

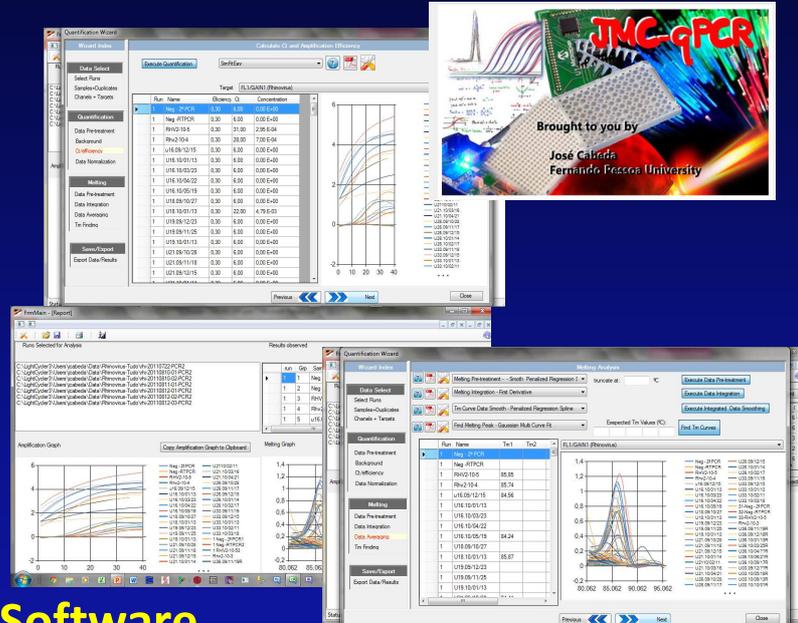
JMCqPCR: Desenvolvimento de Software modular e algoritmos otimizados para análise de qPCR

José Cabeda

Grupo de Imunologia, Centro de Estudos em Biomedicina (CEBIMED),
Faculdade de Ciências da Saúde, Universidade Fernando Pessoa

Resumo

No decurso da análise de dados de qPCR do nosso projecto para o estudo da epidemiologia molecular do rinovirus deparamo-nos com sérias limitações no software disponível: A incapacidade para eliminar o ruído introduzido pelas oscilações eléctricas; a incapacidade para estimar a eficiência das reacções, e a sua diminuição ao longo da reacção; a incapacidade para determinar automaticamente os limites do "background", identificando as amostras que o não possuem em reacções de Nested-PCR; a incapacidade para tabelar e realizar validação cruzada entre os dados de amplificação e os de análise de temperatura de melting. A capacidade para estimar a eficiência das reacções foi já objecto de análise na literatura especializada (1-3), mas os algoritmos descritos encontram-se implementados em plataformas especializadas, não disponíveis na generalidade dos laboratórios de Biologia Molecular. Assim, decidimos criar uma peça de software capaz de ultrapassar todas as limitações descritas, com estrutura modular, por forma a permitir a adição posterior de módulos com novos algoritmos para realizar cada uma das fases da análise de dados em qPCR.



O Software

O fluxo de processamento de dados de qPCR foi dividido em várias fases, cada uma das quais foi implementado sob a forma de módulos que podem ser adicionados ao software, e para os quais o utilizador pode (em tempo de execução) escolher o módulo a utilizar:

- Leitura de dados originais (neste momento encontra-se implementada apenas a leitura de dados de Lightcycler (Software versão 3.5). Para além da fase da leitura de dados, foi ainda implementada uma fase de sistematização dos dados lidos, que permite suplantir omissões nas características das amostras e canais de fluorescência, como a indicação de que amostras são controlos e padrões, que amostras devem ser analisadas, que amostras são duplicados de outras, etc.
- Pré-processamento de dados de fluorescência. Estes módulos permitem fazer ajustes matemáticos aos dados lidos pelo equipamento, com o intuito de eliminar artefactos criados por exemplo por oscilações eléctricas, ou até aumentar a resolução potencial dos dados realizando interpolações.
- Deteção dos intervalos da linha de base. A este respeito foi desenvolvido um novo algoritmo que permite com maior eficácia que os algoritmos já descritos encontrar com precisão os limites da linha de base, identificando os casos em que não se pode realizar correcção a esta linha de base, já que a amplificação se inicia no primeiro ciclo (por exemplo em casos de Nested-PCR)
- Correcção da linha de base: Para esta etapa foram desenvolvidos dois módulos correctores (um baseado na subtracção da melhor recta que descreve a linha de base obtida por regressão linear, e outro baseado no simples deslocamento de toda a curva por subtracção do valor da fluorescência no ciclo 1),

- Quantificação: Foram implementados métodos descritos por outros em que a quantificação ajusta por métodos iterativos (Levenberg-Marquardt e método de Newton) os dados a uma equação que descreve a amplificação e que entra em linha de conta com a eficiência da reacção (SimFITEAV e Dart-PCR; ref. 2-3) ou determinações qualitativas e quantitativas utilizando amostras referência (MaxRatio; ref. 4).
- Normalização de dados quantitativos: Estes módulos (ainda não desenvolvidos) procuram obter quantificações de um gene alvo relativamente à quantidade existente de um ou mais genes de referência
- Pré-tratamento de dados de melting: idêntico ao descrito para os dados de amplificação)
- Derivação dos dados de melting: Tratamento de dados para obtenção das curvas de melting tradicionais (negativo da derivada dos dados originais)
- Smoothing das curvas de Melting: "Penalized Regression Spline" e "Averaging"
- Identificação das curvas de melting: Foi desenvolvido um algoritmo para identificar que curvas apresentam o Tm esperado. O algoritmo ajusta os dados a uma curva de Gauss com Média igual ao Tm esperado, e avalia os valores da Média, Intensidade e largura da curva obtida para aceitar ou rejeitar o ajuste como válido.
- Exportação dos dados analisados. Com estes módulos pretende-se garantir a compatibilidade do software com outros sistemas de análise de dados de PCR, bem como com padrões emergentes como o XXXXX.
- Emissão de Relatório com os resultados, indicando os módulos utilizados. Os dados e gráficos produzidos são ainda exportáveis para o Clipboard, facilitando a sua utilização directa em outras peças de software e utilização para publicações.

Material e Métodos

O Software foi desenvolvido em Visual C# utilizando a plataforma Microsoft Visual Studio Express.

Para o desenvolvimento do programa foram utilizadas bibliotecas de cálculo desenvolvidas por outros ("Alglib" e "PZMath") e disponibilizadas de acesso livre ao abrigo de licenças GNU. O objectivo será que a versão final do programa utilize apenas uma destas bibliotecas.

O Software foi desenvolvido utilizando a língua Inglesa na interacção com o utilizador, devido ao maior potencial de impacto na comunidade de utilizadores desta língua, mas está prevista a existência de módulos de tradução para outras línguas, em particular para a Portuguesa.

Bibliografia

- Rebrikov D.V. & Trofimov D.Y. Applied Biochemistry and Microbiology 42 (2006): 455-463
- Peirson S.N. et al., Nucleic Acids Research 31 (2003): e73
- Batsch A., et al. BMC Bioinformatics 9 (2008): 1-11
- Shain E.B. et al., Nucleic Acids Research 36 (2008): e91