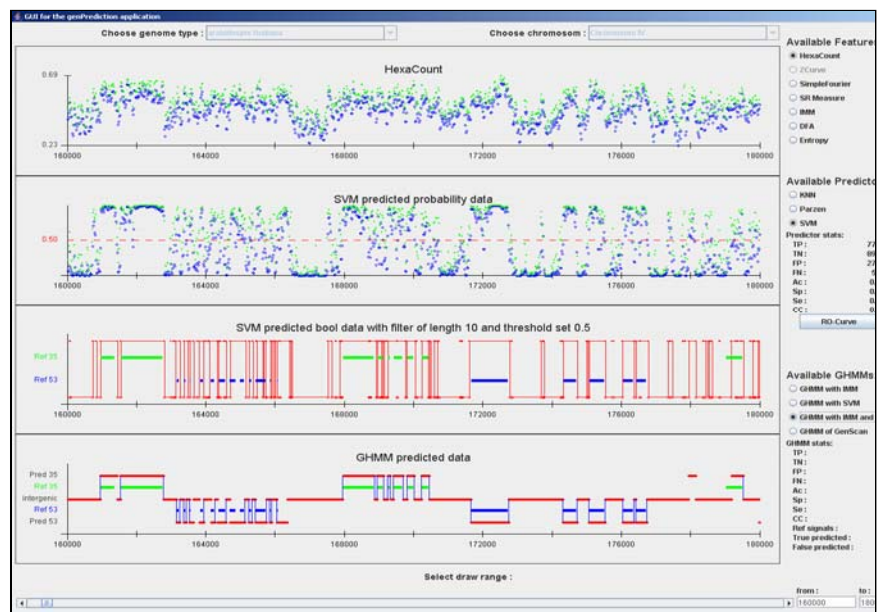
Diplomand / in	Pascal Frei, Philipp Beck
Examinator / in	Prof Dr Guido Schuster
Experte / in	Gabriel Sidler, Eivycor GmbH, Zürich
Themengebiet	Digitale Signalverarbeitung
Projektpartner	Prof. Dr. Tom Ryen University of Stavanger, Stavanger, Norwegen



Aufgabenstellung: Obwohl das menschliche Genom 2003 vollständig sequenziert wurde, heisst das jedoch nicht, dass es damit entschlüsselt ist. Denn viele Sequenzen sind nicht von Bedeutung, so dass die Forscher noch immer auf der Suche nach den codierenden Sequenzen sind. Unsere Aufgabe war nun, mit Hilfe der Support Vektor Maschine (SVM), im Genom diese codierenden Sequenzen zu finden. Auch sollten wir verifizieren ob mithilfe der Kombination von bekanntem eine Verbesserung erreicht wird.

Ziel der Arbeit: Eine Applikation, welche verschiedene bereits bekannte Verfahren zur Erkennung von coding bzw. non coding Sequenzen kombiniert und dadurch eine bessere Genauigkeit erreicht. Diese Applikation sollte zudem einfach zu bedienen, gut dokumentiert und komplett in schnellem C++ Code geschrieben sein.

Lösung: Die Applikation kombiniert sieben bekannte Verfahren zur Erkennung von coding und non coding Abschnitten in einem Genom. Sie kann leicht mit einem weiteren Verfahren erweitert werden. Der Server Teil der Applikation ist komplett in C++ geschrieben, das GUI in Java. In der obersten Zeile des GUI kann eines der bekannten Verfahren ausgewählt werden. Die zweite Zeile zeigt die Wahrscheinlichkeitsausgabe der SVM basierend auf allen bekannten Verfahren. Auf der dritten wird diese mit einem Filter nachbearbeitet und eine Entscheidungsschwelle gesetzt. Um dies noch weiter zu verbessern haben wir der Applikation noch den „generalized hidden Markov model“ Algorithmus hinzugefügt. Dieser bezieht die spezielle Grammatik der Gensequenzen in den Algorithmus ein, die Genvorhersage wird dadurch noch genauer.



Das GUI der Applikation